บทที่ 1
บทนำ

1.1 ประวัติย่อ

สถิติเป็นสาขาวิทยาศาสตร์หนึ่งที่ได้เจริญก้าวหน้ามาจากสมัยโบราณค่อนข้างจะธรรมชาติ ในบันทึกพระคัมภีร์ได้แสดงถึงการใช้ระบบของข้อมูลในการนับจำนวนสัตว์เลี้ยง การวัดพื้นที่ จำนวนเหล็กหล่อ ฯลฯ สำหรับขบวนของรัฐบาลได้มาจากการเก็บภาษีอากรและการนับจำนวนประชากร ในช่วงหลังปี ค.ศ.1600 ก็ได้มีบริษัทประกันภัยได้ใช้หลักความจริงเกี่ยวกับการเกิดการตายมาเป็นวิธีการเพื่อวางโครงชั่วอายุเฉลี่ย ขณะเดียวกันนี้ก็มีนักคณิตศาสตร์กลุ่มเล็ก ๆ ก็กำลังพิจารณาถึงความน่าจะเป็นและโอกาสที่เกี่ยวกับปัญหาของเกมบางชนิด บุคคลเหล่านี้มี Blaise Pascal, Pierre de Fermat และ James Bernoulli ได้พัฒนาแนวความคิดแรกที่นำไปสู่ทฤษฎีของความน่าจะเป็น

ในต้นปี ค.ศ.1700 Abraham Demoivre ได้พัฒนาหลักของกำลังสองน้อยที่สุดและสมการอธิบายถึงโค้งปกติ (รูประฆัง) ระหว่างระยะเวลาเดียวกันนี้ Carl Friedrich Gauss ได้พิสูจน์โค้งปกติจากการศึกษาของเกี่ยวกับค่าคลาดเคลื่อนจากการวัด ต่อมาในศตวรรษเดียวกัน จากการชักชวนของ Adolph Quetelet ได้นำวิธีการทางสถิติมาใช้กันอย่างกว้างขวาง Thomas Bayes ได้เสนอปรัชญาของความน่าจะเป็นที่เกี่ยวกับความคิดฝัน ซึ่งเป็นพื้นฐานของวิทยาศาสตร์การตัดสินใจมีประโยชน์มากโดยเฉพาะงานด้านธุรกิจและเศรษฐกิจ

ในปี ค.ศ.1800 ต่อมามีชาวอังกฤษหลายคนที่มีชื่อเสียงในทางสถิติอย่างเช่น Francis Galton และ Karl Pearson Galton เป็นญาติของ Charles Darwin ได้ใช้สถิติในการศึกษาพันธุกรรม Pearson ได้มีส่วนช่วยเหลือจำนวนมากเกี่ยวกับสถิติรวมถึงแสดงแนวความคิดของส่วนเบี่ยงเบนมาตรฐานที่เป็นงานวิเคราะห์สหสัมพันธ์

ในศตวรรษที่ 20 R.A. Fisher ได้ให้การสนับสนุนเกี่ยวกับการนำเอาวิธีการทางสถิติไปประยุกต์ให้สมบูรณ์ยิ่งขึ้นโดยเฉพาะทางการเกษตร มีอยู่คนหนึ่งคือ William Gossett ผู้ซึ่งได้เสนอ t-statistic หรือ ตัวสถิติ t สำหรับการอนุมานกรณี ตัวอย่างขนาดที่สุ่มเลือกมาจากการแจกแจงปกติที่ไม่ทราบความแปรปรวน

ในสหรัฐอเมริกาจำนวนของหลาย ๆ เหตุการณ์ได้นำสถิติไปสู่ความสนใจอย่างกว้างขวาง อย่างเช่น การแก้ไขการสำรวจสำมะโนประชากร งานเกี่ยวกับการรวบรวมข้อมูลสำมะโนประชากร ซึ่งใช้เวลาหลายเดือนของการสำรวจสำมะโนประชากร ปี ค.ศ. 1920 และ ค.ศ. 1930 นำไปสู่การสุ่มตัวอย่างในการสำรวจสำมะโนประชากรปี ค.ศ.1940 ในปี ค.ศ.1970 ข่าวสารเกือบทั้งหมดได้มาจากตัวอย่าง กำลังกระตุ้นในด้านอื่น ๆ วิธีการทางสถิติได้มาจากคะแนนออกเสียงเลือกตั้งของผู้มีสิทธิออกเสียง การเลือกตัวอย่างโดยขึ้นอยู่กับความน่าจะเป็นก็เป็นเรื่องธรรมดา สำหรับคะแนนออกเสียงเลือกตั้งของผู้มีสิทธิออกเสียง

