บทที่ 1

บทนำ

1.1 <u>ความเป็นมาและความสำคัญของปัญหา</u>

เมื่อเราต้องการทราบคำตอบเกี่ยวกับเรื่องใดเรื่องหนึ่ง เราอาจทำการทดลองเกี่ยวกับเรื่องนั้น เพื่อสังเกตผลลัพธ์ที่ได้ การทดลอง เป็นกระบวนการใด ๆ ที่สามารถกระทำซ้ำได้ภายใต้สภาวการณ์เดียวกัน ผลลัพธ์ของการทดลองที่ถูกบันทึกไว้ เรียกว่า ข้อมูลดิบ การทดลองใด ๆ ที่ถ้าทำซ้ำกันหลาย ๆ ครั้ง โดยใครก็ได้ ซึ่งทำการทดลองภายใต้สภาวการณ์เดียวกัน แล้วให้ผลการทดลองเหมือนกันทุกครั้ง เช่น การทิ้งก้อนวัตถุจากที่สูง การนำก้อนน้ำแข็งมาวางทิ้งไว้ในห้องที่มีอุณหภูมิปกติ การนำแท่งแม่เหล็กสองแท่งมาชนกันโดยหันขั้วเหนือของทั้งสองแท่งเข้าหากัน การกระทำเหล่านี้ไม่ว่าจะให้ใครทำหรือทำซ้ำหลาย ๆ ครั้ง ก็จะให้ผลตรงกัน เรียกการทดลองประเภทนี้ว่า การทดลองที่ทราบผลแน่นอน ส่วนการทดลองอีกประเภท ผลลัพธ์ของการทดลองที่ได้จากการทำการทดลองซ้ำ ๆ กัน อาจแตกต่างกันไม่เหมือนกันทุกครั้ง เช่น การโยนลูกเต๋าแล้วสังเกตแต้มที่หงายขึ้น การเลือกไพ่หนึ่งใบออกจากสำรับที่มีไพ่ครบทุกใบ เรียกการทดลองประเภทนี้ว่า การทดลองเชิงสุ่ม

การทดลองเชิงสุ่มใด ๆ ที่มีผลลัพธ์เป็นไปได้เพียงสองทาง เรียกว่าการทดลองเบอร์นูลี ปกติมักเรียกผลลัพธ์ของการทดลองเบอร์นูลีอย่างหนึ่งว่า ความสำเร็จ และเรียกผลลัพธ์อีกอย่างว่า ความไม่สำเร็จ หรือ ความล้มเหลว

เนื่องจากผลลัพธ์ดังกล่าวเป็นการบอกเหตุการณ์เชิงคุณภาพ การวิเคราะห์ทางสถิติ จำเป็นต้องเปลี่ยนเหตุการณ์เชิงคุณภาพให้เป็นตัวเลข โดยมากนิยมกำหนดให้เลข 1 แทนเหตุการณ์ที่เป็นความสำเร็จ และเลข 0 แทนเหตุการณ์ที่เป็นความไม่สำเร็จ จากการกำหนดในลักษณะดังกล่าวจะได้ ตัวแปรสุ่มเบอร์นูลี

ตัวแปรสุ่มเบอร์นูลี คือ ตัวแปรสุ่มที่มีค่าได้เพียง 2 ค่า คือ 0 และ 1 ค่าความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าเป็น 1 มีค่าเท่ากับ p และค่าความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าเป็น 0 เท่ากับ $1 - p$

