บทที่ 1
บทนำ

1.1 หลักการและเหตุผล

ในการศึกษาวิธีการทางสถิติหรือแม้แต่การนำสถิติไปประยุกต์ใช้กับงานวิจัยในสาขาวิชาต่างๆ สิ่งหนึ่งที่นักสถิติหรือผู้ทำงานวิจัย ควรจะต้องมี คือ ความรู้ทางด้านสถิติคณิตศาสตร์ (Mathematical Statistics) ซึ่งความรู้ดังกล่าวมีส่วนสำคัญในการที่จะทำให้ทราบถึงลักษณะของตัวแปรสุ่ม และการแจกแจงของตัวแปรสุ่ม ซึ่งการทราบการแจกแจงของตัวแปรสุ่มนับเป็นปัจจัยสำคัญที่จะนำไปสู่การเลือกตัวสถิติทดสอบ และวิธีการวิเคราะห์ข้อมูลที่ถูกต้อง และเชื่อถือได้

การแจกแจงของตัวแปรสุ่มที่นักสถิติ หรือนักวิจัยทราบกันเป็นอย่างดี ได้แก่ การแจกแจงปกติ (Normal distribution) ซึ่งมีการนำมาใช้กันอย่างกว้างขวาง รวมทั้ง วิธีการทางสถิติส่วนใหญ่ เช่น การวิเคราะห์ความแปรปรวน (Analysis of variance), การวิเคราะห์การถดถอย (Regression Analysis) มักมีเงื่อนไขหรือข้อสมมุติฐานที่ว่า ข้อมูลจะต้องมีการแจกแจงปกติ ซึ่งในความเป็นจริงนั้นข้อมูลที่มีอยู่ในปัจจุบันไม่ได้มีการแจกแจงปกติทั้งหมด และ หากนักสถิติหรือผู้ทำงานวิจัยนำข้อมูลที่ไม่ได้มีการแจกแจงปกติ ไปทำการศึกษาโดยใช้วิธีการทางสถิติที่มีเงื่อนไขว่า ข้อมูลต้องมีการแจกแจงปกติแล้ว จะทำให้ผลการศึกษาหรืองานวิจัยนั้น มีความคลาดเคลื่อน

ดังได้กล่าวในตอนต้นแล้วว่า ข้อมูลที่ได้พบไม่ได้มีการแจกแจงปกติทั้งหมด แต่อาจจะมีการแจกแจงในลักษณะที่เบ้ (Skewed distribution) ซึ่งการแจกแจงดังกล่าวนี้ มีลักษณะที่สำคัญคือ มีค่าเฉลี่ยต่ำ และความแปรปรวนสูง ซึ่งข้อมูลที่มีความแปรปรวนสูง เรากล่าวว่าเป็นข้อมูลที่มีการแจกแจงแบบหางยาว (Long tailed Distribution)

การแจกแจงแบบหางยาว เป็นการแจกแจงที่มีความแปรปรวนสูงกว่าการแจกแจงปกติ หรือกล่าวว่า เป็นการแจกแจงที่มีส่วนหางของโค้งการแจกแจงลดลงเข้าสู่ศูนย์ช้ามาก เมื่อเทียบกับการแจกแจงปกติ ซึ่งการแจกแจงแบบหางยาวนี้ จะมีสมบัติที่แตกต่างไปจาก การแจกแจงปกติ

ในการตรวจสอบการแจกแจงว่ามีลักษณะหางยาวหรือไม่นั้นเราสามารถทำได้โดยการพิจารณาที่ค่าสัมประสิทธิ์ความโด่ง (Coefficient of kurtosis) โดยที่ ถ้าค่าดังกล่าว มีค่ามากกว่าสาม แสดงว่าการแจกแจงมีลักษณะหางยาว ถ้ามีค่าน้อยกว่าสาม แสดงว่าการแจกแจงมีลักษณะหางสั้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

และ การแจกแจงปกติ มีค่าสัมประสิทธิ์ความโด่ง เท่ากับ สาม ซึ่งค่าสัมประสิทธิ์ความโด่ง มีวิธีการหาดังนี้คือ

$$\text{Coefficient of kurtosis} = \frac{E(X-\mu)^4}{\sigma^4}$$

เมื่อ σ คือค่าเบี่ยงเบนมาตรฐาน

ซึ่งค่าดังกล่าวจะไม่เปลี่ยนแปลงตาม scale และ location ของการแจกแจง เช่น ค่าสัมประสิทธิ์ความโด่งของกลุ่มตัวอย่าง (หรือประชากร) ของค่าอุณหภูมิที่มีหน่วยเป็น องศาฟาเรนไฮต์ จะไม่เปลี่ยนแปลง ถ้าเราแปลงองศาเป็น องศาเซลเซียส เป็นต้น ซึ่งในกรณีเช่นนี้ ค่าเฉลี่ยและความแปรปรวนจะมีการเปลี่ยนแปลง

เนื่องจากมีหลายการแจกแจงที่มีลักษณะหางยาว เช่น การแจกแจงพาเรโต (Pareto distribution) การแจกแจงที (Student't distribution) การแจกแจงลอกนอร์มอล (Lognormal distribution) การแจกแจงโคชี (Cauchy distribution) การแจกแจงลาปลาซ (Laplace distribution) การแจกแจงโลจิสติก (Logistic distribution) ฯลฯ ซึ่งในที่นี้จะทำการศึกษา การแจกแจงลอกนอร์มอล การแจกแจงโคชี และ การแจกแจงลาปลาซ เท่านั้น เนื่องจากทั้งสามการแจกแจงมีการนำไปประยุกต์ใช้ในชีวิตประจำวันมาก รวมทั้งในบางการแจกแจงยังมีลักษณะพิเศษทางสถิติด้วย