มีชาวสหรัฐ จำนวนมากได้ให้การสนับสนุนสถิติแตกต่างกัน George Snedecor เป็นผู้ริเริ่มแนวความคิดปฏิบัติการทางสถิติ โดยตั้งศูนย์อาชีพบริการมหาวิทยาลัยและมวลชน แนวความคิดห้องปฏิบัติการสถิติของ Snedecor ครั้งแรกใช้ที่ Iowa State College ปัจจุบันมีอยู่หลาย ๆ แห่งในหลาย ๆ ประเทศ ในปี ค.ศ.1933 Jersey Neyman และ E.A. Pearson ได้ออกสิ่งพิมพ์ผลลัพธ์บางอย่างที่เป็นรากฐานต่อการทดสอบสมมติฐานหลายประเภท ปี ค.ศ.1940 W.A. Shewhart เป็นตัวจักรในการพัฒนากระบวนการสุ่มตัวอย่างเพื่อใช้ในการควบคุมคุณภาพในผลิตภัณฑ์อุตสาหกรรม

ในปี ค.ศ.1940 และ 1950 ได้ใช้ข้อมูลไปสัมพันธ์กับการพัฒนาของคอมพิวเตอร์ อย่างเช่น ปี ค.ศ.1970 คอมพิวเตอร์สามารถสรุปข้อมูลสำมะโนประชากรภายในสองสามสัปดาห์ ซึ่งแต่ก่อนจะต้องใช้เวลาเป็นเดือนอาจถึงปี ความเร็วของการสรุปข้อมูลการเลือกตั้งในเครือข่ายโทรทัศน์แห่งชาติก็ได้จากการคำนวณของคอมพิวเตอร์ การพัฒนาหลาย ๆ ด้านในอนาคตเกี่ยวกับสถิติไม่ต้องสงสัยว่าจะเกี่ยวโยงกับความก้าวหน้าในด้านวิทยาศาสตร์คอมพิวเตอร์

สถิติได้นำไปใช้กันอย่างกว้างขวางดังแสดงได้จากเนื้อหาภายใน American Statistical Association สถิติสำหรับธุรกิจ สถิติสำหรับเศรษฐศาสตร์ สถิติสำหรับสังคม สถิติสำหรับการศึกษา และ สถิติสำหรับชีววิทยา สถิติที่เกี่ยวกับปัญหาในที่ซึ่งมีข้อมูล ปริมาณและโอกาสของเหตุการณ์เกิดขึ้น เป็นต้น

1.2 สถิติคืออะไร

สถิติมีความหมายได้สองความหมายด้วยกัน ในความหมายแรก สถิติหมายถึงตัวเลขที่ได้รวบรวมขึ้นและหามาได้ในเรื่องใดเรื่องหนึ่ง เพื่อการตรวจสอบ อย่างเช่น เราสนใจจำนวนครั้งที่ชกชนะของแสนศักดิ์ เมืองสุรินทร์ จำนวนประตูที่เปเล่ยิงเข้าประตู เงินเดือนที่หามาได้ของนักธุรกิจที่มีชื่อเสียง จำนวนนักศึกษาใหม่แต่ละปีที่สมัครเข้าศึกษาในมหาวิทยาลัยรามคำแหง ตัวเลขหรือข้อมูลเหล่านี้คือสถิติที่ได้ถูกรวบรวมขึ้นเพื่อวัตถุประสงค์โดยเฉพาะ

สถิติในความหมายที่สองหมายถึงสาขาของการศึกษาที่เกี่ยวกับการรวบรวม การวิเคราะห์ การเสนอ การตีความ พร้อมทั้งการสรุปข้อมูลนั้นเพื่อใช้ในการตัดสินใจ เพราะว่ามีบุคคลจำนวนมากได้ทำการตัดสินใจโดยปราศจากข่าวสารที่สมบูรณ์อยู่บ่อย ๆ การศึกษาสถิติก็เพื่อให้การตัดสินใจได้ถูกต้องยิ่งขึ้น เมื่อไรนักศึกษาได้ศึกษาวิชาสถิติแล้ว นักศึกษาจะทราบว่าข้อมูลข่าวสารอะไรมีความสำคัญต่อผู้ทำการตัดสินใจ การรวบรวมข้อมูลข่าวสารได้อย่างไร และการตีความหมายข้อมูลได้ถูกต้องอย่างไร

นิยาม สถิติเป็นสาขาหนึ่งของการศึกษาเกี่ยวกับการรวบรวมและการวิเคราะห์ข้อมูลเพื่อใช้ในการทำการตัดสินใจ

โดยปกตินักสถิติแบ่งสถิติออกเป็นสองสาขา หนึ่งสถิติภาคพรรณนา เป็นสถิติเกี่ยวกับการรวบรวม การแจกแจงข้อมูล นี้หมายถึงการคำนวณค่าเฉลี่ย เขียนกราฟ การเสนอตาราง ฯลฯ มีบุคคลจำนวนมากคิดว่าสถิติมีเนื้อหาเพียงเท่านี้ อย่างไรก็ตาม สถิติภาคพรรณนาบางครั้งก็มีความสำคัญน้อยกว่าสถิติเชิงอนุมาน ขณะที่เราต้องใช้สถิติภาคพรรณนาเพื่อทำการตัดสินใจ เราพบว่าการทำการตัดสินใจในด้านสถิติแล้ว สถิติภาคพรรณนามีความสำคัญมากกว่า

