บทที่ 1
สถิติเชิงพรรณา
(DESCRIPTIVE STATISTICS)

1.1 ทำไมจึงต้องเรียนสถิติ

วิชาสถิติเป็นเครื่องมือสำหรับการตัดสินใจและช่วยประกอบการวางแผนที่สำคัญอย่างหนึ่ง ในกิจกรรมต่าง ๆ แทบทุกสาขา ไม่ว่าจะเป็นด้านวิทยาศาสตร์ สังคมศาสตร์ รัฐศาสตร์ การแพทย์ การเกษตร เศรษฐกิจ ธุรกิจ อุตสาหกรรม วิศวกรรม ฯลฯ สถิติจะสัมพันธ์กับชีวิตประจำวันอยู่เสมอ สื่อสารมวลชน ไม่ว่าจะเป็นหนังสือพิมพ์ วิทยุ โทรทัศน์ จะมีข้อมูลเกี่ยวกับสถิติต่าง ๆ แทบทุกวัน เช่นสถิติเกี่ยวกับการเกิด การตาย สถิติการเพิ่มผลผลิต สถิติสินค้าส่งออก สถิติอุบัติเหตุบนถนน สถิติแสดงจำนวนอาชญากรรม สถิติผลิตผลการเกษตร สถิติแสดงปริมาณน้ำฝน สถิติคนเป็นโรคต่าง ๆ ฯลฯ.

ทฤษฎีสถิติ ไม่ว่าจะเป็นทฤษฎีความน่าจะเป็น (Probability Theory) ทฤษฎีสถิติ (Theory of Statistics) ฯลฯ รวมทั้งหลักวิชาสถิติอื่น ๆ จัดว่ามีความสำคัญและเป็นประโยชน์ต่อทุกวงการและทุกระดับ ไม่ว่าจะเป็น หน่วยงาน องค์การ ประเทศ หรือระดับนานาชาติ ถ้าข้อมูลที่ได้มาเป็นข้อมูลจริง ได้มาจากตัวอย่างหรือประชากรซึ่งเก็บและรวบรวมโดยใช้หลักวิชาทางสถิติ นำไปวิเคราะห์โดยอาศัยทฤษฎีต่าง ๆ ที่เหมาะสมแล้ว จะมีประโยชน์ในด้านการวางแผนงาน (planning) และช่วยประกอบการวินิจฉัยสั่งการ (Decision making) เป็นอันมาก ในการทำวิจัย นักวิจัยจำเป็นจะต้องมีความรู้ทางสถิติ ทั้งนี้เพราะการวิจัยจะต้องเกี่ยวข้องกับการรวบรวมข้อมูล วิเคราะห์ข้อมูล ตลอดจนการตีความหมายข้อมูลเพื่อให้เป็นข้อสรุปในการวิจัย

ด้วยเหตุที่วิชาสถิติมีประโยชน์ต่อการค้นคว้าทางด้านวิทยาศาสตร์ เพิ่มผลผลิตทางด้านอุตสาหกรรม ควบคุมทางด้านเศรษฐกิจ และกิจกรรมอื่น ๆ ที่เกี่ยวข้องกับสถิติดังกล่าวข้างต้น ซึ่งจะช่วยในการบริหารรัฐ ทำให้ประเทศชาติพัฒนาไปด้วยความมั่นคงและปลอดภัย จึงเป็นสิ่งจำเป็นที่นักศึกษาและบุคคลทั่วไปควรจะได้มีความรู้ทางด้านสถิติ

1.2 ประวัติวิชาสถิติ

สถิติ ตรงกับคำในภาษาอังกฤษว่า "Statistics" ซึ่งเป็นคำที่ตรงกับคำในภาษาเยอรมันว่า "statistik" โดยที่ทั้งสองคำนี้มาจากรากศัพท์เดียวกัน คือ "state" ในปี ค.ศ. 1749 นักปราชญ์ชาวเยอรมันชื่อ Gottfried Achenwall ได้ให้นิยามความหมายของคำว่า "state" ว่าหมายถึงข้อมูลที่เกี่ยวข้องและเป็นประโยชน์ในการบริหารรัฐ ซึ่งเป็นรัฐศาสตร์แขนงหนึ่ง โดยที่ศาสตร์ดังกล่าวนี้มีมาตั้งแต่สมัยโรมัน เรียกเป็นภาษาลาตินว่า "ratio status" ซึ่งหมายถึงการนำเอาข้อเท็จจริงมาแยกแยะนำไปวิเคราะห์ เพื่อช่วยประกอบการบริหารงานของรัฐ ถ้าจะให้ความหมายของ "statistics" ในสมัยดั้งเดิมว่า "the fact of state" ก็คงไม่ผิดนัก ในศตวรรษที่ 16 ประเทศในยุโรปได้เริ่มสนใจในการรวบรวมข้อมูลเกี่ยวกับประชากรและข่าวสารความเป็นไปเกี่ยวกับรัฐ ทั้งนี้ก็เพราะนักปกครองสมัยนั้นพิจารณาเห็นว่า การที่จะให้การบริหารงานและบริการรัฐสัมฤทธิ์ผลได้อย่างสมบูรณ์นั้นจำเป็นที่รัฐจักต้องมีข้อมูล ข้อสนเทศ เกี่ยวกับประชากรอย่างถูกต้องและเพียงพอ และด้วยเหตุที่นักปกครองได้นำเอาหลักการทางสถิติศาสตร์มาใช้ก่อน ถ้าจะกล่าวว่าสถิติศาสตร์กำเนิดมาจากนักปกครองก็คงจะได้.

ในตอนต้นของศตวรรษที่ 17 มีนักคณิตศาสตร์หลายคนได้นำเอาหลักทางคณิตศาสตร์มาประยุกต์ใช้กับการพนัน โดยคำนวณโอกาสที่จะเอาชนะคู่ต่อสู้ ซึ่งหลักการนี้ต่อมาเป็นที่รู้จักกันในนามของทฤษฎีความน่าจะเป็น (Probability Theory)

ในคริสตศักราช 1733 อับราแฮม เดอ มอนว์ (Abraham De Moivre : 1667 - 1754) ได้ค้นพบเส้นโค้งปกติ (normal curve) Carl Friedrick Gauss (1777 - 1855) และ Laplace ได้นำเอาโค้งปกติมาอธิบายการกระจายเกี่ยวกับการวัด ซึ่งเป็นต้นตำรับของทฤษฎีว่าด้วยความคลาดเคลื่อน (Theory of Error) และ Bernoulli ได้ค้นพบทฤษฎี การโน้มเข้าสู่เกณฑ์กลาง (Central Limit Theorem).

ชาวเบลเยี่ยมชื่อ Adolph Quelelet (1796-1874) เป็นคนแรกที่ประยุกต์เอาวิธีการใหม่ ๆ มาใช้ในการรวบรวมข้อมูล (collection of data) ต่อมาในปี ค.ศ. 1893 ชาวอังกฤษสองท่าน คือ Francis Galton (1822 - 1911) และ Karl Pearson (1857-1936) เป็นผู้ที่วางแนวคิดสำคัญ ๆ ทางทฤษฎีและเทคนิคทางสถิติ ซึ่งในยุคนี้มีวิวัฒนาการทางสถิติแผนใหม่กว้างขวางที่สุด เช่น หลักในการทดสอบสมมุติฐาน การใช้ค่าสหสัมพันธ์เชิงเส้น สหสัมพันธ์พหุคูณ การวิเคราะห์ความแปรปรวน และการวางแผนแบบทดลอง เป็นต้น ต่อมาในศตวรรษที่ 20 ผู้ที่มีส่วนในการวางรากฐานให้แก่สถิติสมัยใหม่คือ Sir Ronald Fisher (1890-1962) Jersey Neyman และ Egon Pearson เป็นต้น.

วิวัฒนาการทางสถิติศาสตร์ยังก้าวหน้าต่อไป ยิ่งสภาพสังคมมีความสลับซับซ้อนมากเพียงใด สถิติก็ยิ่งมีความสำคัญมากขึ้นเท่านั้น.

1.3 ประเภทของสถิติ

สถิติในความหมายที่เป็นศาสตร์สามารถแยกออกได้ 3 ประเภท คือ

1.3.1 สถิติเชิงพรรณา (Descriptive Statistics)

เป็นสถิติที่เกี่ยวกับการวิเคราะห์ข้อมูลในรูปของการบรรยาย หรือเป็นการแปลความหมายข้อมูลในรูป บทความ ตารางแผนภูมิ กราฟ รูปภาพ รวมทั้งระเบียบวิธีเบื้องต้นเกี่ยวกับความหมายและการคำนวณหาค่าเฉลี่ยเลขคณิตและความเบี่ยงเบนมาตรฐาน นอกจากนี้ยังเกี่ยวกับการใช้อัตราส่วน เพื่อวิเคราะห์เชิงเปรียบเทียบข้อมูลในรายการต่าง ๆ

1.3.2 สถิติเชิงอนุมาน (Inferential Statistics)

เป็นสถิติที่เกี่ยวกับการคาดหมายประมาณการหรือทดสอบเกี่ยวกับค่าที่แสดงคุณลักษณะทางประชากร (พารามิเตอร์) โดยอาศัยข้อมูลจากเพียงบางส่วนของกลุ่มประชากร (Sample) มาคาดหมายค่าที่แสดงคุณลักษณะดังกล่าว ที่เป็นเช่นนี้ก็เพราะในทางปฏิบัตินั้นเราไม่อาจวัดหรือตรวจสอบข้อมูลจากประชากรทั้งกลุ่มได้.

เหตุที่สถิติอนุมานเป็นเรื่องที่เกี่ยวกับการใช้ข้อมูล ข้อสนเทศ จากกลุ่มตัวอย่างไปอนุมานสู่กลุ่มประชากร ดังนั้นการศึกษาสถิติอนุมานจึงเกี่ยวข้องอยู่กับทฤษฎีความน่าจะเป็นและทฤษฎีสถิติเป็นอย่างมาก และในทางปฏิบัติเราจำเป็นต้องใช้สถิติอนุมานอยู่ตลอดเวลา นักศึกษาจะได้ศึกษาเรื่องราวของสถิติคืออนุมานในบทที่ 2,3,4 และ 5.

1.3.3 สถิติปฏิบัติ

เป็นสถิติเกี่ยวกับระเบียบวิธีของกระบวนการต่าง ๆ ใช้ในการเก็บรวบรวมข้อมูลสถิติ เช่นใช้ในการเก็บข้อมูลโดยวิธีสัมมะโน เก็บจากตัวอย่างหรือจากระเบียน ระเบียบการประมวลผลข้อมูล และระเบียบการวิเคราะห์ผลการสำรวจ เป็นต้น

1.4 ความหมายและขอบข่ายของวิชาสถิติ

1.4.1 ความหมายของวิชาสถิติ

ในความหมายแรกนั้นหมายถึง ตัวเลขหรือยอดรวมที่แทนข้อเท็จจริงต่าง ๆ เช่นตัวเลขแสดงอัตราการเกิดการตายของประชากร อุณหภูมิของอากาศ ปริมาณน้ำฝน ผลผลิตทางการเกษตร เป็นต้น สถิติในอีกความหมายหนึ่งซึ่งถือว่าเป็นศาสตร์ทั้งวิทยาศาสตร์และศิลปศาสตร์เป็นขบวนการที่เกี่ยวข้องกับระเบียบวิธีและเทคนิคต่าง ๆ ดังต่อไปนี้

1) การเก็บรวบรวมข้อมูล (collection of data)
2) การนำเสนอข้อมูล (presentration of data)
3) การวิเคราะห์ข้อมูล (analysis of data)
4) การตีความหมายหรือหาข้อสรุปข้อมูล (interpretation of data)

1.4.2 ขอบข่ายของวิชาสถิติ

วิชาสถิติมีขอบข่ายเกี่ยวข้องกับขบวนการต่าง ๆ ดังนี้

1) การวางแผนการทดลอง (experimental design) หรือการวางแผนสำรวจ (servey design) เพื่อให้ได้ข้อมูลตามความต้องการ มีความแม่นยำและรวดเร็วทันเวลา.

2) การรวบรวมข้อมูล (collecting of data) เป็นการรวบรวมข้อมูลจากการทดลอง การสำรวจ หรือการสังเกตการณ์ ข้อมูลที่ได้อาจเป็นข้อมูลเชิงปริมาณ ซึ่งวัดค่าเป็นตัวเลขหรือปริมาณต่าง ๆ หรือข้อมูลเชิงคุณภาพที่แสดงข้อมูลในรูปของนามธรรม เช่น ดี เลว ชอบ ไม่ชอบ เฉย ๆ เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง เพศ สถานภาพสมรส การอ่านออกเขียนได้ หรืออื่น ๆ

3) การเสนอผลการลงตาราง (tabulating) เป็นการจัดจำพวกสร้างตารางแสดงความสัมพันธ์ระหว่างรายการของข้อมูลที่เก็บมา หรือจำแนกข้อมูลตามตัวแปรที่กำหนดไว้

4) การประมาณค่า (estimation) ได้แก่การคำนวณค่าต่าง ๆ ที่ต้องการทราบ เช่น ยอดรวม ค่าเฉลี่ย อัตราส่วน เปอร์เซนต์ เป็นต้น โดยทำข้อสรุปจากข้อมูลที่ประมวลผลแล้ว

5) การวิเคราะห์ (analysis) เป็นการตรวจสอบความน่าเชื่อถือของค่าต่าง ๆ ที่ได้จากการประมาณ

6) การแปลความหมายและทำข้อสรุป (interpretation and drawing conclusion) ข้อมูล

ที่ประมวลผล ประมาณค่าและวิเคราะห์แล้ว จะต้องนำมาตีความและสรุปผลทั้งตรวจสอบความถูกต้อง ความสมบูรณ์ของขั้นตอนต่าง ๆ.

1.5 **การเก็บรวบรวมข้อมูล** (collection of data)

ก่อนอื่นนักศึกษาควรจะได้ทำความเข้าใจความหมายของกลุ่มประชากร (population) เสียก่อน ในความหมายของคนโดยทั่วไปประชากรหมายถึงคนหรือกลุ่มคน เช่นประชากรของประเทศ แต่ในความหมายทางสถิติ คำว่าประชากรกินความหมายกว้างกว่าความหมายดังกล่าว โดยถือว่ากลุ่มประชากรหมายถึงกลุ่มของหน่วยเบื้องต้น (Collection of elementary Unit) หรือกลุ่มของหน่วยแจงนับ (Collection of enumeration Unit) เมื่อเป็นเช่นนี้ความหมายของกลุ่มประชากรจึงมีความหมายครอบคลุมไปถึงทั้งคน สัตว์ และวัตถุสิ่งของ ทั้งนี้ขึ้นอยู่กับวัตถุประสงค์ของงานว่าจะเก็บรวบรวมข้อมูล อะไร จากใคร เป็นสำคัญ เช่นต้องการรวบรวมข้อมูลเกี่ยวกับทัศนคติของครูในเขตกรุงเทพมหานครต่อบทบาทสตรี กลุ่มประชากรก็คือกลุ่มครูในเขตกรุงเทพ-มหานคร บริษัทผู้ผลิตแบตเตอรี่ต้องการตรวจสอบอายุการใช้งานแบตเตอรี่ที่ผลิต กลุ่มประชากรก็คือแบตเตอรี่ทั้งหมดที่บริษัทดังกล่าวผลิตขึ้นมา ถ้าต้องการทราบว่าโคนมที่เลี้ยงในประเทศไทยให้นมมากน้อยเพียงใด กลุ่มประชากรก็คือโคนมทั้งหมดที่เลี้ยงในประเทศไทย ดังนี้เป็นต้น.

ข้อมูลที่จะจัดเก็บรวบรวมหมายถึงอะไรก็ได้ ขึ้นอยู่กับวัตถุประสงค์ของงาน แต่โดยทั่วไปข้อมูลคือรายละเอียดด้านต่าง ๆ ของหน่วยเบื้องต้น ข้อมูลที่ได้รับอาจเป็นได้ทั้งข้อมูลเชิงปริมาณ (Quantitative data) ซึ่งเป็นรายละเอียดของหน่วยเบื้องต้น ที่วัดค่าออกมาเป็นจำนวนหรือปริมาณ ตามหน่วย ชั่ง ตวง วัด เช่น ความสูง น้ำหนัก ปริมาณนิโคตินในบุหรี่ จำนวนแบคทีเรียในน้ำ 1 ซี.ซี. และอื่น ๆ หรืออาจจะเป็นข้อมูลเชิงคุณภาพ (Qualitative data) ซึ่งเป็นรายละเอียดของหน่วยเบื้องต้นที่ประเมินออกมาในรูปของนามธรรม เช่น ดี เลว เห็นด้วย ไม่เห็นด้วย เป็นต้น ส่วนใหญ่งานวิจัยเกี่ยวกับทัศนคติหรือความคิดเห็นนิยมประเมินเป็นข้อมูลเชิงคุณภาพ.

โดยทั่วไปเรานิยมเก็บรวบรวมข้อมูลด้วยวิธีการต่าง ๆ ดังต่อไปนี้ อาจใช้แบบใดแบบหนึ่ง หรือผสมกัน คือ

1.5.1 **การสำมะโน** (cencus หรือ complete enumeration)

การสำมะโนหมายถึงการเก็บรวบรวมข้อมูลจากหน่วยเบื้องต้นทุกหน่วย เช่น การ

สำมะโนประชากร สำมะโนการเกษตร และสำมะโนเศรษฐกิจและสังคม

การสำมะโนเป็นงานที่สิ้นเปลืองกำลังแรงงาน ทุนทรัพย์ และเวลามาก เช่น การสำมะโนประชากรครั้งหนึ่ง ๆ ต้องใช้งบประมาณ 10-30 ล้านบาท ใช้บุคลากรเป็นแสน ๆ คน และกินเวลาในการรวบรวมข้อมูลหลายเดือน และยังต้องใช้เวลาประมวลผลเป็นปี เช่นสำมะโนประชากรในปี 2513 ทราบผลการสำมะโนปี 2516 ต้องใช้เวลาถึง 3 ปี ดังนี้เป็นต้น ดังนั้นการสำมะโนจึงเป็นวิธีปฏิบัติที่กระทำเฉพาะในงานที่จำเป็นจริง ๆ และในกรณีที่รัฐต้องอาศัยข้อมูลเหล่านั้นมาใช้ในกิจกรรมบริหารงาน เพื่อวางแผนพัฒนาและกิจการอื่น ๆ ที่จำเป็นเท่านั้น.

1.5.2 การสำรวจด้วยตัวอย่าง (Sample Survey)

การสำรวจด้วยตัวอย่างหมายถึงการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างซึ่งเป็นเพียงบางส่วนของกลุ่มประชากร แล้วนำผลลัพธ์ที่ได้ไปคาดหมายพยากรณ์ค่าที่แสดงคุณลักษณะของประชากร (พารามิเตอร์) อีกทีหนึ่ง สถิติในส่วนนี้เรียกว่าสถิติอนุมาน ซึ่งความถูกต้องแม่นยำของการคาดหมายพยากรณ์ขึ้นอยู่กับปัจจัยหลายประการ ที่สำคัญก็คือแผนสำรวจ แผนการวิเคราะห์ สูตรที่ใช้วิเคราะห์ การแปลความหมาย และการควบคุมข้อบกพร่องต่าง ๆ

แผนสำรวจที่ยอมรับกันในทางทฤษฎีเท่าที่ปฏิบัติกันอยู่ในปัจจุบัน มีอยู่ 4 แผน ดังต่อไปนี้

1) Simple Random Sampling (SRS) เป็นแผนสำรวจที่ถือว่าหน่วยสำรวจ (Sampling Unit)[1] ทุกหน่วยมีโอกาสได้รับการเลือกมาเป็นตัวอย่างได้เท่า ๆ กัน หรืออีกนัยหนึ่ง SRS คือแผนสำรวจที่ทุกกลุ่มตัวอย่าง จากทุกตัวอย่างที่เป็นไปได้ มีโอกาสถูกเลือกมาใช้เท่า ๆ กัน

วิธีการในการเลือกตัวอย่างอาจทำโดยการจับฉลาก หรือใช้ตารางเลขสุ่ม (Random number) ก็ได้ แต่ในทางปฏิบัตินิยมใช้ตารางเลขสุ่ม เพราะสะดวกรวดเร็วและถูกต้องกว่า.

