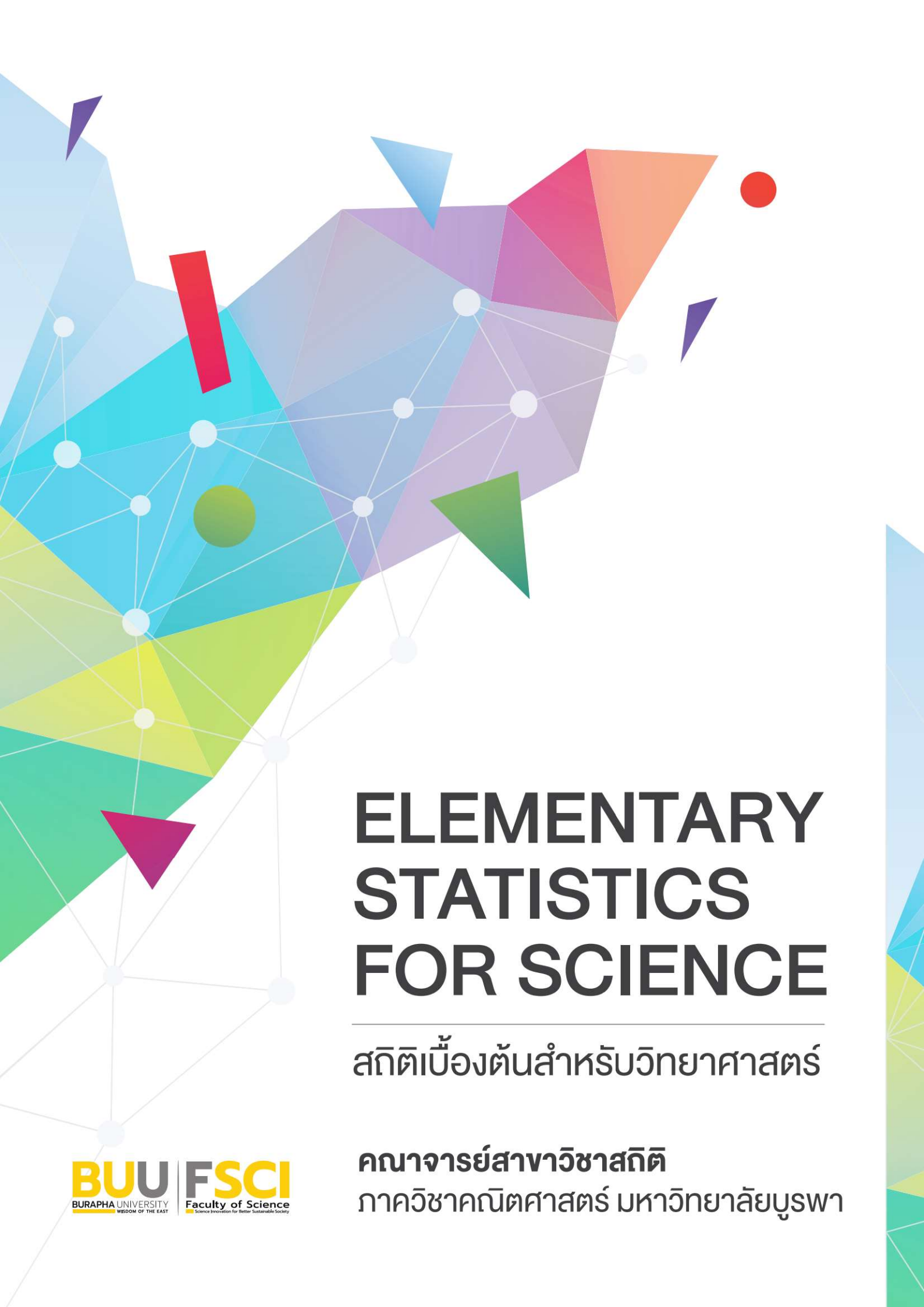




ELEMENTARY STATISTICS FOR SCIENCE

สถิติเบื้องต้นสำหรับวิทยาศาสตร์

คณาจารย์สาขาวิชาสถิติ
ภาควิชาคณิตศาสตร์ มหาวิทยาลัยบูรพา

สถิติเบื้องต้นสำหรับวิทยาศาสตร์

ELEMENTARY STATISTICS FOR SCIENCE

คณาจารย์สาขาวิชาสถิติ
ภาควิชาคณิตศาสตร์ มหาวิทยาลัยบูรพา

สถิติเบื้องต้นสำหรับวิทยาศาสตร์

ISBN (e-book): 978-974-384-620-5

จัดทำโดย ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา
ราคา 150 บาท

คำนำ

หนังสือเล่มนี้จัดทำขึ้นเพื่อให้นิสิตระดับปริญญาตรี มหาวิทยาลัยบูรพา ใช้ประกอบการเรียนการสอนรายวิชา สถิติเบื้องต้นสำหรับวิทยาศาสตร์ ซึ่งคณะผู้เขียนได้ประมวลเอาความรู้และประสบการณ์ในการสอนรายวิชาดังกล่าวเป็นเวลาหลายปีมาเรียบเรียงเป็นภาษาที่เข้าใจง่าย โดยเนื้อหาในเล่มมีรายละเอียดครอบคลุมเกี่ยวกับสถิติเบื้องต้นทั้งหมด ได้แก่ ความรู้เบื้องต้นทางสถิติ ความน่าจะเป็นและการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม การประมาณค่าพารามิเตอร์ การทดสอบสมมติฐาน และการทดสอบไคกำลังสอง อีกทั้งยังมีตัวอย่างผลการวิเคราะห์ข้อมูลด้วยโปรแกรม R ในบางหัวข้อ รวมทั้งแบบฝึกหัดให้ได้ลองฝึกคิดทบทวนความรู้ เพื่อให้เกิดความเข้าใจในบทเรียนอย่างถ่องแท้

ขอขอบคุณทุกท่านที่มีส่วนร่วมในการปรับปรุงแก้ไขหนังสือเล่มนี้ให้สำเร็จด้วยดี ขอคุณภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ที่ให้การสนับสนุน และหวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้จะเป็นประโยชน์กับนิสิต นักศึกษา นักวิจัย และผู้สนใจ และสามารถนำความรู้เกี่ยวกับความรู้เกี่ยวกับสถิติเบื้องต้นไปประยุกต์ใช้ได้อย่างถูกต้องต่อไป

คณาจารย์สาขาวิชาสถิติ

มิถุนายน 2567

สารบัญ

คำนำ	I
1 ความรู้เบื้องต้นทางสถิติ	1
1.1 บทนำ	1
1.1.1 ความหมายของสถิติ	1
1.1.2 สถิติเชิงพรรณนาและสถิติเชิงอนุมาน	2
1.1.3 ประชากรและตัวอย่าง	3
1.2 ชนิดของข้อมูล	4
1.2.1 ข้อมูลเชิงคุณภาพและข้อมูลเชิงปริมาณ	4
1.2.2 การแบ่งชนิดของข้อมูลตามระดับการวัดค่า	4
1.3 การอธิบายข้อมูลด้วยตารางความถี่และแผนภาพ	5
1.3.1 ตารางความถี่และแผนภาพสำหรับข้อมูลเชิงคุณภาพ	5
1.3.2 ตารางความถี่และแผนภาพสำหรับข้อมูลเชิงปริมาณ	10
1.4 การอธิบายข้อมูลเชิงปริมาณด้วยค่าวัดทางสถิติ	16
1.4.1 ค่าวัดแนวโน้มเข้าสู่ส่วนกลาง	16
1.4.2 ค่าวัดการกระจายของข้อมูล	20
2 ความน่าจะเป็น	
และการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม	29
2.1 ความน่าจะเป็น	29
2.1.1 การทดลองความน่าจะเป็น ปริภูมิตัวอย่าง และเหตุการณ์	29
2.1.2 คุณสมบัติของความน่าจะเป็น	31
2.2 การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม	32
2.2.1 ตัวแปรสุ่มและชนิดของตัวแปรสุ่ม	32
2.2.2 การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม	33
2.3 การแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่องที่สำคัญ	33

2.3.1	การแจกแจงเอกรูปไม่ต่อเนื่อง	34
2.3.2	การแจกแจงทวินาม	36
2.3.3	การแจกแจงปัวซอง	40
3	การแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่องที่สำคัญ	47
3.1	การแจกแจงเอกรูปต่อเนื่อง	47
3.2	การแจกแจงปกติ	49
3.3	การแจกแจงที	57
3.4	การแจกแจงไคกำลังสอง	59
3.5	การแจกแจงเอฟ	62
4	การประมาณค่าพารามิเตอร์สำหรับประชากร 1 กลุ่ม	70
4.1	การประมาณค่าพารามิเตอร์	70
4.1.1	การประมาณค่าแบบจุด	71
4.1.2	การประมาณค่าแบบช่วง	71
4.2	การประมาณค่าค่าเฉลี่ยของประชากร	73
4.2.1	กรณีทราบค่า σ^2	73
4.2.2	กรณีไม่ทราบค่า σ^2 และตัวอย่างมีขนาดใหญ่	74
4.2.3	กรณีไม่ทราบค่า σ^2 และตัวอย่างมีขนาดเล็ก	74
4.3	การประมาณค่าสัดส่วนของประชากร (p)	79
4.4	การประมาณค่าความแปรปรวนของประชากร (σ^2)	81
5	การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ สำหรับประชากร 1 กลุ่ม	87
5.1	องค์ประกอบของการทดสอบสมมติฐาน	87
5.1.1	สมมติฐานว่างและสมมติฐานทางเลือก	88
5.1.2	ตัวสถิติทดสอบ	89
5.1.3	ระดับนัยสำคัญ ค่าวิกฤต และบริเวณวิกฤต	90
5.1.4	ค่าพี	91
5.1.5	ประเภทของการทดสอบสมมติฐาน	91
5.1.6	การตัดสินใจและการสรุป	93
5.1.7	ความผิดพลาดแบบที่ 1 และ 2	93
5.2	การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย	94
5.2.1	กรณีทราบค่า σ^2	95
5.2.2	กรณีไม่ทราบค่า σ^2 และตัวอย่างมีขนาดใหญ่	95

5.2.3	กรณีไม่ทราบค่า σ^2 และตัวอย่างมีขนาดเล็ก	96
5.3	การทดสอบสมมุติฐานเกี่ยวกับสัดส่วน	103
5.4	การทดสอบสมมุติฐานเกี่ยวกับความแปรปรวน	105
6	การทดสอบสมมุติฐานและการสร้างช่วงความเชื่อมั่น	
	สำหรับพารามิเตอร์ของประชากร 2 กลุ่ม	111
6.1	การทดสอบสมมุติฐานและช่วงความเชื่อมั่นสำหรับผลต่าง ระหว่างสัดส่วนของประชากร 2 กลุ่ม	111
6.2	การทดสอบสมมุติฐานและช่วงความเชื่อมั่นสำหรับอัตราส่วนความแปรปรวนของประชากร 2 กลุ่ม	118
6.3	การทดสอบสมมุติฐานและช่วงความเชื่อมั่นสำหรับผลต่าง ระหว่างค่าเฉลี่ยของประชากร 2 กลุ่ม เมื่อตัวอย่าง 2 ชุด เป็นอิสระกัน	124
6.3.1	กรณีทราบความแปรปรวนของประชากรทั้ง 2 กลุ่ม	125
6.3.2	กรณีไม่ทราบความแปรปรวนของประชากร และตัวอย่างมีขนาดใหญ่	126
6.3.3	กรณีไม่ทราบความแปรปรวนของประชากร และตัวอย่างมีขนาดเล็ก	127
6.4	การทดสอบสมมุติฐานและช่วงความเชื่อมั่นสำหรับผลต่าง ระหว่างค่าเฉลี่ยของประชากร 2 กลุ่ม เมื่อตัวอย่าง 2 ชุด ไม่เป็นอิสระกัน	145
7	การทดสอบไคกำลังสอง	160
7.1	การทดสอบภาวะสารูปดี	161
7.2	การทดสอบความเป็นอิสระ	164
7.3	การทดสอบภาวะความเท่ากัน	167
	บรรณานุกรม	173
	ภาคผนวก: ตารางสถิติ	175

บทที่ 1

ความรู้เบื้องต้นทางสถิติ

1.1 บทนำ

1.1.1 ความหมายของสถิติ

ผู้ที่เริ่มเรียนวิชาสถิติมักจะถามตัวเองว่าสถิติคืออะไร ทำไมต้องเรียนสถิติ เรียนแล้วนำไปใช้ประโยชน์ในสาขาวิชาของตนได้อย่างไร หรือสถิติเกี่ยวข้องกับชีวิตประจำวันอย่างไร

ในวิชาสถิติเบื้องต้นนี้มีจุดมุ่งหมายเพื่อชี้ให้เห็นถึงคำตอบของคำถามที่กล่าวมาในข้างต้น โดยให้ผู้เรียนได้มีความรู้เกี่ยวกับสถิติในระดับพื้นฐาน สามารถนำความรู้เหล่านี้ไปใช้ได้อย่างถูกต้อง และใช้ความรู้เหล่านี้ในการศึกษาวิธีการทางสถิติขั้นสูงต่อไป

สถิติ (Statistics) คือวิธีการหรือกระบวนการที่ประกอบด้วย การเก็บรวบรวมข้อมูล (Collecting data) จากนั้นนำข้อมูลที่ได้มาจัดการให้อยู่ในรูปแบบที่มีประโยชน์ (Organizing) สรุปสาระสำคัญที่ได้จากข้อมูลที่มีการจัดการแล้ว (Summarizing) นำเสนอสาระสำคัญที่ได้ (Presenting) นำข้อมูลมาวิเคราะห์ทางสถิติ (Analyzing) และแปลผลที่ได้จากการวิเคราะห์ (Interpreting) และท้ายสุดทำการสรุปผลที่ได้จากข้อมูล (Drawing conclusions)

โดยทั่วไปเมื่อกล่าวถึงคำว่า **สถิติ** เรามักนึกถึงตัวเลข หรือตารางตัวเลข เช่น จำนวนทารกแรกเกิดในปี พ.ศ. 2566 อัตราการป่วยเป็นโรคมะเร็งหลอดลมของผู้ที่สูบบุหรี่ แต่ในความเป็นจริงแล้วข้อมูลทางสถิติสามารถเป็นได้ทั้งตัวเลขและข้อความ ซึ่งบอกถึงลักษณะของสิ่งที่สนใจศึกษา เช่น เพศ ระดับการศึกษา หมู่เลือด และ สี เป็นต้น

เกือบจะเรียกได้ว่าคนทุกคนเกี่ยวข้องกับข้อมูล ไม่ว่าจะเป็น นักธุรกิจ นักวิจัย หรือ อาจารย์ ซึ่งข้อมูลเหล่านี้ อาจเกี่ยวข้องกับยอดขายรายเดือน อัตราการปนเปื้อนของสารเคมีในตัวอย่างน้ำ และคะแนนสอบของนักเรียน เป็นต้น

ทำไมเราจึงต้องเรียนสถิติ

มีเหตุผลหลายประการที่ทำให้เราต้องเรียนสถิติ เช่น

1. นิสิตต้องมีความสามารถในการอ่านและมีความเข้าใจเกี่ยวกับวิธีการทางสถิติที่ใช้ในการศึกษาในสาขานั้น ๆ ดังนั้น นิสิตจึงจำเป็นต้องมีความรู้เกี่ยวกับคำศัพท์ สัญลักษณ์ หลักการ และวิธีการทางสถิติที่ใช้ในการศึกษาเหล่านั้น
2. นิสิตและผู้ที่อยู่ในสายอาชีพต่าง ๆ อาจจำเป็นต้องสร้างงานวิจัยในสาขาของตนเองขึ้น และด้วยเหตุที่วิธีการทางสถิติเป็นพื้นฐานของงานวิจัยต่าง ๆ เราจึงต้องสามารถที่จะสร้างแผนแบบการทดลอง เก็บรวบรวมข้อมูล จัดการข้อมูล วิเคราะห์ข้อมูล และสรุปสิ่งที่ได้จากข้อมูล นอกจากนั้นอาจจะต้องมีการพยากรณ์สำหรับประโยชน์ในอนาคตอีกด้วย และเรายังต้องสามารถสื่อสารหรือแปลผลที่ได้จากการศึกษาในรูปแบบของตนเอง และให้ผู้อื่นเข้าใจได้ง่ายด้วย
3. นิสิตและผู้ที่อยู่ในสายอาชีพต่าง ๆ สามารถใช้ความรู้ที่ได้จากสถิติมาช่วยทำให้ตนเองกลายเป็นผู้บริโภคหรือสมาชิกในสังคมที่ดีได้ เช่น เราสามารถตัดสินใจได้อย่างฉลาดในการเลือกซื้อสินค้าโดยอาศัยการศึกษาเกี่ยวกับผู้บริโภค และสามารถทราบถึงการใช้จ่ายหรือการบริหารงานของรัฐบาลโดยอาศัยการศึกษาต่าง ๆ ได้

ประโยชน์ของสถิติกับงานด้านต่าง ๆ

1. ด้านวิทยาศาสตร์

ในงานด้านวิทยาศาสตร์ได้มีการนำวิธีการเชิงสถิติไปประยุกต์ใช้ในเกือบทุกด้าน เช่น งานวิจัยในด้านการแพทย์ซึ่งเป็นงานวิจัยหรือการทดลองเกี่ยวกับยา หรือวิธีการรักษาโรคต่าง ๆ การทดลองเกี่ยวกับการใช้ฮอร์โมนที่มีผลต่อการเจริญเติบโตของพืช หรือการศึกษาเกี่ยวกับปริมาณสารปนเปื้อนในอาหาร เป็นต้น

2. ด้านธุรกิจ

ในงานด้านธุรกิจนั้นจำเป็นต้องใช้ข้อมูลและวิธีการทางสถิติมาช่วยในตัดสินใจและการบริหารหน่วยงาน เช่น ด้านการตลาด การผลิต และการวางแผนในการดำเนินงานในหน่วยงาน โดยการนำข้อมูลมาวิเคราะห์ทางสถิติ เพื่อการพยากรณ์เหตุการณ์ที่จะเกิดขึ้นในอนาคตโดยอาศัยข้อมูลในอดีต

3. งานด้านการวางแผนเพื่อพัฒนาเศรษฐกิจของประเทศ

รัฐบาลสามารถวางแผนการพัฒนาประเทศโดยอาศัยข้อมูลทางสถิติในด้านต่างๆ เช่น สถิติประชากร สถิติการศึกษา สถิติแรงงาน สถิติอุตสาหกรรม สถิติเกษตร เป็นต้น

1.1.2 สถิติเชิงพรรณนาและสถิติเชิงอนุมาน

ในการหาความรู้เกี่ยวกับเหตุการณ์ที่ดูเหมือนว่าจะเกิดขึ้นโดยสุ่ม หรือเหตุการณ์ที่ไม่สามารถทราบผลได้ล่วงหน้า นักสถิติต้องเก็บรวบรวมข้อมูลสำหรับตัวแปร (Variable) ที่สนใจ ซึ่งตัวแปรนี้จะอธิบายถึงการเกิดเหตุการณ์

ที่สนใจนี้ เช่น เพศของลูกค้า รายได้ของผู้ใช้โทรศัพท์มือถือ เป็นต้น

ตัวแปร (Variable) คือลักษณะหรือคุณสมบัติที่ผู้ศึกษาสนใจ โดยลักษณะที่สนใจนี้มีค่า (Aalues) ที่เป็นไปได้ที่แตกต่างกัน

ตัวแปรสุ่ม (Random variable) คือตัวแปรที่ค่าของมันหาได้ด้วยความบังเอิญ (Chance)

ข้อมูล (Data) คือค่าที่เป็นไปได้ของตัวแปรที่อาจได้จากการวัดหรือการสังเกต

ชุดข้อมูล (Data set) คือกลุ่มของข้อมูลที่รวบรวมมาได้

ในทางสถิติ วิธีเชิงสถิติสามารถแบ่งได้เป็นสองกลุ่มหลักขึ้นอยู่กับการนำข้อมูลไปใช้ ดังนี้

1. **สถิติเชิงพรรณนา (Descriptive statistics)** หมายถึงสถิติที่เกี่ยวข้องกับการเก็บรวบรวมข้อมูล จัดการข้อมูล การสรุป และการนำเสนอข้อมูล ซึ่งเป็นการใช้ตัวเลขและแผนภาพรูปแบบต่าง ๆ หรือตารางในการอธิบายลักษณะของข้อมูล และสรุปสาระสำคัญที่ได้จากข้อมูล
2. **สถิติเชิงอนุมาน (Inferential statistics)** หมายถึงสถิติที่เกี่ยวข้องกับการใช้ข้อมูลตัวอย่างในการหาข้อสรุปเกี่ยวกับประชากร โดยอาศัยทฤษฎีความน่าจะเป็น ได้แก่ โอกาสของการเกิดเหตุการณ์หนึ่ง ๆ เข้ามาช่วย ซึ่งวิธีการทางสถิติอนุมาน ได้แก่ การประมาณค่า และการทดสอบสมมุติฐาน เป็นต้น

1.1.3 ประชากรและตัวอย่าง

จากความหมายของสถิติอนุมาน คำว่าตัวอย่างและประชากร คือ

ประชากร (Population) คือกลุ่มของหน่วย (คน สัตว์ สิ่งของ หรือเหตุการณ์) ที่สนใจศึกษา เช่น การศึกษารายได้ของคนไทยในปี พ.ศ. 2566 ในกรณีนี้ประชากรคือคนไทยทุกคน

ตัวอย่าง (Sample) คือกลุ่มย่อย (Subset) ของหน่วยของประชากร ที่ถูกเลือกมาจากประชากรนั้น เช่น การสำรวจรายได้ของคนไทย ตัวอย่างคือคนไทยบางส่วนที่ถูกเลือกเป็นตัวอย่าง

นอกจากนี้ ยังมีคำอีกสองคำที่ใช้ในการศึกษาทางสถิติที่ควรทราบ คือ พารามิเตอร์และค่าสถิติ

พารามิเตอร์ (Parameter) คือ ค่าเป็นตัวเลขที่ใช้ในการอธิบายลักษณะเฉพาะของประชากร เช่น ค่าเฉลี่ยของประชากร ค่าความแปรปรวนของประชากร เป็นต้น

ค่าสถิติ (Statistic) คือค่าที่เป็นตัวเลขที่ใช้อธิบายลักษณะเฉพาะของตัวอย่าง เช่นค่าเฉลี่ยของตัวอย่าง และค่าความแปรปรวนของตัวอย่าง เป็นต้น

1.2 ชนิดของข้อมูล

โดยทั่วไปในการเก็บรวบรวมข้อมูล ข้อมูลที่ได้มักมีทั้งข้อมูลที่มีค่าเป็นตัวเลขและที่ไม่ได้มีค่าเป็นตัวเลข เช่น การเก็บข้อมูลเกี่ยวกับความพึงพอใจของผู้มาใช้บริการของโรงพยาบาล ข้อมูลที่เก็บ เช่น เพศ อายุ ระดับการศึกษา และอาชีพ เป็นต้น ในการวิเคราะห์ข้อมูลเหล่านี้ในทางสถิติไม่สามารถใช้วิธีเดียวกันได้ทั้งหมด วิธีใดจะเป็นวิธีที่เหมาะสมขึ้นอยู่กับชนิดของข้อมูล

1.2.1 ข้อมูลเชิงคุณภาพและข้อมูลเชิงปริมาณ

ในทางสถิติสามารถแบ่งข้อมูลได้เป็นสองชนิดใหญ่ ๆ ได้แก่ ข้อมูลเชิงคุณภาพ และข้อมูลเชิงปริมาณ

1. **ข้อมูลเชิงคุณภาพ (Qualitative data)** คือข้อมูลที่สามารถถูกแบ่งหรือจำแนกเป็นกลุ่ม ๆ ตามลักษณะหรือคุณสมบัติบางประการได้ เช่น เพศ (ชาย หญิง) สถานภาพสมรส (โสด แต่งงาน หม้าย) หมู่เลือด (O A B AB) และขนาดรองเท้า (เช่น 38 39 40 42) เป็นต้น
2. **ข้อมูลเชิงปริมาณ (Quantitative data)** คือข้อมูลที่มีค่าเป็นตัวเลข (Numerical measurements) และสามารถนำค่ามาเรียงลำดับได้ เช่น รายได้ อายุ ความสูง และอุณหภูมิร่างกายของคน เป็นต้น

นอกจากนั้น ข้อมูลเชิงปริมาณยังแบ่งออกได้เป็นสองชนิด คือ

1. **ข้อมูลชนิดไม่ต่อเนื่อง (Discrete data)** ซึ่งมีค่าเป็นจำนวนนับ เช่น จำนวนบุตรในครัวเรือน จำนวนนิสิตในห้องเรียน จำนวนแบคทีเรียในน้ำหนึ่งหยด และจำนวนโทรศัพท์ที่เข้ามาในหนึ่งวัน เป็นต้น
2. **ข้อมูลชนิดต่อเนื่อง (Continuous data)** ซึ่งมีค่าได้ไม่จำกัดระหว่างค่าสองค่าใด ๆ (มีค่าเป็นจำนวนจริง) เช่น อุณหภูมิ อายุ และส่วนสูง เป็นต้น

นอกจากการแบ่งข้อมูลเป็นข้อมูลเชิงคุณภาพและเชิงปริมาณแล้ว เรายังสามารถแบ่งข้อมูลตามระดับการวัดค่าได้ด้วย

1.2.2 การแบ่งชนิดของข้อมูลตามระดับการวัดค่า

1. **ข้อมูลนามบัญญัติ (Nominal data)** เป็นข้อมูลที่มีมาตรวัดที่หยาบที่สุด โดยเป็นข้อมูลเชิงคุณภาพเท่านั้น เช่น เพศ หมู่เลือด เป็นต้น ข้อมูลชนิดนี้ไม่สามารถใช้สถิติขั้นสูงในการวิเคราะห์ได้ ใช้ได้เพียงการนับ การหาร้อยละ การหาความถี่ หรือใช้วิธีเชิงสถิติไม่อิงพารามิเตอร์ (Non-parametric statistics) เท่านั้น
2. **ข้อมูลอันดับ (Ordinal data)** เป็นข้อมูลเชิงคุณภาพ มีลักษณะเรียงตามอันดับ แต่ความแตกต่างระหว่างอันดับของข้อมูลไม่มีความหมาย เช่น การประเมินการสอนของอาจารย์ท่านหนึ่งโดยนิสิตที่เรียนในกลุ่ม

นั้น ซึ่งผลการประเมินจะมีได้สามระดับคือ สอนดีมาก ปานกลาง สอนไม่ดี สถิติที่ใช้กับข้อมูลระดับนี้ได้คือ ความถี่ ร้อยละ การจัดอันดับต่าง ๆ และสถิติไม่อิงพารามิเตอร์

3. **ข้อมูลช่วง (Interval data)** เป็นข้อมูลเชิงปริมาณ เรียงอันดับได้ และสามารถคำนวณหาผลต่างระหว่างข้อมูลได้ แต่ไม่มีค่าเป็นศูนย์อย่างแท้จริง นั่นคือเมื่อค่าของข้อมูลเท่ากับศูนย์ยังสามารถวัดค่าได้อยู่ เช่น อุณหภูมิอากาศ เมื่ออุณหภูมิเท่ากับ 0 องศาเซลเซียส น้ำจะเป็นน้ำแข็ง แต่ไม่ได้หมายความว่าที่อุณหภูมิ 0 องศาเซลเซียส ไม่มีความร้อนอยู่เลย สถิติที่ใช้กับข้อมูลระดับนี้คือ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน การหาค่าสหสัมพันธ์ การทดสอบสมมติฐาน เป็นต้น
4. **ข้อมูลอัตราส่วน (Ratio data)** เป็นข้อมูลเชิงปริมาณที่มีมาตรวัดที่มีลักษณะสมบูรณ์ทุกอย่างและมีศูนย์แท้ เช่น ส่วนสูง 0 เซนติเมตร คือไม่มีความสูงอยู่เลย วิธีเชิงสถิติทุกวิธีสามารถใช้ได้กับข้อมูลชนิดนี้

1.3 การอธิบายข้อมูลด้วยตารางความถี่และแผนภาพ

ในการศึกษาทางสถิติ นักวิจัยต้องรวบรวมข้อมูลเกี่ยวกับตัวแปรที่สนใจศึกษา เช่น ในการศึกษาถึงจำนวนของคนที่ถูกงูพิษกัดในพื้นที่ ๆ หนึ่งในช่วงเวลาหลายปีที่ผ่านมา ซึ่งข้อมูลที่ต้องรวบรวมจะได้มาจากแพทย์ โรงพยาบาล และหน่วยงานด้านสุขภาพอื่น ๆ เพื่ออธิบายถึงลักษณะของข้อมูลที่ได้ การสรุปเกี่ยวกับข้อมูล และการอนุมานต่าง ๆ โดยใช้ข้อมูลนี้ นักวิจัยต้องทำการจัดข้อมูลให้อยู่ในรูปแบบที่มีความหมาย ซึ่งวิธีที่สะดวกที่สุดในการจัดข้อมูลคือการสร้างตารางความถี่

หลังจากที่ได้จัดข้อมูลให้อยู่ในรูปแบบที่มีความหมายแล้ว นักวิจัยต้องแสดงรูปแบบของข้อมูลให้ผู้สนใจเข้าใจได้ง่าย ซึ่งวิธีการนำเสนอข้อมูลที่มีประโยชน์มากที่สุดคือการนำเสนอในรูปแบบของแผนภาพ ซึ่งมีได้หลายรูปแบบ แต่ละรูปแบบจะมีวัตถุประสงค์ในการใช้แตกต่างกัน

อย่างไรก็ตาม วิธีการนำเสนอและนำเสนอข้อมูลไม่ว่าจะเป็นในรูปแบบของค่าสถิติหรือแผนภาพนั้นขึ้นอยู่กับชนิดของตัวแปรที่ศึกษา ตัวแปรเชิงคุณภาพจะมีวิธีการอธิบายโดยใช้ค่าความถี่ หรือร้อยละ ในขณะที่ตัวแปรเชิงปริมาณจะใช้ค่าเฉลี่ย ค่ามัธยฐาน และค่าเบี่ยงเบนมาตรฐาน เป็นต้น ส่วนในเรื่องของแผนภาพ ตัวแปรเชิงคุณภาพจะใช้แผนภูมิแท่ง (Bar chart) หรือแผนภูมิรูปวงกลม (Pie chart) ในการนำเสนอข้อมูล ส่วนตัวแปรเชิงปริมาณจะใช้ฮิสโตแกรม (Histogram) แผนภาพลำต้นและใบ (Stem and leaf plot) หรือแผนภาพกล่อง (Box plot) ในการนำเสนอข้อมูล

1.3.1 ตารางความถี่และแผนภาพสำหรับข้อมูลเชิงคุณภาพ

ตารางความถี่

ตัวแปรเชิงคุณภาพจะแบ่งข้อมูลออกเป็นกลุ่มตามคุณสมบัติที่ศึกษา เช่น แบ่งกลุ่มตัวอย่างเป็น ชาย-หญิง แบ่งตามระดับการศึกษา หรือ แบ่งตามอาชีพ ผู้ศึกษาสามารถทำการแจกแจงได้ว่าในกลุ่มตัวอย่างมีผู้ชายกี่คน คิดเป็นร้อยละเท่าใด และเป็นผู้หญิงกี่คน คิดเป็นร้อยละเท่าใด ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1.1 ในการสำรวจความคิดเห็นของผู้ที่มาใช้บริการในโรงพยาบาลแห่งหนึ่ง ได้สุ่มตัวอย่างผู้ที่มาใช้บริการมาจำนวน 200 คน และรวบรวมข้อมูลต่าง ๆ ได้แก่ เพศ อายุ น้ำหนัก ระดับการศึกษา อาชีพ และระดับความพึงพอใจในการให้บริการของโรงพยาบาลแห่งนี้ ตัวอย่างของข้อมูลแสดงในตารางที่ 1.1

ตารางที่ 1.1: ข้อมูลผู้เข้าใช้บริการในโรงพยาบาล

เพศ	อายุ	น้ำหนัก	ระดับการศึกษา	อาชีพ	ความพึงพอใจ
หญิง	52	68	มัธยมศึกษา	พนักงานของรัฐ	มากที่สุด
หญิง	44	52	มัธยมศึกษา	ข้าราชการ	มากที่สุด
หญิง	41	58	ต่ำกว่ามัธยมศึกษา	พนักงานเอกชน	มากที่สุด
หญิง	46	72	ปริญญาโทขึ้นไป	ข้าราชการ	มาก
หญิง	28	68	ต่ำกว่ามัธยมศึกษา	ข้าราชการ	มากที่สุด
ชาย	45	70	มัธยมศึกษา	ข้าราชการ	มาก
ชาย	33	64	มัธยมศึกษา	พนักงานของรัฐ	มากที่สุด
หญิง	42	61	ต่ำกว่ามัธยมศึกษา	พนักงานของรัฐ	มาก
หญิง	40	70	ต่ำกว่ามัธยมศึกษา	พนักงานของรัฐ	มากที่สุด
ชาย	43	75	ต่ำกว่ามัธยมศึกษา	ข้าราชการ	ปานกลาง
ชาย	37	61	ปริญญาโทขึ้นไป	พนักงานของรัฐ	มาก
...	...	...	...	...	...
หญิง	30	65	มัธยมศึกษา	อาชีพอิสระ	มากที่สุด
หญิง	56	65	ปริญญาตรี	พนักงานของรัฐ	มากที่สุด
หญิง	46	58	ปริญญาโทขึ้นไป	ข้าราชการ	มากที่สุด
ชาย	46	65	ต่ำกว่ามัธยมศึกษา	อาชีพอิสระ	มาก
ชาย	38	63	ต่ำกว่ามัธยมศึกษา	ข้าราชการ	มาก
หญิง	35	68	ต่ำกว่ามัธยมศึกษา	ข้าราชการ	มากที่สุด
หญิง	49	56	มัธยมศึกษา	พนักงานของรัฐ	มากที่สุด
ชาย	38	61	มัธยมศึกษา	ข้าราชการ	ปานกลาง
หญิง	38	64	ต่ำกว่ามัธยมศึกษา	ข้าราชการ	มากที่สุด
หญิง	36	66	ปริญญาโทขึ้นไป	พนักงานของรัฐ	มากที่สุด
หญิง	38	51	ปริญญาโทขึ้นไป	ข้าราชการ	มากที่สุด

จากข้อมูลสามารถนำเสนอได้ในรูปตารางความถี่ตามตัวแปรเพศ ระดับการศึกษา และอาชีพ ดังนี้

ตารางที่ 1.2: ตารางความถี่ของเพศ

เพศ	ความถี่	ความถี่สะสม	ร้อยละ	ร้อยละสะสม
ชาย	58	58	29.00	29.00
หญิง	142	200	71.00	100.00

จากตารางที่ 1.2 จะเห็นได้ว่าผู้ที่มาใช้บริการในโรงพยาบาลจำนวน 200 คน เป็นผู้หญิงจำนวน 142 คน คิดเป็นร้อยละ 71 และผู้ชาย 58 คน คิดเป็นร้อยละ 29

ตารางที่ 1.3: ตารางความถี่ของระดับการศึกษา

การศึกษา	ความถี่	ความถี่สะสม	ร้อยละ	ร้อยละสะสม
ต่ำกว่ามัธยมศึกษา	81	81	40.50	40.50
มัธยมศึกษา	66	147	33.00	73.50
ปริญญาตรี	37	184	18.50	92.00
ปริญญาโทขึ้นไป	16	200	8.00	100.00

จากตารางที่ 1.3 จะเห็นได้ว่าผู้ที่มาใช้บริการในโรงพยาบาลจำนวน 200 คน เป็นผู้จบการศึกษาในระดับต่ำกว่ามัธยมศึกษาจำนวน 81 คน คิดเป็นร้อยละ 40.50 รองลงมาคือระดับมัธยมศึกษาจำนวน 66 คน คิดเป็นร้อยละ 33

ตารางที่ 1.4: ตารางความถี่ของอาชีพ

อาชีพ	ความถี่	ความถี่สะสม	ร้อยละ	ร้อยละสะสม
ข้าราชการ	95	95	47.50	47.50
พนักงานของรัฐ	61	156	30.50	78.00
พนักงานเอกชน	24	180	12.00	90.00
อาชีพอิสระ	20	200	10.00	100.00

จากตารางที่ 1.4 จะเห็นได้ว่าผู้ที่มาใช้บริการในโรงพยาบาลจำนวน 200 คน เป็นข้าราชการจำนวน 95 คน คิดเป็นร้อยละ 47.50 รองลงมาคือพนักงานของรัฐจำนวน 61 คน คิดเป็นร้อยละ 30.50

นอกจากการสร้างตารางความถี่สำหรับตัวแปรเดียวแล้ว ยังสามารถสร้างตารางความถี่สำหรับสองตัวแปรได้ดังนี้

ตารางที่ 1.5: ตารางความถี่จำแนกตามเพศและอาชีพ

เพศ	อาชีพ				รวม
	ข้าราชการ	พนักงานของรัฐ	พนักงานเอกชน	อาชีพอิสระ	
ชาย	23	19	10	6	58
หญิง	72	42	14	14	142
รวม	95	61	24	20	200

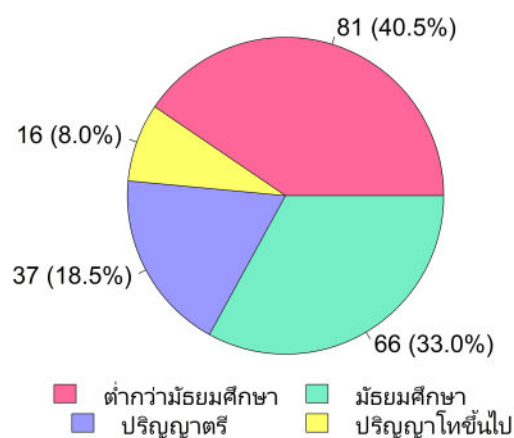
จากตารางที่ 1.5 จะเห็นได้ว่าผู้มาใช้บริการในโรงพยาบาลจำนวน 200 คน เป็นข้าราชการหญิงจำนวน 72 คน รองลงมาคือพนักงานของรัฐหญิงจำนวน 42 คน และน้อยที่สุดคือผู้ที่ประกอบอาชีพอิสระชายจำนวน 6 คน

แผนภาพสำหรับข้อมูลเชิงคุณภาพ

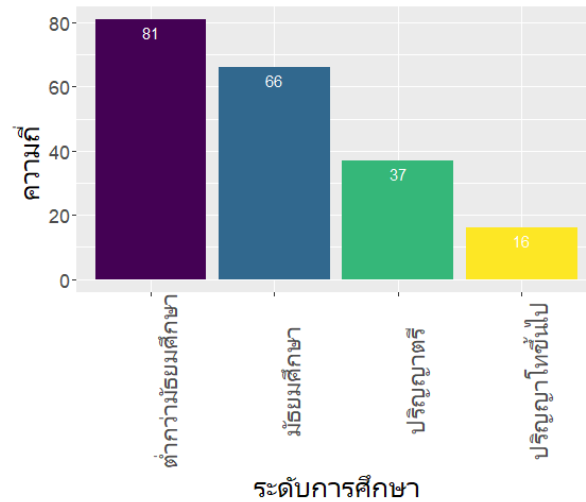
แผนภูมิรูปวงกลม (Pie chart) เป็นแผนภูมิที่นิยมใช้ในการอธิบายข้อมูลเชิงคุณภาพ โดยการสร้างแผนภูมิรูปวงกลมจะแบ่งวงกลมออกเป็น ส่วน ๆ แต่ละส่วนแทนกลุ่มของข้อมูล เช่น แบ่งเพศออกเป็น ชายและหญิง แผนภูมิรูปวงกลมของข้อมูลเพศก็จะแบ่งเป็นสองส่วน หรือแผนภูมิรูปวงกลมของตัวแปรระดับการศึกษาจะถูกแบ่งเป็น 4 ส่วน แทนแต่ละระดับการศึกษา โดยขนาดของพื้นที่ในวงกลมจะแทนจำนวนข้อมูลในแต่ละกลุ่ม

แผนภูมิแท่ง (Bar chart) เป็นแผนภูมิอีกชนิดหนึ่งที่นิยมใช้ในการอธิบายข้อมูลเชิงคุณภาพ โดยแท่งแต่ละแท่งจะแทนแต่ละกลุ่มของข้อมูล ส่วนความสูงของแท่งจะแสดงความถี่หรือจำนวนข้อมูลในแต่ละกลุ่ม

ตัวอย่างที่ 1.2 แผนภูมิรูปวงกลมและแผนภูมิแท่งของตัวแปรระดับการศึกษาจากข้อมูลในตัวอย่างที่ 1.1



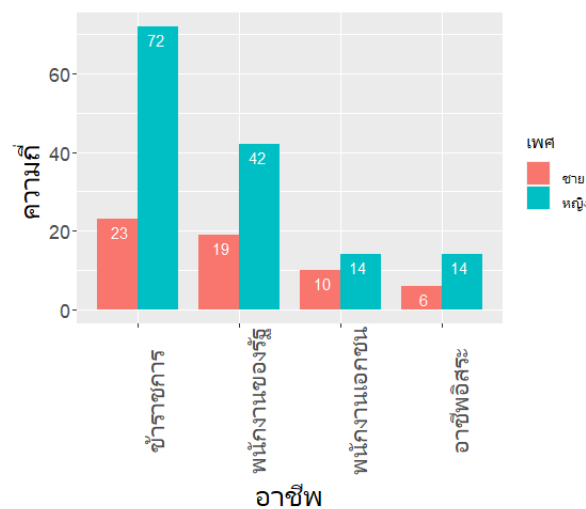
รูปที่ 1.1: แผนภูมิรูปวงกลมของระดับการศึกษา



รูปที่ 1.2: แผนภูมิแท่งของระดับการศึกษา

จากรูป 1.1 และ 1.2 จะเห็นว่าผู้ที่มีระดับการศึกษาด้านมัธยมศึกษาจำนวนมากที่สุด นอกจากการสร้างแผนภาพสำหรับการนำเสนอข้อมูลตัวแปรเดียวแล้ว ยังสามารถสร้างแผนภาพสำหรับการนำเสนอข้อมูลจำแนกตามตัวแปรสองตัว ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1.3 แผนภูมิแท่งจำแนกข้อมูลตามเพศและอาชีพ



รูปที่ 1.3: แผนภูมิแท่งจำแนกตามเพศและอาชีพ

จากรูปที่ 1.3 จะเห็นว่าผู้ที่ใช้บริการในโรงพยาบาลจำนวน 200 คน เป็นข้าราชการหญิงมากที่สุด รองลงมาคือพนักงานของรัฐหญิง และน้อยที่สุดคือผู้ที่ประกอบอาชีพอิสระชาย

1.3.2 ตารางความถี่และแผนภาพสำหรับข้อมูลเชิงปริมาณ

ตารางความถี่

การแจกแจงความถี่ (Frequency distribution) หมายถึงการจัดข้อมูลดิบให้อยู่ในรูปของตาราง โดยการใช้ชั้นหรือกลุ่ม (Class) และความถี่ (Frequency)

ฮิสโทแกรม

จากการพิจารณาลักษณะของข้อมูลโดยการสร้างตารางความถี่แล้ว การสร้างฮิสโทแกรมเป็นอีกวิธีหนึ่งที่ทำให้เราเห็นลักษณะหรือรูปแบบของข้อมูลได้ง่ายมากขึ้น โดยฮิสโทแกรมนั้นสร้างจากตารางความถี่

ฮิสโทแกรม (Histogram) คือแผนภาพที่แสดงถึงลักษณะของข้อมูลโดยการใช้ความสูงของแท่ง แผนภาพแสดงความถี่ของชั้นหรือของกลุ่มข้อมูล โดยแต่ละแท่งแผนภาพจะอยู่ติดกัน

ลักษณะของฮิสโทแกรม

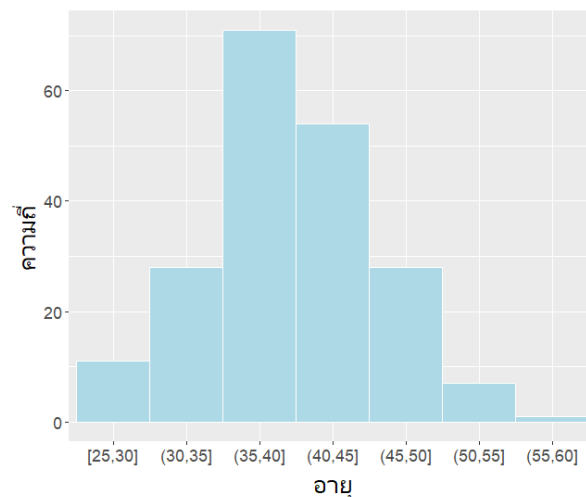
1. มาตรา (Scale) ในแนวนอนของฮิสโทแกรมเป็นค่าเชิงปริมาณและเป็นค่าของข้อมูล
2. มาตราในแนวตั้งของฮิสโทแกรมแสดงความถี่ของชั้น
3. แท่งของแผนภาพที่ต่อเนื่องกันต้องสัมผัสกัน

เนื่องจากแท่งแผนภาพที่ต่อเนื่องกันต้องสัมผัสกัน แท่งแผนภาพต้องเริ่มและสิ้นสุดที่ขอบชั้น (Class boundary) ในการกำหนดค่าในแกนแนวนอนนั้นอาจใช้ค่าขอบเขตของชั้นหรือค่ากึ่งกลางชั้นก็ได้

ตัวอย่างที่ 1.4 จากข้อมูลเกี่ยวกับผู้ใช้บริการในโรงพยาบาลในตารางที่ 1.1

ตารางที่ 1.6: ตารางความถี่ของอายุของผู้ใช้บริการโรงพยาบาล

อายุ	ความถี่	ความถี่สะสม	ร้อยละ	ร้อยละสะสม
[25,30]	11	11	5.5	5.5
(30,35]	28	39	14.0	19.5
(35,40]	71	110	35.5	55.0
(40,45]	54	164	27.0	82.0
(45,50]	28	192	14.0	96.0
(50,55]	7	199	3.5	99.5
(55,60]	1	200	0.5	100.0
รวม	200		100.0	



รูปที่ 1.4: ฮิสโทแกรมของอายุ

จากตารางที่ 1.6 และแผนภาพในรูปที่ 1.4 จะเห็นว่าค่าร้อยละสูงที่สุดอยู่ในชั้น 35 - 40 ซึ่งหมายความว่าผู้ที่มาใช้บริการในโรงพยาบาลแห่งนี้มีอายุระหว่าง 35 - 40 ปี มีจำนวนมากที่สุดคิดเป็นร้อยละ 35.5 รองลงมาคืออายุระหว่าง 40 - 45 ปี คิดเป็นร้อยละ 27

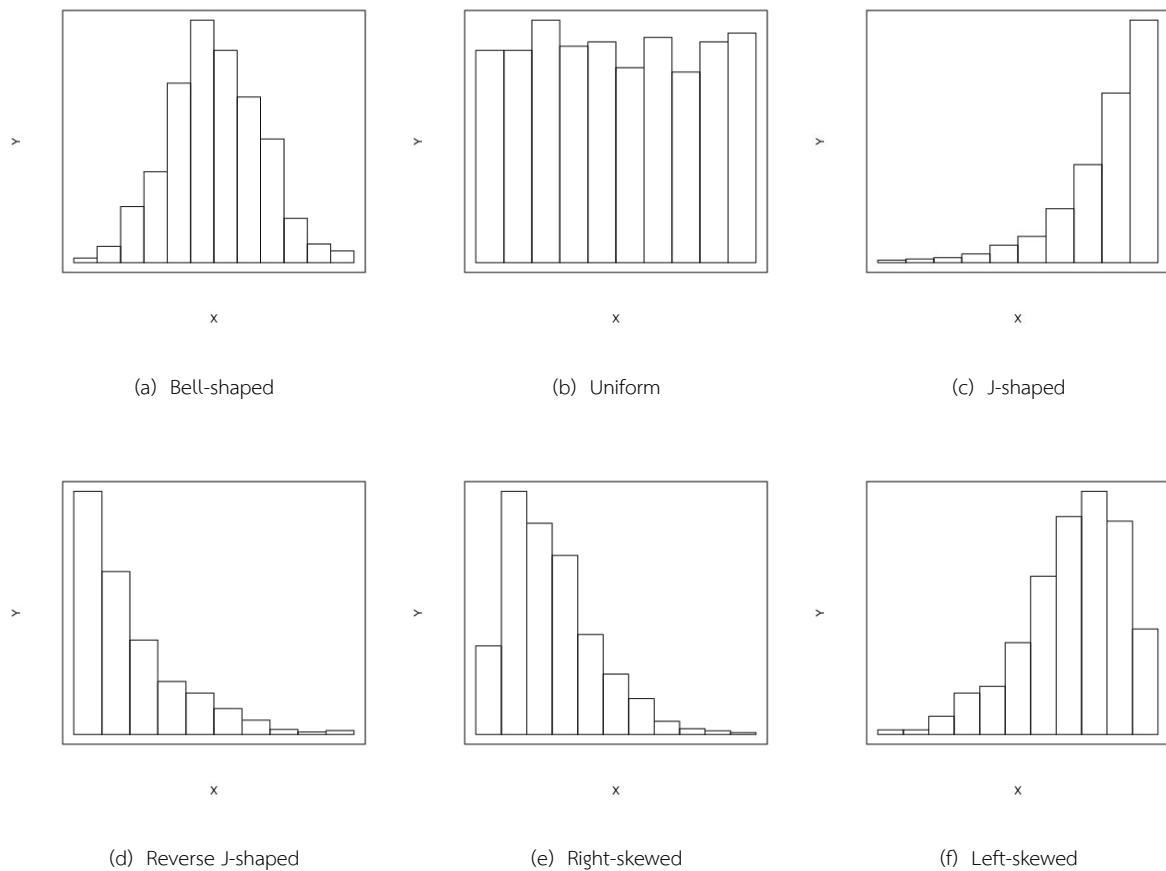
รูปร่างของการแจกแจง

เมื่อทำการอธิบายข้อมูลแล้ว ผู้ศึกษาจำเป็นต้องทราบถึงรูปร่างของการแจกแจงของข้อมูล ซึ่งในบทต่อ ๆ ไปจะเห็นว่ารูปร่างของการแจกแจงจะทำให้ทราบถึงวิธีเชิงสถิติที่จะนำมาใช้ในการวิเคราะห์ข้อมูล

การแจกแจงของข้อมูลมีอยู่ด้วยกันหลายรูปแบบ และวิธีที่ใช้ในการวิเคราะห์การแจกแจงของข้อมูลคือการสร้างฮิสโทแกรมของข้อมูล รูปร่างทั่ว ๆ ไปของการแจกแจงแสดงในรูปที่ 1.5 ซึ่งมีทั้งแบบรูประฆัง (Bell-shaped) แบบเอกรูป (Uniform-shaped) แบบ J (J-shaped) แบบ J กลับด้าน (Reverse J-shaped) แบบบวกหรือเบ้ขวา (Positive or right-skewed shaped) แบบลบหรือเบ้ซ้าย (Negative or left-skewed shaped) เป็นต้น

โดยปกติการแจกแจงของข้อมูลมักไม่ได้มีรูปร่างตรงตามรูปแบบอย่างสมบูรณ์ ดังนั้น เมื่อพิจารณาถึงรูปร่างของการแจกแจงของข้อมูลใด ๆ จึงพิจารณาในภาพรวมเท่านั้น

นอกจากรูปร่างเหล่านี้แล้ว การแจกแจงยังมีรูปร่างอื่น ๆ อีก อย่างไรก็ตาม มีลักษณะของรูปร่างของการแจกแจงบางประการที่ต้องพิจารณาในการวิเคราะห์ข้อมูล นั่นคือเมื่อวิเคราะห์ฮิสโทแกรม ต้องสังเกตรูปร่างของแผนภาพ เช่น แผนภาพนั้นมีจุดสูงสุดจุดเดียวหรือสองจุด แผนภาพมีลักษณะแบนหรืออยู่ในลักษณะรูปตัว U ข้อมูลมีการกระจายมากหรือไม่ในแผนภาพ หรือข้อมูลมีการกระจายตัวเป็นกลุ่มรอบจุดกึ่งกลางของแผนภาพ ข้อมูลมีค่าที่สูงหรือต่ำผิดปกติที่ปลายทั้งสองด้านของแผนภาพหรือไม่ ซึ่งค่านี้อาจเป็นค่านอกเกณฑ์ (Outlier) ได้ มีช่องว่างระหว่างแท่งแผนภาพในฮิสโทแกรมหรือไม่ สุดท้ายข้อมูลมีการรวมกันเป็นกลุ่มที่ปลายด้านใดด้านหนึ่งของแผนภาพหรือไม่ (ชี้ให้เห็นถึงการแจกแจงแบบเบ้)



รูปที่ 1.5: รูปร่างของการแจกแจง

แผนภาพลำต้นและใบ

แผนภาพลำต้นและใบ (Stem and leaf plot) เป็นวิธีหนึ่งที่ใช้ในการจัดรูปแบบและนำเสนอข้อมูลเชิงปริมาณในรูปแบบกราฟิกคล้ายกับฮิสโทแกรม โดยการรวมการเรียงลำดับข้อมูลและแผนภาพเข้าด้วยกัน แผนภาพชนิดนี้มีข้อดีคือมีการแสดงค่าของข้อมูลที่แท้จริงพร้อมกับการแสดงรูปร่างการแจกแจงของข้อมูลโดยยังคงรักษาค่าข้อมูลดั้งเดิมไว้

แผนภาพลำต้นและใบ คือแผนภาพที่สร้างโดยการใช้ส่วนหนึ่งของค่าของข้อมูลเป็นลำต้น (Stem) และอีกส่วนหนึ่งของค่าของข้อมูลเป็นใบ (Leaf) เพื่อสร้างกลุ่มหรือชั้นของข้อมูลขึ้น

สำหรับแผนภาพลำต้นและใบ ค่าของข้อมูลแต่ละค่าจะถูกแบ่งเป็นสองส่วนคือลำต้น (ตัวเลขที่อยู่ด้านซ้ายสุด) และใบ (ตัวเลขที่อยู่ทางขวาสุด) แผนภาพแบบนี้จะมีลักษณะคล้ายกับฮิสโทแกรม โดยจะแสดงถึงรูปร่างการแจกแจงของข้อมูล และมีข้อดีคือแผนภาพชนิดนี้แสดงให้เห็นข้อมูลที่แท้จริง และมีการเรียงลำดับข้อมูลอีกด้วย นอกจากนี้ แผนภาพลำต้นและใบ ยังแสดงให้เห็นถึงข้อมูลที่มีค่านอกเกณฑ์ได้

การสร้างแผนภาพลำต้นและใบ

1. เรียงลำดับค่าของข้อมูลจากน้อยไปหามาก
2. แยกค่าของข้อมูลตามตัวเลขหลักที่อยู่ทางซ้ายสุด
3. สร้างแผนภาพโดยใช้ตัวเลขหลักที่อยู่ทางซ้ายสุดเป็นต้น และตัวเลขของหลักถัดไปเป็นใบ

ตัวอย่างที่ 1.5 ข้อมูลในตารางที่ คือข้อมูลการใช้อินเทอร์เน็ต (Time) ของนิสิตจำนวน 50 คน

ตารางที่ 1.7: เวลาในการใช้อินเทอร์เน็ตของนิสิตจำนวน 50 คน

Time	gender	Time	gender	Time	gender
50	ชาย	41	หญิง	18	ชาย
40	ชาย	78	หญิง	29	ชาย
41	ชาย	56	หญิง	34	หญิง
17	หญิง	72	ชาย	59	ชาย
11	หญิง	56	ชาย	73	ชาย
7	หญิง	17	ชาย	77	หญิง
22	ชาย	7	หญิง	36	ชาย
44	หญิง	69	ชาย	39	ชาย
28	ชาย	30	ชาย	30	หญิง
21	ชาย	80	หญิง	62	หญิง
19	หญิง	56	หญิง	54	หญิง
23	หญิง	29	หญิง	67	หญิง
37	หญิง	33	หญิง	39	หญิง
51	หญิง	46	หญิง	31	ชาย
54	ชาย	31	หญิง	53	หญิง
42	ชาย	39	ชาย	44	หญิง
88	ชาย	20	หญิง		

รูปที่ 1.6 และ 1.6 แสดงแผนภาพลำต้นและใบ ของข้อมูลเวลาในการใช้อินเทอร์เน็ตของนิสิตจำนวน 50 คน

จากรูปที่ 1.6 จะเห็นได้ว่าข้อมูลชุดนี้มีการแจกแจงใกล้เคียงรูประฆัง โดยนิสิตจำนวน 11 คน ใช้อินเทอร์เน็ตเป็นระยะเวลา 30 - 39 นาทีต่อครั้ง และนิสิตอีก 9 คน ใช้เวลา 50 - 59 นาทีต่อครั้ง ซึ่ง 2 ช่วงเวลานี้เป็นเวลา ที่นิสิตใช้อินเทอร์เน็ตมากที่สุด สำหรับรูปที่ 1.6 เป็นแผนภาพลำต้นและใบของข้อมูลเวลาในการใช้อินเทอร์เน็ต

ของนีสิตจำแนกตามเพศ ซึ่งจะเห็นได้ว่าเวลาในการใช้อินเทอร์เน็ตของนีสิตชายมีลักษณะการแจกแจงเบ้ขวา ในขณะที่ของนีสิตหญิงมีลักษณะการแจกแจงคล้ายระฆัง และระยะเวลาที่นีสิตชายและหญิงก็มีการใช้อินเทอร์เน็ตมากที่สุด คือ 30 - 39 นาที

```

1 | 2: represents 12
leaf unit: 1
      n: 50
  2    0 | 77
  7    1 | 17789
 14    2 | 0123899
(11)   3 | 00113467999
(7)    4 | 0112446
 18    5 | 013446669
  9    6 | 279
  6    7 | 2378
  2    8 | 08

```

รูปที่ 1.6: แผนภาพลำต้นและใบของเวลาในการใช้อินเทอร์เน็ตของนีสิต

1 2: represents 12, leaf unit: 1				
		male	female	
		0 77		2
2	87	1 179		5
6	9821	2 039		8
(5)	99610	3 013479	(6)	
(3)	210	4 1446	(4)	
8	9640	5 13466	10	
4	9	6 27	5	
3	32	7 78	3	
1	8	8 0	1	
		9		
n:	22		28	

รูปที่ 1.7: แผนภาพลำต้นและใบของเวลาในการใช้อินเทอร์เน็ตจำแนกตามเพศของนีสิต

แผนภาพกล่อง

แผนภาพกล่อง (Box plot หรือ Box-and-Whisker plot) คือแผนภาพที่ใช้แสดงการกระจายของข้อมูล ซึ่งช่วยให้มองเห็นค่ากลางและการกระจายของข้อมูลได้ง่าย โดยมีส่วนประกอบ ดังนี้

1. **กล่อง (Box)** แสดงค่าควอร์ไทล์ที่ 1 (Q_1) และค่าควอร์ไทล์ที่ 3 (Q_3) โดยความยาวของกล่องแสดงถึงค่าพิสัยระหว่างควอร์ไทล์ (Interquartile range) ที่เป็นค่าวัดการกระจายของข้อมูลร้อยละ 50 ในช่วงกลาง