นิยามสถิติภาคพรรณนา สถิติภาคพรรณนาเป็นสถิติสาขาหนึ่งเกี่ยวกับการรวบรวมและแจกแจงข้อมูล

สถิติในสาขาที่สองเป็นสถิติเชิงอนุมาน สถิติเชิงอนุมานเป็นส่วนของสถิติซึ่งเกี่ยวกับการดึงข้อสรุปบนพื้นฐานข้อมูลข่าวสารสถิติที่รวบรวมขึ้นจากตัวอย่าง ขบวนการทำการตัดสินใจจะเกี่ยวข้องกับความไม่แน่นอน ความไม่แน่นอนอาจถูกรวบรวมขึ้นจากส่วนหนึ่งของตัวแบบการตัดสินใจของเรา ตลอดการใช้ความน่าจะเป็น ด้วยเหตุนี้เราจึงมีเนื้อหาสาระที่สนใจสองสาขา สถิติภาคพรรณนาและสถิติเชิงอนุมาน แต่เราต้องพิจารณาความน่าจะเป็นด้วยเพื่อเชื่อมสองสาขา

นิยามสถิติเชิงอนุมาน สถิติเชิงอนุมานเป็นสถิติสาขาหนึ่ง ซึ่งเกี่ยวกับการดึงข้อสรุปเกี่ยวกับประชากรบนพื้นฐานของตัวอย่าง

โดยปกติการรวบรวมข้อมูลไม่ได้เกี่ยวกับสำมะโนประชากร เมื่อไรที่ใช้เทคนิคสำมะโนประชากรแล้วจะต้องใช้จำนวนของประชากรทั้งหมด นี้จะต้องเสียค่าใช้จ่ายและเวลาจำนวนมาก ซึ่งไม่จำเป็นทุกกรณีเสมอไปด้วย เราเลือกตัวอย่างจากประชากรเพื่อรวบรวมข้อมูลที่ต้องการก็เพียงพอ ประชากรเป็นเสมือนกลุ่มข้อมูลที่เราสนใจ ตัวอย่างเป็นส่วนหนึ่งหรือกลุ่มย่อยของประชากรนั้น เราอาจเลือกตัวอย่างหนึ่งอาจจะเป็นจำนวนเท่าใดก็ได้ เพื่อเป็นตัวแทนที่มีประโยชน์มากที่สุด ตัวอย่างที่ใช้เป็นตัวแทนประชากรที่ดีจะต้องมีประโยชน์ในการทำการตัดสินใจมากกว่าตัวอย่างที่ไม่มีประโยชน์ วิธีการที่ดีที่สุดเพื่อหาตัวอย่างที่ใช้เป็นตัวแทนจะต้องได้มาโดยวิธีสุ่ม ตัวอย่างสุ่มเป็น

ตัวอย่างที่ได้รับเลือกมาโดยที่ทุก ๆ สมาชิกของประชากรได้รับเลือกมามีโอกาสเท่า ๆ กัน

นิยามประชากร ประชากรคือกลุ่มของข้อมูลทั้งหมดซึ่งนักสถิติสนใจในการดึงข้อสรุป

นักสถิติมักนิยามประชากรเป็นกลุ่มของค่าสังเกตซึ่งเขาต้องการดึงข้อสรุปอย่างเช่น คะแนนทดสอบเชาว์ปัญญาของนักเรียนทั้งหมดที่ลงทะเบียนเรียนของโรงเรียนสาธิตมหาวิทยาลัยรามคำแหง หมายเลขจำนวนมากของเวลาที่ทำปฏิกิริยาที่ผู้ทดลองวัดเวลาต่าง ๆ กัน จำนวนห้องในบ้านในกรุงเทพฯ จำนวนเงินที่นักศึกษาแต่ละคนลงทะเบียนเรียนในหนึ่งปี ประชากรแบ่งได้เป็น 2 ประเภทคือ

1. ประชากรจำกัด หมายถึงประชากรที่จำกัดหรือกำหนดจำนวนของสิ่งของประชากรจำกัดอาจประกอบด้วยคะแนน 5 คะแนนหรือผู้มีสิทธิออกเสียงเลือกตั้งทั้งประเทศไทยก็ได้

2. ประชากรไม่จำกัด หมายถึงประชากรที่ไม่จำกัดหรือกำหนดจำนวนของสิ่งของ อย่างเช่น ประชากรไม่จำกัดประกอบด้วยผลลัพธ์ทั้งหมด (หัวและก้อย) ของจำนวนไม่จำกัดของการโยนของเหรียญหนึ่งอัน หรือคณะวิจัยการศึกษากำลังศึกษาผลของวิธีการสอนแบบใหม่ทางคณิตศาสตร์จากโรงเรียนมัธยมศึกษาตอนปลาย ประชากรที่จะศึกษานี้อาจพิจารณาไม่จำกัด เนื่องจากว่าจะรวมโรงเรียนมัธยมศึกษาในประเทศปัจจุบันจนถึงอนาคต

นิยามตัวอย่าง ตัวอย่างเป็นกลุ่มย่อยของประชากรที่ซึ่งเลือกมาเป็นตัวแทนของประชากร

ตัวอย่างอาจมีขนาดเล็กจนกระทั่งประกอบด้วยหนึ่งค่าสังเกต หรืออาจมีขนาดใหญ่เกือบเท่าประชากรที่ขาดไปหนึ่งสมาชิกเท่านั้น

นิยามตัวอย่างสุ่ม ตัวอย่างสุ่มเป็นตัวอย่างหนึ่งที่ได้เลือกมาในวิถีทางนั้น ๆ ที่แต่ละสมาชิกของประชากรมีโอกาสได้รับเลือกเท่า ๆ กันรวมกันเป็นตัวอย่าง

ตัวอย่าง เราสนใจที่จะศึกษาลักษณะนิสัยชอบซื้อของนักศึกษาปีที่หนึ่งของมหาวิทยาลัย ด้วยเหตุนี้ประชากรของเราก็สามารถนิยามได้เป็นเสมือนนักศึกษาปีที่หนึ่งของมหาวิทยาลัยต่าง ๆ ภายในประเทศ จำนวนของชนิดต่าง ๆ ของตัวอย่างอาจได้มาจากประชากรนี้ แต่ตัวอย่างสุ่มน่าจะเป็นตัวแทนเกือบทั้งหมดซึ่งแต่ละสมาชิกของประชากรมีโอกาสได้รับเลือกเท่า ๆ กัน สำหรับวิธีเลือก เราต้องรวบรวมชื่อของนักศึกษาปีที่หนึ่งทั้งหมดในประเทศ เราใส่ชื่อทั้งหมดในกล่องและเลือกขึ้นมา หรือเราใช้คอมพิวเตอร์เลือกตัวอย่างก็ได้ โดยปกติเราจะพบว่าตัวอย่างหนึ่งประกอบด้วยข้อมูลสองสามร้อยก็เป็นการเพียงพอที่จะเป็นตัวแทนประชากรที่ใหญ่มาก ๆ บางครั้งตัวอย่างอาจมีขนาดเล็กกว่าขึ้นอยู่กับชนิดของข้อมูลที่จะรวบรวม

นิยามตัวแปร ตัวแปรเป็นปริมาณหนึ่งที่เปลี่ยนแปลงได้หรือปริมาณหนึ่งที่มีค่าต่าง ๆ กัน

ตัวอย่าง ผลการสอบของนักศึกษาห้าครั้ง จะได้เกรดแปรไปในแต่ละครั้ง ในกรณีที่ตัวแปรคือเกรดของนักศึกษา ตัวแปรอื่น ๆ ก็มีความสูง น้ำหนัก ราคา เวลาของการปฏิกิริยาเป็นต้น ตัวแปรมีอยู่ 2 แบบ

1. ตัวแปรไม่ต่อเนื่อง ตัวแปรไม่ต่อเนื่องสามารถมีค่าที่แน่นอนเท่านั้น อย่างเช่นจำนวนนักเรียนของชั้นซึ่งอาจเป็น 20,22, 30 คนก็ได้ ไม่สามารถเป็น 22.3 หรือ 25.58 หรือเศษส่วน

2. ตัวแปรต่อเนื่อง ตัวแปรต่อเนื่องสามารถมีค่าหนึ่งค่าใดที่อยู่ระหว่างขีดจำกัดหรือช่วงที่ถูกกำหนดขึ้น อย่างเช่น เราอาจวัดความสูงของชายคนหนึ่งอย่างหยาบ ๆ ได้เป็น 62 นิ้วหรือเราอาจวัดได้ 61.7 นิ้วหรือ 61.68 นิ้ว หรือ 61.6832 นิ้วก็ได้ ความสูงสามารถมีค่าอยู่ระหว่างขีดจำกัดที่ถูกกำหนดขึ้นและค่าเหล่านี้เป็นตัวแปรต่อเนื่อง

นิยามตัวคงที่ ตัวคงที่เป็นปริมาณหนึ่งที่มีค่าเดียวเท่านั้นระหว่างการพิจารณาโดยเฉพาะ

ตัวอย่าง สมมติว่าเรากำลังศึกษาขนาดของแต่ละธนาคารและความต้องการเงินฝาก ในเมือง ๆ หนึ่งมีอยู่ห้าธนาคารและเงินฝากกับธนาคารเป็น 24 ล้านบาท 123.5 ล้านบาท 16.8 ล้านบาท 59.7 ล้านบาท และ 230.4 ล้านบาท เรารู้ว่าจำนวนธนาคารไม่เปลี่ยนระหว่างการศึกษา ด้วยเหตุนี้จำนวนห้าใช้แทนค่าคงที่แสดงถึงจำนวนธนาคารที่เราศึกษา สำหรับเงินฝากแปรจากธนาคารหนึ่งไปยังอีกธนาคารหนึ่ง

1.3 ข้อมูลและการรวบรวมข้อมูล

ข้อมูลที่ใช้สำหรับการวิเคราะห์สถิติโดยทั่ว ๆ ไปจะเป็นประเภทหนึ่งประเภทใดของสองประเภท อาจจะเป็นข้อมูลเชิงอนุกรมเวลาหรือข้อมูลเฉพาะระยะเวลาหนึ่ง (cross-section) ข้อมูลเชิงอนุกรมเวลาจะถูกรวบรวมขึ้นสำหรับหน่วยที่ถูกกำหนดขึ้นโดยเฉพาะตลอดระยะเวลา ข้อมูลอาจเป็นรายปี, สามเดือน, รายเดือน รายสัปดาห์หรือรายวัน ข้อมูลเฉพาะเวลาหนึ่งเกี่ยวข้องกับระยะเวลาหนึ่งโดยเฉพาะ แต่ถูกรวบรวมขึ้นจากแต่ละบุคคล แต่ละห้างหุ้นส่วน แต่ละพื้นที่ เป็นต้น

ตัวอย่างของข้อมูลเชิงอนุกรมเวลาได้แก่จำนวนเงินที่ขายได้รายสัปดาห์ของบริษัทหนึ่งโดยเฉพาะ เราควรจะหาข้อมูลทั้ง 52 ค่าสังเกตต่อปี สำหรับตัวอย่างข้อมูลเฉพาะระยะเวลาหนึ่งควรจะเป็นข้อมูลเกี่ยวกับจำนวนเงินที่ขายได้สำหรับหนึ่งสัปดาห์ของพนักงานขายต่าง ๆ ทั้งหมดในบริษัท ถ้าหากว่าบริษัทมีพนักงานขาย 100 คน จำนวนเงินที่ขายได้ของแต่ละคนถูกบันทึกไว้ในหนึ่งสัปดาห์โดยเฉพาะ ดังนั้นระยะเวลาหนึ่งสัปดาห์ควรจะให้ 100 ค่าสังเกตเกี่ยวกับจำนวนเงินที่ขาย ข้อมูลชนิดต่างกันย่อมให้ข้อมูลข่าวสารต่างชนิดกัน

วิธีการรวบรวมข้อมูลที่ใช้กันมี

1. **โดยการสัมภาษณ์** การเก็บรวบรวมข้อมูลวิธีนี้ทำได้โดยส่งเจ้าหน้าที่ที่ได้รับการฝึกอบรมดีแล้วไปสัมภาษณ์ผู้ที่เราต้องการทราบข้อเท็จจริง ผลที่ได้มักจะขึ้นอยู่กับผู้สัมภาษณ์เป็นสำคัญ ถ้าหากว่าผู้สัมภาษณ์ไม่มีความซื่อตรงหรือจริงใจต่องานแล้ว ข้อเท็จจริงที่ได้มาก็ผิดไม่มีประโยชน์ที่จะนำมาวิเคราะห์ต่อไปได้

2. **โดยตั้งกระทู้ถามส่งไปรษณีย์** วิธีนี้ใช้ติดต่อกับคนได้จำนวนมาก ใช้คนงานน้อย ค่าใช้จ่ายน้อยกว่า เป็นวิธีที่ง่ายกว่าและสะดวกกว่า ผู้ตอบมักจะให้คำตอบเปิดเผยกว่าเพราะการกรอกข้อความไม่ได้ทำต่อหน้าคนอื่น แต่มีข้อเสียอยู่บ้างคือใช้กับผู้ไม่รู้หนังสือไม่ได้คำตอบที่ผู้ตอบส่งมาอาจมีข้อบกพร่อง

เราจะใช้ข้อมูลชนิดใดชนิดหนึ่งนั้นขึ้นอยู่กับเราจะใช้ข้อมูลข่าวสารชนิดไหน โดยปกติเราพบว่า การรวบรวมข้อมูลเพื่อจัดให้ข้อมูลข่าวสารนั้นมีประโยชน์สำหรับวัตถุประสงค์เพื่อทำการตัดสินใจ

นิยามข้อมูล ข้อมูลใช้เพื่ออ้างอิงจำนวนตัวเลขข้อมูลข่าวสารที่มีอยู่ในประชากร กลุ่มข้อมูลเป็นกลุ่มของตัวเลข(หรือสถิติ) ใช้เพื่อการวิเคราะห์

การทดลองควรจะถูกออกแบบให้ได้ข้อมูลข่าวสารมากที่สุดเท่าที่สามารถรวบรวมข้อมูลได้

การรวบรวมข้อมูลอาจมาจากแหล่งปฐมภูมิหรือทุติยภูมิ ข้อมูลปฐมภูมิเป็นข้อมูลที่ถูกรวบรวมขึ้นโดยนักสถิติสำหรับใช้โดยเฉพาะข้อมูลปฐมภูมิสิ้นค่าใช้จ่ายและเวลามากในการรวบรวม ข้อมูลทุติยภูมิเป็นข้อมูลที่ได้รับการรวบรวมและพิมพ์อยู่ในบางแหล่งของข้อมูลทุติยภูมิแหล่งข้อมูลที่แตกต่างกันย่อมมีความแตกต่างกัน เรียงจากรายงานสิ่งพิมพ์ของบริษัทจนถึงรัฐบาลหรือการศึกษาโดยองค์การวิจัยเอกชน แหล่งข้อมูลที่สำคัญสำหรับนักสถิติควรจะเป็นแหล่งทุติยภูมิ

แหล่งข้อมูลอาจจะหาได้ง่ายหรือมีจำกัดมาก ขึ้นอยู่กับชนิดของข้อมูลที่หาได้ มีข้อมูลบางชนิดจะต้องเก็บรวบรวมเร็ว ๆ หรือทันทีแต่ยังเสนอใช้ไม่ได้ สำหรับข้อมูลทุติยภูมิจะต้องเสียค่าใช้จ่ายมากในการที่ไปเก็บมาทั้งหมดได้ มีแหล่งข้อมูลอยู่หลาย ๆ แห่งเพื่อที่จะเก็บข้อมูลไหนบ้างที่นำมาใช้ประโยชน์

วิธีการรวบรวมข้อมูลมีอยู่หลายวิธี เราอาจจะใช้โทรศัพท์ โทรทัศน์เป็นเครื่องมือติดต่อข้อเท็จจริงที่ต้องการ หรือส่งเจ้าหน้าที่ไปติดต่อโดยตรงหรือติดต่อทางไปรษณีย์ หรือใช้วิธีอื่น ๆ ก็ได้ แต่การใช้โทรศัพท์ โทรทัศน์ เราได้ข้อเท็จจริงอยู่ในวงจำกัด ซึ่งนำผลลัพธ์ไปสู่ความคลาดเคลื่อน

ที่สำคัญคือสิ้นค่าใช้จ่ายมากไม่ถูกต้องตามหลักสถิติ ซึ่งจะต้องสิ้นค่าใช้จ่าย เสียเวลาและความคลาดเคลื่อนน้อยที่สุด

1.4 การประยุกต์สถิติได้อย่างไร

มีวิธีการประยุกต์สถิติในด้านต่าง ๆ หลายสาขาด้วยวิธีการใช้ที่ธรรมดาที่สุดคือการรวบรวมข้อมูลข่าวสารเกี่ยวกับการดำเนินงานของห้างหุ้นส่วนเพื่อวิเคราะห์ตำแหน่งและการตัดสินใจเกี่ยวกับอนาคต โดยเฉพาะสถิติภาคพรรณนามีความสำคัญในการเปรียบเทียบสภาวะปัจจุบันกับสภาวะในอดีต เราจะพบว่าส่วนของการทำการตัดสินใจในตำราเกี่ยวข้องกับการทดสอบสมมติฐานและการประมาณค่ามีประโยชน์มากในการสร้างสูตรแนวความคิดทางสถิติให้เห็นได้ชัดเจนยิ่งขึ้น และทำให้การตัดสินใจมีความหมายยิ่งขึ้นด้วย การวิเคราะห์การถดถอยและสหสัมพันธ์ใช้สำหรับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสองตัวหรือมากกว่า เราจะเห็นได้โดยการสร้างสูตรความสัมพันธ์นั้น เราสามารถพยากรณ์เหตุการณ์ในอนาคตบนฐานของความสัมพันธ์ระหว่างตัวแปรหรือแนวโน้มของเวลา ดังนั้นสถิติจึงมีความสำคัญมากต่อกระบวนการทำการตัดสินใจ ถ้าหากว่าการรวบรวมที่ถูกต้อง ข้อมูลข่าวสารมีประโยชน์ ผู้ทำการตัดสินใจต้องมีข้อมูลข่าวสารบนพื้นฐานที่ถูกต้อง