---

1. กลุ่มประชากรประกอบด้วยหน่วยพื้นฐาน เรียกว่า Elementary Unit แต่ในบางครั้งเราไม่อาจสำรวจกับ elementary Unit ได้ จำเป็นต้องจัด elementary Unit เป็นกลุ่มแล้วจึงค่อยสำรวจ เรียกว่า Sampling Unit เช่นการสำรวจรายได้ของเกษตรกร elementary Unit คือเกษตรกรแต่ละคน ซึ่งเวลาปฏิบัติจริงเราไม่อาจติดตามสำรวจได้ครบ จำเป็นต้องสำรวจกับครอบครัวเกษตรกร การสำรวจโรคของ โค กระบือ elementary Unit คือ โค กระบือ แต่ละตัว ในทางปฏิบัติเพื่อการสำรวจ เราจะใช้ คอก เป็นหน่วยสำรวจ แต่บางครั้ง elementary Unit กับ Sampling Unit ก็เป็นสิ่งเดียวกัน.

2) Stratified Random Sampling เป็นแผนสำรวจที่จำแนกประชากรออกเป็นกลุ่มประชากรย่อย (Stratum หรือ Subpopulation) ก่อน จากนั้นจึงเลือกตัวอย่างจากกลุ่มประชากรย่อยโดยวิธี SRS วิธีนี้มีประสิทธิภาพสูงกว่าวิธีที่ (1) คือ ถูกต้องแม่นยำกว่า และที่สำคัญ แบบนี้ช่วยให้เราทราบข้อมูลเฉพาะกลุ่มประชากรย่อยได้ ทั้งยังสามารถศึกษาเปรียบเทียบระหว่างกลุ่มประชากรย่อยได้อีกด้วย เช่นการสำรวจค่าใช้จ่ายส่วนตัวของนักศึกษาในมหาวิทยาลัยรามคำแหง จะเห็นว่าหน่วยสำรวจคือ นักศึกษาของมหาวิทยาลัยรามคำแหง (ข้อสังเกต : กรณีนี้หน่วยเบื้องต้นกับหน่วยสำรวจคือสิ่งเดียวกัน) วิธีสำรวจทำได้โดยจำแนกกลุ่มนักศึกษาออกตามคณะ หลังจากนั้นจึงสุ่มตัวอย่างโดยวิธี SRS จากแต่ละคณะ การกระทำเช่นนี้นอกจากทำให้ได้รับข้อมูลรวมของมหาวิทยาลัยแล้ว ยังได้ข้อมูลรายคณะ ทั้งยังสามารถศึกษาเปรียบเทียบระหว่างคณะอีกด้วย.

3) Systematic Sampling เป็นแผนสำรวจที่เลือกตัวอย่างมา 1 หน่วยในทุก ๆ K หน่วย เช่นเลือกตัวอย่าง 1 หน่วยทุก ๆ 10 หน่วย K คือจำนวนประชากรต่อจำนวนตัวอย่าง ซึ่งเท่ากับ $\frac{N}{n}$ N คือจำนวนประชากร n คือจำนวนตัวอย่าง วิธีนี้ใช้ได้ดีในกรณีที่กลุ่มประชากรมีขนาดใหญ่ ซึ่งอาจทำให้กลุ่มตัวอย่างที่เลือกได้ไม่กระจายไปทั่ว ๆ กลุ่มประชากร ถ้าใช้วิธี Systematic กลุ่มตัวอย่างจะกระจายไปทั่วทั้งประชากร เช่น ในการทำสำรวจภาวะเศรษฐกิจของทหารผ่านศึกในนิคมเกษตรกรรมชานุมาน จังหวัดอุบลราชธานี การสำรวจใช้วิธี Systematic ปรากฏว่านิคมดังกล่าวประกอบด้วยครอบครัวทหารผ่านศึก 315 ครอบครัว และผู้วิจัยได้กำหนดขนาดตัวอย่างไว้ประมาณ 34% หรือประมาณ 105 หน่วย ดังนั้น $K = \frac{N}{n} = \frac{315}{105} = 3$ นั่นคือการสำรวจจะสำรวจ 1 บ้านทุก ๆ 3 บ้าน หรือสำรวจ 1 บ้าน เว้น 2 บ้าน

4) Cluster Sampling คือแผนสำรวจที่แบ่งประชากรออกเป็นกลุ่มประชากรย่อย เรียกว่า cluster ก่อน การสุ่มตัวอย่างจะเริ่มด้วยการเลือก cluster ขึ้นมาโดยวิธี SRS จากนั้นจึงเลือกตัวอย่างโดยวิธี SRS เฉพาะ cluster ที่เลือกได้ดังกล่าว.

แผนสำรวจนี้เราสามารถเลือกตัวอย่างซ้อน ๆ กันไปหลาย ๆ ขั้นได้ ถ้ามีการเลือกเพียงสองขั้นเรียกว่า Two Stage Cluster Sampling ถ้าเลือก 3 ขั้นเรียกว่า Three-Stage Cluster Sampling ถ้าเลือก 4 ขั้นเรียกว่า Four-Stage Cluster Sampling.

ตัวอย่างเช่น ในการสำรวจลักษณะการวางแผนชีวิตของเด็กวัยรุ่นในเขตกรุงเทพมหานคร

โดยมุ่งเฉพาะนักเรียนชั้น ม.ศ.3 แผนสำรวจใช้ Three-Stage Cluster Sampling จากเขตการศึกษาใน กทม. มีอยู่ทั้งสิ้น 23 เขต ในการสำรวจมีขั้นตอนดังนี้

ขั้นที่ 1 เลือกเขตการศึกษาขึ้นมาเป็นตัวอย่าง
ขั้นที่ 2 สุ่มโรงเรียนในเขตการศึกษาที่เลือกไว้มาเป็นตัวอย่าง
ขั้นที่ 3 สุ่มนักเรียนจากโรงเรียนที่เลือกไว้มาเป็นตัวอย่าง

เหล่านี้เป็นเพียงหลักการกว้าง ๆ ของแผนสำรวจที่นิยมใช้กันอยู่ในทางปฏิบัตินอกจากวิธีทั้ง 4 นี้แล้ว เรายังสามารถใช้วิธีผสมกันได้อีก เช่นวิธีที่ 2 ผสมกับวิธีที่ 3 วิธีที่ 2 ผสมกับวิธีที่ 4 หรืออื่น ๆ ซึ่งจะไม่กล่าวในที่นี้.

อนึ่ง ยังมีแผนการเลือกตัวอย่างอีกแบบหนึ่งซึ่งนิยมใช้กันมาแต่โบราณกาลตั้งแต่ปีคริสตศักราช 1930 และยังนิยมใช้กันอยู่ในปัจจุบัน โดยเฉพาะในการวิจัยทางธุรกิจหรือวิจัยตลาด แต่เป็นวิธีที่นักสถิติไม่ยอมรับ เพราะมิได้อ้างอิงทฤษฎีความน่าจะเป็น กล่าวคือหน่วยสำรวจมีโอกาสได้รับเลือกขึ้นเป็นตัวอย่างอย่างไม่มีหลักเกณฑ์ แผนดังกล่าวคือ Qouta Sampling และ Judgement Sampling.

Qouta Sampling คือแผนสำรวจที่เลือกตัวอย่างโดยมิได้ยึดถือหลักเกณฑ์ความน่าจะเป็นมาใช้สำรวจ เมื่อกำหนดตัวอย่าง n **ตัวอย่าง และประชากรในข่ายสำรวจไว้เรียบร้อยแล้ว การสำรวจก็ดำเนินไปโดยพนักงานสำรวจเลือกหรือเสาะหาหน่วยตัวอย่างที่มีคุณลักษณะตามที่กำหนดไว้ แล้วสัมภาษณ์หรือสังเกต แล้วแต่กรณี เมื่อหาได้ครบ** n หน่วยก็ถือว่าเสร็จสิ้น นำข้อมูลมาวิเคราะห์ได้.

ตัวอย่างเช่น ในการสำรวจรสนิยมของผู้สูบบุหรี่ ซึ่งทางโรงงานยาสูบต้องการข้อมูลไปใช้ในการปรับปรุงคุณภาพของบุหรี่ ในกรณีเช่นนี้กลุ่มประชากรคือกลุ่มผู้สูบบุหรี่ ในทางปฏิบัติเราจะเห็นว่าเป็นการยากที่จะบันทึกรายชื่อบุคคลผู้สูบบุหรี่รวบรวมเป็นกลุ่มประชากรในข่ายสำรวจ (ทางสถิติเรียกว่ากรอบตัวอย่าง) ดังนั้น นักสำรวจจึงดำเนินการสำรวจโดยวิธีเสาะหาหน่วยสำรวจเอาเอง พบใครสูบบุหรี่ก็เข้าสอบถามหรือสังเกตสอบถามจนครบจำนวนที่ระบุไว้ในขนาดตัวอย่าง เช่น 3,000 คน ก็เป็นอันเสร็จสิ้นหรือในการสำรวจทัศนคติของผู้ติดยาเสพติดต่อวิธีรักษาโรคยาเสพติดแผนปัจจุบัน ณ โรงพยาบาลธัญญรักษ์ โดยมุ่งศึกษาเปรียบ

เทียบระหว่างกลุ่มอายุ ปรากฏว่าผู้ป่วยที่เข้ารับการรักษาพยาบาลในโรงพยาบาล เข้ารับการรักษาในระยะเวลาอันจำกัด และยังย้ายเข้าย้ายออกอยู่ตลอดเวลา จึงไม่อาจจำแนกผู้ป่วยตามกลุ่มอายุได้ นอกจากนี้ผู้ป่วยยังมีระดับอายุต่าง ๆ มีจำนวนไม่แน่นอน ส่วนใหญ่อายุไม่เกิน 35 ปี ทำให้การสร้างบัญชีรายชื่อผู้ป่วยจำแนกตามอายุเป็นไปโดยยาก และถึงแม้จะทำได้ก็ไม่ทันการ ถ้าจะคอยให้ได้จำนวนผู้ป่วยจำแนกตามอายุต่าง ๆ มากพอแล้วจึงค่อยสุ่มตัวอย่าง ก็กระทำไม่ได้ เพราะผู้ป่วยที่เข้ารับการรักษาจะย้ายออกเสียก่อน ดังนี้เมื่อกำหนดว่าจะสุ่มตัวอย่างกลุ่มละ 100 คน ในช่วงอายุไม่เกิน 24 ปี 25-34 ปี และ 45 ปีขึ้นไปการสุ่มตัวอย่างจึงใช้วิธีโควต้า คือถามอายุผู้ป่วยทุกคนที่ขอเข้ารับการรักษาก่อน เพื่อทราบว่าผู้ป่วยอยู่ในกลุ่มอายุใด แล้วจึงสัมภาษณ์เรื่องต่าง ๆ ตามแบบสอบถาม ดำเนินเช่นนี้ไปจนครบกลุ่มอายุละ 100 คน เป็นอันว่าเสร็จสิ้น

Judgment Sampling เป็นแผนการสำรวจที่ขึ้นอยู่กับวิจารณญาณของนักสำรวจเป็นสำคัญ โดยนักสำรวจจะกำหนดขึ้นเองจากประสบการณ์ว่าควรจะใช้ใครเป็นหน่วยสำรวจ และสำรวจเท่าไร วิธีนี้ไม่ยึดแบบแผนทางทฤษฎีแต่อย่างไร อาศัยประสบการณ์ของนักสำรวจเองเป็นสำคัญ ดังนั้นผู้ที่จะใช้วิธีนี้ได้ควรจะเป็นผู้ชำนาญการ หรือเกี่ยวข้องในงานแขนงที่จะสำรวจมานานพอจนรู้จักธรรมชาติของงานและประชากรในข่ายสำรวจดี เช่นนักปกครองอยากทราบพฤติกรรมทางการเมืองของราษฎร แทนที่เขาเองจะสัมภาษณ์โดยตรงจากราษฎร กลับสัมภาษณ์จากกำนันผู้ใหญ่บ้านแทน พนักงานควบคุมคุณภาพสินค้าส่งออก เช่น ข้าวและน้ำตาล แทนที่จะเลือกข้าวและน้ำตาลจากจุดต่าง ๆ ทั่ว ทั้งกระสอบ กลับเลือกข้าวและน้ำตาลจากปากกระสอบหรือเจาะกลางกระสอบ ครูอยากรู้ปัญหาและความต้องการของนักเรียน แทนที่จะสอบถามนักเรียนทั่ว ๆ ไป ก็กลับสอบถามจากหัวหน้าชั้น ดังนี้เป็นต้น

แผนสำรวจทั้งสองนี้ แม้จะไม่ใช้วิธีทางทฤษฎีความน่าจะเป็นในการเลือกตัวอย่าง แต่ในการประมวลผลกลับใช้สูตรการวิเคราะห์ต่าง ๆ ของแผนทั้ง 4 ข้างต้น ซึ่งแย้งกันระหว่างหลักปฏิบัติและหลักทางทฤษฎีอยู่ในตัว

1.5.3 การทะเบียน (Registration)

การเก็บรวบรวมข้อมูลโดยการทะเบียนเป็นการเก็บข้อมูลจากแหล่งที่มีการบันทึกข้อมูลไว้แล้ว ส่วนมากจะเป็นหน่วยงานราชการ ข้อมูลจากทะเบียนนั้นบางประเภทก็สมบูรณ์ดี เช่น ทะเบียนคนเกิด คนตาย ซึ่งเมื่อมีการเกิดการตายจะต้องแจ้งต่อนายทะเบียนท้องถิ่นเพื่อบันทึกลงทะเบียน ทะเบียนรถยนต์ก็จัดว่าสมบูรณ์อย่างหนึ่ง รถยนต์ทุกคันจะต้องนำไปขึ้นทะเบียน

เลขเครื่อง เลขตัวถัง และเจ้าของรถ นอกจากนี้ก็มีทะเบียนบ้าน ทะเบียนทหาร ทะเบียนคนต่างด้าว ทะเบียนปืน เป็นต้น ทะเบียนบางชนิดที่ไม่ได้บันทึกทุกหน่วยของประชากร เช่นทะเบียนสมรส บางคนแต่งงานแล้วไม่ได้ไปจดทะเบียนก็มีเป็นจำนวนมาก ทะเบียนบางประเภทก็มีการบิดเบือนความจริง เช่นทะเบียนการค้า ทุนในการขอจดทะเบียนมักไม่ตรงกับความเป็นจริง ทั้งนี้เพื่อหลีกเลี่ยงภาษีบ้าง เพื่อป้องกันการชดใช้หนี้สินในภายหลังถ้าประสบกับการขาดทุน ทะเบียนดังกล่าวข้างท้ายนี้จึงไม่เหมาะที่จะเก็บไปวิเคราะห์ โดยทั่วไปแล้วข้อมูลในทะเบียนจะทำการแก้ไขให้ทันสมัยอยู่เสมอ

การเก็บรวบรวมข้อมูลที่คล้ายกับการทะเบียนอีกอย่างหนึ่งก็คือการระเบียน (record) ซึ่งเป็นการรวบรวมข้อมูลหรือข้อความที่เป็นข้อเท็จจริงบันทึกไว้เป็นหลักฐาน อาจจะบันทึกไว้เป็นลายลักษณ์อักษร เป็นภาพยนตร์ หรือบันทึกเสียงไว้ เช่น ศิลาจารึกของพ่อขุน ระเบียนวิชาเรียน (Transcript) ของนักเรียน นักศึกษา ภาพยนตร์เหตุการณ์เมื่อสงครามโลกครั้งที่ 2 หรือการบันทึกเสียงคำปราศรัยของบุคคลสำคัญ เป็นต้น

การเก็บข้อมูลจากทะเบียนนั้น ถึงแม้ว่าจะง่ายและสะดวก แต่ก็มีข้อเสีย ถ้าการบันทึกลงทะเบียนไม่สมบูรณ์ ผิดพลาด ก็จะมีผลกระทบกระเทือนต่อการวิเคราะห์และสรุปผลได้

เมื่อรู้ถึงการเก็บรวบรวมข้อมูลแล้ว สิ่งที่ควรรู้อีกอย่างหนึ่งคือ **การออกแบบสอบถาม (Questionaire)**

แบบสอบถามนั้นเป็นเครื่องมือที่จะใช้ในการบันทึก ข้อมูล จากประชากร หรือกลุ่มตัวอย่าง ที่เป็นแหล่งข้อมูล เพื่อรวบรวมนำมาวิเคราะห์ วิธีในการเก็บรวบรวมข้อมูล อาจทำโดยส่งแบบสอบถามไปให้ผู้ตอบกรอกข้อมูลลงในแบบสอบถามเอง แล้วส่งกลับทางไปรษณีย์ วิธีนี้ผู้รวบรวมจะต้องแนบซองและแสตมป์ไปให้พร้อม ถ้าแหล่งข้อมูลอยู่ไม่ไกลนัก อาจจะส่งคนไปเก็บรวบรวมเองก็ได้ หรืออาจจะรวบรวมข้อมูลโดยออกไปสัมภาษณ์แล้วผู้สำรวจกรอกข้อมูลให้ ในกรณีที่ประชากรของแหล่งข้อมูลเขียนหนังสืออ่านหนังสือไม่ได้ หรือมีการศึกษาน้อย หรือแหล่งข้อมูลเป็นวัตถุสิ่งของ ผู้สำรวจจะต้องทำการบันทึกข้อมูลเอง

สิ่งที่สำคัญสำหรับแบบสอบถามก็คือ **การออกแบบ** แบบสอบถามที่ดีจะต้องมีหัวข้อเด่นชัด ภาษาที่ใช้อ่านเข้าใจง่าย ไม่กำกวม หัวข้อในแบบสอบถามไม่มากเกินไป และไม่ควรออกแบบที่จะต้องให้ผู้ตอบเขียนข้อมูลเชิงพรรณนายาว ๆ ถ้าหลีกเลี่ยงได้ควรจะหลีกเลี่ยงเสีย

แบบสอบถามแบบนี้ นอกจากจะทำให้ผู้ตอบเกิดความเบื่อหน่ายไม่อยากที่จะตอบแล้ว ยังยากต่อการสรุปด้วย สิ่งที่จะต้องคำนึงถึงอีกอย่างก็คือ ลักษณะของคำถาม จะต้องมีลักษณะที่เอื้ออำนวยในความสะดวก และรวดเร็วที่จะนำมาวิเคราะห์ด้วย

1.6 การจัดกลุ่มข้อมูลและการแจกแจงความถี่

(Grouping Data and Frequency Distribution)

ข้อมูลที่เก็บรวบรวมไม่ว่าจะได้จากการทำสำมะโน หรือจากการสำรวจตัวอย่าง ข้อมูลนี้เรียกว่าข้อมูลดิบ (Raw data) ถ้าได้มาโดยตรงจากการสัมภาษณ์หรือส่งแบบสอบถามเรียกว่าข้อมูลปฐมภูมิ (Primary data) และถ้าเป็นข้อมูลที่ได้มาจากหน่วยงานอื่นที่เก็บรวบรวมไว้แล้ว เช่น ข้อมูลที่ได้จากทะเบียนต่าง ๆ เรียกว่า ข้อมูลทุติยภูมิ (Secondary data) ข้อมูลดิบเหล่านี้อาจจะเรียกอีกอย่างหนึ่งว่า Ungrouped data คือข้อมูลที่ยังไม่ได้จัดหมวดหมู่ให้เป็นระเบียบ การจัดกลุ่มข้อมูลที่นิยมมี 2 วิธี คือ

1.6.1 การจัดลำดับข้อมูล

เป็นการจัดระเบียบข้อมูลให้เป็นลำดับจากน้อยไปมาก หรือจากมากมาน้อย ซึ่งจะทำให้ข้อมูลเป็นระเบียบ สะดวกในการตรวจสอบ และหาค่าต่าง ๆ ทางสถิติ เช่นรายได้ต่อวันของกรรมกร 10 คน มีดังนี้ 18 35 38 40 37 48 30 54 15 และ 52 บาท ข้อมูลดังกล่าวนี้ยังไม่ได้จัดลำดับให้เป็นระเบียบ นำมาจัดลำดับได้ดังนี้ จัดลำดับเรียงจากน้อยไปมากจะได้ดังนี้ 15 18 30 35 37 38 40 48 52 54 และนำไปจัดลำดับจากมากมาน้อยได้ดังนี้ 54 52 48 40 38 37 35 30 18 15 จะเห็นได้ว่าเมื่อจัดเรียงลำดับแล้วจะรู้ได้ทันทีว่า รายได้ต่ำสุดของกรรมกร 10 คนนี้ คือ 15 บาทต่อวัน และรายได้สูงสุดคือ 54 บาทต่อวัน.

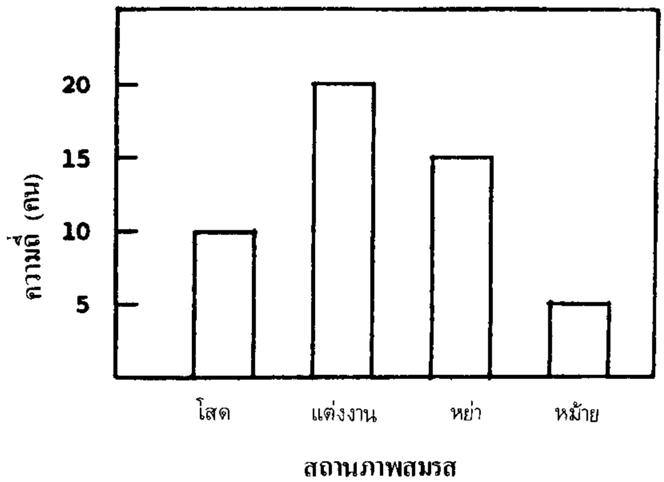
1.6.2 การสร้างตารางแจกแจงความถี่ (Frequency Table)

การแจกแจงความถี่ลงในตารางเป็นการจัดหมวดหมู่ของข้อมูล โดยแบ่งเป็นกลุ่มเป็นชั้น (class) ข้อมูลที่นำมาแจกแจงความถี่แล้วเรียกว่า Grouped data จะมีประโยชน์ทำให้ทราบลักษณะข้อมูลได้ละเอียด และสะดวกในการนำไปคำนวณหาค่าทางสถิติ เพื่อจะนำไปวิเคราะห์ต่อไป.

ตัวอย่างเช่น การสำรวจสถานภาพสมรสคนงานหญิงจำนวน 50 คนของโรงงานแห่งหนึ่งพบว่า มีสถานภาพเป็นโสด 10 คน แต่งงานแล้ว 20 คน หย่า 15 คน และเป็นหม้าย 5 คน นำมาเขียนเป็นตารางแจกแจงความถี่ได้ดังนี้.

ตาราง 1.1 แสดงการแจกแจงความถี่สถานภาพสมรสของคนงานหญิง

สถานภาพสมรส	ความถี่ (คน)
โสด	10
แต่งงาน	20
หย่า	15
หม้าย	5
รวม	50



รูป 1.1 แผนภาพแสดงการแจกแจงความถี่สถานภาพสมรสของคนงานหญิง

1.6.3 วิธีการแจกแจงความถี่

การแจกแจงข้อมูลซึ่งมีเป็นจำนวนมาก ๆ จัดให้เป็นหมวดหมู่โดยการสร้างตารางมีขั้นตอนดังนี้

1) หาพิสัย (Range)

   พิสัย = ค่าสูงสุด - ค่าต่ำสุด

2) หาจำนวนชั้นโดยประมาณ (number of class)

   การหาจำนวนชั้นอาจหาได้จากสูตรทั่วไปได้ดังนี้

   จำนวนชั้น = 1 + 3.3 log N

   โดยที่ N = จำนวนข้อมูลทั้งหมด

   จำนวนชั้นนั้นโดยทั่ว ๆ ไปนิยมใช้ตั้งแต่ 6 - 10 ชั้น

3) หาอันตรภาคชั้น (class interval) อันตรภาคชั้นนั้นโดยทั่วไปใช้สัญลักษณ์ i และหาได้จากสูตร

   $$\text{อันตรภาคชั้น} = \frac{\text{พิสัย}}{\text{จำนวนชั้น}}$$

4) เมื่อได้ค่าอัตรภาคชั้นแล้ว นำไปหาจำนวนชั้นที่แท้จริงอีกครั้งหนึ่ง ดังนี้

   $$\text{จำนวนชั้น} = \frac{\text{พิสัย}}{\text{อันตรภาคชั้น}} + 1$$

5) กำหนดค่ากึ่งกลาง (mid point) ค่ากึ่งกลางของชั้นที่หนึ่ง กำหนดให้ต่ำกว่าค่าต่ำสุดของข้อมูลเล็กน้อย

   สำหรับค่ากึ่งกลางชั้นต่อไปหาได้ดังนี้

   ค่ากึ่งกลางชั้นที่ 2 = ค่ากึ่งกลางชั้นที่ 1 + อันตรภาคชั้น

   ค่ากึ่งกลางชั้นที่ 3 = ค่ากึ่งกลางชั้นที่ 2 + อันตรภาคชั้น

   ค่ากึ่งกลางชั้นที่ 4 = ค่ากึ่งกลางชั้นที่ 3 + อันตรภาคชั้น

   ฯลฯ

   สำหรับการหาค่ากึ่งกลางเมื่อทราบขีดจำกัดล่าง (Lower Limit) และขีดจำกัดบน (Upper Limit) แล้ว หาได้ดังนี้

   $$\text{ค่ากึ่งกลาง} = \frac{\text{ขีดจำกัดล่าง} + \text{ขีดจำกัดบน}}{2}$$

6) หาขีดจำกัด เมื่อทราบอันตรภาคชั้นและค่ากึ่งกลางแล้ว สามารถที่จะหาขีดจำกัดล่าง (Lower Limit) และขีดจำกัดบน (Upper Limit) ได้ดังนี้

$$\text{ขีดจำกัดล่าง} = \text{ค่ากึ่งกลาง} - \frac{\text{อันตรภาคชั้น}}{2}$$

$$\text{ขีดจำกัดบน} = \text{ค่ากึ่งกลาง} + \frac{\text{อันตรภาคชั้น}}{2}$$

7) ตีตารางเพื่อแจกแจงความถี่ ตารางพื้นฐานที่จะนำไปใช้ในการแจกแจงความถี่ มีช่องต่าง ๆ ดังนี้

ก. ช่องขีดจำกัดของชั้น (Class Limit or Class Boundaries)

ข. ค่ากึ่งกลาง (mid point)

ค. ช่องรอยขีดความถี่ (Tallied frequency)

ง. ช่องความถี่ (Frequency)

จ. ช่องความถี่สะสม (Cumulative Frequency)

ตาราง 1.2 แสดงแบบของตารางแจกแจงความถี่

ขีดจำกัดของชั้น (Class Limit)	ค่ากึ่งกลาง (mid point)	รอยขีด (Tally)	ความถี่ (Frequency)	ความถี่สะสม (C.F.)
ก.	ข.	ค.	ง.	จ.

ตัวอย่าง 1.1 คะแนนสอบวิชาคณิตศาสตร์ของนักศึกษาที่สุ่มตัวอย่างขึ้นมาจำนวน 40 คน สอบได้คะแนนดังนี้

68	84	75	82	68	90	62	88	76	93	73	79	88	73	60
93	71	59	85	75	61	65	75	74	62	95	78	63	72	66
78	82	75	94	79	69	74	68	60	86					

ให้สร้างตารางและแจกแจงความถี่ ตามวิธีที่เรียนมา

วิธีทำ

1) หาพิสัย

$$\text{พิสัย} = \text{ค่าสูงสุด} - \text{ค่าต่ำสุด} = 95 - 59 = 36$$

2) หาจำนวนชั้นโดยประมาณ

$$\text{จำนวนชั้น} = 1 + 3.3 \log N = 1 + 3.3 \log 40 = 1 + 3.3\,(1.6201) = 6.4 \approx 6$$

3) หาอันตรภาคชั้น

$$\text{อันตรภาคชั้น} = \frac{\text{พิสัย}}{\text{จำนวนชั้น}} = \frac{36}{6} = 6$$

4) หาจำนวนชั้นที่แท้จริง

$$\text{จำนวนชั้น} = \frac{\text{พิสัย}}{\text{อันตรภาคชั้น}} + 1 = \frac{36}{6} + 1 = 7$$

5) กำหนดค่ากึ่งกลางของชั้นที่หนึ่งให้ต่ำกว่าค่าต่ำสุดเล็กน้อย ในที่นี้ค่าต่ำสุดคือ 59 เพราะฉะนั้นกำหนดให้ค่ากึ่งกลางเท่ากับ 58.5 ค่ากึ่งกลางของชั้นต่อไปจะเพิ่มขึ้นเท่าอันตราภาคชั้น

ค่ากึ่งกลางชั้นที่ 1 = 58.5

ค่ากึ่งกลางชั้นที่ 2 = ค่ากึ่งกลางชั้นที่ 1 + อันตรภาคชั้น = 58.5 + 6 = 64.5

ค่ากึ่งกลางชั้นที่ 3 = ค่ากึ่งกลางชั้นที่ 2 + อันตรภาคชั้น = 64.5 = 6 = 70.5

ค่ากึ่งกลางชั้นที่ 4 = ค่ากึ่งกลางชั้นที่ 3 + อันตรภาคชั้น = 70.5 + 6 = 76.5

ค่ากึ่งกลางชั้นที่ 5 = ค่ากึ่งกลางชั้นที่ 4 + อันตรภาคชั้น = 76.5 + 6 = 82.5

ค่ากึ่งกลางชั้นที่ 6 = ค่ากึ่งกลางชั้นที่ 5 + อันตรภาคชั้น = 82.5 + 6 = 88.5

ค่ากึ่งกลางชั้นที่ 7 = ค่ากึ่งกลางชั้นที่ 6 + อันตรภาคชั้น = 88.5 + 6 = 94.5

6) หาขีดจำกัดล่างและขีดจำกัดบน

ชั้นที่ 1 ขีดจำกัดล่าง = ค่ากึ่งกลาง - $\frac{\text{อัตรภาคชั้น}}{2}$

$= 58.5 - \frac{6}{2}$

$= 58.5 - 3 = 55.5$

ขีดจำกัดบน = ค่ากึ่งกลาง + $\frac{\text{อันตราภาคชั้น}}{2}$

$= 58.5 + \frac{6}{2}$

$= 58.5 + 3 = 61.5$

ชั้นอื่น ๆ ก็หาได้ในทำนองเดียวกัน

7) สร้างตารางแจกแจงความถี่

ตาราง 1.3 แสดงการแจกแจงความถี่คะแนนสอบวิชาคณิตศาสตร์ของนักศึกษาที่สุ่มตัวอย่าง ทั้งหมดจำนวน 40 คน

ขีดจำกัด	ค่ากึ่งกลาง	รอยขีด	ความถี่	ความถี่สะสม
๕๕.๕ - ๖๑.๕	๕๘.๕	////	๔	๔
๖๑.๕ - ๖๗.๕	๖๔.๕	~~////~~	๕	๙
๖๗.๕ - ๗๓.๕	๗๐.๕	~~////~~ ///	๘	๑๗
๗๓.๕ - ๗๙.๕	๗๖.๕	~~////~~ ~~////~~ /	๑๑	๒๘
๗๙.๕ - ๘๕.๕	๘๒.๕	////	๔	๓๒
๘๕.๕ - ๙๑.๕	๘๘.๕	////	๔	๓๖
๙๑.๕ - ๙๗.๕	๙๔.๕	////	๔	๔๐
			๔๐	

การสร้างตารางแจกแจงความถี่อาจสร้างอย่างหยาบ ๆ ดังนี้

1) หาพิสัย โดยที่ พิสัย = ค่าสูงสุด - ค่าต่ำสุด

2) หาจำนวนชั้น จำนวนชั้น = $1 + 3.3 \log N$

3) หาอันตรภาคชั้น อันตรภาคชั้น = $\dfrac{\text{พิสัย}}{\text{จำนวนชั้น}}$

4) กำหนดให้ค่าต่ำสุดของข้อมูลเป็นขีดจำกัดล่างของชั้นที่ 1 ชั้นถัดไปจะแตกต่างกันเท่ากับอันตรภาคชั้น

ตัวอย่าง 1.2 จากคะแนนสอบของนักศึกษาจำนวน 40 คนในตัวอย่างที่ 1 นำมาสร้างตารางแจกแจงความถี่อีกวิธีหนึ่งได้ดังนี้

1) หาพิสัย พิสัย = 95 - 59 = 36

2) หาจำนวนชั้น จำนวนชั้น = $1 + 3.3 \log 40 = 1 + 3.3\ (1.6201) = 6.4 \approx 6$

3) หาอันตรภาคชั้น อันตรภาคชั้น = $\dfrac{\text{พิสัย}}{\text{จำนวนชั้น}} = \dfrac{36}{6} = 6$

4) หาขีดจำกัดล่าง
   ขีดจำกัดล่างของชั้นที่ 1 = 59

ขีดจำกัดล่างของชั้นถัดไปจะเพิ่มขึ้นเท่ากับอันตรภาคชั้น

เมื่อหาค่าต่าง ๆ ได้แล้ว นำมาสร้างตารางแจกแจงความถี่ได้ดังนี้.

ตาราง 1.4 ตารางแจกแจงความถี่แบบง่าย ๆ

ขีดจำกัด	**ค่ากึ่งกลาง**	**รอยขีด**	**ความถี่**	**ความถี่สะสม**
๕๙ - ๖๔	๖๑.๕	~~////~~ //	๗	๗
๖๕ - ๗๐	๖๗.๕	~~////~~ /	๖	๑๓
๗๑ - ๗๖	๗๓.๕	~~////~~ ~~////~~ /	๑๑	๒๔
๗๗ - ๘๒	๗๙.๕	~~////~~ /	๖	๓๐
๘๓ - ๘๘	๘๕.๕	~~////~~	๕	๓๕
๘๙ - ๙๔	๙๑.๕	////	๔	๓๙
			๓๙	

จากตารางที่ 1.4 ขีดจำกัดของข้อมูลในชั้นสุดท้ายคือ (88 - 94) คะแนน ไม่สามารถที่จะรวมคะแนนของนักศึกษาอีกหนึ่งคน คือคนที่ได้ 95 คะแนน เพื่อให้ตารางแจกแจงความถี่รวมเอาข้อมูลทั้งหมดไว้ อาจจะกำหนดขีดจำกัดของชั้นที่ 1 และชั้นสุดท้ายไว้เป็นแบบ Opened limit interval ก็ได้ การสร้างตารางแบบนี้นอกจากเพื่อเหตุผลดังกล่าวแล้วยังช่วยหลีกเลี่ยงไม่ให้ชั้นใดชั้นหนึ่งเป็นศูนย์ ซึ่งไม่เป็นที่นิยมด้วย

จากคะแนนสอบของนักเรียนจำนวน 40 คน นำมาสร้างตารางแจกแจงความถี่ แบบ Opened limit ได้ดังนี้.

ตาราง 1.5 แสดงตารางแจกแจงชนิดขีดจำกัดแบบ Opened limit

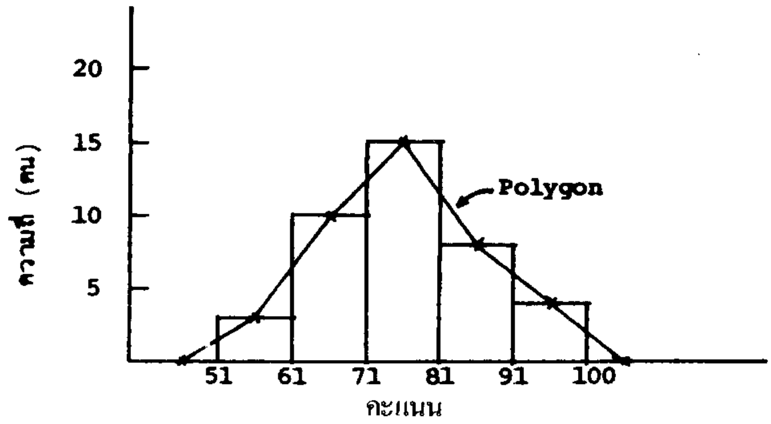
ขีดจำกัด	รอยขีด	ความถี่	ความถี่สะสม
น้อยกว่า ๖๔	𝍸 //	๗	๗
๖๔ - ๗๐	𝍸 /	๖	๑๓
๗๑ - ๗๖	𝍸 𝍸 /	๑๑	๒๔
๗๗ - ๘๒	𝍸 /	๖	๓๐
๘๓ - ๘๘	𝍸	๕	๓๕
มากกว่า ๘๘	𝍸	๕	๔๐
		๔๐	

ตารางแจกแจงความถี่นอกจากที่กล่าวมาแล้วอาจจะสร้างตามวัตถุประสงค์ โดยกำหนดจำนวนชั้นและอันตรภาคชั้นตามต้องการ เช่นจากคะแนนสอบของนักศึกษาในตัวอย่างที่ 1.1 ต้องการสร้างตารางแจกแจงความถี่ให้มี 6 ชั้น และมีอันตรภาคชั้นเท่ากับ 10 จะได้ดังนี้.

ตาราง 1.6 ตารางแจกแจงความถี่กำหนดชั้นและอันตรภาคชั้นตามต้องการ

ขีดจำกัด	ค่ากึ่งกลาง	รอยขีด	ความถี่
๕๑ - ๖๐	๕๕.๕	///	๓
๖๑ - ๗๐	๖๕.๕	𝍸 𝍸	๑๐
๗๑ - ๘๐	๗๕.๕	𝍸 𝍸 𝍸	๑๕
๘๑ - ๙๐	๘๕.๕	𝍸 ///	๘
๙๑ - ๑๐๐	๙๕.๕	////	๔
			๔๐

นำเอาข้อมูลจากตารางแจกแจงความถี่ ตารางที่ 1.6 ไปเขียนกราฟ โดยกำหนดให้แกนตั้งแทน ความถี่ (frequency) ของข้อมูล แกนนอนแทนคะแนนของนักศึกษา



รูป 1.2 กราฟแสดงการแจกแจงความถี่ของคะแนนจากตาราง 1.6

เส้นโยงระหว่างจุดกึ่งกลางของยอด Bar เรียกว่า เส้น Frequency Polygon เพื่อให้กราฟปิดหัวและท้ายดูสวยงาม จึงต่อออกไปยังจุดกึ่งกลางของชั้นที่มีความถี่เป็นศูนย์ซึ่งจะปิดหัวท้ายพอดี

7 การนำเสนอข้อมูล

ข้อมูลที่เก็บรวบรวมได้ เมื่อนำไปจัดหมวดหมู่เป็นยอดตัวเลขแล้ว ข้อมูลบางอย่างอาจจะไม่จำเป็นต้องนำไปวิเคราะห์ทางสถิติ อาจจะนำมาเสนอให้ดูอย่างง่าย ๆ ซึ่งจะจูงใจให้น่าดูมากขึ้น ในการเสนอข้อมูลนี้ ที่นิยมทั่วไปมีอยู่ 6 แบบด้วยกัน แต่ละแบบมีวิธีการเสนอดังนี้

1.7.1 การเสนอโดยบทความ

การเสนอแบบนี้จะเสนอเป็นบทความสั้น ๆ และมียอดตัวเลขผสมอยู่ด้วย ส่วนมากจะปรากฏในรายการวิทยุ หรือในหนังสือรายงานต่าง ๆ เช่น :-

"ปีพุทธศักราช 2522 คาดว่าจะมีนักศึกษาจบ ม.ศ.5 ในปีนี้และที่ตกค้างอยู่ประสงค์จะสมัครสอบคัดเลือกเข้าศึกษาในระดับอุดมศึกษาประมาณ 120,000 คน"

"ปีการศึกษา 2520 นักเรียนในโรงเรียนรัฐบาลและเอกชนของประเทศไทย มีจำนวน

8,358,700 คน เป็นนักเรียนในโรงเรียนรัฐบาลเสีย 7,145,400 คน

ส่วนที่เหลืออีก 1,213,300 คน เป็นนักเรียนในโรงเรียนเอกชน"

1.7.2 การเสนอโดยตาราง

การเสนอแบบตารางนี้ เป็นการเสนอข้อมูลทางสถิติที่นิยมแบบหนึ่ง อ่านง่าย เข้าใจง่าย ตัวอย่างเช่น การเสนอจำนวนประชากรของประเทศไทย ปี พ.ศ. 2520.

