สถิติธุรกิจ
Business Statistics
(BUA3127)

รัมภาภัค ฤกษ์วีระวัฒนา

ตามที่ระบุเนื้อหาของรายวิชา "สถิติธุรกิจ" ได้แบ่งเป็น 12 บทเนื้อหาตามดังนี้:

1. บทนำ
    - ความสำคัญของสถิติในการวิเคราะห์ข้อมูล
    - แนวคิดพื้นฐานเกี่ยวกับการสถิติ
    - สรุป

    เอกสารอ้างอิง
2. การเก็บรวบรวมข้อมูล
    - การเลือกและการสำรวจข้อมูล
    - การออกแบบแบบสำรวจ
    - การกระจายและการจัดเก็บข้อมูล
    - สรุป

    เอกสารอ้างอิง
3. การแสดงข้อมูล
    - การนำเสนอข้อมูลในรูปแบบต่างๆ
        - กราฟและแผนภูมิที่เหมาะสม
        - การสร้างตาราง
    - การนำเสนอข้อมูลที่เข้าใจง่าย
    - สรุป

    เอกสารอ้างอิง
4. การวิเคราะห์ข้อมูล
    - ประเภทของสถิติที่ใช้ในการวิเคราะห์ข้อมูล
    - สถิติบรรยาย
    - สถิติอ้างอิง
    - การสร้างและการอ่านผลการวิเคราะห์
    - สรุป

    เอกสารอ้างอิง

5. ตัวแปรสุ่มและการแจกแจง
    - การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม
    - การประมาณค่าความน่าจะเป็น
    - ตัวแปรสุ่มไม่ต่อเนื่อง
    - ตัวแปรสุ่มต่อเนื่อง
    - สรุป

เอกสารอ้างอิง

6. การประมาณค่า
    - การประมาณค่าแบบจุด
    - การประมาณค่าแบบช่วง
    - สรุป

เอกสารอ้างอิง

7. การทดสอบสมมติฐาน
    - การทดสอบสมมติฐานทางสถิติ
    - การตีความผลการทดสอบ
    - สรุป

เอกสารอ้างอิง

8. การวิเคราะห์ความแปรปรวน
    - การวิเคราะห์การแจกแจงความแปรปรวน
    - การใช้เครื่องมือสถิติในการวิเคราะห์ข้อมูลแปรปรวน
    - สรุป

เอกสารอ้างอิง

9. สถิตินอนพาราเมตริก
    - การคำนวณและการตีความสถิติพาราเมตริก
    - สรุป

เอกสารอ้างอิง

10. การวิเคราะห์การถดถอยและสหสัมพันธ์

- การทดสอบสหสัมพันธ์
- การวิเคราะห์การถดถอยและการใช้โมเดลทางสถิติ
- สรุป

เอกสารอ้างอิง

11. การพยากรณ์และการตัดสินใจทางธุรกิจ
    - การใช้สถิติในการพยากรณ์และการวางแผนทางธุรกิจ
    - การใช้ข้อมูลสถิติในการตัดสินใจทางธุรกิจ
    - สรุป

    เอกสารอ้างอิง
12. ดัชนีและการประยุกต์ใช้
    - ดัชนี
    - การประยุกต์ใช้สถิติในสาขาต่างๆ เช่น การวิเคราะห์การตลาด, การวิเคราะห์การเงิน
    - สรุป

    เอกสารอ้างอิง

บทที่ 1
บทนำ

สถิติมีบทบาทสำคัญอย่างมากในการวิเคราะห์และเข้าใจข้อมูลในหลากหลายด้านของชีวิต เริ่มตั้งแต่การตีความข้อมูลในสถานการณ์ประจำวันไปจนถึงการใช้สถิติในการตัดสินใจทางธุรกิจและการวิจัยทางวิทยาศาสตร์ เครื่องมือและแนวคิดทางสถิติช่วยให้เราสามารถวิเคราะห์และอธิบายข้อมูลในลักษณะที่เป็นรูปแบบ โดยเน้นการสร้างความเข้าใจและข้อสรุปที่ถูกต้อง เป็นเครื่องมือที่สำคัญในการช่วยประกอบการตัดสินใจที่มีความเชื่อถือได้ในทุกๆ ด้านของชีวิตและวิชาชีพ

แนวคิดและทฤษฎีบทนำสถิติ

สถิติเป็นส่วนที่สำคัญในการเรียนรู้เกี่ยวกับวิชานี้ เนื่องจากจะช่วยให้ผู้เรียนเข้าใจถึงความสำคัญดังนั้นจึงควรเข้าใจแนวคิดพื้นฐานของสถิติ มักจะประกอบด้วยเนื้อหาต่อไปนี้

1. **ความสำคัญของสถิติ**: สถิติช่วยอธิบายข้อมูลต่างๆและมีบทบาทสำคัญในชีวิตประจำวันและในการทำงานในหลายสาขา เช่น การบริหารจัดการ, การวิเคราะห์ข้อมูลทางการแพทย์, การวิเคราะห์การตลาด, ฯลฯ
2. **แนวคิดพื้นฐานเกี่ยวกับการสถิติ**: คือการอธิบายเกี่ยวกับแนวคิดพื้นฐานที่ใช้ในการสถิติ เช่น ตัวแปร, การกระจาย, ค่าสถิติพื้นฐาน เป็นต้น
3. **การใช้สถิติในการตัดสินใจ**: โดยเฉพาะในธุรกิจและองค์กร เน้นความสำคัญของการใช้ข้อมูลสถิติในการตัดสินใจอย่างมีประสิทธิภาพ
4. **การทำงานกับข้อมูล**: แสดงถึงขั้นตอนการเก็บรวบรวมข้อมูล, การอ้างอิงแหล่งข้อมูล, การจัดระเบียบข้อมูล, และการตรวจสอบความถูกต้องของข้อมูล
5. **การแสดงข้อมูล**: อธิบายเกี่ยวกับวิธีการนำเสนอข้อมูลทางสถิติให้เข้าใจง่าย เช่น การใช้กราฟ, แผนภูมิ, และตาราง

6. **การวิเคราะห์ข้อมูล**: อธิบายเกี่ยวกับกระบวนการในการวิเคราะห์ข้อมูลที่ได้ เช่น การคำนวณค่าสถิติต่างๆ และการตีความผลลัพธ์

7. **การนำสถิติไปใช้**: อธิบายเกี่ยวกับการนำความรู้สถิติไปใช้ในชีวิตประจำวันและในงานต่างๆ เพื่อการตัดสินใจและการแก้ปัญหา

หากนักศึกษาเข้าใจและตระหนักถึงความสำคัญและประโยชน์ของสถิติในชีวิตจริงจะช่วยสร้างพื้นฐานที่แข็งแกร่งในการเรียนรู้และใช้งานสถิติในอนาคต

ความสำคัญของสถิติในการวิเคราะห์ข้อมูล

สถิติมีความสำคัญอันโดดเด่นในการวิเคราะห์ข้อมูลเนื่องจากมีบทบาทสำคัญในหลากหลายด้านดังนี้:

1. **การสร้างความเข้าใจ**: สถิติช่วยให้เราเข้าใจและสร้างความเข้าใจเกี่ยวกับข้อมูลที่เรามีอยู่ เช่น การหาแนวโน้ม การกระจายของข้อมูลและความสัมพันธ์ระหว่างตัวแปรต่างๆ

2. **การตัดสินใจทางธุรกิจ**: สถิติช่วยให้ผู้บริหารและนักวิเคราะห์ทางธุรกิจตัดสินใจในทิศทางที่ถูกต้อง โดยใช้ข้อมูลที่เป็นหลักฐานและคาดการณ์ที่มีความเชื่อถือได้

3. **การวิเคราะห์ทางวิทยาศาสตร์**: ในการวิจัยทางวิทยาศาสตร์, สถิติช่วยในการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการทดลองและการสร้างทฤษฎีใหม่

4. **การวิเคราะห์ทางการแพทย์**: ในการวิเคราะห์ข้อมูลทางการแพทย์, สถิติช่วยให้เราสามารถทำนายโรค, ประสิทธิภาพของการรักษา, และการสร้างนโยบายสาธารณสุข

5. **การวิเคราะห์ข้อมูลทางสังคม**: สถิติช่วยให้เราเข้าใจแนวโน้มและภาวะสังคมที่สำคัญ เช่น การวิเคราะห์ข้อมูลทางเศรษฐกิจ, การประเมินผลโครงการสังคม

ดังนั้น สถิติเป็นเครื่องมือที่สำคัญในการช่วยให้เราวิเคราะห์และเข้าใจข้อมูลอย่างมีประสิทธิภาพในหลากหลายด้านของชีวิตและสาขาวิชาต่างๆ

แนวคิดพื้นฐานเกี่ยวกับการสถิติ

แนวคิดพื้นฐานในการสถิติเป็นพื้นฐานสำคัญที่เป็นฐานในการเรียนรู้และใช้งานสถิติ ซึ่งประกอบด้วยสิ่งต่อไปนี้:

1. **ตัวแปร (Variables)**: การสถิติมักเกี่ยวข้องกับการศึกษาตัวแปร โดยตัวแปรสามารถเป็นข้อมูลที่สามารถวัดได้หรือเป็นข้อมูลที่อยู่ในรูปแบบของกลุ่ม ตัวแปรสามารถเป็นตัวแปรต้น (Independent variables) และตัวแปรตาม (Dependent variables) ซึ่งมักนำมาใช้ในการวิเคราะห์และสร้างโมเดลทางสถิติ

2. **การกระจาย (Distribution)**: การกระจายของข้อมูลเป็นแนวคิดสำคัญในสถิติ เพราะมันช่วยให้เรารู้ถึงลักษณะการกระจายของข้อมูลว่าเป็นแบบไหน เช่น แบบเบี้ยวหรือแบบกระจายแบบปกติ

3. **ค่าสถิติ (Statistics)**: ค่าสถิติเป็นตัวชี้วัดที่ใช้สำหรับอธิบายและวิเคราะห์ข้อมูล ซึ่งมีหลายประเภท เช่น ค่าเฉลี่ย, ส่วนเบี่ยงเบนมาตรฐาน, ความถี่, และอื่นๆ ซึ่งใช้เพื่อสร้างความเข้าใจเกี่ยวกับข้อมูล (สุบิน ยุระรัช, 2566)

4. **การสร้างโมเดล (Modeling)**: การสถิติมักใช้ในการสร้างโมเดลเพื่อทำนายผลลัพธ์หรือเพื่ออธิบายความสัมพันธ์ระหว่างตัวแปร โดยใช้เครื่องมือทางสถิติต่างๆ เช่น การถดถอยเชิงเส้น, การสร้างโมเดลการจำลอง, หรือการใช้เทคนิคการทำนาย

5. **การทดสอบสมมติฐาน (Hypothesis Testing)**: การทดสอบสมมติฐานเป็นกระบวนการที่สำคัญในการตรวจสอบว่ามีความแตกต่างกันระหว่างกลุ่มข้อมูลหรือไม่ ซึ่งเป็นการใช้สถิติเพื่อตัดสินใจว่ามีความแตกต่างทางสถิติอย่างมีนัยสำคัญหรือไม่

ความหมายของสถิติ

สถิติ ในปัจจุบันมีความหมายกว้างๆ 2 ประการ ด้วยกันคือ

ประการแรก สถิติ หมายถึง ตัวเลขที่ได้จากการรวบรวมขึ้นและคิดคำนวณออกมาเพื่อแสดงให้เห็นข้อเท็จจริงบางอย่าง เช่น สถิติการวิ่ง 100 เมตรของประเทศไทย 10.5 วินาที สถิติการเกิดของคนไทยเมื่อปี 2536 คือ เกิด 1 คนทุก 10 นาที สถิติน้ำฝนที่ตกในภาคได้แต่ละปี เป็นต้น

ประการที่สอง สถิติ หมายถึงวิชาหรือศาสตร์ที่ว่าด้วยหลักการและระเบียบวิชาการทางสถิติซึ่งประกอบด้วยการเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูลและการตีความหมายข้อมูลและการตีความหมายของข้อมูล (สุคนธ์ ประสทธิ์วัฒนเสรี, n.d.)

ศัพท์ต่าง ๆ ที่ใช้วิชาสถิติ

ประชากร (population) หมายถึง สิ่งต่างๆ ทั้งหมดในเรื่องที่เราสนใจศึกษา ประชากรทางสถิติจะเป็นอะไรก็ได้ไม่ว่าคน สัตว์ หรือสิ่งของ(สุเมธ สมภักดี, 2550) ตัวอย่างเช่น

1. ถ้าสนใจศึกษารายได้ของคนไทยที่จบการศึกษา ปวส. ในประเทศไทย คนไทยทั้งประเทศที่จบการศึกษา ปวส. เป็นประชากร

2. ถ้าสนใจเกี่ยวกับรถยนต์เก๋งในประเทศญี่ปุ่น รถยนต์เก๋งของประเทศญี่ปุ่นทั้งหมดเป็นประชากร

3. ถ้าสนใจศึกษาการเจริญเติมโตของช้างเลี้ยงในประเทศไทย ประชากรทางสถิติก็คือ ช้างเลี้ยงในประเทศไทย ทุกตัว

ประชากรสามารถแบ่งได้เป็น 2 ประเภทคือ

ก.ประชากรที่มีจำนวนจำกัด (Finite Population) ได้แก่ ประชากรที่สามารถบอกถึงจำนวนของประชากรได้แน่นอน มีจำนวนเท่าไหร่ เช่น จำนวน

รถอีแต๊กในจังหวัดอุดรธานี จำนวนนักศึกษาของมหาวิทยาลัยราชภัฏสวนสุนันทา จำนวนผู้ติดเชื้อหวัด rsv ในจังหวัดสมุทรสาคร จำนวนประชากรที่กล่าวมานี้สามารถจะหาจำนวนแน่นอนได้ว่าจำนวนเท่าไหร่

ข. ประชากรที่มีจำนวนไม่จำกัด (Infinite Population) ได้แก่ ประชากรที่เราไม่สามารถบอกถึงจำนวนของประชากรได้ว่ามีจำนวนเท่าไร เช่น จำนวนเมล็ดข้าวที่ไม่สมบูรณ์ในผลผลิตข้าวปี 2566 หรือจำนวนหมูในประเทศไทยที่มีพยาธิตัวตืดอยู่ ประชากรเหล่านี้ไม่สามารถจะหาจำนวนที่แน่นอนออกมาได้

ตัวอย่าง (Sample) หมายถึง ส่วนหนึ่งหรือส่วนย่อย (Subset) ของประชากรที่ต้องการศึกษาเพื่อที่จะใช้ตัวอย่างเป็นตัวแทนของประชากรเพราะในทางปฏิบัติแล้วคงไม่สามารถนำทุกหน่วยของ ประชากรมาศึกษาได้ เนื่องจากมีจำนวนมาก ทำให้เสียเวลาเสียค่าใช้จ่ายมาก จึงนำเพียงส่วนหนึ่งของประชากรมาศึกษา เช่น สมมติว่านักศึกษาชั้นมัธยมศึกษาปีที่ 6 ในปี 2566 ทั้งหมด 200,000 คน จะต้องการศึกษาดูว่า นักศึกษาชั้นมัธยมปีที่ 6 จะเสียค่าใช้จ่ายในการศึกษาทั้งหมด เช่น ค่าบำรุง การศึกษา ค่าอาหาร ค่าเครื่องแต่งกาย ค่ากิจกรรม ฯลฯ ปีละเท่าไร ถ้าศึกษาโดยใช้นักศึกษาชั้นมัธยมศึกษาปีที่ 6 ในปี 2566 ทั้งหมด 200,000 คนอาจจะทำได้แต่เสียเวลาค่าใช้จ่ายสูง แต่ถ้าเลือกสุ่มนักศึกษาชั้นมัธยมศึกษาปีที่ 6 ในปี 2566 มา 2,000 คนมาศึกษาหาข้อมูลของค่าใช้จ่ายซึ่งผลที่ได้ ใกล้เคียงความจริงทำให้เสียค่าใช้จ่าย เสียเวลาในการเก็บของมูลน้อยลง นักศึกษาชั้นมัธยมศึกษาปีที่ 6 ในปี 2566 มา 2,000 คนนั้นจึงเป็นตัวอย่าง

ค่าสถิติ (Statistic) หมายถึง ค่าที่เกิดขึ้นมาจากการนำข้อมูลที่ได้จากกลุ่มตัวอย่างที่เราเก็บมา คำนวณเพื่อให้ได้ค่าตามที่ต้องการ การนำข้อมูลจากกลุ่มตัวอย่างมาคำนวณหาร้อยละ ค่าเฉลี่ยเลขาคณิต ค่ามัธยฐาน เป็นต้น เช่น จากตัวอย่างที่ผ่านมาถ้าเก็บข้อมูลโดยการสุ่มนำ นักศึกษามัธยมศึกษา ปีที่ 6 มา 2,000 คนมาศึกษาข้อมูลของค่าใช้จ่าย พบว่าค่าใช้จ่ายโดยเฉลี่ยทั้งหมดเป็น 10,000 บาท นี้คือ ค่าสถิติ

พารามิเตอร์ (Parameter) หมายถึง ค่าที่เกิดจากการคำนวณของข้อมูลที่เป็นทุกหน่วยของ ประชากร เช่น สนใจอายุการใช้งานของเครื่องปรับอากาศยี่ห้อ Daikin รุ่น Max Inverter ตั้งแต่ติดตั้ง จนกระทั่งเสียครั้งแรก สมมติเครื่องปรับอากาศยี่ห้อ Daikin รุ่น Max Inverter มีทั้งหมด 3,000 เครื่อง และเก็บข้อมูลว่าติดตั้งและเสียครั้งแรกเมื่อไรก็สามารถทราบถึงอายุการใช้งานแต่ละเครื่อง โดยการคำนวณหาอายุการใช้งานเฉลี่ยของเครื่องปรับอากาศยี่ห้อ Daikin รุ่น Max Inverter ได้และค่าอายุการใช้งานเฉลี่ยนี้คือพารามิเตอร์

ข้อมูล (Data) หมายถึง ข้อเท็จจริงหรือข่าวสารต่างๆ ที่รวบรวมเพื่อศึกษาเรื่องใดเรื่องหนึ่ง ซึ่ง ข้อมูลอาจจะเป็นตัวเลขหรือไม่เป็นตัวเลขก็ได้ ข้อมูลที่เป็นตัวเลขประกอบกันเป็นกลุ่มหรือข้อความที่เป็น จำนวนมาก เราเรียกว่า ข้อมูลสถิติ (Statistical Data) ข้อมูลเพียงหน่วยเดียวไม่ถือว่าเป็นข้อมูลสถิติ เช่น อายุของนักศึกษา 1 คน แต่อายุของนักศึกษาทุกคนในระดับ ปวส ถือเป็นข้อมูลสถิติ โดยทั่วไป ข้อมูลทางสถิติ โดยทั่วไปข้อมูลแบ่งได้เป็น 2 ประเภท ดังนี้ คือข้อมูลเชิงปริมาณ และข้อมูลเชิงคุณภาพ (Salkind, N. J. & Frey, B. B. ,2020).

1. ข้อมูลเชิงประมาณ (Quantitative data) หมายถึง ข้อมูลที่เก็บได้เป็นตัวเลขที่สามารถบอก ถึงขนาดหรือปริมาณหรือจำนวนได้ เช่น ข้อมูลที่เกี่ยวกับ อายุ จำนวนบุตร เงินเดือน ความสูง น้ำหนัก เป็นต้น

สมมติให้นักเรียนห้อง ปวส. การเงิน ชั้นปีที่ 1 ห้อง 2 มี 15 คน ซึ่งแต่ละคนได้คะแนนสอบวิชาบัญชี ดังนี้

50, 63, 39, 50, 65, 47, 17, 51, 61, 58, 46, 51, 42, 45, 35

ข้อมูลซึ่งเป็นตัวเลขทั้ง 15 ตัว นั้นเป็นชุดของข้อมูลซึ่งเป็นคะแนนวิชาบัญชีของนักเรียนปวส. การเงิน ชั้นปีที่ 1 ห้อง 2 บางครั้งเรียนตัวเลขแต่ละตัวว่าค่าสังเกต (Observation)

2. ข้อมูลเชิงคุณภาพ ได้แก่ ข้อมูลที่แสดงถึงคุณลักษณะของสิ่งนั้น เช่น นักเรียนจำแนกตาม เพศ คือ ชายหรือหญิง ขนาดของเสื้อยืด **S, M, L, XL** ขนาดรองเท้าเบอร์ **35 36 37 38 ... 45**

ลักษณะข้อมูลทางสถิติจำแนกตามแหล่งของข้อมูลทางสถิติ ซึ่งแบ่งได้เป็น 2 ชนิด คือ

1. ข้อมูลเชิงปฐมภูมิ คือ ข้อมูลที่ได้มาจากประชากรที่ต้องการศึกษาโดยตรงและเก็บ รวบรวมโดยการจดบันทึกหรือพิมพ์ เช่น สนใจการแก้ปัญหาสิ่งแวดล้อมของลำคลองในกรุงเทพมหานครของคนกรุงเทพมหานครแล้ว และสอบถามการแก้ปัญหาสิ่งแวดล้อมข้อมูลที่ได้มาจากแหล่งให้ข้อมูลโดยตรง จะเรียกข้อมูลที่ได้นี้ว่าข้อมูลปฐมภูมิ

2. ข้อมูลทุติยภูมิ คือ ข้อมูลที่ได้แหล่งข้อมูลอื่นที่ไม่ได้จากประชากรโดยตรง โดยแหล่งข้อมูลนี้ ได้เก็บข้อมูลที่สนใจไว้แล้วและเพียงไปนำมาใช้งาน เช่น ข้อมูลของสำนักงานสถิติแห่งชาติ หรือข้อมูล จากกรมวิชาการ กระทรวงเกษตรและสหกรณ์ ถือว่าเป็นข้อมูลทุติยภูมิ เช่น ถ้าอยากทราบปริมาณสัตว์ใหญ่ที่นำมาเป็นอาหารของคน ได้แก่ ไก่หมู โค กระบือ เป็นต้น เราอาจจะหาข้อมูลนี้ได้จากหนังสือ สถิติการเกษตรของประเทศไทยปีต่างๆ โดยที่ไม่ต้องไปนับจำนวน ไก่ หมู โค กระบือจากแหล่ง ต่างๆ ทั่วประเทศ

สถิติที่มีความหมายเป็นวิชาหรือศาสตร์ การดำเนินการทางสถิติที่เกี่ยวกับข้อมูลเป็นจำนวนมาก โดยมีขั้นตอนในการปฏิบัติ 4 ขั้นตอนดังนี้

ขั้นที่ 1 การเก็บรวบรวมข้อมูล คือ การรวบรวมข้อเท็จจริงจากตัวอย่างที่ผู้วิจัยสนใจในเรื่องนั้นมาข้อมูล

ขั้นที่ 2 การนำเสนอข้อมูล คือ การนำเอาข้อมูลที่เก็บรวบรวมมาได้แสดง ซึ่งอาจจะแสดงในรูปของกราฟหรือแผนภูมิ ตารางแจกแจงความถี่ เพื่อที่นำไปผลที่แสดงไปวิเคราะห์ และสรุปผลในขั้นต่อไป

ขั้นที่ 3 การวิเคราะห์ข้อมูล คือ การนำข้อมูลที่เก็บรวบรวมมาได้คำนวณเพื่อให้ได้ค่าสถิติที่เรา ต้องการที่จะใช้ประโยชน์ต่อไป เช่น การหาค่าเฉลี่ย การหาค่าส่วนเบี่ยงเบนมาตรฐาน

ขั้นที่ 4 การตีความหมายข้อมูล คือ การนำข้อมูลที่ผ่านการวิเคราะห์มาให้ความหมาย การสรุป และอธิบาย ลักษณะของข้อมูลที่ได้วิเคราะห์หรือคำนวณค่าเป็นค่าสถิติต่างๆ แล้วเพื่อให้ได้ผลสรุป แล้วจึงนำไปใช้ประโยชน์ต่อไป (สายชล สินสมบูรณ์ทอง, 2560)

สถิติเป็นวิชาหรือศาสตร์ที่มีความสำคัญในการวิเคราะห์ข้อมูล โดยมีขั้นตอนหลักในการปฏิบัติดังนี้: การเก็บรวบรวมข้อมูล, การนำเสนอข้อมูล, การวิเคราะห์ข้อมูล, และการตีความหมายข้อมูล การเก็บรวบรวมข้อมูลเป็นขั้นตอนแรกที่สำคัญโดยต้องรวบรวมข้อมูลที่สำคัญและเป็นประโยชน์จากตัวอย่างที่สนใจ เพื่อนำข้อมูลนั้นมานำเสนอต่อไปในขั้นตอนถัดไป ขั้นตอนการนำเสนอข้อมูลเป็นการแสดงผลข้อมูลที่เก็บรวบรวมมา ในรูปแบบต่างๆ เช่น กราฟ แผนภูมิ หรือตาราง เพื่อให้ข้อมูลนั้นมีความชัดเจนและเข้าใจได้ง่าย ขั้นตอนถัดไปคือการวิเคราะห์ข้อมูลโดยการนำข้อมูลมาคำนวณเพื่อให้ได้ค่าสถิติที่เหมาะสม เช่น ค่าเฉลี่ย หรือค่าส่วนเบี่ยงเบนมาตรฐาน เพื่อใช้ในการวิเคราะห์และสรุปผลในขั้นตอนต่อไป สุดท้ายคือการตีความหมายข้อมูล โดยนำข้อมูลที่ผ่าน การวิเคราะห์มาให้ความหมาย และสรุปผลให้เข้าใจได้ง่าย การดำเนินการทางสถิตินี้เป็นส่วนสำคัญในการใช้ข้อมูลให้เกิดประโยชน์และสามารถนำไปใช้ในการตัดสินใจอย่างมีเหตุผลได้อย่างเหมาะสม

ความสำคัญของของสถิติต่อธุรกิจ

สถิติมีความสำคัญอย่างมากต่อธุรกิจในหลายด้าน เพียงแก่บทบาทที่เสริมสร้างความเข้าใจในข้อมูลและช่วยในการตัดสินใจที่ดีขึ้นดังนี้:

1. **การวิเคราะห์ทางการตลาด**: สถิติช่วยในการวิเคราะห์ข้อมูลการตลาด เช่น การศึกษาความต้องการของลูกค้า, การวิเคราะห์ความสำเร็จของแคมเปญโฆษณา, และการวิเคราะห์ผลของกิจกรรมการตลาดต่างๆ(วัชราภรณ์ อารีรัตนศักดิ์และคณะ, 2563)

2. **การบริหารจัดการธุรกิจ**: สถิติช่วยในการวิเคราะห์ข้อมูลทางธุรกิจ เช่น การวิเคราะห์รายได้และกำไร, การประเมินประสิทธิภาพของการผลิตและการประเมินความเสี่ยงทางธุรกิจ

3. **การตัดสินใจเชิงกลยุทธ์**: สถิติช่วยในการตัดสินใจทางกลยุทธ์ เช่น การวิเคราะห์ข้อมูลเพื่อการวางแผนยุทธ์การตลาด, การพัฒนาผลิตภัณฑ์, และการเตรียมการต่างๆ สำหรับอุตสาหกรรม

4. **การควบคุมคุณภาพ**: สถิติช่วยในการวิเคราะห์และประเมินคุณภาพของผลิตภัณฑ์หรือบริการ เพื่อให้มั่นใจได้ว่ามีคุณภาพมาตรฐานและตอบสนองต่อความต้องการของลูกค้า

5. **การประเมินผล**: สถิติช่วยในการประเมินผลและการวิเคราะห์ข้อมูลเพื่อการปรับปรุง การใช้สถิติในการทำแบบสำรวจหรือการวิเคราะห์ข้อมูลจากลูกค้าช่วยให้ธุรกิจสามารถปรับปรุงสินค้าหรือบริการของตนได้อย่างต่อเนื่อง

การนำสถิติมาใช้ในธุรกิจช่วยให้ผู้บริหารและนักวิเคราะห์ธุรกิจสามารถตัดสินใจอย่างมั่นใจและมีประสิทธิภาพมากยิ่งขึ้นในการจัดการและพัฒนาธุรกิจของตน

ตัวอย่าง: การตัดสินใจเพื่อการวางแผนการผลิต

บริษัท XYZ ผลิตสินค้าอิเล็กทรอนิกส์และกำลังพิจารณาเพิ่มการผลิตสินค้าใหม่เข้าสู่ไลน์ผลิตของตน ก่อนที่จะตัดสินใจเรื่องนี้ พวกเขาต้องการทราบว่ามีความต้องการจำนวนสินค้าเท่าใดจากตลาดในอนาคต โดยใช้การวิเคราะห์ข้อมูลจากข้อมูลการขายของปีก่อนหน้านี้ พวกเขาสามารถใช้สถิติเพื่อ:

1. **การคาดการณ์ปริมาณการขายในอนาคต**: การวิเคราะห์แนวโน้มของยอดขายในอดีตจะช่วยให้พวกเขาสร้างแบบจำลองที่คาดการณ์ปริมาณการขายในอนาคตได้อย่างแม่นยำ
2. **การประมาณค่าความเสี่ยง**: การใช้เทคนิคสถิติเพื่อประมาณค่าความเสี่ยงที่อาจเกิดขึ้น เช่น การประมาณค่าความเสี่ยงในการลงทุนในการขยายกำลังการผลิตใหม่
3. **การวิเคราะห์ผลกระทบ**: การทำนายผลกระทบของการเพิ่มการผลิตสินค้าใหม่ต่อกำไรและกำลังการผลิตรวม ซึ่งจะช่วยในการตัดสินใจเกี่ยวกับการลงทุน

ดังนั้นการใช้สถิติในธุรกิจช่วยให้บริษัท XYZ มีข้อมูลที่แม่นยำและมั่นใจในการตัดสินใจเพื่อวางแผนการผลิตใหม่อย่างมีประสิทธิภาพ

ตัวอย่าง: การวิเคราะห์ข้อมูลการตลาดเพื่อวางแผนการโปรโมชั่น

บริษัท ABC เป็นบริษัทที่ผลิตเครื่องใช้ไฟฟ้าและกำลังพิจารณาการจัดโปรโมชั่นเพื่อเพิ่มยอดขายในช่วงเดือนสุดท้ายของปี เพื่อวางแผนโปรโมชั่นให้เหมาะสมและมีผลสัมฤทธิ์ พวกเขาจึงใช้สถิติเพื่อ:

1. **การวิเคราะห์ข้อมูลลูกค้า**: การวิเคราะห์ข้อมูลลูกค้าเกี่ยวกับการซื้อที่ผ่านมา เช่น วันเวลาที่ทำการซื้อ, ประเภทสินค้าที่ซื้อบ่อยที่สุด, และจำนวนเงินที่ใช้ในการซื้อ เพื่อให้เข้าใจลักษณะและพฤติกรรมของลูกค้า
2. **การประมาณค่าความน่าสนใจของโปรโมชั่น**: การใช้สถิติเพื่อประมาณค่าความน่าสนใจของโปรโมชั่น เช่น ความน่าจะเป็นที่ลูกค้าจะสนใจและเข้ามาซื้อสินค้าที่มีโปรโมชั่น
3. **การวิเคราะห์ผลกระทบ**: การทำนายผลกระทบของการจัดโปรโมชั่นต่อยอดขาย และการประเมินผลในการดำเนินการโปรโมชั่น
4. **การทดลองสองกลุ่ม (A/B Testing)**: การใช้สถิติในการทดลองสองกลุ่มเพื่อทดสอบประสิทธิภาพของแต่ละโปรโมชั่น และเลือกโปรโมชั่นที่มีผลสูงสุด

ดังนั้นการใช้สถิติในการวิเคราะห์และวางแผนโปรโมชั่นในบริษัท ABC ช่วยให้บริษัทมีการตัดสินใจทางการตลาดที่มีความแม่นยำและมั่นใจได้ในการเพิ่มยอดขายและสร้างความพึงพอใจให้กับลูกค้าได้อย่างเป็นระบบ

สรุป

ในบทนี้ได้ศึกษาถึงความสำคัญของสถิติในการวิเคราะห์ข้อมูลและการตัดสินใจในธุรกิจ โดยได้สรุปถึงวิธีการใช้สถิติในหลายด้านของธุรกิจ เช่น การวิเคราะห์ทางการตลาดเพื่อวางแผนการโปรโมชั่น, การตัดสินใจเชิงกลยุทธ์ในการบริหารจัดการธุรกิจ, การควบคุมคุณภาพผลิตภัณฑ์หรือบริการ, และการวางแผนการผลิต การใช้สถิติในธุรกิจช่วยให้ผู้บริหารและนักวิเคราะห์สามารถตัดสินใจอย่างมั่นใจและมีประสิทธิภาพในการจัดการและพัฒนาธุรกิจของตนอย่างเป็นระบบ

เอกสารอ้างอิง

วัชราภรณ์ อารีรัตนศักดิ์และคณะ. (2563). ความสัมพันธ์เชิงสาเหตุของคุณภาพการบริการ ความพึงพอใจ ความไว้เนื้อเชื่อใจ ที่มีต่อการกลับมาใช้บริการซ้ำของลูกค้าในธุรกิจร้านอาหาร. *วารสารสหวิทยาการเพื่อการพัฒนา มหาวิทยาลัยราชภัฏอุตรดิตถ์.* 10(2) 73-85.

สายชล สินสมบูรณ์ทอง. (2560). *สถิติเบื้องต้น*. (พิมพ์ครั้งที่ 11). กรุงเทพฯ: จามจุรีโปรดักส์.

สุคนธ์ ประสทธิ์วัฒนเสรี. (n.d.). *สถิติเบื้องต้น*. ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

สุบิน ยุระรัช. (2566). แนวทาง การสอนสถิติสำหรับนักศึกษาปริญญาเอกที่วิชาเอกไม่ใช่สถิติ. *วารสารนวัตกรรมและการจัดการ*. 8(1) 106-118.

สุเมธ สมภักดี. (2550). ทฤษฎีการเลือกตัวอย่าง. ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

บทที่ 2
การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเป็นกระบวนการที่มีความสำคัญอย่างมากในการทำงานทางธุรกิจ เนื่องจากมีบทบาทสำคัญในการเตรียมข้อมูลที่เป็นพื้นฐานสำหรับการวิเคราะห์และการตัดสินใจในระดับต่างๆ การเลือกและการสำรวจข้อมูลเป็นขั้นตอนแรกที่สำคัญในกระบวนการเก็บข้อมูล เพื่อให้ได้ข้อมูลที่เป็นประโยชน์และเชื่อถือได้ การออกแบบแบบสำรวจมุ่งเน้นการเตรียมคำถามและโครงสร้างแบบสำรวจให้เหมาะสมเพื่อเก็บข้อมูลที่ต้องการ และการกระจายและการจัดเก็บข้อมูลมีความสำคัญในการสร้างระบบเก็บข้อมูลที่มีประสิทธิภาพและปลอดภัย เพื่อให้สามารถเข้าถึงและนำข้อมูลไปใช้งานได้อย่างสะดวกและรวดเร็ว ดังนั้น กระบวนการเก็บรวบรวมข้อมูลเป็นหัวใจสำคัญที่เตรียมพร้อมสำหรับการวิเคราะห์ข้อมูลและการตัดสินใจในองค์กร

การเลือกและการสำรวจข้อมูล

การเลือกและการสำรวจข้อมูลเป็นขั้นตอนที่สำคัญในการเก็บรวบรวมข้อมูลที่มีคุณภาพและมีประสิทธิภาพสำหรับการวิเคราะห์และการใช้ประโยชน์จากข้อมูลในองค์กร การเลือกข้อมูลเริ่มต้นจากการระบุวัตถุประสงค์และคำถามวิจัยที่ต้องการตอบ (วราพรรณ ธิยานันท์และกฤษณะ ไวยมัย ,2023) เพื่อให้สามารถเลือกเก็บข้อมูลที่เกี่ยวข้องและมีประโยชน์ตามความต้องการ เช่น หากต้องการศึกษาความพึงพอใจของลูกค้าในผลิตภัณฑ์ ก็จะต้องระบุว่าต้องการทราบถึงปัจจัยที่มีผลต่อความพึงพอใจ และเลือกเก็บข้อมูลที่เกี่ยวข้องเช่น ความพึงพอใจในคุณภาพสินค้า การบริการหลังการขาย หรือราคาของสินค้า

การสำรวจข้อมูลเป็นขั้นตอนที่ใช้เครื่องมือและวิธีการต่างๆ เพื่อเก็บข้อมูลจากแหล่งต่างๆ อย่างเชื่อถือได้ เช่น การใช้แบบสอบถาม การสัมภาษณ์

หรือการส่งอีเมล์สอบถาม เพื่อรวบรวมข้อมูลจากกลุ่มเป้าหมาย การสำรวจข้อมูลต้องมีการวางแผนและออกแบบอย่างเหมาะสมเพื่อให้ได้ข้อมูลที่ถูกต้องและเป็นประโยชน์ โดยคำถามที่ถูกต้องและคำนวณการตัดสินใจในการสำรวจข้อมูลจะช่วยให้ได้ข้อมูลที่มีคุณภาพและสามารถนำมาวิเคราะห์และใช้ประโยชน์ได้อย่างเหมาะสมในการตัดสินใจในธุรกิจ