โดยทั่วไปเมื่อกล่าวถึงการแจกแจงของประชากร จะหมายถึงการแจกแจงของตัวแปรสุ่มที่ใช้นิยามประชากรนั้น การแจกแจงทวินาม เป็นการแจกแจงแบบไม่ต่อเนื่องชนิดหนึ่งที่พบกันมากในทางปฏิบัติ เป็นที่ทราบกันดีว่าจากการทดลองเบอร์นูลี ซึ่งมีความน่าจะเป็นของความสำเร็จในการทดลองแต่ละครั้งเท่ากับ p ถ้าทำการทดลองซ้ำ ๆ กัน n ครั้ง โดยที่การทดลองแต่ละครั้งจะต้องเป็นอิสระจากกัน และค่าความน่าจะเป็นของความสำเร็จในการทดลองจะต้องไม่เปลี่ยนแปลง (คงที่เท่ากันทุกครั้ง) แล้วจะได้ว่า ผลรวมของความสำเร็จที่ได้จากการทดลองเบอร์นูลี n ครั้งดังกล่าว จะมีการแจกแจงทวินาม ที่มีค่าพารามิเตอร์เท่ากับ (n,p) เมื่อ n คือจำนวนครั้งของการทดลองเบอร์นูลี และ p เป็นค่าสัดส่วนของความสำเร็จในประชากร เรียกตัวแปรสุ่มที่กำหนดจากผลรวมของความสำเร็จดังกล่าวว่า ตัวแปรสุ่มทวินาม ค่าที่เป็นไปได้ของตัวแปรสุ่มทวินาม จะมีได้ทั้งสิ้น $n + 1$ ค่า โดยมีค่าเป็นเลขจำนวนเต็มตั้งแต่ 0 ถึง n และกล่าวว่า จำนวนค่าสังเกตทั้งหมดของตัวแปรสุ่มทวินาม ซึ่งอาจมีจำนวนถึงค่าอนันต์ได้ คือ ประชากรที่มีการแจกแจงทวินาม

การอนุมานเชิงสถิติ เป็นเรื่องของการประยุกต์ใช้ทฤษฎีความน่าจะเป็นในการวิเคราะห์ข้อมูลที่ได้จากตัวอย่าง เพื่อใช้บรรยายหรือสรุปลักษณะบางประการของประชากร สำหรับในประชากรที่มีการแจกแจงทวินาม ประเด็นความสนใจของนักวิจัยจะมุ่งไปที่การประมาณค่าพารามิเตอร์ p ซึ่งเป็นค่าสัดส่วนของความสำเร็จในประชากร

ในการประมาณค่าพารามิเตอร์ของประชากรโดยใช้ข้อมูลจากตัวอย่าง สามารถทำได้โดยการหาตัวประมาณซึ่งเป็นฟังก์ชันของตัวอย่าง เมื่อแทนค่าของค่าสังเกตที่ได้จากตัวอย่างลงไปในตัวประมาณ จะได้ค่าประมาณของพารามิเตอร์ การประมาณค่าพารามิเตอร์ ที่ใช้กันในการอนุมานเชิงสถิติ แบ่งออกเป็น 2 ชนิด คือ การประมาณค่าแบบจุด และการประมาณค่าแบบช่วง ทฤษฎีหรือตัวสูตรต่าง ๆ ที่เกี่ยวกับการประมาณค่าสัดส่วนของความสำเร็จในประชากรที่มีการแจกแจงทวินาม ทั้งแบบจุด และ แบบช่วง สามารถค้นคว้าได้จากหนังสือสถิติในทุกระดับ

จากการรวบรวมค้นคว้าของผู้วิจัย พบว่า ตำราภาษาไทย และ ตำราภาษาอังกฤษ โดยทั่วไปกล่าวถึงการประมาณค่าแบบช่วงสำหรับค่าพารามิเตอร์ p ไว้เฉพาะกรณีที่ค่าขนาดตัวอย่าง n มีค่ามาก เพื่อที่จะได้ทำการอนุมานผ่านทฤษฎีค่าจำกัดสู่ส่วนกลาง โดยได้หลีกเลี่ยงที่จะกล่าวถึงการประมาณค่าแบบช่วง ในกรณีที่ขนาดตัวอย่าง n มีค่าน้อย ที่มีกล่าวถึงบ้าง ก็เป็นการกล่าวถึงวิธีที่ใช้การตรวจค่า จากตารางการแจกแจงแบบทวินาม ซึ่งมีข้อจำกัดที่ความละเอียดของค่า p ในตารางมีไม่พอ ทำให้หาช่วงความเชื่อมั่นที่มีค่าสัมประสิทธิ์ความเชื่อมั่นเท่ากับค่าที่ต้องการไม่ได้

ด้วยเหตุผลดังกล่าว ยังคงมีผู้ใช้ทฤษฎีค่าจำกัดสู่ส่วนกลางในการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ p ในกรณีที่ขนาดตัวอย่างมีค่าไม่มาก เพื่อให้ได้ช่วงความเชื่อมั่นที่มีค่าสัมประสิทธิ์ความเชื่อมั่นเท่ากับที่ต้องการ ซึ่งโดยมากจะเป็นค่าที่นิยมใช้กันในหมู่นักวิจัย ทั้ง ๆ ที่ในความเป็นจริงแล้ว ช่วงที่หาได้จากวิธีดังกล่าว ไม่ได้มีคุณสมบัติตามที่คาดหวัง เพราะว่าขนาดตัวอย่าง ไม่สอดคล้องกับข้อตกลงของทฤษฎี

สำหรับในการวิจัยครั้งนี้ ผู้วิจัยสนใจการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ p ของประชากรที่มีการแจกแจงทวินาม ซึ่งนิยามจากตัวแปรสุ่มทวินามที่มีค่าพารามิเตอร์เท่ากับ (n,p) เมื่อขนาดตัวอย่าง n มีค่าเล็ก

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์หลัก 2 ประการ ดังนี้

1.2.1 ทำตารางสำเร็จรูป แสดงช่วงความเชื่อมั่นสำหรับค่าพารามิเตอร์ p ของประชากรที่มีการแจกแจงทวินาม ในกรณีที่ขนาดตัวอย่าง n มีค่าตั้งแต่ 1 ถึง 30

1.2.2 เพื่อศึกษา ผลที่เกิดขึ้นจากการใช้สูตรการประมาณโดยใช้การแจกแจงปกติ ในการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ p ของประชากรที่มีการแจกแจงทวินาม เมื่อขนาดตัวอย่าง n เล็ก (n = 1 ถึง 30) โดยจะพิจารณาถึงความเหมาะสมของการใช้สูตรการประมาณโดยใช้การแจกแจงปกติ ภายใต้เงื่อนไขที่กำหนด

1.3 ข้อตกลงเบื้องต้น

1.3.1 สำหรับการแจกแจงทวินามที่มีพารามิเตอร์ (n,p) ในที่นี้จะถือว่า ขนาดตัวอย่าง n เป็นพารามิเตอร์ที่ทราบค่า ส่วนค่าสัดส่วนความสำเร็จในประชากร p เป็นพารามิเตอร์ที่ไม่ทราบค่า ดังนั้นเป้าหมายของการอนุมานเชิงสถิติ สำหรับการวิจัยครั้งนี้ จะเป็นการอนุมานเกี่ยวกับค่า p เท่านั้น

1.3.2 การวิจัยครั้งนี้ถือว่า ค่าสัมประสิทธิ์ความเชื่อมั่นเป็นเกณฑ์ในการตัดสินความเชื่อถือได้ของช่วงความเชื่อมั่นที่สร้างขึ้น

1.4 สมมติฐานของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ทำการตั้งสมมติฐานไว้ 5 ข้อ คือ

1.4.1 ช่วงความเชื่อมั่นที่สร้างขึ้นจากวิธีการประมาณโดยใช้การแจกแจงปกติ จะมีระดับความเชื่อมั่นต่ำกว่า ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

1.4.2 ช่วงความเชื่อมั่นที่สร้างขึ้นจากสูตรที่แท้จริงซึ่งสัมพันธ์กับการแจกแจงเอฟ จะมีระดับความเชื่อมั่นไม่ต่ำกว่า ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

1.4.3 เมื่อค่าผลคูณ np และ n(1-p) มีค่ามากกว่า 5 ทั้งคู่ และค่า p มีค่าอยู่ในช่วง [0.40,0.60] จะสามารถใช้สูตรการประมาณโดยใช้การแจกแจงปกติได้