การแจกแจงลอกนอร์มอลเริ่มเป็นที่รู้จักตั้งแต่ปี พ.ศ. 2422 ซึ่งเป็นที่เชื่อกันว่า Galton และ McAlister เป็นนักวิทยาศาสตร์กลุ่มแรกที่ใช้การแจกแจงดังกล่าว ดังนั้นบางครั้งการแจกแจงลอกนอร์มอลจึงถูกเรียกว่า การแจกแจง Galton-McAlister นอกจากนี้ในทางเศรษฐศาสตร์บางครั้งเรียกว่า การแจกแจง Gobb-Douglas การแจกแจงลอกนอร์มอลถูกนำไปประยุกต์ใช้กับเหตุการณ์ต่างๆ ในปัจจุบันอย่างกว้างขวาง เช่น ในทางเศรษฐศาสตร์ได้นำการแจกแจงลอกนอร์มอลไปอธิบายการกระจายรายได้ของประชากร รวมทั้งบริษัทประกันภัยได้นำไปอธิบายการกระจายของเงินผลตอบแทนที่บริษัทต้องจ่ายให้กับลูกค้าผู้ทำประกันภัยในแต่ละปี ทางด้านวิศวกรรมได้นำการแจกแจงลอกนอร์มอลไปอธิบายการกระจายความเร็วลมในสถานที่ต่างๆเพื่อศึกษาความเร็วลมและนำพลังงานลมกลับมาใช้ประโยชน์ต่อไป เป็นต้น

สำหรับการแจกแจงโคชีและการแจกแจงลาปลาซนั้นแม้ว่าจะมีการนำไปประยุกต์ใช้น้อยกว่าการแจกแจงลอกนอร์มอล แต่การแจกแจงทั้งสองชนิดก็มีลักษณะพิเศษและมีประโยชน์มากในบางสายงาน เช่น การแจกแจงโคชีหรือเรียกอีกชื่อหนึ่งว่า การแจกแจงลอเรนซ์ (Lorentz distribution) มีลักษณะพิเศษทางสถิติคือ ไม่มีค่าเฉลี่ย ความแปรปรวน สัมประสิทธิ์ความเบ้ และสัมประสิทธิ์ความโด่ง ซึ่งการแจกแจงโคชีนี้มีการนำไปประยุกต์ใช้ในทางฟิสิกส์ เช่น นำไปอธิบายความหนาแน่นของสนามแม่เหล็กไฟฟ้า เป็นต้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ส่วนการแจกแจงลาปลาซหรืออาจเรียกได้อีกชื่อหนึ่งว่าการแจกแจงดับเบิลเอกซ์โปเนนเชียล (Double exponential distribution) ได้ถูกนำไปอธิบายการกระจายของอัตราการเจริญเติบโตของบริษัทที่ทำอุตสาหกรรมการผลิต

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาที่มา คุณสมบัติ และทฤษฎีบทที่สำคัญ ตลอดจนการประยุกต์ใช้การแจกแจงที่มีลักษณะหางยาว ได้แก่ การแจกแจงลอกนอร์มอล การแจกแจงโคชี และการแจกแจงลาปลาซ

1.3 ขอบเขตและวิธีวิจัย

1. ขอบเขตการศึกษา

ในการศึกษาการแจกแจงของตัวแปรสุ่มที่มีลักษณะหางยาวจะศึกษาทฤษฎีและลักษณะที่สำคัญ ตลอดจนแนวทางในการประยุกต์ใช้ของการแจกแจงลอกนอร์มอล การแจกแจงโคชี และการแจกแจงลาปลาซ

2. วิธีวิจัย

2.1 ศึกษาการแจกแจงของตัวแปรสุ่มที่มีลักษณะหางยาว

2.2 รวบรวมสมบัติและลักษณะที่สำคัญเกี่ยวกับการแจกแจงของตัวแปรสุ่มที่มีลักษณะหางยาว ตลอดจนแนวทางในการประยุกต์ใช้

1.4 นิยามศัพท์

1. ฟังก์ชันความหนาแน่นน่าจะเป็นของตัวแปรสุ่ม X แทนด้วยสัญลักษณ์ f(x)
2. ฟังก์ชันความหนาแน่นน่าจะเป็นสะสมของตัวแปรสุ่ม X แทนด้วยสัญลักษณ์ F(x)
3. ค่าคาดหวังของตัวแปรสุ่ม X แทนด้วยสัญลักษณ์ E(x)
4. ความแปรปรวนของตัวแปรสุ่ม X แทนด้วยสัญลักษณ์ V(x)

1.5 ประโยชน์ที่จะได้รับจากการศึกษาเชิงทฤษฎีและ/หรือ เชิงประยุกต์

1. เพื่อเป็นแหล่งอ้างอิงสำหรับผู้ที่สนใจศึกษาการแจกแจงที่มีลักษณะหางยาว
2. เพื่อเป็นพื้นฐานในการนำไปประยุกต์ใช้กับข้อมูลที่มีการแจกแจงในลักษณะหางยาว

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved