

บทที่ 1 ความหมายของสถิติและข้อมูล

1. ความหมายของสถิติศาสตร์และข้อมูล

- ข้อมูลที่เกิดขึ้นจริงในปัจจุบันมีหลายรูปแบบ และสามารถนำมาใช้ให้เกิดประโยชน์ในด้านต่างๆ เช่น ในการดำเนินธุรกิจ ผู้ประกอบการอาจมีความต้องการที่จะปรับปรุงคุณภาพสินค้าที่ผลิตขึ้น หรือหาแนวทางใหม่ๆ ในการบริการแก่ลูกค้า ดังนั้น ผู้ประกอบการจึงต้องมีการเก็บข้อมูลต่างๆ ที่เกี่ยวกับธุรกิจของตน ไม่ว่าจะเป็นการผลิตสินค้า การเก็บรักษาสินค้า และการขนส่งสินค้า การจ้างบุคลากรในส่วนต่าง ๆ รายรับและรายจ่ายในแต่ละ ประเภท และอื่น ๆ หากพิจารณาการประกอบกิจการธุรกิจร้านขายเครื่องดื่มและขนมหวาน ผู้ประกอบการจะต้องเก็บข้อมูลปริมาณวัตถุดิบที่ต้องใช้ในการผลิต เช่น เมล็ดกาแฟ นม น้ำตาล แป้งสาลี เนย ผลไม้หรือธัญพืชที่ใช้ในการทำขนม ราคาและเวลาที่ใช้ในการสั่งซื้อ วัตถุดิบแต่ละชนิด แรงงานหรือทรัพยากรอื่น ๆ ที่ต้องใช้ในการผลิต เครื่องดื่มและขนม แต่ละชนิด ปริมาณลูกค้าที่เข้ามาใช้บริการในแต่ละช่วงเวลาของวัน สัปดาห์ เดือน หรือช่วงเทศกาลต่าง ๆ รวมทั้งระยะเวลาที่ลูกค้าเข้ามาใช้บริการในร้าน นอกจากนี้ ผู้ประกอบการ อาจสอบถามความพึงพอใจของลูกค้า โดยเก็บข้อมูลเกี่ยวกับเพศ อายุ เครื่องดื่มหรือขนม ที่ลูกค้านิยม ระดับความพึงพอใจต่อการให้บริการของพนักงาน รวมทั้งความเห็นอื่น ๆ เกี่ยวกับร้าน แล้วนำข้อมูลเหล่านี้มาวิเคราะห์และใช้ประกอบการพิจารณาเพื่อออกแบบ ผลิตภัณฑ์ใหม่ให้ตอบสนองความต้องการของลูกค้าได้มากขึ้น หรือปรับปรุงการให้บริการ การตัดสินใจทางธุรกิจบนพื้นฐานของการวิเคราะห์ข้อมูลดังตัวอย่างข้างต้นเป็นปัจจัยสำคัญ ที่ส่งผลต่อความสำเร็จของธุรกิจต่าง ๆ ในปัจจุบัน และความรู้ที่อยู่เบื้องหลังการวิเคราะห์ ข้อมูลดังกล่าวก็คือความรู้ทางด้านสถิติศาสตร์

1.1 สถิติศาสตร์

- สถิติศาสตร์ (statistics) หมายถึง วิชาที่ว่าด้วยการเก็บรวบรวมข้อมูล วิเคราะห์ข้อมูล และสรุปผลจากข้อมูลที่เกี่ยวข้อง เพื่อนำมาตอบคำถาม อธิบายปรากฏการณ์หรือประเด็นที่น่าสนใจ

ประโยชน์ของสถิติในด้านต่างๆ เช่น

- การศึกษา คือ รัฐบาลสามารถนำข้อมูลเกี่ยวกับประชากรก่อนวัยเรียนและวัยเรียน บุคลากรทางการศึกษา ปริมาณการผลิตและพัฒนาครูในแต่ละสาขาวิชา จำนวนสถาน ศึกษา และค่าใช้จ่ายในแต่ละระดับการศึกษา มาวิเคราะห์เพื่อกำหนดนโยบาย และวางแผนพัฒนาการศึกษาและการกระจายโอกาสทางการศึกษาให้กับเด็ก และเยาวชนในระดับการศึกษาขั้นพื้นฐาน
- การเกษตร รัฐบาลสามารถนำข้อมูลเกี่ยวกับครัวเรือนที่ทำการเกษตร พื้นที่การเพาะปลูก ผลผลิตทางการเกษตร ราคาสินค้าเกษตรกรรม เครื่องมือเครื่องใช้ทางการ เกษตร แหล่งน้ำและการชลประทาน และความต้องการสินค้าทางการเกษตร ของผู้บริโภค มาวิเคราะห์เพื่อให้คำแนะนำในการวางแผนการเพาะปลูกกับ เกษตรกรในฤดูกาลถัดไปได้

- การผลิต การวางแผนผลิตสินค้าออกจำหน่ายควรเก็บข้อมูลและวิเคราะห์ว่าสินค้านั้นเป็น ที่ต้องการของตลาดหรือไม่ ผู้ผลิตมีกำลังการผลิตเท่าใด ควรจำหน่ายสินค้าที่ใด และทำอย่างไรผลผลิตจะมีอย่างต่อเนื่องและได้สินค้าที่มีคุณภาพ

- การควบคุม คุณภาพสินค้า การควบคุมคุณภาพสินค้าที่ผลิตให้มีมาตรฐานตามที่กำหนด โดยเฉพาะอย่างยิ่ง ในยุคที่มีการแข่งขันสูง การรักษามาตรฐานของสินค้ายังมีความสำคัญมากยิ่งขึ้น อย่างไรก็ตาม การตรวจสอบคุณภาพของสินค้าที่ผลิตทุกชิ้นก่อนส่งออกจำหน่าย นับเป็นภาระที่มากเกินไปที่จะกระทำได้ในการควบคุมคุณภาพสินค้าที่ผลิตให้ เป็นไปตามมาตรฐานที่กำหนดโดยไม่ต้องตรวจสอบคุณภาพของสินค้าที่ผลิต ทุกชิ้นเป็นสิ่งที่สามารถกระทำได้ โดยกำหนดวิธีการตามขั้นตอนของกระบวนการ ทางสถิติศาสตร์ กล่าวคือ ในขั้นตอนการเก็บข้อมูล สถิติศาสตร์จะกำหนดวิธีการ เก็บรวบรวมข้อมูลที่เหมาะสมว่าควรเป็นการเลือกตัวอย่างโดยการสุ่มตัวอย่าง สินค้าจำนวนหนึ่งจากจำนวนสินค้าทั้งหมดที่ผลิตในแต่ละรุ่น โดยสามารถ กำหนดขนาดตัวอย่างที่จะเลือกและสามารถใช้วิธีการวิเคราะห์ซึ่งสร้างเกณฑ์ การพิจารณาว่า หากมีสินค้าที่ชำรุดในสินค้าตัวอย่างทั้งหมดที่เลือกมาจำนวน ไม่เกินเท่าใดแล้วจะถือว่าสินค้านั้นได้มาตรฐานตามที่กำหนด

- ผู้บริโภค การตัดสินใจเลือกซื้อสินค้าสามารถนำสถิติศาสตร์มาใช้ โดยเก็บข้อมูลเกี่ยวกับ สินค้าที่ต้องการ แล้วนำข้อมูลที่ได้ไปวิเคราะห์หรือเปรียบเทียบข้อมูลจากร้านค้า ต่าง ๆ ที่สำรวจเพื่อนำมาตัดสินใจว่าจะซื้อสินค้านั้นหรือไม่

- การพยากรณ์ การพยากรณ์เกี่ยวกับฝนโดยอาศัยข้อมูลจากกรมอุตุนิยมวิทยา ซึ่งใช้ข้อมูลทั้งใน อดีตและปัจจุบัน ตลอดจนวิธีวิเคราะห์ทางสถิติศาสตร์เข้ามาช่วย ทำให้สามารถ คาดเดาสถานการณ์ล่วงหน้าได้ว่าฝนจะตกหนักหรือไม่

- ชีวิตประจำวัน จดบันทึกรายรับรายจ่ายในแต่ละเดือนเพื่อใช้ในการวิเคราะห์พฤติกรรม การใช้จ่ายซึ่งจะช่วยให้สามารถปรับวิธีการใช้จ่ายเงินและควบคุมค่าใช้จ่ายได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

1.2 คำสำคัญในสถิติศาสตร์

- ประชากร (population) หมายถึง กลุ่มของหน่วยทั้งหมดในเรื่องที่สนใจศึกษา หน่วยในที่นี้อาจเป็น คน สัตว์ หรือสิ่งของ

- ตัวอย่าง (sample) หมายถึง กลุ่มย่อยของประชากรที่ถูกเลือกมาเป็นตัวแทนของประชากร โดยทั่วไป มีวัตถุประสงค์เพื่อใช้ตัวอย่างในการสรุปผลเกี่ยวกับลักษณะของประชากรที่สนใจ

ตัวอย่างเช่น ถ้ากำหนดให้ประชากรคือผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ตัวอย่างได้ดังนี้ พ.ศ. 2562 อาจกำหนด

1. ผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน
 2. ผู้ป่วยโรคเบาหวานในภาคกลางของประเทศไทยใน พ.ศ. 2562
- ตัวแปร (variable) หมายถึง ลักษณะบางประการของประชากรหรือตัวอย่างที่สนใจศึกษา
 - ข้อมูล (data) หมายถึง ข้อความจริงเกี่ยวกับเรื่องใดเรื่องหนึ่งที่สามารถใช้ในการสรุปผลในเรื่องที่สนใจ ศึกษา อาจเป็นได้ทั้งตัวเลขหรือไม่ใช่ตัวเลข หรืออาจหมายถึงค่าของตัวแปรที่สนใจศึกษา
 - พารามิเตอร์ (parameter) หมายถึง ค่าวัดที่แสดงลักษณะของประชากร ซึ่งเป็นค่าคงตัวที่คำนวณ หรือประมวล จากข้อมูลทั้งหมดของประชากร

ตัวอย่างเกี่ยวกับพารามิเตอร์

ถ้ากำหนดให้ประชากรคือผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 พารามิเตอร์ อาจเป็นได้ดังนี้

1. อายุเฉลี่ยของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562
 2. ฐานนิยมของอาชีพของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562
- ค่าสถิติ (statistic) เป็นค่าคงตัวที่พิจารณาจากข้อมูลของตัวอย่าง โดยมีวัตถุประสงค์เพื่ออธิบายลักษณะ ของ ตัวอย่างนั้นหรือเพื่อประมาณค่าของพารามิเตอร์แล้วนำไปใช้ในการอธิบายลักษณะของประชากร

ตัวอย่างเกี่ยวกับค่าสถิติ

ถ้ากำหนดให้ประชากรคือผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 และเลือก ตัวอย่างคือผู้ป่วย โรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน ค่าสถิติอาจเป็นได้ดังนี้

1. อายุเฉลี่ยของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน
2. ฐานนิยมของอาชีพของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน

1.3 ประเภทของข้อมูล

ข้อมูลที่จะนำมาใช้ศึกษาสามารถแบ่งได้หลายประเภทที่สำคัญมีดังนี้

1. การแบ่งประเภทของข้อมูลตามแหล่งที่มาของข้อมูล
2. การแบ่งประเภทของข้อมูลตามระยะเวลาที่จัดเก็บ

3. การแบ่งประเภทของข้อมูลตามลักษณะของข้อมูล

1.3.1 การแบ่งประเภทของข้อมูลตามแหล่งที่มาของข้อมูล การแบ่งประเภทของข้อมูลตามแหล่งที่มาของข้อมูล เป็นการแบ่งประเภทของข้อมูล โดยคำนึงว่า ผู้ใช้ข้อมูลเป็นผู้จัดเก็บข้อมูลเอง หรือเป็นข้อมูลที่บุคคลหรือหน่วยงานอื่น เป็นผู้จัดเก็บแล้วผู้ใช้เพียงแค่นำมาใช้ จึงแบ่งข้อมูลออกเป็น 2 ประเภท ดังนี้

- ข้อมูลปฐมภูมิ (primary data) คือข้อมูลที่ใช้ดำเนินการเก็บรวบรวมจากเจ้าของ ข้อมูลหรือต้นกำเนิดของข้อมูลโดยตรง

ตัวอย่างของข้อมูลปฐมภูมิ

1. ข้อมูลความพึงพอใจในสินค้าที่เจ้าของสินค้าเก็บรวบรวมจากผู้ซื้อสินค้าหรือผู้บริโภค เป็นข้อมูลปฐมภูมิของเจ้าของสินค้า
2. ข้อมูลการรักษาพยาบาลที่โรงพยาบาลบันทึกไว้ในประวัติผู้ป่วย โรงพยาบาล เป็นข้อมูลปฐมภูมิของ
3. ข้อมูลที่นักเรียนบันทึกจากการทดลองฟิสิกส์ในห้องปฏิบัติการของโรงเรียน เป็นข้อมูล ปฐมภูมิของนักเรียนที่ทำการทดลอง

- ข้อมูลทุติยภูมิ (secondary data) คือข้อมูลที่ใช้ไม่ได้ดำเนินการเก็บรวบรวมจากเจ้าของข้อมูลหรือต้นกำเนิดของข้อมูลโดยตรง แต่ใช้ข้อมูลที่บุคคลหรือหน่วยงานอื่น เก็บรวบรวมมา ซึ่งส่วนใหญ่ผู้ใช้มักจะใช้ข้อมูลที่เก็บรวบรวมโดยภาครัฐซึ่งเป็นการเก็บ รวบรวมข้อมูลตามภารกิจของหน่วยงาน

ตัวอย่างของข้อมูลทุติยภูมิ

1. สำนักงานตรวจคนเข้าเมืองจัดเก็บข้อมูลนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทย ในช่วงไตรมาสหนึ่ง ข้อมูลนี้เป็นข้อมูลปฐมภูมิของสำนักงานตรวจคนเข้าเมือง แต่เป็นข้อมูล ทุติยภูมิสำหรับผู้วิจัยหรือผู้ศึกษาอื่นที่นำข้อมูลนี้มาวิเคราะห์เพื่อหาคำตอบที่สนใจ
2. โรงเรียนจัดเก็บข้อมูลทะเบียนประวัติของนักเรียน ข้อมูลนี้เป็นข้อมูลปฐมภูมิของโรงเรียน เนื่องจากโรงเรียนเป็นผู้จัดเก็บจากนักเรียนซึ่งเป็นเจ้าของข้อมูล แต่เมื่อโรงเรียนส่งรายงาน ข้อมูลไปยังเขตพื้นที่การศึกษา ข้อมูลนี้จะเป็นข้อมูลทุติยภูมิของเขตพื้นที่การศึกษา

ข้อดีของข้อมูลทุติยภูมิ คือผู้ใช้ไม่ต้องเสียเวลาและค่าใช้จ่ายในการเก็บรวบรวมข้อมูลเอง สามารถนำข้อมูลที่มีผู้อื่นเก็บรวบรวมไว้แล้วมาใช้ได้เลย แต่อย่างไรก็ตามผู้ใช้จะต้องระมัดระวัง ในการนำข้อมูลประเภทนี้มาใช้ เนื่องจากมีโอกาสผิดพลาดได้ง่าย

1.3.2 การแบ่งประเภทของข้อมูลตามระยะเวลาที่จัดเก็บ

การแบ่งประเภทของข้อมูลตามระยะเวลาที่จัดเก็บเป็นการแบ่งประเภทของข้อมูลโดยพิจารณา จากช่วงเวลาที่มีข้อมูลเกิดขึ้นและมีการจัดเก็บ โดยแบ่งประเภทของข้อมูลออกเป็นข้อมูล อนุกรมเวลาและข้อมูลตัดขวาง

- ข้อมูลอนุกรมเวลา (time series data) คือชุดข้อมูลที่เกิดขึ้นและจัดเก็บตามลำดับเวลาต่อเนื่องกันไปตลอดช่วง ๆ หนึ่ง

ข้อมูลอนุกรมเวลาสามารถแสดงการเปลี่ยนแปลงหรือการเคลื่อนไหวของข้อมูลที่น่าสนใจ เมื่อเวลาเปลี่ยนไป ข้อมูลประเภทนี้จึงเป็นประโยชน์สำหรับการทำวิจัยระยะยาว เนื่องจาก ทำให้ผู้วิจัยเห็นแนวโน้มของเรื่องนั้นได้ ตัวอย่างเช่น ยอดขายสินค้ารายเดือนตั้งแต่เดือน มกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2561 คือยอดขายสินค้าที่เกิดขึ้นและมีการบันทึก ในแต่ละเดือนว่าเป็นเท่าใดตลอดช่วงเวลานั้น ซึ่งเมื่อนำข้อมูลรายเดือนมาเรียงต่อกันตาม ลำดับเวลา จะแสดงการเคลื่อนไหวขึ้นลงของยอดขายสินค้ารายเดือนในช่วงนั้นว่าเป็นอย่างไร เดือนใดมียอดขายสูงสุดและเดือนใดมียอดขายต่ำสุด ลักษณะการเปลี่ยนแปลงหรือเคลื่อนไหว เมื่อเวลาเปลี่ยนแปลงไปเป็นอย่างไร

- ข้อมูลตัดขวาง (cross-sectional data) คือข้อมูลที่บอกสถานะหรือสภาพของสิ่ง ที่สนใจ ณ จุดหนึ่งของเวลา ตัวอย่างของข้อมูลตัดขวาง

1. จำนวนประชากรของประเทศไทย ณ วันที่ 13 เมษายน พ.ศ. 2562 โดยอาจแสดงว่า มีประชากรรวมทั้งสิ้นกี่คน เป็นเพศชายและเพศหญิงกี่คน เป็นประชากรอายุน้อยกว่า 1 ปี กี่คน เป็นประชากรอายุ 1-5 ปี กี่คน เป็นประชากรสูงอายุคือตั้งแต่ 65 ปีขึ้นไป กี่คน อาศัยอยู่ในแต่ละจังหวัดจำนวนเท่าใด -
2. รายงานผลการศึกษานักเรียนเมื่อสิ้นภาคการศึกษา เนื่องจากการเป็นารแสดง สถานภาพทางการเรียนของนักเรียน ณ วันประกาศผลเมื่อสิ้นภาคการศึกษา

1.3.3 การแบ่งประเภทของข้อมูลตามลักษณะของข้อมูล

การแบ่งประเภทของข้อมูลตามลักษณะของข้อมูลเป็นการแบ่งประเภทของข้อมูลโดยพิจารณา ว่าข้อมูลนั้นแสดงถึงปริมาณของสิ่ง ๆ หนึ่งหรือไม่ โดยแบ่งออกได้เป็นข้อมูลเชิงปริมาณและ ข้อมูลเชิงคุณภาพ

- ข้อมูลเชิงปริมาณ (quantitative data) คือข้อมูลที่ได้จากการวัดหรือการนับค่า โดย แสดงเป็นตัวเลขหรือปริมาณที่สามารถนำไปบวก ลบ คูณ หรือหาร และเปรียบเทียบ กันได้

ตัวอย่างของข้อมูลเชิงปริมาณ เช่น คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่ง จำนวน นักเรียนที่ใช้บริการห้องสมุดของโรงเรียนในแต่ละวันในภาคการศึกษาที่แล้ว ยอดขายรถยนต์ รายเดือน

- ข้อมูลเชิงคุณภาพ (qualitative data) คือข้อมูลที่แสดงลักษณะ ประเภท สมบัติในเชิงคุณภาพ และอื่น ๆ ที่ไม่สามารถวัดค่าเป็นตัวเลขที่นำมาบวก ลบ คูณ หรือหารกันได้

ตัวอย่างของข้อมูลเชิงคุณภาพ เช่น อาชีพของผู้ปกครองนักเรียนห้องหนึ่ง หมายเลขโทรศัพท์ ของนักเรียนห้องหนึ่ง
ความคิดเห็นที่ประชาชนในท้องถิ่นมีต่อเรื่องหนึ่ง ๆ ความพึงพอใจใน การใช้บริการห้องสมุดของนักเรียนว่าพอใจ
มากน้อยเพียงไร (พอใจมากที่สุด พอใจมาก พอใจ ปานกลาง พอใจน้อย พอใจน้อยที่สุด)

1.4 สถิติศาสตร์เชิงพรรณนาและสถิติศาสตร์เชิงอนุมาน

การวิเคราะห์ข้อมูลแบ่งได้เป็นสถิติศาสตร์เชิงพรรณนาและสถิติศาสตร์เชิงอนุมาน โดยมีรายละเอียด
ดังต่อไปนี้

- สถิติศาสตร์เชิงพรรณนา (descriptive statistics) เป็นการวิเคราะห์ข้อมูลที่สรุปสาระสำคัญของ ข้อมูล
ชุดหนึ่ง ซึ่งเป็นข้อมูลเชิงคุณภาพหรือข้อมูลเชิงปริมาณ เพื่ออธิบายลักษณะหรือสภาพ ของข้อมูลชุดนั้นว่าเป็น
อย่างไร โดยทั่วไปข้อมูลเชิงคุณภาพจะใช้นำเสนอด้วยตารางความถี่ แผนภูมิแท่ง ฐานนิยม และอื่น ๆ
ส่วนข้อมูลเชิงปริมาณจะใช้นำเสนอด้วยฮิสโทแกรม แผนภาพกล่อง ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย และอื่น ๆ

ตัวอย่างของการใช้สถิติศาสตร์เชิงพรรณนา

1. คะแนนต่ำสุดในการสอบกลางภาควิชาคณิตศาสตร์ คือ 7 คะแนน
2. นักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งในกรุงเทพมหานครที่มีภูมิลำเนาอยู่ทาง ภาคเหนือคิด
เป็นร้อยละ 15 ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ทั้งหมดของโรงเรียนแห่งนี้

- สถิติศาสตร์เชิงอนุมาน (inferential statistics) เป็นการวิเคราะห์ข้อมูลที่ใช้ทฤษฎีที่เกี่ยวกับ ความ
น่าจะเป็นในการหาข้อสรุปเกี่ยวกับลักษณะของประชากรโดยใช้ข้อมูลจากตัวอย่างที่ได้มาจาก ประชากรนั้น

ตัวอย่างของการใช้สถิติศาสตร์เชิงอนุมาน

- ในการสำรวจพฤติกรรมการเดินทางท่องเที่ยวของคนไทยที่มีอายุ 15 ปีขึ้นไป ใน พ.ศ. 2559 ซึ่งจัดทำ
โดยสำนักงานสถิติแห่งชาติร่วมกับการท่องเที่ยวแห่งประเทศไทย (ททท.) ได้มีการเก็บรวบรวมข้อมูล ในเดือน
มกราคม - กุมภาพันธ์ พ.ศ. 2560 จากคนไทยที่มีอายุ 15 ปีขึ้นไปทั่วประเทศ ที่เลือก เป็นตัวอย่างจำนวน 63,060
คน และได้ข้อสรุปดังแสดงในรูปที่ 2 ซึ่งอาจสรุปได้ว่าวัตถุประสงค์หลัก อันดับที่ 1 ในการเดินทางท่องเที่ยวของคน
ไทยทั้งประเทศที่มีอายุ 15 ปีขึ้นไป ใน พ.ศ. 2559 คือการ เยี่ยมครอบครัว ญาติ/เพื่อน ซึ่งคิดเป็น 33.1%

บทที่ 2 การเก็บรวบรวมและการนำเสนอข้อมูล

1. การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยตารางความถี่

1.1 การวิเคราะห์ข้อมูลเชิงคุณภาพ

- ข้อมูลเชิงคุณภาพเป็นข้อมูลที่แสดงลักษณะ ประเภท สมบัติในเชิงคุณภาพ และอื่น ๆ ที่ไม่สามารถวัดค่าเป็นตัวเลขที่นำมาบวก ลบ คูณ หรือหารกันได้ เช่น ถ้าพิจารณาตัวแปร คือเพศของผู้ใช้บริการในร้านอาหารแห่งหนึ่ง ข้อมูลที่เป็นไปได้คือหญิงหรือชาย โดยทั่วไป การเก็บข้อมูลเชิงคุณภาพที่มีจำนวนมากอาจมีข้อมูลที่ซ้ำกันอยู่ ดังนั้นจึงต้องมีการจัดระเบียบ ข้อมูล เพื่อให้สังเกตลักษณะของข้อมูลได้ง่ายขึ้นและสามารถหาข้อสรุปที่มีความหมาย และนำไปใช้ประโยชน์ได้

ในทางสถิติศาสตร์จะวิเคราะห์ข้อมูลเชิงคุณภาพโดยพิจารณาจากความถี่และฐานนิยม ดังบท นิยามต่อไปนี้

- ความถี่ (frequency) คือ จำนวนครั้งของการเกิดข้อมูลข้อมูลหนึ่งและค่าของตัวแปร ค่าหนึ่ง
- ฐานนิยม (mode) คือ ข้อมูลที่มีจำนวนครั้งของการเกิดซ้ำกันมากที่สุดหรือข้อมูลที่มี ความถี่สูงสุดที่มากกว่า 1

1.2 การนำเสนอข้อมูลเชิงคุณภาพด้วยตารางความถี่

- การนำเสนอข้อมูลเชิงคุณภาพในรูปตารางความถี่ (frequency table) ที่จะกล่าวถึงในที่นี้ ได้แก่ ตารางความถี่จำแนกทางเดียว (one-way frequency table) และ ตารางความถี่ จำแนกสองทาง (two-way frequency table)

ตารางความถี่จำแนกทางเดียว หมายถึง

- ตารางความถี่จำแนกทางเดียวเป็นการนำเสนอข้อมูลในรูปตารางโดยแสดงข้อมูลและความถี่ ของข้อมูลของตัวแปรเพียงหนึ่งตัว มักใช้ในการนำเสนอข้อมูลเชิงคุณภาพเพื่อสรุปลักษณะ ที่สนใจหรือเปรียบเทียบความถี่ของแต่ละข้อมูล

ความถี่สัมพัทธ์ (relative frequency) คือ

- สัดส่วนของความถี่ของแต่ละข้อมูล เทียบ กับผลรวมของความถี่ทั้งหมด

ตารางความถี่จำแนกสองทาง หมายถึง

- ตารางความถี่จำแนกสองทางเป็นการนำเสนอข้อมูลในรูปตารางซึ่งมีตัวแปรที่สนใจศึกษา 2 ตัว โดยแสดงความถี่ของข้อมูลเชิงคุณภาพของแต่ละตัวแปรที่สนใจศึกษาในรูปตาราง

2. การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยแผนภาพ

- ในหัวข้อนี้ จะทบทวนเกี่ยวกับการวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยแผนภาพ ซึ่งอยู่ในรูป แผนภูมิรูปภาพ แผนภูมิรูปร่างกลม และแผนภูมิแท่ง ที่ได้ศึกษามาแล้วในระดับมัธยมศึกษาตอนต้น

แผนภูมิรูปภาพ

- แผนภูมิรูปภาพ (pictogram) เป็นการนำเสนอข้อมูลโดยใช้รูปภาพหรือสัญลักษณ์แสดงความถี่ของแต่ละข้อมูล ซึ่งจะต้องกำหนดในแผนภูมิว่ารูปภาพหรือสัญลักษณ์หนึ่งรูปนั้นแทนความถี่เท่าใด

แผนภูมิรูปร่างกลม

- แผนภูมิรูปร่างกลม (pie chart or circular chart) เป็นการนำเสนอข้อมูลโดยใช้พื้นที่ภายในของรูปร่างกลมแทนความถี่ของข้อมูลทั้งหมด และแสดงสัดส่วนของความถี่ของแต่ละข้อมูลด้วยพื้นที่แต่ละส่วน ภายในรูปร่างกลมซึ่งแบ่งด้วยรัศมี โดยสัดส่วนของความถี่ของแต่ละข้อมูลเท่ากับสัดส่วนของขนาดของ มุมที่จุดศูนย์กลางของรูปร่างกลม โดยทั่วไปจะแสดงสัดส่วนของความถี่ของข้อมูลด้วยความถี่สัมพัทธ์ ในรูปร้อยละ

แผนภูมิแท่ง

- แผนภูมิแท่ง (bar chart) เป็นการนำเสนอข้อมูลด้วยแท่งรูปสี่เหลี่ยมมุมฉากในแนวตั้งหรือแนวนอน โดยใช้ความสูงหรือความยาวของแท่งรูปสี่เหลี่ยมมุมฉากแต่ละรูปแสดงความถี่ของข้อมูลของแต่ละ ตัวแปรที่สนใจศึกษา

- แผนภูมิแท่งมีได้หลายแบบ เช่น แผนภูมิแท่งเชิงเดียว แผนภูมิแท่งพหุคูณ แผนภูมิแท่งส่วนประกอบ

1. แผนภูมิแท่งเชิงเดียว

- แผนภูมิแท่งเชิงเดียว (simple bar chart) เป็นแผนภูมิแท่งซึ่งมีตัวแปรที่สนใจศึกษาเพียงหนึ่งตัว โดยแสดงข้อมูลและความถี่ของข้อมูล เพื่อแสดงการเปรียบเทียบความถี่ของแต่ละข้อมูล เช่น จากตัวอย่างที่ 6 ถ้ามีผู้ตอบแบบสำรวจทั้งหมด 1,000 คน จะสามารถแสดงจำนวนของผู้ใช้สื่อสังคม ออนไลน์หลักแต่ละอย่าง (ความถี่) ได้ดังนี้

2. แผนภูมิแท่งพหุคูณ

- แผนภูมิแท่งพหุคูณ (multiple bar chart) เป็นแผนภูมิแท่งซึ่งมีตัวแปรที่สนใจศึกษาตั้งแต่ 2 ตัว ขึ้นไป (ในที่นี้จะศึกษาเพียง 2 ตัวเท่านั้น) โดยแสดงข้อมูลของแต่ละตัวแปรที่สนใจศึกษาบนแกน เดียวกัน และแสดงความถี่รวมของข้อมูลของตัวแปรที่สนใจศึกษาด้วยแท่งรูปสี่เหลี่ยมมุมฉาก

3. แผนภูมิแท่งส่วนประกอบ

- แผนภูมิแท่งส่วนประกอบ (component bar chart) เป็นแผนภูมิแท่งที่แสดงจำนวนรวมและส่วนประกอบของจำนวนรวมนั้น โดยการแบ่งเป็นส่วนย่อย ๆ

บทที่ 3 การวัดค่ากลาง

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ

- ข้อมูลเชิงปริมาณเป็นข้อมูลที่พบได้มากในการดำเนิน ธุรกิจ การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณสามารถทำได้โดยใช้ค่าวัดทางสถิติ หรืออาจนำเสนอด้วย ตารางความถี่ ฮิสโทแกรม แผนภาพจุด แผนภาพลำต้น และใบ แผนภาพกล่อง หรือแผนภาพการกระจาย เพื่อนำไปใช้ในการแปลความหมายและตัดสินใจ ต่อไป ตัวอย่างในธุรกิจร้านขายเครื่องดื่มและของหวาน ผู้ประกอบการต้องพิจารณาข้อมูลเกี่ยวกับปริมาณ วัตถุดิบที่ใช้ในการผลิต ราคาและเวลาที่ใช้ในการสั่งซื้อวัตถุดิบแต่ละชนิด โดยอาจใช้ค่ากลางของข้อมูล ที่เหมาะสมในการคาดการณ์ ปริมาณวัตถุดิบแต่ละ ชนิดที่ต้องใช้ในแต่ละวัน สัปดาห์ หรือเดือน เพื่อที่จะดำเนินการสั่งซื้อวัตถุดิบต่าง ๆ ได้อย่างเหมาะสม ไม่เกิดเหตุการณ์ที่วัตถุดิบหมดหรือสั่งซื้อด้วยปริมาณที่มากเกินไป นอกจากนี้ ผู้ประกอบการอาจใช้ข้อมูลเกี่ยวกับปริมาณลูกค้าที่เข้ามาใช้บริการหรือซื้อสินค้า และระยะเวลาที่ลูกค้าเข้ามาใช้บริการในร้าน ในการวิเคราะห์พฤติกรรม การบริโภคของลูกค้าที่อาจเปลี่ยนแปลงไปตามแต่ละช่วงเวลาในแต่ละวัน หรือแต่ละสัปดาห์ เช่น ช่วงเวลาเร่งด่วนในตอนเช้า ช่วงพักกลางวัน และช่วงวันหยุดสุดสัปดาห์ อาจมีลูกค้า เข้ามาใช้บริการจำนวนมาก ทำให้ลูกค้าต้องใช้เวลามากขึ้นในการรอรับบริการ ผู้ประกอบการ สามารถใช้ข้อมูลเหล่านี้ประกอบการวางแผนการจัดกำลังคนในแต่ละช่วงเวลา แต่ละวัน และตัดสินใจจ้างพนักงานเพิ่มได้อย่างเหมาะสม

1. การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

- ในบทที่ผ่านมา นักเรียนได้ศึกษาเกี่ยวกับการแจกแจงความถี่ของข้อมูลเชิงคุณภาพมาแล้ว สำหรับ ในบทนี้นักเรียนจะได้ศึกษาการแจกแจงความถี่ของข้อมูลเชิงปริมาณซึ่งเป็นข้อมูลที่ได้จากการวัดหรือการนับค่า โดยแสดงเป็นตัวเลขหรือปริมาณที่สามารถนำไปบวก ลบ คูณ หรือหาร และเปรียบเทียบกัน ได้อย่างมีความหมาย ข้อมูลเชิงปริมาณสามารถใช้การแจกแจงความถี่เพื่อจัดระเบียบและนำเสนอข้อมูล ได้เช่นเดียวกับข้อมูลเชิงคุณภาพ โดยการเขียนตารางความถี่สำหรับข้อมูลเชิงปริมาณมี 2 แบบ ได้แก่

1.1 ตารางความถี่แบบไม่ได้แบ่งข้อมูลเป็นช่วง ซึ่งเหมาะสำหรับใช้ในกรณีที่ค่าที่เป็นไปได้ของข้อมูล มีจำนวนน้อย

2.2 ตารางความถี่แบบแบ่งข้อมูลเป็นช่วง ซึ่งเหมาะสำหรับใช้ในกรณีที่ค่าที่เป็นไปได้ของข้อมูลมีจำนวนมาก

- ในกรณีที่ค่าของคะแนนที่เป็นไปได้มีจำนวนมาก เช่น ในการสอบวิชาคณิตศาสตร์ซึ่งมีคะแนนเต็ม 100 คะแนน โดยครูให้คะแนนเป็นจำนวนเต็ม ถ้าเขียนตารางความถี่โดยใช้ทุกค่าของคะแนนที่เป็นไปได้ จะมีมากถึง 101 ค่า ซึ่งยากต่อการนำเสนอ ด้วยเหตุนี้จึงแบ่งข้อมูลที่เป็นไปได้ทั้งหมดออกเป็นช่วง ๆ และ เรียกแต่ละช่วงว่า อंतरภาคชั้น (class interval)

ขั้นตอนการเขียนตารางความถี่ของข้อมูลเชิงปริมาณที่มีข้อมูลทั้งหมดเป็นจำนวนเต็ม

1. กำหนดจำนวนอันตรภาคชั้นเป็น k ชั้น

2. กำหนดค่าเริ่มต้นและค่าสุดท้ายที่ครอบคลุมทุกค่าของข้อมูล โดยที่ค่าเริ่มต้นคือค่าต่ำสุดหรือค่า ที่น้อยกว่าค่าต่ำสุดของข้อมูล และค่าสุดท้ายคือค่าสูงสุดหรือค่าที่มากกว่าค่าสูงสุดของข้อมูล

3. คำนวณความกว้างของอันตรภาคชั้น โดยหาได้จาก

ค่าสุดท้าย - ค่าเริ่มต้น ส่วนด้วยจำนวนอันตรภาคชั้น

ถ้าค่าที่คำนวณได้ไม่เป็นจำนวนเต็มให้ปัดเศษขึ้นเป็นจำนวนเต็มเสมอ

4. กำหนดอันตรภาคชั้นโดยที่

- ชั้นแรกมีค่าเริ่มต้นที่กำหนดในข้อ 2 ถึงจำนวนที่ได้จากการนำค่าเริ่มต้นที่กำหนดในข้อ 2 บวกกับ ความกว้างของอันตรภาคชั้นลบด้วย 1

- ชั้นที่สองมีค่าเริ่มต้นเป็นค่าสุดท้ายของชั้นแรกบวกด้วย 1 ถึงค่าเริ่มต้นของชั้นที่สองบวกกับ ความกว้างของอันตรภาคชั้นลบด้วย 1

- ทำเช่นนี้ไปเรื่อย ๆ จนถึงชั้นที่ k (ในกรณีที่ ค่าสุดท้าย - ค่าเริ่มต้นส่วนด้วยจำนวนอันตรภาคชั้น

เป็นจำนวนเต็ม ค่าสุดท้าย ของชั้นที่ k จะไม่เท่ากับค่าสุดท้ายที่กำหนดในข้อ 2 แต่ต้องมากกว่าหรือเท่ากับข้อมูลทุกค่า)

5. หาจำนวนข้อมูลทั้งหมดที่อยู่ในแต่ละอันตรภาคชั้น โดยทำรอยขีดแทนจำนวนไว้ในแต่ละ อันตรภาคชั้น โดยปกติมักใช้รอยขีด | แทนหนึ่งค่า และเพื่อความสะดวกในการนับจำนวนข้อมูล ที่อยู่ในแต่ละอันตรภาคชั้น

6. นับจำนวนข้อมูลจากรอยขีดที่ทำในข้อ 5 แล้วบันทึกจำนวนข้อมูลลงในช่องความถี่ของแต่ละอันตรภาคชั้น

2. การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยแผนภาพ

- ในการวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ นอกจากจะใช้ตารางความถี่แล้ว ยังสามารถใช้แผนภาพได้อีกด้วย โดยการใช้แผนภาพแสดงการแจกแจงความถี่ของข้อมูล จะทำให้เห็นการแจกแจงของข้อมูล ได้ชัดเจนมากกว่าการพิจารณาจากตารางความถี่

ฮิสโทแกรม

- ฮิสโทแกรม (histogram) เป็นการนำเสนอข้อมูลเชิงปริมาณที่สร้างจากตารางความถี่ โดยใช้แท่งสี่เหลี่ยมมุมฉากที่เรียงติดกันบนแกนนอน เมื่อแกนนอนแทนค่าของข้อมูล ความสูงของแท่งสี่เหลี่ยม มุมฉากจะแสดงความถี่ของข้อมูล ซึ่งการแสดงความถี่ของข้อมูลอาจนำเสนอความถี่ของข้อมูลเพียง ค่าเดียวหรือข้อมูลในแต่ละอันตรภาคชั้น โดยความกว้างของแท่งสี่เหลี่ยมมุมฉากแต่ละแท่งจะสอดคล้อง กับความกว้างของแต่ละอันตรภาคชั้นของตารางความถี่

ขอบล่างของชั้น (lower class boundary) คือ

- ค่ากึ่งกลางระหว่างค่าของข้อมูลที่มากที่สุด ในชั้นก่อนหน้ากับค่าของข้อมูลที่น้อยที่สุดในชั้นนั้น

ขอบบนของชั้น (upper class boundary) คือ

- ค่ากึ่งกลางระหว่างค่าของข้อมูลที่มากที่สุด ในชั้นนั้นกับค่าของข้อมูลที่น้อยที่สุดในชั้นถัดไป

แผนภาพจุด

- แผนภาพจุด (dot plot) เป็นการนำเสนอข้อมูลเชิงปริมาณโดยใช้จุดหรือวงกลมเล็ก ๆ แทนข้อมูล แต่ละตัว เขียนเรียงไว้เหนือเส้นในแนวนอนที่มีสเกล จุดหรือวงกลมเล็ก ๆ ดังกล่าวจะเรียงกันใน แนวตั้งตรงกับตำแหน่งซึ่งแสดงค่าของข้อมูลแต่ละตัว

แผนภาพลำต้นและใบ

- แผนภาพลำต้นและใบ (stem and leaf plot) เป็นการนำเสนอข้อมูลเชิงปริมาณโดยใช้แผนภาพที่มีการแสดงข้อมูลโดยเรียงลำดับจากน้อยไปมากและแบ่งการแสดงข้อมูลออกเป็นสองส่วนที่เรียกว่า ส่วนลำต้น และ ส่วนใบ ในที่นี้กำหนดส่วนใบเป็นเลขโดดในหลักหน่วย และตัวเลขที่เหลือเป็นส่วนลำต้น เช่น 298 จะมี 29 เป็นส่วนลำต้น และ 8 เป็นส่วนใบ

แผนภาพกล่อง

- แผนภาพกล่อง (box plot) เป็นการนำเสนอข้อมูลเชิงปริมาณที่แสดงตำแหน่งสำคัญของข้อมูล ซึ่งประกอบด้วยค่าต่ำสุด ค่าสูงสุด และควอร์ไทล์ (quartile) นอกจากนี้ แผนภาพกล่องสามารถ ใช้ในการตรวจสอบว่ามีข้อมูลที่แตกต่างไปจากข้อมูลส่วนใหญ่หรือไม่ โดยจะเรียกข้อมูลดังกล่าวว่า ค่านอกเกณฑ์ (outlier)

ขั้นตอนการเขียนแผนภาพกล่อง

1. เรียงข้อมูลจากน้อยไปมาก จากนั้นหาค่าต่ำสุดและค่าสูงสุดของข้อมูล
2. หาควอร์ไทล์ที่ 1 (Q1) ควอร์ไทล์ที่ 2 (Q2) และควอร์ไทล์ที่ 3 (Q3) โดยที่
 - ควอร์ไทล์ที่ 1 (Q1) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณหนึ่งในสี่ของจำนวน ข้อมูลทั้งหมด
 - ควอร์ไทล์ที่ 2 (Q2) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณครึ่งหนึ่งของจำนวน ข้อมูลทั้งหมด หรือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งหมด
 - ควอร์ไทล์ที่ 3 (Q3) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณสามในสี่ของจำนวน ข้อมูลทั้งหมด

ถ้า n เป็นจำนวนข้อมูลทั้งหมด สามารถหาตำแหน่งของควอร์ไทล์ได้ดังนี้

Q1 อยู่ในตำแหน่งที่ $n+1$ ส่วน 4

Q2 อยู่ในตำแหน่งที่ $2(n+1)$ ส่วน 4

และ Q3 อยู่ในตำแหน่งที่ $3(n+1)$ ส่วน 4

3. หาค่า $Q1 - 1.5(Q3 - Q1)$ และ $Q3 + 1.5(Q3 - Q1)$
4. พิจารณาว่าชุดข้อมูลมีค่านอกเกณฑ์หรือไม่ โดยในที่นี้ค่านอกเกณฑ์คือข้อมูลที่มีค่าน้อยกว่า $Q1 - 1.5(Q3 - Q1)$ หรือข้อมูลที่มีค่ามากกว่า $Q3 + 1.5(Q3 - Q1)$

แผนภาพกล่องกับการกระจายของข้อมูล

- นอกจากแผนภาพกล่องจะสามารถใช้ในการตรวจสอบว่าชุดข้อมูลมีค่านอกเกณฑ์หรือไม่ ยังสามารถใช้ในการอธิบายลักษณะการกระจายของข้อมูลได้อีกด้วย การกระจายของข้อมูลจะทำให้เห็นว่า โดยภาพรวมแล้วข้อมูลมีการเกาะกลุ่มกันหรือไม่ ถ้าข้อมูลมีการกระจายมาก แสดงว่าข้อมูลมีค่า แตกต่างกันมากหรือข้อมูลไม่เกาะกลุ่มกัน แต่ถ้าข้อมูลมีการกระจายน้อย แสดงว่าข้อมูลมีค่าใกล้เคียง กันมากหรือข้อมูลเกาะกลุ่มกัน

แผนภาพการกระจาย

- การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณข้างต้นพิจารณาเฉพาะข้อมูลเชิงปริมาณของตัวแปรเพียงหนึ่งตัว แต่ในหัวข้อนี้จะพิจารณาข้อมูลเชิงปริมาณของตัวแปรสองตัวว่าตัวแปรทั้งสองมีความสัมพันธ์ เชิงเส้นกันหรือไม่ เนื่องจากเมื่อมีตัวแปรสองตัว ไม่จำเป็นที่ตัวหนึ่งต้องเป็นตัวแปรต้นและอีกตัว ต้องเป็นตัวแปรตาม การตรวจสอบความสัมพันธ์เชิงเส้นของตัวแปรทั้งสองจะพิจารณาโดยใช้ แผนภาพการกระจาย

- แผนภาพการกระจาย (scatter plot) คือแผนภาพที่เกิดจากการลงจุดที่แสดงค่าของตัวแปรคู่หนึ่ง รูปแบบการกระจายของจุดต่าง ๆ ที่ปรากฏในแผนภาพจะแสดงถึงรูปแบบความสัมพันธ์ระหว่าง สองตัวแปรนั้น

3. ค่าวัดทางสถิติ

- ในการวิเคราะห์ข้อมูลโดยใช้วิธีการของสถิติศาสตร์เชิงพรรณนา นอกจากจะทำได้โดยการเขียนตารางความถี่หรือแผนภาพต่าง ๆ ตามที่ได้กล่าวมาแล้ว ยังสามารถใช้ค่าวัดทางสถิติซึ่งเป็นค่าที่ได้จากการ นำข้อมูลทั้งหมดหรือข้อมูลบางส่วนมาคำนวณเพื่อใช้ในการวิเคราะห์ข้อมูล โดยค่าวัดทางสถิติจะช่วยให้เห็นภาพรวมของข้อมูลและช่วยให้สามารถจดจำข้อสรุปเกี่ยวกับข้อมูลนั้น ๆ ได้ง่ายขึ้น ซึ่งจะเป็น ประโยชน์ในการนำไปประกอบการตัดสินใจหรือการวางแผนต่าง ๆ

- ค่าวัดทางสถิติประกอบด้วยค่าสถิติและพารามิเตอร์ ซึ่งเป็นค่าที่วัดลักษณะโดยประมาณหรือคำนวณจากข้อมูลเหมือนกัน แต่แตกต่างกันตรงที่ค่าสถิติได้จากการพิจารณาข้อมูลของตัวอย่าง ในขณะที่ พารามิเตอร์ได้จากการพิจารณาข้อมูลทั้งหมดของประชากร โดยค่าสถิติและพารามิเตอร์ที่จะศึกษา ในหัวข้อนี้ ได้แก่ ค่ากลางของข้อมูล ค่าวัดการกระจาย และค่าวัดตำแหน่งที่ของข้อมูล

3.1 ค่ากลางของข้อมูล

- ค่ากลางของข้อมูลมีหลายชนิด เช่น ค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม ซึ่งนักเรียนได้ศึกษา มาแล้วในระดับมัธยมศึกษาตอนต้น ค่ากลางแต่ละชนิดต่างก็มีข้อดี ข้อเสีย และมีความเหมาะสม ในการนำไปใช้ไม่เหมือนกัน ขึ้นอยู่กับลักษณะการแจกแจงของข้อมูลและวัตถุประสงค์ของ ผู้ใช้ข้อมูลนั้น ๆ

- ในทางสถิติจะใช้ค่ากลางของข้อมูลเป็นตัวแทนของข้อมูลทั้งหมด เพื่อให้เข้าใจภาพรวม และสะดวกในการจดจำข้อสรุปเกี่ยวกับข้อมูลนั้น ๆ เช่น ผู้อำนวยการโรงเรียนแห่งหนึ่งต้องการ ทราบผลการเรียนวิชาคณิตศาสตร์ของนักเรียนในระดับชั้นต่าง ๆ ในปีที่ผ่านมา ผู้อำนวยการ อาจไม่จำเป็นต้องทราบผลการเรียนวิชา

คณิตศาสตร์ของนักเรียนแต่ละคนในแต่ละระดับชั้น แต่อาจพิจารณาเบื้องต้นจากค่าเฉลี่ยเลขคณิตของคะแนนวิชาคณิตศาสตร์ของนักเรียนแต่ละ ระดับชั้น

ค่าเฉลี่ยเลขคณิต

- สูตร ค่าเฉลี่ยเลขคณิต = ผลรวมของข้อมูล ÷ จำนวนข้อมูล
- ค่าเฉลี่ยเลขคณิต (arithmetic mean) เป็นค่าที่ได้จากการหารผลรวมของข้อมูลทั้งหมด ด้วยจำนวนข้อมูลที่มี
- ค่าเฉลี่ยเลขคณิตของประชากร (population mean) เขียนแทนด้วย μ (อ่านว่า มิว)
- ค่าตัวอย่าง (sample mean) เขียนแทนด้วย \bar{x} (อ่านว่า เอ็กซ์บาร์)

เพื่อความสะดวกจะใช้ตัวอักษรกรีกตัวพิมพ์ใหญ่ Σ (อ่านว่า ซิกมา) เป็นสัญลักษณ์แสดง การบวก

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก

- ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก (weighted arithmetic mean) เหมาะสำหรับใช้ในกรณีที่ ข้อมูลแต่ละค่า มีความสำคัญไม่เท่ากัน เช่น การหาค่าเฉลี่ยเลขคณิตของคะแนนสอบ 4 วิชา ซึ่งแต่ละวิชามีหน่วยกิตไม่เท่ากัน ถ้าใช้วิธีการหาค่าเฉลี่ยเลขคณิตโดยไม่ถ่วงน้ำหนัก อาจทำให้ ค่าที่ได้นำไปสู่ข้อสรุปที่คลาดเคลื่อนจากความเป็นจริง เพราะข้อมูลแต่ละค่ามีความสำคัญ ไม่เท่ากัน ขึ้นอยู่กับน้ำหนักของแต่ละข้อมูล

มัธยฐาน

เมื่อนำข้อมูลทั้งหมดมาเรียงลำดับจากน้อยไปมากหรือจากมากไปน้อย จะเรียกค่าที่อยู่ ในตำแหน่งกึ่งกลางของข้อมูลว่า **มัธยฐาน (median)**

ถ้าข้อมูลมี n ตัว การหามัธยฐานทำได้โดยเรียงลำดับข้อมูลจากน้อยไปมากหรือจากมาก ไปน้อย จะได้มัธยฐานอยู่ในตำแหน่งที่ $n+1$ ส่วนด้วย 2 นั่นคือ

- ถ้า n เป็นจำนวนที่ มัธยฐานคือข้อมูลที่อยู่กึ่งกลาง
- ถ้า n เป็นจำนวนคู่ มัธยฐานคือค่าเฉลี่ยเลขคณิตของข้อมูลสองตัวที่อยู่กึ่งกลาง

ฐานนิยม

- ฐานนิยม คือ ข้อมูลที่มีจำนวน ครั้งของการเกิดซ้ำกันมากที่สุดหรือข้อมูลที่มีความถี่สูงสุดที่มากกว่า 1 ซึ่งบทนิยามของฐานนิยม ดังกล่าวสามารถใช้กับข้อมูลเชิงปริมาณได้เช่นกัน

- ข้อมูลบางชุดอาจไม่มีฐานนิยม เช่น ในกรณีที่ข้อมูลมีความถี่เป็น 1 เท่ากันหมด นอกจากนี้ ข้อมูลบางชุดอาจมีฐานนิยมมากกว่า 1 ค่า เช่น ในกรณีที่ข้อมูลมากกว่า 1 ข้อมูล ที่มีความถี่ สูงสุดเท่ากัน

- อย่างไรก็ตาม ในที่นี้จะพิจารณาเฉพาะชุดข้อมูลที่มีฐานนิยมเพียงค่าเดียว

ข้อสังเกตที่สำคัญเกี่ยวกับค่ากลางชนิดต่างๆ

1. ฐานนิยมจะมีค่าตรงกับค่าใดค่าหนึ่งของข้อมูลชุดนั้น ในขณะที่ค่าเฉลี่ยเลขคณิตและมัธยฐาน อาจไม่ใช่ค่าใดค่าหนึ่งของข้อมูลชุดนั้น
2. โดยปกติค่าเฉลี่ยเลขคณิตมักเป็นค่ากลางที่นิยมมากที่สุด แต่ถ้าชุดข้อมูลมีข้อมูลที่แตกต่าง จากข้อมูลตัวอื่นมาก จะมีผลต่อค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ แต่จะไม่มีผลต่อมัธยฐาน และฐานนิยม
3. สำหรับข้อมูลเชิงคุณภาพ จะสามารถหาค่ากลางได้เฉพาะฐานนิยมเท่านั้น ไม่สามารถหา ค่าเฉลี่ยเลขคณิต และมัธยฐานได้
4. ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมไม่จำเป็นต้องมีค่าเท่ากัน ทั้งนี้ ค่าเฉลี่ยเลขคณิต และมัธยฐานจะมีค่าที่ไม่สูงหรือต่ำเกินไปเมื่อเทียบกับค่าของข้อมูลทั้งหมด ในขณะที่ ฐานนิยมอาจเป็นค่าสูงสุดหรือค่าต่ำสุดของชุดข้อมูลนั้นได้

การพิจารณาเลือกใช้ค่ากลางของข้อมูลไม่มีกฎเกณฑ์ที่แน่ชัด แต่ควรเลือกให้เหมาะสมกับ วัตถุประสงค์ในการนำไปใช้และลักษณะของข้อมูล รวมทั้งต้องพิจารณาข้อดีและข้อเสียของ ค่ากลางแต่ละชนิด หากเลือกใช้ค่ากลางที่ไม่เหมาะสม อาจทำให้สรุปผลหรือตัดสินใจผิดพลาดได้

ตัวอย่างค่าเฉลี่ยเลขคณิตระดับง่าย

ตัวอย่างค่าเฉลี่ยระดับง่ายที่ 1. คะแนนสอบ: 50, 60, 70, 80 จงหาค่าเฉลี่ยของคะแนนสอบนี้

วิธีทำ

$$\text{ผลรวม} = 50 + 60 + 70 + 80 = 260$$

$$\text{จำนวนข้อมูล} = 4$$

$$\text{ค่าเฉลี่ย} = 260 \div 4 = 65$$

ตอบ ค่าเฉลี่ยเลขคณิต = 65 คะแนน

ตัวอย่างค่าเฉลี่ยระดับง่ายที่ 2. น้ำหนักผลไม้: 1, 2, 3, 4, 5 กิโลกรัม จงหาค่าเฉลี่ยของน้ำหนักผลไม้

วิธีทำ

$$\text{ผลรวม} = 1 + 2 + 3 + 4 + 5 = 15$$

$$\text{จำนวนข้อมูล} = 5$$

$$\text{ค่าเฉลี่ย} = 15 \div 5 = 3$$

ตอบ ค่าเฉลี่ยเลขคณิต = 3 กิโลกรัม

ตัวอย่างค่าเฉลี่ยเลขคณิตระดับปานกลาง

ตัวอย่างค่าเฉลี่ยระดับปานกลางที่ 1. คะแนนเก็บ: 40, 50, 55, 60, 65, 70, 80 จงหาค่าเฉลี่ยของคะแนนเก็บทั้งหมด

วิธีทำ

$$\text{ผลรวม} = 420$$

$$\text{จำนวนข้อมูล} = 7$$

$$\text{ค่าเฉลี่ย} = 420 \div 7 = 60$$

ตอบ ค่าเฉลี่ยเลขคณิต = 60 คะแนน

ตัวอย่างค่าเฉลี่ยระดับปานกลางที่ 2. ค่าใช้จ่ายรายวันของนายเอ: 50, 50, 60, 60, 70, 200 จงหาค่าเฉลี่ยค่าใช้จ่ายทั้งหมด

วิธีทำ

$$\text{ผลรวม} = 490$$

$$\text{จำนวนข้อมูล} = 6$$

$$\text{ค่าเฉลี่ย} = 490 \div 6 = 81.67$$

ตอบ ค่าเฉลี่ยเลขคณิต = 81.67 บาท

ตัวอย่างค่าเฉลี่ยเลขคณิตระดับยาก

ตัวอย่างค่าเฉลี่ยระดับยากที่ 1. ทีมฟุตบอลชายของโรงเรียนแห่งหนึ่งมีสมาชิกทั้งหมด 24 คน โดยความสูง (เซนติเมตร) ของ สมาชิกแต่ละคน แสดงได้ดังนี้

165 178 170 168 167 167 180 175 181 164 179 158

177 163 165 172 180 191 185 176 175 183 177 179

จงหาค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

วิธีทำ

รวมข้อมูล: นำตัวเลขความสูงทั้ง 24 ตัวมาบวกกันทั้งหมด ในโจทย์นี้รวมได้เท่ากับ 4,175

หารด้วยจำนวนคน: นำผลรวมที่ได้มาหารด้วย 24

คือ $4,175 \div 24 = 173.96$ เซนติเมตร

ดังนั้น ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้มีค่าประมาณ 173.96 เซนติเมตร

ตัวอย่างค่าเฉลี่ยระดับยากที่ 2. โรงเรียนแห่งหนึ่งกำหนดว่านักเรียนจะได้เกรด 4 วิชาคณิตศาสตร์ ก็ต่อเมื่อนักเรียนได้คะแนน เฉลี่ยจากการสอบย่อย 6 ครั้ง ไม่ต่ำกว่า 80 คะแนน ถ้าค่าเฉลี่ยเลขคณิตจากการสอบย่อย 5 ครั้ง ของนักเรียนคนหนึ่งเท่ากับ 77 คะแนน จงหาว่าในการสอบย่อยครั้งที่ 6 นักเรียนคนนี้ จะต้องได้คะแนนอย่างน้อยเท่าใด จึงจะได้เกรด 4