1.4.4 เมื่อค่าผลคูณ np และ n(1-p) มีค่ามากกว่า 5 ทั้งคู่ และค่า p มีค่าอยู่ในช่วง [0.25,0.40) U (0.60,0.75] สามารถอนุโลมใช้สูตรการประมาณโดยใช้การแจกแจงปกติได้

1.4.5 เมื่อค่าผลคูณ np และ n(1-p) มีค่ามากกว่า 5 ทั้งคู่ และค่า p มีค่าอยู่ในช่วง (0.00,0.25) U (0.75,1.00) การใช้สูตรการประมาณโดยใช้การแจกแจงปกติ ไม่มีความเหมาะสม

1.5 ขอบเขตของการวิจัย

สำหรับการวิจัยครั้งนี้ จะทำการเปรียบเทียบ ช่วงความเชื่อมั่นที่สร้างขึ้นจากวิธีการประมาณโดยใช้การแจกแจงปกติ กับช่วงความเชื่อมั่นที่สร้างขึ้นจากสูตรที่แท้จริงซึ่งสัมพันธ์กับการแจกแจงเอฟ เท่านั้น

1.5.1 ขนาดตัวอย่างที่ใช้ศึกษา กำหนดให้มีขนาดเท่ากับ 1 ถึง 30 ซึ่งในทางสถิติ จัดว่าเป็นตัวอย่างขนาดเล็ก

1.5.2 เนื่องจากค่าระดับความเชื่อมั่น จะแตกต่างกันไปตามค่าของพารามิเตอร์ p จึงได้กำหนดค่า p ให้มีค่าตั้งแต่ 0.05 โดยทำการเพิ่มค่าขึ้นทีละ 0.05 รวมเป็นค่า p ทั้งสิ้น 19 ค่า

1.5.3 ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้มาจากการจำลองขึ้นในเครื่องมินิคอมพิวเตอร์รุ่น Micro VAX3600 ที่ศูนย์คอมพิวเตอร์ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย โดยทำการสร้างข้อมูลให้มีสภาพตามข้อ 1.5.1 และ 1.5.2 การจำลองได้ใช้โปรแกรมที่เขียนขึ้นด้วยภาษา FORTRAN 77

1.5.4 ใช้ค่าสัมประสิทธิ์ความเชื่อมั่น 8 ระดับ คือ 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.98 และ 0.99 สำหรับขนาดตัวอย่าง n หนึ่งค่า จะมีช่วงความเชื่อมั่นที่แตกต่างกัน 8 ช่วง

1.6 วิธีดำเนินการวิจัย

ในการศึกษา ได้แบ่งประเด็นในการตีปัญหาออกเป็น 4 ขั้นตอน ดังนี้

1.6.1 ทำตารางการแจกแจงความน่าจะเป็นแบบเอฟ เนื่องจากสูตรที่แท้จริงซึ่งสัมพันธ์กับการแจกแจงเอฟ ต้องใช้ค่าตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นแบบเอฟ ณ ค่าองศาอิสระและเปอร์เซนต์ไตล์ต่าง ๆ ซึ่งบางค่าของเปอร์เซนต์ไตล์ ยังไม่มีการสร้างตารางไว้ จึงมีความจำเป็นที่จะต้องสร้างค่าเอฟในตารางเพิ่มเติม เฉพาะค่าเปอร์เซนต์ไตล์ที่มีตารางอยู่แล้ว เพื่อให้ได้ค่าเอฟครบ ตามที่ต้องใช้แทนค่าในสูตรสร้างช่วงความเชื่อมั่น

สำหรับค่าเปอร์เซนต์ไตล์ที่ยังไม่มีการทำตารางเอาไว้ จะทำตารางการแจกแจงความน่าจะเป็นแบบเอฟ ณ ค่าเปอร์เซนต์ไตล์นั้น ๆ ทั้งตาราง โดยมีขอบเขตเฉพาะค่าองศาอิสระ ที่ต้องใช้แทนค่าในสูตรเท่านั้น

1.6.2 จากตารางการแจกแจงความน่าจะเป็นแบบเอฟ ที่สร้างขึ้นเพิ่มเติม จะนำไปเปรียบเทียบกับของเก่าที่มีอยู่แล้ว เพื่อเป็นการยืนยันความถูกต้องเชื่อถือได้

1.6.3 ทำการจำลองค่าตัวแปรสุ่มทวินามที่มีค่าพารามิเตอร์เท่ากับ (n,p) ให้สอดคล้องตามที่กำหนดไว้ในขอบเขตของการวิจัย (หัวข้อ 1.5) แล้วทำการคำนวณช่วงความเชื่อมั่นสำหรับค่าพารามิเตอร์ p จาก

ก). สูตรการประมาณโดยใช้การแจกแจงปกติ

ข). สูตรที่แท้จริงซึ่งสัมพันธ์กับการแจกแจงเอฟ

ในแต่ละค่าของพารามิเตอร์ (n,p) จะทำการสร้างค่าตัวแปรสุ่มซ้ำจำนวน 1000 ครั้ง แล้วทำการประมาณค่าระดับความเชื่อมั่นที่ได้ จากวิธีการคำนวณช่วงความเชื่อมั่น 2 วิธีดังกล่าว อนึ่งในการจำลองค่าตัวแปรสุ่ม จะทำการทดสอบว่าค่าตัวแปรสุ่มที่จำลองขึ้นว่ามีการแจกแจงแบบทวินามหรือไม่

สำหรับกรณีที่พบว่าตัวแปรสุ่มที่จำลองขึ้นไม่ได้มีการแจกแจงทวินาม จะไม่นำค่าของตัวแปรสุ่มดังกล่าวมาใช้ โดยจะทำการจำลองค่าของตัวแปรสุ่มใหม่ จนกว่าจะผ่านการทดสอบว่า ตัวแปรสุ่มที่จำลองขึ้นมีการแจกแจงทวินามจริง จึงจะนำค่าของตัวแปรสุ่ม 1000 ค่า ดังกล่าว ไปใช้ในการประมาณค่าระดับความเชื่อมั่น

1.6.4 ทำการวินิจฉัยค่าระดับความเชื่อมั่นที่ประมาณได้ โดยจะทำการเปรียบเทียบ ค่าระดับความเชื่อมั่นที่ได้จากสูตรที่แท้จริงซึ่งสัมพันธ์กับการแจกแจงเอฟ กับ สูตรการประมาณโดยใช้การแจกแจงปกติ จากนั้นจะทำการสรุปผลการวิจัย

1.7 ประโยชน์ที่ได้รับจากการวิจัย

1.7.1 ผู้ทำการวิจัยสามารถนำช่วงที่สร้างขึ้น ไปใช้ในการประมาณค่าแบบช่วงสำหรับค่าสัดส่วนของความสำเร็จในประชากรที่ศึกษาได้ ในกรณีที่การวิจัยมีความจำเป็นที่จะต้องใช้ตัวอย่างขนาดเล็ก ($n = 1$ ถึง 30)

1.7.2 ผู้ทำการวิจัยสามารถใช้สูตรการประมาณค่าแบบช่วงสำหรับพารามิเตอร์ p ของประชากรที่มีการแจกแจงทวินามได้อย่างเหมาะสม เมื่อมีขนาดตัวอย่างเล็กและใหญ่

1.7.3 การวิจัยครั้งนี้จะมีผลพลอยได้จากการวิจัยออกมาด้วย คือ จะได้ตารางการแจกแจงความน่าจะเป็นแบบเอฟ ที่ได้สร้างเพิ่มเติมขึ้น

1.7.4 การวิจัยครั้งนี้สามารถใช้เป็นแนวทางในการศึกษา การประมาณค่าแบบช่วงสำหรับพารามิเตอร์ ของการแจกแจงแบบไม่ต่อเนื่องชนิดอื่น ๆ สำหรับกรณีตัวอย่างขนาดเล็ก