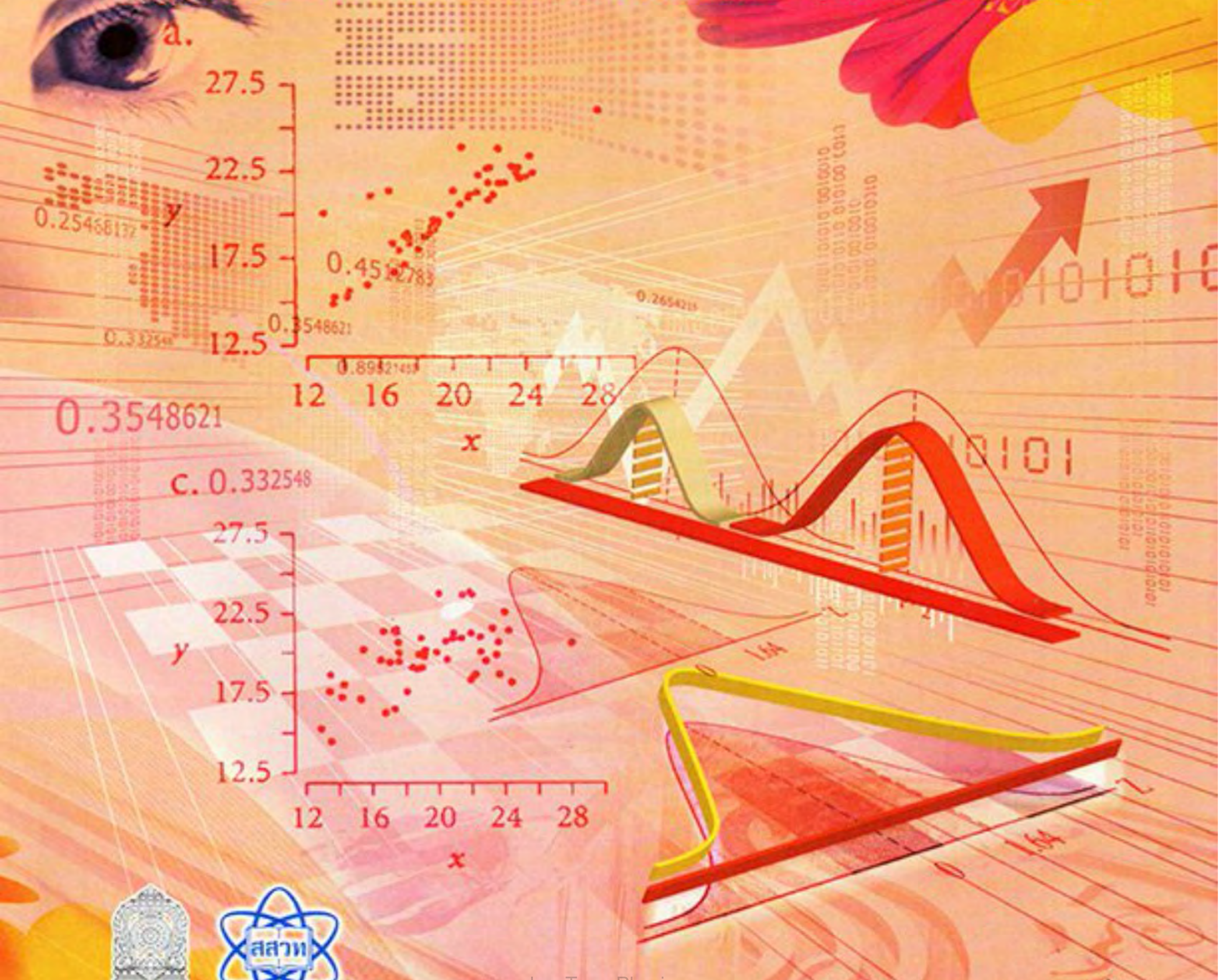
หนังสือเรียนรายวิชาเพิ่มเติม

คณิตศาสตร์ เล่ม ๕

ชั้นมัธยมศึกษาปีที่ ๔-๖

กลุ่มสาระการเรียนรู้คณิตศาสตร์

ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑



กระทรวงศึกษาธิการ

สสวท

๓๖.-

หนังสือเรียนรายวิชาเพิ่มเติม
คณิตศาสตร์ เล่ม ๕
ชั้นมัธยมศึกษาปีที่ ๔–๖

กลุ่มสาระการเรียนรู้คณิตศาสตร์

ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑

จัดทำโดย

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
กระทรวงศึกษาธิการ

ISBN 978-974-01-9626-6

พิมพ์ครั้งที่หนึ่ง ๓๐๐,๐๐๐ เล่ม

พ.ศ. ๒๕๕๔

องค์การค้าของ สกสค. จัดพิมพ์จำหน่าย

พิมพ์ที่โรงพิมพ์ สกสค. ลาดพร้าว

๒๒๔๙ ถนนลาดพร้าว วังทองหลาง กรุงเทพมหานคร

ประกาศสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

เรื่อง อนุญาตให้ใช้สื่อการเรียนรู้ในสถานศึกษา

ด้วยสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ได้มอบหมายให้สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยีจัดทำโครงสร้างหลักสูตรรายวิชาเพิ่มเติม และจัดทำหนังสือเรียนรายวิชาเพิ่มเติม คณิตศาสตร์ เล่ม ๕ กลุ่มสาระการเรียนรู้คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ ๔-๖ สำหรับให้สถานศึกษาพิจารณาเลือกใช้เทียบเคียงกับหลักสูตรของสถานศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานได้พิจารณาแล้วอนุญาตให้ใช้ในสถานศึกษาได้

ประกาศ ณ วันที่ ๑๕ ธันวาคม พ.ศ. ๒๕๕๓

(นายชินภัทร ภูมิรัตน)

เลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน

คำนำ

หนังสือเรียน รายวิชาเพิ่มเติม คณิตศาสตร์ เล่ม ๕ ชั้นมัธยมศึกษาปีที่ ๔-๖ นี้ กลุ่มสาระการเรียนรู้คณิตศาสตร์ จัดทำขึ้นโดยสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี ซึ่งได้จัดทำโครงสร้างหลักสูตรรายวิชาเพิ่มเติม ที่ประกอบด้วย โครงสร้างรายวิชาเพิ่มเติมและคำอธิบายรายวิชาที่มีทั้งผลการเรียนรู้และสาระการเรียนรู้เพิ่มเติม เพื่อให้สถานศึกษาได้เทียบเคียงกับหลักสูตรของสถานศึกษา และพิจารณาเลือกใช้หนังสือนี้ประกอบการจัดการเรียนรู้ให้สอดคล้องกับหลักสูตรสถานศึกษาของตนได้ตามความเหมาะสม ซึ่งสถานศึกษาสามารถใช้เป็นแนวทางในการออกแบบกิจกรรมการเรียนรู้ ให้ความรู้ความเข้าใจผู้เรียนนำไปสู่ทักษะการคิด วิเคราะห์ สังเคราะห์ ตามความสามารถและความแตกต่างระหว่างบุคคลของผู้เรียนได้ ในการจัดทำหนังสือเรียนเล่มนี้ ได้รับความร่วมมือจากคณาจารย์ ผู้ทรงคุณวุฒิ ผู้เชี่ยวชาญด้านคณิตศาสตร์จากสถาบันต่างๆ ทั้งภาครัฐและเอกชนเป็นอย่างดี

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน หวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้จะเป็นประโยชน์ต่อการจัดการเรียนรู้ เพื่อประยุกต์ใช้พัฒนาการเรียนรู้ของผู้เรียนได้อย่างเหมาะสม ขอขอบคุณสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี ตลอดจนบุคคลและหน่วยงานที่มีส่วนเกี่ยวข้องในการจัดทำหนังสือไว้ ณ โอกาสนี้

(นายชินภัทร ภูมิรัตน)
เลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน
๑๕ ธันวาคม ๒๕๕๓

คำชี้แจง

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ได้รับมอบหมายจากกระทรวงศึกษาธิการ ให้พัฒนาหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระการเรียนรู้คณิตศาสตร์ กลุ่มสาระการเรียนรู้วิทยาศาสตร์ รวมทั้งสาระการออกแบบและเทคโนโลยีและสาระเทคโนโลยีสารสนเทศในกลุ่มสาระการเรียนรู้การงานอาชีพและเทคโนโลยี ตลอดจนจัดทำสื่อการเรียนรู้ตามหลักสูตรดังกล่าว

หนังสือเรียนรายวิชาเพิ่มเติม คณิตศาสตร์ สำหรับระดับมัธยมศึกษาตอนปลาย มีด้วยกันทั้งหมด 6 เล่ม จัดทำขึ้นเพื่อให้ผู้เรียนสามารถเรียนรู้และพัฒนาตนเอง นำความรู้ทางคณิตศาสตร์ไปพัฒนาชีวิต และเป็นเครื่องมือในการเรียนรู้คณิตศาสตร์ตลอดจนศาสตร์อื่น ๆ ในระดับที่สูงขึ้น ทั้งนี้สถานศึกษาสามารถปรับใช้เนื้อหาจากหนังสือเรียนทั้ง 6 เล่มนี้ เพื่อจัดการเรียนการสอนรายวิชาเพิ่มเติม คณิตศาสตร์ ระดับมัธยมศึกษาตอนปลาย ได้ตามความเหมาะสม

หนังสือเรียนรายวิชาเพิ่มเติม คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 4 – 6 เล่ม 5 ประกอบด้วยการวิเคราะห์ข้อมูลเบื้องต้น การแจกแจงปกติ และความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล ซึ่งเป็นเนื้อหาสาระตามมาตรฐานการเรียนรู้ตามที่กำหนดไว้ในหลักสูตร และคำนึงถึงความต่อเนื่องกับมาตรฐานการเรียนรู้ของระดับมัธยมศึกษาตอนต้นของ สสวท. เป็นหลัก อย่างไรก็ตามผู้สอนสามารถปรับบทเรียนให้เหมาะสมกับศักยภาพของผู้เรียนแต่ละกลุ่ม

การจัดทำหนังสือเรียนคณิตศาสตร์เล่มนี้ สสวท. ได้รับความร่วมมืออย่างดียิ่งจากคณาจารย์ ผู้ทรงคุณวุฒิ นักวิชาการ และครูผู้สอน จากหลายหน่วยงาน ทั้งภาครัฐและเอกชน สสวท. จึงขอขอบคุณทุกท่านไว้ ณ ที่นี้ และหวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้จะเป็นประโยชน์ต่อการศึกษาคณิตศาสตร์ อันเป็นรากฐานสำคัญของการพัฒนาทรัพยากรมนุษย์ของชาติต่อไป หากมีข้อเสนอแนะใดที่จะทำให้หนังสือเรียนเล่มนี้สมบูรณ์ยิ่งขึ้น โปรดแจ้งให้สาขาคณิตศาสตร์มัธยมศึกษา สสวท. ทราบด้วย จักขอบคุณยิ่ง

(นางพรพรรณ ไวทยางกูร)
ผู้อำนวยการ
สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
กระทรวงศึกษาธิการ

คำอธิบายรายวิชาเพิ่มเติม

คณิตศาสตร์เพิ่มเติม ๕ **กลุ่มสาระการเรียนรู้คณิตศาสตร์**

ชั้นมัธยมศึกษาปีที่ ๔ – ๖ **เวลา ๘๐ ชั่วโมง** **จำนวน ๒.๐ หน่วยกิต**

ศึกษา และฝึกทักษะและกระบวนการทางคณิตศาสตร์อันได้แก่ การแก้ปัญหา การให้เหตุผล การสื่อสาร การสื่อความหมายทางคณิตศาสตร์ และการนำเสนอ การเชื่อมโยงความรู้ต่างๆ ทางคณิตศาสตร์ และเชื่อมโยงคณิตศาสตร์กับศาสตร์อื่นๆ และมีความคิดริเริ่มสร้างสรรค์ ในสาระต่อไปนี้

การวิเคราะห์ข้อมูลเบื้องต้น การวัดค่ากลางของข้อมูล ได้แก่ ค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม ค่าเฉลี่ยเรขาคณิต และค่าเฉลี่ยฮาร์มอนิก การวัดตำแหน่งที่หรือตำแหน่งสัมพัทธ์ของข้อมูล การวัดการกระจายของข้อมูล ได้แก่ การวัดการกระจายสัมบูรณ์ การวัดการกระจายสัมพัทธ์ ความสัมพันธ์ระหว่างการแจกแจงความถี่ ค่ากลาง และการกระจายของข้อมูล

การแจกแจงปกติ ค่ามาตรฐาน การแจกแจงปกติและเส้นโค้งปกติ

ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล การวิเคราะห์ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล แผนภาพการกระจาย การประมาณค่าของค่าคงตัวโดยใช้วิธีกำลังสองน้อยที่สุด ความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่อยู่ในรูปอนุกรมเวลา

ผลการเรียนรู้

๑. เลือกวิธีวิเคราะห์ข้อมูลเบื้องต้นและอธิบายผลการวิเคราะห์ข้อมูลได้ถูกต้อง
๒. นำความรู้เรื่องการวิเคราะห์ข้อมูลไปใช้ได้
๓. นำความรู้เรื่องค่ามาตรฐานไปใช้ในการเปรียบเทียบข้อมูล
๔. หาพื้นที่ใต้โค้งปกติและนำความรู้เกี่ยวกับพื้นที่ใต้เส้นโค้งปกติไปใช้ได้
๕. เข้าใจความหมายของการสร้างความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่ประกอบด้วยสองตัวแปร
๖. สร้างความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่ประกอบด้วยสองตัวแปรที่อยู่ในรูปอนุกรมเวลาโดยใช้เครื่องคำนวณ
๗. ใช้ความสัมพันธ์เชิงฟังก์ชันของข้อมูลพยากรณ์ค่าตัวแปรตามเมื่อกำหนดตัวแปรอิสระให้

รวมทั้งหมด ๗ ผลการเรียนรู้

สารบัญ

บทที่ 1
การวิเคราะห์ข้อมูลเบื้องต้น

ก่อนทำการวิเคราะห์ข้อมูลชุดใด นักเรียนควรทราบที่มาของข้อมูลหรือวิธีการเก็บรวบรวมข้อมูล ประเภทของข้อมูลและข้อกำหนดที่สำคัญของข้อมูล เพื่อช่วยให้สามารถเลือกวิธีวิเคราะห์ข้อมูลและนำสารสนเทศ (information) ที่ได้ไปตัดสินใจวางแผนได้ตรงกับจุดประสงค์อย่างถูกต้อง

สารสนเทศ

ที่มาของข้อมูลมาจากวิธีเก็บรวบรวมข้อมูล 3 แหล่งใหญ่ ๆ คือ

1) จากข้อมูลที่มีอยู่แล้วในทะเบียนหรือแหล่งที่ทำข้อมูลไว้แล้ว เช่น จากสำนักงานสถิติแห่งชาติ

2) จากการสำรวจประชากรหรือตัวอย่าง เช่น จากนักเรียนทุกคนในโรงเรียน

3) จากการทดลองหรือสังเกตผลจากการทดลองเฉพาะทาง เช่น การทดลองในห้องปฏิบัติการ

เมื่อทราบแหล่งที่มาของข้อมูลแล้ว ให้ตรวจสอบว่า ตัวแปรหรือข้อมูลของแต่ละแหล่งที่มาเป็นข้อมูลประเภทใด ซึ่งประเภทของข้อมูลมี 2 ประเภท ได้แก่

1) ข้อมูลเชิงปริมาณ (quantitative data) เช่น คะแนน น้ำหนัก ส่วนสูง รายได้ ปริมาณน้ำในเขื่อน เป็นต้น

2) ข้อมูลเชิงคุณภาพ (qualitative data) เช่น เพศ ศาสนา ระดับความคิดเห็น เป็นต้น

วัตถุประสงค์ในการวิเคราะห์ข้อมูลจะทำให้ทราบว่า จะต้องวิเคราะห์เพื่อทราบภาพโดยรวม หรือลักษณะกว้าง ๆ ของข้อมูลโดยใช้สถิติเชิงพรรณนา (descriptive statistics) หรือจะต้องศึกษาความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลโดยใช้สถิติเชิงอนุมาน (inferential statistics)

การตรวจสอบและจัดข้อมูลเพื่อการวิเคราะห์มีประโยชน์เบื้องต้น คือ สามารถตรวจสอบความสมบูรณ์ ความถูกต้อง และภาพรวมของข้อมูลทั้งหมดได้ โดยอาจใช้วิธีการแจกแจงความถี่และหาความถี่สัมพัทธ์ของข้อมูล ถ้าเป็นข้อมูลเชิงปริมาณควรใช้ฮิสโทแกรม แผนภาพต้น-ใบ (stem-and-leaf plot) หรือแผนภาพกล่อง (box plot) แต่ถ้าเป็นข้อมูลเชิงคุณภาพอาจใช้แผนภูมิแท่งหรือฮิสโทแกรมที่จำแนกตามกลุ่มต่าง ๆ เช่น เพศ ศาสนา ฯลฯ จากการตรวจสอบ

สถิติเชิงพรรณนาและสถิติเชิงอนุมาน

แผนภาพต้น - ใบ

แผนภาพกล่อง

แหล่งที่มาของข้อมูล

ข้อมูลเชิงปริมาณและข้อมูลเชิงคุณภาพ

ข้อมูลด้วยวิธีดังกล่าวนี้ นักเรียนจะทราบว่าข้อมูลทั้งชุดมีค่าผิดพลาดหรือไม่ เช่น มีการเขียนหรือพิมพ์ข้อมูลผิดหรือไม่ ซึ่งดูได้จากฮิสโทแกรมหรือการทำรอยขีด (tally) การตรวจดูว่าข้อมูลมีค่าสูงหรือต่ำผิดปกติหรือไม่ ถ้ามี มีกี่จำนวน ดูได้จากแผนภาพต้น-ใบ หรือแผนภาพกล่อง รายละเอียดเพิ่มเติมเกี่ยวกับวิธีทำฮิสโทแกรมและแผนภาพต่าง ๆ สามารถศึกษาได้จากหนังสือเรียนรายวิชาพื้นฐานคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 4–6 เล่ม 3 หรือจากหนังสือสถิติในหัวข้อสถิติพรรณนาโดยทั่วไป อย่างไรก็ตามขอเน้นว่า ฮิสโทแกรม แผนภาพต้น-ใบ และแผนภาพกล่อง นอกจากจะช่วยตรวจสอบดูข้อมูลเบื้องต้นได้แล้ว ยังสามารถให้สารสนเทศที่สำคัญ ๆ เกี่ยวกับเรื่องสัดส่วนของข้อมูลที่สามารถจำแนกสัดส่วนตามกลุ่มหรือช่วงของข้อมูลที่สนใจศึกษา และลักษณะของการแจกแจงของข้อมูลว่า มีรูปแบบเกาะกลุ่มหรือเบ้ทางใดด้วย โดยเป็นการวิเคราะห์ข้อมูลแบบพรรณนาซึ่งมีประโยชน์ในเบื้องต้น นอกจากนี้ควรทำการวิเคราะห์ข้อมูลแบบมีการคำนวณค่าสถิติต่าง ๆ เพิ่มเติม เพื่อหาสารสนเทศให้มากขึ้น

ตัวอย่างของการเลือกวิธีวิเคราะห์ข้อมูล โดยทั่วไปขึ้นอยู่กับประเภทของข้อมูลว่า เป็นข้อมูลเชิงปริมาณหรือคุณภาพ เช่น ในการวัดค่ากลาง ถ้าเป็นข้อมูลเชิงปริมาณและไม่มีข้อมูลค่าใดค่าหนึ่งหรือหลายค่าสูงหรือต่ำผิดปกติจะนิยมใช้ค่าเฉลี่ยเลขคณิต ถ้าเป็นข้อมูลเชิงคุณภาพใช้ฐานนิยม (ถ้าหาได้) หรือเลือกวิธีวิเคราะห์แบบอื่น ๆ ที่เป็นวิธีวิเคราะห์ข้อมูลขั้นสูงขึ้น ซึ่งต้องพิจารณาว่าเป็นข้อมูลระดับประชากรหรือตัวอย่างด้วย อย่างไรก็ตามสำหรับระดับเบื้องต้นนี้ ทั้งข้อมูลระดับประชากรและระดับตัวอย่างจะวิเคราะห์โดยใช้สถิติพรรณนาและเลือกสูตรให้ตรงกับระดับประชากรหรือระดับตัวอย่างให้ถูกต้อง

ข้อมูลที่สนใจศึกษาอาจจะเป็นข้อมูลเชิงปริมาณหรือเชิงคุณภาพ เป็นข้อมูลที่สำรวจในระดับประชากรหรือระดับตัวอย่าง หรือเป็นข้อมูลที่สังเกตจากการทดลอง สามารถตรวจสอบจากแหล่งที่มา ซึ่งแต่ละแหล่งอาจมีข้อมูลอยู่หลายประเภท โดยทั่วไปถ้าเป็นข้อมูลจากทะเบียน เช่น จากสำนักงานสถิติแห่งชาติมักเป็นข้อมูลระดับประเทศ รวบรวมไว้ในลักษณะที่แจกแจงความถี่แล้ว ซึ่งอาจเป็นข้อมูลเชิงปริมาณบ้าง เชิงคุณภาพบ้าง ระดับประชากรหรือระดับตัวอย่างบ้าง เช่น จำนวนผู้ถือครองที่ดินทำการเกษตรจำแนกตามรายได้จากผลผลิตทางการเกษตร พ.ศ. 2546 (แสดงไว้ในตารางที่ 1) การวิเคราะห์ข้อมูลเบื้องต้น เช่น สนใจค่ากลางและการวัดการกระจายต้องเลือกใช้สูตรค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานที่ข้อมูลมีการแจกแจงความถี่ และใช้สัญลักษณ์ในระดับประชากรด้วย เพราะว่าเป็นข้อมูลเชิงปริมาณที่ได้

แจกแจงความถี่ไว้แล้ว และสำรวจในระดับประชากรนั่นเอง ถ้าเป็นข้อมูลในระดับตัวอย่าง เช่น มูลค่าความเสียหาย (เฉพาะที่สัมภาษณ์ได้) เป็นรายจังหวัดจากธรณีพิบัติภัย พ.ศ. 2547 (แสดงไว้ในตารางที่ 2) การวิเคราะห์ข้อมูลเบื้องต้นจะเลือกสูตรค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานในระดับตัวอย่างที่ไม่ได้แจกแจงความถี่ เนื่องจากเป็นข้อมูลเชิงปริมาณที่ไม่ได้แจกแจงความถี่และได้สำรวจในระดับตัวอย่าง (เฉพาะตารางที่ 2 นี้เป็นตัวอย่างเจาะจง ไม่รวมถึงตัวอย่างที่เป็นตัวแทนทั้งประชากร) และถ้าเป็นข้อมูลจากการทดลอง เช่น ในการทดลองเพื่อตรวจสอบส่วนประกอบของเมทิลแอลกอฮอล์ (methyl alcohol) ของสารผสมชนิดหนึ่งจากห้องปฏิบัติการ 4 แห่ง เพื่อศึกษาว่า ร้อยละของเมทิลแอลกอฮอล์จากแต่ละแห่ง มีระดับสูงเพียงไร แตกต่างกันหรือไม่ โดยอาศัยข้อมูลระดับตัวอย่างจากห้องปฏิบัติการแห่งละ 3 หน่วย (แสดงไว้ในตารางที่ 3) การวิเคราะห์ข้อมูลเบื้องต้นนี้ ยังคงเลือกค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานในระดับตัวอย่างที่ใช้สูตรซึ่งไม่ได้แจกแจงความถี่ เพราะเป็นข้อมูลเชิงปริมาณในระดับตัวอย่างและเป็นข้อมูลดิบของแต่ละหน่วยโดยตรง สำหรับการวิเคราะห์ข้อมูลเบื้องต้นจะยกตัวอย่างข้อมูลเชิงปริมาณเป็นสำคัญเพราะใช้บ่อยและเป็นพื้นฐานของการวิเคราะห์ข้อมูลแบบอื่น ๆ ในระดับสูงต่อไป

ตารางที่ 1 จำนวนผู้ถือครองที่ดินทำการเกษตร (คน) จำแนกตามรายได้จากผลผลิตทางการเกษตรในปี พ.ศ. 2536 2541 และ 2546

รายได้จากผลผลิตทางการเกษตร	พ.ศ.		
	2536	2541	2546
ยังไม่มีรายได้จากผลผลิต	124,353	97,797	141,018
มีรายได้จากผลผลิต (บาท)	5,519,176	5,479,503	5,667,105
ต่ำกว่า 5,001	841,098	334,585	352,284
5,001 – 10,000	1,485,238	905,539	665,802
10,001 – 20,000	1,369,204	1,254,615	1,124,539
20,001 – 50,000	1,259,620	1,774,818	1,981,426
50,001 – 100,000	387,758	814,057	981,610
100,001 ขึ้นไป	176,258	395,889	561,444

ที่มา : โครงการสำมะโนการเกษตร พ.ศ. 2546 สารสถิติปีที่ 16 ฉบับที่ 3 เดือนมีนาคม พ.ศ. 2548

ตารางที่ 2 มูลค่าความเสียหายจากธรณีพิบัติภัย พ.ศ.2547 (เฉพาะที่สัมภาษณ์ได้) เป็นรายจังหวัด

จังหวัด	รวม	ความเสียหาย (ล้านบาท)				
		ที่ดิน	บ้าน/อาคาร สิ่งปลูกสร้าง	อุปกรณ์	ยานพาหนะ	อื่น ๆ
รวม	**1,942.8**	**262.5**	**772.8**	**372.8**	**304.0**	**230.7**
กระบี่	321.3	17.1	99.5	71.9	64.8	68.0
พังงา	1,077.4	122.6	557.3	154.0	161.0	82.5
ระนอง	203.3	102.1	26.2	36.8	30.6	7.6
ตรัง	43.0	10.1	5.9	14.4	9.9	2.7
ภูเก็ต	188.6	1.1	83.3	41.4	29.6	33.2
สตูล	109.2	9.5	0.6	54.3	8.1	36.7

ที่มา : สารสถิติปีที่ 16 ฉบับที่ 3 เดือนมีนาคม พ.ศ. 2548

ตารางที่ 3 ร้อยละของเมทิลแอลกอฮอล์จำแนกตามห้องปฏิบัติการ

หน่วยทดลอง	ห้องปฏิบัติการ			
	LAB1	LAB2	LAB3	LAB4
1	85.06	84.99	84.48	84.10
2	85.25	84.28	84.72	84.55
3	84.87	84.88	85.10	84.05

ที่มา : Devore and Farnum (2005)

การวิเคราะห์ข้อมูลเบื้องต้นเป็นการวิเคราะห์ข้อมูลเพื่อทราบลักษณะโดยรวมของข้อมูลเกี่ยวกับเรื่องนั้น ๆ ส่วนใหญ่นิยมใช้การแจกแจงความถี่ (frequency distribution) ของข้อมูล ค่ากลาง (central value) ของข้อมูล และการกระจาย (dispersion) ของข้อมูลซึ่งเป็นวิธีหรือเครื่องมือที่สำคัญในการวิเคราะห์ข้อมูลทุกระดับ ในบทเรียนนี้จะกล่าวถึง การวัดค่ากลางของข้อมูล การวัด

ตำแหน่งที่ของข้อมูล และการวัดการกระจายของข้อมูล สำหรับการหาตำแหน่งที่ของข้อมูลเกี่ยวกับเรื่องใดเรื่องหนึ่งมีความสำคัญมากในการทราบลักษณะโดยรวมของข้อมูลโดยตำแหน่งที่บางตำแหน่งสามารถให้ค่ากลางของข้อมูลชุดนั้นได้ การหาตำแหน่งที่ของข้อมูลใช้ทั้งกรณีที่ข้อมูลมีการแจกแจงในรูปสมมาตรและไม่สมมาตรคือ มีลักษณะเบ้ (skewed distribution) ซึ่งในกรณีที่ข้อมูลมีความเบ้ควรหาตำแหน่งที่ของข้อมูลเสมอ เพื่อใช้เป็นสารสนเทศของชุดข้อมูลร่วมกับค่าสถิติอื่น ๆ ด้วย

1.1 การวัดค่ากลางของข้อมูล (measures of central value)

การวิเคราะห์ข้อมูลเบื้องต้นนอกจากจะทำโดยการแจกแจงความถี่ดังที่ได้กล่าวมาแล้ว การหาค่ากลางมาเป็นตัวแทนของข้อมูลทั้งหมดจะทำให้สะดวกในการจดจำข้อสรุปเรื่องราวเกี่ยวกับข้อมูลนั้น ๆ ได้มากขึ้น เช่น ผู้อำนวยการต้องการทราบผลการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นต่าง ๆ ของปีที่ผ่านมาว่าเป็นอย่างไร แทนที่ผู้อำนวยการจะต้องทราบระดับคะแนนของวิชาคณิตศาสตร์ของนักเรียนแต่ละคนในแต่ละชั้น อาจทราบเพียงค่ากลางหรือค่าเฉลี่ยของระดับคะแนนของวิชาคณิตศาสตร์ของนักเรียนในชั้นนั้น ๆ ซึ่งเป็นการเพียงพอที่จะทราบผลการเรียนคณิตศาสตร์ของนักเรียนในแต่ละชั้นว่าเป็นอย่างไร ในการพิจารณาของรัฐบาลเกี่ยวกับผลผลิตทางการเกษตรของพืชบางชนิด รายได้ของประชากร และราคาสินค้าต่าง ๆ ก็เช่นเดียวกัน อาจพิจารณาในระดับเบื้องต้นจากค่ากลางของข้อมูลเหล่านี้ได้ นอกจากนี้ค่ากลางของข้อมูลบางชนิดยังนำไปใช้ช่วยในการวิเคราะห์เรื่องอื่น ๆ ของข้อมูล เช่น การวัดการกระจายหรือความแปรปรวนของข้อมูล ตลอดจนการนำไปใช้ในการวัดความสัมพันธ์เชิงฟังก์ชันของข้อมูลและการวิเคราะห์ข้อมูลชั้นสูงต่อไป เพื่อให้การสรุปสารสนเทศของข้อมูลมีความชัดเจนยิ่งขึ้น

ค่ากลางของข้อมูลมีหลายชนิด เช่น ค่าเฉลี่ยเลขคณิต (arithmetic mean) ค่าเฉลี่ยเรขาคณิต (geometric mean) ค่าเฉลี่ยฮาร์มอนิก (harmonic mean) ค่ากึ่งกลางพิสัย (mid–range) มัธยฐาน (median) และฐานนิยม (mode) ค่ากลางแต่ละชนิดต่างก็มีข้อดี ข้อเสีย และมีความเหมาะสมในการนำไปใช้ไม่เหมือนกัน ขึ้นอยู่กับลักษณะการแจกแจงของข้อมูลและวัตถุประสงค์ของผู้ใช้ข้อมูลนั้น ๆ แต่ค่ากลางของข้อมูลที่นิยมใช้กันมีอยู่ 3 ชนิด คือ ค่าเฉลี่ย

เลขคณิต มัธยฐาน และฐานนิยม การคำนวณค่ากลางทั้งสามชนิดนี้โดยทั่ว ๆ ไปแบ่งออกได้เป็น 2 กรณีใหญ่ ๆ คือ

การหาค่ากลางของข้อมูลที่ไม่ได้แจกแจงความถี่ (ungrouped data) ซึ่งค่าที่ได้เป็นค่ากลางที่ถูกต้องแน่นอนของข้อมูลชุดนั้น

การหาค่ากลางของข้อมูลที่แจกแจงความถี่ (grouped data) ซึ่งค่าที่ได้เป็นค่ากลางโดยประมาณของข้อมูลชุดนั้น

1.1.1 ค่าเฉลี่ยเลขคณิต (arithmetic mean)

ค่าเฉลี่ยเลขคณิตเหมาะที่จะนำมาใช้เป็นค่ากลางของข้อมูล เมื่อข้อมูลนั้น ๆ ไม่มีค่าใดค่าหนึ่งหรือหลาย ๆ ค่าซึ่งสูงหรือต่ำกว่าค่าอื่น ๆ มาก เช่น คะแนนสอบวิชาคณิตศาสตร์ของนักเรียน 10 คน เป็นดังนี้ 70, 72, 68, 3, 71, 74, 70, 67, 73 และ 5 ซึ่ง 3 และ 5 ถือว่าเป็นค่าที่ต่ำผิดปกติ การหาค่าเฉลี่ยเลขคณิตจากข้อมูลที่มีค่าสูงหรือต่ำผิดปกติ จะได้ค่ากลางที่สูงหรือต่ำผิดปกติ ค่าเฉลี่ยเลขคณิตในกรณีเช่นนี้จึงไม่เป็นค่ากลางที่ดีของข้อมูลชุดนั้น (อาจใช้ค่ากลางอื่นแทน เช่น มัธยฐาน)

การหาค่าเฉลี่ยเลขคณิตของข้อมูลที่ไม่ได้แจกแจงความถี่

การหาค่าเฉลี่ยเลขคณิตของข้อมูลที่ไม่ได้แจกแจงความถี่ หาได้โดยตรงจากข้อมูลที่มีอยู่ทั้งหมดโดยการหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูล กล่าวคือ ถ้าให้ $x_1, x_2, x_3, \ldots, x_N$ เป็นข้อมูล N จำนวนจากประชากร (population) หรือใช้ $x_1, x_2, x_3, \ldots, x_n$ เป็นข้อมูลเพียง n จำนวนจากตัวอย่าง (sample) ซึ่งเป็นตัวแทนของประชากร ดังนั้น

ค่าเฉลี่ยเลขคณิตของประชากร (population mean)

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$ หรือ

ค่าเฉลี่ยเลขคณิตของตัวอย่าง (sample mean)

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

μ (อ่านว่า "มิว") และ $\overline{X}$ (อ่านว่า "เอ็กซ์ บาร์") มีความหมายต่างกัน โดยสามารถสรุปได้ดังนี้

$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$ คือค่าเฉลี่ยเลขคณิตของประชากร ซึ่งเป็นพารามิเตอร์ (parameter) หรือค่าจริงแบบหนึ่งของประชากร เรียกว่า พารามิเตอร์ μ และ N แทนจำนวนหน่วยของประชากร (population units)

$\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ คือค่าเฉลี่ยเลขคณิตของตัวอย่าง ซึ่งเป็นตัวประมาณค่า (estimator) ของพารามิเตอร์ μ โดยที่ $\overline{X}$ คำนวณจากข้อมูลตัวอย่างที่เป็นตัวแทนของประชากร และ n แทนจำนวนหน่วยของตัวอย่าง (sample units)

หมายเหตุ ถ้าโจทย์ไม่ได้ระบุว่าเป็นข้อมูลของระดับตัวอย่างให้ใช้สูตรของระดับประชากร ส่วนสัญลักษณ์ $\sum_{i=1}^{N} x_i$ และ $\sum_{i=1}^{n} x_i$ ใช้แทนผลบวกของข้อมูล x_i ทุก ๆ ค่าจาก $i = 1$ ถึง $i = N$ หรือ n แล้วแต่กรณีของประชากรหรือตัวอย่างตามลำดับ และสัญลักษณ์ $\sum$ เป็นอักษรกรีก เรียกว่า "capital sigma" ในที่นี้จะอ่านว่า "ซัมเมชัน"

สมบัติของ $\sum$ ที่ควรทราบมีดังนี้

ถ้า c เป็นค่าคงตัวใด ๆ

1) $\sum_{i=1}^{N} c = Nc$

2) $\sum_{i=1}^{N} cx_i = c\sum_{i=1}^{N} x_i$

3) $\sum_{i=1}^{N} (x_i + y_i) = \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} y_i$

4) $\sum_{i=1}^{N} (x_i - y_i) = \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} y_i$

สมบัติเหล่านี้จะเป็นจริงทั้งข้อมูลระดับประชากร (N) และข้อมูลระดับตัวอย่าง (n) นอกจากนี้จะสังเกตได้ว่า การคำนวณค่าเฉลี่ยเลขคณิตข้างต้นใช้สูตรคล้ายกัน จึงต้องทำความเข้าใจและแยกสูตรให้ชัดเจน คือ ถ้าเป็นประชากร N หน่วยให้ใช้ μ แต่ถ้าเป็นตัวอย่าง n หน่วยให้ใช้ $\overline{X}$

ประชากรและกลุ่มตัวอย่าง

ตัวอย่างที่ 1 จากการตรวจสอบราคาข้าวโพดที่ใช้เลี้ยงสัตว์ในจังหวัดนครราชสีมาที่โรงงานรับซื้อในปี พ.ศ. 2547 โดยตรวจสอบเพียงบางโรงงาน เพื่อนำมาเป็นตัวอย่างจำนวน 10 โรงงาน ปรากฏว่า ราคาข้าวโพดที่ใช้เลี้ยงสัตว์ซึ่งโรงงานรับซื้อต่อกิโลกรัม (บาท) เป็นดังนี้

4.57 4.42 5.28 6.80 7.08 4.82 5.48 4.95 7.20 4.43

จงหาราคาเฉลี่ยต่อกิโลกรัมของข้าวโพดเลี้ยงสัตว์ที่โรงงานรับซื้อ

วิธีทำ ราคาข้าวโพดที่ใช้เลี้ยงสัตว์เฉลี่ยต่อกิโลกรัม คือ $\overline{X} = \dfrac{\sum_{i=1}^{10} x_i}{10}$

จะได้ $\overline{X} = \dfrac{4.57+4.42+5.28+6.80+7.08+4.82+5.48+4.95+7.20+4.43}{10}$

$= \dfrac{55.03}{10}$

$= 5.503$

≈ 5.50

ดังนั้น ราคาข้าวโพดที่ใช้เลี้ยงสัตว์เฉลี่ยต่อกิโลกรัมประมาณ 5.50 บาท

อาจกล่าวได้ว่า จากข้อมูลตัวอย่างที่สำรวจมาจากโรงงานที่รับซื้อข้าวโพดที่ใช้เลี้ยงสัตว์ในจังหวัดนครราชสีมา มีราคาค่ากลาง 5.50 บาทต่อกิโลกรัม และถ้าต้องการตรวจสอบเพิ่มเติมว่าราคาข้าวโพดที่ใช้เลี้ยงสัตว์ที่แต่ละโรงงานรับซื้อมีราคาใกล้เคียงกับค่ากลางหรือไม่สามารถทำได้ด้วยการวัดการกระจายของข้อมูลซึ่งจะได้ศึกษาต่อไป

การหาค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก (weighted arithmetic mean)

การหาค่าเฉลี่ยเลขคณิตถ่วงน้ำหนักนี้ใช้ในกรณีที่ข้อมูลแต่ละค่ามีความสำคัญไม่เท่ากัน เช่น การหาค่าเฉลี่ยเลขคณิตของคะแนนสอบ 4 วิชาที่แต่ละวิชาในแต่ละสัปดาห์ใช้เวลาเรียนไม่เท่ากัน หรือการหาค่าเฉลี่ยเลขคณิตของราคาสินค้าชนิดเดียวกัน แต่มีน้ำหนักหรือปริมาณการขายต่างกัน ถ้าจะใช้วิธีการหาค่าเฉลี่ยเลขคณิตธรรมดา นั่นคือไม่ถ่วงน้ำหนัก อาจทำให้ค่าเฉลี่ยที่ได้คลาดเคลื่อนไปจากที่ควรจะเป็นจริง ซึ่งอาจจะน้อยกว่าหรือมากกว่าที่ควรจะเป็นจริงก็ได้ ทั้งนี้ขึ้นอยู่กับน้ำหนักของข้อมูลแต่ละค่าที่นำมาใช้เป็นสำคัญ

ถ้าให้ $w_1, w_2, w_3, \dots, w_N$ เป็นความสำคัญหรือน้ำหนักของค่าจากการสังเกต $x_1, x_2, x_3, \dots, x_N$ ตามลำดับ แล้ว

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก
$$\mu = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_Nx_N}{w_1 + w_2 + w_3 + \cdots + w_N}$$
$$= \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$$

ตัวอย่างที่ 2 ในการทดสอบทักษะและกระบวนการทางคณิตศาสตร์ของนักเรียนคนหนึ่งซึ่งมีคะแนนการทดสอบและความสำคัญของคะแนนทั้งหมดรวม 5 ด้าน จากคะแนนเต็ม 100 คะแนน ดังข้อมูลในตาราง จงหาคะแนนเฉลี่ยของการทดสอบทักษะและกระบวนการทางคณิตศาสตร์ของนักเรียนคนนี้

ด้านที่	**ทักษะและกระบวนการทางคณิตศาสตร์**	**คะแนนที่สอบได้**	**ความสำคัญของคะแนน**
1	การแก้ปัญหา	54	30
2	การให้เหตุผล	65	20
3	การสื่อสาร การสื่อความหมาย และการนำเสนอ	70	15
4	การเชื่อมโยงความรู้ทางคณิตศาสตร์	55	20
5	ความคิดริเริ่มสร้างสรรค์	75	15
รวม			**100**

วิธีทำ ค่าจากการสังเกตมี 5 ค่า คือ $x_1 = 54$, $x_2 = 65$, $x_3 = 70$, $x_4 = 55$ และ $x_5 = 75$

ความสำคัญของคะแนน คือ $w_1 = 30$, $w_2 = 20$, $w_3 = 15$, $w_4 = 20$ และ $w_5 = 15$

คะแนนเฉลี่ยของการทดสอบทักษะและกระบวนการทางคณิตศาสตร์ของนักเรียนคนนี้

คือ $$\mu = \frac{\sum_{i=1}^{5} w_i x_i}{\sum_{i=1}^{5} w_i}$$

จะได้ $$\mu = \frac{30(54)+20(65)+15(70)+20(55)+15(75)}{30+20+15+20+15}$$

$$= \frac{6,195}{100}$$

$$= 61.95$$

ดังนั้น เมื่อคำนึงถึงความสำคัญหรือน้ำหนักของทักษะและกระบวนการทางคณิตศาสตร์ทั้ง 5 ด้าน คะแนนเฉลี่ยของนักเรียนคนนี้ คือ 61.95 คะแนน

การกำหนดความสำคัญหรือน้ำหนักของค่าจากการสังเกตแต่ละค่า เช่น ในตัวอย่างที่ 2 กำหนดให้ $w_1 = 30$, $w_2 = 20$, $w_3 = 15$, $w_4 = 20$ และ $w_5 = 15$ ถ้าจะกำหนดให้ $w_1 = 2$, $w_2 = \frac{4}{3}$, $w_3 = 1$, $w_4 = \frac{4}{3}$ และ $w_5 = 1$ หรือกำหนดให้มีค่าอื่นใดที่เป็นสัดส่วนเดียวกันกับความสำคัญของค่าจากการสังเกตที่กำหนดไว้เดิม ค่าเฉลี่ยเลขคณิตที่ได้จะไม่เปลี่ยนแปลง นั่นคือ 61.95 คะแนนเท่าเดิม

การหาค่าเฉลี่ยเลขคณิตของข้อมูลที่แจกแจงความถี่

ในกรณีที่มีข้อมูลจำนวนมาก และไม่มีข้อมูลดิบแต่ละหน่วยเช่นข้อมูลที่รายงานจากทะเบียนต่าง ๆ ในลักษณะที่ได้แจกแจงความถี่แล้ว ถ้าให้ f_1 เป็นความถี่ของค่าจากการสังเกต x_1 หรืออันตรภาคชั้นที่ 1 ที่มี x_1 เป็นจุดกึ่งกลางชั้น f_2 เป็นความถี่ของค่าจากการสังเกต x_2 หรืออันตรภาคชั้นที่ 2 ที่มี x_2 เป็นจุดกึ่งกลางชั้น เรื่อยไปจนถึง f_k เป็นความถี่ของค่าจากการสังเกต x_k หรืออันตรภาคชั้นที่ k ที่มี x_k เป็นจุดกึ่งกลางชั้น แล้วสามารถคำนวณค่าเฉลี่ยได้ดังนี้

$$\mu = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \cdots + f_kx_k}{f_1 + f_2 + f_3 + \cdots + f_k}$$

$$= \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

$$= \frac{\sum_{i=1}^{k} f_i x_i}{N}$$

เมื่อ N เป็นจำนวนค่าจากการสังเกตทั้งหมด หรือ $N = \sum_{i=1}^{k} f_i$

x_i เป็นจุดกึ่งกลางของชั้นที่ i

f_i เป็นความถี่ของอันตรภาคชั้นที่ i

k เป็นจำนวนอันตรภาคชั้น

หมายเหตุ ถ้าเป็นค่าเฉลี่ยเลขคณิตของตัวอย่างที่เป็นตัวแทนของประชากรยังคงใช้สูตรของค่าเฉลี่ยเลขคณิตแบบเดิม แต่เปลี่ยน μ เป็น $\overline{X}$ และหารด้วย n แทน N

การคำนวณหาค่าเฉลี่ยเลขคณิตโดยวิธีนี้ ใช้สูตรทำนองเดียวกันกับการคำนวณหาค่าเฉลี่ยเลขคณิตโดยวิธีถ่วงน้ำหนักนั่นเอง โดยที่ความสำคัญหรือน้ำหนักในที่นี้คือ ความถี่ของแต่ละอันตรภาคชั้น ค่าของข้อมูลที่อยู่ในแต่ละอันตรภาคชั้นจะประมาณด้วยจุดกึ่งกลางของแต่ละชั้น หรือจะสมมุติว่าค่าทุกค่ามีค่าเท่ากับจุดกึ่งกลางของแต่ละอันตรภาคชั้น ทั้งนี้ถือว่าจุดกึ่งกลางของอันตรภาคชั้นโดยทั่ว ๆ ไปเป็นตัวแทนที่ดีที่สุดของค่าจากการสังเกตที่มีอยู่ในอันตรภาคชั้นนั้น ๆ

ในปัจจุบันมีการนำคอมพิวเตอร์มาช่วยคำนวณได้โดยสะดวก การหาค่าเฉลี่ยเลขคณิตของข้อมูลที่ต้องแจกแจงความถี่โดยอาศัยการแบ่งอันตรภาคชั้นจึงไม่นิยมใช้ การใช้คอมพิวเตอร์สามารถหาค่าเฉลี่ยของข้อมูลดิบทั้งหมดได้โดยตรง และยังให้ความถูกต้องมากกว่าไม่ว่าข้อมูลจะมีมากเพียงใด ดังนั้น การหาค่าเฉลี่ยเลขคณิตของข้อมูลที่แจกแจงความถี่ จึงใช้สำหรับกรณีที่ไม่มีข้อมูลดิบทุกหน่วย แต่มีเพียงข้อมูลเฉพาะที่ได้แจกแจงความถี่ของแต่ละอันตรภาคชั้นไว้แล้วเท่านั้น เช่น การคำนวณค่าเฉลี่ยดังตัวอย่างที่ 3

ตัวอย่างที่ 3 เงินเดือนของบุคลากรทั้งหมดในสถาบันแห่งหนึ่ง จำนวน 120 คน ซึ่งมีข้อมูลจากทะเบียนที่ให้มาโดยมีการแจกแจงความถี่ไว้แล้วดังนี้

เงินเดือน (บาท)	**จำนวนพนักงาน (คน)**
5,000 – 6,999	10
7,000 – 8,999	11
9,000 – 10,999	25
11,000 – 12,999	20
13,000 – 14,999	19
15,000 – 16,999	20
17,000 – 18,999	10
19,000 – 20,999	5

จงหาค่าเฉลี่ยเลขคณิตของเงินเดือนบุคลากรทั้งหมดในสถาบันแห่งนี้

วิธีทำ เพื่อความสะดวกในการคำนวณหาค่าเฉลี่ยเลขคณิตควรสร้างตาราง ดังนี้

เงินเดือน (บาท)	**จุดกึ่งกลาง (x_i)**	**จำนวนพนักงาน (f_i)**	**$f_i x_i$**
5,000 – 6,999	5,999.5	10	59,995.0
7,000 – 8,999	7,999.5	11	87,994.5
9,000 – 10,999	9,999.5	25	249,987.5
11,000 – 12,999	11,999.5	20	239,990.0
13,000 – 14,999	13,999.5	19	265,990.5
15,000 – 16,999	15,999.5	20	319,990.0
17,000 – 18,999	17,999.5	10	179,995.0
19,000 – 20,999	19,999.5	5	99,997.5
รวม		$\sum_{i=1}^{8} f_i = 120$	$\sum_{i=1}^{8} f_i x_i = 1,503,940$

$$\mu = \frac{\sum_{i=1}^{8} f_i x_i}{\sum_{i=1}^{8} f_i}$$

$$= \frac{1,503,940}{120}$$

$$= 12,532.8333\ldots$$

ดังนั้น เงินเดือนเฉลี่ยของบุคลากรทั้ง 120 คน ประมาณ 12,532.83 บาท

(หรือกล่าวได้ว่าบุคลากรของสถาบันแห่งนี้ได้เงินเดือนที่มีค่ากลางประมาณ 12,532.83 บาท)

สมบัติที่สำคัญของค่าเฉลี่ยเลขคณิต

สมบัติที่สำคัญของค่าเฉลี่ยเลขคณิตมีดังนี้

1) ค่าเฉลี่ยเลขคณิตเมื่อคูณกับจำนวนข้อมูลทั้งหมดไม่ว่าจะเป็นประชากรขนาด N หรือตัวอย่างขนาด n จะเท่ากับผลรวมของข้อมูลทุก ๆ ค่า ตามลำดับดังนี้

$$\sum_{i=1}^{N} x_i = N\mu$$

และ $$\sum_{i=1}^{n} x_i = n\overline{X}$$

2) ผลรวมของผลต่างระหว่างแต่ละค่าของข้อมูลกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น ๆ จะเท่ากับ 0 กล่าวคือ

$$\sum_{i=1}^{N} (x_i - \mu) = 0$$

และ $$\sum_{i=1}^{n} (x_i - \overline{X}) = 0$$

3) ผลรวมของกำลังสองของผลต่างของแต่ละค่าของข้อมูลกับจำนวนจริง M ใด ๆ จะมีค่าน้อยที่สุดเมื่อ M เท่ากับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น กล่าวคือ

$$\sum_{i=1}^{N} (x_i - M)^2 \text{ น้อยที่สุด เมื่อ } M = \mu$$

และ $$\sum_{i=1}^{n} (x_i - M)^2 \text{ น้อยที่สุด เมื่อ } M = \overline{X}$$

หรืออาจเขียนได้อีกอย่างหนึ่งว่า เมื่อ M เป็นจำนวนจริงใด ๆ

$$\sum_{i=1}^{N}(x_i - \mu)^2 \leq \sum_{i=1}^{N}(x_i - M)^2$$

และ
$$\sum_{i=1}^{n}(x_i - \overline{X})^2 \leq \sum_{i=1}^{n}(x_i - M)^2$$

4) ค่าเฉลี่ยเลขคณิตของข้อมูลชุดใด ๆ จะต้องอยู่ระหว่างค่าจากการสังเกตที่น้อยที่สุด และค่าจากการสังเกตที่มากที่สุดในข้อมูลชุดนั้นกล่าวคือ

$$x_{min} < \mu < x_{max}$$

และ
$$x_{min} < \overline{X} < x_{max}$$

เมื่อ x_{min} และ x_{max} เป็นค่าจากการสังเกตที่น้อยที่สุดและค่าจากการสังเกตที่มากที่สุดในข้อมูลชุดนั้น ตามลำดับ

5) ถ้าตัวแปร Y สัมพันธ์กับตัวแปร X ในรูปฟังก์ชันเชิงเส้น นั่นคือ

ถ้า $y_i = ax_i + b$ เมื่อ i คือ 1, 2, 3, ... , N และ a, b เป็นค่าคงตัวใด ๆ

แล้ว $\mu_Y = a\mu_X + b$ เมื่อ μ_Y คือค่าเฉลี่ยเลขคณิตของข้อมูล y_i

μ_X คือค่าเฉลี่ยเลขคณิตของข้อมูล x_i

หรือ $\overline{Y} = a\overline{X} + b$

ในที่นี้จะแสดงวิธีพิสูจน์สมบัติข้อ 3) เฉพาะกรณีข้อมูลระดับประชากรขนาด N หน่วย

พิสูจน์ สมบัติข้อ 3)

ให้ M เป็นจำนวนจริงใด ๆ

$$\sum_{i=1}^{N}(x_i - M)^2 = \sum_{i=1}^{N}(x_i - \mu + \mu - M)^2$$

$$= \sum_{i=1}^{N}\{(x_i - \mu) + (\mu - M)\}^2$$

$$= \sum_{i=1}^{N}\{(x_i-\mu)^2+2(x_i-\mu)(\mu-M)+(\mu-M)^2\}$$

$$= \sum_{i=1}^{N}(x_i-\mu)^2+2(\mu-M)\sum_{i=1}^{N}(x_i-\mu)+\sum_{i=1}^{N}(\mu-M)^2$$

แต่ $\sum_{i=1}^{N}(x_i-\mu) = 0$ (ตามสมบัติข้อ 2))

และ $\sum_{i=1}^{N}(\mu-M)^2 = N(\mu-M)^2$

ดังนั้น $\sum_{i=1}^{N}(x_i-M)^2 = \sum_{i=1}^{N}(x_i-\mu)^2+N(\mu-M)^2$

แต่เนื่องจาก $N(\mu-M)^2 \geq 0$ เสมอ

ดังนั้น $\sum_{i=1}^{N}(x_i-\mu)^2 \leq \sum_{i=1}^{N}(x_i-M)^2$

ตัวอย่างต่อไปนี้แสดงการใช้สมบัติข้อ 5)

ตัวอย่างที่ 4 บริษัทแห่งหนึ่งต้องการกำหนดเบี้ยประกันอุบัติเหตุรถยนต์โดยคำนวณจากความสัมพันธ์ $Y = 0.73X + 2,500$ เมื่อ Y เป็นตัวแปรแทนเบี้ยประกันอุบัติเหตุรถยนต์มีหน่วยเป็นบาท และ X เป็นตัวแปรแทนอายุการใช้งานของรถยนต์มีหน่วยเป็นปี จากการสำรวจในปีที่ผ่านมา พบว่าอายุการใช้งานของรถยนต์ส่วนมากเป็น 2, 3.5, 5, 6.5, 7, 8, 9, 6, 8.5 และ 10 ปี จงหาค่าเฉลี่ยเลขคณิตของเบี้ยประกันอุบัติเหตุรถยนต์ของบริษัทนี้

วิธีทำ จากโจทย์กำหนด Y เป็นตัวแปรแทนเบี้ยประกันภัยอุบัติเหตุรถยนต์มีหน่วยเป็นบาท และ X เป็นตัวแปรแทนอายุการใช้งานของรถยนต์มีหน่วยเป็นปี จากความสัมพันธ์ของ X และ Y คือ

$$Y = 0.73X + 2,500$$

ในที่นี้ $y_i = 0.73x_i + 2,500$ เมื่อ i คือ 1, 2, 3, ... , 10

ดังนั้น $\overline{Y} = 0.73\overline{X} + 2,500$

แต่ $\overline{X} = \dfrac{2+3.5+5+6.5+7+8+9+6+8.5+10}{10}$

$$= \frac{65.5}{10}$$

$$= 6.55$$

ดังนั้น $\overline{Y} = 0.73(6.55) + 2{,}500 = 2{,}504.7815$

นั่นคือ ค่าเฉลี่ยของเบี้ยประกันอุบัติเหตุรถยนต์ของบริษัทนี้ประมาณ 2,504.78 บาท

ค่าเฉลี่ยเลขคณิตรวม (combined arithmetic mean)

ในการวิเคราะห์ข้อมูลของตัวแปรเดียวกันจากตัวอย่างหลาย ๆ ชุดที่สุ่มมาจากประชากรเดียวกันและหาค่าเฉลี่ยเลขคณิตของตัวอย่างแต่ละชุดไว้แล้ว หากผู้วิเคราะห์ต้องการทราบค่าเฉลี่ยเลขคณิตของข้อมูลทั้งหมด โดยนับรวมเป็นชุดเดียวกันก็สามารถหาได้จากค่าเฉลี่ยเลขคณิตของข้อมูลแต่ละชุด ที่คำนวณไว้แล้ว กล่าวคือ

ถ้า $\overline{X}_1, \overline{X}_2, \overline{X}_3, ..., \overline{X}_k$ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่ 1, 2, 3, … , k ตามลำดับ

$n_1, n_2, n_3, \ldots, n_k$ เป็นจำนวนค่าจากการสังเกตในข้อมูลชุดที่ 1, 2, 3, … , k ตามลำดับ

ค่าเฉลี่ยเลขคณิตรวม

$$\overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3 + \cdots + n_k\overline{X}_k}{n_1 + n_2 + n_3 + \cdots + n_k}$$

$$= \frac{\sum_{i=1}^{k} n_i\overline{X}_i}{\sum_{i=1}^{k} n_i}$$

และถ้าข้อมูลเป็นระดับประชากร การคำนวณยังคงใช้สูตรทำนองเดียวกันแต่เปลี่ยน $\overline{X}$ เป็น μ และ n เป็น N ในข้อมูลแต่ละชุด

ตัวอย่างที่ 5 คะแนนเฉลี่ยจากการสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 จากตัวอย่างจำนวนสามห้องของโรงเรียนแห่งหนึ่งเป็น 50, 65 และ 76 คะแนน โดยมีจำนวนนักเรียนในแต่ละห้องเป็น 40, 50 และ 30 คน ตามลำดับ จงหาค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 ทั้งสามห้อง

วิธีทำ ให้ n_i และ $\overline{X}_i$ แทนจำนวนนักเรียนและคะแนนเฉลี่ยจากการสอบ เมื่อ i คือห้องที่ 1, 2 และ 3 ตามลำดับ และให้ $n_1 = 40$, $\overline{X}_1 = 50$; $n_2 = 50$, $\overline{X}_2 = 65$ และ $n_3 = 30$, $\overline{X}_3 = 76$

จะได้ ค่าเฉลี่ยเลขคณิตรวม $\overline{X} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3}$

$$= \frac{(40\times 50)+(50\times 65)+(30\times 76)}{40+50+30} = \frac{7,530}{120}$$

$$= 62.75$$

ดังนั้น คะแนนสอบวิชาคณิตศาสตร์โดยเฉลี่ยของนักเรียนชั้นมัธยมศึกษาปีที่ 4 ทั้งสามห้องคือ 62.75 คะแนน

หมายเหตุ จากตัวอย่างที่ 5 จะสังเกตว่าค่าเฉลี่ยเลขคณิตรวมที่ได้จะอยู่ระหว่าง 50 และ 76 คะแนนซึ่งเป็นคะแนนเฉลี่ยต่ำสุดและสูงสุด ตามลำดับ

ตัวอย่างที่ 6 คะแนนสอบวิชาคณิตศาสตร์โดยเฉลี่ยของนักเรียนในโรงเรียนมัธยมศึกษาแห่งหนึ่งได้ผลดังตาราง

ชั้น	จำนวนนักเรียน	คะแนนเฉลี่ย
ม.1	50	65
ม.2	40	70
ม.3	45	60
ม.4	50	75
ม.5	60	50
ม.6	50	70
รวม	295	390

จงหาค่าเฉลี่ยเลขคณิตรวมของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนในโรงเรียนนี้

วิธีทำ ให้ n_i และ μ_i แทนจำนวนนักเรียนและคะแนนเฉลี่ยของการสอบ เมื่อ i คือ ชั้นมัธยมศึกษาปีที่ 1, 2, 3, 4, 5 และ 6 ตามลำดับ

จะได้ว่า ค่าเฉลี่ยเลขคณิตรวม

$$\mu = \frac{N_1\mu_1 + N_2\mu_2 + N_3\mu_3 + N_4\mu_4 + N_5\mu_5 + N_6\mu_6}{N_1+N_2+N_3+N_4+N_5+N_6}$$

$$= \frac{(50\times 65)+(40\times 70)+(45\times 60)+(50\times 75)+(60\times 50)+(50\times 70)}{50+40+45+50+60+50}$$

$$= \frac{19,000}{295}$$

$$\approx 64.41$$

ดังนั้น คะแนนเฉลี่ยรวมวิชาคณิตศาสตร์ของนักเรียนในโรงเรียนแห่งนี้ประมาณ 64.41 คะแนน กล่าวคือ ค่ากลางของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนโรงเรียนนี้ประมาณ 64.41 คะแนน

หมายเหตุ จะเห็นว่าสามารถคำนวณค่าเฉลี่ยเลขคณิตรวม โดยใช้สูตรเดียวกันทั้งสำหรับข้อมูลระดับประชากรทั้งหมด หรือสำหรับข้อมูลบางส่วนที่เป็นตัวอย่าง ดังนั้น จึงควรใช้สัญลักษณ์ให้ชัดเจนเพื่อแสดงความแตกต่างกันของค่าเฉลี่ยและจำนวนข้อมูลของแต่ละแบบ นั่นคือ μ และ N หรือ $\overline{X}$ และ n ตามลำดับ ในระดับเบื้องต้นนี้ ถ้าไม่ได้ระบุชัดเจนว่าเป็นข้อมูลตัวอย่าง ให้ทำการวิเคราะห์ข้อมูลในระดับประชากร

แบบฝึกหัด 1.1 ก

1. ถ้า $x_1 = 1,\ x_2 = 3,\ x_3 = 4,\ x_4 = 7,\ x_5 = 0,\ f_1 = 10,\ f_2 = 15,\ f_3 = 5,\ f_4 = 8,\ f_5 = 6$ และ $c = 2$ จงหาค่าของ

   1) $\sum_{i=1}^{10} c$ 2) $\sum_{i=1}^{5} (x_i - 2)^3$

   3) $\sum_{i=1}^{3} (f_i x_i + c)$ 4) $\sum_{i=1}^{4} (x_i - 3)(x_i + 3)$

2. ถ้า $\sum_{i=1}^{5} y_i = 10$ และ $\sum_{i=1}^{5} y_i^2 = 30$ จงหาค่าของ $\sum_{i=1}^{5} (5y_i - 50)$ และ $\sum_{i=1}^{5} (y_i - 3)^2$

3. ถ้า $\sum_{i=1}^{4} x_i = 5$ $\sum_{i=1}^{4} y_i = -2$ และ $\sum_{i=1}^{4} x_i y_i = 4$ จงหาค่าของ $\sum_{i=1}^{4} (x_i + 1)(4y_i - 3)$

4. จงเขียนผลบวกของพจน์ต่อไปนี้โดยใช้เครื่องหมาย Σ

   1) $2x_1^2 + 2x_2^2 + 2x_3^2 + \cdots + 2x_{10}^2$

2) $(x_1-\overline{X})f_1 + (x_2-\overline{X})f_2 + (x_3-\overline{X})f_3 + \cdots + (x_k-\overline{X})f_k$

3) $\frac{1}{n}\left\{(y_1-\overline{Y})^2 f_1 + (y_2-\overline{Y})^2 f_2 + (y_3-\overline{Y})^2 f_3 + \cdots + (y_k-\overline{Y})^2 f_k\right\}$

5. จงแสดงว่า

$$\sum_{i=1}^{N}(x_i - 3y_i + 2z_i + 1) \;=\; \sum_{i=1}^{N}x_i - 3\sum_{i=1}^{N}y_i + 2\sum_{i=1}^{N}z_i + N$$

6. ในการทดสอบวิชาสถิติของนักเรียนห้องหนึ่งจำนวน 50 คน ปรากฏว่ามีนักเรียนจำนวน 40 คน ที่ได้คะแนนอยู่ระหว่าง 60 – 80 คะแนน ส่วนอีก 10 คนที่เหลือ ได้คะแนนอยู่ระหว่าง 90 – 100 คะแนน

   1) จงหาค่าเฉลี่ยเลขคณิตโดยประมาณของคะแนนสอบวิชาสถิติของนักเรียนทั้ง 50 คนนี้

   2) ถ้าทราบว่านักเรียน 40 คน ได้ค่าเฉลี่ยเลขคณิตของคะแนนสอบเป็น 75 คะแนน และอีก 10 คนที่เหลือได้ค่าเฉลี่ยเลขคณิตของคะแนนสอบเป็น 95 คะแนน ค่าเฉลี่ยเลขคณิตของคะแนนสอบของนักเรียนทั้งหมด 50 คน ที่คำนวณได้ จะเท่ากับค่าเฉลี่ยเลขคณิตที่หาได้ในข้อ 1) หรือไม่

7. โรงเรียนแห่งหนึ่งมีชั้นมัธยมศึกษาปีที่ 5 อยู่ 4 ห้อง โดยมีนักเรียน 40, 45, 50 และ 45 คน ตามลำดับ จากการวัดส่วนสูงของนักเรียนแต่ละคนแล้วคำนวณค่าเฉลี่ยเลขคณิตของส่วนสูงของนักเรียนแต่ละห้องปรากฏว่าได้ 165, 168, 167 และ 164 เซนติเมตร ตามลำดับ จงหาค่าเฉลี่ยเลขคณิตรวมของส่วนสูงของนักเรียนชั้นมัธยมศึกษาปีที่ 5 ทั้งหมดของโรงเรียนนี้

8. ถ้าความสัมพันธ์ระหว่างราคาซื้อ (B) และราคาขาย (S) ของสินค้าชนิดหนึ่งเป็น $S = 10 + 1.4B$ และพ่อค้า 10 ราย ซื้อสินค้าดังกล่าวมาด้วยราคา 80, 85, 70, 80, 75, 78, 82, 86, 79 และ 69 บาท ตามลำดับ จงหาค่าเฉลี่ยเลขคณิตของราคาขายของสินค้าชนิดนี้

9. ธนากรต้องการซื้อไข่ไก่ขนาดใหญ่จำนวน 100 ฟอง จึงสำรวจราคาไข่ไก่ขนาดเดียวกันนี้จากร้านค้าใกล้บ้านสามร้าน ได้ข้อมูลราคาไข่ไก่จากร้านที่ 1, 2 และ 3 เป็น 2.30, 2.00 และ 1.70 บาท ตามลำดับ ถ้าเขาซื้อไข่ไก่จากร้านที่ 1, 2 และ 3 จำนวน 50, 30 และ 20 ฟอง ตามลำดับ โดยเฉลี่ยแล้วธนากรซื้อไข่ไก่มาฟองละเท่าไร

10. จงพิสูจน์สมบัติที่สำคัญของค่าเฉลี่ยเลขคณิตข้อ 1), 2), 4) และ 5)

1.1.2 มัธยฐาน (median)

มัธยฐาน คือ ค่าที่มีตำแหน่งอยู่กึ่งกลางของข้อมูลทั้งหมด เมื่อเรียงลำดับข้อมูลจากค่าน้อยที่สุดไปหาค่ามากที่สุด หรือจากค่ามากที่สุดไปหาค่าน้อยที่สุด นั่นคือมัธยฐานเป็นค่าที่แสดงให้ทราบว่ามีจำนวนข้อมูลที่มากกว่าและน้อยกว่าค่ามัธยฐานอยู่เท่า ๆ กัน ดังนั้นมัธยฐานเป็นค่าที่แบ่งข้อมูลที่เรียงลำดับแล้วออกเป็น 2 ส่วน โดยมีข้อมูลจำนวนที่มากกว่าและน้อยกว่าค่ามัธยฐานร้อยละ 50 ค่ามัธยฐานอาจเป็นค่าใดค่าหนึ่งของข้อมูลซึ่งเป็นค่าจากการสังเกตหรืออาจเป็นค่าที่คำนวณขึ้นมาใหม่ที่ไม่ตรงกับค่าของข้อมูลในชุดนั้น ๆ ก็ได้

จุดเด่นของการใช้ค่ามัธยฐาน คือ ค่ามัธยฐานเป็นค่าเหมาะสมที่จะนำมาใช้เป็นค่ากลางของข้อมูล เมื่อข้อมูลนั้น ๆ มีค่าใดค่าหนึ่งหรือหลาย ๆ ค่า ซึ่งสูงหรือต่ำกว่าค่าอื่น ๆ มาก หรือต้องการทราบว่าค่าที่เป็นไปได้ค่าใดของข้อมูลนั้น ๆ มีจำนวนค่าสังเกตที่มากกว่าและน้อยกว่าค่านี้อยู่ประมาณเท่า ๆ กัน

การหามัธยฐานของข้อมูลที่ไม่ได้แจกแจงความถี่

การหามัธยฐานโดยวิธีนี้หาโดยการเรียงข้อมูลที่มีอยู่ทั้งหมดจากค่าน้อยไปหาค่ามากหรือจากค่ามากไปหาค่าน้อยอย่างใดอย่างหนึ่ง แล้วดูว่าค่าของข้อมูลค่าใดอยู่ตรงกึ่งกลางของข้อมูลทั้งหมด ค่านั้นเป็นมัธยฐานของข้อมูลชุดนั้น ในกรณีที่มีจำนวนข้อมูลทั้งหมดอยู่เป็นจำนวนคู่ มัธยฐานจะอยู่ระหว่างข้อมูลสองค่าที่อยู่กึ่งกลางของข้อมูลทั้งหมด ซึ่งนิยมใช้ค่าเฉลี่ยของข้อมูลสองค่านั้น ในกรณีที่จำนวนข้อมูลทั้งหมดเป็นจำนวนคี่ มัธยฐานคือค่าที่อยู่ตรงกึ่งกลางของข้อมูลที่เรียงลำดับทั้งหมด ดังนั้นมัธยฐานอาจเป็นค่าที่ปรากฏอยู่ในข้อมูลชุดนั้นหรือไม่ก็ได้ เช่น มัธยฐานของข้อมูล 12, 13, 15, 17 และ 18 คือ 15 ส่วนมัธยฐานของข้อมูล 66, 63, 63, 62, 61, 60, 60 และ 60 คือ $\frac{62+61}{2} = 61.5$

โดยทั่วไป ถ้าจัดเรียงข้อมูลชุดหนึ่งซึ่งมี N ค่า มัธยฐานจะอยู่ในตำแหน่งที่ $\frac{N+1}{2}$ เช่น ถ้าจำนวนข้อมูลชุดหนึ่งมี 11 ค่า เมื่อจัดเรียงแล้ว มัธยฐานจะอยู่ตำแหน่งที่ $\frac{11+1}{2} = 6$ ถ้าจำนวนข้อมูลมี 14 ค่า เมื่อจัดเรียงแล้ว มัธยฐานจะอยู่ตำแหน่งที่ $\frac{14+1}{2} = 7.5$

ในปัจจุบันการหาค่าสถิติต่าง ๆ สามารถใช้เครื่องคอมพิวเตอร์มาช่วยในการคำนวณได้โดยง่าย การหาค่ามัธยฐานก็เช่นกัน ถ้ามีข้อมูลดิบครบทุกหน่วย นิยมใช้วิธีหาค่ามัธยฐานของข้อมูลที่ไม่ได้แจกแจงความถี่มากกว่าวิธีหาค่ามัธยฐานของข้อมูลที่แจกแจงความถี่แล้ว

ตัวอย่างที่ 7 จงหามัธยฐานของจำนวนเงินฝากของธนาคารพาณิชย์ในประเทศไทยในรอบ 10 ปีที่ผ่านมา ตั้งแต่ พ.ศ. 2536 ถึง พ.ศ. 2545 ซึ่งมีจำนวนเงินฝากในแต่ละปีดังในตาราง

พ.ศ.	**จำนวนเงินฝาก (ล้านล้านบาท)**
2536	2.43
2537	2.76
2538	3.25
2539	3.68
2540	4.31
2541	4.69
2542	4.67
2543	4.91
2544	5.11
2545	5.22

ที่มา : ธนาคารแห่งประเทศไทย

วิธีทำ เมื่อเรียงจำนวนเงินฝากของธนาคารพาณิชย์ในประเทศไทยแต่ละปี จากจำนวนเงินฝากที่น้อยที่สุดจนถึงจำนวนเงินฝากที่มากที่สุดจะได้ดังนี้

2.43 2.76 3.25 3.68 4.31 4.67 4.69 4.91 5.11 5.22

มัธยฐานอยู่ในตำแหน่งที่ $\frac{10+1}{2} = 5.5$

ดังนั้น มัธยฐานของจำนวนเงินฝากของธนาคารพาณิชย์ในประเทศไทยเท่ากับ $\frac{4.31+4.67}{2} = \frac{8.98}{2} = 4.49$ ล้านล้านบาท หรือ 4,490,000 ล้านบาท กล่าวคือ ธนาคารพาณิชย์ในประเทศไทยมีจำนวนเงินฝากประมาณปีละ 4,490,000 ล้านบาท

การหามัธยฐานของข้อมูลที่แจกแจงความถี่

สำหรับข้อมูลที่แจกแจงความถี่ การหามัธยฐานในกรณีนี้จะให้ค่าโดยประมาณ เนื่องจากไม่สามารถนำข้อมูลมาเรียงกันได้ แต่ทราบว่าในแต่ละอันตรภาคชั้นมีข้อมูลอยู่จำนวนเท่าใด (ความถี่) ก่อนอื่นจึงต้องหาว่ามัธยฐานตกอยู่ในอันตรภาคชั้นใด โดยพิจารณาจากความถี่สะสม แล้วหาค่าโดยประมาณของมัธยฐานจากอันตรภาคชั้นนั้น ถ้าข้อมูลชุดที่พิจารณามีผลรวมของความถี่เป็น N มัธยฐานคือค่าที่แสดงให้ทราบว่ามีจำนวนข้อมูลอยู่ต่ำกว่าค่านี้อยู่ $\frac{N}{2}$ จำนวน และมีจำนวนข้อมูลอยู่สูงกว่าค่านี้อยู่ $\frac{N}{2}$ จำนวน

หมายเหตุ ในปัจจุบันวิธีนี้ไม่นิยมใช้แล้ว เนื่องจากว่ามีเครื่องคอมพิวเตอร์ช่วยในการคำนวน จึงควรหามัธยฐานโดยตรงจากข้อมูลดิบทุกหน่วยไม่ต้องแจกแจงความถี่ อย่างไรก็ตามในระดับพื้นฐานชั้นมัธยมศึกษาควรทราบสูตรที่ใช้ในกรณีนี้ ซึ่งควรใช้เฉพาะกรณีที่ไม่มีโอกาสทราบข้อมูลดิบทุกหน่วย

ตัวอย่างที่ 8 จงหามัธยฐานของปริมาณข้าวที่บริษัทข้าวสยามส่งออกไปขายยังต่างประเทศเป็นระยะเวลา 22 ปี ซึ่งมีการแจกแจงความถี่ดังตาราง

ปริมาณข้าวส่งออก (แสนตัน)	ความถี่ (จำนวนปีที่ส่งออก)	ความถี่สะสม
0.80 – 0.99	1	1
1.00 – 1.19	3	4
1.20 – 1.39	6	10
1.40 – 1.59	9	19
1.60 – 1.79	0	19
1.80 – 1.99	1	20
2.00 – 2.19	2	22

วิธีทำ จำนวนปีที่ส่งออกทั้งหมดคือ 22 ดังนั้นครึ่งหนึ่งของจำนวนข้อมูลคือ 11 ซึ่งไม่ตรงกับความถี่สะสมในอันตรภาคชั้นใด จึงจะเทียบหาว่าความถี่สะสม 11 ตรงกับค่าที่เป็นไปได้ค่าใดจากตารางดังนี้

อันตรภาคชั้น 1.20 – 1.39 มีความถี่สะสม 10 หมายความว่า มีข้อมูลต่ำกว่า 1.395 (ซึ่งเป็นขอบบน* ของอันตรภาคชั้นนี้) อยู่ 10 จำนวน

อันตรภาคชั้น 1.40 – 1.59 มีความถี่สะสม 19 หมายความว่า มีข้อมูลต่ำกว่า 1.595 อยู่ 19 จำนวน

เพื่อให้ได้ค่าที่มีข้อมูลต่ำกว่าอยู่ 11 จำนวน จึงต้องคิดเทียบจากอันตรภาคชั้น 1.40 – 1.59 ซึ่งมีความถี่ 9 โดยถือว่าค่าของข้อมูลที่ตกอยู่ในอันตรภาคชั้นนี้ทั้ง 9 ค่ากระจายกันอยู่อย่างสม่ำเสมอ โดยใช้วิธีเทียบส่วนดังนี้

จากตารางความถี่สะสมเพิ่มขึ้นจาก 10 เป็น 19 คือ 9 ค่าที่เป็นไปได้เพิ่มขึ้นจาก 1.395 เป็น 1.595 คือ 0.200

ดังนั้นถ้าความถี่สะสมเพิ่มขึ้นจาก 10 เป็น 11 ค่าที่เป็นไปได้เพิ่มขึ้นคือ $\frac{0.200 \times 1}{9} = 0.022$ และมัธยฐานคือ $1.395 + 0.022 = 1.417$

นั่นคือ ค่าโดยประมาณของมัธยฐานคือ 1.417 แสนตัน

หมายเหตุ ขอบบน (upper boundary) คือค่ากึ่งกลางระหว่างค่าที่มากที่สุดในอันตรภาคชั้นนั้นกับค่าที่น้อยที่สุดของอันตรภาคชั้นถัดไปหนึ่งชั้น ถ้าเป็นขอบบนของช่วงคะแนนสูงสุดให้ถือเสมือนว่ามีอันตรภาคชั้นที่สูงกว่าชั้นนั้นอีกหนึ่งชั้น

ขอบล่าง (lower boundary) คือค่ากึ่งกลางระหว่างค่าที่มากที่สุดในอันตรภาคชั้นก่อนหน้านั้นหนึ่งชั้นกับค่าที่น้อยที่สุดของอันตรภาคชั้นนั้น ถ้าเป็นขอบล่างของช่วงคะแนนต่ำสุดให้ถือเสมือนว่ามีอันตรภาคชั้นที่ต่ำกว่าชั้นนั้นอีกหนึ่งชั้น

* คำอธิบายของขอบบนดูได้จากหมายเหตุ

ในกรณีที่ครึ่งหนึ่งของจำนวนข้อมูลตรงกับความถี่สะสมในอันตรภาคชั้นใดจะใช้ขอบบนของอันตรภาคชั้นนั้นเป็นมัธยฐาน เช่นในตัวอย่างที่ 8 ถ้าไม่มีอันตรภาคชั้นสุดท้ายคือ 2.00 – 2.19 จำนวนข้อมูลทั้งหมดคือ 20 และครึ่งหนึ่งของจำนวนข้อมูลจะเท่ากับ 10 ซึ่งตรงกับความถี่สะสมของอันตรภาคชั้น 1.20 – 1.39 ในกรณีนี้มัธยฐานคือขอบบนของอันตรภาคชั้นซึ่งเท่ากับ 1.395 แสนตัน

การหามัธยฐานของข้อมูลที่แจกแจงความถี่โดยวิธีในตัวอย่างที่ 8 เมื่อนำมาเขียนเป็นสูตรจะได้ดังนี้

$$\text{มัธยฐาน} = L + \left(\frac{\frac{N}{2} - \sum f_L}{f_M}\right) I$$

$$\text{หรือ มัธยฐาน} = U - \left(\frac{\sum f_U - \frac{N}{2}}{f_M}\right) I$$

เมื่อ L และ U เป็นขอบล่างและขอบบนของอันตรภาคชั้นที่มีมัธยฐานอยู่ ตามลำดับ

- N เป็นผลรวมของความถี่ทั้งหมด ($\sum_{i=1}^{k} f_i$)
- $\sum f_L$ เป็นผลรวมของความถี่ของทุกอันตรภาคชั้นที่เป็นช่วงคะแนนต่ำกว่าชั้นที่มีมัธยฐานอยู่
- $\sum f_U$ เป็นผลรวมของความถี่ของอันตรภาคชั้นที่มีมัธยฐานอยู่และทุกชั้นที่เป็นช่วงคะแนนต่ำกว่า
- f_M เป็นความถี่ของอันตรภาคชั้นที่มีมัธยฐานอยู่
- I เป็นความกว้างของอันตรภาคชั้นที่มีมัธยฐานอยู่

การหามัธยฐานโดยใช้สูตรใดสูตรหนึ่งในสองสูตรนี้จะให้ผลลัพธ์เท่ากัน ผู้เรียนควรลองหามัธยฐานในตัวอย่างที่ 8 โดยใช้สูตรทั้งสอง แล้วเปรียบเทียบผลลัพธ์ที่ได้ และสูตรที่ใช้ยังสามารถใช้ได้ทั้งสำหรับประชากร N หน่วย และสำหรับตัวอย่าง n หน่วย ด้วย

สมบัติของมัธยฐาน

สมบัติที่สำคัญข้อหนึ่งของมัธยฐานคือ ผลรวมของค่าสัมบูรณ์ของผลต่างระหว่างข้อมูลแต่ละค่ากับมัธยฐานของข้อมูลชุดนั้นจะมีค่าน้อยที่สุด กล่าวคือ

$$\sum_{i=1}^{N}|x_i - \text{มัธยฐาน}| \quad \text{มีค่าน้อยที่สุด}$$

ถ้าแทนมัธยฐานด้วยจำนวนอื่น จะได้ผลรวมมากกว่าหรือเท่ากับค่านี้เสมอ เช่น ถ้ากำหนดข้อมูล 3, 4, 6, 13, 8, 2 แล้วมัธยฐานของข้อมูลชุดนี้คือ 5

และ $$\sum_{i=1}^{6}|x_i - 5| = |3-5| + |4-5| + |6-5| + |13-5| + |8-5| + |2-5| = 18$$

แต่ถ้าเปลี่ยน 5 เป็นจำนวนอื่น เช่น 7 หรือ 9 จะได้ผลรวมเป็น 20 และ 26 ตามลำดับ หรือถ้าเปลี่ยน 5 เป็น 4 จะได้ผลรวมเป็น 18 นั่นคือ ถ้าแทนค่ามัธยฐานด้วยจำนวนอื่น จะได้ผลรวมมากกว่าหรือเท่ากับที่ใช้มัธยฐาน

แบบฝึกหัด 1.1 ข

1. ค่าใช้จ่ายรายวัน (บาท) ของนักเรียน 10 คน เป็นดังนี้

   11 15 22 36 11 18 22 22 16 28

   จงหามัธยฐานของข้อมูลชุดนี้ และมีนักเรียนกี่คนที่มีค่าใช้จ่ายรายวันมากกว่ามัธยฐานที่หาได้

2. จงหามัธยฐานของจำนวนผู้มีงานทำจำแนกตามประเภทอุตสาหกรรมในปี พ.ศ. 2546 ดังตาราง

ประเภทอุตสาหกรรม	จำนวนผู้มีงานทำ (พันคน)
การประมง	466.4
การทำเหมืองแร่และหิน	44.3
การผลิต	5,270.9
การก่อสร้าง	1,724.4
โรงแรมและภัตตาคาร	2,148.8
การศึกษา	974.0
การขนส่ง สถานที่เก็บสินค้าและคมนาคม	1,080.8

ที่มา : สำนักงานสถิติแห่งชาติ กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร

3. ถ้า $x_1, x_2, x_3, \dots, x_N$ เป็นข้อมูลซึ่งเรียงจากน้อยไปหามาก หรือมากไปหาน้อย จงหามัธยฐาน เมื่อ 1) N เป็นจำนวนคู่ 2) N เป็นจำนวนคี่

1.1.3 ฐานนิยม (mode)

ฐานนิยม คือ ค่าของข้อมูลที่มีความถี่สูงสุด ใช้เป็นค่ากลางของข้อมูลอีกชนิดหนึ่ง นอกเหนือจากค่าเฉลี่ยเลขคณิตและมัธยฐานที่ได้กล่าวมาแล้ว ส่วนมากฐานนิยมจะใช้กับข้อมูลเชิงคุณภาพมากกว่าข้อมูลเชิงปริมาณ

ฐานนิยมเหมาะสมกับการใช้เป็นค่ากลางของข้อมูล เมื่อข้อมูลนั้น ๆ เป็นข้อมูลเชิงคุณภาพ เช่น ขนาดรองเท้า ขนาดยางรถยนต์ ฯลฯ หรือข้อมูลที่แจกแจงความถี่ตามกลุ่มหรือช่วงต่าง ๆ และโดยเฉพาะเมื่อมีข้อมูลที่มีค่าสูงหรือต่ำผิดปกติรวมอยู่ด้วย

การหาฐานนิยมของข้อมูลที่ไม่ได้แจกแจงความถี่

ฐานนิยมของข้อมูลชนิดนี้หาได้จากการดูว่าข้อมูลค่าใดจากข้อมูลที่มีอยู่ทั้งหมดมีความถี่สูงสุดหรือปรากฏบ่อยครั้งที่สุด ข้อมูลนั้นจะเป็นฐานนิยมของข้อมูลชุดนั้น

ตัวอย่างที่ 9 จงหาฐานนิยมของอายุนักเรียนที่มาเข้าค่ายคณิตศาสตร์ จำนวน 15 คน ซึ่งมีอายุดังนี้ 5, 8, 7, 6, 7, 8, 12, 11, 10, 11, 8, 6, 8, 7 และ 8 ปี

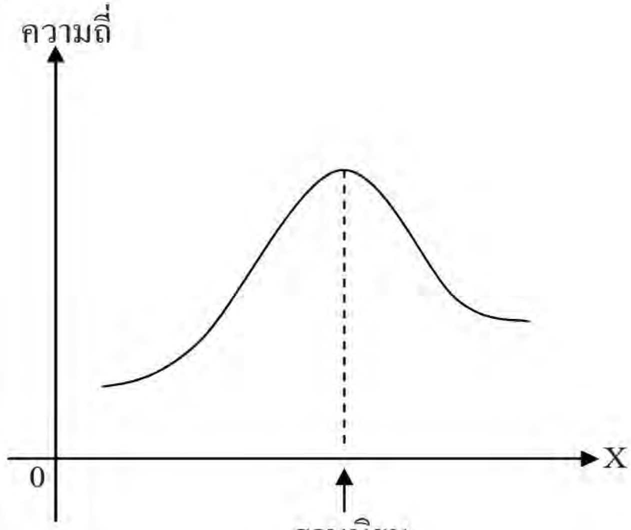
วิธีทำ ฐานนิยมของอายุนักเรียนที่มาเข้าค่ายคณิตศาสตร์ทั้ง 15 คน คือ 8 ปี เพราะนักเรียนที่มาเข้าค่ายคณิตศาสตร์ที่มีอายุ 8 ปี มีจำนวนมากที่สุดคือ 5 คน

การหาฐานนิยมโดยวิธีดังกล่าว ข้อมูลบางชุดอาจจะไม่มีฐานนิยมเลยก็ได้ เช่น ข้อมูลที่ประกอบด้วย 5, 8, 9, 10, 12, 18, 16, 20 จะไม่มีฐานนิยมเลยเพราะข้อมูลแต่ละค่ามีความถี่เท่ากันหมด ข้อมูลอีกชุดหนึ่งประกอบด้วย 13, 16, 20, 25, 20, 26, 25 มีข้อมูลที่มีความถี่สูงสุดเท่ากันสองค่าคือ 20 และ 25 ในกรณีนี้จะถือได้ว่าข้อมูลชุดนี้ไม่มีฐานนิยม แต่ในกรณีที่มีจำนวนข้อมูลมากเพียงพอ และมีความถี่สูงสุดมากกว่า 1 ค่า เช่น โรงงานแห่งหนึ่งสำรวจการซื้อเสื้อของประชากร พบว่าเสื้อขนาด S และ L มีผู้นิยมซื้อมากที่สุด ในกรณีนี้โรงงานอาจจะใช้ข้อมูลนี้ในการผลิตสินค้าเหมือนกับเป็นฐานนิยมได้

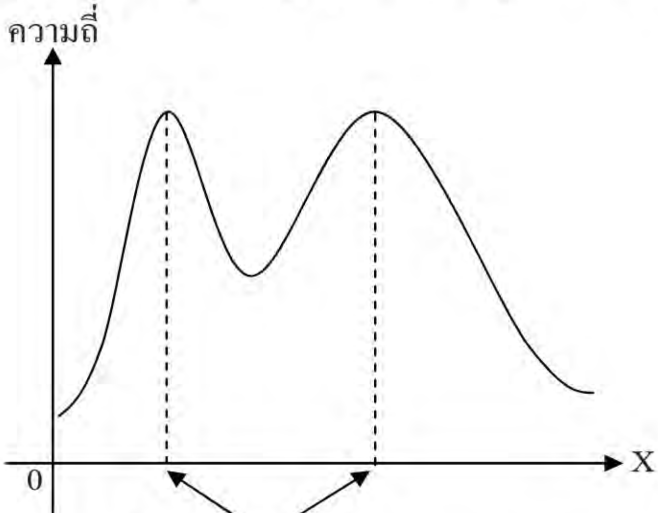
หมายเหตุ ในกรณีที่ข้อมูลชุดหนึ่งมีข้อมูลที่มีความถี่สูงสุดมากกว่า 1 ค่า อาจหาตัวแปรเชิงคุณภาพอื่น ๆ ที่เกี่ยวข้อง เช่น เพศ ชั้นเรียน ฯลฯ มาแบ่งข้อมูลออกเป็นคนละชุด เพื่อทำให้แต่ละชุดมีฐานนิยมเพียง 1 ค่า

การหาฐานนิยมของข้อมูลที่แจกแจงความถี่

ถ้าเขียนเส้นโค้งความถี่ของข้อมูลที่แจกแจงความถี่ ฐานนิยมคือ ค่าของ X ที่อยู่ตรงกับจุดสูงสุดบนเส้นโค้งความถี่ ดังรูป



แต่ถ้าเส้นโค้งความถี่มีจุดสูงสุดสองจุด ข้อมูลชุดนั้นจะไม่มีฐานนิยม ดังรูป



สำหรับการหาฐานนิยม ถ้ามีข้อมูลครบทุกหน่วย ควรใช้การหาฐานนิยมโดยตรงจากข้อมูลที่ไม่ได้แจกแจงความถี่ เนื่องจากค่าที่คำนวณได้ในกรณีที่ข้อมูลแจกแจงความถี่จะให้ค่าโดยประมาณ

สำหรับการคำนวณหาฐานนิยมของข้อมูลที่แจกแจงความถี่ทำได้หลายวิธี วิธีหนึ่งคือหาจุดกึ่งกลางของอันตรภาคชั้นที่มีฐานนิยมอยู่ ค่าที่หาได้จะเป็นค่าของฐานนิยมโดยประมาณ ส่วนการหาว่าฐานนิยมอยู่ในอันตรภาคชั้นใดนั้น จะต้องพิจารณาด้วยว่าอันตรภาคชั้นแต่ละชั้นมีความกว้างเท่ากันหรือไม่ ในกรณีที่ความกว้างของอันตรภาคชั้นทุกชั้นเท่ากัน อันตรภาคชั้นที่มีฐานนิยมคืออันตรภาคชั้นที่มีความถี่สูงสุด ส่วนในกรณีที่ความกว้างของอันตรภาคชั้นไม่เท่ากันทุกชั้น ให้หารความถี่ด้วยความกว้างของแต่ละอันตรภาคชั้น โดยอันตรภาคชั้นที่ผลหารมีค่ามากที่สุด จะเป็นอันตรภาคชั้นที่มีฐานนิยมอยู่

ตัวอย่างที่ 10 จากข้อมูลของนักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 40 คน เกี่ยวกับจำนวนเงินที่จ่ายเป็นค่าอาหารในแต่ละสัปดาห์ซึ่งแสดงได้ดังตาราง

จำนวนเงินที่จ่ายเป็นค่าอาหาร (บาท)	จำนวนนักเรียน (คน)
0 – 49	4
50 – 99	7
100 – 149	15
150 – 199	10
200 – 249	3
250 – 299	1

จงหาฐานนิยมโดยประมาณของจำนวนเงินที่จ่ายเป็นค่าอาหารของนักเรียนทั้ง 40 คน ในแต่ละสัปดาห์

วิธีทำ เนื่องจากความกว้างของอันตรภาคชั้นเท่ากันทุกชั้น ดังนั้นอันตรภาคชั้นที่มีฐานนิยมอยู่คืออันตรภาคชั้นที่มีความถี่สูงสุด

อันตรภาคชั้นที่มีความถี่สูงสุดคือ 100 – 149

จุดกึ่งกลางของอันตรภาคชั้นนี้คือ $\frac{100+149}{2} = 124.50$

นั่นคือ ฐานนิยมของจำนวนเงินที่จ่ายเป็นค่าอาหารของนักเรียนทั้ง 40 คน ในแต่ละสัปดาห์ประมาณ 124.50 บาท

ข้อสังเกตและหลักเกณฑ์ที่สำคัญในการใช้ค่ากลางชนิดต่าง ๆ

1) ค่าเฉลี่ยเลขคณิตเป็นค่ากลางที่ได้จากการนำทุก ๆ ค่าของข้อมูลมาเฉลี่ย แต่มัธยฐานเป็นเพียงค่ากลางที่ใช้ตำแหน่งที่ (position) ของข้อมูลบางค่าเท่านั้น ส่วนฐานนิยมเป็นค่ากลางที่พิจารณาจากข้อมูลที่มีความถี่สูงสุด

2) ถ้าในจำนวนข้อมูลทั้งหมดมีข้อมูลบางค่าที่มีค่าสูงหรือต่ำกว่าข้อมูลอื่น ๆ มาก จะมีผลกระทบต่อค่าเฉลี่ยเลขคณิต กล่าวคือ อาจจะทำให้ค่าเฉลี่ยเลขคณิตมีค่าสูงหรือต่ำกว่าข้อมูลที่มีอยู่ส่วนใหญ่ แต่จะไม่มีผลกระทบต่อมัธยฐาน ดังนั้น กรณีเช่นนี้ควรใช้มัธยฐาน

3) มัธยฐานและฐานนิยมใช้เมื่อต้องการทราบค่ากลางของข้อมูลทั้งหมดโดยประมาณและรวดเร็ว ทั้งนี้เนื่องจากการหามัธยฐานและฐานนิยมบางวิธีไม่จำเป็นต้องมีการคำนวณ

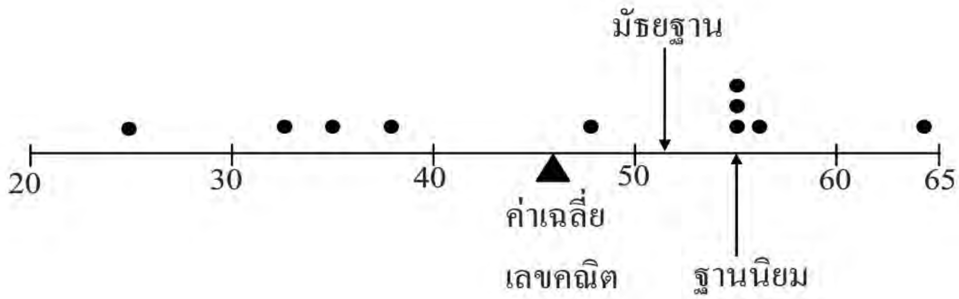
4) ถ้าการแจกแจงความถี่ของข้อมูลประกอบด้วยอันตรภาคชั้นที่มีช่วงเปิดซึ่งอาจเป็นชั้นต่ำสุดหรือชั้นสูงสุดชั้นใดชั้นหนึ่งหรือทั้งสองชั้น การหาค่ากลางโดยใช้ค่าเฉลี่ยเลขคณิตไม่สามารถหาได้ แต่สามารถหามัธยฐานหรือฐานนิยมได้

5) การแจกแจงความถี่ของข้อมูลที่มีความกว้างของแต่ละอันตรภาคชั้นไม่เท่ากัน อาจมีผลทำให้ค่าเฉลี่ยเลขคณิตหรือฐานนิยมคลาดเคลื่อนไปจากที่ควรจะเป็นได้บ้าง แต่ไม่มีผลกระทบต่อการหามัธยฐาน

6) ในกรณีที่ข้อมูลเป็นประเภทข้อมูลเชิงคุณภาพ จะสามารถหาค่ากลางได้เฉพาะฐานนิยมเท่านั้น แต่ไม่สามารถหาค่าเฉลี่ยเลขคณิตหรือมัธยฐานได้

7) ในกรณีที่สามารถนำข้อมูลมาเรียงลำดับได้ ควรหาค่ากลางคือมัธยฐานก่อน และถ้าเป็นข้อมูลเชิงปริมาณที่มีค่าต่อเนื่อง (continuous random variable) ควรใช้ค่าเฉลี่ยเลขคณิตแทนมัธยฐานจะเหมาะสมกว่า

8) ลักษณะเฉพาะของค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม อาจแสดงด้วยข้อมูล 10 ค่าต่อไปนี้ 25 33 35 38 48 55 55 55 56 และ 64 โดยเขียนเป็นแผนภาพได้ดังนี้



จะเห็นว่า ค่าเฉลี่ยเลขคณิตเป็นค่ากลางที่แบ่งน้ำหนักข้อมูล 2 ด้านให้มีสมดุล ส่วนมัธยฐานเป็นค่าที่อยู่ตรงกลางของข้อมูลที่เรียงจากน้อยไปมาก (หรือมากไปน้อย) และฐานนิยมเป็นค่ากลางที่อยู่ตรงจุดที่มีความถี่ของข้อมูลหรือจำนวนข้อมูลที่มากที่สุด

ค่ากลางที่คำนวณโดยใช้ทุก ๆ ค่าของข้อมูล นอกจากค่าเฉลี่ยเลขคณิต แล้วยังมีค่าเฉลี่ยเรขาคณิต (geometric mean) และค่าเฉลี่ยฮาร์มอนิก (harmonic mean) ซึ่งจะกล่าวถึงค่ากลางสองชนิดนี้เพียงย่อ ๆ เนื่องจากการคำนวณยุ่งยากกว่าการหาค่ากลางอื่น ๆ ที่ได้กล่าวมาแล้ว ค่าเฉลี่ยเรขาคณิตมักนิยมใช้เมื่อข้อมูลมีลักษณะเบ้ขวา หรือมีข้อมูลที่มีค่าต่ำมากรวมอยู่ ซึ่งข้อมูลลักษณะดังกล่าวส่วนมากจะเป็นข้อมูลทางชีววิทยา

1.1.4 ค่าเฉลี่ยเรขาคณิต (geometric mean)

ถ้า $x_1, x_2, x_3, \ldots, x_N$ เป็นข้อมูล N จำนวนซึ่งเป็นจำนวนบวกทุกจำนวน

ค่าเฉลี่ยเรขาคณิต $\text{G.M.} = \sqrt[N]{x_1 x_2 x_3 \ldots x_N}$

ในกรณีที่ x_i มีความถี่ f_i และ $\sum_{i=1}^{k} f_i = N$

$$\text{G.M.} = \sqrt[N]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \ldots x_k^{f_k}}$$

หมายเหตุ เนื่องจากการหาค่า G.M. ต้องมีการคำนวณหากรณฑ์ที่ N ของจำนวน ซึ่งทำให้การใช้สูตรดังกล่าวข้างต้นไม่สะดวกในกรณีที่ข้อมูลมีค่ามาก ๆ และต้องใช้เครื่องคิดเลขในการคำนวณ ดังนั้นเพื่อความสะดวกจึงใช้ลอการิทึมช่วย โดยจะได้สูตรในการหาค่า G.M. ดังนี้

สำหรับข้อมูลที่ไม่ได้แจกแจงความถี่

$$\log \text{G.M.} = \frac{1}{N}\sum_{i=1}^{N} \log x_i$$

สำหรับข้อมูลที่แจกแจงความถี่แล้ว

$$\log \text{G.M.} = \frac{1}{N}\sum_{i=1}^{k} f_i \log x_i \qquad \text{โดยที่} \quad \sum_{i=1}^{k} f_i = N$$

เมื่อ x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i โดยที่ i คือ 1, 2, ... , k

f_i แทนความถี่ของข้อมูลอันตรภาคชั้นที่ i

k แทนจำนวนอันตรภาคชั้น

ตัวอย่างที่ 11 จงหาค่าเฉลี่ยเรขาคณิตของ 7.96, 13.82, 22.95, 35.34, 36.54, 32.23, 30.65, 33.46, 31.21, 35.83, 32.16, 33.87, 30.82, 34.26, 35.82

วิธีทำ

log 7.96	=	0.9009	log 13.82	=	1.1405
log 22.95	=	1.3608	log 35.34	=	1.5483
log 36.54	=	1.5628	log 32.23	=	1.5083
log 30.65	=	1.4864	log 33.46	=	1.5245

log 31.21	=	1.4943	log 35.83	=	1.5543
log 32.16	=	1.5073	log 33.87	=	1.5298
log 30.82	=	1.4895	log 34.26	=	1.5348
log 35.82	=	1.5541			

$$\log \text{G.M.} = \frac{1}{15}\sum_{i=1}^{15} \log x_i$$

$$= \frac{21.6966}{15}$$

$$= 1.44644$$

$$\text{G.M.} = 27.95375$$

ดังนั้น ค่าเฉลี่ยเรขาคณิตของข้อมูลชุดนี้ประมาณ 27.95

ตัวอย่างที่ 12 ถ้าสิรินภามีเงินใช้จ่าย 9,300 บาท หรือ 93 หน่วย และต้องการใช้ให้หมดภายใน 5 วัน โดยใช้จ่ายในวันที่ 1 – 5 ดังนี้ 48 24 12 6 3 จงหาค่าเฉลี่ยเรขาคณิตของการใช้จ่ายต่อวัน

วิธีทำ $\log \text{G.M.} = \frac{1}{5}\sum_{i=1}^{5} \log x_i$

$$= \frac{\log 48 + \log 24 + \log 12 + \log 6 + \log 3}{5}$$

$$= \frac{1}{5}(1.6812 + 1.3802 + 1.0792 + 0.7782 + 0.4771)$$

$$= 1.07918$$

$$\text{G.M.} = 11.99997$$

ดังนั้น ค่าเฉลี่ยเรขาคณิตของการใช้จ่ายต่อวันประมาณ 1,200 บาท

หมายเหตุ ค่าเฉลี่ยเรขาคณิตมีประโยชน์ในการหาค่าเฉลี่ยของข้อมูลในกรณีที่ค่าของข้อมูลสูงหรือต่ำกว่าค่าอื่น ๆ รวมอยู่บางค่าหรือหลายค่ามาก กรณีเช่นนี้ค่าเฉลี่ยเรขาคณิตใช้เป็นค่ากลางของข้อมูลได้ดีกว่าค่าเฉลี่ยเลขคณิต เนื่องจากค่าที่สูงหรือต่ำมากเหล่านี้จะมีผลกระทบต่อค่าเฉลี่ยเรขาคณิตของข้อมูลไม่มากนัก

1.1.5 ค่าเฉลี่ยฮาร์มอนิก (harmonic mean)

ถ้า $x_1, x_2, x_3, \ldots, x_N$ เป็นข้อมูล N จำนวนซึ่งเป็นจำนวนบวกทุกจำนวน

ค่าเฉลี่ยฮาร์มอนิก

$$\text{H.M.} = \frac{1}{\frac{1}{N}\left\{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_N}\right\}} = \frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}$$

ในกรณีที่ x_i มีความถี่ f_i และ $\sum_{i=1}^{k} f_i = N$

$$\text{H.M.} = \frac{1}{\frac{1}{N}\left\{f_1\frac{1}{x_1}+f_2\frac{1}{x_2}+f_3\frac{1}{x_3}+\cdots+f_k\frac{1}{x_k}\right\}} = \frac{N}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

เมื่อ k แทนจำนวนอันตรภาคชั้น

x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

หมายเหตุ 1. สูตรการคำนวณค่าเฉลี่ยเรขาคณิตและค่าเฉลี่ยฮาร์มอนิกสามารถทำได้ทั้งสำหรับประชากร N หน่วย และตัวอย่าง n หน่วย

2. ค่าเฉลี่ยฮาร์มอนิกมีประโยชน์ในการหาค่าเฉลี่ยของข้อมูลที่เป็นอัตราส่วน เช่น ระยะทางต่อชั่วโมง จำนวนหน่วยต่อ 1 บาท ราคาสินค้าต่อ 1 โหล ฯลฯ

ตัวอย่างที่ 13 ในโรงงานแห่งหนึ่ง นาย ก ทำงานหนึ่งหน่วยแล้วเสร็จในเวลา 4 นาที นาย ข นาย ค นาย ง และนาย จ ทำงานหน่วยเดียวกันนี้เสร็จในเวลา 5, 6, 10 และ 12 นาที ตามลำดับ จงหาค่าเฉลี่ยของอัตราการทำงานของคนทั้งห้านี้ และอยากทราบว่าใน 6 ชั่วโมง ทั้งห้าคนนี้จะทำงานได้รวมทั้งสิ้นกี่หน่วย

วิธีทำ ข้อมูลชุดนี้กำหนดให้ในรูปเวลาที่ใช้ต่องานหนึ่งหน่วย คือ 4, 5, 6, 10 และ 12 นาที

นั่นคือ นาย ก ทำงาน 1 นาที ได้งาน $\frac{1}{4}$ หน่วย

นาย ข ทำงาน 1 นาที ได้งาน $\frac{1}{5}$ หน่วย

นาย ค ทำงาน 1 นาที ได้งาน $\frac{1}{6}$ หน่วย

นาย ง ทำงาน 1 นาที ได้งาน $\frac{1}{10}$ หน่วย

นาย จ ทำงาน 1 นาที ได้งาน $\frac{1}{12}$ หน่วย

ถ้าต้องการพิจารณาผลงานเฉลี่ยต่อหนึ่งหน่วยเวลา ค่าเฉลี่ยที่เหมาะสมคือ

ค่าเฉลี่ยฮาร์มอนิก H.M. $= \dfrac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}$

$$= \frac{5}{\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{10}+\frac{1}{12}}$$

$$= \frac{25}{4}$$

ดังนั้น ค่าเฉลี่ยของอัตราการทำงานของทั้งห้าคนนี้คือ $\frac{25}{4}$ นาที ต่องานหนึ่งหน่วย

และใน 360 นาที คนทั้งห้าคนนี้จะทำงานได้ $5\times\dfrac{360}{\frac{25}{4}} = 288$ หน่วย

นอกจากนี้ ยังมีวิธีการหาค่ากลางของข้อมูลทั้งหมดอย่างคร่าว ๆ และเสียเวลาในการหาน้อย นั่นคือ ค่ากึ่งกลางพิสัย (mid-range) ซึ่งมีสูตรที่ใช้ในการคำนวณดังนี้

ค่ากึ่งกลางพิสัย $= \dfrac{x_{max} + x_{min}}{2}$

เมื่อ x_{max} และ x_{min} เป็นค่าสูงสุดและค่าต่ำสุดของข้อมูลทั้งหมดตามลำดับ

ในการพิจารณาเลือกใช้ค่ากลางชนิดต่าง ๆ คือ ค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม ค่าเฉลี่ยเรขาคณิต ค่าเฉลี่ยฮาร์มอนิก หรือค่ากึ่งกลางพิสัย ขึ้นอยู่กับตัวประกอบที่สำคัญหลายประการ เช่น การกระจายหรือความแตกต่างระหว่างค่าของข้อมูลแต่ละค่า จำนวนข้อมูลที่มีอยู่ ชนิดของค่ากลางที่ต้องการหรือวัตถุประสงค์ที่จะนำค่ากลางนั้นไปใช้ประโยชน์ตลอดจนลักษณะของการแจกแจงของข้อมูลว่ามีลักษณะใดด้วย

แบบฝึกหัด 1.1 ค

1. จงหาฐานนิยมของอายุ (ปี) ของเด็ก 15 คน ซึ่งมีอายุดังต่อไปนี้
   8, 7, 6, 5, 8, 9, 7, 7, 6, 5, 7, 9, 7, 8, 8

2. ถ้าครอบครัว 20 ครอบครัว บริโภคไข่ไก่ (ฟอง) ต่อเดือนตามจำนวนต่อไปนี้ 60, 32, 51, 60, 48, 52, 46, 38, 35, 60, 48, 54, 44, 49, 47, 48, 48, 44, 65, 60
   จงหาฐานนิยมและค่ากึ่งกลางพิสัยของจำนวนไข่ไก่ที่แต่ละครอบครัวบริโภคต่อเดือน

3. จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของเงินเดือนพนักงาน 7 คน ซึ่งมีเงินเดือนดังนี้ 3500, 3600, 3450, 3500, 3400, 3500 และ 21000 บาท นักเรียนคิดว่าค่ากลางที่หาได้ชนิดใดควรเป็นตัวแทนของเงินเดือนของพนักงานทั้ง 7 คนนี้ได้ดีกว่ากัน จงอธิบาย

4. เครื่องบินลำหนึ่งเดินทางได้ระยะ d_1, d_2 และ d_3 ไมล์ ด้วยอัตราเร็ว v_1, v_2 และ v_3 ไมล์ต่อชั่วโมง ตามลำดับ จงแสดงว่าอัตราเร็วเฉลี่ยเป็นไปตามสูตร $v = \dfrac{d_1 + d_2 + d_3}{\dfrac{d_1}{v_1} + \dfrac{d_2}{v_2} + \dfrac{d_3}{v_3}}$
   และถ้า $d_1 = 2500$, $d_2 = 1200$, $d_3 = 500$, $v_1 = 500$, $v_2 = 400$ และ $v_3 = 250$ จงหา v

5. จงพิจารณาว่าข้อความที่กำหนดให้เป็นจริงหรือเป็นเท็จ ถ้าเป็นเท็จจงบอกเหตุผล
   1) ค่าเฉลี่ยเลขคณิตของจำนวนนักเรียนในแต่ละห้องของโรงเรียนแห่งหนึ่ง เท่ากับ 35 คน หมายความว่ามีครึ่งหนึ่งของจำนวนห้องเรียนทั้งหมดที่มีจำนวนนักเรียนต่ำกว่า 35 คน และอีกครึ่งหนึ่งมีจำนวนนักเรียนมากกว่า 35 คน
   2) ในกรณีที่ค่าจากการสังเกตค่าหนึ่งค่าใดสูงหรือต่ำกว่าค่าจากการสังเกตอื่น ๆ ในข้อมูลชุดเดียวกันมาก มัธยฐานย่อมใช้เป็นตัวกลางของข้อมูลชุดนั้นได้ดีกว่าค่าเฉลี่ยเลขคณิต
   3) ค่าเฉลี่ยของข้อมูลชุดใด ๆ จะต้องอยู่ระหว่างค่าจากการสังเกตน้อยที่สุดและค่าจากการสังเกตมากที่สุดของข้อมูลชุดนั้นเสมอ

4) ฐานนิยมของข้อมูลชุดใด ๆ จะมีจำนวนมากกว่าหนึ่งค่าไม่ได้

5) ค่าเฉลี่ยเลขคณิตของข้อมูลชุดหนึ่งจะมีค่ามากกว่ามัธยฐานและฐานนิยมของข้อมูลชุดนั้นเสมอ

6. จากข้อมูลในตารางเป็นข้อมูลตัวอย่าง จงหาค่าเฉลี่ยเลขคณิตโดยประมาณของมูลค่าความเสียหายจากธรณีพิบัติภัยในปี 2547 จำแนกตามประเภทของความเสียหายพร้อมทั้งหาค่าเฉลี่ยเลขคณิตของความเสียหายของแต่ละจังหวัดและจังหวัดใดมีค่าเฉลี่ยของความเสียหายสูงที่สุด

จังหวัด	รวม	ความเสียหาย (ล้านบาท)				
		ที่ดิน	บ้าน/อาคาร สิ่งปลูกสร้าง	อุปกรณ์	ยานพาหนะ	อื่น ๆ
รวม	**1,942.8**	**262.5**	**772.8**	**372.8**	**304.0**	**230.7**
กระบี่	321.3	17.1	99.5	71.9	64.8	68.0
พังงา	1,077.4	122.6	557.3	154.0	161.0	82.5
ระนอง	203.3	102.1	26.2	36.8	30.6	7.6
ตรัง	43.0	10.1	5.9	14.4	9.9	2.7
ภูเก็ต	188.6	1.1	83.3	41.4	29.6	33.2
สตูล	109.2	9.5	0.6	54.3	8.1	36.7

ที่มา : สารสถิติปีที่ 16 ฉบับที่ 3 เดือนมีนาคม พ.ศ. 2548

7. จากข้อ 6 จะสามารถหาค่ากลางโดยใช้มัธยฐาน ฐานนิยม หรือค่ากึ่งกลางพิสัยได้หรือไม่ ถ้าได้ จงหาค่ากลางเหล่านั้นและเปรียบเทียบผลลัพธ์กับข้อ 6 ถ้าไม่ได้ จงให้เหตุผล

แบบฝึกหัดระคน

1. พิจารณาสมบัติของค่าเฉลี่ยเลขคณิตและสมบัติของมัธยฐานแล้ว อธิบายว่าความแตกต่างระหว่างการใช้ค่าเฉลี่ยเลขคณิตและมัธยฐานคืออะไร

2. ถ้าข้อมูลมีการแจกแจงแบบสมมาตร (symmetrical) หรือมีการแจกแจงแบบเบ้ (skewed) แล้วจงอธิบายว่าค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมจะแตกต่างกันอย่างไร

3. ในหน่วยงานของบริษัทแห่งหนึ่ง มีพนักงาน 12 คน และหัวหน้างาน 1 คน โดยพนักงานแต่ละคนบริจาคเงินเพื่อการกุศล ดังนี้

   20, 40, 40, 60, 30, 15, 20, 25, 50, 30, 45, 60

   ส่วนหัวหน้างานบริจาคเงิน 500 บาท จงหาค่าเฉลี่ยเลขคณิตและมัธยฐานของเงินบริจาคของคนในหน่วยงานนี้ และอธิบายว่าควรเลือกค่ากลางชนิดใดสำหรับใช้เป็นตัวแทนที่ดีสำหรับเงินบริจาคของคนกลุ่มนี้

4. ในการทดสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่งซึ่งมีนักเรียน 42 คน พบว่าคะแนนสอบของนักเรียนมีค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมเป็น 32, 30 และ 28 คะแนนตามลำดับ ถ้ามีนักเรียนลาออกสองคน ซึ่งนักเรียนสองคนนี้สอบได้ 36 และ 28 คะแนน แล้วค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของคะแนนสอบของนักเรียนที่เหลือเปลี่ยนแปลงหรือไม่อย่างไร

5. จงหาค่าเฉลี่ยเลขคณิตและมัธยฐานของข้อมูลในแต่ละข้อต่อไปนี้

1) 1 2 3	2) 1 2 6
3) 1 2 9	4) 1 2 297
5) 1 2 3 4	6) 1 2 3 4 5
7) 1 2 3 4 5 6	8) 1 2 3 4 5 6 ... 98 99

6. จงเปรียบเทียบผลลัพธ์ที่ได้จากแบบฝึกหัดข้อ 5 และบอกเหตุผลจากการสังเกตผลลัพธ์ที่ได้จากข้อย่อย 4) และข้อย่อย 8)

7. จากข้อมูลในตารางเป็นร้อยละของเมทิลแอลกอฮอล์จำแนกตามห้องปฏิบัติการ 4 แห่ง จงตรวจสอบว่า ข้อมูลดังกล่าวเป็นข้อมูลเชิงปริมาณหรือไม่ ควรใช้ค่ากลางชนิดใดเพื่อเป็นตัวแทนของแต่ละห้องปฏิบัติการ และเปรียบเทียบค่ากลางนั้นจากแต่ละห้องปฏิบัติการ

หน่วยทดลอง	ห้องปฏิบัติการ			
	LAB1	LAB2	LAB3	LAB4
1	85.06	84.99	84.48	84.10
2	85.25	84.28	84.72	84.55
3	84.87	84.88	85.10	84.05

ที่มา : Devore and Farnum (2005)

1.2 การวัดตำแหน่งที่หรือตำแหน่งสัมพัทธ์ของข้อมูล (measures of relative standing)

โดยทั่ว ๆ ไป เมื่อกล่าวถึงตำแหน่งหรือลำดับที่ของข้อมูลชุดหนึ่ง เช่น ในการสรุปเหรียญรางวัลจากการแข่งขันกีฬาโอลิมปิก มักจะสรุปโดยเรียงลำดับที่โดยพิจารณาจากประเภทและจำนวนเหรียญรางวัลซึ่งแต่ละประเทศได้รับ ในการกล่าวถึงตำแหน่งที่ของประเทศใดประเทศหนึ่งโดยไม่มีข้อมูลอื่นประกอบ ผู้รับข่าวสารจะไม่สามารถทราบได้ว่าตำแหน่งที่ที่กล่าวถึงนั้นอยู่ตรงส่วนไหนของข้อมูล เช่นในปี พ.ศ. 2547 ประเทศไทยได้ส่งนักกีฬาเข้าร่วมการแข่งขันกีฬาโอลิมปิก ครั้งที่ 28 ณ กรุงเอเธนส์ ประเทศกรีซ ระหว่างวันที่ 13 – 29 สิงหาคม 2547 ผลปรากฏว่าประเทศไทยได้รับรางวัล 3 เหรียญทอง 1 เหรียญเงิน และ 4 เหรียญทองแดง และเป็นอันดับที่ 25 ถ้าไม่ทราบว่ามีประเทศที่เข้าร่วมการแข่งขันครั้งนี้กี่ประเทศก็จะไม่ทราบว่าประเทศไทยอยู่ในตำแหน่งที่ดีหรือไม่ แต่ถ้ามีข้อมูลเพิ่มเติมว่าในการแข่งขันครั้งนี้มีประเทศที่เข้าร่วมการแข่งขันทั้งสิ้น 202 ประเทศ จะสรุปได้ว่าทีมจากประเทศไทยได้ตำแหน่งดี

นอกจากนี้เพื่อเป็นการช่วยให้การกล่าวถึงตำแหน่งที่เป็นไปอย่างมีความหมาย กล่าวคือ สามารถบอกได้ทันทีว่าตำแหน่งนั้นดีหรือไม่เพียงไรในกลุ่ม จึงได้มีการพัฒนาวิธีการบอกตำแหน่ง โดยใช้เปอร์เซ็นไทล์ เดไซล์ และควอร์ไทล์ ดังจะได้กล่าวต่อไป

ควอร์ไทล์ เดไซล์ และเปอร์เซ็นไทล์ (quartiles, deciles, and percentiles)

เมื่อนำชุดข้อมูลชุดหนึ่งมาเรียงค่าของข้อมูลจากน้อยไปหามาก ค่าที่ตรงกับจุด 3 จุด ที่แบ่งข้อมูลออกเป็น 4 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่า ๆ กัน เรียกว่า ควอร์ไทล์ที่หนึ่ง (Q_1) ควอร์ไทล์ที่สอง (Q_2) และควอร์ไทล์ที่สาม (Q_3) ตามลำดับ

ดังนั้น ควอร์ไทล์ที่หนึ่ง เป็นค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณหนึ่งในสี่ของข้อมูลทั้งหมด หรือร้อยละ 25

ควอร์ไทล์ที่สอง เป็นค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณหนึ่งในสองของข้อมูลทั้งหมด หรือร้อยละ 50 หรือกล่าวได้ว่าเป็นค่าที่อยู่ตรงกึ่งกลางข้อมูลที่เรียงแล้ว ดังนั้น ควอร์ไทล์ที่สองคือมัธยฐานนั่นเอง หรือค่าที่แบ่งข้อมูลที่มีค่าน้อยกว่า Q_2 และข้อมูลที่มีค่ามากกว่า Q_2 ให้มีจำนวนเท่ากัน

ควอร์ไทล์ที่สาม เป็นค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณสามในสี่ของจำนวนข้อมูลทั้งหมด หรือร้อยละ 75

ความเข้าใจในเรื่องควอร์ไทล์ตามที่กล่าวมานี้อาจนำไปใช้กับเดไซล์ซึ่งเป็นค่าของข้อมูล ณ จุด 9 จุด ที่แบ่งข้อมูลซึ่งเรียงจากน้อยไปหามากออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่า ๆ กัน เรียกค่าที่เรียงลำดับจากค่าน้อยไปหาค่ามากว่า เดไซล์ที่หนึ่ง (D_1) เดไซล์ที่สอง (D_2) เดไซล์ที่สาม (D_3) ... เดไซล์ที่เก้า (D_9) ดังนั้น เดไซล์ที่เจ็ด (D_7) คือค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณเจ็ดในสิบของจำนวนข้อมูลทั้งหมด

ในทำนองเดียวกัน ค่าของข้อมูล ณ จุด 99 จุด ที่แบ่งข้อมูลซึ่งเรียงจากน้อยไปหามากออกเป็น 100 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่า ๆ กัน เรียกว่าเปอร์เซ็นไทล์ ดังนั้น เปอร์เซ็นไทล์ที่ 65 (P_{65}) คือค่าที่มีข้อมูลน้อยกว่าค่านี้อยู่ประมาณหกสิบห้าในร้อยของจำนวนข้อมูลทั้งหมด

การหาควอร์ไทล์ เดไซล์ และเปอร์เซ็นไทล์ของข้อมูลที่ไม่ได้แจกแจงความถี่

วิธีหาเหมือนกับการหามัธยฐานของข้อมูลที่ไม่ได้แจกแจงความถี่ กล่าวคือเรียงข้อมูลจากค่าที่น้อยที่สุดไปหาค่าที่มากที่สุด แล้วหาตำแหน่งที่ของควอร์ไทล์ เดไซล์ หรือเปอร์เซ็นไทล์ที่ต้องการ ถ้า N เป็นจำนวนข้อมูลทั้งหมดในประชากร หรือถ้าเป็นข้อมูลตัวอย่างให้ใช้ n แทน N ตำแหน่งต่าง ๆ จะหาได้ดังนี้

Q_1 อยู่ในตำแหน่งที่ $\frac{N+1}{4}$

Q_2 อยู่ในตำแหน่งที่ $\frac{2(N+1)}{4}$

Q_3 อยู่ในตำแหน่งที่ $\frac{3(N+1)}{4}$

D_1 อยู่ในตำแหน่งที่ $\frac{N+1}{10}$

D_2 อยู่ในตำแหน่งที่ $\frac{2(N+1)}{10}$

D_3 อยู่ในตำแหน่งที่ $\frac{3(N+1)}{10}$

$\vdots$ $\vdots$

D_9 อยู่ในตำแหน่งที่ $\frac{9(N+1)}{10}$

P_1 อยู่ในตำแหน่งที่ $\frac{N+1}{100}$

P_2 อยู่ในตำแหน่งที่ $\frac{2(N+1)}{100}$

P_3 อยู่ในตำแหน่งที่ $\frac{3(N+1)}{100}$

$\vdots$ $\vdots$

P_{99} อยู่ในตำแหน่งที่ $\frac{99(N+1)}{100}$

เมื่อหาตำแหน่งที่แล้ว หาค่าของข้อมูลที่ตรงกับตำแหน่งควอร์ไทล์ เดไซล์ หรือเปอร์เซ็นไทล์ที่ต้องการ โดยวิธีเดียวกับการหามัธยฐานของข้อมูลที่ไม่ได้แจกแจงความถี่

หมายเหตุ ในทางสถิติไม่นิยมหาเปอร์เซ็นไทล์ของข้อมูลที่มีจำนวนน้อย ในกรณีที่มีจำนวนข้อมูลน้อยค่าเปอร์เซ็นไทล์จะเป็นค่าประมาณที่มีความถูกต้องน้อย เช่น P_{30} หรือ P_{90} ของข้อมูลซึ่งประกอบด้วยค่าต่าง ๆ เพียง 6 ค่า ได้แก่ 8, 17, 23, 42, 61, 63 ในกรณีที่หาค่าเปอร์เซ็นไทล์สูงหรือต่ำมาก เช่น P_1 หรือ P_{99} ของข้อมูลที่มีจำนวนน้อย ค่าเปอร์เซ็นไทล์ที่ได้อาจจะต่ำกว่าค่าต่ำสุดของข้อมูลหรือสูงกว่าค่าสูงสุดของข้อมูลที่มีอยู่ก็ได้

ตัวอย่างที่ 1 ปริมาณการส่งออกข้าวของประเทศไทยไปยังประเทศต่าง ๆ ในทวีปเอเชีย 10 ประเทศในปี พ.ศ. 2546 แสดงไว้ในตาราง

ประเทศ	ปริมาณข้าวส่งออก (พันตัน)
กัมพูชา	42
จีน	255
ญี่ปุ่น	156
ไต้หวัน	22
ฟิลิปปินส์	389
มาเลเซีย	316
สิงคโปร์	270
อินโดนีเซีย	764
บรูไน	20
เกาหลีเหนือ	95

ที่มา : สภาหอการค้าแห่งประเทศไทย สำนักงานมาตรฐานสินค้า

จงหา 1. ควอร์ไทล์ที่หนึ่งและควอร์ไทล์ที่สามของปริมาณการส่งออกข้าวของประเทศไทยไปยังประเทศต่าง ๆ ในทวีปเอเชียทั้ง 10 ประเทศ

2. ประเทศใดบ้างที่ประเทศไทยส่งออกข้าวไปในปริมาณที่ต่ำกว่าควอร์ไทล์ที่หนึ่งที่หาได้ในข้อ 1

3. ประเทศใดบ้างที่ประเทศไทยส่งออกข้าวไปในปริมาณที่สูงกว่าควอร์ไทล์ที่สามที่หาได้ในข้อ 1

วิธีทำ จากตาราง ปริมาณการส่งออกข้าวของประเทศไทยไปยัง 10 ประเทศ เมื่อนำปริมาณการส่งออกข้าวของประเทศไทยไปยัง 10 ประเทศมาเรียงจากน้อยไปหามากจะได้ดังนี้

20 22 42 95 156 255 270 316 389 และ 764

1. เนื่องจาก Q_1 อยู่ในตำแหน่งที่ $\frac{10+1}{4} = 2.75$

   ดังนั้น Q_1 มีค่าอยู่ระหว่าง 22 กับ 42

   ตำแหน่งต่างกัน $3-2=1$ ปริมาณการส่งออกข้าวเพิ่มขึ้น $42-22 = 20$ พันตัน

   ตำแหน่งต่างกัน $2.75-2 = 0.75$

   ปริมาณการส่งออกข้าวเพิ่มขึ้น $\frac{0.75 \times 20}{1} = 15$ พันตัน

   นั่นคือ $Q_1 = 22+15 = 37$ พันตัน หรือ 37,000 ตัน

   และเนื่องจาก Q_3 อยู่ในตำแหน่งที่ $\frac{3(10+1)}{4} = 8.25$

   ดังนั้น Q_3 มีค่าอยู่ระหว่าง 316 กับ 389

   ตำแหน่งต่างกัน $9-8=1$ ปริมาณการส่งออกข้าวเพิ่มขึ้น $389-316=73$ พันตัน

   ตำแหน่งต่างกัน $8.25-8 = 0.25$

   ปริมาณการส่งออกข้าวเพิ่มขึ้น $\frac{0.25 \times 73}{1} = 18.25$ พันตัน

   นั่นคือ $Q_3 = 316+18.25 = 334.25$ พันตัน หรือ 334,250 ตัน

2. ประเทศที่ประเทศไทยส่งออกข้าวไปในปริมาณที่ต่ำกว่าควอร์ไทล์ที่หนึ่งของทั้ง 10 ประเทศก็คือ ประเทศที่ประเทศไทยส่งออกข้าวไปในปริมาณที่ต่ำกว่า 37 พันตัน หรือต่ำกว่า 37,000 ตัน ซึ่งได้แก่ ประเทศไต้หวัน และประเทศบรูไน

3. สำหรับประเทศที่ประเทศไทยส่งออกข้าวไปในปริมาณที่สูงกว่าควอร์ไทล์ที่สามของทั้ง 10 ประเทศก็คือ ประเทศที่ประเทศไทยส่งออกข้าวไปในปริมาณที่สูงกว่า 334.25 พันตัน หรือสูงกว่า 334,250 ตัน ซึ่งได้แก่ ประเทศฟิลิปปินส์ และอินโดนีเซีย

การหาควอร์ไทล์ เดไซล์ และเปอร์เซ็นไทล์ของข้อมูลที่แจกแจงความถี่

การหาควอร์ไทล์ เดไซล์ และเปอร์เซ็นไทล์ของข้อมูลที่แจกแจงความถี่ คล้ายกับการหามัธยฐานของข้อมูลที่แจกแจงความถี่แล้ว ซึ่งข้อมูลที่นำมาใช้คำนวณไม่ได้เป็นข้อมูลที่แน่นอน แต่เป็นข้อมูลที่เป็นอันตรภาคชั้น ดังนั้น ในการคำนวณหาควอร์ไทล์ เดไซล์ และเปอร์เซ็นไทล์ของข้อมูลที่แจกแจงความถี่จึงได้ค่าโดยประมาณเท่านั้น กล่าวคือหาตำแหน่งที่ของควอร์ไทล์ เดไซล์ หรือเปอร์เซ็นไทล์ของข้อมูลจำนวน N หน่วยที่ต้องการโดยใช้สูตรซึ่งสรุปได้ดังนี้

ตำแหน่งที่ของ Q_r คือ $\frac{rN}{4}$ เมื่อ $r \in \{1, 2, 3\}$

ตำแหน่งที่ของ D_r คือ $\frac{rN}{10}$ เมื่อ $r \in \{1, 2, 3, ..., 9\}$

ตำแหน่งที่ของ P_r คือ $\frac{rN}{100}$ เมื่อ $r \in \{1, 2, 3, ..., 99\}$

เมื่อหาตำแหน่งที่ได้แล้ว คำนวณหาค่าของข้อมูลที่ตรงกับตำแหน่งที่ควอร์ไทล์ เดไซล์ หรือเปอร์เซ็นไทล์ที่ต้องการ โดยใช้วิธีคำนวณ หรือหาจากเส้นโค้งความถี่สะสม (ogive)

ตัวอย่างที่ 2 จากตารางแจกแจงความถี่แสดงผลผลิตข้าวเฉลี่ยต่อไร่ (กิโลกรัม) จาก 50 จังหวัดของประเทศไทย ในปีเพาะปลูก 2544/45 ดังตาราง

ผลผลิตข้าวเฉลี่ยต่อไร่ (กิโลกรัม)	ความถี่ (จำนวนจังหวัด)
258 – 312	5
313 – 367	8
368 – 422	1
423 – 477	6
478 – 532	10
533 – 587	7
588 – 642	3
643 – 697	5
698 – 752	4
753 – 807	1

1. ผลผลิตข้าวเฉลี่ยต่อไร่ต้องเป็นเท่าไรจึงต่ำกว่าผลผลิตข้าวเฉลี่ยต่อไร่ของประมาณหนึ่งในสี่ของทั้ง 50 จังหวัด
2. ผลผลิตข้าวเฉลี่ยต่อไร่ต้องเป็นเท่าไรจึงสูงกว่าผลผลิตข้าวเฉลี่ยต่อไร่ของประมาณสามในสี่ของทั้ง 50 จังหวัด

วิธีทำ จากตารางแจกแจงความถี่แสดงผลผลิตข้าวเฉลี่ยต่อไร่จาก 50 จังหวัดของประเทศไทยในปีเพาะปลูก 2544/45 ข้างต้น นำมาสร้างตารางความถี่สะสมได้ดังนี้

ผลผลิตข้าวเฉลี่ยต่อไร่ (กิโลกรัม)	**ความถี่ (จำนวนจังหวัด)**	**ความถี่สะสม**
258 – 312	5	5
313 – 367	8	13
368 – 422	1	14
423 – 477	6	20
478 – 532	10	30
533 – 587	7	37
588 – 642	3	40
643 – 697	5	45
698 – 752	4	49
753 – 807	1	50

1. ผลผลิตข้าวเฉลี่ยต่อไร่ที่ต่ำกว่าผลผลิตข้าวเฉลี่ยต่อไร่ของประมาณหนึ่งในสี่ของทั้ง 50 จังหวัด คือ ผลผลิตข้าวเฉลี่ยต่อไร่ต่ำกว่า Q_1

   เนื่องจากตำแหน่งที่ของ Q_1 คือ $\frac{50}{4} = 12.5$ เป็นตำแหน่งที่อยู่ระหว่างความถี่สะสม 5 และ 13 ในอันตรภาคชั้น 258 – 312 และ 313 – 367 ตามลำดับ

   ความถี่สะสมต่างกัน 8 ผลผลิตข้าวเฉลี่ยต่อไร่เพิ่มขึ้น 55 กิโลกรัม

ความถี่สะสมต่างกัน 7.5 ผลผลิตข้าวเฉลี่ยต่อไร่เพิ่มขึ้น $\frac{7.5 \times 55}{8} \approx 51.56$ กิโลกรัม

ดังนั้น $Q_1 = 312.50 + 51.56 = 364.06$ กิโลกรัม

นั่นคือ ผลผลิตข้าวเฉลี่ยต่อไร่ที่ต่ำกว่าผลผลิตข้าวเฉลี่ยต่อไร่ของประมาณหนึ่งในสี่ของทั้ง 50 จังหวัด คือ ผลผลิตข้าวโดยเฉลี่ยที่ต่ำกว่า 364.06 กิโลกรัมต่อไร่

2. ผลผลิตข้าวเฉลี่ยต่อไร่ที่สูงกว่าผลผลิตข้าวเฉลี่ยต่อไร่ของประมาณสามในสี่ของทั้ง 50 จังหวัด คือผลผลิตข้าวเฉลี่ยต่อไร่สูงกว่า Q_3

เนื่องจากตำแหน่งที่ของ Q_3 คือ $\frac{3 \times 50}{4} = 37.5$ เป็นตำแหน่งที่อยู่ระหว่างความถี่สะสม 37 และ 40 ในอันตรภาคชั้น 533 – 587 และ 588 – 642 ตามลำดับ

ความถี่สะสมต่างกัน 3 ผลผลิตข้าวเฉลี่ยต่อไร่เพิ่มขึ้น 55 กิโลกรัม

ความถี่สะสมต่างกัน 0.5 ผลผลิตข้าวเฉลี่ยต่อไร่เพิ่มขึ้น $\frac{0.5 \times 55}{3} \approx 9.17$ กิโลกรัม

นั่นคือ $Q_3 = 587.50 + 9.17 = 596.67$ กิโลกรัม

ดังนั้น ผลผลิตข้าวเฉลี่ยต่อไร่ที่สูงกว่าผลผลิตข้าวเฉลี่ยต่อไร่ของประมาณสามในสี่ของทั้ง 50 จังหวัด คือ ผลผลิตข้าวโดยเฉลี่ยสูงกว่า 596.67 กิโลกรัมต่อไร่

การหาควอร์ไทล์ เดไซล์ และเปอร์เซ็นไทล์จากกราฟ

การหาควอร์ไทล์ เดไซล์ และเปอร์เซ็นไทล์ นอกจากจะทำได้โดยวิธีการคำนวณตามที่ได้กล่าวมาแล้ว ยังสามารถทำได้โดยอาศัยกราฟที่ได้จากข้อมูลที่มีการแจกแจงความถี่ ซึ่งเป็นค่าประมาณเช่นเดียวกับการคำนวณจากข้อมูลที่มีการแจกแจงความถี่แล้ว โดยการเขียนเส้นโค้งความถี่สะสมที่มีแกนตั้งแทนความถี่สะสม และแกนนอนแทนข้อมูล เช่น ถ้าต้องการหาค่า P_r ในขั้นตอนแรกหาจุดบนแกนตั้งที่แทน $\frac{rN}{100}$ แล้วลากส่วนของเส้นตรงขนานกับแกนนอนให้ตัดกับเส้นโค้งความถี่สะสมจากจุดตัดนี้ ลากส่วนของเส้นตรงขนานกับแกนตั้งตัดแกนนอนที่จุดใด ค่าที่อ่านจากแกนนอนที่จุดนั้น คือ P_r ดังจะแสดงในตัวอย่างที่ 3

ถ้าต้องการหาตำแหน่งเปอร์เซ็นไทล์ของค่าใด ๆ ก็ใช้วิธีการกลับกัน กล่าวคือจากค่าที่ต้องการหาตำแหน่งเปอร์เซ็นไทล์บนแกนนอนลากส่วนของเส้นตรงให้ขนานกับแกนตั้งตัดเส้นโค้งความถี่สะสม จากจุดตัดลากส่วนของเส้นตรงให้ขนานกับแกนนอนตัดแกนตั้งที่จุดใด จุดนั้นคือค่า

ของ $\frac{rN}{100}$ แล้วนำไปคำนวณหาค่า r โดยแทนค่า N เพื่อความสะดวกในการอ่านค่าควรหาร้อยละของความถี่สะสมสัมพัทธ์ แล้วเขียนแกนตั้งใหม่ให้ขนานกับแกนตั้งเดิมแสดงร้อยละของความถี่สะสมสัมพัทธ์ ดังรูปในตัวอย่างที่ 3 ค่าร้อยละของความถี่สะสมสัมพัทธ์จะแสดงค่า r ของ P_r จะทำให้สามารถอ่านค่า r จากแกนตั้งใหม่ได้ทันที

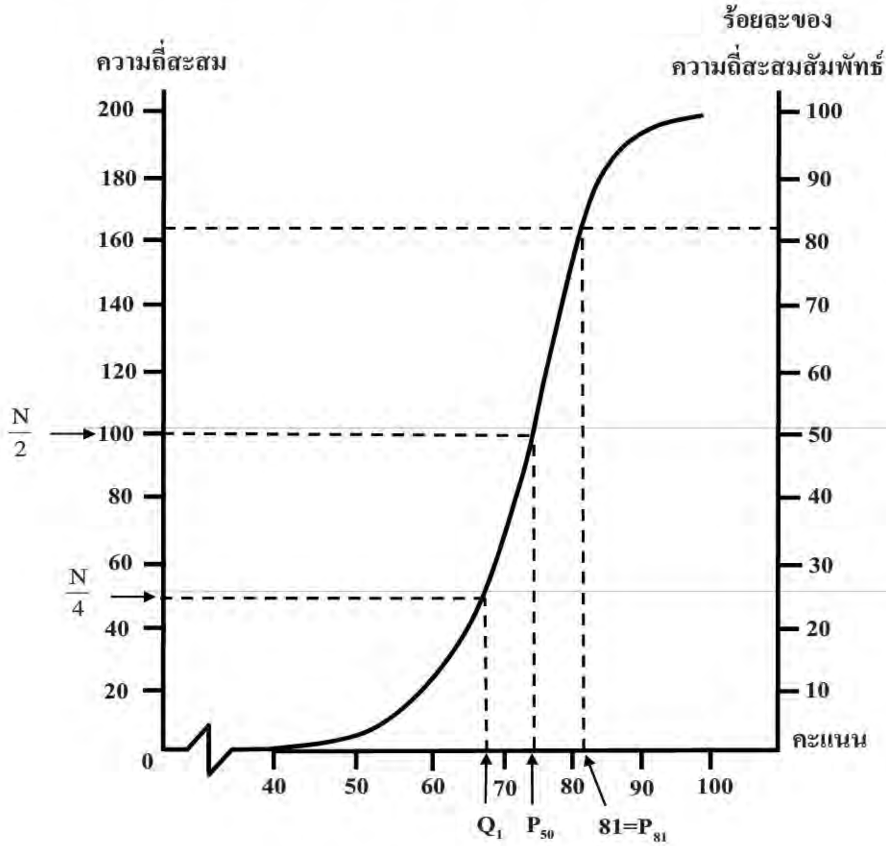
ตัวอย่างที่ 3 ตารางต่อไปนี้ แสดงความถี่ ความถี่สะสม และร้อยละของความถี่สะสมสัมพัทธ์ของคะแนนสอบของนักเรียน 200 คน

คะแนนสอบ	ความถี่	ความถี่สะสม	ร้อยละของความถี่สะสมสัมพัทธ์
95 – 99	1	200	100.0
90 – 94	6	199	99.5
85 – 89	8	193	96.5
80 – 84	33	185	92.5
75 – 79	40	152	76.0
70 – 74	50	112	56.0
65 – 69	26	62	31.0
60 – 64	14	36	18.0
55 – 59	10	22	11.0
50 – 54	6	12	6.0
45 – 49	4	6	3.0
40 – 44	2	2	1.0
รวม	**200**		

1) จงหาควอร์ไทล์ที่ 1 และเปอร์เซ็นไทล์ที่ 50
2) นักเรียนที่สอบได้น้อยกว่า 81 คะแนนมีประมาณกี่เปอร์เซ็นต์ของนักเรียนทั้งหมด
3) จักรพันธ์เป็นผู้ที่สอบได้คะแนนสูงกว่านักเรียนคนอื่น ๆ 100 คน จักรพันธ์สอบได้กี่คะแนน

4) มณีนุชเป็นผู้ที่สอบได้คะแนนสูงกว่านักเรียนคนอื่น ๆ เพียง 50 คน มณีนุชสอบได้กี่คะแนน

วิธีทำ จากตารางเขียนเส้นโค้งความถี่สะสมได้ดังรูป



1) จากกราฟ Q_1 มีค่าประมาณ 67 คะแนน

P_{50} มีค่าประมาณ 73 คะแนน

2) คะแนน 81 เป็นค่าประมาณของ P_{81}

เนื่องจาก $P_{81} \approx 81$ ดังนั้นนักเรียนที่สอบได้น้อยกว่า 81 คะแนน มีประมาณ 81% ของนักเรียนทั้งหมด

3) คะแนนที่จักรพันธ์สอบได้ P_{50} ซึ่งคิดเป็นคะแนน 73 คะแนน เป็นคะแนนที่จักรพันธ์สอบได้คะแนนสูงกว่านักเรียนคนอื่น ๆ อยู่ 100 คน

4) คะแนนที่มณีนุชสอบได้ คือ Q_1 ซึ่งคิดเป็นคะแนน 67 คะแนน เป็นคะแนนที่มณีนุชสอบได้คะแนนสูงกว่านักเรียนคนอื่น ๆ เพียง 50 คน

ข้อสังเกต ~~า ยุ หา มเ~~

1. ค่าที่ได้จากการใช้กราฟอาจมีความคลาดเคลื่อนไปจากที่ควรจะเป็นจริง ถ้าความละเอียดในการเขียนกราฟมีไม่มากพอ
2. การหาเปอร์เซ็นไทล์ของข้อมูลที่มีจำนวนน้อยเกินไปจะทำให้ค่าเปอร์เซ็นไทล์ที่คำนวณได้มีความเชื่อถือได้น้อย เนื่องจากสูตรหรือวิธีการที่ใช้คำนวณจะใช้สำหรับประมาณค่าเปอร์เซ็นไทล์ของข้อมูลที่มีจำนวนมากพอ
3. ค่าวัดตำแหน่งของควอร์ไทล์ เดไซล์ หรือเปอร์เซ็นไทล์ที่คำนวณได้เป็นค่าโดยประมาณ วิธีการที่ใช้คำนวณหาค่าวัดตำแหน่งของควอร์ไทล์ เดไซล์ หรือเปอร์เซ็นไทล์ อาจทำได้หลายวิธี แต่ค่าวัดตำแหน่งที่คำนวณได้จะใกล้เคียงกัน ถ้าจำนวนข้อมูลที่นำมาใช้มีมากพอ
4. ในทางปฏิบัติ เมื่อมีข้อมูลคะแนนของนักเรียนแต่ละคนอาจไม่จำเป็นต้องแบ่งอันตรภาคชั้นของคะแนน เนื่องจากสามารถใช้เครื่องคอมพิวเตอร์หาควอร์ไทล์ เดไซล์ หรือเปอร์เซ็นไทล์ของข้อมูลได้โดยตรงและสะดวก ดังนั้นวิธีคำนวณที่อาศัยอันตรภาคชั้นนั้นเป็นเพียงทางเลือกที่อาจทำได้ เพราะทำให้เห็นภาพรวมอยู่ในตาราง
5. นอกจากวิธีการวัดตำแหน่งที่หรือการวัดตำแหน่งสัมพัทธ์ดังกล่าวข้างต้น ยังมีอีกวิธีหนึ่งที่นิยมใช้มาก คือ คะแนนมาตรฐาน (standard score) ซึ่งรายละเอียดของคะแนนมาตรฐานจะแสดงไว้ในบทที่ 2 ต่อไป

แบบฝึกหัด 1.2

1. ผลการทดสอบความสามารถในการคิดอย่างมีวิจารณญาณของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 40 คนของโรงเรียนแห่งนี้ เป็นดังนี้ (คะแนนเต็ม 100 คะแนน)

96	78	80	76	84	77	74	85	65	69
82	53	45	67	58	54	56	62	56	54
43	48	49	50	60	65	54	51	55	60
65	66	75	98	97	63	92	94	76	78

1) นักเรียนจะต้องสอบได้กี่คะแนน จึงจะมีนักเรียนประมาณครึ่งหนึ่งของชั้นได้คะแนนต่ำกว่า
2) นักเรียนจะต้องสอบได้กี่คะแนน จึงจะมีนักเรียนประมาณหนึ่งในสี่ของชั้นได้คะแนนสูงกว่า
3) นักเรียนจะต้องสอบได้กี่คะแนน จึงจะมีผู้ที่สอบได้คะแนนน้อยกว่าอยู่หกในสิบ

2. การสอบแข่งขันคณิตคิดเร็วในสัปดาห์วิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 ของโรงเรียนแห่งหนึ่ง โดยใช้ข้อสอบในการแข่งขัน 30 ข้อ มีจำนวนนักเรียนเข้าแข่งขัน 30 คน เวลา (นาที) ที่ใช้ในการทำข้อสอบของนักเรียนที่เข้าแข่งขันมีดังนี้

70	50	45	60	55	40	43	49	52	51
65	75	80	72	73	44	62	58	53	61
30	35	42	39	48	46	57	63	58	69

1) ถ้ามีจำนวนนักเรียนที่เข้าแข่งขันที่ใช้เวลาในการทำข้อสอบน้อยกว่าสมชายและสมศักดิ์ประมาณร้อยละ 55 และ 68 ตามลำดับ สมชายและสมศักดิ์ใช้เวลาเท่าไรในการทำข้อสอบ

2) จงหาเวลาที่ดวงจันทร์ใช้ในการทำข้อสอบเมื่อทราบว่ามีนักเรียนที่เข้าแข่งขันประมาณแปดในสิบที่ใช้เวลาในการทำข้อสอบน้อยกว่าดวงจันทร์

3) ถ้านักเรียนประมาณหนึ่งในสี่ใช้เวลาในการทำข้อสอบน้อยกว่านักเรียนที่เข้าแข่งขันทั้งหมดจะได้รับรางวัลเป็นกล่องดินสอ อยากทราบว่านักเรียนคนที่ได้รับรางวัลใช้เวลาในการทำข้อสอบมากที่สุดเท่าไร

3. ตารางแสดงจำนวนนักเรียนจำแนกตามช่วงคะแนนที่สอบได้เป็นดังนี้

ช่วงคะแนน	**ความถี่ (จำนวนนักเรียน)**
55 – 64	3
65 – 74	21
75 – 84	78
85 – 94	182
95 – 104	305
105 – 114	209
115 – 124	81
125 – 134	21
135 – 144	5
รวม	**905**

1) จงหา Q_2, D_5 และ P_{50} แล้วเปรียบเทียบค่าที่ได้

2) จงหา $Q_1, (D_1 + D_3)$ และ P_{25} แล้วเปรียบเทียบค่าที่ได้

4. นักเรียนชั้นมัธยมศึกษาปีที่ 5 สองห้องทำข้อสอบวิชาคณิตศาสตร์ฉบับเดียวกันได้ผลดังตาราง

ช่วงคะแนน	จำนวนนักเรียน (คน)	
	ห้อง ก.	ห้อง ข.
71 – 75	1	1
66 – 70	0	0
61 – 65	2	3
56 – 60	4	3
51 – 55	3	4
46 – 50	7	6
41 – 45	6	5
36 – 40	4	5
31 – 35	5	4
26 – 30	2	3
21 – 25	2	0
16 – 20	3	4
11 – 15	0	1
6 – 10	1	0
1 – 5	0	1

1) จงหาเปอร์เซ็นไทล์ที่ 25 ของคะแนนสอบของนักเรียนแต่ละห้อง และหาเปอร์เซ็นไทล์ที่ 50 ของคะแนนสอบของนักเรียนทั้งหมด แล้วเปรียบเทียบคะแนนที่ได้
2) นักเรียนคนหนึ่งในห้อง ก. สอบได้คะแนนเท่ากับ Q_3 ถ้าเขาอยู่ห้อง ข. เขาจะสอบได้คะแนนสูงกว่านักเรียนห้อง ข. มากกว่าครึ่งห้องหรือไม่

5. จากตารางแจกแจงความถี่ต่อไปนี้

ช่วงคะแนน	46 – 55	56 – 65	66 – 75	76 – 85	86 – 95	96 – 105
ความถี่	3	4	8	9	4	2

1) จงหา Q_1, Q_3, D_2 และ D_9
2) จงเปรียบเทียบ Q_2 และ $\frac{1}{2}(Q_1 + Q_3)$

6. ตารางต่อไปนี้เป็นคะแนนสอบวิชาคณิตศาสตร์ของนักเรียน 120 คน

คะแนน	จำนวนนักเรียน (คน)
30 – 39	1
40 – 49	4
50 – 59	10
60 – 69	22
70 – 79	45
80 – 89	30
90 – 99	8

1) จงหาคะแนนต่ำสุดของกลุ่มนักเรียนที่ได้คะแนนสูงสุด ซึ่งนักเรียนกลุ่มนี้คิดเป็น 20% ของนักเรียนทั้งชั้น
2) จงหาคะแนนสูงสุดของกลุ่มนักเรียนที่ได้คะแนนต่ำสุด ซึ่งนักเรียนกลุ่มนี้คิดเป็น 15% ของนักเรียนทั้งชั้น
3) นักเรียนคนหนึ่งสอบได้ 75 คะแนน เขาได้คะแนนเป็นเปอร์เซ็นไทล์ที่เท่าใด

1.3 การวัดการกระจายของข้อมูล (measures of dispersion)

จากความหมาย การคำนวณ และการใช้ค่ากลางชนิดต่าง ๆ ถ้าพิจารณาให้ละเอียดจะเห็นว่า การทราบแต่เพียงค่ากลางของข้อมูลไม่เพียงพอที่จะอธิบายการแจกแจงของข้อมูลชุดนั้น ค่ากลางแต่ละชนิดมิได้บอกให้ทราบว่า ค่าจากการสังเกตทั้งหลายในข้อมูลชุดนั้นต่างจากค่ากลางมากน้อยเพียงใด และค่าส่วนใหญ่อยู่รวมกลุ่มกันหรือกระจายออกไป สมมุติว่าคะแนนสอบวิชาหนึ่งของนักเรียน 2 ห้อง ซึ่งใช้ข้อสอบชุดเดียวกันมีค่าเฉลี่ยเลขคณิตเท่ากัน คือ 67 แต่ห้องแรกมีคะแนนสูงสุด 72 และคะแนนต่ำสุด 62 ส่วนห้องหลังมีคะแนนสูงสุด 97 และคะแนนต่ำสุด 25 จะเห็นว่าคะแนนสูงสุดกับคะแนนต่ำสุดของห้องแรกต่างกันเพียง 10 คะแนน แต่ห้องหลังคะแนนต่างกันถึง 72 คะแนน แสดงว่าห้องหลังนี้มีการกระจายของคะแนนมากกว่าห้องแรกมาก ซึ่งอาจกล่าวได้ว่านักเรียนห้องแรกส่วนใหญ่สอบได้คะแนนใกล้เคียงกัน แต่นักเรียนห้องหลังสอบได้คะแนนแตกต่างกันมาก เพื่อให้เห็นลักษณะของข้อมูลชัดเจนขึ้น จึงจำเป็นต้องทราบทั้งค่ากลางและค่าซึ่งแสดงการกระจายของข้อมูลด้วย

โดยทั่ว ๆ ไป การวัดการกระจายของข้อมูลแบ่งออกได้เป็น 2 วิธี คือ

1. การกระจายสัมบูรณ์ (absolute variation) คือ การวัดการกระจายของข้อมูลเพียงชุดเดียว เพื่อดูว่าในข้อมูลชุดนั้นแต่ละค่ามีความแตกต่างกันมากหรือน้อยเพียงไร การวัดการกระจายสัมบูรณ์ที่นิยมใช้กันอยู่มี 4 ชนิด คือ

   1) พิสัย (range)
   2) ส่วนเบี่ยงเบนควอร์ไทล์ หรือกึ่งช่วงควอร์ไทล์ (quartile deviation หรือ semi-interquartile range)
   3) ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation)
   4) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

2. การกระจายสัมพัทธ์ (relative variation) คือ การหาค่าเพื่อเปรียบเทียบการกระจายระหว่างข้อมูลมากกว่า 1 ชุด โดยใช้อัตราส่วน เช่น อัตราส่วนระหว่างค่าการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้น ๆ การวัดการกระจายสัมพัทธ์ของข้อมูลแต่ละชุด เพื่อนำไปใช้ในการเปรียบเทียบการกระจายของข้อมูลระหว่างชุดมีอยู่ 4 ชนิด คือ

1) สัมประสิทธิ์ของพิสัย (coefficient of range)

2) สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ (coefficient of quartile deviation)

3) สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (coefficient of average deviation)

4) สัมประสิทธิ์ของความแปรผัน (coefficient of variation)

การวัดการกระจายวิธีต่าง ๆ จะได้กล่าวโดยละเอียดในหัวข้อต่อไป

1.3.1 การวัดการกระจายสัมบูรณ์ (measures of absolute variation)

พิสัย (range)

พิสัย คือ ค่าที่ใช้วัดการกระจายที่ได้จากผลต่างระหว่างข้อมูลที่มีค่าสูงสุดและข้อมูลที่มีค่าต่ำสุด

ถ้า $x_1, x_2, x_3, \ldots, x_n$ เป็นค่าของข้อมูลชุดหนึ่ง พิสัยของข้อมูลชุดนี้เท่ากับ $x_{max} - x_{min}$
เมื่อ x_{max} และ x_{min} เป็นค่าสูงสุดและค่าต่ำสุดของข้อมูลชุดนี้ตามลำดับ

ตัวอย่างที่ 1 จงหาพิสัยของผลผลิตน้ำตาลของ 6 ประเทศที่ผลิตได้ในเดือนกันยายน พ.ศ. 2547 จากตารางต่อไปนี้

ประเทศ	จีน	สหรัฐอเมริกา	ไทย	อินเดีย	ออสเตรเลีย	บราซิล
ผลผลิต (ล้านตัน)	10.92	8.07	7.00	15.06	5.11	27.66

ที่มา : กรมการค้าต่างประเทศ

วิธีทำ จากตาราง ผลต่างของผลผลิตน้ำตาลมากที่สุดและน้อยที่สุดของทั้ง 6 ประเทศ คือ

$27.66 - 5.11 = 22.55$

ดังนั้น พิสัยของผลผลิตน้ำตาลของ 6 ประเทศที่ผลิตได้ในเดือนกันยายน พ.ศ. 2547 คือ 22.55 ล้านตัน

ถ้าเป็นข้อมูลที่แจกแจงความถี่โดยแบ่งเป็นอันตรภาคชั้นต่าง ๆ แล้ว

พิสัย คือ ผลต่างระหว่างขอบบนของอันตรภาคชั้นของข้อมูลที่มีค่าสูงสุดและขอบล่างของอันตรภาคชั้นของข้อมูลที่มีค่าต่ำสุด

ถ้าอันตรภาคชั้นแรกหรืออันตรภาคชั้นสุดท้าย อันตรภาคชั้นใดชั้นหนึ่งหรือทั้งสองอันตรภาคชั้นเป็นอันตรภาคชั้นเปิด ย่อมจะหาพิสัยไม่ได้

การวัดการกระจายโดยใช้พิสัยนี้เป็นวิธีวัดการกระจายอย่างคร่าว ๆ เพราะค่าที่ได้หามาจากค่าของข้อมูลเพียงสองค่าเท่านั้น ค่าอื่น ๆ ของข้อมูลไม่ได้นำมาใช้ในการคำนวณหาพิสัยเลย ดังนั้น ถ้าค่าของข้อมูลค่าใดค่าหนึ่งมีค่ามากหรือน้อยผิดปกติจากค่าของข้อมูลอื่น ๆ ก็อาจมีผลทำให้การวัดการกระจายโดยใช้พิสัยมีค่าสูงกว่าที่ควรจะเป็นจริงมาก ความถูกต้องที่ได้จากการวัดการกระจายโดยวิธีนี้จึงอาจจะมีน้อยเมื่อเปรียบเทียบกับการวัดการกระจายโดยวิธีอื่น ๆ ที่ใช้ค่าของข้อมูลทั้งหมดที่มีอยู่ แต่การวัดการกระจายโดยใช้พิสัยมีข้อดีที่สามารถวัดได้สะดวกและรวดเร็ว ส่วนใหญ่จึงมักใช้วัดการกระจายของข้อมูลในกรณีซึ่งไม่ต้องการความถูกต้องมากนัก

ส่วนเบี่ยงเบนควอร์ไทล์ หรือกึ่งช่วงควอร์ไทล์ (quartile deviation หรือ semi- interquartile range)

ส่วนเบี่ยงเบนควอร์ไทล์ คือ ค่าที่ใช้วัดการกระจายที่หาได้จากครึ่งหนึ่งของผลต่างระหว่างควอร์ไทล์ที่สาม (Q_3) และควอร์ไทล์ที่หนึ่ง (Q_1) นั่นคือ

$$\text{ส่วนเบี่ยงเบนควอร์ไทล์ (Q.D.)} = \frac{Q_3 - Q_1}{2}$$

ตัวอย่างที่ 2 จงหาส่วนเบี่ยงเบนควอร์ไทล์ จากตารางแจกแจงความถี่ แสดงผลผลิตต่อไร่โดยประมาณของถั่วเหลือง (กิโลกรัม) จาก 47 จังหวัดของประเทศไทยมีดังนี้

ผลผลิตถั่วเหลืองเฉลี่ยต่อไร่ (กิโลกรัม)	ความถี่ (จำนวนจังหวัด)	ความถี่สะสม
120 – 138	3	3
139 – 157	1	4
158 – 176	6	10
177 – 195	8	18
196 – 214	19	37
215 – 233	2	39
234 – 252	7	46
253 – 271	1	47

วิธีทำ เนื่องจากตำแหน่งที่ Q_1 คือ $\frac{47}{4} = 11.75$

จะได้ $Q_1 \;=\; 176.5 + \frac{1.75 \times 19}{8} = 180.66$

และเนื่องจากตำแหน่งที่ Q_3 คือ $3(\frac{47}{4}) = 35.25$

จะได้ $Q_3 \;=\; 195.5 + \frac{17.25 \times 19}{19} = 212.75$

ดังนั้น ส่วนเบี่ยงเบนควอร์ไทล์ (Q.D.) $= \frac{Q_3 - Q_1}{2}$

$$= \frac{212.75 - 180.66}{2}$$

$$= 16.045$$

นั่นคือ ผลผลิตถั่วเหลืองเฉลี่ยต่อไร่ใน 47 จังหวัดของประเทศไทยมีค่าการกระจายของข้อมูลที่วัดโดยส่วนเบี่ยงเบนควอร์ไทล์มีค่าประมาณ 16 กิโลกรัม

การวัดการกระจายของข้อมูลโดยใช้ส่วนเบี่ยงเบนควอร์ไทล์มีส่วนดีเพราะกรณีที่มีข้อมูลบางค่าสูงหรือต่ำกว่าข้อมูลอื่น ๆ ในชุดเดียวกันมาก จะมีผลกระทบต่อพิสัยแต่จะไม่มีผลกระทบต่อส่วนเบี่ยงเบนควอร์ไทล์ เนื่องจากมิได้นำเอาข้อมูลที่มีค่าต่ำกว่า Q_1 หรือสูงกว่า Q_3 มาคำนวณด้วย ในกรณีที่การแจกแจงความถี่ของข้อมูลมีอันตรภาคชั้นแรก หรืออันตรภาคชั้นสุดท้ายในตารางเป็นอันตรภาคชั้นเปิด ก็สามารถหาส่วนเบี่ยงเบนควอร์ไทล์ได้ เพราะไม่ต้องเกี่ยวข้องกับอันตรภาคชั้นแรกและอันตรภาคชั้นสุดท้ายในตารางแต่อย่างใด ต่างกับการหาส่วนเบี่ยงเบนชนิดอื่น ๆ ที่จะกล่าวต่อไป แต่ส่วนเบี่ยงเบนควอร์ไทล์มีข้อเสียเนื่องจากไม่ได้ใช้ข้อมูลทั้งหมดที่มีอยู่มาคำนวณ ใช้เฉพาะข้อมูลที่มีค่าใกล้เคียงหรือเท่ากับข้อมูลในตำแหน่งที่ของควอร์ไทล์ที่หนึ่งและสามเท่านั้น จึงเป็นการวัดการกระจายซึ่งไม่ละเอียดนัก

ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation)

ส่วนเบี่ยงเบนเฉลี่ย คือ ค่าที่ใช้วัดการกระจายของข้อมูลที่ได้จากการเฉลี่ยค่าสัมบูรณ์ของความแตกต่างระหว่างค่าของข้อมูลแต่ละค่าจากค่ากลางของข้อมูลชุดนั้น ค่ากลางที่ใช้อาจเป็นค่าเฉลี่ยเลขคณิตหรือมัธยฐานก็ได้ แต่ส่วนมากนิยมใช้ค่าเฉลี่ยเลขคณิต

การหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่ไม่ได้แจกแจงความถี่

ถ้า $x_1, x_2, x_3, \ldots, x_n$ เป็นข้อมูลตัวอย่าง n จำนวน และมีค่าเฉลี่ยเลขคณิตเป็น $\overline{X}$ แล้ว

$$\text{ส่วนเบี่ยงเบนเฉลี่ย (M.D.)} = \frac{|x_1 - \overline{X}| + |x_2 - \overline{X}| + |x_3 - \overline{X}| + \cdots + |x_n - \overline{X}|}{n}$$

$$= \frac{\sum_{i=1}^{n} |x_i - \overline{X}|}{n}$$

ตัวอย่างที่ 3 ปริมาณน้ำฝน (มิลลิเมตร) ของจังหวัดต่างๆ 10 จังหวัดในภาคตะวันออกเฉียงเหนือ ในปี พ.ศ. 2545 แสดงไว้ในตาราง จงหาส่วนเบี่ยงเบนเฉลี่ยของปริมาณน้ำฝน ดังกล่าว

จังหวัด	**ปริมาณน้ำฝน (มิลลิเมตร)**
ขอนแก่น	1,402.6
ชัยภูมิ	927.5
นครพนม	2,995.9
มุกดาหาร	1,901.7
ร้อยเอ็ด	1,357.2
เลย	1,414.8
สกลนคร	1,888.6
สุรินทร์	1,857.9
หนองคาย	2,247.5
อุดรธานี	1,777.0

ที่มา : กรมอุตุนิยมวิทยา กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร

วิธีทำ จาก $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1402.6 + 927.5 + 2995.9 + \cdots + 1777.0}{10} = 1,777.07$

จะได้ $\text{M.D.} = \frac{\sum_{i=1}^{n} \left| x_i - \overline{X} \right|}{n} = \frac{374.47 + 849.57 + 1218.83 + \cdots + 0.07}{10} = 401.25$

นั่นคือ ส่วนเบี่ยงเบนเฉลี่ยของปริมาณน้ำฝนใน 10 จังหวัดของภาคตะวันออกเฉียงเหนือ ในปี พ.ศ. 2545 เท่ากับ 401.25 มิลลิเมตร หรือกล่าวได้ว่าโดยเฉลี่ยแล้วปริมาณน้ำฝนของแต่ละจังหวัดต่างจากปริมาณน้ำฝนโดยเฉลี่ยของทั้ง 10 จังหวัด 401.25 มิลลิเมตร

การหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่แจกแจงความถี่

ถ้า $x_1, x_2, x_3, \ldots, x_k$ เป็นจุดกึ่งกลางของอันตรภาคชั้นต่าง ๆ k ชั้น ซึ่งมีความถี่ $f_1, f_2, f_3, \ldots, f_k$ ตามลำดับ n เป็นจำนวนข้อมูลตัวอย่างทั้งหมด และถ้าข้อมูลชุดนี้มีค่าเฉลี่ยเลขคณิตเป็น $\overline{X}$ แล้ว สามารถคำนวณส่วนเบี่ยงเบนเฉลี่ย (โดยประมาณ) ได้จากสูตร

$$\text{ส่วนเบี่ยงเบนเฉลี่ย (M.D.)} = \frac{\sum_{i=1}^{k} f_i \left| x_i - \overline{X} \right|}{\sum_{i=1}^{k} f_i}$$

$$= \frac{\sum_{i=1}^{k} f_i \left| x_i - \overline{X} \right|}{n}$$

เมื่อ k แทนจำนวนอันตรภาคชั้น

f_i แทนความถี่ของอันตรภาคชั้นที่ i

x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

ตัวอย่างที่ 4 จงหาส่วนเบี่ยงเบนเฉลี่ยของผลผลิตข้าวเฉลี่ยต่อไร่ของ 50 จังหวัด ที่เป็นตัวอย่างหรือตัวแทนทั้งหมดของประเทศไทยในปีเพาะปลูก 2544/45 จากตารางแจกแจงความถี่ในตัวอย่างที่ 2 หัวข้อ 1.2

วิธีทำ เพื่อความสะดวกในการคำนวณหาค่าส่วนเบี่ยงเบนเฉลี่ย จึงสร้างตารางดังนี้

ผลผลิตข้าวเฉลี่ยต่อไร่ (กิโลกรัม)	ความถี่ (f_i)	จุดกึ่งกลางชั้น (x_i)	$f_i x_i$	$\left\|x_i - \overline{X}\right\|$	$f_i \left\|x_i - \overline{X}\right\|$
258 – 312	5	285	1,425	216.7	1,083.5
313 – 367	8	340	2,720	161.7	1,293.6
368 – 422	1	395	395	106.7	106.7
423 – 477	6	450	2,700	51.7	310.2
478 – 532	10	505	5,050	3.3	33.0
533 – 587	7	560	3,920	58.3	408.1
588 – 642	3	615	1,845	113.3	339.9
643 – 697	5	670	3,350	168.3	841.5
698 – 752	4	725	2,900	223.3	893.2
753 – 807	1	780	780	278.3	278.3
รวม	50		25,085		5,588

จากตารางแจกแจงความถี่ข้างต้น จะหาค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนเฉลี่ยได้ดังนี้

$$\overline{X} = \frac{\sum_{i=1}^{10} f_i x_i}{n} \qquad \text{M.D.} = \frac{\sum_{i=1}^{10} f_i \left| x_i - \overline{X} \right|}{n}$$

$$= \frac{25,085}{50} \qquad = \frac{5,588}{50}$$

$$= 501.7 \qquad = 111.76$$

นั่นคือ ส่วนเบี่ยงเบนเฉลี่ยโดยประมาณของผลผลิตข้าวเฉลี่ยต่อไร่ของ 50 จังหวัดของประเทศไทย ในปีเพาะปลูก 2544/45 มีค่าเท่ากับ 111.76 กิโลกรัม หรือโดยเฉลี่ยแล้วผลผลิตข้าวเฉลี่ยต่อไร่ของแต่ละจังหวัดต่างจากผลผลิตข้าวโดยเฉลี่ยของทั้ง 50 จังหวัดประมาณ 111.76 กิโลกรัมต่อไร่

เนื่องจากการวัดการกระจายโดยใช้ส่วนเบี่ยงเบนเฉลี่ยนี้ ได้นำค่าของข้อมูลทุก ๆ ค่าหรือทุกค่าของข้อมูลมีตัวแทนมาใช้ในการคำนวณ จึงทำให้การวัดการกระจายโดยวิธีนี้ดีกว่าการวัดการกระจายโดยใช้พิสัย หรือส่วนเบี่ยงเบนควอร์ไทล์ ซึ่งใช้เฉพาะค่าของข้อมูลเพียงบางค่าเท่านั้น แต่ในกรณีที่ข้อมูลมีจำนวนมาก ๆ จะมีปัญหาในการคำนวณโดยเฉพาะเมื่อต้องใช้เครื่องคำนวณที่มีความสามารถต่ำ ซึ่งจะทำให้หาค่าสัมบูรณ์ของผลต่างค่อนข้างลำบาก จึงนิยมวัดการกระจายโดยใช้ส่วนเบี่ยงเบนมาตรฐานซึ่งจะได้กล่าวถึงในลำดับต่อไป

แบบฝึกหัด 1.3 ก

1. จงหาทุกควอร์ไทล์และส่วนเบี่ยงเบนควอร์ไทล์ของข้อมูลต่อไปนี้
   1) 1 2 3 4 5 6
   2) 1 2 3 4 5 6 7
   3) 1 2 3 4 5 6 7 8
   4) 1 2 3 4 5 6 7 8 9

2. จากตารางราคาสินค้าชนิดหนึ่งที่จำหน่ายในห้างสรรพสินค้า 5 แห่ง เป็นดังนี้

ห้างสรรพสินค้า	โชคชัย	สินทวี	เซนโก้	ทรัพย์ทวี	แจ่มฟ้า
ราคาสินค้า (บาท)	2,000	2,025	1,995	1,990	2,000

นักเรียนคิดว่าถ้าวัดการกระจายของข้อมูลชุดนี้โดยใช้พิสัย ค่าที่วัดได้จะมีความถูกต้องพอที่จะเชื่อถือได้หรือไม่ เพราะเหตุใด

3. ในปี พ.ศ. 2543 การไฟฟ้าฝ่ายผลิตรายงานการผลิตไฟฟ้าจากโรงงานไฟฟ้าพลังน้ำแต่ละเขื่อนได้ดังนี้

เขื่อน	**กำลังผลิต (เมกะวัตต์)**
ภูมิพล	743.90
สิริกิติ์	500.00
อุบลรัตน์	25.20
สิรินธร	36.00
จุฬาภรณ์	40.00
น้ำพุง	6.00
ศรีนครินทร์	720.00
วชิราลงกรณ์	300.00
ท่าทุ่งนา	38.00
แก่งกระจาน	17.50
บางลาง	72.00
บ้านสันติ	1.28
ห้วยกุ่ม	1.06
แม่งัดสมบูรณ์ชล	9.00
รัชชประภา	240.00
ปากมูล	136.00

ที่มา : การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย

จงหาส่วนเบี่ยงเบนควอร์ไทล์และส่วนเบี่ยงเบนเฉลี่ยของปริมาณการผลิตกำลังไฟฟ้าจากโรงงานไฟฟ้าพลังน้ำข้างต้น

4. ปริมาณการผลิตไม้สักในประเทศไทย จำแนกตามจังหวัดในปี พ.ศ. 2545 เป็นดังนี้

จังหวัด	ปริมาณที่ผลิต (ลูกบาศก์เมตร)
ตาก	6,284
เพชรบูรณ์	884
แม่ฮ่องสอน	678
เชียงใหม่	426
กำแพงเพชร	50
เชียงราย	45
อุตรดิตถ์	44
น่าน	39

ที่มา : กรมป่าไม้ กระทรวงเกษตรและสหกรณ์

1) จงหาพิสัย ส่วนเบี่ยงเบนควอร์ไทล์ และส่วนเบี่ยงเบนเฉลี่ย
2) เปรียบเทียบค่าที่ได้จากการวัดทั้งสามวิธี นักเรียนคิดว่าควรใช้พิสัยในการวัดการกระจายของข้อมูลชุดนี้หรือไม่ เพราะเหตุใด
3) นักเรียนคิดว่าวิธีการวัดการกระจายในข้อ 1) วิธีใดใช้วัดการกระจายของข้อมูลได้เหมาะสมที่สุด เพราะเหตุใด

5. จากการศึกษาอัตราเร็ว (เมตร / ชั่วโมง) ในการวิ่งของสัตว์เลี้ยงลูกด้วยนมจำนวนทั้งหมด 18 ตัว ซึ่งประกอบด้วย สัตว์ป่า 12 ตัว และสัตว์เลี้ยง 6 ตัว เมื่อคำนวณค่า Q_1 และ Q_3 และมัธยฐาน ได้ผลดังนี้

	Q_1	มัธยฐาน	Q_3
สัตว์เลี้ยง	30	37	40
สัตว์ป่า	27.5	36	43.5

อยากทราบว่า สัตว์เลี้ยงหรือสัตว์ป่ามีการกระจายของข้อมูลมากกว่าอีกฝ่ายหนึ่ง

6. จงหาส่วนเบี่ยงเบนควอร์ไทล์และส่วนเบี่ยงเบนเฉลี่ยของรายได้ของคนงานหญิง 400 คน จากตารางแจกแจงความถี่ต่อไปนี้ แล้วเปรียบเทียบค่าที่ได้กับพิสัยของข้อมูลชุดนี้

รายได้ (บาท)	จำนวนคนงาน (คน)
1,500 – 1,599	20
1,600 – 1,699	70
1,700 – 1,799	120
1,800 – 1,899	100
1,900 – 1,999	60
2,000 – 2,099	20
2,100 – 2,199	10

ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

การวัดการกระจายของข้อมูลโดยใช้ส่วนเบี่ยงเบนมาตรฐาน เป็นวิธีที่นักสถิติยอมรับว่าเป็นวิธีที่ใช้วัดการกระจายได้ดีที่สุด เมื่อเปรียบเทียบกับวิธีวัดการกระจายทั้งสามวิธีที่ได้กล่าวมาแล้ว คือ พิสัย ส่วนเบี่ยงเบนควอร์ไทล์ และส่วนเบี่ยงเบนเฉลี่ย ทั้งนี้เนื่องจากการวัดการกระจายโดยวิธีนี้ใช้ข้อมูลทุก ๆ ค่าหรือนำตัวแทนของข้อมูลทุกค่ามาคำนวณ และขจัดปัญหาในการที่ต้องใช้ค่าสัมบูรณ์ให้หมดไป การวัดการกระจายโดยวิธีนี้นอกจากจะให้ค่าการกระจายที่มีความละเอียดถูกต้อง และเชื่อถือได้มากที่สุดแล้วยังสามารถนำไปใช้ในการวิเคราะห์ข้อมูลสถิติในชั้นสูงต่อไป ซึ่งการวัดการกระจายแบบอื่นนำไปใช้ไม่ได้

การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่ไม่ได้แจกแจงความถี่

ถ้า $x_1, x_2, x_3, \ldots, x_N$ เป็นข้อมูลของประชากร N หน่วย และมีค่าเฉลี่ยเลขคณิตเป็น μ แล้ว ส่วนเบี่ยงเบนมาตรฐานของประชากร หรือ σ (อ่านว่า "ซิกมา") สามารถคำนวณได้ดังนี้

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

หรือ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}x_i^2}{N} - \mu^2}$$

โดยที่ μ แทนค่าเฉลี่ยเลขคณิตของประชากร

และ N แทนจำนวนข้อมูลทั้งหมดของประชากร

นอกจากการใช้สัญลักษณ์ σ แล้ว อาจใช้สัญลักษณ์ S.D. หรือ s ในกรณีที่ไม่สามารถศึกษาข้อมูลทั้งหมดของประชากร และข้อมูลที่ใช้เป็นข้อมูลจากตัวอย่างซึ่งเป็นตัวแทนของประชากรแล้ว ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง (sample standard deviation หรือ s) ซึ่งใช้เป็นตัวประมาณของ σ คำนวณได้ดังนี้

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n-1}}$$

หรือ

$$s = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - n\overline{X}^2}{n-1}}$$

โดยที่ $\overline{X}$ แทนค่าเฉลี่ยเลขคณิตของตัวอย่าง

และ n แทนจำนวนข้อมูลทั้งหมดของตัวอย่าง

อนึ่งในทางปฏิบัติ ข้อมูลที่ใช้มักเป็นข้อมูลที่เป็นตัวอย่างของประชากร จึงนิยมใช้สูตร s แทน σ เพราะโดยทั่วไปเราไม่ทราบค่า σ ดังนั้นในการวิเคราะห์ข้อมูลเชิงสถิติจึงนิยมใช้ s เป็นตัวประมาณของ σ อยู่เสมอ (ส่วนการพิสูจน์สมบัติของ s เช่น สูตรของ s ตัวส่วนที่หารด้วย $n-1$ แทนการหารด้วย n แสดงวิธีไว้ในทฤษฎีทางสถิติและไม่ได้อยู่ในหลักสูตรจึงยังไม่กล่าวถึงในระดับนี้)

อย่างไรก็ตามในเบื้องต้น อาจทราบเพียงว่า สูตรการคำนวณส่วนเบี่ยงเบนมาตรฐานของตัวอย่างมีตัวหาร 2 แบบ คือ แบบที่หารด้วย $n-1$ และแบบที่หารด้วย n เช่น ที่เห็นได้จากสูตรบนเครื่องคิดเลขบางเครื่องที่อาจใช้สัญลักษณ์ σ_{n-1} และ σ_n ตามลำดับ การหารด้วย $n-1$ จะให้ค่า S.D. หรือ s ข้างต้นสูงกว่าค่าของ s ที่ใช้ตัวหารเป็น n และการหารด้วย $n-1$ ยังสนับสนุนการอนุมานหรือการอ้างอิงเชิงสถิติ (statistical inference) ในเรื่องสมบัติต่าง ๆ ของตัวประมาณ ถ้าสูตรของ s ที่ใช้ตัวหาร 2 แบบนี้ ให้ผลลัพธ์ต่างกันมาก อาจบอกได้ว่าขนาดตัวอย่างที่ใช้เล็กเกินไป และถ้าขนาดตัวอย่างมากขึ้นผลลัพธ์ดังกล่าวจะใกล้เคียงกัน ดังนั้น เมื่อตัวอย่างมีขนาดใหญ่มาก หรือในระดับประชากรค่าที่คำนวณได้จากสูตร 2 สูตร มีค่าไม่ต่างกัน ในทางปฏิบัติจึงนิยมใช้สูตรที่มีตัวหาร $n-1$ มากกว่าใช้ n หนังสือเล่มนี้จึงใช้สูตรที่มีตัวหาร $n-1$ แทน n เสมอ

ตัวอย่างที่ 5 จงหาส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของทีมนักเรียนที่เข้าร่วมแข่งขันคณิตศาสตร์โอลิมปิก ครั้งที่ 45 (พ.ศ. 2547) ณ ประเทศกรีซจากประเทศในเอเชียตะวันออกเฉียงใต้ทั้งหมดที่เข้าร่วมแข่งขันซึ่งมีผลสอบดังนี้

ประเทศ	คะแนน (x_i)
เวียดนาม	196
สิงคโปร์	139
ไทย	99
อินโดนีเซีย	61
มาเลเซีย	34
ฟิลิปปินส์	16
รวม	**545**

วิธีทำ จากตารางคะแนนสอบวิชาคณิตศาสตร์ นำไปหา x_i^2 จะได้ดังนี้

ประเทศ	คะแนน (x_i)	x_i^2
เวียดนาม	196	38,416
สิงค์โปร์	139	19,321
ไทย	99	9,801
อินโดนีเซีย	61	3,721
มาเลเซีย	34	1,156
ฟิลิปปินส์	16	256
รวม	**545**	**72,671**

จากคะแนนสอบข้างต้น จะได้

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{545}{6}$$

และจะได้

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2} = \sqrt{\frac{72,671}{6} - \left(\frac{545}{6}\right)^2} \approx 62.1381$$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของทีมนักเรียนทั้งหมดจากประเทศในเอเชียตะวันออกเฉียงใต้ ที่ส่งทีมนักเรียนเข้าร่วมแข่งขันคณิตศาสตร์โอลิมปิก ครั้งที่ 45 ประมาณ 62 คะแนน

การหาส่วนเบี่ยงเบนมาตรฐานของประชากรที่แจกแจงความถี่

การหาส่วนเบี่ยงเบนมาตรฐานของประชากรที่แจกแจงความถี่แล้วเป็นการหาค่าโดยประมาณ เพราะต้องใช้จุดกึ่งกลางของแต่ละอันตรภาคชั้นเป็นตัวแทนข้อมูลซึ่งจะใช้วิธีเดียวกันกับการหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่ไม่แจกแจงความถี่ ดังนี้

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i - \mu)^2}{N}}$$

หรือ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{N} - \mu^2}$$

เมื่อ x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

f_i แทนความถี่ของอันตรภาคชั้นที่ i

k แทนจำนวนอันตรภาคชั้น

N แทนจำนวนข้อมูลทั้งหมดในประชากร หรือผลรวมของความถี่ของทุก ๆ อันตรภาคชั้น

μ แทนค่าเฉลี่ยเลขคณิตของประชากร

ความแปรปรวนของประชากร (population variance)

ความแปรปรวน คือ กำลังสองของส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวนของประชากรที่ไม่ได้แจกแจงความถี่ หาได้โดยใช้สูตร

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} = \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2$$

และความแปรปรวนของประชากรที่แจกแจงความถี่ หาได้โดยใช้สูตร

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{N} = \frac{\sum_{i=1}^{k} f_i x_i^2}{N} - \mu^2$$

เมื่อ k แทนจำนวนอันตรภาคชั้น

x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

f_i แทนความถี่ของอันตรภาคชั้นที่ i

การหาส่วนเบี่ยงเบนมาตรฐานของตัวอย่างที่แจกแจงความถี่

การหาส่วนเบี่ยงเบนมาตรฐาน (s) จากตัวอย่างที่มีจำนวนมากหรือน้อยก็ตาม ทำได้ทำนองเดียวกันกับกรณีของข้อมูลที่ไม่แจกแจงความถี่ ซึ่งค่าที่ได้จะเป็นค่าประมาณ และในปัจจุบันถ้ามีข้อมูลดิบทุกหน่วย (ข้อมูลของแต่ละหน่วยที่ยังไม่แจกแจงความถี่) สามารถใช้คอมพิวเตอร์ช่วยคำนวณได้สะดวก ไม่ว่าข้อมูลจะมีจำนวนมากหรือน้อยเพียงไร อย่างไรก็ตาม ถ้ามีกรณีที่ข้อมูลไม่ใช่ข้อมูลดิบทุกหน่วย แต่เป็นข้อมูลที่อาจมาจากแหล่งทุติยภูมิอื่น ๆ ซึ่งข้อมูลมีการแจกแจงความถี่เป็นอันตรภาคชั้น หรือเป็นกลุ่มมาแล้ว สามารถให้สูตรการหาส่วนเบี่ยงเบนมาตรฐานดังนี้

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \overline{X})^2}{n-1}}$$

หรือ

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2 - n\overline{X}^2}{n-1}}$$

โดยที่ x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

f_i แทนความถี่ของอันตรภาคชั้นที่ i

$\overline{X}$ แทนค่าเฉลี่ยเลขคณิตของตัวอย่าง

n แทนจำนวนตัวอย่างทั้งหมด ($n = \sum_{i=1}^{k} f_i$)

k แทนจำนวนอันตรภาคชั้นหรือจำนวนกลุ่ม

ความแปรปรวนของตัวอย่าง (sample variance)

ความแปรปรวนของตัวอย่างทั้งกรณีไม่แจกแจงความถี่และแจกแจงความถี่ คือ กำลังสองของส่วนเบี่ยงเบนมาตรฐานของตัวอย่างในทั้งสองกรณีตามลำดับ

ดังนั้น ความแปรปรวนของตัวอย่าง คำนวณได้ดังนี้

กรณีไม่แจกแจงความถี่

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n-1}$$

หรือ
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{X}^2}{n-1}$$

กรณีมีการแจกแจงความถี่

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \overline{X})^2}{n-1}$$

หรือ
$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2 - n\overline{X}^2}{n-1}$$

โดยที่ x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

f_i แทนความถี่ของอันตรภาคชั้นที่ i

$\overline{X}$ แทนค่าเฉลี่ยเลขคณิตของตัวอย่าง

n แทนจำนวนตัวอย่างทั้งหมด ($n = \sum_{i=1}^{k} f_i$)

k แทนจำนวนอันตรภาคชั้นหรือจำนวนกลุ่ม

ตัวอย่างที่ 6 จงหาความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานโดยประมาณจากจำนวนวันหยุดเรียนของนักเรียน 50 คน ที่แสดงไว้ในตารางต่อไปนี้

จำนวนวันที่หยุดเรียน (วัน)	**จำนวนนักเรียน (คน)**
0 – 2	15
3 – 5	20
6 – 8	12
9 – 11	2
12 – 14	1

วิธีทำ จากข้อมูลข้างต้นนำมาสร้างตารางใหม่ได้ดังนี้

จำนวนวันที่หยุดเรียน	**ความถี่ (f_i)**	**จุดกึ่งกลาง (x_i)**	x_i^2	f_ix_i	$f_ix_i^2$
0 – 2	15	1	1	15	15
3 – 5	20	4	16	80	320
6 – 8	12	7	49	84	588
9 – 11	2	10	100	20	200
12 – 14	1	13	169	13	169
	$\sum_{i=1}^{5} f_i = 50$			$\sum_{i=1}^{5} f_ix_i = 212$	$\sum_{i=1}^{5} f_ix_i^2 = 1,292$

ดังนั้น $\overline{X} = \dfrac{\sum_{i=1}^{5} f_ix_i}{n} = \dfrac{212}{50}$

จะได้ $s^2 = \dfrac{\sum_{i=1}^{5} f_ix_i^2 - n\overline{X}^2}{n-1} = \dfrac{1,292 - 50\left(\dfrac{212}{50}\right)^2}{50-1} = \dfrac{1404}{175} \approx 8.02$

และ $s = \sqrt{\frac{1404}{175}} \approx 2.83$

นั่นคือ จำนวนวันหยุดเรียนของนักเรียนแต่ละคนจะต่างจากจำนวนวันหยุดเรียนโดยเฉลี่ยของนักเรียนทั้ง 50 คน อยู่ประมาณ 3 วัน และความแปรปรวนของจำนวนวันหยุดเรียนของนักเรียนทั้ง 50 คน ประมาณ 8 วัน^2

ตัวอย่างที่ 7 จงหาส่วนเบี่ยงเบนมาตรฐานของอายุขัย* (longevity) ของสัตว์เลี้ยงลูกด้วยนม 10 ประเภท ประเภทละ 1 ตัว ดังในตารางต่อไปนี้

สัตว์เลี้ยง (ตัวที่)	**ประเภท**	**อายุขัย (ปี)**
1	แมว	12
2	วัว	15
3	สุนัข	12
4	ลา	12
5	แพะ	8
6	หนูตะเภา	4
7	ม้า	20
8	หมู	10
9	กระต่าย	5
10	แกะ	12
รวม		**110**

ที่มา : Watkins et al. (2004)

* อายุขัย หมายถึง การสิ้นอายุ ความตาย อัตรากำหนดอายุจนสิ้นอายุ

วิธีทำ จากข้อมูลข้างต้นนำมาสร้างตารางใหม่ได้ดังนี้

สัตว์เลี้ยง (ตัวที่)	**ประเภท**	**อายุขัย (ปี) (x_i)**	**ค่าเฉลี่ย ($\overline{X}$)**	$x_i - \overline{X}$	$(x_i - \overline{X})^2$
1	แมว	12	11	1	1
2	วัว	15	11	4	16
3	สุนัข	12	11	1	1
4	ลา	12	11	1	1
5	แพะ	8	11	–3	9
6	หนูตะเภา	4	11	–7	49
7	ม้า	20	11	9	81
8	หมู	10	11	–1	1
9	กระต่าย	5	11	–6	36
10	แกะ	12	11	1	1
รวม		**110**		**0**	**196**

จะได้ $s^2 = \dfrac{\sum_{i=1}^{10}(x_i - \overline{X})^2}{n-1}$

$= \dfrac{196}{9}$

ดังนั้น $s = \sqrt{\dfrac{196}{9}} = \dfrac{14}{3}$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของอายุขัยของสัตว์เลี้ยงลูกด้วยนมประมาณ $\frac{14}{3}$ ปี หรือ 4 ปี 8 เดือน

สมบัติที่สำคัญของส่วนเบี่ยงเบนมาตรฐาน

1. ส่วนเบี่ยงเบนมาตรฐานจะมากกว่าหรือเท่ากับศูนย์

2. ถ้าคำนวณหาส่วนเบี่ยงเบนมาตรฐานโดยใช้ค่ากลางของข้อมูลชนิดอื่น ๆ ที่ไม่ใช่ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานที่ได้จะมีค่ามากกว่าส่วนเบี่ยงเบนมาตรฐานที่ใช้ค่าเฉลี่ยเลขคณิตเสมอ นั่นคือ

$$\sqrt{\frac{\sum_{i=1}^{N}(x_i - a)^2}{N}} > \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

อสมการนี้เป็นจริงเสมอ เมื่อ a เป็นจำนวนจริงใด ๆ ที่ไม่เท่ากับค่าเฉลี่ยเลขคณิต

สรุปสัญลักษณ์ที่เกี่ยวกับค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานและจำนวนข้อมูลที่ใช้เป็นดังนี้

	ประชากร (พารามิเตอร์)	ตัวอย่าง (ตัวประมาณ)
ค่าเฉลี่ยเลขคณิต	μ	$\overline{X}$
ส่วนเบี่ยงเบนมาตรฐาน	σ	s หรือ S.D.
จำนวนข้อมูล	N	n

แบบฝึกหัด 1.3 ข

1. จงหาพิสัย ส่วนเบี่ยงเบนมาตรฐาน ส่วนเบี่ยงเบนเฉลี่ย และส่วนเบี่ยงเบนควอร์ไทล์ของราคาเครื่องสำอางชนิดหนึ่ง ซึ่งจากการสำรวจร้านค้ามาเป็นตัวอย่าง 8 แห่ง ได้ราคาของเครื่องสำอางดังนี้ 410 415 425 410 640 400 410 และ 410 บาท ตามลำดับ แล้วพิจารณาว่าการวัดการกระจายของข้อมูลวิธีใดเหมาะสมกับข้อมูลที่สุด

2. จงหาส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนจากข้อมูลที่เกี่ยวกับปริมาณน้ำฝนของจังหวัดต่าง ๆ ในภาคตะวันออกเฉียงเหนือในปี พ.ศ. 2545 ดังตาราง

จังหวัด	ปริมาณน้ำฝน (มิลลิเมตร)
ขอนแก่น	1,402.6
ชัยภูมิ	927.5
นครพนม	2,995.9
มุกดาหาร	1,901.7
ร้อยเอ็ด	1,357.2
เลย	1,414.8
สกลนคร	1,888.6
สุรินทร์	1,857.9
หนองคาย	2,247.5
อุดรธานี	1,777.0

ที่มา : กรมอุตุนิยมวิทยา กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร

3. จงหาค่าเฉลี่ยเลขคณิตและเปรียบเทียบการกระจายของราคาสินค้าชนิดหนึ่งที่ขายตามร้านต่าง ๆ ในสองท้องที่ซึ่งขายในราคาที่ต่างกันดังนี้
ท้องที่ที่หนึ่ง (บาท) 50, 52, 45, 55, 54, 48, 53
ท้องที่ที่สอง (บาท) 40, 50, 51, 52, 51, 51, 62, 53, 49

4. ครอบครัวหนึ่งมีบุตร 4 คน ถ้าอายุของบิดา มารดา และบุตรทั้งสี่คน เป็น 45, 42, 20, 17, 16 และ 14 ปีตามลำดับ จงหาส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของอายุของสมาชิกทุกคนในครอบครัวนี้ และในอีก 5 ปีข้างหน้า ส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของอายุสมาชิกทุกคนในครอบครัวนี้จะเป็นอย่างไร

5. ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่เป็นกลุ่มตัวอย่างของประชากรขนาด 20 รายการ ชุดหนึ่งเป็น 10 และ 2 ตามลำดับ ภายหลังพบว่ามีการบันทึกข้อมูลรายการหนึ่งผิดพลาดไป โดย 12 บันทึกเป็น 8 จงคำนวณหาค่าเฉลี่ยเลขคณิต และส่วนเบี่ยงเบนมาตรฐานที่ถูกต้อง

6. เมื่อใช้ $n-1$ เป็นตัวหารจะให้ผลลัพธ์มากขึ้นหรือน้อยลงจากการใช้ n เป็นตัวหารในสูตรส่วนเบี่ยงเบนมาตรฐาน s_1 และ s_2 และจากส่วนเบี่ยงเบนมาตรฐานของตัวอย่างที่ได้เรียนแล้วจะใช้สูตรใดไปประมาณส่วนเบี่ยงเบนมาตรฐานของประชากร (σ)

$$s_1 = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n-1}} \quad \text{และ} \quad s_2 = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n}}$$

7. จากข้อมูลของตัวอย่างที่ 7 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ $\frac{14}{3}$ ปี เปรียบเสมือนช่วงระยะเวลาที่ห่างจากค่าเฉลี่ย 11 ปี สำหรับอายุขัยโดยเฉลี่ยของสัตว์เลี้ยงลูกด้วยนมในตัวอย่างใช่หรือไม่ จงอธิบายเพิ่มเติม เช่น บางตัวอาจมีอายุขัย $11-\frac{14}{3} = \frac{19}{3}$ ปี หรือบางตัวอาจมีอายุขัย $11+\frac{14}{3} = \frac{47}{3}$ ปี ใช่หรือไม่ โดยให้เหตุผลประกอบ

8. จากข้อ 7 จงบอกหน่วยของส่วนเบี่ยงเบนมาตรฐาน มัธยฐาน พิสัย และกึ่งช่วงควอร์ไทล์

1.3.2 การวัดการกระจายสัมพัทธ์ (measures of relative variation)

ในการเปรียบเทียบข้อมูลตั้งแต่สองชุดขึ้นไปเพื่อตัดสินว่าชุดใดมีการกระจายมาก ชุดใดมีการกระจายน้อย ถ้าใช้ค่าที่ได้จากการวัดการกระจายสัมบูรณ์ของข้อมูลแต่ละชุดมาเปรียบเทียบกันย่อมตัดสินได้ยาก เช่น ข้อมูลชุดหนึ่งมีค่าตั้งแต่ 0 ถึง 10 มีส่วนเบี่ยงเบนมาตรฐาน 2.2 อีกชุดหนึ่งมีค่าตั้งแต่ 200 ถึง 800 มีส่วนเบี่ยงเบนมาตรฐาน 60.5 ถ้าจะตัดสินว่าข้อมูลชุดที่หนึ่งมีการกระจายน้อยกว่าข้อมูลชุดที่สองก็อาจจะไม่ถูกต้องนัก เพราะค่าของข้อมูลสองชุดนี้ต่างกันมาก

ค่ากลางและค่าแสดงการกระจายย่อมจะต่างกันมากด้วย เพื่อให้การเปรียบเทียบมีความหมาย จึงนิยมหาอัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้น ๆ (แต่ไม่ใช้ค่ากลางเป็นตัวหารในกรณีที่หาพิสัย หรือส่วนเบี่ยงเบนควอร์ไทล์ เพราะการวัดการกระจายทั้งสองไม่ได้วัดจากค่ากลาง) แล้วจึงนำอัตราส่วนที่หาได้มาเปรียบเทียบกัน อัตราส่วนเหล่านี้มีชื่อเรียกนำหน้าว่า สัมประสิทธิ์ เช่น อัตราส่วนของส่วนเบี่ยงเบนเฉลี่ยกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดเดียวกันเรียกว่า สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย สัมประสิทธิ์ที่ใช้วัดการกระจายสัมพัทธ์มีดังนี้

1. สัมประสิทธิ์ของพิสัย $= \dfrac{x_{max} - x_{min}}{x_{max} + x_{min}}$
2. สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$
3. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง $= \dfrac{M.D.}{\overline{X}}$
4. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของประชากร $= \dfrac{M.D.}{\mu}$
5. สัมประสิทธิ์ของการแปรผันของตัวอย่าง $= \dfrac{s}{\overline{X}}$
6. สัมประสิทธิ์ของการแปรผันของประชากร $= \dfrac{\sigma}{\mu}$

สัมประสิทธิ์ของการแปรผัน เป็นวิธีวัดการกระจายสัมพัทธ์ที่นิยมใช้กันมากที่สุด และมักเขียนในรูปเปอร์เซ็นต์โดยคูณด้วย 100 เช่น ข้อมูลชุดหนึ่งมี $s = 10$ และ $\overline{X} = 30$ สัมประสิทธิ์ของการแปรผันของข้อมูลชุดนี้คือ 0.33 หรือ 33%

ตัวอย่างที่ 8 จงเปรียบเทียบการกระจายของราคาสินค้าสองชนิดที่ได้จากการสำรวจตัวอย่างของร้านค้าที่ขายสินค้าดังกล่าว

ราคาสินค้าชนิดที่ 1 (บาท)	6	7	9	8	12
ราคาสินค้าชนิดที่ 2 (บาท)	50	52	49	55	44

วิธีทำ การคำนวณหาค่าสถิติเพื่อเปรียบเทียบราคาสินค้า แสดงไว้ในตาราง

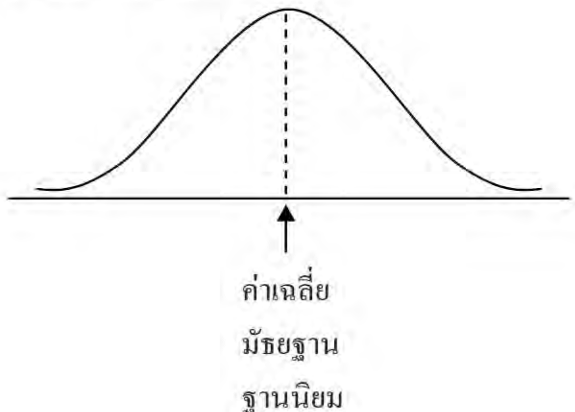
ค่าสถิติ	สินค้าชนิดที่ 1	สินค้าชนิดที่ 2
ควอร์ไทล์ที่ 1 (Q_1)	6.50	46.50
ควอร์ไทล์ที่ 3 (Q_3)	10.50	53.50
ค่าเฉลี่ยเลขคณิต ($\overline{X}$)	8.40	50
ส่วนเบี่ยงเบนเฉลี่ย (M.D.)	1.68	2.80
ส่วนเบี่ยงเบนมาตรฐาน (s)	2.30	4.06
สัมประสิทธิ์ของพิสัย $\left(\frac{x_{max} - x_{min}}{x_{max} + x_{min}}\right)$	$\frac{12-6}{12+6} = 0.33$	$\frac{55-44}{55+44} = 0.11$
สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ $\left(\frac{Q_3 - Q_1}{Q_3 + Q_1}\right)$	$\frac{10.50-6.50}{10.50+6.50} = 0.24$	$\frac{53.50-46.50}{53.50+46.50} = 0.07$
สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง $\left(\frac{M.D.}{\overline{X}}\right)$	$\frac{1.68}{8.40} = 0.20$	$\frac{2.80}{50} = 0.06$
สัมประสิทธิ์ของการแปรผันของตัวอย่าง $\left(\frac{s}{\overline{X}}\right)$	$\frac{2.30}{8.40} = 0.27$	$\frac{4.06}{50} = 0.08$

จากการเปรียบเทียบการวัดการกระจายสัมพัทธ์ของราคาสินค้าสองชนิดโดยวิธีต่าง ๆ กัน จะเห็นได้ว่าผลที่ได้จากการวัดแต่ละวิธีจะเหมือนกัน คือ ราคาสินค้าชนิดที่ 1 มีการกระจายของราคามากกว่าสินค้าชนิดที่ 2 ควรสังเกตด้วยว่า ค่าที่ใช้วัดการกระจายสัมพัทธ์ไม่มีหน่วย

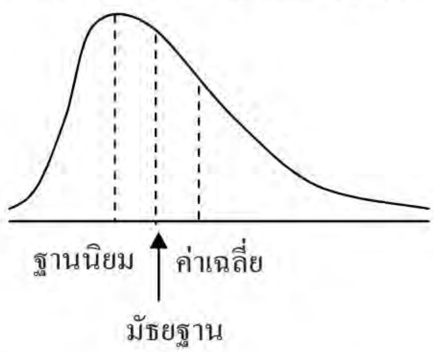
1.3.3 ความสัมพันธ์ระหว่างการแจกแจงความถี่ ค่ากลาง และการกระจายของข้อมูล

จากเรื่องการแจกแจงความถี่ของข้อมูลที่ได้กล่าวมาแล้ว จะเห็นว่าโดยทั่ว ๆไป รูปแบบของเส้นโค้งความถี่อาจแบ่งออกได้เป็น 3 แบบ คือ

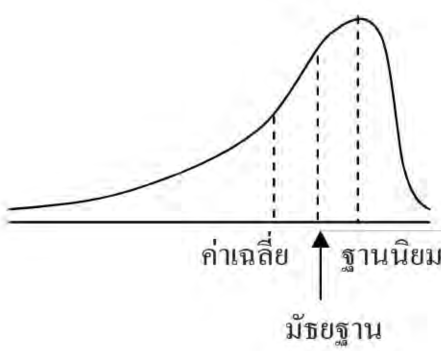
1. เส้นโค้งปกติ (normal curve)



2. เส้นโค้งเบ้ลาดทางขวาหรือทางบวก (positively skewed curve)



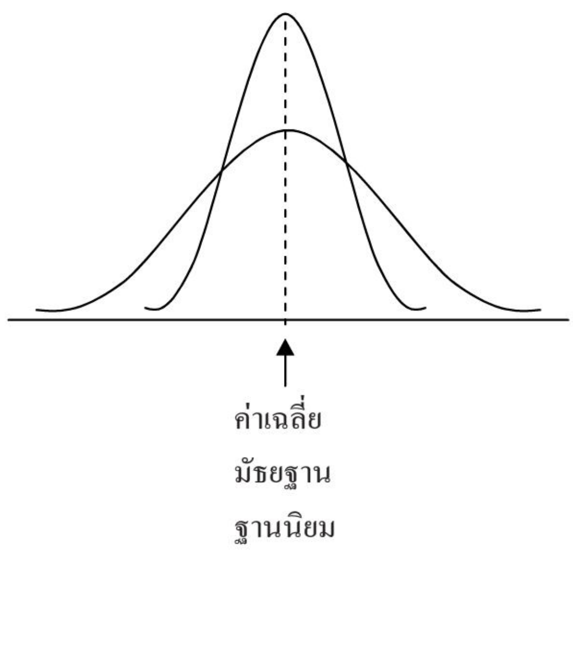
3. เส้นโค้งเบ้ลาดทางซ้ายหรือทางลบ (negatively skewed curve)



เส้นโค้งความถี่ของข้อมูลที่พบอยู่เสมอ ไม่ว่าจะเป็นข้อมูลทางด้านประชากร เกษตร สังคม เศรษฐกิจ หรือวิทยาศาสตร์ ส่วนใหญ่เป็นข้อมูลที่เกิดขึ้นหรือเป็นไปตามธรรมชาติ และจะมีเส้นโค้งความถี่เป็นเส้นโค้งปกติ เช่น ข้อมูลเกี่ยวกับส่วนสูง น้ำหนัก ราคา และผลผลิตทางการเกษตร มักจะมีลักษณะเป็นเส้นโค้งปกติ

ถ้าข้อมูลชุดใดมีเส้นโค้งความถี่เป็นเส้นโค้งปกติ จะมีค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม อยู่ที่จุดเดียวกัน คือ จุดที่มีความถี่สูงสุด ถ้าเส้นโค้งความถี่เป็นเส้นโค้งเบ้ลาดทางขวาหรือส่วนของเส้นโค้งที่มีความชันน้อยอยู่ทางด้านขวาแล้ว ค่าเฉลี่ยเลขคณิตจะมีค่ามากที่สุด รองลงมาเป็นมัธยฐาน และฐานนิยมตามลำดับ ในกรณีที่เส้นโค้งความถี่เป็นเส้นโค้งเบ้ลาดทางซ้ายหรือส่วนของเส้นโค้งที่มีความชันน้อยอยู่ทางด้านซ้าย ฐานนิยมจะมีค่ามากที่สุด รองลงมาเป็นมัธยฐาน และค่าเฉลี่ยเลขคณิตมีค่าน้อยที่สุด

เส้นโค้งปกติจะมีความโด่งมากหรือน้อยขึ้นอยู่กับการกระจายของข้อมูล ถ้าข้อมูลมีการกระจายมากเส้นโค้งปกติจะมีความโด่งน้อยหรือค่อนข้างแบน แต่ถ้าข้อมูลมีการกระจายน้อยเส้นโค้งปกติจะมีความโด่งมากหรือค่อนข้างสูงดังปรากฏในรูปต่อไปนี้



แบบฝึกหัด 1.3 ค

1. จงเปรียบเทียบการกระจายของอายุบุตรสองครอบครัว โดยที่อายุบุตรทั้งสองครอบครัวเป็นดังนี้

   อายุบุตรในครอบครัวที่หนึ่ง (ปี) 6, 5, 3, 1

   อายุบุตรในครอบครัวที่สอง (ปี) 25, 24, 22, 21, 17

   1) ใช้สัมประสิทธิ์ของพิสัย
   2) ใช้สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์
   3) ใช้สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย
   4) ใช้สัมประสิทธิ์ของการแปรผัน

   ผลของการเปรียบเทียบที่ได้จากการใช้วิธีทั้ง 4 นี้ เหมือนกันหรือไม่

2. ถ้าราคาข้าวเปลือกและราคาข้าวสารต่อถังของร้านค้าข้าวที่สำรวจมาเป็นตัวอย่าง 6 ร้าน ในท้องที่แห่งหนึ่ง เป็นดังนี้

ราคาข้าวเปลือก (บาท)	72	75	73	74	76	71
ราคาข้าวสาร (บาท)	115	118	112	114	117	110

   จงหาสัมประสิทธิ์ของการแปรผันและสัมประสิทธิ์ของพิสัยของราคาข้าวเปลือกและราคาข้าวสาร พร้อมทั้งเปรียบเทียบการกระจายของราคาข้าวเปลือกและราคาข้าวสาร

3. จากการสอบถามนักเรียนชั้น ป.2 ป.6 ม.3 และ ม.6 ของโรงเรียนแห่งหนึ่งถึงจำนวนเงินที่ผู้ปกครองให้มาใช้ที่โรงเรียนในแต่ละวันปรากฏว่า ค่าเฉลี่ยเลขคณิต และค่าความแปรปรวนของจำนวนเงินที่นักเรียนในแต่ละชั้นได้มาใช้เป็นดังนี้

	ป.2	ป.6	ม.3	ม.6
ค่าเฉลี่ยเลขคณิต (บาท)	18	20	22	25
ความแปรปรวน (บาท^2)	24	40	40	51

   จงเปรียบเทียบการกระจายของจำนวนเงินที่นักเรียนในแต่ละชั้นได้มาใช้ในแต่ละวันและอธิบายความหมายของค่าที่หาได้

4. สัมประสิทธิ์ของพิสัยของความสูงของนักเรียนในชั้นหนึ่งเป็น 0.0625 ถ้าความสูงของนักเรียนที่สูงที่สุดในชั้นเป็น 170 เซนติเมตร จงหาความสูงของนักเรียนคนที่เตี้ยที่สุดในชั้น

5. ข้อมูลชุดหนึ่งมีสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 0.12 และส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 8.5 จงหาสัมประสิทธิ์ของการแปรผัน ถ้าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

6. จงพิจารณาข้อความที่กำหนดให้ต่อไปนี้ว่าเป็นจริงหรือเป็นเท็จ ถ้าเป็นเท็จจงบอกเหตุผล
   1) พิสัยของข้อมูลชุดใด ๆ อาจจะมีค่ามากกว่าข้อมูลที่มีค่ามากที่สุดของข้อมูลชุดนั้นก็ได้
   2) ควอร์ไทล์ที่สองมีค่าเป็นสองเท่าของควอร์ไทล์ที่หนึ่ง และควอร์ไทล์ที่สามมีค่าเป็นสองเท่าของควอร์ไทล์ที่สอง
   3) ส่วนเบี่ยงเบนควอร์ไทล์ของข้อมูลชุดใดจะมีค่าเท่ากับมัธยฐานของข้อมูลชุดนั้นเสมอ
   4) ส่วนเบี่ยงเบนเฉลี่ยอาจมีค่าน้อยกว่าศูนย์ก็ได้
   5) ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดหนึ่งมีค่ามาก ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนั้นอาจจะมีค่าน้อยกว่าศูนย์ได้
   6) ส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของข้อมูลชุดเดียวกันอาจจะมีค่าเท่ากันก็ได้
   7) ถ้าทุก ๆ ค่าของข้อมูลชุดหนึ่งเท่ากัน ส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของข้อมูลชุดนั้นจะเท่ากับศูนย์เสมอ
   8) โดยทั่วไป การวัดการกระจายของข้อมูลโดยใช้ส่วนเบี่ยงเบนมาตรฐานมีความถูกต้องและเชื่อถือได้มากที่สุด เมื่อเทียบกับการวัดการกระจายแบบอื่น ๆ
   9) ถ้าข้อมูลชุดหนึ่งมีส่วนเบี่ยงเบนมาตรฐานมากกว่าข้อมูลอีกชุดหนึ่ง แสดงว่า ข้อมูลชุดนั้นมีการกระจายมากกว่าข้อมูลอีกชุดหนึ่งเสมอ

7. ถ้ามีข้อมูลผิดปกติ เช่น ค่าสูงหรือต่ำกว่าค่าส่วนใหญ่ จะมีผลกระทบทำให้ค่ากลางค่าใดและการวัดการกระจายค่าใดมีการเปลี่ยนแปลงไปมาก เนื่องจากค่าผิดปกตินั้น (outliers) แต่กลับไม่กระทบหรือมีผลกระทบน้อยต่อค่ากลางและค่าวัดการกระจายค่าใด จงอธิบาย

8. จงหาสัมประสิทธิ์ของการแปรผันของความเสียหายรวมของแต่ละจังหวัดจากตารางต่อไปนี้

จังหวัด	รวม	ความเสียหาย (ล้านบาท)				
		ที่ดิน	บ้าน/อาคาร สิ่งปลูกสร้าง	อุปกรณ์	ยานพาหนะ	อื่น ๆ
รวม	**1,942.8**	**262.5**	**772.8**	**372.8**	**304.0**	**230.7**
กระบี่	321.3	17.1	99.5	71.9	64.8	68.0
พังงา	1,077.4	122.6	557.3	154.0	161.0	82.5
ระนอง	203.3	102.1	26.2	36.8	30.6	7.6
ตรัง	43.0	10.1	5.9	14.4	9.9	2.7
ภูเก็ต	188.6	1.1	83.3	41.4	29.6	33.2
สตูล	109.2	9.5	0.6	54.3	8.1	36.7

ที่มา : สารสถิติปีที่ 16 ฉบับที่ 3 เดือนมีนาคม พ.ศ. 2548

9. จงหาสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของร้อยละของเมทิลแอลกอฮอล์จำแนกตามห้องปฏิบัติการและนำผลที่ได้มาเปรียบเทียบกัน จากตารางต่อไปนี้

หน่วยทดลอง	ห้องปฏิบัติการ			
	LAB1	LAB2	LAB3	LAB4
1	85.06	84.99	84.48	84.10
2	85.25	84.28	84.72	84.55
3	84.87	84.88	85.10	84.05

ที่มา : Devore and Farnum (2005)

บทที่ 2
การแจกแจงปกติ

ในบทที่ 1 เป็นการวิเคราะห์ข้อมูลเบื้องต้นโดยใช้ค่ากลางของข้อมูลและการวัดตำแหน่งที่หรือตำแหน่งสัมพัทธ์ ตลอดจนการกระจายของข้อมูล ซึ่งเป็นเรื่องจำเป็นเพราะการเลือกใช้ค่าสถิติและคำนวณได้ถูกต้องเหมาะสม สามารถบอกสารสนเทศเกี่ยวกับเรื่องที่สนใจศึกษาและนำไปเป็นข้อมูลสำหรับการตัดสินใจ หรือวางแผนต่อไปได้อย่างเหมาะสม

นอกจากเรื่องหลัก ๆ ในทางสถิติได้แก่ ค่ากลาง การวัดการกระจายของข้อมูลแล้ว เรื่องสำคัญอีกอย่างหนึ่ง คือ รูปแบบของการแจกแจงของข้อมูล รูปแบบหรือลักษณะของการแจกแจงข้อมูลบอกให้ทราบว่า ข้อมูลในเรื่องนั้นทั้งชุดมีลักษณะปกติ สมมาตร หรือมีความเบ้ หรือมีความโด่งอย่างไร หรือไม่สามารถจัดรูปแบบใด ๆ ได้เลย ผู้เรียนได้รู้จักลักษณะการแจกแจงของข้อมูลบ้างแล้ว โดยสังเกตจาก ฮิสโทแกรม แผนภาพต้น-ใบ เส้นโค้งความถี่ ฯลฯ เรื่องต่อไปควรศึกษาเพิ่มเติมเกี่ยวกับลักษณะการแจกแจงของข้อมูลอีกลักษณะหนึ่ง ซึ่งใช้ลักษณะของการแจกแจงหลักสำหรับการนำไปประยุกต์กับการแจกแจงลักษณะอื่น ๆ ต่อไปได้ นั่นคือ การแจกแจงปกติ (normal distribution)

โดยทั่วไป ถ้าต้องการหาสารสนเทศให้ครบถ้วน ในประเด็นหลัก ๆ ของสถิติเรื่องหนึ่งเรื่องใดแล้วควรพิจารณา 3 เรื่อง คือ ค่ากลาง การกระจาย และลักษณะการแจกแจงของข้อมูล ซึ่งเมื่อใช้ครบทั้งสามอย่างนี้แล้ว จะช่วยชี้ให้เห็นลักษณะของข้อมูลชัดเจนยิ่งขึ้นว่า ข้อมูลทั้งชุดบอกสารสนเทศอะไร และด้วยข้อมูลดังกล่าว ควรวางแผนหรือตัดสินใจทำอย่างไรต่อไป รูปแบบของการแจกแจงจึงเป็นเรื่องสำคัญเรื่องหนึ่ง

ประโยชน์ของการแจกแจงปกติที่ควรทราบในเบื้องต้น คือ การใช้การแจกแจงปกติมาช่วยในการหาความน่าจะเป็นของข้อมูลที่สนใจ ในบทที่ 2 นี้ จะกล่าวถึงการแจกแจงปกติและเส้นโค้งปกติ (normal curve) ซึ่งเกี่ยวข้องกับคะแนนมาตรฐาน (standard score หรือ z - score) คะแนนมาตรฐานมีประโยชน์หลายประการ เช่น ใช้ในการวัดตำแหน่งที่หรือตำแหน่งสัมพัทธ์ของข้อมูล ใช้ในการเปรียบเทียบข้อมูล และใช้เป็นพื้นฐานของการคำนวณพื้นที่ใต้เส้นโค้งปกติ

และความน่าจะเป็น ดังนั้น ควรศึกษาเรื่องการแจกแจงปกติไปพร้อม ๆ กับเรื่องคะแนนมาตรฐาน และการหาพื้นที่ใต้เส้นโค้งปกติด้วย

2.1 คะแนนมาตรฐาน

การเปรียบเทียบข้อมูลตั้งแต่สองข้อมูลขึ้นไปที่มาจากข้อมูลคนละชุดว่ามีความแตกต่างกันหรือไม่เพียงไร บางครั้งไม่สามารถเปรียบเทียบโดยตรงได้ ทั้งนี้เนื่องจากค่าเฉลี่ยเลขคณิตของข้อมูลแต่ละชุด และส่วนเบี่ยงเบนมาตรฐานมักจะไม่เท่ากัน เช่น ต้องการเปรียบเทียบผลการเรียนวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ของนักเรียนคนใดคนหนึ่งในชั้นว่าเรียนวิชาไหนได้ดีกว่ากัน แม้ว่าจะทำได้โดยดูจากคะแนนสอบของวิชาทั้งสอง โดยปรับให้มีคะแนนเต็มเท่ากัน ถ้าคะแนนสอบของวิชาไหนดีกว่าก็สรุปผลว่านักเรียนคนนั้นเรียนวิชานั้นได้ดีกว่า ซึ่งจะเห็นได้ว่าเป็นการสรุปผลที่ยังไม่ถูกต้องนักเพราะค่าเฉลี่ยเลขคณิตหรือส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาทั้งสองของนักเรียนทั้งหมดในชั้นอาจจะไม่เท่ากัน ทั้งนี้อาจเนื่องมาจากเนื้อหาหรือข้อสอบของทั้งสองวิชามีความยากง่ายต่างกัน หรือครูผู้สอนแต่ละวิชามีวิธีการสอนที่จะทำให้นักเรียนมีความเข้าใจในวิชานั้น ๆ ต่างกัน เป็นต้น ดังนั้นเพื่อที่จะให้การเปรียบเทียบมีความถูกต้องมากขึ้น จึงมีความจำเป็นต้องแปลงคะแนนของวิชาทั้งสองที่นักเรียนคนนั้นสอบได้ให้เป็นคะแนนมาตรฐานหรือค่ามาตรฐาน (ซึ่งมีค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานเท่ากันเสียก่อน) แล้วจึงเปรียบเทียบคะแนนมาตรฐานของวิชาทั้งสอง การแปลงค่าของข้อมูลของตัวแปรแต่ละตัวให้เป็นคะแนนมาตรฐานนี้ โดยทั่วไปคือการแปลงข้อมูลให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเลขคณิต 0 และส่วนเบี่ยงเบนมาตรฐาน 1

ถ้า x_i เป็นค่าที่ i ของตัวแปร X แล้ว คะแนนมาตรฐานของ x_i คือ

$$z_i = \frac{x_i - \mu}{\sigma} \quad \text{เมื่อ } i \text{ คือ } 1, 2, 3, \ldots, N$$

โดยที่ x_i แทน ค่าที่ i ของตัวแปร X

μ แทน ค่าเฉลี่ยเลขคณิตของประชากร

σ แทน ส่วนเบี่ยงเบนมาตรฐานของประชากร

N แทน จำนวนประชากร

หรือ $z_i = \dfrac{x_i - \overline{X}}{s}$ เมื่อ i คือ 1, 2, 3, ... , n

โดยที่ x_i แทน ค่าที่ i ของตัวแปร X

$\overline{X}$ แทน ค่าเฉลี่ยเลขคณิตของตัวอย่าง

s แทน ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

n แทน จำนวนตัวอย่าง

อาจกล่าวได้ว่าคะแนนมาตรฐานเป็นค่าที่บอกให้ทราบว่า ความแตกต่างระหว่างค่าของข้อมูลนั้น ๆ กับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้นว่าเป็นกี่เท่าของส่วนเบี่ยงเบนมาตรฐาน ดังนั้น คะแนนมาตรฐานจึงเป็นค่าวัดตำแหน่งที่ของข้อมูลหรือค่าวัดตำแหน่งสัมพัทธ์ของข้อมูลที่นิยมใช้อีกวิธีหนึ่งด้วย

ตัวอย่าง 1

ตัวอย่างที่ 1 นักเรียนคนหนึ่งสอบวิชาภาษาอังกฤษและวิชาคณิตศาสตร์ซึ่งมีคะแนนเต็ม 100 คะแนนเท่ากัน ได้ 72 คะแนน และ 75 คะแนน ตามลำดับ ถ้าค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนห้องนี้เป็น 70 และ 10 คะแนน ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์เป็น 73 และ 16 คะแนน ตามลำดับ จงเปรียบเทียบดูว่านักเรียนคนนี้เรียนวิชาไหนได้ดีกว่ากัน

วิธีทำ จากคะแนนมาตรฐานของ x_i คือ $z_i = \dfrac{x_i - \mu}{\sigma}$

ดังนั้น คะแนนมาตรฐานของคะแนนสอบวิชาภาษาอังกฤษ คือ $\dfrac{72-70}{10} = 0.2$

คะแนนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์คือ $\dfrac{75-73}{16} = 0.125$

เนื่องจากคะแนนมาตรฐานของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนคนนี้สูงกว่าคะแนนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ ดังนั้นนักเรียนคนนี้เรียนวิชาภาษาอังกฤษได้ดีกว่าวิชาคณิตศาสตร์

ตัวอย่างที่ 2 คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่งมีค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานเป็น 73 และ 16 คะแนน ตามลำดับ ถ้าคะแนนมาตรฐานของคะแนนสอบวิชานี้ของนักเรียนคนหนึ่งในห้องนี้ คือ 0.2 อยากทราบว่านักเรียนคนนี้สอบได้กี่คะแนน

วิธีทำ จากคะแนนมาตรฐานของ x_i คือ $z_i = \frac{x_i - \mu}{\sigma}$

จะได้ $0.2 = \frac{x - 73}{16}$

ดังนั้น $x = (0.2 \times 16) + 73 = 76.2$

นั่นคือ นักเรียนคนนี้สอบวิชาคณิตศาสตร์ได้ 76.2 คะแนน

ตัวอย่างที่ 3 บริษัทแห่งหนึ่งซึ่งเป็นตัวแทนจำหน่ายสินค้า 3 ชนิด พนักงาน 2 คนของบริษัทนี้ขายสินค้าทั้งสามชนิด ได้ยอดรวมเท่ากันดังตาราง จงพิจารณาว่าใครขายเก่งกว่ากัน

พนักงาน	จำนวนสินค้าที่ขายได้ (ชิ้น)		
	ชนิดที่ 1	ชนิดที่ 2	ชนิดที่ 3
คนที่ 1	10	4	7
คนที่ 2	6	7	8
ค่าเฉลี่ยเลขคณิต	7	6	10
ส่วนเบี่ยงเบนมาตรฐาน	3	2	1

วิธีทำ ให้ x_i แทนจำนวนสินค้าชนิดที่ i เมื่อ i คือ 1, 2, 3

μ_i แทนค่าเฉลี่ยเลขคณิตของจำนวนสินค้าชนิดที่ i

σ_i แทนส่วนเบี่ยงเบนมาตรฐานของจำนวนสินค้าชนิดที่ i

จะได้ $\mu_1 = 7, \sigma_1 = 3$; $\mu_2 = 6, \sigma_2 = 2$ และ $\mu_3 = 10, \sigma_3 = 1$

จากคะแนนมาตรฐานของ x_i คือ $z_i = \frac{x_i - \mu_i}{\sigma_i}$

หาคะแนนมาตรฐานของจำนวนสินค้าที่พนักงานคนที่ 1 ขายได้ ดังนี้

$z_1 = \frac{10-7}{3} = 1,\ z_2 = \frac{4-6}{2} = -1$ และ $z_3 = \frac{7-10}{1} = -3$

ดังนั้น คะแนนมาตรฐานเฉลี่ยของจำนวนสินค้าชนิดที่ 1, 2 และ 3 ที่พนักงานคนที่ 1 ขายได้ คือ $\frac{1+(-1)+(-3)}{3} = -1$

หาคะแนนมาตรฐานของจำนวนสินค้าที่พนักงานคนที่ 2 ขายได้ ดังนี้

$z_1 = \frac{6-7}{3} = \frac{-1}{3}$, $z_2 = \frac{7-6}{2} = \frac{1}{2}$ และ $z_3 = \frac{8-10}{1} = -2$

ดังนั้น คะแนนมาตรฐานเฉลี่ยของจำนวนสินค้าชนิดที่ 1, 2 และ 3 ที่พนักงานคนที่ 2 ขายได้ คือ $\frac{(-\frac{1}{3})+\frac{1}{2}+(-2)}{3} = -\frac{11}{18}$

เนื่องจากคะแนนมาตรฐานเฉลี่ยของจำนวนสินค้าทั้งสามชนิดที่พนักงานคนที่ 2 ขายได้มากกว่า

คะแนนมาตรฐานเฉลี่ยของจำนวนสินค้าทั้งสามชนิดที่พนักงานคนที่ 1 ขายได้

ดังนั้น สรุปได้ว่า พนักงานคนที่ 2 ขายสินค้าทั้งสามชนิดได้เก่งกว่าพนักงานคนที่ 1

ข้อสังเกตเกี่ยวกับคะแนนมาตรฐาน

1. คะแนนมาตรฐานของข้อมูลใด ๆ จะเป็นจำนวนบวกหรือจำนวนลบก็ได้ ทั้งนี้ขึ้นอยู่กับค่าของข้อมูลนั้น ๆ กับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้นว่าค่าใดจะมากกว่ากัน

2. คะแนนมาตรฐานของข้อมูลที่มีการแจกแจงปกติหรือใกล้เคียงปกติ โดยทั่วไปจะมีค่าตั้งแต่ –3 ถึง 3 แต่อาจจะมีคะแนนมาตรฐานของข้อมูลบางค่าที่น้อยกว่า –3 หรือมากกว่า 3 ได้

3. เมื่อแปลงทุก ๆ ค่าในข้อมูลชุดใดชุดหนึ่งที่เป็นข้อมูลระดับประชากรให้เป็นคะแนนมาตรฐาน แล้วนำคะแนนมาตรฐานเหล่านี้มาคำนวณหาค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐาน จะได้ค่าเฉลี่ยเลขคณิต 0 และส่วนเบี่ยงเบนมาตรฐาน 1 เสมอ

แบบฝึกหัด 2.1

1. จงหาค่าของ X จากสูตรของคะแนนมาตรฐาน โดยใช้ข้อมูลต่อไปนี้
   1) $Z = 2$ ค่าเฉลี่ยเลขคณิต 20 ส่วนเบี่ยงเบนมาตรฐาน 5
   2) $Z = -1$ ค่าเฉลี่ยเลขคณิต 25 ส่วนเบี่ยงเบนมาตรฐาน 3
   3) $Z = -1.5$ ค่าเฉลี่ยเลขคณิต 100 ส่วนเบี่ยงเบนมาตรฐาน 10
   4) $Z = 2.5$ ค่าเฉลี่ยเลขคณิต -10 ส่วนเบี่ยงเบนมาตรฐาน 0.2

2. ด.ช.วิชัย สอบได้คะแนนวิชาคณิตศาสตร์ในชั้น ม.3 และ ม.4 เป็น 75 คะแนนและ 80 คะแนน ตามลำดับ ถ้าค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนวิชาคณิตศาสตร์ของนักเรียนทุกคนในชั้น ม.3 ที่ ด.ช.วิชัย เรียนอยู่เป็น 70 คะแนน และ 15 คะแนน และค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของนักเรียนทุกคนในชั้น ม.4 เป็น 80 คะแนนและ 20 คะแนน ตามลำดับ ด.ช.วิชัย เรียนวิชาคณิตศาสตร์ในชั้นไหนได้ดีกว่ากัน

3. ในการทดสอบเวลาที่ใช้ในการวิ่งแข่งระยะทาง 100 เมตรของนักกีฬาในโรงเรียนแห่งหนึ่ง เพื่อคัดเลือกตัวแทนไปแข่งขันกับโรงเรียนอื่นโดยถือว่าผู้ที่ผ่านการทดสอบจะต้องได้คะแนนมาตรฐานของเวลาที่ใช้ไม่มากกว่า 1.0 ถ้าจากผลการทดสอบปรากฏว่านักกีฬาที่ใช้เวลามากกว่า 12 วินาที ไม่ผ่านการทดสอบ ถามว่าในการทดสอบคราวนี้ค่าเฉลี่ยเลขคณิตของเวลาที่ใช้ในการวิ่งของนักกีฬาทั้งหมดเป็นเท่าไร ถ้าส่วนเบี่ยงเบนมาตรฐานของเวลาที่ใช้ในการวิ่งของนักกีฬาเป็น 1.1 วินาที

4. ถ้าคะแนนสอบวิชาต่าง ๆ ของ ด.ญ.จิตรา ค่าเฉลี่ยเลขคณิต และส่วนเบี่ยงเบนมาตรฐานของคะแนนแต่ละวิชาของนักเรียนทั้งหมดในชั้นที่ ด.ญ.จิตรา เรียนอยู่เป็นดังนี้

วิชา	คะแนนที่สอบได้	ค่าเฉลี่ยเลขคณิต	ส่วนเบี่ยงเบนมาตรฐาน
ภาษาไทย	80	85	15
ภาษาอังกฤษ	60	75	20
วิทยาศาสตร์	70	65	5

ด.ญ.จิตรา เรียนวิชาไหนได้ดีกว่ากัน

5. ในโรงงานอุตสาหกรรมแห่งหนึ่งต้องการรับสมัครคนงาน โดยมีข้อแม้ว่า คนงานที่บริษัทจะรับเข้าทำงานจะต้องมีคะแนนมาตรฐานของอายุตั้งแต่ 2.0 ขึ้นไป ถ้าค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของอายุของคนงานทั้งหมดที่มาสมัครเข้าทำงานเป็น 25 ปี และ 2 ปี ตามลำดับ คนงานที่มีอายุตั้งแต่เท่าไรขึ้นไปจึงจะมีโอกาสได้รับเลือกเข้าเป็นคนงานของโรงงานอุตสาหกรรมนี้

6. ในการสอบคัดเลือกเข้าทำงานในหน่วยงานแห่งหนึ่งซึ่งมีวิชาที่จะต้องสอบ 3 วิชา ถ้าผู้สมัครเข้าสอบคัดเลือกจำนวน 2 คน คือ นาย ก และนางสาว ข ได้คะแนนในแต่ละวิชาเป็นดังนี้

	วิชาที่ 1	วิชาที่ 2	วิชาที่ 3
นาย ก	70	75	75
นางสาว ข	75	50	95

จงหาว่า นาย ก หรือ นางสาว ข ใครได้ตำแหน่งที่ในการสอบคัดเลือกดีกว่ากัน ถ้าค่าเฉลี่ยเลขคณิตของวิชาที่ 1 วิชาที่ 2 และวิชาที่ 3 ของคะแนนของผู้สมัครสอบทั้งหมดเป็น 70, 70 และ 80 คะแนน และส่วนเบี่ยงเบนมาตรฐานเป็น 5, 10 และ 15 คะแนน ตามลำดับ ถ้าหน่วยงานแห่งนี้ตั้งหลักเกณฑ์ไว้ว่า ผู้ที่จะได้รับเลือกเข้าทำงานจะต้องได้คะแนนมาตรฐานเฉลี่ยของคะแนนทั้งสามวิชา ไม่ต่ำกว่า 0 ถามว่า นาย ก และนางสาว ข จะได้รับเลือกเข้าทำงานหรือไม่

7. ในการสอบแข่งขันชิงทุนการศึกษา นายประพันธ์ สอบได้ที่ 1 และได้คะแนน 650 คะแนน นางสาวมะลิวัลย์ สอบได้ที่ 10 และได้คะแนน 540 คะแนน ถ้าคะแนนมาตรฐานของนายประพันธ์ และ นางสาวมะลิวัลย์ เป็น 3 และ 1.9 ตามลำดับ จงหาค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบครั้งนี้

8. ประเทศสหรัฐอเมริกาให้ข้อมูลผู้ป่วยโรคหัวใจและโรคมะเร็งต่อประชากร 100,000 คนต่อปีของ 50 รัฐ พบว่ามีผู้ป่วยเสียชีวิตด้วยโรคดังกล่าวโดยแสดงเป็นค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานดังนี้

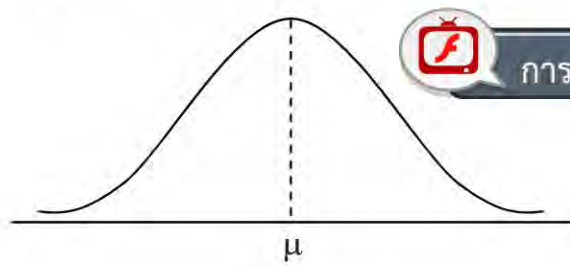
	ค่าเฉลี่ยเลขคณิต	ส่วนเบี่ยงเบนมาตรฐาน
โรคหัวใจ	289	54
โรคมะเร็ง	200	31

1) ถ้าในรัฐอลาสกา (Alaska) มีผู้ป่วยเสียชีวิตด้วยโรคหัวใจ 90 คนต่อประชากร 100,000 คน โรคหัวใจในรัฐอลาสกาจะมีความรุนแรงมากหรือน้อยกว่ารัฐอื่น ๆ หรือไม่
2) ถ้าในรัฐแคลิฟอร์เนีย (California) มีผู้ป่วยเสียชีวิตด้วยโรคหัวใจ 240 คน และ โรคมะเร็ง 166 คน ต่อประชากร 100,000 คนเช่นกัน โรคใดมีความรุนแรงกว่ากันเมื่อเทียบกับที่พบในรัฐอื่น ๆ

2.2 การแจกแจงปกติและเส้นโค้งปกติ

จากเรื่องฮิสโทแกรม พื้นที่ของรูปสี่เหลี่ยมผืนผ้าแต่ละรูปในฮิสโทแกรมแทนความถี่ของแต่ละอันตรภาคชั้น ถ้าเขียนรูปหลายเหลี่ยมของความถี่และปรับรูปหลายเหลี่ยมของความถี่ให้เป็นเส้นโค้งเรียบ จะได้เส้นโค้งความถี่ซึ่งพื้นที่ใต้เส้นโค้งความถี่จะแทนความถี่ของค่าจากการสังเกตทั้งหมด

เส้นโค้งความถี่ที่พบบ่อยมักมีลักษณะเป็นรูประฆังซึ่งเรียกว่า เส้นโค้งปกติ การแจกแจงความถี่ของข้อมูลซึ่งเส้นโค้งที่ได้มีลักษณะเป็นรูประฆัง เรียกว่า การแจกแจงปกติ สมการของเส้นโค้งนี้ขึ้นอยู่กับค่า 2 ค่า คือ ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐาน ถ้ากำหนดค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานให้จะสามารถหาสมการของเส้นโค้งปกติได้และเขียนรูปได้ดังในรูปที่ 1



รูปที่ 1

จากรูปที่ 1 จะเห็นว่า ลักษณะของเส้นโค้งปกติเป็นรูประฆัง ซึ่งเป็นรูปสมมาตรโดยมีเส้นประเป็นแกนสมมาตร จึงเรียกว่า การแจกแจงปกติ ขนาดของการแจกแจงปกติหรือรูประฆัง

จะต่างกันเพียงไรขึ้นอยู่กับค่าเฉลี่ยเลขคณิต μ และส่วนเบี่ยงเบนมาตรฐาน σ (ในกรณีข้อมูลประชากร) และสมการของเส้นโค้งปกติ คือ

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{เมื่อ } -\infty < x < \infty$$

$$\text{หรือ} \quad y_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \qquad \text{เมื่อ } i \text{ คือ } 1, 2, 3, \ldots, N$$

โดยที่ $\pi \approx 3.1416$, $e \approx 2.718$

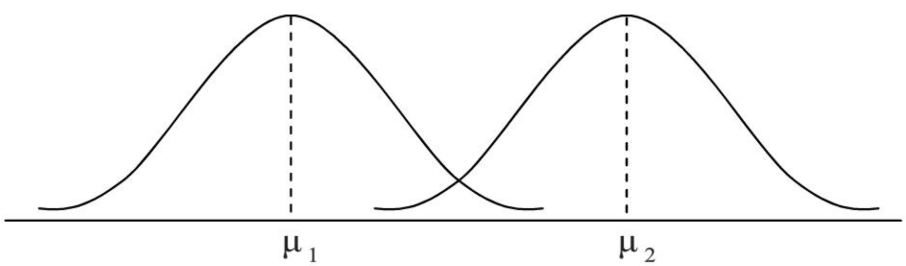
μ แทนค่าเฉลี่ยเลขคณิตของประชากร

σ แทนส่วนเบี่ยงเบนมาตรฐานของประชากร

N แทนจำนวนประชากร

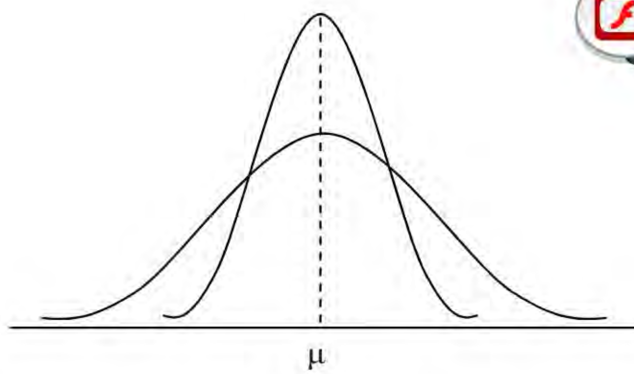
ในการเขียนกราฟ ให้แทนข้อมูล x_i หนึ่งค่าลงในสมการข้างต้นซึ่งกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของประชากรให้ จะได้ค่า y_i ที่สอดคล้องกับค่า x_i เป็นคู่ ๆ และเมื่อแทนค่า x_i ทุกค่า จะได้ ค่า y_i ทุกค่า นำค่า (x_i, y_i) ไปเขียนกราฟ จะได้กราฟของการแจกแจงปกติ ดังรูปที่ 1

เส้นโค้งปกติ 2 รูป ซึ่งมีส่วนเบี่ยงเบนมาตรฐานเท่ากัน แต่ค่าเฉลี่ยเลขคณิตไม่เท่ากันจะมีลักษณะเหมือนกันแต่ตั้งอยู่บนตำแหน่งที่ต่างกัน ดังรูปที่ 2



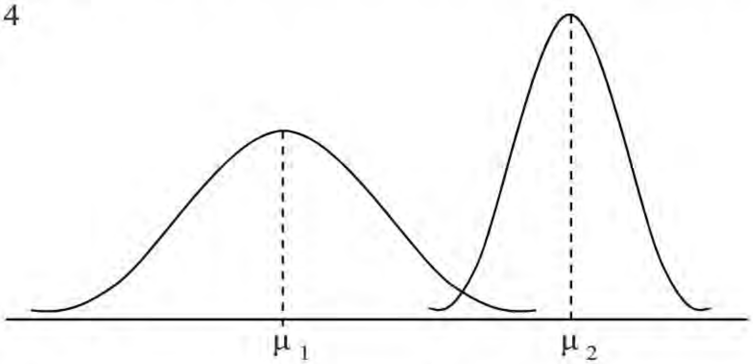
รูปที่ 2

เส้นโค้งปกติ 2 รูป ซึ่งมีค่าเฉลี่ยเลขคณิตเท่ากันแต่ส่วนเบี่ยงเบนมาตรฐานต่างกัน จะมีจุดที่แสดงค่าเฉลี่ยเลขคณิตอยู่ที่ตำแหน่งเดียวกันบนแกนนอน แต่เส้นโค้งที่มีส่วนเบี่ยงเบนมาตรฐานมากกว่าจะเตี้ยกว่า ดังนั้น ถ้าข้อมูลมีการกระจายมากเส้นโค้งจะเตี้ยลงและขยายฐานกว้างขึ้นดังรูปที่ 3



รูปที่ 3

เส้นโค้งปกติ 2 รูป ซึ่งมีค่าเฉลี่ยเลขคณิตต่างกันและส่วนเบี่ยงเบนมาตรฐานต่างกันมีลักษณะดังรูปที่ 4



รูปที่ 4

สมบัติของเส้นโค้งปกติ

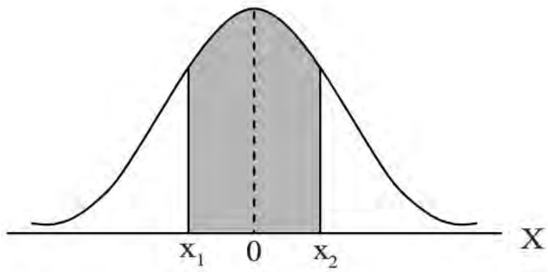
1) ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม จะเท่ากัน และจะอยู่ ณ จุดที่เส้นตรงที่ลากผ่านจุดโด่งสุดของเส้นโค้งนั้นตั้งฉากกับแกนนอน

2) เส้นโค้งจะมีเส้นตั้งฉากกับแกนนอนที่ลากผ่านค่าเฉลี่ยเลขคณิตเป็นแกนสมมาตร

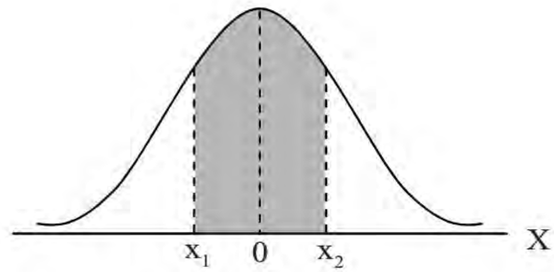
3) เส้นโค้งจะเข้าใกล้แกนนอน เมื่อต่อปลายเส้นโค้งทั้งสองข้างให้ห่างจากค่าเฉลี่ยเลขคณิตออกไป แต่จะไม่ตัดแกนนอน

4) พื้นที่ใต้เส้นโค้งปกติมีค่าเท่ากับ 1 เสมอ

5) พื้นที่ที่อยู่เหนือค่าใดค่าหนึ่งของ X จะเป็น 0 เสมอ จะได้ว่าพื้นที่ใต้เส้นโค้งปกติซึ่งอยู่ระหว่างค่าของ X ในช่วงปิด $[x_1, x_2]$ จะเท่ากับพื้นที่ใต้เส้นโค้งปกติซึ่งอยู่ระหว่างค่าของ X ในช่วงเปิด (x_1, x_2)



พื้นที่ใต้เส้นโค้งปกติซึ่งอยู่ระหว่าง
ค่าของ X ในช่วงปิด $[x_1, x_2]$

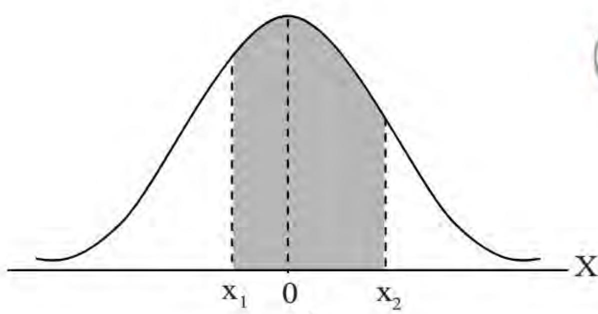


พื้นที่ใต้เส้นโค้งปกติซึ่งอยู่ระหว่าง
ค่าของ X ในช่วงเปิด (x_1, x_2)

พื้นที่ใต้เส้นโค้งปกติ

ถ้าทราบสมการของเส้นโค้งปกติ จะสามารถหาพื้นที่ใต้เส้นโค้งปกติซึ่งอยู่ระหว่างค่า X สองค่าใด ๆ คือ x_1 และ x_2 ได้ โดยใช้วิธีการของแคลคูลัสซึ่งค่อนข้างยุ่งยาก

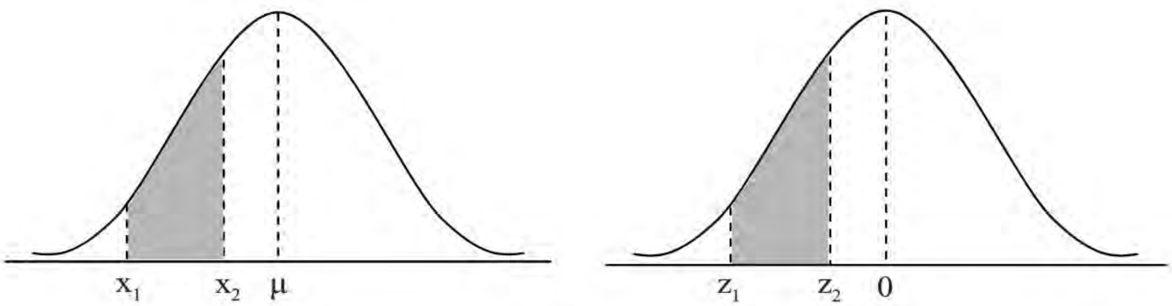
พื้นที่ใต้เส้นโค้งปกติ



รูปที่ 5

ในทางปฏิบัติ จะหาพื้นที่ใต้เส้นโค้งปกติได้โดยใช้ตารางแสดงพื้นที่ใต้เส้นโค้งปกติ แต่เนื่องจากเป็นไปไม่ได้ที่จะสร้างตารางหลาย ๆ ตารางมาแสดงพื้นที่ใต้เส้นโค้งปกติซึ่งมีค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานต่างกัน ดังนั้น จึงใช้วิธีแปลงค่า X ให้เป็นคะแนนมาตรฐาน Z โดยใช้สูตร $Z = \frac{X-\mu}{\sigma}$ ซึ่งจะพิสูจน์ได้ว่าค่าเฉลี่ยเลขคณิตของ Z คือ 0 และส่วนเบี่ยงเบนมาตรฐานของ Z คือ 1

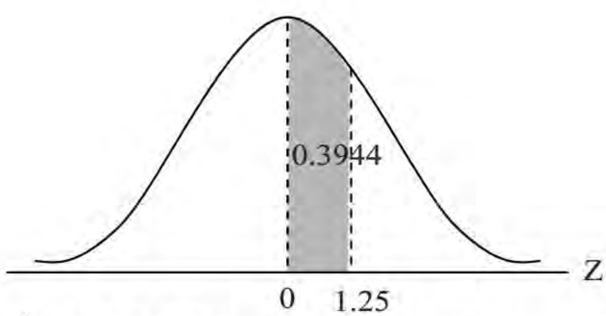
พื้นที่ใต้เส้นโค้งปกติระหว่าง x_1 และ x_2 จะเท่ากับพื้นที่ใต้เส้นโค้งปกติระหว่าง z_1 และ z_2 เมื่อ $z_1 = \frac{x_1-\mu}{\sigma}$ และ $z_2 = \frac{x_2-\mu}{\sigma}$



รูปที่ 6

หมายเหตุ เนื่องจากตารางที่ใช้ในการหาพื้นที่ใต้เส้นโค้งปกติมีหลายรูปแบบ แต่ในหนังสือเรียนเล่มนี้จะใช้ตารางแสดงพื้นที่ใต้เส้นโค้งปกติที่แสดงเฉพาะค่าของ Z ที่เป็นจำนวนบวก

เส้นโค้งปกติซึ่งได้จากชุดข้อมูลที่มีค่าเฉลี่ยเลขคณิตเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 เรียกว่า เส้นโค้งปกติมาตรฐาน ในการหาพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 ถึง Z ใด ๆ จะใช้ตารางแสดงพื้นที่ใต้เส้นโค้งปกติมาตรฐาน ซึ่งแสดงพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และค่าอื่น ๆ ของ Z คือ 0.01, 0.02, 0.03, ... , 3.88, 3.89 เช่น พื้นที่ใต้เส้นโค้งปกติระหว่างคะแนนมาตรฐาน 0 และ 1.25 ที่อ่านได้จากตารางคือ 0.3944

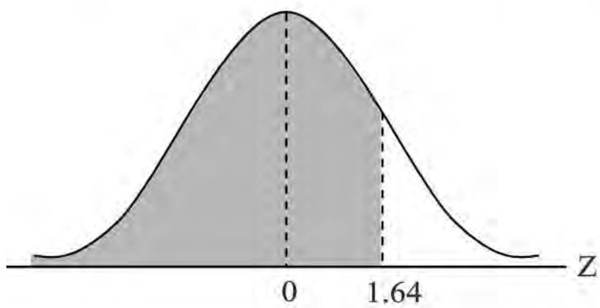


รูปที่ 7 พื้นที่ใต้เส้นโค้งปกติมาตรฐาน ในช่วง $0 < Z < 1.25$

ตารางพื้นที่ใต้เส้นโค้งปกติมาตรฐานที่แสดงในภาคผนวกไม่มีค่า Z ที่เป็นจำนวนลบ แต่สามารถหาพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่าง Z ที่เป็นจำนวนลบและศูนย์ได้ เนื่องจากเส้นที่ตั้งฉากกับแกนนอนที่ลากผ่านคะแนนมาตรฐาน 0 เป็นแกนสมมาตรของเส้นโค้งปกติ เช่น

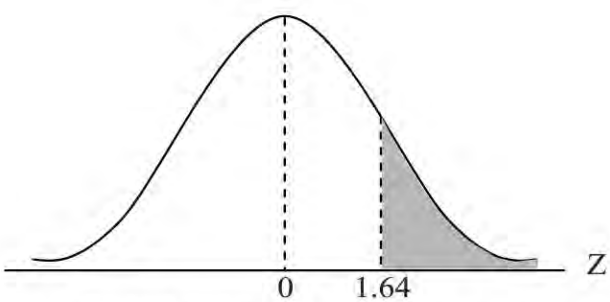
พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน –1.25 และ 0 หาได้จากการหาพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 ถึง 1.25 ซึ่งเท่ากับ 0.3944 เช่นเดียวกัน

เนื่องจากพื้นที่ใต้เส้นโค้งปกติมาตรฐานเท่ากับ 1 ดังนั้นพื้นที่ทางขวามือของคะแนนมาตรฐาน 0 กับพื้นที่ทางซ้ายมือของคะแนนมาตรฐาน 0 เท่ากันคือ 0.5 อาศัยความรู้ดังกล่าวจะหาพื้นที่ใต้เส้นโค้งปกติมาตรฐานทางขวามือหรือซ้ายมือของค่า Z ใด ๆ และพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างค่า Z สองค่าใด ๆ ได้ เช่น



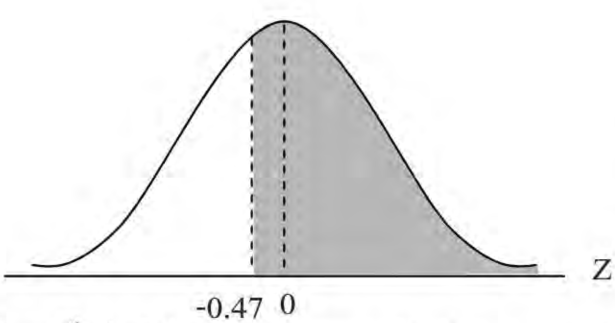
รูปที่ 8 พื้นที่ใต้เส้นโค้งปกติมาตรฐานในช่วง $Z < 1.64$

พื้นที่ใต้เส้นโค้งปกติมาตรฐานทางซ้ายมือของคะแนนมาตรฐาน 1.64 เท่ากับพื้นที่ใต้เส้นโค้งโค้งปกติมาตรฐานเมื่อ $Z < 1.64$ ซึ่งเท่ากับ 0.5 บวกกับพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 ถึง 1.64 จะได้พื้นที่เท่ากับ $0.5 + 0.4495 = 0.9495$



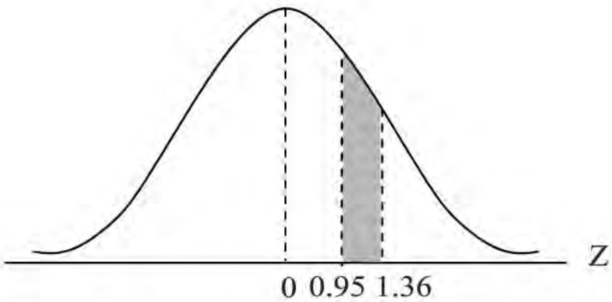
รูปที่ 9 พื้นที่ใต้เส้นโค้งปกติมาตรฐานในช่วง $Z > 1.64$

พื้นที่ใต้เส้นโค้งปกติมาตรฐานทางขวามือของคะแนนมาตรฐาน 1.64 เท่ากับพื้นที่ใต้เส้นโค้งปกติมาตรฐาน เมื่อ $Z > 1.64$ ซึ่งเท่ากับ 0.5 ลบด้วยพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 1.64 คือ $0.5 - 0.4495 = 0.0505$



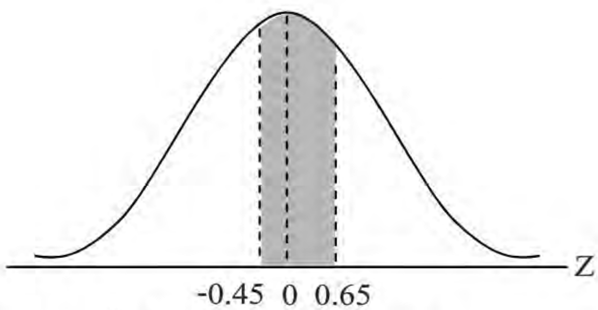
รูปที่ 10 พื้นที่ใต้เส้นโค้งปกติมาตรฐานในช่วง $Z > -0.47$

พื้นที่ใต้เส้นโค้งปกติมาตรฐานทางขวามือของคะแนนมาตรฐาน –0.47 เท่ากับ 0.5 บวกกับพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน –0.47 และ 0 คือ $0.5 + 0.1808 = 0.6808$



รูปที่ 11 พื้นที่ใต้เส้นโค้งปกติมาตรฐาน ในช่วง $0.95 < Z < 1.36$

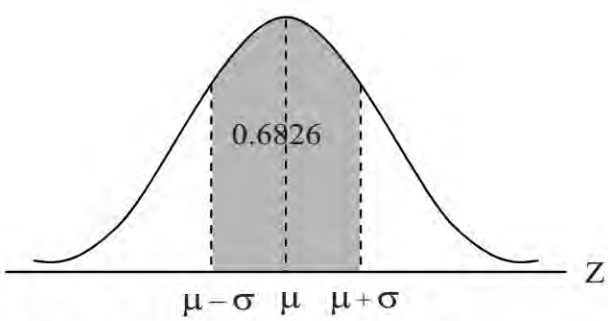
พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0.95 และ 1.36 คือ ผลต่างของพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 1.36 กับพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 0.95 นั่นคือ พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0.95 และ 1.36 คือ $0.4131 - 0.3289 = 0.0842$



รูปที่ 12 พื้นที่ใต้เส้นโค้งปกติมาตรฐาน ในช่วง $-0.45 < Z < 0.65$

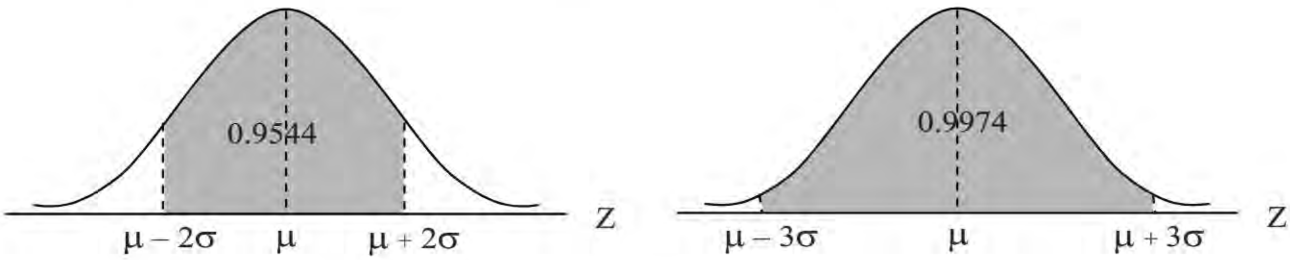
พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน -0.45 และ 0.65 คือ พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน -0.45 และ 0 บวกพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 0.65 คือ $0.1736 + 0.2422 = 0.4158$

ข้อสังเกต เนื่องจากพื้นที่ใต้เส้นโค้งปกติเท่ากับ 1 พิจารณาพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่าง $\mu - \sigma$ ถึง $\mu + \sigma$ ระหว่าง $\mu - 2\sigma$ ถึง $\mu + 2\sigma$ และระหว่าง $\mu - 3\sigma$ ถึง $\mu + 3\sigma$ ได้ดังนี้



รูปที่ 13 พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่าง $\mu - \sigma$ ถึง $\mu + \sigma$

พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่าง $\mu - \sigma$ ถึง $\mu + \sigma$ เท่ากับ 0.6826 หรือประมาณ 68.26% ของพื้นที่ทั้งหมดใต้เส้นโค้งปกติมาตรฐาน ซึ่งหมายความว่าประมาณ 68.26% ของจำนวนข้อมูลทั้งหมดตกอยู่ในช่วง $\mu - \sigma$ ถึง $\mu + \sigma$



รูปที่ 14 พื้นที่ใต้เส้นโค้งปกติมาตรฐาน ระหว่าง $\mu \pm 2\sigma$ และระหว่าง $\mu \pm 3\sigma$

พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่าง $\mu - 2\sigma$ และ $\mu + 2\sigma$ เท่ากับ 0.9544 หรือประมาณ 95.44% ของพื้นที่ทั้งหมด และพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่าง $\mu - 3\sigma$ และ $\mu + 3\sigma$ เท่ากับ 0.9974 หรือประมาณ 99.74% ของพื้นที่ทั้งหมด นั่นคือ ถ้าการแจกแจงข้อมูลเป็นเส้นโค้งปกติ ค่าของข้อมูลเกือบทั้งหมดจะตกอยู่ในช่วง $\mu - 3\sigma$ ถึง $\mu + 3\sigma$

ตัวอย่างที่ 1 คะแนนสอบของนักเรียนกลุ่มหนึ่งมีการแจกแจงปกติโดยมีค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐาน 50 และ 10 คะแนน ตามลำดับ อยากทราบว่า นักเรียนที่สอบได้ 60 คะแนน จะมีตำแหน่งเปอร์เซ็นไทล์ที่เท่าไร

วิธีทำ ให้ x เป็นคะแนนสอบของนักเรียน

จาก $Z = \dfrac{X - \mu}{\sigma}$

จะได้ $Z = \dfrac{60 - 50}{10} = 1$

จากตาราง พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 1 เท่ากับ 0.3413

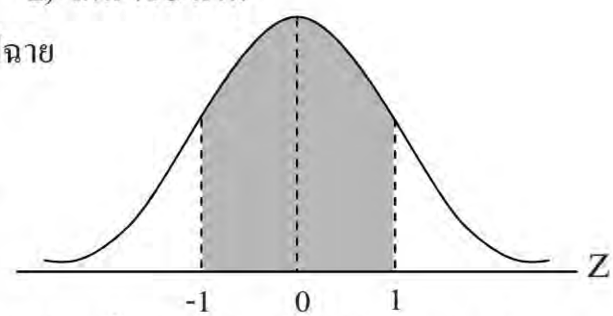
ดังนั้น พื้นที่ใต้เส้นโค้งปกติมาตรฐานเมื่อ $Z < 1$ เท่ากับ $0.5 + 0.3413 = 0.8413$ หรือ 84.13%

ดังนั้น ตำแหน่งเปอร์เซ็นไทล์ของคะแนน 60 คือ 84.13

ตัวอย่างที่ 2 อายุการใช้งานของถ่านไฟฉายชนิดหนึ่งมีการแจกแจงปกติ โดยค่าเฉลี่ยเลขคณิตเป็น 756 นาที และส่วนเบี่ยงเบนมาตรฐานเป็น 35 นาที จงหาเปอร์เซ็นต์ของถ่านไฟฉายที่ใช้ได้นาน

1) ระหว่าง 721 ถึง 791 นาที　　2) เกิน 798 นาที

วิธีทำ 1) ให้ X เป็นอายุการใช้งานของถ่านไฟฉาย



$$Z = \frac{X-\mu}{\sigma}$$

จะได้ $z_1 = \frac{721-756}{35} = -1$

และ $z_2 = \frac{791-756}{35} = 1$

จากตารางพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 1 เท่ากับ 0.3413

และพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ -1 เท่ากับ 0.3413

ดังนั้น พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน -1 และ 1 คือ $0.3413 + 0.3413 = 0.6826$

นั่นคือ มีถ่านไฟฉาย 68.26% ที่ใช้ได้นานระหว่าง 721 นาที ถึง 791 นาที

2) เมื่ออายุการใช้งานของถ่านไฟฉายเป็น 798 นาที

จะได้ $Z = \frac{798-756}{35} = 1.2$

จะได้ พื้นที่ใต้เส้นโค้งปกติมาตรฐาน เมื่อ $Z > 1.2$ เท่ากับ $0.5 - 0.3849 = 0.1151$

นั่นคือ มีถ่านไฟฉาย 11.51% ที่ใช้ได้นานเกิน 798 นาที

ตัวอย่างที่ 3 ครูคนหนึ่งให้ระดับคะแนนนักเรียนในการสอบดังนี้

ระดับคะแนน A ถ้านักเรียนได้คะแนนตั้งแต่ $\mu+1.5\sigma$ ขึ้นไป

ระดับคะแนน B ถ้านักเรียนได้คะแนนอยู่ในช่วง $[\mu+0.5\sigma,\ \mu+1.5\sigma]$

ระดับคะแนน C ถ้านักเรียนได้คะแนนอยู่ในช่วง $[\mu-0.5\sigma,\ \mu+0.5\sigma]$

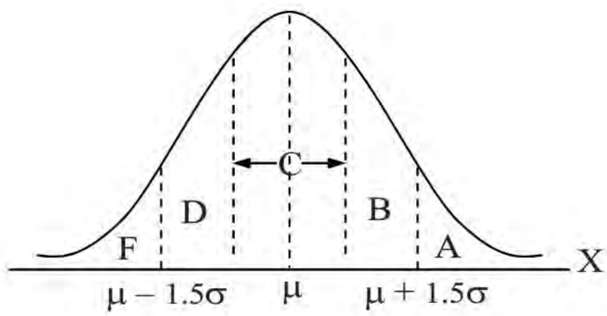
ระดับคะแนน D ถ้านักเรียนได้คะแนนอยู่ในช่วง $[\mu-1.5\sigma,\ \mu-0.5\sigma]$

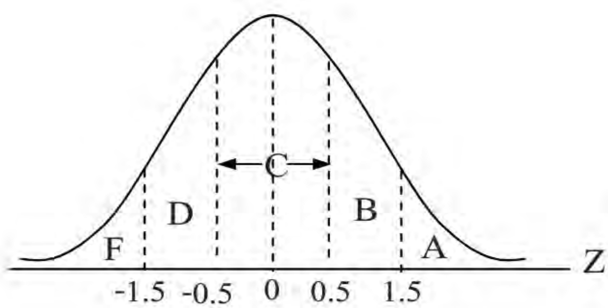
ระดับคะแนน F ถ้านักเรียนได้คะแนนน้อยกว่า $\mu-1.5\sigma$

สมมติว่าคะแนนสอบมีการแจกแจงปกติ อยากทราบว่าแต่ละระดับคะแนนมีนักเรียนสอบได้กี่เปอร์เซ็นต์

วิธีทำ แปลงคะแนนดิบให้เป็นคะแนนมาตรฐาน z ดังนี้

หาคะแนนมาตรฐานของคะแนน $\mu+1.5\sigma$, $\mu+0.5\sigma$, $\mu-0.5\sigma$ และ $\mu-1.5\sigma$ จากสูตร $Z = \dfrac{X-\mu}{\sigma}$ จะได้คะแนนมาตรฐานเป็น 1.5, 0.5, –0.5 และ –1.5 ตามลำดับ





จากตาราง พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 1.5 เท่ากับ 0.4332

ดังนั้น พื้นที่ใต้เส้นโค้งปกติมาตรฐานเมื่อ $Z > 1.5$ คือ $0.5 - 0.4332 = 0.0668$ หรือ 6.68% นั่นคือ มีนักเรียน 6.68% ได้ระดับคะแนน A

พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0 และ 0.5 คือ 0.1915

ดังนั้น พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน 0.5 และ 1.5 คือ $0.4332 - 0.1915 = 0.2417$ หรือ 24.17% นั่นคือ มีนักเรียน 24.17% ที่ได้ระดับคะแนน B

พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน –0.5 และ 0.5 คือ $0.1915 + 0.1915 = 0.3830$ หรือ 38.30% นั่นคือ มีนักเรียน 38.30% ที่ได้ระดับคะแนน C

เนื่องจากเส้นโค้งปกติเป็นรูปสมมาตร โดยมีเส้นตรงที่ตั้งฉากที่คะแนนมาตรฐาน 0 เป็นแกนสมมาตร ดังนั้น พื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างคะแนนมาตรฐาน –1.5 และ –0.5 เท่ากับ 0.2417 หรือ 24.17% นั่นคือ มีนักเรียน 24.17% ที่ได้ระดับคะแนน D

พื้นที่ใต้เส้นโค้งปกติมาตรฐานเมื่อ $Z < -1.5$ เท่ากับ $0.5 - 0.4332 = 0.0668$ หรือ 6.68% นั่นคือ มีนักเรียน 6.68% ที่ได้ระดับคะแนน F

แบบฝึกหัด 2.2

1. ถ้าข้อมูลชุดหนึ่งมีการแจกแจงปกติโดยมีค่าเฉลี่ยเลขคณิตเป็น 400 หน่วยและส่วนเบี่ยงเบนมาตรฐานเป็น 100 หน่วย อยากทราบว่ามีกี่เปอร์เซ็นต์ของข้อมูลซึ่งมีค่า

   1) มากกว่า 538
   2) มากกว่า 179
   3) น้อยกว่า 356
   4) น้อยกว่า 621
   5) ระหว่าง 318 และ 671
   6) ระหว่าง 484 และ 565
   7) ระหว่าง 249 และ 297

2. ในการบรรจุกาแฟชนิดหนึ่งลงขวดให้มีน้ำหนักสุทธิ 115 กรัม ถ้าน้ำหนักของกาแฟที่บรรจุมีการแจกแจงปกติ และมีน้ำหนักเฉลี่ยเท่ากับ 115.5 กรัม ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.3 กรัม อยากทราบว่ามีกี่เปอร์เซ็นต์ที่กาแฟในแต่ละขวดมีน้ำหนัก

   1) ระหว่าง 115 กรัม และ 115.5 กรัม
   2) ระหว่าง 114.9 กรัม และ 115.5 กรัม
   3) ระหว่าง 115.2 กรัม และ 115.9 กรัม
   4) ระหว่าง 114.7 กรัม และ 115 กรัม
   5) มากกว่า 115.5 กรัม
   6) น้อยกว่า 115 กรัม

3. คะแนนทดสอบความถนัดทางคณิตศาสตร์ (Mathematics Attitude Test) สำหรับกลุ่มนักเรียนหญิงมีค่าเฉลี่ยเลขคณิตเป็น 60 คะแนน ส่วนเบี่ยงเบนมาตรฐานเป็น 10 คะแนน และกลุ่มนักเรียนชายมีค่าเฉลี่ยเลขคณิตเป็น 64 คะแนน ส่วนเบี่ยงเบนมาตรฐานเป็น 8 คะแนน ถ้าคะแนนของแต่ละกลุ่มมีการแจกแจงปกติ อยากทราบว่า

   1) ถ้านายผไทสอบได้ 62 คะแนน คะแนนของเขาเป็นตำแหน่งเปอร์เซ็นไทล์ที่เท่าไรของคะแนนในกลุ่มนักเรียนชาย
   2) ถ้านางสาวอาภัสราสอบได้ 73 คะแนน คะแนนของเธอเป็นตำแหน่งเปอร์เซ็นไทล์ที่เท่าไรในกลุ่มนักเรียนหญิง และเป็นตำแหน่งเปอร์เซ็นไทล์ที่เท่าไรในกลุ่มนักเรียนชาย

4. การแจกแจงของคะแนนสอบครั้งหนึ่งเป็นการแจกแจงปกติโดยมีค่าเฉลี่ยเลขคณิต 72 คะแนน และส่วนเบี่ยงเบนมาตรฐาน 12 คะแนน จงหา

   1) คะแนนที่เป็นเปอร์เซ็นไทล์ที่ 25
   2) คะแนนที่เป็นเปอร์เซ็นไทล์ที่ 90

5. ในการผลิตแผ่นพลาสติกของบริษัทแห่งหนึ่งปรากฏว่าความหนาแน่นของแผ่นพลาสติกมีการแจกแจงแบบปกติโดยมีความหนาเฉลี่ย 0.0625 เซนติเมตร ความแปรปรวนเป็น 0.00000625 เซนติเมตร^2 จงหาว่าแผ่นพลาสติกที่ผลิตได้มีความหนาอยู่ระหว่าง 0.0595 เซนติเมตร และ 0.0659 เซนติเมตรมีกี่เปอร์เซ็นต์

6. ให้ x เป็นความคลาดเคลื่อนในรอบ 24 ชั่วโมงของนาฬิกาที่ผลิตโดยโรงงานแห่งหนึ่ง ถ้าความคลาดเคลื่อนมีการแจกแจงปกติ และมีค่าเฉลี่ยเลขคณิต 0.00 วินาที ความแปรปรวน 0.160 วินาที^{2} อยากทราบว่า x ซึ่งทำให้ 50.04% ของนาฬิกาทั้งหมดที่ผลิตได้โดยมีความคลาดเคลื่อนระหว่าง x กับ 0.136 วินาที

7. น้ำหนักสุทธิของกระป๋องบรรจุถั่วที่ผลิตโดยบริษัทแห่งหนึ่งมีการแจกแจงแบบปกติ โดยมีน้ำหนักสุทธิเฉลี่ยเป็น 12.00 กรัม ถ้ากระป๋องที่มีน้ำหนักสุทธิน้อยกว่า 11.88 กรัม มีอยู่ 11.51% จงหาความแปรปรวนของน้ำหนักสุทธิของกระป๋องบรรจุถั่วที่ผลิตโดยบริษัทนี้

8. ถ้า x แทนคะแนนที่สนใจศึกษาและ P แทนพื้นที่ใต้เส้นโค้งปกติมาตรฐานของคะแนนที่ต่ำกว่า x จงหาค่า a , b , c และ d จากข้อมูลที่กำหนดให้ต่อไปนี้

ค่าเฉลี่ยเลขคณิต	ส่วนเบี่ยงเบนมาตรฐาน	x	พื้นที่ใต้เส้นโค้งปกติมาตรฐาน (P)
3	1	2	d
10	2	c	0.18
a	3	6	0.09
10	b	12	0.60

9. คะแนนสอบ SAT (SAT Scores) มีการแจกแจงปกติโดยมีค่าเฉลี่ยเลขคณิต 505 คะแนน และส่วนเบี่ยงเบนมาตรฐาน 111 คะแนน จงหาพื้นที่ใต้เส้นโค้งปกติมาตรฐานในข้อต่อไปนี้
   1) คะแนน SAT อยู่ระหว่าง 400 และ 600
   2) คะแนน SAT มากกว่า 700
   3) คะแนน SAT น้อยกว่า 450

บทที่ 3
ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล

ในบทที่ 1 และบทที่ 2 เป็นการศึกษาข้อมูลเชิงปริมาณของตัวแปรเดียวหรือของแต่ละตัวแปรในข้อมูลแต่ละชุด ส่วนในบทที่ 3 นี้ เป็นการศึกษาข้อมูลเชิงปริมาณของตัวแปรทีละสองตัวแปรพร้อมกัน โดยมุ่งตรวจสอบความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลของสองตัวแปรนั้น ๆ เพื่อประโยชน์ในการใช้ตัวแปรตัวหนึ่งไปพยากรณ์ตัวแปรอีกตัวหนึ่งภายใต้สมการที่แสดงความสัมพันธ์เชิงฟังก์ชันที่คำนวณได้ โดยทั่วไปสมการที่ใช้แสดงความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลอาจประกอบด้วยตัวแปรหลายตัว แต่ในระดับเบื้องต้นนี้เป็นการศึกษากรณีพื้นฐานของสองตัวแปรเท่านั้น

สมการที่ใช้แสดงความสัมพันธ์เชิงฟังก์ชันในบทที่ 3 นี้ มีรูปแบบของความสัมพันธ์ที่มีกราฟเป็นเส้นตรงและไม่เป็นเส้นตรง ดังนั้น การเลือกว่าจะใช้สมการรูปแบบใดจึงจำเป็นต้องตรวจสอบโดยการลงจุดสร้างแผนภาพการกระจาย (scatterplots) ระหว่างข้อมูลของ 2 ตัวแปรก่อน เพื่อตรวจดูรูปแบบที่เหมาะสม แล้วจึงวิเคราะห์ข้อมูลในลำดับต่อไป

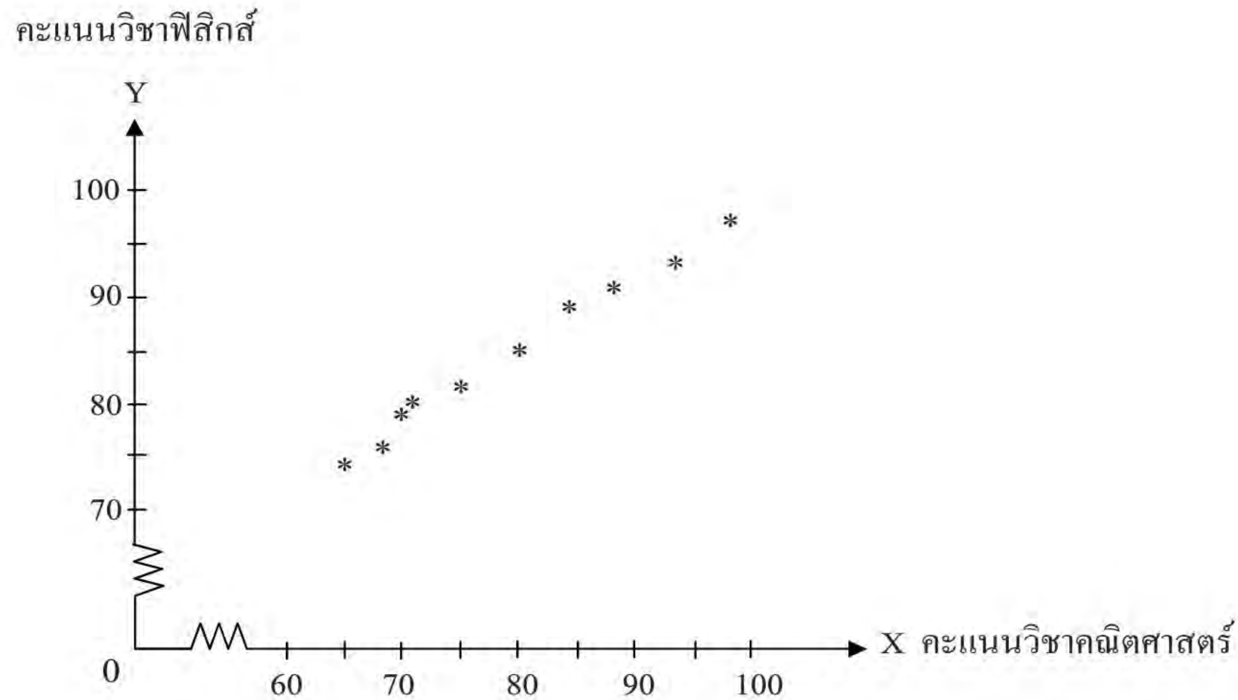
3.1 การวิเคราะห์ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล

ในการวิเคราะห์ข้อมูลบ่อยครั้งมีข้อมูลเชิงปริมาณที่ประกอบด้วยตัวแปรตั้งแต่สองตัวขึ้นไป และตัวแปรเหล่านั้นมีความเกี่ยวข้องกันอยู่ เช่น รายได้และรายจ่ายของครอบครัว ส่วนสูงและน้ำหนักของเด็กแรกเกิด ความเกี่ยวข้องกันของตัวแปรจะมีลักษณะที่ค่าของตัวแปรหนึ่งขึ้นอยู่กับอีกตัวแปรหนึ่ง เช่น รายจ่ายจะขึ้นอยู่กับรายได้ ส่วนสูงขึ้นอยู่กับน้ำหนัก กรณีเช่นนี้จะเรียกตัวแปรที่แสดงรายได้หรือน้ำหนักว่า **ตัวแปรอิสระ** (independent variables) เรียกตัวแปรที่แสดงรายจ่ายหรือส่วนสูงว่า **ตัวแปรตาม** (dependent variables)

สมมติว่า ข้อมูลต่อไปนี้แสดงคะแนนวิชาฟิสิกส์ และวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 5 จำนวน 10 คน ของโรงเรียนแห่งหนึ่ง

นักเรียนคนที่	1	2	3	4	5	6	7	8	9	10
คะแนนวิชาคณิตศาสตร์	84	87	98	71	75	65	93	68	70	80
คะแนนวิชาฟิสิกส์	89	91	97	80	82	74	93	76	78	85

ถ้าให้คะแนนวิชาคณิตศาสตร์เป็นตัวแปรอิสระแทนด้วย X และคะแนนวิชาฟิสิกส์เป็นตัวแปรตามแทนด้วย Y เมื่อใช้คู่อันดับ (x, y) โดย x เป็นคะแนนวิชาคณิตศาสตร์ y เป็นคะแนนวิชาฟิสิกส์ของนักเรียนแต่ละคน ซึ่งคือคู่อันดับ (84, 89), (87, 91), (98, 97), ... , (70, 78), (80, 85) เป็นพิกัดของจุดจะได้จุดดังในรูปต่อไปนี้



จะเห็นว่าการกระจายของจุด (x, y) ซึ่งใช้แสดงค่าต่าง ๆ ของตัวแปรทั้งสองอยู่ในลักษณะที่พอจะประมาณหรือแทนด้วยเส้นตรงได้ และสามารถสร้างความสัมพันธ์เชิงฟังก์ชันที่อยู่ในรูปเส้นตรงได้โดยใช้ข้อมูลที่มีอยู่ จุดประสงค์สำคัญของการสร้างความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลสองชุด ก็เพื่อใช้สมการความสัมพันธ์เชิงฟังก์ชันของ x และ y ในการพยากรณ์ค่าของตัวแปรตามเมื่อทราบค่าของตัวแปรอิสระ เช่น ถ้าทราบ

คะแนนวิชาคณิตศาสตร์ (x) ของนักเรียนชั้นมัธยมศึกษาปีที่ 5 คนหนึ่งในโรงเรียนแห่งหนึ่ง ก็สามารถพยากรณ์คะแนนวิชาฟิสิกส์ (y) ของนักเรียนคนนั้นได้ โดยแทน x ลงในสมการเพื่อคำนวณหา y นั่นเอง

ในการสร้างความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลเชิงปริมาณนี้ ข้อมูลเชิงปริมาณที่นำมาสร้างความสัมพันธ์จะต้องประกอบด้วยค่าจากการสังเกตเป็นจำนวนมากพอสมควร เช่น ตั้งแต่ 10 ค่าขึ้นไป เพราะถ้าค่าจากการสังเกตมีจำนวนน้อยแล้ว ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลเชิงปริมาณของตัวแปรสองตัวที่สร้างขึ้นอาจจะไม่สามารถแทนความสัมพันธ์ที่ควรจะเกิดขึ้นจริง ๆ ระหว่างตัวแปรทั้งสอง จะเป็นผลทำให้การพยากรณ์ค่าของตัวแปรตามที่ต้องการทราบอาจคลาดเคลื่อนไปจากที่ควรจะเป็นจริงมาก

โดยทั่ว ๆ ไป ความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่ประกอบด้วยตัวแปรสองตัวแปรอาจแบ่งออกเป็นสองชนิดใหญ่ ๆ คือ

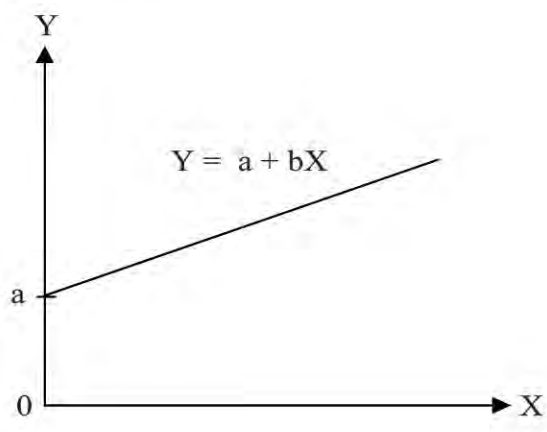
1. ความสัมพันธ์เชิงฟังก์ชันที่กราฟเป็นเส้นตรง มีสมการทั่วไปของความสัมพันธ์เชิงฟังก์ชันเป็น

$$Y = a + bX$$

เมื่อ Y เป็นตัวแปรตาม และ X เป็นตัวแปรอิสระ

a และ b เป็นค่าคงตัวที่ต้องการหา ซึ่ง a เป็นระยะตัดแกน Y และ b เป็นความชันของเส้นตรง หรือค่าของ Y ที่เปลี่ยนไปเมื่อ X เปลี่ยนไปหนึ่งหน่วย

เมื่อทราบค่า a และ b กราฟของความสัมพันธ์เชิงเส้นตรงบางกรณีจะมีลักษณะตามรูปข้างล่าง



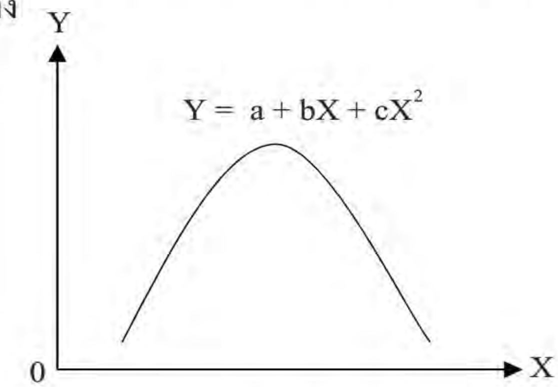
2. ความสัมพันธ์เชิงฟังก์ชันที่กราฟไม่เป็นเส้นตรง ในที่นี้จะกล่าวเฉพาะความสัมพันธ์ที่มีกราฟเป็นพาราโบลาและความสัมพันธ์ในรูปฟังก์ชันเอกซ์โพเนนเชียล ซึ่งสมการของความสัมพันธ์ที่มีกราฟเป็นพาราโบลาเป็นดังนี้

$$Y = a + bX + cX^2$$

เมื่อ Y เป็นตัวแปรตาม และ X เป็นตัวแปรอิสระ

a, b และ c เป็นค่าคงตัวที่จะต้องหา

เมื่อทราบค่า a, b และ c กราฟของความสัมพันธ์เชิงพาราโบลาบางกรณีจะมีลักษณะตามรูปข้างล่าง



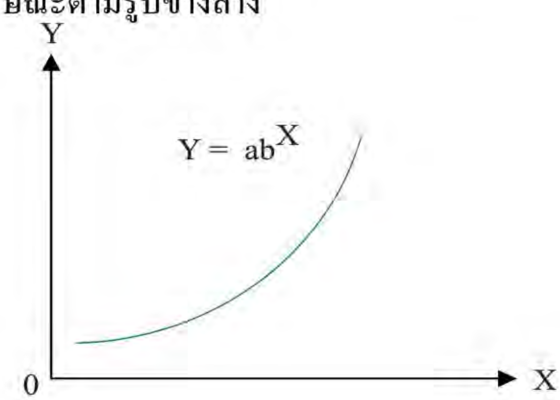
สมการของความสัมพันธ์ในรูปฟังก์ชันเอกซ์โพเนนเชียลเป็นดังนี้

$$Y = ab^X$$

เมื่อ Y เป็นตัวแปรตาม และ X เป็นตัวแปรอิสระ

a และ b เป็นค่าคงตัวที่จะต้องหา

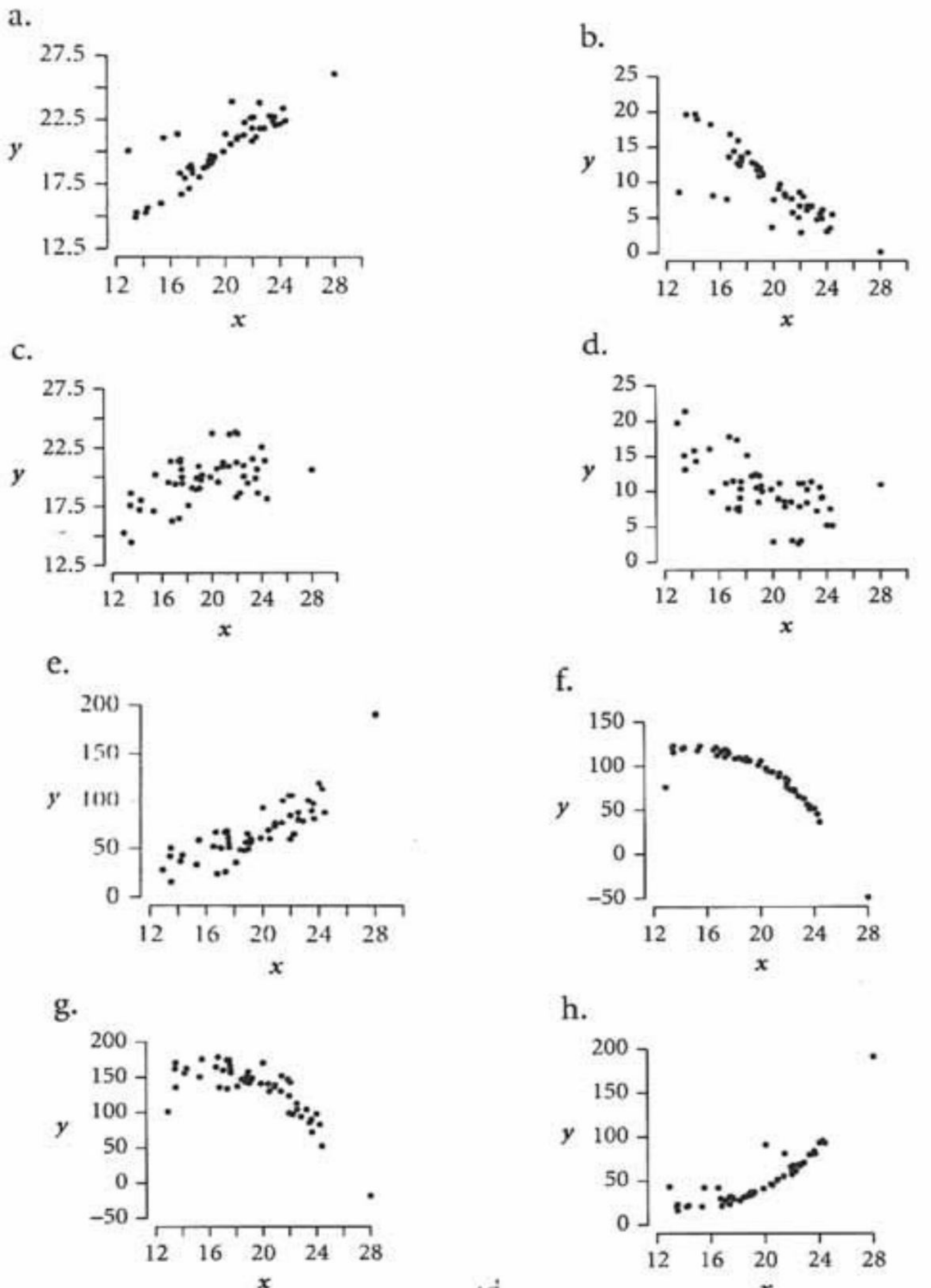
เมื่อทราบค่า a และ b กราฟของความสัมพันธ์ในรูปฟังก์ชันเอกซ์โพเนนเชียลบางกรณีจะมีลักษณะตามรูปข้างล่าง



3.2 แผนภาพการกระจาย

แผนภาพการกระจาย

ในการสร้างความสัมพันธ์เชิงฟังก์ชันของข้อมูลเชิงปริมาณที่ประกอบด้วยตัวแปรสองตัวจากข้อมูลที่มีอยู่ทั้งหมด หรือจากตัวอย่างข้อมูลที่เลือกมาเป็นตัวแทนของข้อมูลที่มีอยู่ มีความจำเป็นที่จะต้องตรวจดูรูปแบบของความสัมพันธ์ที่เกิดขึ้นระหว่างตัวแปรทั้งสอง เพื่อที่จะนำมาใช้ในการกำหนดความสัมพันธ์เชิงฟังก์ชันที่จะสร้างขึ้น รูปแบบของความสัมพันธ์นี้พิจารณาได้จากกราฟที่สร้างจากข้อมูลที่มีอยู่ทั้งหมดหรือจากข้อมูลตัวอย่างที่เลือกมาเป็นตัวแทนของข้อมูลที่มีอยู่ ซึ่งเรียกว่า แผนภาพการกระจายของข้อมูลนั้นๆ บางครั้งการพิจารณาจากแผนภาพการกระจายไม่สามารถบอกลงไปได้แน่นอนว่ามีรูปของความสัมพันธ์เป็นแบบใด เนื่องจากลักษณะการกระจายของข้อมูลไม่สามารถจัดเข้าในรูปความสัมพันธ์เชิงฟังก์ชันใด ๆ ได้ หรืออาจจะมีลักษณะของความสัมพันธ์ที่ใกล้เคียงกับรูปของความสัมพันธ์เชิงฟังก์ชันสองรูป อาจจะอนุโลมให้อยู่ในรูปของความสัมพันธ์เชิงฟังก์ชันที่เป็นเส้นตรง หรือเป็นฟังก์ชันเอกซ์โพเนนเซียลก็ได้ ในกรณีนี้ถ้าผู้สร้างความสัมพันธ์มีความรู้ความชำนาญเกี่ยวกับข้อมูลชนิดนั้น ๆ อาจจะบอกได้ว่าควรจะสร้างความสัมพันธ์เชิงฟังก์ชันให้อยู่ในรูปใด จึงจะเหมาะสมกับสิ่งที่ควรจะเป็นมากที่สุด หรือถ้าผู้สร้างความสัมพันธ์ไม่ต้องการความละเอียดถูกต้องจากการพยากรณ์ค่าตัวแปรตาม จากความสัมพันธ์ที่สร้างขึ้นนั้นมากนักก็อาจจะเลือกใช้รูปของความสัมพันธ์เชิงฟังก์ชันที่ง่ายต่อการคำนวณ สรุปได้ว่าการกำหนดรูปแบบของความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณทั้งสองโดยพิจารณาจากแผนภาพการกระจาย (รูปที่ 1) จะขึ้นอยู่กับความรู้ความชำนาญเกี่ยวกับเรื่องที่นำมาสร้างความสัมพันธ์ของผู้สร้างความสัมพันธ์นั้นและความละเอียดถูกต้องของค่าทำนายที่ต้องการเป็นสำคัญ



รูปที่ 1

ที่มา : Watkins et al. (2004)

แผนภาพการกระจาย 8 รูปนี้ แต่ละรูปแสดงแนวโน้ม (trend) ของความสัมพันธ์เชิงฟังก์ชัน ระหว่างตัวแปรอิสระ (X) และตัวแปรตาม (Y) ในรูปของกราฟที่มีแนวโน้ม 2 แบบหลัก ๆ ดังนี้

แบบที่ 1 คือ แบบเป็นเส้นตรง ซึ่งสามารถพิจารณาเป็น 2 กรณี คือ กรณีที่มีแนวโน้มทางบวก (positive trend) หมายความว่า เมื่อค่าของ X เพิ่มขึ้น ค่าของ Y จะเพิ่มขึ้นด้วย และกรณีที่มีแนวโน้มทางลบ (negative trend) หมายความว่า เมื่อค่าของ X เพิ่มขึ้น ค่าของ Y กลับลดลง

แบบที่ 2 คือ แบบไม่เป็นเส้นตรง ซึ่งก็พิจารณาเป็น 2 กรณี เช่นเดียวกันกับแนวโน้มของความสัมพันธ์ที่มีลักษณะเป็นเส้นตรง คือ อาจมีแนวโน้มทางบวกหรือทางลบได้

ดังนั้น ในแต่ละลักษณะหรือรูปแบบของความสัมพันธ์ที่เป็นแบบเส้นตรงหรือไม่เป็นเส้นตรง (linear or curved) นั้น เมื่อพิจารณาได้ว่าความสัมพันธ์นั้น ๆ มีแนวโน้มทางบวกหรือทางลบแล้ว จุดสำคัญยังต้องพิจารณาสารสนเทศเพิ่มเติมให้ได้ว่า แนวโน้มความสัมพันธ์ที่เกิดขึ้นหรือที่พบมีระดับใด เช่น มาก (strong) ปานกลาง (moderate) หรือน้อย (weak) (ดูรูปที่ 1 ประกอบ)

จากรูปที่ 1 จะเห็นว่า รูป a, b และ e มีรูปแบบของความสัมพันธ์เป็นเส้นตรง ส่วนรูป c, d, f, g และ h มีรูปแบบของความสัมพันธ์ไม่เป็นเส้นตรง ส่วนแนวโน้มของความสัมพันธ์ของรูป a, b และ e นั้น มี 2 แบบ คือ รูป a และ e มีแนวโน้มเป็นเส้นตรงทางบวก ส่วนรูป b มีแนวโน้มเป็นเส้นตรงทางลบ และสำหรับรูปที่เหลือ คือ c, d, f, g และ h ก็แบ่งออกได้ 2 แบบเช่นเดียวกัน คือ รูป c และ h มีแนวโน้มไม่เป็นเส้นตรงทางบวก ส่วนรูป d, f และ g มีแนวโน้มไม่เป็นเส้นตรงทางลบ

นอกจากนี้การดูแผนภาพการกระจาย ถ้าทราบว่ามีลักษณะของความสัมพันธ์และความสัมพันธ์นั้น ๆ มีแนวโน้มเป็นทางบวกหรือทางลบแล้ว ยังสามารถดูลักษณะของความสัมพันธ์ของจุดต่าง ๆ ว่า สัมพันธ์กันมาก ปานกลาง หรือน้อย ซึ่งพิจารณาดังนี้ ถ้าจุดต่าง ๆ หรือกลุ่มของจุด (points cluster) อยู่ใกล้รอบ ๆ เส้นที่คาดไว้ (imaginary line) ไม่ว่าจะเป็นแบบเส้นตรงหรือไม่เป็นเส้นตรงก็ตาม จะชี้ให้เห็นว่ามีความสัมพันธ์มาก ถ้ากลุ่มของจุดหรือจุด กระจายห่างจากเส้นที่คาดไว้ จะชี้ให้เห็นว่า มีความสัมพันธ์น้อย นอกจากนี้อาจมีความสัมพันธ์ปานกลางหรืออาจจะไม่มีความสัมพันธ์กันถ้าแผนภาพการกระจายปราศจากรูปแบบแนวโน้มใด ๆ คือ กระจายเต็มทุกทิศทาง

3.3 การประมาณค่าของค่าคงตัวโดยใช้วิธีกำลังสองน้อยสุด

เมื่อทราบรูปแบบของความสัมพันธ์เชิงฟังก์ชันของข้อมูลจากการสร้างแผนภาพการกระจายของข้อมูลว่าควรอยู่ในรูปแบบใด เช่น เส้นตรง พาราโบลา หรือฟังก์ชันเอกซ์โพเนนเชียลแล้ว ขั้นต่อไปคือการหาสมการของความสัมพันธ์ โดยเริ่มจากสมการรูปใดรูปหนึ่งได้แก่ $Y = a + bX$ $Y = a + bX + cX^2$ หรือ $Y = ab^X$ แล้วหาค่าประมาณของค่าคงตัวในสมการที่เลือก เมื่อทราบค่าคงตัวแล้ว จึงสามารถนำไปใช้ในการพยากรณ์ค่าตัวแปรตามเมื่อทราบหรือกำหนดค่าตัวแปรอิสระได้

ในการหาค่าคงตัวที่ปรากฏอยู่ในสมการทั่วไปนั้น หาได้โดยอาศัยหลักว่า ถ้าจะให้สมการของความสัมพันธ์ที่สร้างขึ้นสามารถใช้แทนความสัมพันธ์ที่เกิดขึ้นได้ดีแล้ว ผลรวมของความแตกต่างระหว่างค่าที่ได้จากความสัมพันธ์เชิงฟังก์ชันที่สร้างขึ้นกับค่าที่เกิดขึ้นจริงทุก ๆ ค่าควรจะน้อยที่สุด แต่เนื่องจากความแตกต่างระหว่างค่าที่ได้จากความสัมพันธ์เชิงฟังก์ชันแต่ละค่ากับค่าที่เกิดขึ้นจริงอาจเป็นจำนวนบวกบ้าง ลบบ้าง ขึ้นอยู่กับแต่ละค่าของความสัมพันธ์กับค่าที่เกิดขึ้นจริงว่าค่าใดจะมากกว่ากัน ทำให้ผลรวมของความแตกต่างดังกล่าวอาจจะหักล้างกันหมดพอดี ดังนั้นผลรวมของความแตกต่างที่เกิดขึ้นเป็นศูนย์ซึ่งไม่ตรงกับความจริง เพราะความแตกต่างที่เกิดขึ้นระหว่างค่าจากความสัมพันธ์และค่าที่เกิดขึ้นจริงไม่ว่าจะมากกว่าหรือน้อยกว่าจะเป็นความแตกต่าง ซึ่งถือว่าเป็นความคลาดเคลื่อนทั้งสิ้น ไม่สามารถหักล้างกันจนความคลาดเคลื่อนที่เกิดขึ้นหมดไปได้ ดังนั้น เพื่อที่จะไม่ให้ผลรวมของความแตกต่างที่เกิดขึ้นเป็นศูนย์ จึงต้องหากำลังสองของความแตกต่างที่เกิดขึ้นเสียก่อน แล้วจึงหาผลรวมของกำลังสองของความแตกต่างที่เกิดขึ้น ถ้าผลรวมนี้น้อยที่สุดจะถือว่าความสัมพันธ์เชิงฟังก์ชันนั้นเหมาะสมสำหรับการประมาณค่าตัวแปรตามมากที่สุด วิธีดังกล่าวนี้เรียกว่า **วิธีกำลังสองน้อยสุด** (method of least-squares) กล่าวคือ

ถ้าให้ y_i เป็นค่าของตัวแปรตาม Y หรือค่าจากการสังเกตที่เกิดขึ้นจริงตัวที่ i เมื่อ i คือ 1, 2, 3, ... , n

$\hat{y}_i$ (อ่านว่า “วายแฮท” ตัวที่ i) เป็นค่าของตัวแปรตาม Y ที่ได้จากความสัมพันธ์เชิงฟังก์ชันระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X ที่ได้จากการแทนค่าตัวแปร X ด้วย x_i

ต้องการทำให้ $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ มีค่าน้อยสุด (เมื่อ n เป็นจำนวนตัวอย่างที่ใช้หรือจำนวนข้อมูลทั้งหมดที่นำมาสร้างความสัมพันธ์)

วิธีที่จะทำให้ $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ มีค่าน้อยที่สุดจะไม่แสดงไว้ในที่นี้ เนื่องจากต้องใช้ความรู้คณิตศาสตร์ชั้นสูง แต่จากการใช้วิธีกำลังสองน้อยที่สุดดังกล่าวจะได้สมการซึ่งเรียกว่า **สมการปกติ** (normal equations) โดยมีจำนวนสมการเท่ากับจำนวนค่าคงตัว ที่ต้องการหา กล่าวคือ

1. สมการเส้นตรงมีรูปสมการทั่วไปคือ $Y = a + bX$ และมีสมการปกติคือ

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i \quad \text{---------- (1)}$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad \text{---------- (2)}$$

2. สมการพาราโบลา มีรูปสมการทั่วไปคือ $Y = a + bX + cX^2$ และมีสมการปกติคือ

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 \quad \text{---------- (1)}$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 \quad \text{---------- (2)}$$

$$\sum_{i=1}^{n} x_i^2 y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 \quad \text{---------- (3)}$$

3. สมการเอกซ์โพเนนเชียลมีรูปสมการทั่วไปคือ $Y = ab^X$ หรือสมการเชิงเส้นของลอการิทึมที่เกี่ยวข้อง $\log Y = \log a + X \log b$ ซึ่งมีสมการปกติคือ

$$\sum_{i=1}^{n} \log y_i = n \log a + (\log b)\sum_{i=1}^{n} x_i \quad \text{---------- (1)}$$

$$\sum_{i=1}^{n} x_i \log y_i = (\log a)\sum_{i=1}^{n} x_i + (\log b)\sum_{i=1}^{n} x_i^2 \quad \text{---------- (2)}$$

เมื่อ (x_i, y_i) เป็นค่าจากการสังเกตของตัวแปรอิสระและตัวแปรตามคู่ที่ i เมื่อ i คือ 1, 2, 3, ... , n และ a, b, c เป็นค่าคงตัวที่ต้องการหา

ค่าคงตัวของสมการของความสัมพันธ์เชิงฟังก์ชันแบบต่าง ๆ หาได้จากการแก้สมการปกติหลังจากที่ได้แทนค่าตัวแปรอิสระและตัวแปรตามที่นำมาสร้างความสัมพันธ์นั้นแล้ว

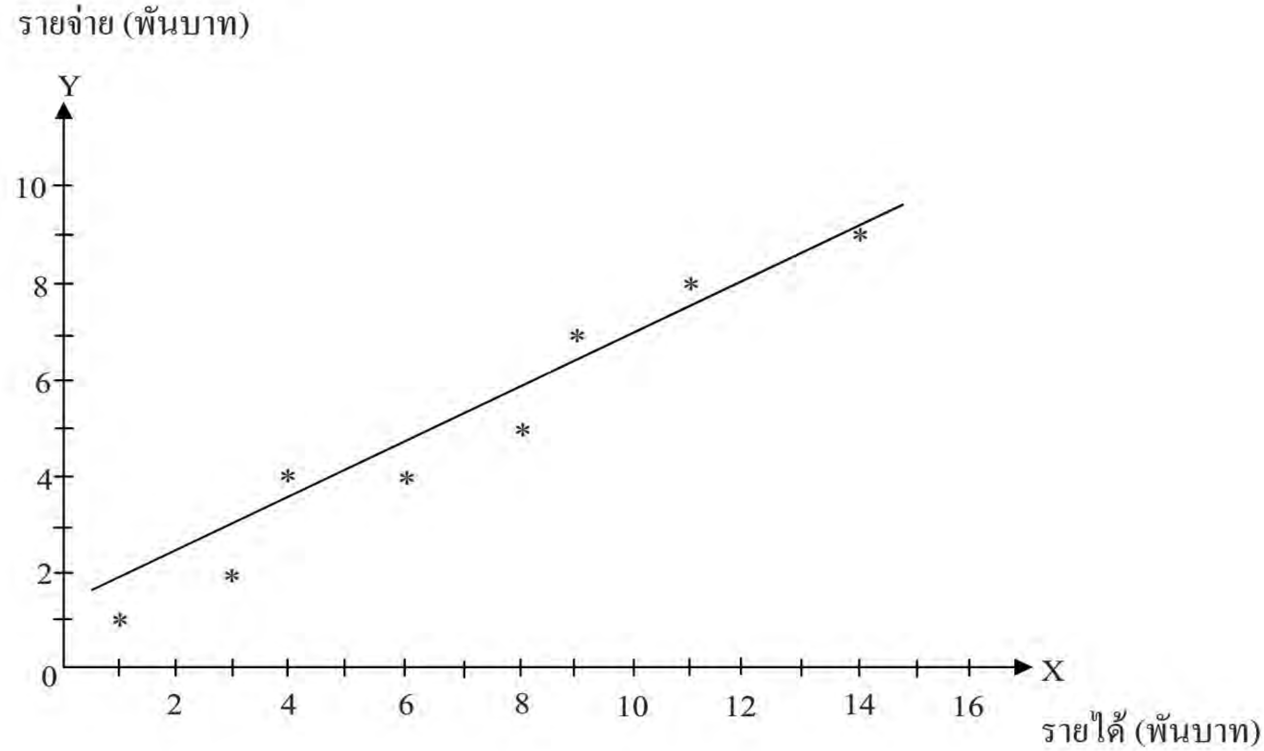
หมายเหตุ สมการต่าง ๆ ที่กล่าวถึงในหัวข้อ 3.3 เป็นสมการที่ใช้ในการประมาณหรือพยากรณ์ค่าตัวแปรตาม เมื่อทราบหรือกำหนดค่าตัวแปรอิสระ นั่นคือ ถ้าต้องการประมาณหรือพยากรณ์ค่าของตัวแปรใดก็ต้องให้ตัวแปรตัวนั้นเป็นตัวแปรตาม และตัวแปรอีกตัวหนึ่งเป็นตัวแปรอิสระ

ตัวอย่างที่ 1 จากการสอบถามรายจ่ายของครอบครัว 8 ครอบครัว ในท้องที่แห่งหนึ่งที่มีรายได้ต่อเดือนตั้งแต่ครอบครัวละ 1,000 บาท จนถึง 14,000 บาท ปรากฏผลเป็นดังนี้

รายได้ (พันบาท) : X	1	3	4	6	8	9	11	14
รายจ่าย (พันบาท) : Y	1	2	4	4	5	7	8	9

1) จงเขียนแผนภาพการกระจายของข้อมูล
2) จงเขียนกราฟที่ใช้แสดงความสัมพันธ์ระหว่างรายจ่ายและรายได้ของครอบครัวซึ่งอาศัยอยู่ในท้องที่แห่งนี้จากแผนภาพการกระจายของข้อมูลในข้อ 1)
3) จงหาสมการของความสัมพันธ์เชิงฟังก์ชันสำหรับการประมาณรายจ่ายจากรายได้ของครอบครัว
4) ถ้าครอบครัวหนึ่งในท้องที่แห่งนั้นมีรายได้เดือนละ 5,000 บาท ให้พยากรณ์รายจ่ายของครอบครัวนั้น
5) ถ้าครอบครัวหนึ่งในท้องที่แห่งนั้นมีรายจ่ายเดือนละ 3,500 บาท ให้พยากรณ์รายได้ของครอบครัวนั้น

วิธีทำ 1) และ 2)



3) เมื่อพิจารณากราฟแผนภาพการกระจายที่ใช้แสดงความสัมพันธ์ระหว่างรายจ่ายและรายได้ของครอบครัวที่อาศัยอยู่ในท้องที่แห่งนี้ อาจอนุโลมได้ว่าอยู่ในรูปแบบเส้นตรง สมการปกติของความสัมพันธ์เชิงฟังก์ชันที่มีกราฟเป็นเส้นตรงมีรูปสมการทั่วไปเป็น $Y = a + bX$ และมีสมการปกติคือ

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i \qquad \text{----------- (1)}$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \qquad \text{----------- (2)}$$

จากสมการ (1) และ (2) หาพจน์ต่าง ๆ ที่ใช้ในการคำนวณหาค่าคงตัว a และ b

x_i	y_i	x_i^2	$x_i y_i$
1	1	1	1
3	2	9	6
4	4	16	16
6	4	36	24
8	5	64	40
9	7	81	63
11	8	121	88
14	9	196	126
$\sum_{i=1}^{8} x_i = 56$	$\sum_{i=1}^{8} y_i = 40$	$\sum_{i=1}^{8} x_i^2 = 524$	$\sum_{i=1}^{8} x_i y_i = 364$

แทน $\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} y_i, \sum_{i=1}^{n} x_i^2$ และ $\sum_{i=1}^{n} x_i y_i$ ด้วยค่าในตาราง ในสมการ (1) และ (2) จะได้

$$40 = 8a + 56b \qquad \text{----------} (3)$$

$$364 = 56a + 524b \qquad \text{----------} (4)$$

จาก (3) และ (4) คำตอบของระบบสมการคือ $b \approx 0.636$ และ $a \approx 0.545$
ดังนั้น สมการสำหรับการประมาณรายจ่ายจากรายได้ของครอบครัวใด ๆ ที่อาศัยอยู่ในท้องที่แห่งนี้ คือ

$$\hat{Y} = 0.545 + 0.636X^{*}$$

เมื่อ X และ Y แทนรายได้และรายจ่ายของครอบครัวในท้องที่แห่งนั้น ตามลำดับ

* หนังสือเล่มนี้จะใช้ $\hat{X}$ หรือ $\hat{Y}$ ในสมการที่ใช้ในการพยากรณ์

4) ถ้าครอบครัวหนึ่งซึ่งอาศัยอยู่ในท้องที่แห่งนี้มีรายได้ 5,000 บาท แสดงว่าต้องแทน X ในสมการทำนายนี้ด้วย 5 จะได้ $\hat{Y} = 0.545 + 0.636(5) = 3.725$
ดังนั้น ถ้าครอบครัวหนึ่งมีรายได้ 5,000 บาท จะทำนายได้ว่าครอบครัวนี้มีรายจ่ายประมาณ 3,725 บาท

5) ในการพยากรณ์รายได้ของครอบครัวที่มีรายจ่าย 3,500 บาท จะต้องสลับให้รายได้ (X) เป็นตัวแปรตาม และรายจ่าย (Y) เป็นตัวแปรอิสระ นั่นคือสมการปกติของความสัมพันธ์ระหว่างรายได้และรายจ่ายจะต้องประมาณด้วยสมการ $X = a + bY$ โดยที่ผลรวมของกำลังสองของความแตกต่างระหว่างค่าที่ประมาณได้จากความสัมพันธ์ที่สร้างขึ้นกับค่าที่เกิดขึ้นจริงทุกค่าหรือ $\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$ มีค่าน้อยที่สุด จะได้สมการปกติเป็น

$$\sum_{i=1}^{n} x_i = an + b\sum_{i=1}^{n} y_i \qquad \text{----------} (5)$$

$$\sum_{i=1}^{n} y_i x_i = a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} y_i^2 \qquad \text{----------} (6)$$

จากสมการ (5) และ (6) หาพจน์ต่าง ๆ ที่ใช้ในการคำนวณหาค่าคงตัว a และ b

x_i	y_i	y_i^2	$x_i y_i$
1	1	1	1
3	2	4	6
4	4	16	16
6	4	16	24
8	5	25	40
9	7	49	63
11	8	64	88
14	9	81	126
$\sum_{i=1}^{8} x_i = 56$	$\sum_{i=1}^{8} y_i = 40$	$\sum_{i=1}^{8} y_i^2 = 256$	$\sum_{i=1}^{8} x_i y_i = 364$

แทน $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} y_i$, $\sum_{i=1}^{n} y_i^2$ และ $\sum_{i=1}^{n} x_i y_i$ ด้วยค่าในตาราง ในสมการ (5) และ (6) จะได้

$$56 = 8a + 40b \qquad \text{----------} (7)$$

$$364 = 40a + 256b \qquad \text{----------} (8)$$

จากสมการ (7) และ (8) คำตอบของระบบสมการคือ

$b = 1.5$ และ $a = -0.5$

ดังนั้นสมการสำหรับการประมาณรายได้จากรายจ่ายของครอบครัวในท้องที่แห่งนี้คือ

$$\hat{X} = -0.5 + 1.5Y$$

ถ้ารายจ่ายของครอบครัวหนึ่งในท้องที่แห่งนี้เท่ากับ 3,500 บาท แสดงว่าต้องแทน Y ในสมการทำนายนี้ด้วย 3.5 จะได้

$$\hat{X} = -0.5 + 1.5(3.5) = 4.75$$

ดังนั้น ถ้ารายจ่ายของครอบครัวหนึ่งเป็น 3,500 บาท จะทำนายได้ว่าครอบครัวนี้มีรายได้ประมาณ 4,750 บาท

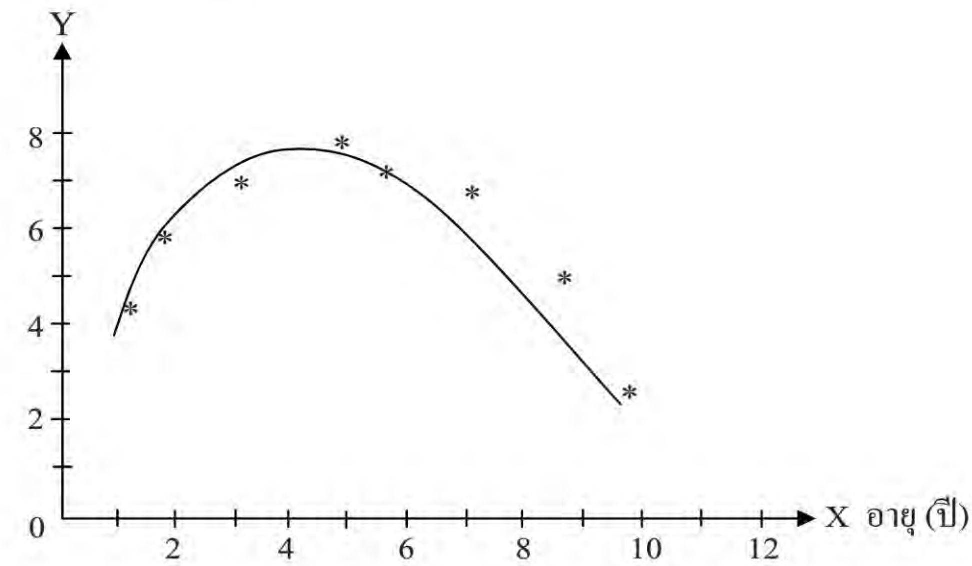
ตัวอย่างที่ 2 จากการศึกษาความสัมพันธ์ระหว่างอายุ (ปี) ของสัตว์ชนิดหนึ่งและปริมาณอาหาร (กิโลกรัม) ที่ใช้เลี้ยงสัตว์ชนิดดังกล่าวต่อสัปดาห์ได้ข้อมูลดังนี้

อายุ (ปี) : X	1.2	1.8	3.1	4.9	5.7	7.1	8.6	9.8
ปริมาณอาหารต่อสัปดาห์ (กิโลกรัม) : Y	4.5	5.9	7.0	7.8	7.2	6.8	4.5	2.7

1) จงเขียนแผนภาพการกระจายของข้อมูลและเขียนกราฟที่ใช้แสดงความสัมพันธ์ดังกล่าว
2) จงหาสมการที่ใช้ประมาณปริมาณอาหารต่อสัปดาห์จากอายุสัตว์
3) ถ้าอายุของสัตว์ชนิดนี้เป็น 4 ปี จะต้องเลี้ยงสัตว์โดยใช้ปริมาณอาหารต่อสัปดาห์เป็นเท่าไร

วิธีทำ 1)

ปริมาณอาหาร (กิโลกรัม)



2) เมื่อพิจารณากราฟจากแผนภาพการกระจายที่ใช้แสดงความสัมพันธ์ระหว่างปริมาณอาหารที่ใช้เลี้ยงสัตว์ต่อสัปดาห์และอายุของสัตว์ อาจอนุโลมได้ว่าอยู่ในรูปพาราโบลาซึ่งสมการปกติของความสัมพันธ์เชิงฟังก์ชันที่มีกราฟเป็นพาราโบลา มีสมการทั่วไปเป็น $Y = a + bX + cX^2$ คือ

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 \quad \text{---------- (1)}$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 \quad \text{---------- (2)}$$

$$\sum_{i=1}^{n} x_i^2 y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 \quad \text{---------- (3)}$$

จากสมการปกติ (1) – (3) หาพจน์ต่าง ๆ ที่ใช้ในการคำนวณหาค่าคงตัว a, b และ c

x_i	y_i	x_i^2	x_i^3	x_i^4	x_iy_i	$x_i^2y_i$
1.2	4.5	1.44	1.728	2.074	5.40	6.480
1.8	5.9	3.24	5.832	10.498	10.62	19.116
3.1	7.0	9.61	29.791	92.352	21.70	67.270
4.9	7.8	24.01	117.649	576.480	38.22	187.278
5.7	7.2	32.49	185.193	1055.600	41.04	233.928
7.1	6.8	50.41	357.911	2541.168	48.28	342.788
8.6	4.5	73.96	636.056	5470.082	38.70	332.820
9.8	2.7	96.04	941.192	9223.682	26.46	259.308
$\sum_{i=1}^{8} x_i = 42.2$	$\sum_{i=1}^{8} y_i = 46.4$	$\sum_{i=1}^{8} x_i^2 = 291.20$	$\sum_{i=1}^{8} x_i^3 = 2275.352$	$\sum_{i=1}^{8} x_i^4 = 18971.936$	$\sum_{i=1}^{8} x_iy_i = 230.42$	$\sum_{i=1}^{8} x_i^2y_i = 1448.988$

แทน $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} y_i$, $\sum_{i=1}^{n} x_i^2$, $\sum_{i=1}^{n} x_i^3$, $\sum_{i=1}^{n} x_i^4$, $\sum_{i=1}^{n} x_iy_i$ และ $\sum_{i=1}^{n} x_i^2y_i$ ด้วยค่าในตาราง ในสมการ (1) – (3) จะได้

$$46.4 = 8a + 42.2b + 291.20c \qquad \text{----------- (4)}$$

$$230.42 = 42.2a + 291.20b + 2275.352c \qquad \text{----------- (5)}$$

$$1448.988 = 291.20a + 2275.352b + 18971.936c \qquad \text{----------- (6)}$$

จากสมการ (4) – (6) คำตอบของระบบสมการคือ

$a \approx 2.588$, $b \approx 2.065$ และ $c \approx -0.211$

ดังนั้น สมการสำหรับการประมาณปริมาณอาหารที่ใช้เลี้ยงสัตว์ต่อสัปดาห์จากอายุของสัตว์ คือ $\hat{Y} = 2.588 + 2.065X - 0.211X^2$

3) ถ้าอายุของสัตว์ชนิดนี้เป็น 4 ปี แสดงว่าต้องแทน X ในสมการทำนายนี้ด้วย 4

จะได้ $\hat{Y} = 2.588 + 2.065(4) - 0.211(4)^2 = 7.472$

ดังนั้น ถ้าอายุของสัตว์ชนิดนี้เป็น 4 ปี จะต้องใช้ปริมาณอาหารเลี้ยงสัตว์ต่อสัปดาห์ประมาณ 7.472 กิโลกรัม

ตัวอย่างที่ 3 ในการวิจัยเกี่ยวกับโรคอ้วน (obesity) ของคนไข้จำนวน 35 คนที่ต้องลดน้ำหนัก และศึกษาความสัมพันธ์ระหว่างปริมาณน้ำหนัก (กิโลกรัม) ที่ลดลงกับปริมาณไทรกลีเซอไรด์ (Triglyceride) (มิลลิกรัม/เดซิลิตร) ที่เปลี่ยนแปลงไปเมื่อได้ลดน้ำหนักตัวเองจนครบ 8 สัปดาห์ จงใช้ข้อมูลที่แสดงไว้ในตาราง ตอบคำถามต่อไปนี้

1) จงเขียนแผนภาพการกระจายของข้อมูลปริมาณน้ำหนักที่ลดลง (X) และปริมาณไทรกลีเซอไรด์ที่เปลี่ยนแปลง (Y) เมื่อคนไข้ลดน้ำหนักเป็นเวลา 8 สัปดาห์
2) จงสังเกตว่า แผนภาพการกระจายมีข้อมูลเล็กหรือใหญ่ผิดปกติหรือไม่ และมีลักษณะของกราฟแสดงความสัมพันธ์ระหว่าง X และ Y อย่างไร
3) จงหาสมการที่ใช้ประมาณ Y จาก X
4) จงพยากรณ์ค่า Y เมื่อน้ำหนักลดลง 5 กิโลกรัม
5) ถ้าสนใจพยากรณ์ค่า Y เมื่อ $X \geq 10$ จะสมเหตุสมผลหรือไม่ เพราะเหตุใด

ลำดับที่	base(X)	wk8(X)	base(Y)	wk8(Y)	X	Y	X^2	XY
1	84.0	82.4	90	131	–1.6	41	2.56	–65.6
2	88.8	87.0	137	82	–1.8	–55	3.24	99.0
3	87.0	81.8	182	152	–5.2	–30	27.04	156.0
4	84.5	80.4	72	72	–4.1	0	16.81	0.0
5	69.4	69.0	143	126	–0.4	–17	0.16	6.8
6	104.7	102.0	96	157	–2.7	61	7.29	–164.7
7	90.0	87.6	115	88	–2.4	–27	5.76	64.8
8	89.4	86.8	124	123	–2.6	–1	6.76	2.6
9	95.2	92.8	188	255	–2.4	67	5.76	–160.8
10	108.1	100.9	167	87	–7.2	–80	51.84	576.0
11	93.9	90.2	143	213	–3.7	70	13.69	–259.0
12	83.4	75.0	143	102	–8.4	–41	70.56	344.4
13	104.4	102.9	276	313	–1.5	37	2.25	–55.5
14	103.7	95.7	84	84	–8.0	0	64.00	0.0
15	99.2	99.2	142	135	0.0	–7	0.00	0.0
16	95.6	88.5	64	114	–7.1	50	50.41	–355.0
17	126.0	123.2	226	152	–2.8	–74	7.84	207.2
18	103.7	95.5	199	120	–8.2	–79	67.24	647.8
19	133.1	130.8	212	156	–2.3	–56	5.29	128.8
20	85.0	80.0	268	250	–5.0	–18	25.00	90.0
21	83.8	77.9	111	107	–5.9	–4	34.81	23.6
22	104.5	98.3	132	117	–6.2	–15	38.44	93.0
23	76.8	73.2	165	96	–3.6	–69	12.96	248.4
24	90.5	88.9	57	63	–1.6	6	2.56	–9.6
25	106.9	103.7	163	131	–3.2	–32	10.24	102.4
26	81.5	78.9	111	54	–2.6	–57	6.76	148.2
27	96.5	94.9	300	241	–1.6	–59	2.56	94.4
28	103.0	97.2	192	124	–5.8	–68	33.64	394.4
29	127.5	124.7	176	215	–2.8	39	7.84	–109.2
30	103.2	102.0	146	138	–1.2	–8	1.44	9.6
31	113.5	115.0	446	795	1.5	349	2.25	523.5
32	107.0	99.2	232	63	–7.8	–169	60.84	1318.2
33	106.0	103.5	255	204	–2.5	–51	6.25	127.5
34	114.9	105.3	187	144	–9.6	–43	92.16	412.8
35	103.4	96.0	154	96	–7.4	–58	54.76	429.2
รวม					–137.70	–398.00	801.01	5069.20

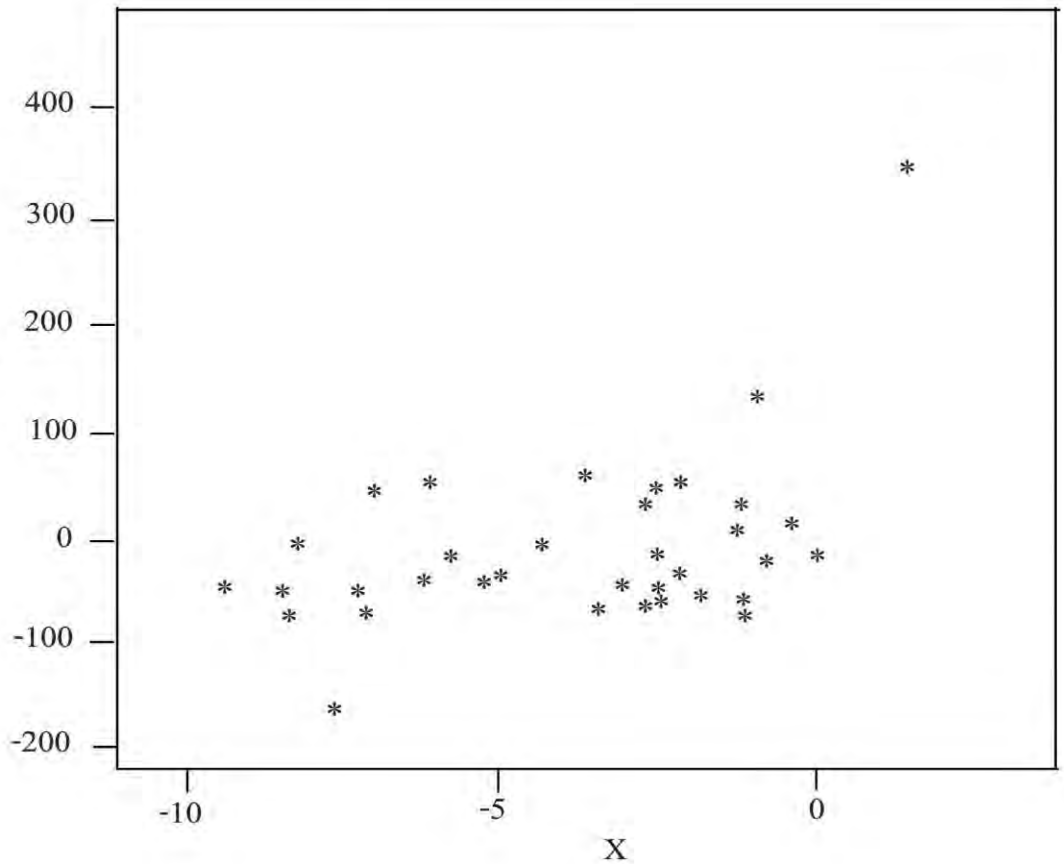
base(X) : น้ำหนักก่อนการทดลอง　　base(Y) : ปริมาณไทรกลีเซอไรด์ก่อนการทดลอง

wk8(X) : น้ำหนักหลังการทดลองได้ 8 สัปดาห์   wk8(X) : ปริมาณไทรกลีเซอไรด์หลังการทดลอง

X : ปริมาณน้ำหนักที่ลดลง โดยที่ X = wk8(X) – base(X)

Y : ปริมาณไทรกลีเซอไรด์ที่เปลี่ยนแปลง โดยที่ Y = wk8(Y) – base(Y)

วิธี   1) แผนภาพการกระจายของข้อมูลระหว่าง X และ Y



2) จากแผนภาพการกระจายในข้อ 1) พบว่ามีจุดผิดปกติ 1 จุดคือ (1.5, 349) ซึ่งอยู่ในลำดับที่ 31 ในตารางข้างต้น และแผนภาพการกระจายมีลักษณะกราฟแสดงความสัมพันธ์ระหว่าง X และ Y คล้ายเชิงเส้นซึ่งมีสมการทั่วไปคือ Y = a + bX

3) สมการปกติของความสัมพันธ์ที่อยู่ในรูปสมการเชิงเส้น $Y = a + bX$ คือ

$$\sum_{i=1}^{35} Y_i = an + \sum_{i=1}^{35} bX_i \qquad \text{----------- (1)}$$

$$\sum_{i=1}^{35} X_i Y_i = a\sum_{i=1}^{35} X_i + b\sum_{i=1}^{35} X_i^2 \qquad \text{----------- (2)}$$

จากตารางจะได้

$$\sum_{i=1}^{35} X_i = -137.70, \qquad \sum_{i=1}^{35} Y_i = -398.00,$$

$$\sum_{i=1}^{35} X_i Y_i = 5069.20 \text{ และ } \sum_{i=1}^{35} X_i^2 = 801.01$$

แทนผลรวมในสมการ (1) และ (2) เพื่อหาค่าคงตัว a และ b

$$-398.00 = 35.00a - 137.70b \qquad \text{----------- (3)}$$

$$5069.20 = -137.70a + 801.01b \qquad \text{----------- (4)}$$

จากสมการ (3) และ (4) คำตอบของระบบสมการคือ $a \approx 41.79$ และ $b \approx 13.51$
ดังนั้น สมการที่ใช้ประมาณค่าของ Y จากค่าของ X คือ

$$\hat{Y} = 41.79 + 13.51X$$

4) เมื่อน้ำหนักลดลง 5 กิโลกรัม แสดงว่าต้องแทน X ในสมการทำนายด้วย -5
จะได้ $\hat{Y} = 41.79 + 13.51(-5) = -25.76$
ดังนั้น เมื่อน้ำหนักลดลง 5 กิโลกรัม ปริมาณไทรกลีเซอไรด์จะลดลงประมาณ 25.76 มิลลิกรัม/เดซิลิตร

5) เนื่องจากค่าของ $X \geq 10$ จะอยู่นอกช่วงข้อมูลที่นำมาทดลอง
ดังนั้น จึงยังไม่สมเหตุสมผลที่จะพยากรณ์ค่า Y จากค่าของ $X \geq 10$
เพราะค่าของ X ที่อยู่นอกช่วงของข้อมูลที่นำมาศึกษาเมื่อนำมาพยากรณ์ค่าของ Y อาจทำให้ค่าพยากรณ์ไม่แม่นยำ

หมายเหตุ จากข้อ 2) พบว่ามีข้อมูลผิดปกติอยู่ด้วย ทำให้สมการทำนายในข้อ 3) อาจมีหรือให้ผลการทำนายที่มีความน่าเชื่อถือน้อย

สรุปการวิเคราะห์ความสัมพันธ์เชิงฟังก์ชัน ถ้าทราบรูปแบบหรือลักษณะของความสัมพันธ์จากแผนภาพการกระจายแล้ว จะสร้างสมการแสดงความสัมพันธ์โดยการประมาณค่าคงตัวต่าง ๆ เช่น ความชันของเส้นตรง ระยะตัดแกน Y หรือค่าคงตัวแบบอื่น ๆ เมื่อกราฟมีรูปแบบไม่เป็นเส้นตรง จุดประสงค์หลักคือการนำสมการมาหาสารสนเทศ 2 ประการดังนี้

1. สรุปความสัมพันธ์ของสมการหรือรูปแบบที่ได้ โดยอธิบายความสัมพันธ์ระหว่างสองตัวแปรนั่นเอง เช่น ความชันของเส้นตรงให้สารสนเทศว่า Y โดยเฉลี่ยจะเปลี่ยนไปเท่ากับความชันของเส้นตรงที่คำนวณได้ ถ้า X เปลี่ยนไป 1 หน่วย

2. พยากรณ์ค่าของ Y โดยใช้ค่าของ X ที่สนใจหนึ่งค่าซึ่งค่าของ X ที่สนใจนั้น ถ้าอยู่ในพิสัยของตัวแปร X ที่นำมาสร้างสมการแล้วจะเรียกว่า การพยากรณ์ในช่วง (interpolation) แต่ถ้าค่าของ X อยู่นอกพิสัยของตัวแปร X แล้ว จะเรียกว่า การพยากรณ์นอกช่วง (extrapolation) ซึ่งประการหลังจะพบความคลาดเคลื่อนมากกว่าประการแรก และจากประโยชน์ที่ได้นี้จึงอาจเรียกตัวแปรอิสระ (X) ได้อีกชื่อหนึ่งว่า ตัวแปรอธิบาย (explanatory variable) หรือตัวแปรพยากรณ์ (predictor variable) ส่วนตัวแปรตาม (Y) อาจเรียกว่า ตัวแปรตอบสนอง (response variable)

3.4 ความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่อยู่ในรูปอนุกรมเวลา

ข้อมูลที่อยู่ในรูปอนุกรมเวลา (time series) คือ ข้อมูลที่แสดงความเปลี่ยนแปลงตามลำดับก่อนหลังของช่วงเวลาที่ข้อมูลนั้น ๆ เกิดขึ้น ซึ่งปกติแล้วข้อมูลนั้น ๆ มักจะเกิดขึ้นในช่วงเวลาเท่า ๆ กัน เช่น ปริมาณข้าวที่ประเทศไทยผลิตได้ในแต่ละปี จำนวนเงินที่ร้านค้าแห่งหนึ่งขายได้ในแต่ละเดือน หรืออุณหภูมิเฉลี่ยในแต่ละวันของจังหวัดกาญจนบุรี เป็นต้น จากตัวอย่างที่ยกมาแสดงนี้ จะเห็นได้ว่า ปริมาณข้าวที่ประเทศไทยผลิตได้ จำนวนเงินที่ร้านค้าแห่งหนึ่งขายได้ หรืออุณหภูมิเฉลี่ยในแต่ละวันของจังหวัดกาญจนบุรี เป็นข้อมูลที่มีความเกี่ยวข้องกับช่วงเวลาทั้งสิ้น นั่นคือ ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลที่สนใจศึกษา (Y) กับช่วงเวลาที่ข้อมูลนั้นเกิด (t) เขียนได้เป็น

$$Y = f(t)$$

เมื่อ t เป็นตัวแปรอิสระ และ Y เป็นตัวแปรตาม

ในการสร้างความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่อยู่ในรูปอนุกรมเวลานี้ ใช้วิธีเดียวกันกับวิธีที่ใช้ในการสร้างความสัมพันธ์เชิงฟังก์ชันของตัวแปรอิสระและตัวแปรตามที่ได้กล่าวมาแล้ว นั่นคือ กำหนดรูปของความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูลที่สนใจศึกษากับช่วงเวลาที่ข้อมูลนั้น ๆ เกิดขึ้น จากแผนภาพการกระจายของข้อมูลที่มีอยู่หรือที่เลือกขึ้นมาเป็นตัวแทนจากข้อมูลที่มีอยู่ทั้งหมด แล้วประมาณค่าของค่าคงตัวจากความสัมพันธ์ที่ได้กำหนดไว้นั้น โดยใช้วิธีกำลังสองน้อยสุดที่ได้กล่าวมาแล้ว

สำหรับการแทนค่า t ของช่วงเวลาที่เท่า ๆ กัน ซึ่งอาจจะเป็นวัน เดือน พุทธศักราช หรือ คริสต์ศักราช นั้น โดยทั่ว ๆไป ถ้าจำนวนช่วงเวลาที่นำมาสร้างความสัมพันธ์เป็นจำนวนคี่มักจะกำหนดให้ช่วงเวลาที่อยู่ตรงกลางเป็น 0 และช่วงเวลาที่อยู่ถัดขึ้นไปก่อนหน้าช่วงเวลาที่กำหนดให้เป็น 0 นี้เป็น –1, –2, –3, ... ตามลำดับ ส่วนช่วงเวลาที่อยู่ถัดลงมา ซึ่งเป็นช่วงเวลาที่เกิดขึ้นภายหลังจะกำหนดเป็น 1, 2, 3, ... ตามลำดับ ดังนั้น ในหัวข้อนี้ จึงเป็นการศึกษาความสัมพันธ์เชิงฟังก์ชัน โดยเน้นเฉพาะแนวโน้มตามเวลาที่ใช้เป็นตัวแปรอิสระ (t) เพื่อใช้รูปแบบสมการที่ประมาณได้มาพยากรณ์ค่าของตัวแปรตาม (Y) ต่อไป

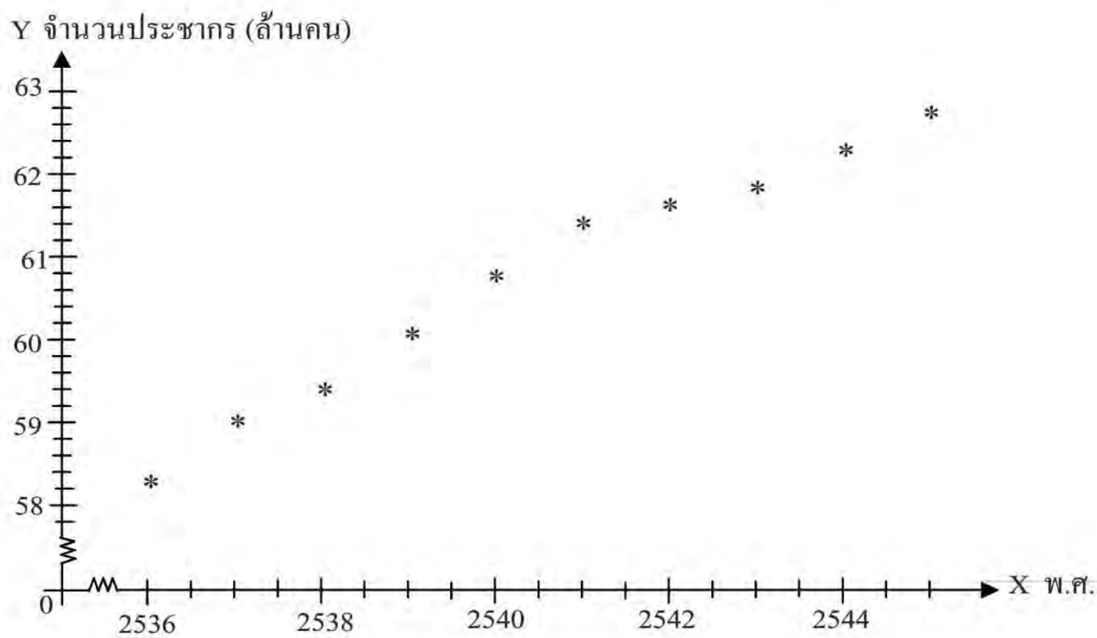
ในกรณีที่จำนวนช่วงเวลาที่นำมาสร้างความสัมพันธ์เป็นจำนวนคู่มักจะแทนสองช่วงเวลาที่อยู่ตรงกลางเป็น –1 และ 1 และแทนช่วงเวลาที่อยู่ถัดขึ้นไปก่อนหน้าช่วงเวลาที่กำหนดให้เป็น –3, –5, –7, ... ตามลำดับ ส่วนช่วงเวลาที่เกิดขึ้นภายหลังแทนด้วย 3, 5, 7, ... ตามลำดับ การกำหนดค่าดังกล่าวขึ้นมาใช้แทนตัวแปรซึ่งเป็นช่วงเวลานี้ ก็เพื่อให้การคำนวณหาค่าคงตัวทำได้สะดวกและรวดเร็วขึ้น เพราะผลรวมของทุก ๆ ค่าของ t จะเท่ากับ 0 และในการทำนายค่าของตัวแปรตาม จะต้องเปลี่ยนช่วงเวลาให้อยู่ในรูปของค่า t ที่กำหนดให้โดยวิธีดังกล่าวด้วย

ตัวอย่างที่ 1 จงสร้างความสัมพันธ์เชิงฟังก์ชันของจำนวนประชากรของประเทศไทย และจงใช้ความสัมพันธ์ที่หาได้พยากรณ์จำนวนประชากรของประเทศไทยในปี พ.ศ. 2555 จากข้อมูลระหว่าง ปี พ.ศ. 2536 – 2545 ที่กำหนดให้ต่อไปนี้

พ.ศ.	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545
จำนวนประชากร (ล้านคน)	58.35	59.10	59.46	60.12	60.82	61.47	61.66	61.88	62.31	62.80

ที่มา : กรมการปกครอง กระทรวงมหาดไทย

วิธีทำ เขียนแผนภาพการกระจายระหว่างจำนวนประชากรและช่วงเวลาได้ดังนี้



จากแผนภาพการกระจายของข้อมูล อาจอนุโลมได้ว่าความสัมพันธ์เชิงฟังก์ชันระหว่างจำนวนประชากรของท้องที่แห่งนี้ (Y) และช่วงเวลาของข้อมูล (t) อยู่ในรูปฟังก์ชันเอกซ์โพเนนเชียลซึ่งสมการเอกซ์โพเนนเชียลมีรูปสมการทั่วไปคือ $Y = ab^t$ ซึ่งมีสมการเชิงเส้นของลอการิทึมที่เกี่ยวข้องคือ $\log Y = \log a + t \log b$ และสมการปกติคือ

$$\sum_{i=1}^{n} \log y_i = n \log a + (\log b)\sum_{i=1}^{n} t_i \qquad \text{----------- (1)}$$

$$\sum_{i=1}^{n} t_i \log y_i = (\log a)\sum_{i=1}^{n} t_i + (\log b)\sum_{i=1}^{n} t_i^2 \qquad \text{----------- (2)}$$

จากสมการปกติ (1) และ (2) หาพจน์ต่าง ๆ ที่ใช้ในการคำนวณค่าคงตัว $\log a$ และ $\log b$ ดังตารางต่อไปนี้

พ.ศ.	จำนวนประชากร (y_i)	$\log y_i$	t_i	t_i^2	$t_i \log y_i$
2536	58.35	1.7661	–9	81	–15.8940
2537	59.10	1.7716	–7	49	–12.4012
2538	59.46	1.7742	–5	25	–8.8710
2539	60.12	1.7790	–3	9	–5.3370
2540	60.82	1.7841	–1	1	–1.7841
2541	61.47	1.7887	1	1	1.7887
2542	61.66	1.7900	3	9	5.3700
2543	61.88	1.7916	5	25	8.9580
2544	62.31	1.7946	7	49	12.5622
2545	62.80	1.7980	9	81	16.1820
รวม		$\sum_{i=1}^{10} \log y_i \approx 17.8378$	$\sum_{i=1}^{10} t_i = 0$	$\sum_{i=1}^{10} t_i^2 = 330$	$\sum_{i=1}^{10} t_i \log y_i \approx 0.5736$

แทนค่า $\sum_{i=1}^{n} t_i$, $\sum_{i=1}^{n} t_i^2$, $\sum_{i=1}^{n} \log y_i$, $\sum_{i=1}^{n} t_i \log y_i$ ในสมการ (1) และ (2) ด้วยค่าในตารางจะได้

$$17.8378 = 10 \log a + 0 \qquad \text{---------- (3)}$$

$$0.5736 = 0 + 330 \log b \qquad \text{---------- (4)}$$

จาก (3) และ (4) คำตอบของระบบสมการคือ

$$\log a = \frac{17.8378}{10} = 1.7838$$

$$\log b = \frac{0.5736}{330} = 0.0017$$

ดังนั้น สมการสำหรับการประมาณจำนวนประชากรของประเทศไทย คือ

$$\log \hat{Y} = 1.7838 + 0.0017t$$

เนื่องจาก ปี พ.ศ. 2555 แสดงว่าต้องแทน t ด้วย 29 ในสมการทำนาย

จะได้ $\log \hat{Y} = 1.7838 + 0.0017(29) = 1.8331$

ดังนั้น $\hat{Y} = 68.0926$

นั่นคือ จำนวนประชากรของประเทศไทยในปี พ.ศ. 2555 จะมีประมาณ 68.0926 ล้านคน หรือประมาณ 68,092,600 คน

หมายเหตุ จากตัวอย่างที่ 1 อาจพิจารณาได้ว่าข้อมูลดังกล่าวมีความสัมพันธ์เชิงเส้นซึ่งมีสมการทั่วไปคือ $Y = a + bt$ เมื่อสร้างสมการปกติและคำนวณพจน์ต่าง ๆ ที่เกี่ยวข้องจากตารางแล้วจะได้ สมการทำนายในรูปเชิงเส้นนี้ คือ

$\hat{Y} = 60.797 + 0.242t$

ดังนั้น ปี พ.ศ. 2555 แสดงว่าต้องแทน t ด้วย 29 ในสมการทำนาย

จะได้ $\hat{Y} = 60.797 + 0.242(29) = 67.815$

นั่นคือ จำนวนประชากรของประเทศไทยในปี พ.ศ. 2555 จะมีประมาณ 67,815,000 คน

แบบฝึกหัด 3.4

1. ตารางต่อไปนี้แสดงปริมาณสัตว์น้ำ (พันตัน) ที่สำคัญ 10 ชนิดที่จับได้ในธรรมชาติในประเทศไทย ในปี พ.ศ.2542 และ พ.ศ.2543

พ.ศ. 2542	164.1	47.9	51.6	59.9	10.1	14.1	44.3	182.8	134.7	206.0
พ.ศ. 2543	152.9	35.2	58.2	53.4	11.0	12.8	42.6	164.0	143.1	197.9

ที่มา : กรมประมง กระทรวงเกษตรและสหกรณ์

1) จงเขียนแผนภาพการกระจายของข้อมูล
2) จงเขียนกราฟที่ใช้แสดงความสัมพันธ์ระหว่างปริมาณสัตว์น้ำแต่ละชนิดที่จับได้ในปี พ.ศ.2542 และ พ.ศ.2543 ในข้อ 1)
3) ถ้าปริมาณสัตว์น้ำชนิดหนึ่งที่จับได้ในปี พ.ศ.2542 เท่ากับ 50,000 ตัน ให้พยากรณ์ปริมาณสัตว์ชนิดนี้ที่จับได้ในปี พ.ศ.2543
4) ถ้าปริมาณสัตว์น้ำชนิดหนึ่งที่จับได้ในปี พ.ศ.2543 เท่ากับ 85,000 ตัน ให้พยากรณ์ปริมาณสัตว์ชนิดนี้ที่จับได้ในปี พ.ศ.2542

2. ราคายางพาราแผ่นดิบชั้น 3 ที่เกษตรกรไทยขายได้ในปี พ.ศ.2546 เป็นดังนี้

เดือน	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.
ราคา (บาท/กก.)	32.41	36.32	40.00	37.29	36.64	38.03	36.59	38.22	39.41	45.41	43.39	41.36

ที่มา : สถาบันวิจัยยาง

จงพยากรณ์ราคายางพาราแผ่นดิบชั้น 3 ที่เกษตรกรไทยขายได้แต่ละเดือนในปี พ.ศ.2547

3. ข้อมูลต่อไปนี้แสดงปริมาณการนำเข้าข้าว (หมื่นตัน) ที่ประเทศหนึ่งในทวีปเอเชียนำเข้าจากประเทศไทยตั้งแต่ปี พ.ศ. 2538 – 2545

พ.ศ.	2538	2539	2540	2541	2542	2543	2544	2545
ปริมาณ (หมื่นตัน)	23.7	23.9	24.8	25.5	25.2	24.6	23.1	23.5

ที่มา : สำนักการค้าข้าวต่างประเทศ

จงพยากรณ์ปริมาณการนำเข้าข้าวที่ประเทศนี้นำเข้าจากประเทศไทย ในปี พ.ศ. 2550

4. ถ้าจำนวนแสดงความสัมพันธ์ระหว่างต้นทุนการผลิตสินค้าต่อหน่วย (บาท) กับจำนวนสินค้าที่ผลิตได้ของโรงงานผลิตของเด็กเล่นแห่งหนึ่งเป็นดังนี้

จำนวนที่ผลิต : X	1	2	3	4	5	6	7	8	9	10
ต้นทุนการผลิตต่อหน่วย : Y	58	56	55	50	45	40	37	30	26	20

จงพยากรณ์ต้นทุนการผลิตสินค้าต่อหน่วย ถ้าจำนวนสินค้าที่ผลิตได้เป็น 12 หน่วย

5. ในการทอดลูกเต๋าสองลูก 12 ครั้ง ปรากฏว่าแต้มที่ได้จากการทอดลูกเต๋าในแต่ละครั้งเป็นดังนี้

ลูกเต๋าลูกที่ 1	1	3	4	2	5	1	6	5	2	4	5	6
ลูกเต๋าลูกที่ 2	2	2	3	1	4	2	4	6	2	5	5	4

1) จงเขียนกราฟแสดงความสัมพันธ์ระหว่างแต้มที่ได้จากการทอดลูกเต๋าสองลูกในแต่ละครั้งความสัมพันธ์เชิงฟังก์ชันระหว่างแต้มที่ได้จากการทอดลูกเต๋าสองลูกควรอยู่ในรูปแบบใด
2) ท่านคิดว่าโดยแท้จริงแล้วแต้มที่ได้จากการทอดลูกเต๋าทั้งสองในแต่ละครั้งควรจะมีความสัมพันธ์กันหรือไม่

6. มูลค่าของสินค้าที่ประเทศไทยนำเข้าจากต่างประเทศระหว่างปี พ.ศ.2536 – 2545 เป็นดังนี้

พ.ศ.	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545
มูลค่าสินค้า (พันล้านบาท)	11.71	13.69	18.35	18.57	19.24	17.74	19.10	24.94	27.55	27.75

1) จงหาสมการที่เหมาะสมเพื่อใช้ประมาณมูลค่าสินค้าที่ประเทศไทยนำเข้าจากต่างประเทศ
2) จงพยากรณ์มูลค่าสินค้าที่ประเทศไทยนำเข้าจากต่างประเทศระหว่างปี พ.ศ. 2546 – 2550

7. จงพิจารณาข้อความที่กำหนดให้ว่าเป็นจริงหรือเป็นเท็จ ถ้าเป็นเท็จจงบอกเหตุผล
   1) ถ้าข้อมูลประกอบด้วยตัวแปรสองตัวแล้ว ตัวแปรทั้งสองนั้นจะต้องมีความสัมพันธ์เชิงฟังก์ชันกันเสมอ
   2) ในการกำหนดรูปสมการเพื่อสร้างความสัมพันธ์ระหว่างตัวแปรสองตัวควรจะสร้างแผนภาพการกระจายของข้อมูลทุกครั้ง
   3) ในการสร้างความสัมพันธ์เชิงฟังก์ชันของข้อมูลที่ประกอบด้วยตัวแปร 2 ตัว ถ้าตัวแปรตัวใดตัวหนึ่งเป็นตัวแปรเชิงคุณภาพแล้ว จะไม่สามารถสร้างความสัมพันธ์ระหว่างตัวแปรทั้งสองได้
   4) การทำนายค่าของตัวแปรตัวหนึ่งจากสมการแสดงความสัมพันธ์ระหว่างตัวแปร 2 ตัว เมื่อทราบค่าของตัวแปรอีกตัวหนึ่ง ค่าที่ได้จากการทำนายของตัวแปรนั้นจะต้องเท่ากับค่าที่ควรจะเป็นจริงเสมอ
   5) ในการสร้างความสัมพันธ์เชิงฟังก์ชันของข้อมูล ไม่ว่าข้อมูลจะมีจำนวนน้อยเพียงใดก็สามารถสร้างความสัมพันธ์ได้เสมอ
   6) ข้อมูลอนุกรมเวลา คือข้อมูลที่ประกอบด้วยตัวแปรเพียงตัวเดียว และตัวแปรนี้ใช้แทนเวลาเท่านั้น
   7) ในการกำหนดค่าของตัวแปรที่ใช้แทนเวลา จะต้องกำหนดให้เป็นจำนวนบวกเท่านั้น
   8) ถ้าความสัมพันธ์ระหว่างรายได้ X และรายจ่าย Y โดยเฉลี่ยต่อเดือนของครอบครัวที่อาศัยอยู่ในจังหวัดชลบุรีเป็น $\hat{Y} = 0.85X$ ครอบครัวของนายชำนาญซึ่งอาศัยอยู่ในจังหวัดระยองและมีรายได้เดือนละ 2,000 บาท จะมีรายจ่ายประมาณ 1,700 บาท
   9) การกำหนดรูปแบบความสัมพันธ์เชิงฟังก์ชันจากแผนภาพการกระจายของข้อมูลอาจมีความแตกต่างกันได้ ถ้าผู้กำหนดเป็นคนละคนกัน
   10) สมการปกติที่ใช้ในการประมาณค่าคงตัว จากความสัมพันธ์เชิงฟังก์ชันระหว่างตัวแปรที่อยู่ในรูปสมการแบบต่าง ๆ จะต้องมีจำนวนอย่างน้อยเท่ากับจำนวนค่าคงตัว

8. จากตารางแสดงข้อมูลเกี่ยวกับระยะเวลา (X : สัปดาห์) ที่คนงานใช้ฝึกทักษะในการเชื่อมต่ออุปกรณ์อิเล็กทรอนิกส์ขนาดเล็ก และข้อมูลจำนวนชิ้นงานที่คนงานทำไม่สำเร็จ (Y : ชิ้น) คนงานทั้งหมดเป็นตัวอย่างที่สุ่มมาจากโรงงานแห่งหนึ่ง 12 คน

คนงานคนที่	1	2	3	4	5	6	7	8	9	10	11	12
ระยะเวลาฝึกทักษะ (X)	7	9	6	14	8	12	10	4	2	11	1	8
จำนวนชิ้นงานที่ทำไม่สำเร็จ (Y)	26	20	28	16	23	18	24	26	38	22	32	25
$\hat{Y}$												
$Y-\hat{Y}$												

1) จงเขียนแผนภาพการกระจายของข้อมูล Y และ X
2) ถ้าสมการที่ใช้ประมาณจำนวนชิ้นงานที่ทำไม่สำเร็จจากระยะเวลาฝึกทักษะคนงาน คือ

$$\hat{Y} = 35.57 - 1.40X$$

จงอธิบายแนวโน้มการเปลี่ยนแปลงของประมาณจำนวนชิ้นงานที่ทำไม่สำเร็จ ($\hat{Y}$) เมื่อระยะเวลาฝึกทักษะคนงาน (X) มีการเปลี่ยนแปลง
3) จงเติมค่า $\hat{Y}$ ในตารางให้สมบูรณ์ โดยการแทนค่า X ในสมการด้วยข้อมูลของคนงานแต่ละคนและหาค่าของ $Y-\hat{Y}$
4) จงอธิบายว่าสมการทำนายที่เหมาะสมหรือไม่เหมาะสมในการพยากรณ์ มีความเกี่ยวข้องกับค่าของ $Y-\hat{Y}$ ที่อาจจะมีค่าสูงหรือค่าต่ำ อย่างไร

9. จากตัวอย่างที่ 3 ในหัวข้อ 3.3 และข้อมูลในตาราง ถ้าตัดข้อมูลผิดปกติในลำดับที่ 31 ออกไป จงตอบคำถามข้อ 1) – 4) ใหม่ และจงเปรียบเทียบผลลัพธ์ที่ได้จากตัวอย่างที่ 3 กับผลลัพธ์ที่ได้นี้ โดยอภิปรายผลลัพธ์ของค่าผิดปกติว่าอาจมีผลกระทบต่อสมการความสัมพันธ์ที่ได้หรือไม่

ลำดับที่	base(X)	wk8(X)	base(Y)	wk8(Y)	X	Y	X^2	(X)(Y)
1	84.0	82.4	90	131	–1.6	41	2.56	–65.6
2	88.8	87.0	137	82	–1.8	–55	3.24	99.0
3	87.0	81.8	182	152	–5.2	–30	27.04	156.0
4	84.5	80.4	72	72	–4.1	0	16.81	0.0
5	69.4	69.0	143	126	–0.4	–17	0.16	6.8
6	104.7	102.0	96	157	–2.7	61	7.29	–164.7
7	90.0	87.6	115	88	–2.4	–27	5.76	64.8
8	89.4	86.8	124	123	–2.6	–1	6.76	2.6
9	95.2	92.8	188	255	–2.4	67	5.76	–160.8
10	108.1	100.9	167	87	–7.2	–80	51.81	576.0
11	93.9	90.2	143	213	–3.7	70	13.69	–259.0
12	83.4	75.0	143	102	–8.4	–41	70.56	344.4
13	104.4	102.9	276	313	–1.5	37	2.25	–55.5
14	103.7	95.7	84	84	–8.0	0	64.00	0.0
15	99.2	99.2	142	135	0.0	–7	0.00	0.0
16	95.6	88.5	64	114	–7.1	50	50.41	–355.0
17	126.0	123.2	226	152	–2.8	–74	7.84	207.2
18	103.7	95.5	199	120	–8.2	–79	67.24	647.8
19	133.1	130.8	212	156	–2.3	–56	5.29	128.8
20	85.0	80.0	268	250	–5.0	–18	25.00	90.0
21	83.8	77.9	111	107	–5.9	–4	34.81	23.6
22	104.5	98.3	132	117	–6.2	–15	38.44	93.0
23	76.8	73.2	165	96	–3.6	–69	12.96	248.4
24	90.5	88.9	57	63	–1.6	6	2.56	–9.6
25	106.9	103.7	163	131	–3.2	–32	10.24	102.4
26	81.5	78.9	111	54	–2.6	–57	6.76	148.2
27	96.5	94.9	300	241	–1.6	–59	2.56	94.4
28	103.0	97.2	192	124	–5.8	–68	33.64	394.4
29	127.5	124.7	176	215	–2.8	39	7.84	–109.2
30	103.2	102.0	146	138	–1.2	–8	1.44	9.6
31	113.5	115.0	446	795	1.5	349	2.25	523.5
32	107.0	99.2	232	63	–7.8	–169	60.84	1318.2
33	106.0	103.5	255	204	–2.5	–51	6.25	127.5
34	114.9	105.3	187	144	–9.6	–43	92.16	412.8
35	103.4	96.0	154	96	–7.4	–58	54.76	429.2

บรรณานุกรม

ส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, สถาบัน. (2544). **หนังสือเรียนคณิตศาสตร์ ค 016 ชั้นมัธยมศึกษาตอนปลาย หลักสูตรมัธยมศึกษาตอนปลาย พุทธศักราช 2524(ฉบับปรับปรุง พ.ศ. 2533)**. พิมพ์ครั้งที่ 9. กรุงเทพมหานคร : โรงพิมพ์คุรุสภาลาดพร้าว.

ส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, สถาบัน. (2545). **หนังสือเรียนคณิตศาสตร์ ค 014 ชั้นมัธยมศึกษาตอนปลาย หลักสูตรมัธยมศึกษาตอนปลาย พุทธศักราช 2524(ฉบับปรับปรุง พ.ศ. 2533)**. พิมพ์ครั้งที่ 12. กรุงเทพมหานคร : โรงพิมพ์คุรุสภาลาดพร้าว.

ส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, สถาบัน. (2545). **หนังสือเรียนคณิตศาสตร์ ค 012 ชั้นมัธยมศึกษาตอนปลาย หลักสูตรมัธยมศึกษาตอนปลาย พุทธศักราช 2524(ฉบับปรับปรุง พ.ศ. 2533)**. พิมพ์ครั้งที่ 11. กรุงเทพมหานคร : โรงพิมพ์คุรุสภาลาดพร้าว.

สถิติแห่งชาติ, สำนักงาน. (2547). **สมุดสถิติรายปีประเทศไทย พ.ศ. 2546**. กรุงเทพมหานคร : โรงพิมพ์ดอกเบี้ย.

สถิติแห่งชาติ, สำนักงาน. (2548). **สารสถิติ ปีที่ 16 ฉบับที่ 3 เดือนมีนาคม 2548**. กรุงเทพมหานคร : โรงพิมพ์ดอกเบี้ย.

Ann E.Watkins, Richard L. Scheaffer, George W. Cobb. (2004). **Statistics in Action**. CA, Key Curriculum Press.

Devore J. and Farnum N. (2005). **Applied Statistics for Engineers and Scientists**. CA, Thomson Brooks / Cole.

Mendenhall et al (1981). **Mathematical Statistics with Applications**. Boston. Dukberry Press. (2nd ed.)

ภาคผนวก

ตารางค่าลอการิทึม

N	0	1	2	3	4	5	6	7	8	9
1.0	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
1.1	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
1.2	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
1.3	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
1.4	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
1.5	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
1.6	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
1.7	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
1.8	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
1.9	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
2.0	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
2.1	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
2.2	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
2.3	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
2.4	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
2.5	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
2.6	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
2.7	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
2.8	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
2.9	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757

ตารางค่าลอการิทึม (ต่อ)

N	0	1	2	3	4	5	6	7	8	9
3.0	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
3.1	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
3.2	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
3.3	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
3.4	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
3.5	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
3.6	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
3.7	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
3.8	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
3.9	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
4.0	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
4.1	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
4.2	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
4.3	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
4.4	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
4.5	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
4.6	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
4.7	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
4.8	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
4.9	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981

ตารางค่าลอการิทึม (ต่อ)

N	0	1	2	3	4	5	6	7	8	9
5.0	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
5.1	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
5.2	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
5.3	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
5.4	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
5.5	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
5.6	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
5.7	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
5.8	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
5.9	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
6.0	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
6.1	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
6.2	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
6.3	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
6.4	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
6.5	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
6.6	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
6.7	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
6.8	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
6.9	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445

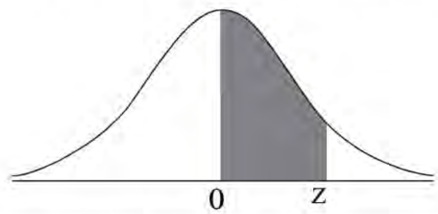
ตารางค่าลอการิทึม (ต่อ)

N	0	1	2	3	4	5	6	7	8	9
7.0	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
7.1	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
7.2	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
7.3	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
7.4	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
7.5	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
7.6	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
7.7	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
7.8	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
7.9	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
8.0	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
8.1	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
8.2	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
8.3	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
8.4	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
8.5	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
8.6	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
8.7	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
8.8	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
8.9	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538

ตารางค่าลอการิทึม (ต่อ)

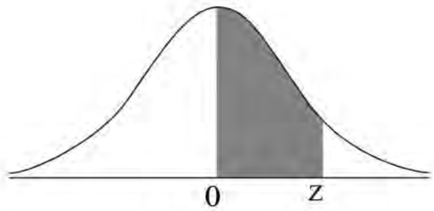
N	0	1	2	3	4	5	6	7	8	9
9.0	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
9.1	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
9.2	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
9.3	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
9.4	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
9.5	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
9.6	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
9.7	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
9.8	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
9.9	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

ตารางแสดงพื้นที่ใต้เส้นโค้งปกติ



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

ตารางแสดงพื้นที่ใต้เส้นโค้งปกติ (ต่อ)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

บัญชีศัพท์

		หน้า
แนวโน้มทางบวก	positive trend	108
แนวโน้มทางลบ	negative trend	108
กลุ่มของจุด	points cluster	108
เส้นที่คาดไว้	imaginary line	108
วิธีกำลังสองน้อยสุด	method of least squares	109
สมการปกติ	normal equations	110
การพยากรณ์ในช่วง	interpolation	122
การพยากรณ์นอกช่วง	extrapolation	122
ตัวแปรอธิบาย, ตัวแปรพยากรณ์	explanatory variable, predictor variable	122
ตัวแปรตอบสนอง	response variable	122
อนุกรมเวลา	time series	123

บัญชีสัญลักษณ์

บทที่ 1

$\overline{X}$	ค่าเฉลี่ยเลขคณิตของตัวอย่าง
μ	ค่าเฉลี่ยเลขคณิตของประชากร
$\sum$	ผลบวกของข้อมูล
G.M.	ค่าเฉลี่ยเรขาคณิต
H.M.	ค่าเฉลี่ยฮาร์มอนิก
Q.D.	ส่วนเบี่ยงเบนควอร์ไทล์
M.D.	ส่วนเบี่ยงเบนเฉลี่ย
σ	ส่วนเบี่ยงเบนมาตรฐานของประชากร
S.D. หรือ s	ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
σ^2	ความแปรปรวนของประชากร
s^2	ความแปรปรวนของตัวอย่าง
N	จำนวนประชากร
n	จำนวนตัวอย่าง

บทที่ 2

z	คะแนนมาตรฐาน

คณะกรรมการจัดทำสื่อการเรียนการสอนวิชาคณิตศาสตร์เพิ่มเติม

ระดับมัธยมศึกษาตอนปลาย

นายประสาท สอ้านวงศ์	ข้าราชการบำนาญ
นางวีรานันท์ พงศาภักดี	มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์
นายกมล บุษบา	มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต
นางสาวสิริพร ทิพย์คง	มหาวิทยาลัยเกษตรศาสตร์
นายธนาวุฒิ ศรีวิรักษ์	มหาวิทยาลัยทักษิณ
นายวิโรจน์ รัตตากร	โรงเรียนสาธิตแห่งมหาวิทยาลัยเชียงใหม่
นางสาวจำเริญ เจียวหวาน	โรงเรียนมหิดลวิทยานุสรณ์
นางสาวขวัญตา พันธุ์บ้านแหลม	โรงเรียนศรีวิชัยวิทยา
นางสาวรุ่งฟ้า จันท์จารุภรณ์	มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
นายสุรชัย บุญเรือง	โรงเรียนเบ็ญจะมะมหาราช
นางนงนุช ผลทวี	โรงเรียนทับปุดวิทยา
นางสาวราตรี คำเทียนทอง	ข้าราชการบำนาญ
นายกันตภณ คูหาพัฒนกุล	มหาวิทยาลัยเกษตรศาสตร์
นายสุปรีดี แดงสกุล	
นางสาวจารุวรรณ แสงทอง	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นายดนัย ยังคง	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางชมัยพร ตั้งตน	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวอลงกรณ์ ตั้งสงวนธรรม	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวนวลจันทร์ ผมอุดทา	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวศศิวรรณ เมลืองนนท์	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวปฐมาภรณ์ อวชัย	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นายอนุชิต อารมณ์สาวะ	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี

คณะบรรณาธิการ

นายประสาท สอ้านวงศ์

ผู้จัดพิมพ์ต้นฉบับ

นางสาวปิยาภรณ์ ทองมาก	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี

คำสำคัญในการสืบค้น

คำสำคัญในการสืบค้น

คำสำคัญในการสืบค้น

คำสำคัญในการสืบค้น

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
กระทรวงศึกษาธิการ

ศึกษาภัณฑ์พาณิชย์
พิมพ์ที่โรงพิมพ์ สกสค. ลาดพร้าว
นายสันติภาพ อินทรพัฒน์ ผู้พิมพ์และผู้โฆษณา
๕๕๐๐๒๒๔
www.suksapan.or.th