1.5 เครื่องหมายของสัญลักษณ์

นักศึกษาจำนวนมากได้ประสบกับความยุ่งยากในการศึกษาวิชาสถิติ อย่างน้อยที่สุดเขาก็จะพูดว่าเขาไม่สามารถที่จะเรียนรู้ชั้นสูงขึ้นในสาขาวิชาที่เขาเลือกถ้าปราศจากความเข้าใจในวิธีการสถิติหรือพื้นฐานทางคณิตศาสตร์ ความมุ่งหมายของหนังสือเล่มนี้เพื่อต้องการให้นักศึกษามีความรู้ขั้นมูลฐานทางคณิตศาสตร์ซึ่งจะนำไปใช้ศึกษาวิชาสถิติเบื้องต้นอันประกอบด้วยพีชคณิต การบวก ลบ คูณ หาร ชั้นมัธยมศึกษา ซึ่งจะช่วยให้ผู้อ่านคุ้นเคยกับปัญหาธรรมดาดังตัวอย่าง การนำเสนอของจำนวนเลขด้วยสัญลักษณ์ x, y และ z คำไขของสมการอย่างง่าย การแทนค่า การใช้สัญลักษณ์แบบฟังก์ชัน ล็อกและตาราง

สูตรทั้งหมดเราศึกษาเพื่อไปประยุกต์กับกลุ่มของข้อมูลต่าง ๆ เราจะต้องเสนอตัวเลขหรือค่าสังเกตที่จะนำสูตรเหล่านี้ไปประยุกต์กับสัญลักษณ์ทั่ว ๆ ไป อย่างเช่น x, y และ z จะมีการดัดแปลงบ้างเล็กน้อยในกรณีที่จะนำไปสู่ความยุ่งยาก เนื่องจากตัวอักษรไม่เพียงพอเราจึงต้องใช้ subscripts ดังตัวอย่าง ชั่วโมงอายุของหลอดไฟฟ้าชนิด A 5 ดวง มี 985, 863, 1024, 972 และ 746 ชั่วโมง ชั่วโมงอายุของหลอดไฟฟ้าชนิด B 5 ดวง มี 982, 1071, 993, 929 และ 785 ชั่วโมง เรา

ให้หลอดไฟฟ้าชนิด A 5 ดวง แทนด้วยสัญลักษณ์ x_1, x_2, x_3, x_4 และ x_5 หลอดไฟฟ้าชนิด B แทนด้วย y_1, y_2, y_3, y_4 และ y_5 โดยใช้อักษรต่างกันระหว่างสองตัวอย่าง ถ้าเราต้องการจะกล่าวถึงจำนวนหนึ่งจำนวนใด ของจำนวนเหล่านี้โดยทั่ว ๆ ไป เราใช้ x_i หรือ y_j เมื่อ i และ j เรียกว่า subscripts ของตัวแปรค่าซึ่งในตัวอย่างนี้เป็นได้จากค่า 1,2,3,4 หรือ 5 (i = 1, 2, 3, 4 หรือ 5; j = 1, 2, 3, 4 หรือ 5)

แทนที่จะเขียน subscripts i และ j เราสามารถใช้อักษรอื่น ๆ ก็ได้อย่างเช่น k, l, n,..... และใช้อักษรหรือสัญลักษณ์อื่น ๆ แทนอักษร x และ y ได้ ทั้งนี้ขึ้นอยู่กับชนิดหรือลักษณะของข้อมูลที่เหมาะสมใช้กับอักษรชนิดใดเราอาจจะเขียน x_{45} และ x_{73} สำหรับข้อมูลตัวที่สี่สิบห้าและเจ็ดสิบสาม

เพื่อที่จะทำสูตรให้ง่ายเข้าในโอกาสที่มีข้อมูลจำนวนมากเราให้สัญลักษณ์ Σ (เป็นอักษรกรีกอ่านว่า ซิกม่า) หมายถึง "ผลบวก" ดังคำจำกัดความเราเขียนได้

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + \ldots\ldots + x_N \qquad \ldots\ldots\ldots(1)$$

ซึ่งอ่านว่า "ผลบวกของ x_i i เริ่มจาก 1 ถึง N" จากคำจำกัดความนี้ เราอาจจะเขียนดังตัวอย่าง

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

หรือ $$\sum_{i=1}^{5} x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$$

$$\sum_{i=1}^{5} y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2$$

$$\sum_{i=2}^{5} x_i f_i = x_2 f_2 + x_3 f_3 + x_4 f_4 + x_5 f_5$$

ตัวอย่างสุดท้าย subscripts เริ่มจาก 2 และลงท้ายด้วย subscripts ด้วย 5 เครื่องหมายผลบวกจะปรากฏอยู่ในสูตรอีกมากที่จะช่วยพิสูจน์กฎขั้นมูลฐานบางกฎดังต่อไปนี้

กฎข้อ ก. บวกของผลบวก (หรือผลต่าง) ของเทอม สองเทอมหรือมากกว่าจะเท่ากับผลบวก (หรือผลต่าง) ของแต่ละเทอม

$$\sum_{i=1}^{N} (x_i + y_i + z_i) = \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} y_i + \sum_{j=1}^{N} z_i \qquad \ldots\ldots\ldots(2)$$