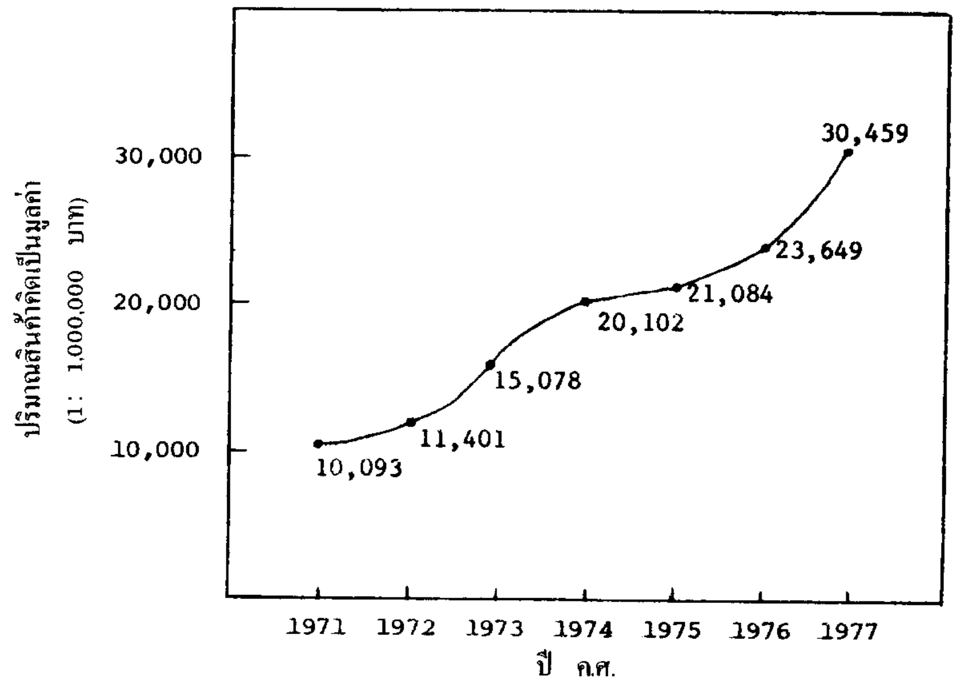
ตาราง 1.7 แสดงจำนวนประชากรประเทศไทยแยกตามอายุ

อายุ (ปี)	ปี พ.ศ. ๒๕๒๐	
	จำนวน (พัน)	เปอร์เซ็น
๐ - ๔	๖,๗๓๗	๑๕.๓
๕ - ๙	๖,๔๐๒	๑๔.๕
๑๐ - ๑๔	๕,๗๐๙	๑๓.๐
๑๕ - ๑๙	๔,๘๒๗	๑๑.๐
๒๐ - ๒๔	๔,๐๗๓	๙.๓
๒๕ - ๒๙	๓,๔๘๔	๗.๙
๓๐ - ๓๔	๒,๗๘๙	๖.๓
๓๕ - ๓๙	๒,๑๐๔	๔.๘
๔๐ - ๔๔	๑,๗๘๐	๔.๐
๔๕ - ๔๙	๑,๕๙๗	๓.๖
๕๐ - ๕๔	๑,๓๒๕	๓.๐
๕๕ - ๕๙	๑,๐๓๖	๒.๔
๖๐ ขึ้นไป	๒,๑๗๗	๔.๙
รวม	๔๔,๐๓๙	๑๐๐.๐

ที่มา : สถาบันประชากรศาสตร์ จุฬาฯ สสช.

1.7.3 การเสนอโดยกราฟเส้น

การเสนอแบบกราฟเส้นนี้ เหมาะสำหรับข้อมูลที่มีค่าต่อเนื่อง (continuous) โดยทั่วไปมักจะเปรียบเทียบกับเวลา ตัวอย่างเช่น ปริมาณสินค้าที่ประเทศไทยสั่งจากประเทศญี่ปุ่น ตั้งแต่ปีคริสตศักราช 1971 - 1977 มีมูลค่าต่าง ๆ ดังแสดงในกราฟนี้.



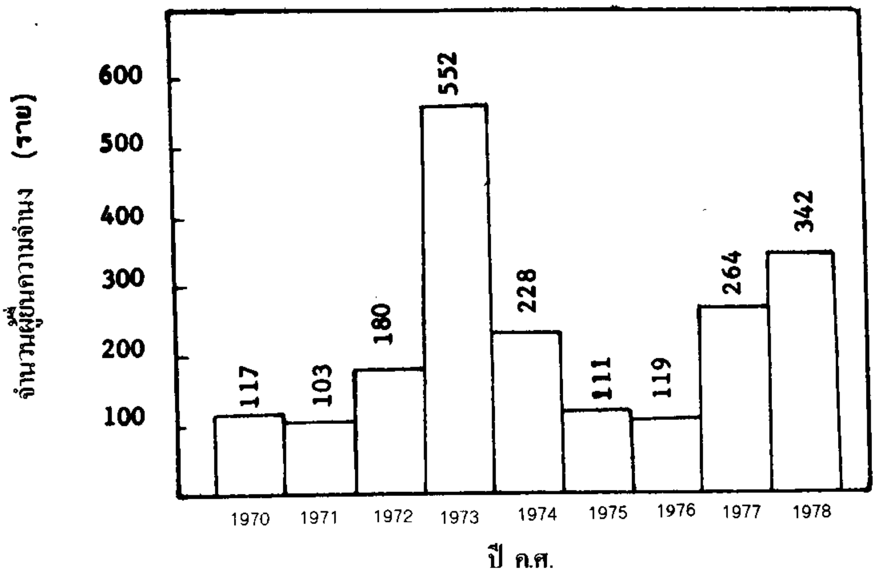
รูป 1.3 กราฟแสดงจำนวนสินค้าที่ประเทศไทยสั่งจากญี่ปุ่น

ที่มา : กรมศุลกากร กระทรวงการคลัง

1.7.4 การเสนอโดยกราฟแท่ง (Bar Chart)

การเสนอแบบกราฟแท่งหรือแผนภูมิแท่ง จะเสนอตามแกนนอนหรือแกนตั้งก็ได้ ปริมาณข้อมูลแสดงโดยขนาดความสูงของแท่งสี่เหลี่ยมผืนผ้า แต่ละแท่งมีขนาดความกว้างเท่ากัน และเขียนเรียงติดต่อกัน หรือเว้นช่วงเล็กน้อยก็ได้ แล้วแต่ความต้องการ ถ้าต้องการเปรียบเทียบข้อมูลแต่ละประเภท ให้เขียนติดต่อกันแล้วระบายสีให้แตกต่างกัน จะช่วยจูงใจให้น่าดูและเปรียบเทียบเห็นได้ชัดขึ้น.

ตัวอย่างข้อมูลเกี่ยวกับจำนวนผู้ยื่นความจำนงขอลงทุนในประเทศไทย ตั้งแต่ปีคริสต-ศักราช 1970 - 1978 เสนอโดยกราฟแท่งดังนี้



รูป 1.4 เปรียบเทียบจำนวนผู้ยื่นความจำนงขอลงทุนในประเทศไทย ระหว่างปี ค.ศ. 1970-1978
ที่มา : หนังสือ Business Time Supplement, Feb. 26, 1979.

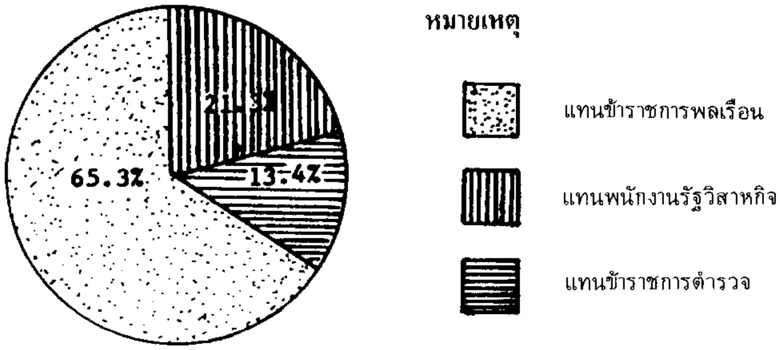
1.7.5 การเสนอโดยกราฟวงกลม

การเสนอโดยกราฟวงกลมหรือเรียกอีกอย่างว่า แผนภาพกง ทำได้โดยแบ่งวงกลม ซึ่งมี 360° เทียบให้เท่ากับ 100% นำเอาข้อมูลแต่ละประเภทไปหาอัตราส่วนของเปอร์เซนต์ แล้วจึงนำไปเทียบเป็นองศา เพื่อจะนำไปเขียนลงในกราฟวงกลม ตัวอย่างเช่น จำนวนข้าราชการพลเรือน ตำรวจ และพนักงานรัฐวิสาหกิจ ในปี พ.ศ. 2520 มีดังนี้.

ตาราง 1.8 เปรียบเทียบจำนวนข้าราชการพลเรือน ตำรวจ และพนักงานรัฐวิสาหกิจ ปี พ.ศ. 2520

	ปีงบประมาณ 2520		
	จำนวน	เปอร์เซนต์	องศา
ข้าราชการพลเรือน	550,000	65.32	235.15
ข้าราชการตำรวจ	113,000	13.42	48.32
พนักงานรัฐวิสาหกิจ	179,000	21.26	76.53
รวม	842,000	100	360

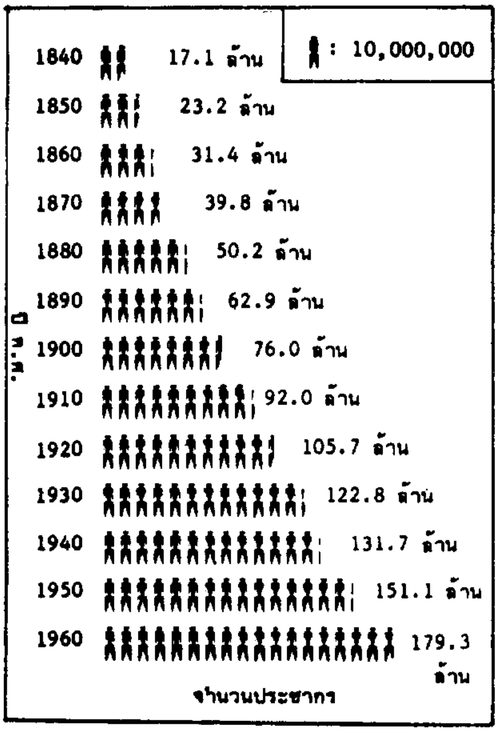
ข้อมูลจากตาราง 1.8 นำมาเขียนกราฟวงกลมได้ดังนี้



รูป 1.5 เปรียบเทียบข้าราชการพลเรือน ตำรวจ และพนักงานรัฐวิสาหกิจ ปี พ.ศ. 2520

1.7.6 การเสนอโดยรูปภาพ (Pictograph)

การเสนอแบบกราฟรูปภาพนี้ ใช้รูปภาพจริง เสนอตามประเภทของข้อมูล ถ้าข้อมูลเป็นคนก็ใช้ภาพคน ข้อมูลเป็นรถยนต์ก็ใช้รถยนต์ ข้อมูลเป็นสัตว์เลี้ยงเช่นโค กระบือ ก็แสดงภาพโค กระบือ เป็นต้น การเสนอจะเสนอในลักษณะของแผนที่หรือตารางก็แล้วแต่ความเหมาะสม ตัวอย่างเช่นการเสนอจำนวนประชากรของประเทศสหรัฐอเมริกา เสนอโดยกราฟรูปภาพดังนี้.



รูป 1.6 แสดงจำนวนประชากรของประเทศสหรัฐอเมริกา ตั้งแต่ปี ค.ศ. 1840-1960.
ที่มา : Bureau of the Cencus.

1.8 การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measure of Central Tendency)

การวัดแนวโน้มเข้าสู่ส่วนกลางคือ การหาค่าตัวกลางซึ่งเป็นเครื่องมือทางสถิติเบื้องต้นที่มีประโยชน์มาก ใช้ในการวัดหาค่าเฉลี่ยของข้อมูลต่าง ๆ เครื่องมือในการวัดค่าตัวกลางที่นิยมแพร่หลายมีอยู่ 3 อย่างคือ

1.8.1 มัชฌิมเลขคณิต (Arthmatic Mean)

เป็นตัวกลางที่ใช้หาค่าเฉลี่ยของข้อมูลที่สำคัญที่สุด นิยมมากที่สุดและสามารถนำไปใช้ประยุกต์ในการคำนวณค่าต่าง ๆ ทางสถิติได้อย่างกว้างขวาง มัชฌิมเลขคณิตเหมาะที่จะใช้ข้อมูลที่มีการกระจายอย่างสม่ำเสมอ ค่าสูงสุดและต่ำสุดไม่แตกต่างกันมาก ข้อมูลที่มีบางตัวผิดปกติ ไม่ควรวัดค่าตัวกลางด้วยมัชฌิมเลขคณิต เพราะการวัดด้วยวิธีนี้จะวัดทุกตัวของข้อมูลซึ่งจะรวมเอาข้อมูลที่ผิดปกติเข้าไปด้วย.

1.8.2 มัธยฐาน (Median)

มัธยฐานเป็นค่าตัวกลางของข้อมูล ซึ่งครึ่งหนึ่งของข้อมูลมีค่าต่ำกว่าหรือเท่ากับ และมากกว่าหรือเท่ากับค่านี้ มัธยฐานเป็นการวัดค่าตัวกลางโดยเรียงข้อมูลให้เป็นลำดับค่าที่อยู่ตรงกลางคือมัธยฐาน จะเห็นได้ว่ามัธยฐานใช้หาค่าตัวกลางโดยพิจารณาตำแหน่งที่ไม่ได้พิจารณาค่าของข้อมูล ดังนั้นข้อมูลที่มีค่าผิดปกติจะไม่มีผลกระทบกระเทือนมัธยฐาน ตัวอย่าง เช่น ในชุมชนแห่งหนึ่งมีนายทุนเป็นเจ้าของโรงงานอยู่หนึ่งครอบครัว มีรายได้สูงมาก นอกนั้นเป็นครอบครัวคนทำงานในโรงงานแห่งนี้มีรายได้ต่ำ การหารายได้เฉลี่ยต่อครอบครัวของชุมชนนี้ โดยหาแบบมัชฌิมเลขคณิต จะได้รายได้เฉลี่ยมีค่าสูง สูงกว่ารายได้จริงของครอบครัวส่วนมาก ที่เป็นเช่นนี้เพราะรายได้ของครอบครัวเจ้าของโรงงานสูง เมื่อนำมาเฉลี่ยแล้วจึงทำให้ครอบครัวอื่น ๆ มีรายได้เฉลี่ยสูงไปด้วย ถ้าใช้มัธยฐานหารายได้เฉลี่ยจะได้ค่าใกล้เคียงกับรายได้จริงของครอบครัวส่วนใหญ่.

1.8.3 ฐานนิยม (Mode)

การหาค่าตัวกลางโดยวิธีนี้ จะพิจารณาค่าของข้อมูลที่มีความถี่มากที่สุด พิจารณาเฉพาะกลุ่มของข้อมูลที่เป็นตัวแทนมากที่สุด ไม่ได้พิจารณาทุกตัว เพราะฉะนั้นข้อมูลที่มีค่าผิดปกติก็จะไม่มีผลต่อการหาฐานนิยม เช่นเดียวกับการหามัธยฐาน ข้อมูลยิ่งมีจำนวนมาก การหาค่าตัวกลางโดยฐานนิยมยิ่งดี นักออกแบบผู้ผลิต นักธุรกิจ ได้อาศัยฐานนิยมเป็นตัวชี้บอกว่าควรจะออกแบบอย่างไร ผลิตสินค้าประเภทไหน จึงจะเป็นที่ถูกใจลูกค้าและพอเพียงกับความต้องการ ร้านค้าก็

อาศัยฐานนิยมช่วยชี้บอกถึงความนิยมของประชาชนในสินค้าประเภทต่าง ๆ ชนิดไหน ขนาดไหน สีอะไร ขายดีที่สุด ซึ่งจะได้เตรียมสั่งซื้อมาขายได้พอเพียงกับความต้องการ เช่น ร้านขายนาฬิกาก็อยากรู้ว่านาฬิกาชนิดไหนราคาเท่าไรที่มีคนซื้อมากที่สุด ร้านขายเสื้อผ้าก็ต้องการทราบ ขนาด สี ของเสื้อผ้าที่มีคนใช้และนิยมมากที่สุด ขายดีที่สุด ซึ่งสิ่งเหล่านี้ฐานนิยมจะช่วยได้เป็นอย่างดี.

การคำนวณหาค่าตัวกลางต่าง ๆ สำหรับข้อมูลที่จัดเป็นหมวดหมู่และไม่ได้จัดเป็นหมวดหมู่ มีวิธีการคำนวณดังต่อไปนี้

1) สำหรับข้อมูลที่ไม่ได้จัดเป็นหมวดหมู่ (Ungrouped Data)

ก. การหาค่ามัชฌิมเลขคณิต

ในการคำนวณหามัชฌิมเลขคณิต คำนวณหาได้โดยใช้สูตรดังนี้

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n}$$

$$= \sum_{i=1}^{n} X_i / n$$

โดยที่ X คือมัชฌิมเลขคณิต

$X_1, X_2, X_3, \ldots. X_n$ เป็นข้อมูล n ข้อมูล

X_i เป็นข้อมูลตัวที่ i

n คือจำนวนข้อมูลทั้งหมด

Σ คือผลรวม

พิจารณาชุดของตัวเลขต่อไปนี้ ซึ่งประกอบด้วยเลข 4 4 5 7 10 มีอยู่ 5 จำนวนด้วยกัน นำไปหามัชฌิมเลขคณิตได้

จากสูตร $$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$= \frac{4 + 4 + 5 + 7 + 10}{5}$$

$$= \frac{30}{5} = 6$$

นั่นคือ มัชฌิมเลขคณิตของเลขชุดนี้ เท่ากับ 6

ข. การหาค่ามัธยฐาน

การหาค่ามัธยฐาน หาได้โดยการเรียงข้อมูลให้เป็นลำดับ จะเรียงจากน้อยไปมาก หรือจากมากไปน้อยก็ได้ ค่าของข้อมูลที่มีตำแหน่งอยู่ตรงกลางคือ มัธยฐาน เช่นต้องการหามัธยฐานของข้อมูล n ข้อมูล มัธยฐานจะอยู่ตำแหน่งที่ $\frac{n+1}{2}$

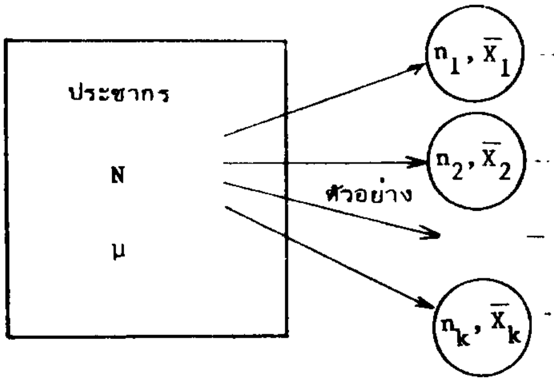
ตัวอย่างเช่น ตัวเลขชุดหนึ่งประกอบด้วยเลข 4 4 5 7 10 เลขชุดนี้มี 5 ตัว เพราะฉะนั้น n = 5 หาตำแหน่งของมัธยฐานจะอยู่ตำแหน่งที่ $\frac{n+1}{2} = \frac{5+1}{2} = 3$ มัธยฐานจะอยู่ตำแหน่งที่ 3 ของข้อมูลที่เรียงลำดับ ในที่นี้คือ 5 นั่นคือ มัธยฐานของเลขชุดนี้เท่ากับ 5