วิธีทำ

1. หาคะแนนรวมเดิม 5 ครั้งแรก

โจทย์บอกว่าสอบไปแล้ว 5 ครั้ง ได้คะแนนเฉลี่ย 77 คะแนน

วิธีคิด: เอาคะแนนเฉลี่ยคูณจำนวนครั้ง $77 \times 5 = 385$ คะแนน

สรุป: ตอนนี้เรามีคะแนนเก็บในมือแล้ว 385 คะแนน

2. หาคะแนนรวมเป้าหมาย 6 ครั้ง

โจทย์บอกว่าถ้าอยากได้เกรด 4 ต้องได้คะแนนเฉลี่ย 6 ครั้ง ไม่ต่ำกว่า 80 คะแนน

วิธีคิด: เอาคะแนนเฉลี่ยที่ต้องการคูณจำนวนครั้งทั้งหมด $80 \times 6 = 480$ คะแนน

สรุป: สอบครบ 6 ครั้ง เราต้องทำคะแนนรวมให้ได้อย่างน้อย 480 คะแนน

3. หาคะแนนที่ขาดไป ครั้งที่ 6

นำคะแนนเป้าหมายมาลบคะแนนที่มีอยู่เดิม เพื่อดูว่าต้องทำเพิ่มอีกเท่าไร

วิธีคิด: $480 - 385 = 95$ คะแนน

ดังนั้น ในการสอบย่อยครั้งที่ 6 นักเรียนคนนี้จะต้องได้คะแนนอย่างน้อย 95 คะแนน จึงจะได้เกรด 4 วิชาคณิตศาสตร์

ตัวอย่างมัธยฐานระดับง่าย

ตัวอย่างมัธยฐานระดับง่ายที่ 1. ข้อมูลชุดหนึ่งคือ 3, 5 และ 7 จงหามัธยฐานของข้อมูลชุดนี้ได้เท่าใด

วิธีทำ

ข้อมูลเรียงแล้วเป็น 3, 5, 7

มีข้อมูล 3 ค่า (จำนวนคี่)

ค่ากลางคือ 5

ตอบ มัธยฐานเท่ากับ 5

ตัวอย่างมัธยฐานระดับง่ายที่ 2. ข้อมูลชุดหนึ่งคือ 10, 20, 30, 40 และ 50 จงหามัธยฐานของข้อมูลชุดนี้ได้เท่าใด

วิธีทำ

ข้อมูลมี 5 ค่า

ค่ากลางคือ 30

ตอบ มัธยฐานเท่ากับ 30

ตัวอย่างมัธยฐานระดับปานกลาง

ตัวอย่างมัธยฐานระดับปานกลางที่ 1. ข้อมูลชุดหนึ่งคือ 15, 20, 25 และ 30 จงหามัธยฐานของข้อมูลชุดนี้ได้เท่าใด

วิธีทำ

ข้อมูลมี 4 ค่า (จำนวนคู่)

$$\text{มัธยฐาน} = (20 + 25) \div 2 = 22.5$$

ตอบ มัธยฐานเท่ากับ 22.5

ตัวอย่างมัธยฐานระดับปานกลางที่ 2. คะแนนสอบของนักเรียนคือ 55, 60, 65, 70, 75 และ 80 จงหามัธยฐานของคะแนนสอบได้เท่าใด

วิธีทำ

ข้อมูลมี 6 ค่า (จำนวนคู่)

$$\text{มัธยฐาน} = (65 + 70) \div 2 = 67.5$$

ตอบ มัธยฐานเท่ากับ 67.5 คะแนน

ตัวอย่างมัธยฐานระดับยาก

ตัวอย่างมัธยฐานระดับยากที่ 1. ความสูง (เซนติเมตร) ของนักเรียนหญิงจำนวน 11 คน แสดงได้ดังนี้

164 158 167 160 163 159 162 161 155 170 168

จงหามัธยฐานของข้อมูลชุดนี้

วิธีทำ

เรียงความสูงของนักเรียนหญิง 11 คน จากน้อยไปมาก ได้ดังนี้

155 158 159 160 161 162 163 164 167 168 170

เนื่องจากมัธยฐานอยู่ในตำแหน่งที่ $11 \div 2 = 6$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ 162 เซนติเมตร

ตัวอย่างมัธยฐานระดับยากที่ 2. ระยะเวลา (นาทีก) ที่ใช้ในการเดินทางจากบ้านไปโรงเรียนของนักเรียน 6 คน แสดงได้ดังนี้

32 15 45 12 90 25

จงหามัธยฐานของข้อมูลชุดนี้

วิธีทำ

เรียงระยะเวลาที่ใช้ในการเดินทางจากบ้านไปโรงเรียนของนักเรียน 6 คน จากน้อยไปมาก ได้ดังนี้

12 15 25 32 45 90

เนื่องจากมัธยฐานอยู่ในตำแหน่งที่ $6+1$ ส่วน $2 = 3.5$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ ค่าเฉลี่ยเลขคณิตของข้อมูลในตำแหน่งที่ 3 และ 4 ซึ่งคือ $25+32$ ส่วน $2 = 28.5$ นาที

จากตัวอย่างข้างต้น จะเห็นว่า 90 เป็นค่าที่สูงกว่าค่าส่วนใหญ่ของข้อมูลชุดนี้ แต่ไม่มีผลต่อลำดับของข้อมูล ดังนั้น 90 จึงไม่มีผลต่อมัธยฐานของข้อมูลชุดนี้

ตัวอย่างฐานนิยมระดับง่าย

ตัวอย่างมัธยฐานระดับง่ายที่ 1. จำนวนดินสอที่นักเรียนแต่ละคนนำมาโรงเรียน คือ 1, 2, 2 และ 3 แท่ง
จงหาฐานนิยมของจำนวนดินสอได้เท่าใด

วิธีทำ

เลข 2 ปรากฏซ้ำมากที่สุด

ตอบ ฐานนิยมเท่ากับ 2 แท่ง

ตัวอย่างมัธยฐานระดับง่ายที่ 2. จำนวนนักเรียนที่มาสายในแต่ละวัน คือ 5, 5, 6 และ 7 คน
จงหาฐานนิยมของจำนวนนักเรียนที่มาสายได้เท่าใด

วิธีทำ

เลข 5 ปรากฏซ้ำมากที่สุด

ตอบ ฐานนิยมเท่ากับ 5 คน

ตัวอย่างฐานนิยมระดับปานกลาง

ตัวอย่างมัธยฐานระดับปานกลางที่ 1. จำนวนนักเรียนที่ใช้ห้องสมุดในแต่ละวัน คือ 20, 25, 25, 30 และ 35 คน
จงหาฐานนิยมของจำนวนนักเรียนที่ใช้ห้องสมุดได้เท่าใด

วิธีทำ

เลข 25 ปรากฏซ้ำมากที่สุด

ตอบ ฐานนิยมเท่ากับ 25 คน

ตัวอย่างมัธยฐานระดับปานกลางที่ 2. จำนวนลูกค้าที่เข้าใช้บริการร้านกาแฟในแต่ละช่วงเวลา คือ
10, 20, 20, 30, 30 และ 30 คน

จงหาฐานนิยมของจำนวนลูกค้าได้เท่าใด

วิธีทำ

เลข 30 ปรากฏซ้ำมากที่สุด

ตอบ ฐานนิยมเท่ากับ 30 คน

ตัวอย่างฐานนิยระดับยาก

ตัวอย่างมัธฐานระดับยากที่ 1. จำนวนครั้งที่นักเรียนเข้าร่วมกิจกรรมในแต่ละเทอม คือ
15, 15, 18, 18, 23 และ 23 ครั้ง

จงพิจารณาฐานนิยของจำนวนครั้งที่เข้าร่วมกิจกรรมได้เท่าใด

วิธีทำ

ทุกค่าปรากฏซ้ำเท่ากัน

ตอบ ไม่มีฐานนิย

ตัวอย่างมัธฐานระดับยากที่ 2. อายุ (ปี) ของนักเรียนที่มาเข้าค่ายคณิตศาสตร์ จำนวน 15 คน แสดงได้ดังนี้

5 8 7 6 7 8 12 11 10 11 8 6 8 7 8

จงหาฐานนิยของข้อมูลชุดนี้

วิธีทำ จากข้อมูลที่กำหนดให้ จะได้

อายุ 5 ปี มีความถี่เป็น 1

อายุ 6 ปี มีความถี่เป็น 2

อายุ 7 ปี มีความถี่เป็น 3

อายุ 8 ปี มีความถี่เป็น 5

อายุ 10 ปี มีความถี่เป็น 1

อายุ 11 ปี มีความถี่เป็น 2

อายุ 12 ปี มีความถี่เป็น 1

จะเห็นว่า อายุ 8 ปี มีความถี่สูงสุด

ดังนั้น ฐานนิยของข้อมูลชุดนี้ คือ 8 ปี