ถ้าเราต้องการใช้เครื่องหมายลบแทนเครื่องหมายบวก เราก็สามารถทำได้โดยเปลี่ยนเครื่องหมายทั้งสองข้างของสมการ การพิสูจน์กฎข้อ ก. จะละเว้นไว้ให้เป็นแบบฝึกหัดของนักศึกษาต่อไป

กฎข้อ ข.

$$\sum_{i=1}^{N} kx_i = k\sum_{i=1}^{N} x_i \qquad \text{.....(3)}$$

ผลบวกของค่าคงที่ k เท่าของตัวแปรค่า x_i จะเท่ากับค่าคงที่ k เท่าของผลบวกตัวแปรค่าดังพิสูจน์ได้

$$\begin{aligned}\sum_{i=1}^{N} kx_i &= kx_1 + kx_2 + kx_3 + \text{.....} + kx_N \\ &= k(x_1 + x_2 + x_3 + \text{.....} + x_N) \\ &= k\sum_{i=1}^{N} x_i\end{aligned}$$

กฎข้อ ค.

$$\sum_{i=1}^{N} k = Nk \qquad \text{.....(4)}$$

ผลบวกของค่าคงที่ k จาก 1 ถึง N จะเท่ากับผลคูณของค่าคงที่ k กับ N พิสูจน์ได้เช่นเดียวกับกฎข้อ ข.

$$\sum_{i=1}^{N} kx_i = kx_1 + kx_2 + kx_3 + \text{.....} + kx_N$$

เราสามารถให้ค่าทั้งหมดของ x_i เท่ากับ 1 จะได้

$$\sum_{i=1}^{N} k.1 = k.1 + k.1 + k.1 + \text{.....} + k.1$$

$$\sum_{i=1}^{N} k = Nk$$

หรือเราสามารถจะกล่าวได้ว่าค่าคงที่ k ไม่ได้ขึ้นอยู่กับอักษรชี้บอกเราก็สามารถเขียนได้ทันที

$$\sum_{i=1}^{N} k = k + k + k + \ldots\ldots\ldots + k$$
$$= Nk$$

แบบฝึกหัด

1. พิสูจน์กฎข้อ ก. ในกรณีที่มีตัวแปรค่า 3 ค่า x, y และ z และ i เริ่มจาก 1 ถึง 4
2. จงกระจายค่าต่อไปนี้ให้สมบูรณ์ (ปราศจากเครื่องหมายผลบวก)

(ก) $\sum_{i=1}^{8} x_i$ (ข) $\sum_{i=1}^{4} x_i^2$ (ค) $\sum_{j=1}^{3} (x_j + y_j)$

(ง) $\sum_{j=1}^{5} (x_j - a)$ (จ) $\sum_{i=2}^{6} x_i y_i$ (ฉ) $\sum_{i=1}^{5} x_i^2 f_i$

3. จงแสดงค่าเหล่านี้ให้อยู่ในรูปของผลบวก

(ก) $z_1 + z_2 + z_3 \ldots\ldots + z_{20}$ (ข) $x_1f_1 + x_2f_2 + x_3f_3 + x_4f_4 + x_5f_5 + x_6f_6$

(ค) $x_2^3y_2 + x_3^3y_3 + x_4^3y_4 + \ldots\ldots + x_7^3y_7$

(ง) $(x_1 - y_1) + (x_2 - y_2) + \ldots\ldots + (x_m - y_m)$

4. กำหนดให้

$x_1 = 3$ $x_2 = -2$ $x_3 = 1$ $x_4 = 0$ $x_5 = 2$

$f_1 = 2$ $f_2 = 8$ $f_3 = 20$ $f_4 = 12$ $f_5 = 3$

$y_1 = 7$ $y_2 = -4$ $y_3 = 8$ $y_4 = 5$ $y_5 = -3$

จงคำนวณค่าแต่ละค่าต่อไปนี้

(ก) $\sum_{i=1}^{5} x_i$ (ข) $\sum_{i=2}^{4} y_i$ (ค) $\sum_{i=1}^{5} x_i f_i$

(ง) $\sum_{i=1}^{5} x_i^2$ (จ) $\sum_{i=1}^{4} (x_i y_i)$ (ฉ) $\sum_{i=1}^{5} x_i^2 y_i$

5. จงพิสูจน์ว่า

$$\sum_{i=1}^{N} (x_i - k) = \sum_{i=1}^{N} x_i - Nk$$

6. โดยทั่ว ๆ ไป

$\left(\sum_{i=1}^{N} x_i \right)^2 = \sum_{i=1}^{n} x_i^2$ เป็นจริงหรือไม่?

7. จงพิสูจน์ว่า

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