ในกรณีที่ข้อมูลเป็นจำนวนคู่เช่น 10 4 5 4 7 12 นำมาเรียงลำดับใหม่จะได้ 4 4 5 7 10 12 ตำแหน่งของมัธยฐานจะอยู่ที่ $\frac{n+1}{2} = \frac{6+1}{2} = 3.5$ ซึ่งเป็นค่าระหว่าง 5 และ 7 กรณีเช่นนี้จะต้องเอาค่าทั้งสองมาหาค่าเฉลี่ย เพราะฉะนั้นค่าของมัธยฐานของเลขชุดนี้จะเท่ากับ $\frac{5+7}{2} = 6$

ค. การหาค่าฐานนิยม

การหาค่าฐานนิยม ให้พิจารณาโดยตรวจดูค่าของข้อมูลที่ซ้ำกันมากที่สุด ค่านั้นคือ ฐานนิยม เช่น เลขชุดหนึ่งประกอบด้วยตัวเลข 4 4 5 7 10 ค่าที่ซ้ำกันมากที่สุดของเลขชุดนี้คือ 4 นั่นคือ ฐานนิยมของเลขชุดนี้ = 4

ในการศึกษาเรื่องค่าตัวกลางมีสัญลักษณ์อยู่ตัวหนึ่ง ที่นักศึกษาควรรู้จักคือ μ ซึ่งใช้แทนค่าเฉลี่ยของประชากร ในทางปฏิบัติการหาค่า μ หาได้ยาก จึงใช้หาค่าเฉลี่ยจากตัวอย่าง คือ $\overline{X}$ แทน ถ้าต้องการทราบค่า μ ก็สุ่มตัวอย่างมาหลาย ๆ ชุด แล้วนำเอาค่าเฉลี่ยตัวอย่างแต่ละชุดไปหาค่าเฉลี่ยอีกครั้งหนึ่ง จะได้ค่าใกล้เคียง μ เช่นประชากรมีจำนวน N ข้อมูล สุ่มตัวอย่างขึ้นมา k ชุด



$$\bar{X}_1 = \frac{X_1 + X_2 + X_3 + \ldots + X_{n_1}}{n_1}$$

$$\bar{X}_2 = \frac{X_1 + X_2 + \ldots + X_{n_2}}{n_2}$$

$$\bar{X}_k = \frac{X_1 + X_2 + \ldots + X_{n_k}}{n_k}$$

จะหาค่าเฉลี่ยของประชากรได้ดังนี้ $\mu = \dfrac{\bar{X}_1 + \bar{X}_2 + \ldots\ldots\ \bar{X}_k}{k}$

ในกรณีที่จำนวนประชากรมีไม่มากนำมาหาค่าเฉลี่ยทุกประชากร μ จะเท่ากับ $\bar{X}$

$$\mu = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

เปรียบเทียบผลต่างของข้อมูลกับตัวกลาง

ตาราง 1.9 แสดงผลต่างระหว่างข้อมูลกับมัชฌิมเลขคณิต

ข้อมูล	มัชฌิม	ผลต่าง	ผลต่างยกกำลังสอง
4	6	-2	4
4	6	-2	4
5	6	-1	1
7	6	+1	1
10	6	+4	16
ผลรวม		0	26

ตาราง 1.10 แสดงผลต่างระหว่างข้อมูลกับมัธยฐาน

ข้อมูล	มัธยฐาน	ผลต่าง	ผลต่างยกกำลังสอง
4	5	-1	1
4	5	-1	1
5	5	0	0
7	5	+2	4
10	5	+5	25
ผลรวม		5	31

ตาราง 1.11 แสดงผลต่างระหว่างข้อมูลกับฐานนิยม

ข้อมูล	ฐานนิยม	ผลต่าง	ผลต่างยกกำลังสอง
4	4	0	0
4	4	0	0
5	4	+1	1
7	4	+3	9
10	4	+6	36
ผลรวม		10	46

จากตารางที่ 1.9, 1.10 และ 1.11 ที่แสดงผลต่างระหว่างข้อมูลกับ มัชฌิมเลขคณิต มัธยฐานและฐานนิยม สรุปได้ดังนี้

(1) ในเซทของข้อมูลใด ๆ ผลรวมของผลต่างระหว่างข้อมูลกับมัชฌิมจะเท่ากับศูนย์ หรือในทางตรงกันข้าม ถ้าผลรวมของผลต่างระหว่างข้อมูลกับค่าใดค่าหนึ่งเท่ากับศูนย์ ค่านั้นคือ มัชฌิมเลขคณิต.

(2) ในเซทของข้อมูลใด ๆ ผลรวมของผลต่างกำลังของข้อมูลกับมัชฌิม จะน้อยกว่า ผลรวมของผลต่างกำลังสองของข้อมูลกับค่าใด ๆ หรือในทางตรงกันข้าม ค่าใด ๆ ในเซทของ ข้อมูลที่มีผลรวมของผลต่างยกกำลังสองมีค่าน้อยที่สุด ค่านั้นคือ มัชฌิมเลขคณิต.

2) สำหรับข้อมูลที่จัดหมวดหมู่แล้ว (Grouped data)

ข้อมูลที่มีเป็นจำนวนมาก ๆ จะหาค่าตัวกลางตามวิธีข้างต้นนี้ยุ่งยาก จะต้องนำข้อมูล มาจัดหมวดหมู่ แจกแจงเป็นตาราง แล้วจึงหาค่าต่าง ๆ ที่ต้องการ ซึ่งจะทำให้สะดวกและเร็วขึ้น โดยมีวิธีคำนวณดังนี้.

ก. การหามัชฌิมเลขคณิต

ในการหาค่ามัชฌิมเลขคณิตของข้อมูลที่จัดเป็นตารางแจกแจงความถี่ แล้ว หาได้ จากสูตรดังนี้

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + \cdots\cdots + f_kX_k}{n}$$

$$= \frac{\sum_{i=1}^{k} f_iX_i}{n}$$

โดยที่ f_i คือ ความถี่ของข้อมูลชั้นที่ i

X_i คือ ค่ากึ่งกลางของชั้นที่ i

k คือ จำนวนชั้นของตารางแจกแจงความถี่

n คือ จำนวนข้อมูลทั้งหมด

ตัวอย่าง 1.3 การหาค่ามัชฌิมเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษา จำนวน 40 คน จากข้อมูลที่ได้แจกแจงในตาราง 1.3

ขีดจำกัด	ค่ากึ่งกลาง (X_i)	ความถี่(f_i)	f_iX_i
55.5 - 61.5	58.5	4	234.0
61.5 - 67.5	64.5	5	322.5
67.5 - 73.5	70.5	8	564.0
73.5 - 79.5	76.5	11	841.5
79.5 - 85.5	82.5	4	330.0
85.5 - 91.5	88.5	4	354.0
91.5 - 97.5	94.5	4	378.0
รวม		40	3,024

จากสูตร

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

แทนค่า

$$= \frac{3,024}{40} = 75.6$$

นั่นคือ มัชฌิมเลขคณิตของนักศึกษาเท่ากับ 75.6 **ตอบ**

การหาค่ามัชฌิมเลขคณิตอาจหาได้อีกวิธีหนึ่ง ซึ่งเป็นวิธีลัด หาได้โดยกำหนดให้ค่ากึ่งกลางของชั้นที่มีความถี่มากที่สุด ซึ่งเป็นชั้นที่มีตัวแทนของข้อมูลมากที่สุด **เป็นมัชฌิมสมมติ** (Guess Mean) ข้อมูลที่ชั้นนี้จะมีความเบี่ยงเบนเท่ากับศูนย์ และในชั้นถัดไปจะมีความเบี่ยงเบนเป็น 1, 2, 3, ... และ – 1, – 2, – 3, ...แล้วแต่การจัดชั้น วิธีการคำนวณหาได้ดังนี้.

$$\bar{X} = a + i\bar{d}$$

โดยที่

$$\bar{d} = \frac{\sum_{i=1}^{k} f_i d_i}{n}$$

และ a คือ มัชฌิมสมมติ

i คือ อันตรภาคชั้น

f_i คือ ความถี่ของชั้นที่ i

d_i คือ ค่าเบี่ยงเบนจากมัชฌิมสมมติ $= \frac{X_i - a}{i}$

X_i คือ ค่ากึ่งกลางชั้นที่ i

k คือ จำนวนชั้นของตารางแจกแจงความถี่

n คือ จำนวนข้อมูลทั้งหมด

ตัวอย่าง 1.4 การหามัชฌิมเลขคณิตโดยวิธีลัด จากข้อมูลตาราง 1.3

ขีดจำกัด	ค่ากึ่งกลาง (X_i)	ความถี่ (f_i)	d_i	$f_i d_i$
55.5 - 61.5	58.5	4	-3	-12
61.5 - 67.5	64.5	5	-2	-10
67.5 - 73.5	70.5	8	-1	- 8
73.5 - 79.5	76.5	11	0	0
79.5 - 85.5	82.5	4	1	4
85.5 - 91.5	88.5	4	2	8
91.5 - 97.5	94.5	4	3	12
รวม		40		-6

จากสูตร $\bar{X} = a + i\bar{d}$

ชั้นที่มีค่ากึ่งกลางเท่ากับ 76.5 เป็นชั้นที่มีความถี่มากที่สุด กำหนดให้ชั้นนี้เป็นมัชฌิมสมมุติ ดังนั้นชั้นนี้จะมีความเบี่ยงเบนเป็น ศูนย์

จากตาราง $a = 76.5$

$$i = 6$$

$$\bar{d} = \frac{\sum_{i=1}^{k} f_i d_i}{n} = \frac{-6}{40}$$

แทนค่า $\bar{X} = 76.5 + (6)\left(\frac{-6}{40}\right)$

$= 76.5 - 0.9$

$= 75.6$

นั่นคือ มัชฌิมเลขคณิตหาโดยวิธีลัด $= 75.6$ คะแนน **ตอบ**

ข. การหามัธยฐาน (Median)

การคำนวณหาค่ามัธยฐาน สำหรับข้อมูลที่จัดหมวดหมู่เป็นตารางแจกแจงความถี่ จะหาได้โดยการเทียบบัญญัติไตรยางค์ หรือหาโดยใช้สูตรดังนี้

$$\text{Median} = L + \left[\frac{\frac{n}{2} - \Sigma f_i}{f_{med.}}\right] i$$

โดยที่ L คือ ขีดจำกัดล่างของชั้นที่มัธยฐานอยู่

n คือ จำนวนข้อมูลทั้งหมด

Σf_i คือ ผลรวมความถี่ตั้งแต่ชั้นแรกจนถึงชั้นก่อนชั้นมัธยฐาน

$f_{med.}$ คือ ความถี่ของชั้นที่ median อยู่ ซึ่งจะอยู่ตำแหน่งที่ $\frac{n}{2}$

i คือ อันตรภาคชั้น

ตัวอย่าง 1.5 การหามัธยฐานของคะแนนสอบวิชาคณิตศาสตร์จาก**ตาราง** 1.3

ขีดจำกัด	ความถี่	ความถี่สะสม
55.5 - 61.5	4	4
61.5 - 67.5	5	9
67.5 - 73.5	8	17
73.5 - 79.5	11	28
79.5 - 85.5	4	32
85.5 - 91.5	4	36
91.5 - 97.5	4	40
รวม	40	

(1) หาค่ามัธยฐาน โดยวิธีเทียบบัญญัติไตรยางค์

ตำแหน่งมัธยฐานจะอยู่ตรงข้อมูลที่ $\frac{n+1}{2} = \frac{40+1}{2} = 20.5$

เพราะฉะนั้นมัธยฐานจะอยู่ในชั้นที่ขีดจำกัด = (73.5 – 79.5)

ตำแหน่งข้อมูลแตกต่างกัน = 28 – 17 = 11 ตำแหน่ง

ค่าของข้อมูลแตกต่างกัน = 79.5 – 73.5 = 6 คะแนน

ถ้าตำแหน่งของข้อมูลแตกต่างกัน = 20.5 – 17 = 3.5 ตำแหน่ง

ค่าของข้อมูลจะแตกต่างกัน $= \frac{6 \times 3.5}{11}$ = 1.9 คะแนน

$\therefore$ มัธยฐานจะเท่ากับ 73.5 + 1.9 = 75.4 คะแนน

ตอบ

(2) หาค่ามัธยฐานโดยวิธีใช้สูตร

$$\text{Median} = L + \left[\frac{\frac{n}{2} - \Sigma f_i}{f_{med.}}\right] i$$

$L = 73.5$

$n = 40$

$\Sigma f_i = 4 + 5 + 8 = 17$

$f_{med.} = 11$

$i = 6$

แทนค่า $\text{Med.} = 73.5 + \left[\frac{\frac{40}{2} - 17}{11}\right] \times 6$

$= 73.5 + (\frac{3}{11}) \times 6$

$= 73.5 + 1.64$

$= 75.14$

นั่นคือ มัธยฐานของคะแนนสอบวิชาคณิตศาสตร์เท่ากับ 75.14 คะแนน

ตอบ

ค. การหาฐานนิยม

ฐานนิยมของข้อมูลที่จัดหมวดหมู่แล้ว จะอยู่ในชั้นที่มีความถี่มากที่สุด การคำนวณหาค่าฐานนิยม หาได้โดยใช้สูตรดังนี้

$$\text{ฐานนิยม} = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) i$$

โดยที่ L คือขีดจำกัดล่างของชั้นที่ฐานนิยมอยู่ ซึ่งจะมีความถี่มากที่สุด

Δ_1 คือผลต่างของความถี่ของชั้นที่มีฐานนิยม กับชั้นที่ถัดขึ้นมา

Δ_2 คือผลต่างของความถี่ของชั้นที่มีฐานนิยมกับชั้นที่ถัดลงไป

i คืออันตรภาคชั้น

ตัวอย่าง 1.6 การหาฐานนิยมของคะแนนสอบวิชาคณิตศาสตร์ ของนักศึกษาจำนวน 40 คน จากตาราง 1.3

ขีดจำกัด	ความถี่
55.5 - 61.5	4
61.5 - 67.5	5
67.5 - 73.5	8
73.5 - 79.5	11
79.5 - 85.5	4
85.5 - 91.5	4
91.5 - 97.5	4
รวม	40

ฐานนิยมจะอยู่ในชั้นที่มีความถี่มากที่สุด คือชั้น

(73.5 – 79.5) ซึ่งมีความถี่ = 11

$\Delta_1 = 11 - 8 = 3$

$\Delta_2 = 11 - 4 = 7$

$i = 6$

$L = 73.5$

จากสูตร $\text{ฐานนิยม} = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) i$

แทนค่า $\qquad = 73.5 + \left(\frac{3}{3+7}\right) 6$

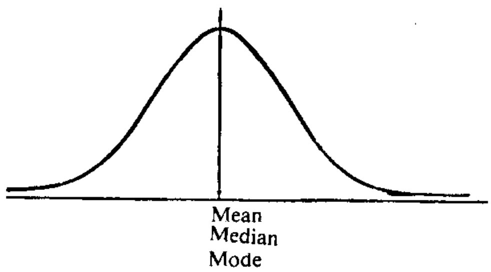
$$= 73.5 + 1.8 = 75.3$$

นั่นคือ ฐานนิยมของคะแนนสอบวิชาคณิตศาสตร์เท่ากับ 75.3 คะแนน

ตอบ

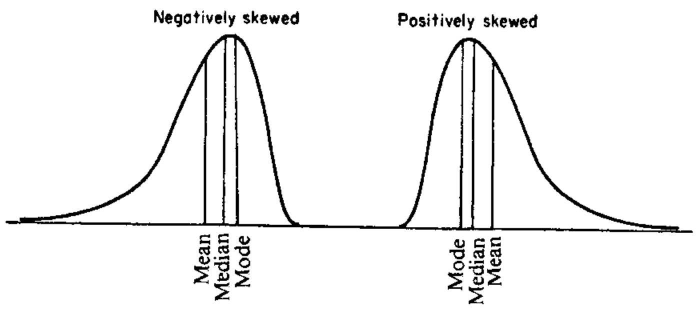
1.8.4 ความสัมพันธ์ระหว่างมัชฌิมเลขคณิต มัธยฐาน และฐานนิยม

ข้อมูลที่มีการกระจายอย่างสมมาตร คือเมื่อนำมาเขียนกราฟแล้วจะได้กราฟเท่ากันทั้งสองข้าง ดังในรูปที่ 1.7 ค่า Mean Median และ Mode จะเท่ากันหรืออยู่ใกล้เคียงกัน



รูปที่ 1.7 กราฟของข้อมูลที่มีการกระจายอย่างสมมาตร

ข้อมูลที่มีการกระจายไม่เท่ากัน เมื่อนำมาเขียนกราฟแล้วจะเห็นว่า จะเบ้ไปข้างใดข้างหนึ่ง ค่า Mean Median และ Mode จะแตกต่างกัน ดังในรูป 1.8 และ 1.9



รูป 1.8 การกระจายของข้อมูลเบ้ขวา **รูป 1.9** การกระจายของข้อมูลเบ้ซ้าย

ในกรณีที่การกระจายของข้อมูลมีความเบ้ ความสัมพันธ์ของ Mean Median และ Mode จะมีดังนี้

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

1.9 การวัดความกระจาย (Measure of Disperson)

ในการพิจารณาลักษณะของข้อมูลแต่ละชุด จะพิจารณาเฉพาะค่าตัวกลางอย่างเดียวนั้นยังไม่เพียงพอ จะต้องพิจารณาเปรียบเทียบดูการกระจายของข้อมูลด้วย ข้อมูลชุดใดที่มีการกระจายมากแสดงว่าข้อมูลชุดนั้นประกอบด้วยตัวเลขซึ่งมีค่าน้อยและมาก หรือมีค่าต่ำและสูงปนกันอยู่ หรือมีความแตกต่างกันมาก ส่วนข้อมูลที่มีการกระจายน้อยแสดงว่าข้อมูลชุดนั้นมีตัวเลขซึ่งมีค่าใกล้เคียงกัน หรือมีความแตกต่างกันน้อย ดังนั้น เพื่อให้การวิเคราะห์ข้อมูลสมบูรณ์ขึ้น จึงจำเป็นต้องวัดค่าตัวกลาง และวัดความกระจายของข้อมูลด้วยเสมอ การวัดความกระจายมีอยู่ 4 วิธีคือ

1.9.1 พิสัย (Range)

เป็นการวัดความกระจายอย่างหยาบ ๆ วัดได้ง่าย เหมาะสำหรับข้อมูลจำนวนน้อย ๆ หาได้โดย หาผลต่างของค่าสูงสุดและต่ำสุด ดังนี้

พิสัย = ค่าสูงสุด − ค่าต่ำสุด

เช่น คะแนนสอบวิชาสถิติของนักเรียน 5 คนมีดังนี้ 50 85 70 25 60 คะแนน

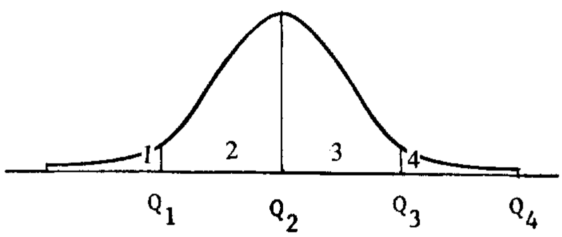
พิสัย = 85 − 25 = 60 คะแนน

นั่นคือ ความกระจายของคะแนนสอบของนักเรียน 5 คนนี้เท่ากับ 60 คะแนน

จะเห็นได้ทันทีว่า คะแนนสอบของนักเรียนแตกต่างกันมาก การหาค่าความกระจายแบบนี้ไม่ละเอียด เพราะเป็นการวัดจากข้อมูลเพียงสองค่าเท่านั้น

1.9.2 ส่วนเบี่ยงเบนควอไทล์ (Quartile Deviation)

เป็นการวัดความกระจายโดยแบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน แต่ละส่วนจะประกอบด้วยข้อมูล 25% การวัดความกระจายโดยวิธีนี้ดีกว่าพิสัย เพราะวัดจากข้อมูลถึง 50%



การหาค่า Quartile Deviation หาได้จากสูตร

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Q_1 และ Q_3 หาได้เหมือน Median โดยใช้สูตรดังนี้

$$Q_1 = L_1 + \left[\frac{\frac{n}{4} - (\Sigma f)_1}{f_1}\right] i$$

$$Q_3 = L_3 + \left[\frac{\frac{3n}{4} - (\Sigma f)_3}{f_3}\right] i$$

โดยที่ L_1 คือขีดจำกัดล่างที่มี Q_1

L_3 คือขีดจำกัดล่างที่มี Q_3

$(\Sigma f)_1$ คือ ผลรวมความถี่ของชั้นแรกจนถึงชั้นก่อน Q_1

$(\Sigma f)_3$ คือ ผลรวมความถี่ของชั้นแรกจนถึงชั้นก่อน Q_3

f_1 คือ ความถี่ของชั้นที่มี Q_1

f_3 คือ ความถี่ของชั้นที่มี Q_3

i คือ อันตรภาคชั้น

ตัวอย่าง 1.7 การหาความกระจายคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษา 40 คน ดังข้อมูลในตาราง 1.3

Q_1 จะอยู่ตำแหน่งที่ $\frac{n}{4} = \frac{40}{4} = 10$

Q_3 จะอยู่ตำแหน่งที่ $\frac{3n}{4} = \frac{3 \times 40}{4} = 30$

ขีดจำกัด	ความถี่	ความถี่สะสม
55.5 - 61.5	4	4
61.5 - 67.5	5	9
67.5 - 73.5	8 ← f_1	17
73.5 - 79.5	11	28
79.5 - 85.5	4 ← f_3	32
85.5 - 91.5	4	36
91.5 - 97.5	4	40
รวม	40	

จากสูตร $Q_1 = L_1 + \left[\dfrac{\frac{n}{4} - (\Sigma f)_1}{f_1}\right] i$

จากตารางแจกแจงความถี่จะได้

$n = 40$

$L_1 = 67.5$

$(\Sigma f)_1 = 9$

$f_1 = 8$

$i = 6$

แทนค่า $Q_1 = 67.5 + \left[\dfrac{\frac{40}{4} - (9)}{8}\right] 6$

$= 68.25$

จากสูตร $Q_3 = L_3 + \left[\dfrac{\frac{3n}{4} - (\Sigma f)_3}{f_3}\right] i$

จากตารางแจกแจงความถี่จะได้

$(\Sigma f)_3 = 28$

$f_3 = 4$

แทนค่า $Q_3 = 79.5 + \left[\dfrac{\frac{3 \times 40}{4} - 28}{4}\right] 6$

$= 82.50$

เมื่อหาค่า Q_1 และ Q_3 ได้แล้ว นำค่าที่ได้ไปหาค่าส่วนเบี่ยงเบนควอไทล์ ดังนี้

จากสูตร $Q.D = \dfrac{Q_3 - Q_1}{2}$

$= \dfrac{82.50 - 68.25}{2}$

$= \dfrac{14.25}{2}$

$= 7.12$

นั่นคือ ความกระจายของคะแนนสอบเท่ากับ 7.12 คะแนน **ตอบ**

1.9.3 ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation หรือ Average Deviation)

การวัดความกระจายโดยวิธีนี้วัด ดีกว่าสองวิธีแรก เป็นการวัดที่เกี่ยวกับข้อมูลทุกตัว

โดยศึกษาดูความเบี่ยงเบนของข้อมูลดิบกับมัชฌิมเลขคณิต ในการคำนวณทำได้ดังนี้

1) สำหรับข้อมูลที่ยังไม่ได้จัดหมวดหมู่ (Ungrouped Data)

ในการคำนวณหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่ไม่ได้จัดหมวดหมู่ หาได้โดยใช้สูตร

$$\text{M.D.} = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}$$

โดยที่ X_i คือ ข้อมูลแต่ละตัว

$\bar{X}$ คือมัชฌิมเลขคณิตของข้อมูล

$|X_i - \bar{X}|$ คือ ค่า Absolute ของข้อมูลที่เบี่ยงเบนไปจากมัชฌิม (ไม่คิดเครื่องหมาย)

n คือ จำนวนข้อมูล

การหาความกระจายแบบนี้พิจารณาดูคะแนนสอบวิชาสถิติของนักศึกษา 5 คน ดังนี้ 50 85 70 25 60 คะแนน

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$= \frac{50 + 85 + 70 + 25 + 60}{5} = \frac{290}{5} = 58$$

$$\text{M.D.} = \frac{\sum_{i=1}^{k} |X_i - \bar{X}|}{n}$$

แทนค่า ; $= \dfrac{|50 - 58| + |85 - 58| + |70 - 58| + |25 - 58| + |60 - 58|}{5}$

$$= \frac{8 + 27 + 12 + 33 + 2}{5}$$

$$= \frac{82}{5} = 16.4 \text{ คะแนน}$$

นั่นคือ ส่วนเบี่ยงเบนเฉลี่ยของคะแนนสอบวิชาสถิติ = 16.4 คะแนน **ตอบ**

2) สำหรับข้อมูลที่จัดหมวดหมู่แล้ว

ในการคำนวณหาส่วนเบี่ยงเบนเฉลี่ย ที่ข้อมูลจัดหมวดหมู่ เป็นตารางแจกแจงความถี่ หาได้โดยใช้สูตร

$$\text{M.D.} = \frac{\sum_{i=1}^{k} f_i |X_i - \bar{X}|}{n}$$

โดยที่ f_i คือความถี่ของข้อมูลชั้นที่ i

X_i คือค่ากึ่งกลางของข้อมูลชั้นที่ i

k คือจำนวนชั้น

n คือ จำนวนข้อมูล

ตัวอย่าง 1.8 การหาส่วนเบี่ยงเบนเฉลี่ยของคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษาจำนวน 40 คน จากข้อมูลที่ได้แจกแจงในตาราง 1.3

ขีดจำกัด	ค่ากึ่งกลาง X_i	ความถี่ f_i	f_iX_i	$\lvert X_i - \bar{X} \rvert$	$f_i \lvert X_i - \bar{X} \rvert$
55.5 - 61.5	58.5	4	234.0	17.1	68.4
61.5 - 67.5	64.5	5	322.5	11.1	55.5
67.5 - 73.5	70.5	8	564.0	5.1	40.8
73.5 - 79.5	76.5	11	841.5	0.9	9.9
79.5 - 85.5	82.5	4	330.0	6.9	27.6
85.5 - 91.5	88.5	4	354.0	12.9	51.6
91.5 - 97.5	94.5	4	378.0	18.9	76.6
รวม		40	3024		329.4

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

$$= \frac{3024}{40} = 75.6 \text{ คะแนน}$$

$$\text{M.D.} = \frac{\sum_{i=1}^{k} f_i \mid X_i - \bar{X} \mid}{n}$$

$$= \frac{329.4}{40}$$

$$= 8.235 \text{ คะแนน}$$

นั่นคือ ความกระจายของคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษาซึ่งวัดโดยส่วนเบี่ยงเบนเฉลี่ย ได้เท่ากับ 8.235 คะแนน **ตอบ**

1.9.4 **ส่วนเบี่ยงเบนมาตรฐาน** (Standard Deviation)

การวัดความกระจายโดยวิธีนี้เป็นวิธีที่นิยมมากที่สุด การคำนวณคล้ายกับการหาส่วนเบี่ยงเบนเฉลี่ย แต่ได้แก้ปัญหาความเบี่ยงเบนของข้อมูลที่ติดลบโดยใช้ยกกำลังสองแทนเครื่องหมาย Absolute ซึ่งค่าเบี่ยงเบนที่ยกกำลังสองแล้วจะเป็นบวกทั้งหมด

1) **สำหรับข้อมูลที่ยังไม่ได้จัดหมวดหมู่** (Ungrouped Data)

ข้อมูลที่ยังไม่ได้จัดหมวดหมู่ มีวิธีคำนวณได้โดยใช้สูตรดังนี้

$$\text{ส่วนเบี่ยงเบนมาตรฐานของประชากร } (\sigma) = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

$$\text{ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง } (S) = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$$

โดยที่ X_i คือข้อมูลดิบ

$\bar{X}$ คือมัชฌิมเลขคณิตของข้อมูล

N คือจำนวนข้อมูลของประชากร

n คือจำนวนตัวอย่าง

สำหรับประชากรที่มีจำนวนไม่มาก สามารถนำมาหาค่าเฉลี่ยได้ทุกประการ ค่าเฉลี่ยของประชากรจะเท่ากับ มัชฌิมเลขคณิตซึ่งจะหาส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

$$\text{ส่วนเบี่ยงเบนมาตรฐานของประชากร} = \sigma = S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{N}}$$

จากคะแนนสอบวิชาสถิติของนักศึกษา ที่สุ่มตัวอย่างขึ้นมา 5 คน มีคะแนน 50 85 70 25 60 คะแนน นำมาหาส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

$$\bar{X} = \frac{\Sigma X_i}{n} = \frac{50 + 85 + 70 + 25 + 60}{5} = \frac{290}{5} = 58$$

$$S = \sqrt{\frac{(50-58)^2 + (85-58)^2 + (70-58)^2 + (25-58)^2 + (60-58)^2}{5-1}}$$

$$= \sqrt{\frac{64 + 729 + 144 + 1089 + 4}{4}} = \sqrt{\frac{2030}{4}} = 22.52$$

นั่นคือ ความกระจายของคะแนนสอบวิชาสถิติ เมื่อวัดโดยส่วนเบี่ยงเบนมาตรฐานจะเท่ากับ 22.52 คะแนน

2) สำหรับข้อมูลที่จัดหมวดหมู่แล้ว (Grouped Data)

ข้อมูลที่จัดหมวดหมู่แล้ว นำไปคำนวณหาส่วนเบี่ยงเบนมาตรฐาน โดยใช้สูตรดังนี้

ส่วนเบี่ยงเบนมาตรฐานของประชากร $= \sigma = \sqrt{\dfrac{\sum_{i=1}^{k} f_i (X_i - \mu)^2}{n}}$

ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง $= S = \sqrt{\dfrac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^2}{n-1}}$

ในกรณีที่ข้อมูลมีจำนวนมาก การคำนวณโดยใช้สูตรนี้ จะยุ่งยากในการคำนวณไม่เป็นที่นิยม นักสถิติได้หันมาใช้สูตรในการคำนวณวิธีลัด ซึ่งที่นิยมมี 2 แบบคือ

$$S = \sqrt{\frac{\Sigma f_i X_i^2}{n} - \left(\frac{f_i X_i}{n}\right)^2} \qquad \text{สูตร (1)}$$

และ
$$S = i\sqrt{\frac{\Sigma f_i d_i^2}{n} \quad \left(\frac{\Sigma f_i d_i}{n}\right)^2} \qquad \text{สูตร (2)}$$

โดยที่ $d = \dfrac{X_i - a}{i}$

i คือ อันตรภาคชั้น

f_i คือ ความถี่ชั้นที่ i

X_i คือ กึ่งกลางของชั้นที่ i

n คือ จำนวนข้อมูล

k คือ จำนวนชั้น

ตัวอย่าง 1.9 การหาส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษา 40 คน ดังมีข้อมูลในตาราง 1.3 หาโดยใช้สูตร (1)

ค่ากึ่งกลาง(X_i)	ความถี่ (f_i)	f_iX_i	X_i^2	$f_iX_i^2$
58.5	4	234.0	3,422.2	13,688.8
64.5	5	322.5	4,160.2	20,801.0
70.5	8	564.0	4,970.2	39,761.6
76.5	11	841.5	5,852.2	64,374.2
82.5	4	330.0	6,806.2	27,224.8
88.5	4	354.0	7,832.2	31,328.8
94.5	4	378.0	8,930.2	35,720.8
รวม	40	3,024		232,900

จากสูตร $S = \sqrt{\frac{\Sigma f_i x_i^2}{n} - \left(\frac{\Sigma f_i x_i}{n}\right)^2}$

$\Sigma f_i x_i^2 = 232,900$

$\Sigma f_i x_i = 3,024$

$n = 40$

แทนค่า $S = \sqrt{\frac{232,000}{40} - \left(\frac{3,024}{40}\right)^2}$

$= \sqrt{107.14} = 10.35$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐาน หาโดยวิธีลัด ใช้สูตรที่ 1 ได้เท่ากับ 10.35 คะแนน

ตัวอย่าง 1.10 การหาส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของนักศึกษา 40 คน โดยใช้สูตรคำนวณวิธีลัด สูตรที่ 2

ค่ากึ่งกลาง (X_i)	f_i	d_i	$f_i d_i$	$f_i d_i^2$
58.5	4	−3	−12	36
64.5	5	−2	−10	20
70.5	8	−1	−8	8
76.5	11	0	0	0
82.5	4	+1	4	4
88.5	4	+2	8	16
94.5	4	+3	12	36
รวม			−6	120

จากสูตร $S = i\sqrt{\frac{\Sigma f_i d_i^2}{n} - \frac{(\Sigma f_i d_i)^2}{n}}$

$\Sigma f_i d_i^2 = 120$

$\Sigma f_i d_i = -6$

$n = 40$

แทนค่า $S = 6\sqrt{\frac{120}{40} - \frac{(-6)^2}{40}} = 6\sqrt{3 - 0.0225}$

$= 6 \times 1.725 \qquad = 10.35$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ ซึ่งหาโดยวิธีลัด โดยใช้สูตรที่ 2 ได้เท่ากับ 10.35 คะแนน

เปรียบเทียบการวัดความกระจาย

1) ข้อมูลที่ไม่ได้จัดหมวดหมู่ของคะแนนสอบของนักเรียน 5 คน ซึ่งสอบได้ 50 85 70 25 และ 60 คะแนน

2) คะแนนสอบของนักเรียนที่จัดหมวดหมู่เป็นตารางแจกแจงความถี่

ตามตารางที่ 1.3

	ความกระจาย	
	คะแนนไม่จัดหมวดหมู่ (1)	คะแนนที่จัดหมวดหมู่ (2)
1. พิสัย	60	-
2. ส่วนเบี่ยงเบนควอไทล์	-	14.25
3. ส่วนเบี่ยงเบนเฉลี่ย	16.4	8.235
4. ส่วนเบี่ยงเบนมาตรฐาน	23.9	10.35

1.10 การหาตำแหน่งของข้อมูล Percentiles, Deciles และ Quartiles

Percentiles, Deciles และ Quartiles ใช้สำหรับเรียกตำแหน่งที่ของข้อมูลเพื่อให้ทราบว่าข้อมูลแต่ละข้อมูลนั้นจะมีตำแหน่งอยู่ ณ ที่ใด เป็นอันดับเท่าไรของกลุ่มข้อมูลนั้น ๆ

1.10.1 Percentiles

ใช้สำหรับเรียกตำแหน่งของข้อมูล เพื่อให้เราทราบว่าข้อมูลนั้นมีตำแหน่งเป็นอันดับที่เท่าไร โดยแบ่งข้อมูลทั้งหมดออกเป็น 100 ส่วนเท่า ๆ กัน

1.10.2 Deciles

ใช้สำหรับเรียกตำแหน่งของข้อมูล เพื่อให้เราทราบว่าข้อมูลนั้นมีตำแหน่งเป็นอันดับที่เท่าไร โดยแบ่งข้อมูลนั้นออกเป็น 10 ส่วนเท่า ๆ กัน

1.10.3 Quartiles

ใช้สำหรับเรียกตำแหน่งของข้อมูล เพื่อให้เราทราบว่าข้อมูลนั้นมีตำแหน่งเป็นอันดับที่เท่าไร โดยแบ่งข้อมูลนั้นออกเป็น 4 ส่วนเท่า ๆ กัน

ความสัมพันธ์ระหว่างเปอร์เซนไทล์ เดไซล์ และควอไทล์

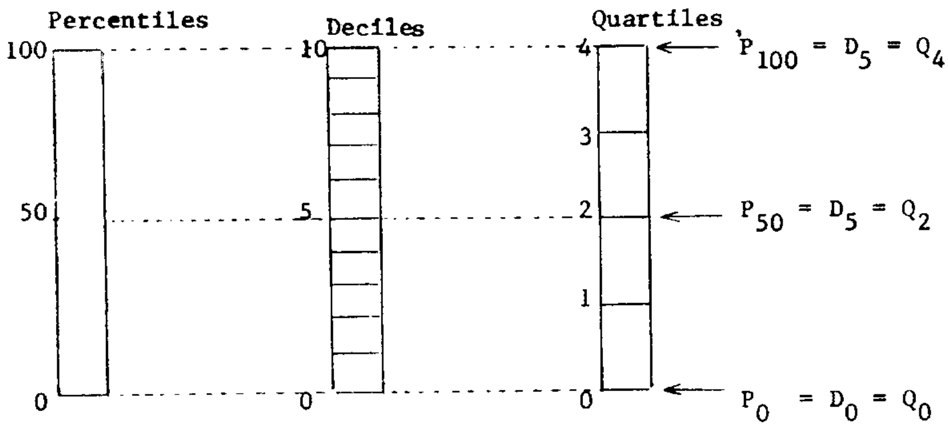
ในการหาตำแหน่ง Percentiles, Deciles และ Quartiles จากรูป 1.10 P, D และ Q มีความสัมพันธ์กันดังนี้ :-

$P_O = D_O = Q_O$, $P_{50} = D_5 = Q_2$ และ $P_{100} = D_{10} = Q_4$

นอกจากนี้ยังมีความสัมพันธ์ระหว่าง P และ Q โดยที่ $P_{25} = Q_1 =$ และ $P_{75} = Q_3$

และความสัมพันธ์ระหว่าง P กับ D โดยที่ $D_1 = P_{10}$, $D_2 = P_{20}$, $D_3 = P_{30}$ ฯลฯ

Quartiles ที่ 1 คือตำแหน่งของข้อมูลซึ่ง 25% ของข้อมูลมีค่าน้อยกว่าตำแหน่งนี้ และ 75% ของข้อมูล มีค่ามากกว่า ซึ่งที่จุดนี้เรียกว่า Percentile ที่ 25 เช่นกัน ในด้านการศึกษานิยมใช้ Quartiles และ Deciles ในการประเมินผล เช่น พิจารณาคะแนนนักเรียนที่ต่ำกว่า 25% หมายถึงคะแนนไม่เกิน Quartile ที่ 1 ถ้าคะแนนของนักเรียนสอบได้ 25% พอดี หมายถึงคะแนนของนักเรียนอยู่ที่ตำแหน่ง Quartile ที่ 1 ในทำนองเดียวกัน ถ้านักเรียนสอบได้คะแนนตำแหน่ง Quartiles ที่ 3 หมายถึงนักเรียนสอบได้คะแนน 75% พอดี และถ้าผลสอบของคะแนนของนักเรียนอยู่สูงกว่า Quartile ที่ 3 หมายความว่าคะแนนจะอยู่ระหว่าง 75% ถึง 100%



รูป 1.10 เปรียบเทียบ Percentiles, Deciles และ Quartiles

การคำนวณหา Percentiles, Deciles **และ** Quartiles

ในการหาตำแหน่งที่ของข้อมูล หาได้โดยการเทียบบัญญัติไตรยางค์ ทั้งข้อมูลที่จัดกลุ่มและยังไม่ได้จัดกลุ่ม

1) **สำหรับข้อมูลที่ยังไม่ได้จัดหมวดหมู่** (Ungrouped Data)

ตัวอย่าง 1.11 พิจารณาข้อมูลชุดหนึ่งมีเลขอยู่ 6 ตัว 2, 9, 10, 4, 5, 6

ก. วิธีหาเปอร์เซนไทล์

การหา Percentile ให้เรียงลำดับข้อมูลจากน้อยไปหามากดังนี้ 2, 4, 5, 6, 9, 10 แล้วนำไปเปรียบเทียบอัตราส่วนกับ 100

สมมุติต้องการหา P_{50}

จากตำแหน่งเทียบที่ 100 เท่ากับข้อมูลจริงในตำแหน่งที่ 6

จากตำแหน่งเทียบที่ 50 เท่ากับข้อมูลจริงในตำแหน่งที่ $\frac{6 \times 50}{100} = 3$

∴ ข้อมูลจริงในตำแหน่งที่ 3 = 5

นั่นคือ P_{50} ของข้อมูล เท่ากับ 5

สมมุติต้องการหา P_{30}

จากตำแหน่งเทียบ 100 เท่ากับข้อมูลจริงในตำแหน่งที่ 6

จากตำแหน่งเทียบ 30 เท่ากับข้อมูลจริงในตำแหน่งที่ $\frac{6 \times 30}{100} = 1.8$

P_{30} จะอยู่ระหว่างเลข 2 และ 4 ต้องหาค่าโดยวิธี interpolate

ตำแหน่งคะแนนจริงต่างกัน 1 ตำแหน่ง ค่าของข้อมูลต่างกัน = 2

ถ้าคะแนนจริงต่างกัน .8 ตำแหน่ง ค่าของข้อมูลต่างกัน = $\frac{2 \times .8}{1}$

$= 1.6$

∴ P_{30} ของข้อมูลเท่ากับ $2 + 1.6 = 3.6$

ข. วิธีหาเดไซล์

การหาตำแหน่งที่และค่าของ Decile ก็ทำเหมือน Percentile แต่เทียบ 10 ส่วน แทน 100 ส่วน โดยเรียงลำดับของข้อมูลจากน้อยไปมาก ดังนี้ 2, 4, 5, 6, 9, 10

สมมุติต้องการหา D_5

จากตำแหน่งเทียบที่ 10 เท่ากับ ค่าของข้อมูลจริงในตำแหน่งที่ 6

จากตำแหน่งเทียบที่ 5 เท่ากับ ค่าของข้อมูลจริงในตำแหน่งที่ $\frac{6 \times 5}{10} = 3$

∴ ข้อมูลจริงตำแหน่งที่ 3 คือ 5

นั่นคือ D_5 ของข้อมูลเท่ากับ 5

สมมุติต้องการหา D_9

จากตำแหน่งเทียบที่ 10 เท่ากับ ค่าของคะแนนจริงตำแหน่งที่ 6

จากตำแหน่งเทียบที่ 9 เท่ากับ ค่าของคะแนนจริงตำแหน่งที่ $\frac{6 \times 9}{10} = 5.4$

∴ ข้อมูลที่ D_9 จะอยู่ระหว่าง 9 และ 10 ต้องหาค่าโดยวิธี interpolate

ตำแหน่งคะแนนจริงต่างกัน 1 ตำแหน่ง ค่าของข้อมูลต่างกัน 1

ตำแหน่งคะแนนจริงต่างกัน (5.4-5) = .4 ค่าของข้อมูลต่างกัน $\frac{1 \times .4}{1} = .4$

∴ D_9 ของข้อมูลจะเท่ากับ $9 + .4 = 9.4$

ค. วิธีหา Quartiles

การหาตำแหน่งที่ของ Quartiles ก็ทำเช่นเดียวกับการหา Percentiles และ Deciles แต่ใช้เทียบอัตราส่วนตำแหน่งของข้อมูลสุดท้ายเท่ากับ 4 แทนในการหานั้นต้องเรียงลำดับของข้อมูล เช่น ต้องการหา Quartiles ของเลขซึ่งเรียงลำดับแล้วดังนี้ 2, 4, 5, 6, 9, 10

สมมุติต้องการหา Q_2

จากตำแหน่งเทียบ 4 เท่ากับ ค่าของข้อมูลจริงในตำแหน่งที่ 6

จากตำแหน่งเทียบ 2 เท่ากับ ค่าของข้อมูลจริงในตำแหน่งที่ $\frac{6 \times 2}{4} = 3$

$\therefore$ ข้อมูลจริง ตำแหน่งที่ 3 = 5

นั่นคือ Q_2 เท่ากับ 5

สมมุติต้องการหา Q_3

จากตำแหน่งเทียบที่ 4 เท่ากับ ค่าของข้อมูลจริงตำแหน่งที่ 6

จากตำแหน่งเทียบที่ 3 เท่ากับ ค่าของข้อมูลจริงตำแหน่งที่ $\frac{6 \times 3}{4} = 4.5$

Q_2 จะอยู่ระหว่าง 6 และ 9 หาค่าระหว่างโดยวิธี interpolate ดังนี้

ตำแหน่งของข้อมูลจริงต่างกัน 1 ตำแหน่ง ค่าของข้อมูลจะต่างกัน 3

ตำแหน่งของข้อมูลจริงต่างกัน (4.5 - 4) = .5 ค่าของข้อมูลจะต่างกัน $\frac{3 \times .5}{1}$

$$= 1.5$$

$\therefore$ ข้อมูลที่ $Q_3 = 6 + 1.5 = 7.5$

2) สำหรับข้อมูลที่จัดหมวดหมู่แล้ว (Grouped Data)

สำหรับข้อมูลที่จัดหมวดหมู่เป็นตารางแจกแจงความถี่ ในการหาตำแหน่งที่ของ Percentiles, Deciles และ Quartiles หาได้โดยเทียบบัญญัติไตรยางค์เช่นกัน ในที่นี้จะหาตำแหน่งที่ของ Percentiles และ Deciles ส่วนตำแหน่งที่ของ Quartiles ให้ดูหัวข้อ 1.9 เรื่องการวัดความกระจาย

ตัวอย่าง 1.12 การหา Percentile และ Decile คะแนนสอบวิชาคณิตศาสตร์ของนักศึกษาจำนวน 40 คน ตามข้อมูลในตาราง 1.3

ขีดจำกัด	ค่ากึ่งกลาง	ความถี่	ความถี่สะสม
55.5 - 61.5	58.5	4	4
61.5 - 67.5	64.5	5	9
67.5 - 73.5	70.5	8	17
73.5 - 79.5	76.5	11	28
79.5 - 85.5	82.5	4	32
85.5 - 91.5	88.5	4	36
91.5 - 97.5	94.5	4	40
รวม		40	40

สมมุติหาค่า P_{50}

ตำแหน่งเทียบที่ 100 เท่ากับ ค่าของคะแนนจริงตำแหน่งที่ 40

ตำแหน่งเทียบที่ 50 เท่ากับ ค่าของคะแนนจริงตำแหน่งที่ $\frac{40\times 50}{100} = 20$

ข้อมูลตำแหน่งที่ 20 จะอยู่ในชั้นของข้อมูลที่มีค่าระหว่าง (73.5 - 79.5)

จากข้อมูล ตำแหน่งของข้อมูลต่างกัน = 28 - 17 = 11 ตำแหน่ง

ค่าของคะแนนจริงต่างกัน = 79.5-73.5 = 6 คะแนน

จากข้อมูล ตำแหน่งของข้อมูลต่างกัน = 20 - 17 = 3 ตำแหน่ง

ค่าของคะแนนจริงจะต่างกัน $= \frac{6\times 3}{11} = 1.64$ คะแนน

$\therefore P_{50}$ ของข้อมูลมีค่า = 73.5+1.64

= 75.14 คะแนน **ตอบ**

หาค่า D_3

ตำแหน่งเทียบที่ 10 เท่ากับ ค่าของคะแนนจริงตำแหน่งที่ 40

ตำแหน่งเทียบที่ 3 เท่ากับ ค่าของคะแนนจริงตำแหน่งที่ $\frac{40\times 3}{10} = 12$

ข้อมูลตำแหน่งที่ 12 จะอยู่ในชั้นของข้อมูล มีค่าระหว่าง (67.5 - 73.5)

จากข้อมูล ตำแหน่งของข้อมูลต่างกัน = 17 - 9 = 8 ตำแหน่ง

ค่าของคะแนนจริงจะต่างกัน	= 73.5-67.5 = 6 คะแนน	
จากข้อมูล ตำแหน่งของข้อมูลต่างกัน	= 12 - 9 = 3 ตำแหน่ง	
ค่าของคะแนนจริง จะต่างกัน	$= \frac{6\times3}{8} = 2.25$ คะแนน	
$\therefore D_3$ ของข้อมูล มีค่า	= 67.5+2.25	
	= 69.75 คะแนน	**ตอบ**

หมายเหตุ การหา P, D และ Q จะหาได้โดยใช้สูตรก็ได้ แต่ค่อนข้างจะยุ่งยากจึงไม่ได้กล่าวไว้ในที่นี้

1 อัตราส่วน สัดส่วน และเปอร์เซนต์ (Ratio, Proportion and Percent)

ความสัมพันธ์ระหว่างข้อมูลต่าง ๆ เช่น ข้อมูลผลิตผลการเกษตร จำนวนประชากร รายได้ ฯลฯ ในลักษณะอัตราส่วน สัดส่วน และเปอร์เซนต์มีประโยชน์ในการเปรียบเทียบ และสรุปผลทางสถิติ ซึ่งง่ายและสะดวกในการนำไปใช้

1.11.1 อัตราส่วน (Ratio)

อัตราส่วนเป็นการเปรียบเทียบ ซึ่งส่วนมากใช้กับข้อมูลอย่างเดียวกัน เช่น เปรียบเทียบจำนวนประชากร คะแนนสอบนักเรียน รายได้ของประชาชน เป็นต้น อัตราส่วนอาจจะเรียกอีกอย่างหนึ่งว่า เศษส่วน เช่น เศษส่วน $\frac{a}{b}$ เขียนอีกอย่างหนึ่งได้ดังนี้ a : b เรียกว่า a ต่อ b โดยที่ b เป็นฐาน (base) ถ้าเปรียบเทียบกับฐาน 100 เรียกว่า เปอร์เซ็นต์ ถ้าฐานเป็น 1,000 เรียกว่า ต่อพัน (permille)

ตัวอย่างเช่น หมู่บ้านแห่งหนึ่งมีประชากรทั้งสิ้น 753 คน เป็นเพศชายเสีย 251 คน เพศหญิง 502 คน จำนวนประชากรของหมู่บ้านแห่งนี้จะมีอัตราส่วนของเพศชายต่อเพศหญิงจะเป็น 251 : 502 = 1 : 2 การเปรียบเทียบเพศชายและเพศหญิงดังกล่าวข้างต้น เป็นการเปรียบเทียบจำนวน ต่อจำนวนอย่างเดียวกัน

อัตราส่วนที่เปรียบเทียบความสัมพันธ์ของข้อมูลชนิดหนึ่ง ต่อหนึ่งหน่วยของข้อมูลอีกชนิดหนึ่ง เรียกว่า อัตรา (rate) เช่น เปรียบเทียบระยะทางกับหนึ่งหน่วยของเวลาเป็นอัตราความเร็ว เปรียบเทียบรายได้ของประชากรต่อเวลาหนึ่งเดือนเป็นอัตราเงินเดือน เปรียบเทียบจำนวนประชากรต่อหนึ่งหน่วยพื้นที่เป็นความหนาแน่นของประชากร ตัวอย่างเช่น หมู่บ้านที่

กล่าวข้างต้นนี้มีประชากร 753 คน มีเนื้อที่ 30 ตารางกิโลเมตร หมู่บ้านแห่งนี้จะมี ความหนาแน่นของประชากร เท่ากับ 753 คน/30 ตารางกิโลเมตร เท่ากับ 25.1 คน/ตารางกิโลเมตร

อัตราส่วนของผลผลิต จะเป็นผลผลิตทางด้านเกษตรกรรม หรืออุตสาหกรรม เช่น การเปรียบเทียบผลผลิตในรอบปีต่าง ๆ หรือ ในหนึ่งหน่วยของเวลา เพื่อดูแนวโน้มของผลผลิตว่าเพิ่มขึ้นหรือลดลง หรือ อาจจะให้เปรียบเทียบผลิตผลที่เสียหายก็ได้เช่นกัน

อัตราส่วนคดีอาชญากรรม เป็นการเปรียบเทียบคดีอาชญากรรมประเภทต่าง ๆ ในรอบปี เปรียบเทียบกับปีอื่น ๆ เพื่อดูแนวโน้มของอาชญากรรมประเภทต่าง ๆ ว่าสูงขึ้นหรือลดลง เพื่อจะช่วยประกอบในการวินิจฉัยวางแผนหาทางป้องกันต่อไป

1.11.2 สัดส่วน (Proportion)

สัดส่วนเป็นอัตราส่วนชนิดหนึ่ง แต่เป็นอัตราส่วนระหว่างส่วนหนึ่งของข้อมูลต่อข้อมูลทั้งหมด สัดส่วนแต่ละสัดส่วนจะมีค่าน้อยกว่าหนึ่งเสมอ เช่น ประชากรกลุ่มหนึ่งประกอบด้วยข้อมูล 2 กลุ่ม กลุ่มที่หนึ่งมีจำนวนข้อมูล a หน่วย กลุ่มที่สองมีจำนวนข้อมูล b หน่วย

สัดส่วนของกลุ่มที่หนึ่ง $= \frac{a}{a+b}$

สัดส่วนของกลุ่มที่สอง $= \frac{b}{a+b}$

ตัวอย่างจากจำนวนประชากรในหมู่บ้านที่มีประชากรทั้งสิ้น 753 คน
เป็นเพศชาย 251 คน และ เพศหญิง 502 คน นำมาหาสัดส่วนได้ดังนี้

สัดส่วนเพศชาย $= \frac{251}{753} = \frac{1}{3}$

สัดส่วนเพศหญิง $= \frac{502}{753} = \frac{2}{3}$

อัตราการเกิดและอัตราการตาย (Birth rates and Death rates)

เป็นการหาสัดส่วนของจำนวนคนเกิด คนตาย ในรอบปีเทียบกับจำนวนประชากรทั้งหมดในกลางปีนั้น เช่น สัดส่วนของคนเกิด คนตาย ของประชากรในตำบลหนึ่งมีดังนี้

ตารางที่ 1.12 แสดงอัตราการเกิดการตาย

ปี พ.ศ.	จำนวนคนเกิด	จำนวนคนตาย	จำนวนประชากรในกลางปี	การเกิด		การตาย	
				สัดส่วน	เปอร์เซนต์	สัดส่วน	เปอร์เซนต์
2519	155	45	6,552	$\frac{155}{6,552}$	2.36	$\frac{45}{6,552}$	0.68
2520	162	64	6,712	$\frac{162}{6,712}$	2.41	$\frac{64}{6,712}$	0.95
2521	186	68	6,811	$\frac{186}{6,811}$	2.73	$\frac{68}{6,811}$	0.99

1.11.3 เปอร์เซนต์ (Percent)

เปอร์เซนต์หรือร้อยละ ก็เป็นอัตราส่วนชนิดหนึ่งซึ่งเทียบต่อ 100 อัตราส่วนเป็นเปอร์เซนต์เป็นที่รู้จักกันแพร่หลายทั้งในหน่วยงานราชการและวงการธุรกิจเอกชน การคำนวณก็ทำได้ง่าย โดยเอา 100 ไปคูณอัตราส่วนที่ต้องการหาผลลัพธ์ที่ได้จะเป็นเปอร์เซนต์

ตัวอย่างเช่น ในโรงพยาบาลแห่งหนึ่ง มีคนไข้ในอยู่ 750 คน แยกเป็นคนไข้ประเภทต่าง ๆ ดังนี้ คนไข้โรคทรวงอก 180 คน คนไข้ระบบทางเดินอาหาร 154 คน คนไข้ระบบประสาท 145 คน คนไข้โรคตา หู คอ จมูก 112 คน ที่เหลือเป็นคนไข้โรคอื่น ๆ 159 คน เราหาเปอร์-เซนต์ของคนไข้ประเภทต่าง ๆ ได้ดังนี้

คนไข้โรคทรวงอก $= \frac{180}{750}(100) = 24\%$

คนไข้โรคทางเดินอาหาร $= \frac{154}{750}(100) = 20.53\%$

คนไข้โรคประสาท $= \frac{145}{750}(100) = 19.33\%$

คนไข้โรค ตา หู คอ จมูก $= \frac{112}{750}(100) = 14.93\%$

คนไข้โรคอื่น ๆ $= \frac{159}{750}(100) = 21.20\%$

สรุปผลจำนวนคนไข้แยกประเภทเป็นเปอร์เซนต์ได้ดังนี้

ตารางที่ 1.13 แสดงเปอร์เซ็นต์ของคนไข้ประเภทต่าง ๆ

ประเภทคนไข้	จำนวน	สัดส่วน	เปอร์เซ็นต์
โรคทรวงอก	180	0.240	24.00
โรคทางเดินอาหาร	154	0.205	20.53
โรคประสาท	145	0.193	19.33
โรค ตา หู คอ จมูก	112	0.149	14.93
โรคอื่น ๆ	159	0.212	21.20
รวม	750	1	100

เลขดรรชนี (Index number)

เป็นตัวเลขที่คำนวณขึ้นเพื่อใช้เปรียบเทียบดูแนวโน้มของสิ่งต่าง ๆ เช่น เปรียบเทียบดูราคาสินค้า ค่าครองชีพ รายได้ การเกิดคดีอาชญากรรม การเกิดอุบัติเหตุ เป็นต้น โดยพิจารณาในช่วงเวลาต่าง ๆ เป็นเดือน หรือเป็นปี โดยทั่วไปนิยมเทียบเป็นเปอร์เซนต์ ในการสร้างเลขดรรชนีให้กำหนดตัวเลขของข้อมูลที่ต้องการทราบปีใดปีหนึ่งเป็นฐานใช้ในการเปรียบเทียบกับปีอื่น ๆ โดยที่ในปีนั้น เลขดรรชนีกำหนดให้เท่ากับ 100 (ดูตาราง 1.15)

การเลือกฐานและการเปลี่ยนฐาน

ฐานในการเปรียบเทียบเพื่อคิดเปอร์เซนต์ นั้นมีผลต่อการลดหรือเพิ่มจำนวน ถ้าไม่ทำความเข้าใจให้ดีอาจจะนำไปสู่ความผิดพลาดในการสรุปผลได้ เช่นเลขจำนวน 10 เพิ่มขึ้น 100% จำนวนใหม่จะเป็น 20 จำนวนใหม่นี้ถ้าต้องการลดลงเป็นจำนวนเดิม ต้องลดลง 50% ไม่ใช่ลดลง 100% เนื่องจากมีการเปลี่ยนฐานจาก 10 เป็น 20 ในทำนองเดียวกับจำนวนเดิม 10 นี้ ถ้าลดลง 50% จะได้จำนวนใหม่เป็น 5 จำนวนใหม่นี้จะเพิ่มให้เป็น 10 เท่าเดิม ต้องเพิ่ม 100% จึงจะได้เท่าจำนวนเดิม

ตัวอย่างการเพิ่มและลดเปอร์เซนต์เมื่อมีการเปลี่ยนฐาน

ตาราง 1.14 การเพิ่มและลดเปอร์เซ็นต์เมื่อมีการเปลี่ยนฐาน

จำนวนเดิม	เปอร์เซนต์เพิ่ม	จำนวนใหม่	เปอร์เซนต์ลดเพื่อให้ได้จำนวนเดิม
10	1.00	10.10	0.99
10	5.00	10.50	4.76
10	10.00	11.00	9.00
10	25.00	12.50	20.00
10	33.33	13.33	25.00
10	50.00	15.00	33.33
10	100.00	20.00	50.00
10	200.00	30.00	66.67
10	500.00	60.00	83.33

ตัวอย่างเกี่ยวกับการเพิ่มและลดเปอร์เซ็นต์เมื่อฐานไม่เปลี่ยนแปลง

พิจารณาการส่งสินค้าออกจากประเทศไทยไปประเทศสหรัฐอเมริกา ระหว่างปี 2512-2517 มีดังนี้

ปี พ.ศ.	2512	2513	2514	2515	2516	2517
จำนวนล้านบาท	2,168	1,985	2,264	2,841	3,261	3,983

เปรียบเทียบการส่งสินค้าออกในปีอื่น ๆ กับปี 2512 ซึ่งสินค้าออกเป็นมูลค่า 2,168 ล้านบาท จำนวนนี้จะเป็นฐานคงที่นำไปเปรียบเทียบกับปีอื่น ๆ

การส่งสินค้าออกปี 2512 = $\frac{2,168}{2,168}$ (100) = 100% ของการส่งสินค้าออกปี 2512

การส่งสินค้าออกปี 2513 = $\frac{1,985}{2,168}$ (100) = 91.55% ของการส่งสินค้าออกปี 2512

การส่งสินค้าออกปี 2514 = $\frac{2,264}{2,168}$ (100) 104.42% ของการส่งสินค้าออกปี 2512

การส่งสินค้าออกปี 2515 = $\frac{2,841}{2,168}$ (100) = 131.04% ของการส่งสินค้าออกปี 2512

การส่งสินค้าออกปี 2516 = $\frac{3,261}{2,168}(100)$ = 150.41% ของการส่งสินค้าออกปี 2512

การส่งสินค้าออกปี 2517 = $\frac{3,983}{2,168}(100)$ = 183.71% ของการส่งสินค้าออกปี 2512

นำมาเสนอโดยสรุปเป็นตารางได้ดังนี้

ตาราง 1.5 แสดงการเปรียบเทียบเปอร์เซนต์เมื่อฐานคงที่

ปี พ.ศ.	มูลค่าที่ส่งออก (ล้านบาท)	เปอร์เซนต์ ของปี 2512	เปอร์เซนต์ ที่เปลี่ยนแปลง
2512	2,168	100	-
2513	1,985	91.55	-8.45
2514	2,264	104.42	4.42
2515	2,841	131.04	31.04
2516	3,261	150.41	50.41
2517	3,983	183.71	83.71

หมายเหตุ เปอร์เซ็นต์เปลี่ยนแปลงมีเครื่องหมายเป็นบวก หมายถึงมูลค่าการส่งสินค้าออกเพิ่มขึ้น ถ้าเครื่องหมายเป็นลบหมายถึงมูลค่าการส่งสินค้าออกลดลง

ตัวอย่างเกี่ยวกับการเพิ่มและลดเปอร์เซ็นต์เมื่อฐานเปลี่ยนแปลง พิจารณาจากตัวอย่างที่แล้ว ในการส่งสินค้าออกจากประเทศไทยไปสหรัฐอเมริกา

เปรียบเทียบการส่งสินค้าออกในปีต่าง ๆ กับปีที่แล้วมา ซึ่งจะเห็นได้ว่าฐานที่นำมาพิจารณาเปรียบเทียบจะเปลี่ยนแปลงดังนี้

การส่งสินค้าออกในปี 2513 = $\frac{1,985}{2,168}(100)$ = 91.55% ของการส่งสินค้าออกในปี 2512

การส่งสินค้าออกในปี 2514 = $\frac{2,264}{1,985}(100)$ = 114.05% ของการส่งสินค้าออกในปี 2513

การส่งสินค้าออกในปี 2515 = $\frac{2,841}{2,264}(100)$ = 125.48% ของการส่งสินค้าออกในปี 2514

การส่งสินค้าออกในปี 2516 = $\frac{3,261}{2,841}(100)$ = 114.78% ของการส่งสินค้าออกในปี 2515

การส่งสินค้าออกในปี 2517 = $\frac{3,983}{3,261}$ (100) = 122.14% ของการส่งสินค้าออกในปี 2516

นำมาเสนอโดยสรุปเป็นตารางได้ดังนี้

ตารางที่ 1.16 แสดงการเปรียบเทียบเปอร์เซ็นต์เมื่อฐานเปลี่ยนแปลง

ปี พ.ศ.	มูลค่าที่ส่งออก (ล้านบาท)	เปอร์เซ็นต์เทียบกับปีที่แล้ว	เปอร์เซ็นต์เปลี่ยนแปลงจากปีที่แล้ว
2512	2,168	-	-
2513	1,985	91.55	-8.45
2514	2,264	114.05	14.05
2515	2,841	125.48	25.48
2516	3,261	114.78	14.78
2517	3,983	122.14	22.14

หมายเหตุ เปอร์เซ็นต์ที่เปลี่ยนแปลง มีเครื่องหมายบวก หมายถึงปริมาณการส่งสินค้าออกมีมูลค่าเพิ่มขึ้นจากปีที่แล้ว ถ้าเครื่องหมายเป็นลบ หมายถึงมูลค่าการส่งสินค้าออกลดลงจากปีที่แล้ว

แบบฝึกหัดที่ 1

1) เลขชุดหนึ่งประกอบด้วยเลข 5, 3, 6, 5, 4, 5, 2, 8, 6, 5, 4, 8, 3, 4, 5, 4, 8, 2, 5, 4 จงหามัชฌิมเลขคณิต มัธยฐาน และ ฐานนิยม ของเลขชุดนี้
2) จงหา Mean, Mediam และ Mode อายุของเด็กต่อไปนี้ 20, 17, 16, 15, 19, 19, 18, 12, 13, 13, 17, 16, 14, 14, 16, 15 ปี
3) จงหาค่า Mean, Mediam และ Mode ของเลขต่อไปนี้ 2, 4, 8, 16, 32, 64, 128, 256, 512 และ 1,204
4) เลขชุดหนึ่งประกอบด้วยเลข 6.2, 5.1, 8,7, 6.2, 4.1, 3.3, 5.4, 6.2, 6.7 และ 9.6

   ก. จงหา Mean, Mediam และ Mode

   ข. คูณทุกตัวด้วย 10 แล้วหา Mean, Mediam และ Mode

   ค. บวกทุกตัวด้วย 10 แล้วหา Mean, Mediam และ Mode
5) ในการวัด I.Q. ของนักเรียนจำนวน 25 คน วัดได้ดังนี้

   100, 83, 88, 81, 83, 96, 105, 108, 78, 102, 97, 113, 126, 94, 85, 119, 67, 91, 88, 99, 88, 72, 77, 88, 114

   ก. จงหา I.Q. เฉลี่ยของนักเรียน

   ข. นักเรียนส่วนมาก มี I.Q. เท่าไร

   ค. นักเรียนที่มี I.Q. ต่ำกว่า I.Q. เฉลี่ย มีกี่เปอร์เซ็นต์ของนักเรียนทั้งหมด
6) ค่าเฉลี่ยของเลข 35 จำนวน เท่ากับ 55 จงหาผลบวกของเลข 35 จำนวนนั้น
7) ผลการสอบของนักศึกษาคนหนึ่งสอบวิชา การใช้ห้องสมุด, กฎหมาย, รัฐศาสตร์ และวิชาวิทยาศาสตร์ ได้คะแนน 70, 65, 80 และ 75 ตามลำดับ โดยที่จำนวนเครดิตของวิชาการใช้ห้องสมุด 1 เครดิต, กฎหมาย 2 เครดิต, รัฐศาสตร์ 3 เครดิต และวิทยาศาสตร์ 4 เครดิต ให้หาคะแนนเฉลี่ยของนักศึกษาผู้นี้
8) โรงงานแห่งหนึ่งมีคนงานได้รับค่าแรงงานวันละ 30 บาท อยู่ 60 คน และได้รับค่าแรงวันละ 40 บาท อยู่ 20 คน

   ก. จงหารายได้เฉลี่ยของคนงานต่อวัน

   ข. ถ้าคนงาน 60 คน มีรายได้เฉลี่ยวันละ 30 บาท ต่อคน และคนงานอีก 20 คน มีรายได้เฉลี่ยวันละ 40 บาท ต่อคน จงหาค่าแรงเฉลี่ยของคนงานแต่ละคนจะเป็นเท่าไร
9) จากโจทย์ในข้อ 8. ให้หารายได้เฉลี่ย มัธยฐาน และ ฐานนิยมของคนงาน

10) คนงานของบริษัทสุรามหาชน มีรายได้เฉลี่ยคนละ 3,000 บาท ต่อเดือน ต่อมาเนื่องจากการเปลี่ยนแปลงสภาวะทางเศรษฐกิจ ทำให้ค่าครองชีพสูงขึ้น ทางบริษัทได้พิจารณาจ่ายค่าครองชีพพนักงานอีกคนละ 500 บาท ต่อเดือน อยากทราบว่ารายได้เฉลี่ยต่อเดือนของคนงานใหม่จะเป็นเท่าไร

11) แม่ค้าได้ซื้อส้มมาขาย มีส้มอยู่ 3 ชนิด ชนิดราคา 8 บาท มี 30 กก. ชนิดราคา 10 บาท มี 20 กก. และชนิดราคา 12 บาท มี 10 กก. แม่ค้านำส้ม 3 ชนิดมาปนกัน แล้วนำไปขาย อยากทราบว่า ควรจะขายในราคา กก.ละเท่าไร จึงจะไม่ขาดทุน

12) สมมุติว่าคุณเข้าไปในร้านค้าเพื่อซื้อเทอร์โมมิเตอร์ ซึ่งมีอยู่เป็นจำนวนมาก แต่ปรากฏว่าเทอร์โมมิเตอร์แต่ละอันชี้บอกอุณหภูมิไม่ตรงกัน อยากทราบว่า คุณจะเลือกอย่างไรจึงจะได้เทอร์โมมิเตอร์ที่ดีที่สุด

13) ความสูงของเด็กนักเรียน 10 คน วัดได้ดังนี้ 162, 160, 150, 175, 155, 164, 168, 159, 85 และ 175 เซนติเมตร

ก. จงหา Mean, Mediam และ Mode ในความสูงของนักเรียน

ข. จงเปรียบเทียบค่าตัวกลางที่วัดจากข้อ ก. มีข้อดี ข้อเสียอย่างไร

14) จงสร้างตารางแจกแจงความถี่ของค่าแรงคนงานของโรงงานอุตสาหกรรมแห่งหนึ่ง จำนวน 100 คน ได้รับค่าแรงเป็นสัปดาห์ โดยแต่ละคนได้รับค่าแรงแตกต่างกันดังนี้

490, 470, 510, 480, 500, 460, 530, 460, 450, 500, 500, 470, 560, 510, 460, 470, 540, 530, 500, 490, 500, 510, 500, 600, 510, 460, 480, 520, 520, 460, 610, 520, 490, 500, 450, 470, 540, 510, 600, 560, 500, 520, 440, 490, 450, 500, 400, 460, 540, 470, 500, 550, 550, 470, 480, 530, 500, 490, 450, 500, 500, 510, 470, 540, 430, 530, 550, 500, 530, 520, 520, 510, 470, 510, 480, 450, 440, 500, 520, 490, 510, 510, 470, 530, 490, 460, 610, 490, 520, 480, 390, 460, 520, 510, 570, 490, 450, 500

15) จากตารางแจกแจงความถี่ในข้อ 14 จงคำนวณหา Mean, Mediam และ Mode

16) จากข้อมูลในข้อ 14 ให้หา Mediam และ Mode จากข้อมูลดิบ แล้วเปรียบเทียบค่ากับข้อ 15

17) นักวิจัยได้ศึกษาถึงความสัมพันธ์ระหว่างอายุกับผลงานที่ได้รับการยอมรับ และพิมพ์ออกเผยแพร่ ของนักประพันธ์ จำนวน 241 คน ซึ่งแจกแจงเป็นตารางได้ดังนี้

ช่วงอายุ	จำนวนประชากร
20 - 30	20
30 - 40	73
40 - 50	80
50 - 60	44
60 - 70	22
70 - 80	2

จากตารางแจกแจงความถี่อายุนักประพันธ์ จงคำนวณหา Mean, Mediam และ Mode

18) ในการวัดน้ำหนักของนักศึกษาหญิงปีหนึ่ง ในมหาวิทยาลัยรามคำแหง จำนวน 100 คน วัดน้ำหนักแล้วนำมาแจกแจงเป็นตารางได้ดังนี้

ช่วงน้ำหนัก (กก.)	ความถี่
40 - 42	5
43 - 45	18
46 - 48	42
49 - 51	27
52 - 54	8
รวม	100

ก. จงคำนวณหา Mean, Mediam และ Mode

ข. ถ้านักศึกษาที่มีน้ำหนักเกิน Mean จัดว่าเป็นคนค่อนข้างอ้วน อยากทราบว่าจะมีคนค่อนข้างอ้วนกี่เปอร์เซนต์

19) ในการสอบวิชาสถิติ 103 ซึ่งมีคะแนนเต็ม 100 คะแนน มีนักศึกษาลงทะเบียนจำนวน 4,372 คน คะแนนสอบของนักศึกษานำมาแจกแจงเป็นตารางได้ดังนี้

ช่วงคะแนน	จำนวนคน
ต่ำกว่า 25	222
25 - 29	405
30 - 34	508
35 - 44	1,045
45 - 54	947
55 - 64	663
65 - 75	416
มากกว่า 75	166
รวม	4,372

ก. คะแนนเฉลี่ยโดยประมาณของนักศึกษาที่สอบวิชาสถิติ 103 เป็นเท่าไร หาโดยวิธีไหน

ข. ถ้าพิจารณาให้นักศึกษาที่สอบได้คะแนนเฉลี่ยตั้งแต่ 59.5 คะแนนขึ้นไปสอบได้จะมีนักศึกษาสอบได้จำนวนเท่าไร และสอบตกจำนวนเท่าไร

ค. ถ้ากำหนดให้ผู้ที่สอบได้คะแนนมากกว่า 75 คะแนน ได้เกรด G จะมีนักศึกษาได้เกรด G กี่เปอร์เซนต์

20) จงเปรียบเทียบข้อดี ข้อเสีย ของ Mean, Mediam และ Mode

21) เลขชุดหนึ่งประกอบด้วยเลข 2, 3, 6, 8 และ 10 จงหา

ก. พิสัย ข. ส่วนเบี่ยงเบนเฉลี่ย ค. ส่วนเบี่ยงเบนมาตรฐาน

22) จากชุดของตัวเลขในข้อ 3 จงหา

ก. พิสัย ข. ส่วนเบี่ยงเบนเฉลี่ย ค. ส่วนเบี่ยงเบนมาตรฐาน

23) อายุของเด็กในข้อ 2 จงหา

ก. พิสัย ข. ส่วนเบี่ยงเบนเฉลี่ย ค. ส่วนเบี่ยงเบนมาตรฐาน

24) ผลการสอบวิชาคณิตศาสตร์ของนักเรียนสิบคน มีส่วนเบี่ยงเบนมาตรฐานเท่ากับศูนย์ หมายความว่าอย่างไร

25) ผลการสอบของนักศึกษาจำนวน 103 คน นำมาแจกแจงเป็นตารางความถี่ได้ดังนี้

ช่วงคะแนน	ความถี่ (คน)
96 - 100	23
91 - 95	19
86 - 90	11
81 - 85	6
76 - 80	15
71 - 75	22
66 - 70	7
รวม	103 คน

จงหาส่วนเบี่ยงเบนเฉลี่ย, ส่วนเบี่ยงเบนควอไทล์ และ ส่วนเบี่ยงเบนมาตรฐาน

26) ในการหาความสัมพันธ์ระหว่างอายุ กับผลงานของนักประพันธ์ซึ่งได้แจกแจงเป็นตาราง ในข้อ 17 จงหาส่วนเบี่ยงเบนควอไทล์ และส่วนเบี่ยงเบนมาตรฐาน

27) จากตารางการแจกแจงน้ำหนักของนักศึกษาหญิงในตัวอย่างข้อ 18 จงหาส่วนเบี่ยงเบนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

28) ในการสอบเก็บคะแนนวิชาคณิตศาสตร์สองครั้ง มีนักศึกษาสอบได้ดังนี้

ช่วงคะแนน	ทดสอบครั้งที่1 ความถี่ (คน)	ทดสอบครั้งที่2 ความถี่ (คน)
30 - 39	1	-
40 - 49	4	2
50 - 59	10	11
60 - 69	22	38
70 - 79	45	52
80 - 89	30	17
90 -100	8	0

จงหาค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐานของคะแนนในการสอบแต่ละครั้ง

29) จากผลการคำนวณในข้อ 28 จงเปรียบเทียบผลการสอบทั้งสองครั้งของนักศึกษา

30) พิสัย (Range) คืออะไร หาได้อย่างไร มีประโยชน์อย่างไร

31) เลขชุดหนึ่งประกอบด้วยเลข 3, 5, 6, 6, 7, 10 และ 12 จงหา P_{50}, Q_2, D_5

32) จากผลการวัด I.Q. ของนักเรียนในข้อ 5 จงหา P_{75} และ Q_3

33) จงหา P_{10} และ D_1 ของอายุเด็กในข้อ 2

34) ค่าใช้จ่ายต่อสัปดาห์ของนักศึกษาในมหาวิทยาลัยแห่งหนึ่ง สุ่มตัวอย่างขึ้นมา 25 คน นำมาแจกแจงได้ตามตารางข้างล่างนี้ ให้หา P_{25} , D_2 และ Q_3

ช่วงค่าใช้จ่าย (บาท)	ความถี่
0 - 99.99	1
100.00 - 199.99	5
200.00 - 299.99	10
300.00 - 399.99	6
400.00 - 499.99	3
รวม	25

35) จากคะแนนสอบวิชาสถิติของนักศึกษาในข้อ 19 จงหาว่า D_6 ตรงกับคนสอบลำดับที่เท่าไร ได้คะแนนเท่าไร และ Q_3 ตรงกับคนสอบลำดับที่เท่าไร ได้คะแนนสอบเท่าไร

36) ในการสอบวิชาสถิติของนักศึกษาซึ่งมีข้อมูลในข้อ 19 ถ้ากำหนดให้นักศึกษาที่สอบได้คะแนนสูงสุดเรียงลำดับลงมาถึงลำดับที่ 1,000 สอบได้ นอกนั้นสอบตก อยากทราบว่าคนที่สอบได้คนสุดท้ายได้คะแนนเท่าไร พิจารณาโดยประมาณ

37) โรงพยาบาลแห่งหนึ่งได้บันทึกข้อมูลเกี่ยวกับคนไข้โรคมะเร็งที่มารับการรักษาในรอบ 5 ปี แต่ละปีมีคนไข้ดังนี้

ปี พ.ศ.	จำนวนคนไข้
2516	52
2517	56
2518	65
2519	79
2520	94

ก. จงหาอัตราส่วนของคนไข้ในปีต่าง ๆ เมื่อเทียบกับปี 2516
ข. จงหาอัตราส่วนของคนไข้ในปีต่าง ๆ เมื่อเทียบกับปีที่แล้วมา

38) จำนวนคนไข้โรคมะเร็งในข้อ 37 อัตราส่วนที่ได้คิดเป็นเปอร์เซนต์ได้เท่าไร

39) จงหาอัตราเงินเดือนขั้นสุดท้ายของนายขาว นายเขียว นายดำ และนายแดง ซึ่งทางบริษัทประสบภาวะขาดทุน เพื่อให้บริษัทอยู่รอดได้ ทางบริษัทจึงขอลดเงินเดือนคนงานทุกคน 20% ของเงินเดือน ต่อมาทางบริษัทมีฐานะการเงินดีขึ้น จึงขึ้นเงินเดือนให้พนักงานทุกคนคนละ 20% ของเงินเดือนที่ได้รับ

พนักงาน	อัตราเงินเดือน
นายขาว	5,000
นายเขียว	4,500
นายดำ	3,000
นายแดง	2,500

40) จากข้อ 39 ถ้าจะให้เงินเดือนของนายขาว นายเขียว นายดำ และนายแดง เท่าเดิม ทางบริษัทจะต้องเพิ่มเงินเดือนให้เขาแต่ละคน คนละกี่เปอร์เซนต์

หนังสืออ้างอิง

1. บุญเสริม วีสกุล, **วิธีเก็บและประมวลข้อมูล**, สำนักพิมพ์ไทยวัฒนาพาณิชย์, กรุงเทพฯ, พ.ศ. 2517
2. ยุวดี นนทรีย์, **สถิติเบื้องต้น**, โรงพิมพ์มหาวิทยาลัยรามคำแหง, กรุงเทพฯ, พ.ศ. 2520
3. วีนัส พีชวณิชย์ และสมจิต วัฒนาชยากุล, **สถิติสำหรับนักสังคมศาสตร์**, โรงพิมพ์มหาวิทยาลัยธรรมศาสตร์, กรุงเทพฯ, พ.ศ. 2519
4. อนันต์ ศรีโสภา, **สถิติเบื้องต้น**, สำนักพิมพ์ไทยวัฒนาพานิช, กรุงเทพฯ, พ.ศ. 2521
5. คณะกรรมการพัฒนาเศรษฐกิจ และสังคมแห่งชาติ, **สถิติด้านประชากร กำลังคน และการมีงานทำ**, กรุงเทพฯ, พ.ศ. 2520
6. Weinberg and Schumaker, **Statistics an intuitive approach,** Brooks/Cole Publishing Company, Carifornia, 1969
7. G. Hole and J. Jessen, **Basic Statistics,** John Willey and Sons, Inc., 1971