วิธีการเลือกข้อมูล

การเลือกข้อมูลเป็นขั้นตอนที่สำคัญในการเก็บรวบรวมข้อมูลเพื่อให้ได้ข้อมูลที่เป็นประโยชน์และมีคุณภาพสำหรับการวิเคราะห์และการใช้ประโยชน์ในองค์กร คือขั้นตอนหลักในการเลือกข้อมูล:

1. **ระบุวัตถุประสงค์และคำถามวิจัย**: กำหนดวัตถุประสงค์ของการเก็บข้อมูลและคำถามวิจัยที่ต้องการตอบ เพื่อให้เก็บข้อมูลที่เป็นประโยชน์และเกี่ยวข้องกับความต้องการของการวิเคราะห์หรือการตัดสินใจที่จะทำในธุรกิจ
2. **กำหนดประเภทข้อมูลที่ต้องการ**: ระบุประเภทข้อมูลที่ต้องการเก็บ เช่น ข้อมูลปริมาณ, ข้อมูลคุณภาพ, ข้อมูลสภาพเศรษฐกิจ, หรือข้อมูลทางสังคม
3. **เลือกแหล่งข้อมูลที่เหมาะสม**: ค้นหาแหล่งข้อมูลที่เหมาะสมและเชื่อถือได้ เช่น ฐานข้อมูลออนไลน์, ข้อมูลจากหน่วยงานราชการ, หรือการสำรวจข้อมูลจากผู้ใช้บริการ
4. **วางแผนกระบวนการเก็บข้อมูล**: วางแผนกระบวนการเก็บข้อมูลอย่างรอบคอบโดยคำนึงถึงวิธีการเก็บข้อมูล เช่น การสำรวจข้อมูลโดยใช้แบบสอบถาม, การสร้างฐานข้อมูลออนไลน์, หรือการนำเข้าข้อมูลจากแหล่งอื่น
5. **ทดสอบและปรับปรุงกระบวนการ**: ทดสอบและปรับปรุงกระบวนการเก็บข้อมูลเพื่อให้ได้ข้อมูลที่มีคุณภาพและเชื่อถือได้ โดยอาจต้องทดสอบการสำรวจข้อมูลกับกลุ่มเป้าหมายก่อนการนำไปใช้จริง

การเลือกข้อมูลอย่างมีประสิทธิภาพจะช่วยให้ได้ข้อมูลที่เป็นประโยชน์และเชื่อถือได้สำหรับการวิเคราะห์และการตัดสินใจในธุรกิจ

วิธีการรวบรวมข้อมูล

การรวบรวมข้อมูลเป็นขั้นตอนสำคัญในการทำสถิติ มีหลายวิธีในการรวบรวมข้อมูล ต่อไปนี้คือวิธีการที่พบบ่อย

1. การสำรวจข้อมูล: การสำรวจข้อมูลโดยตรงจากแหล่งที่มา เช่น การสำรวจความคิดเห็นของลูกค้าหรือการสำรวจข้อมูลในงานวิจัย

2. การใช้เอกสารและบันทึกข้อมูล: การรวบรวมข้อมูลจากเอกสารต่างๆ เช่น รายงานการขาย, ใบเสร็จ, และเอกสารอื่นๆ ที่เกี่ยวข้อง

3. การสืบค้นข้อมูล: การค้นหาข้อมูลจากแหล่งต่างๆ เช่น ฐานข้อมูลออนไลน์, เว็บไซต์, หรือห้องสมุด

4. การใช้ชุดข้อมูลอื่น: การใช้ข้อมูลที่มีอยู่แล้วจากแหล่งอื่น เช่น ข้อมูลทางการศึกษาจากกระทรวงศึกษาธิการหรือข้อมูลทางการแพทย์จากหน่วยงานที่เกี่ยวข้อง

5. การสำรวจตามสัญญาณ: การรวบรวมข้อมูลจากสัญญาณหรือเครื่องมือตรวจวัด เช่น การรวบรวมข้อมูลจากเซ็นเซอร์หรือเครื่องวัดอุณหภูมิ

6. การสำรวจการกระทำ: การรวบรวมข้อมูลโดยตรงจากการกระทำของบุคคลหรือองค์กร เช่น การเก็บข้อมูลการจ่ายเงินจากธนาคารหรือการเก็บข้อมูลการใช้บริการจากหน่วยงานต่างๆ

7. การใช้วิธีอื่นๆ: การรวบรวมข้อมูลโดยใช้วิธีอื่นๆ อาจเป็นการสำรวจกลุ่มเป้าหมายโดยใช้แบบสำรวจออนไลน์หรือการนัดหมายสัมภาษณ์

การเลือกวิธีการรวบรวมข้อมูลขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของการศึกษาหรือการวิจัยที่กำลังดำเนินการ

ประชากรและกลุ่มตัวอย่าง

ในงานวิจัยหรือการสำรวจข้อมูล, "ประชากร" (population) หมายถึง กลุ่มทั้งหมดของบุคคลหรือองค์กรที่ต้องการศึกษาหรือสำรวจข้อมูล เป็นกลุ่มที่มีลักษณะที่ต้องการให้ข้อมูลและทำการวิจัย

"กลุ่มตัวอย่าง" (sample) คือ ส่วนหนึ่งของประชากรที่ถูกเลือกมาเพื่อทำการวิจัยหรือสำรวจข้อมูล มักจะเลือกมาเพียงบางส่วนเท่านั้นเนื่องจากการสำรวจทั้งหมดของประชากรอาจไม่เป็นไปตามความเป็นจริง การเลือกกลุ่มตัวอย่างที่เป็น Representative of the population จะช่วยให้สามารถทำข้อสรุปหรือนำเสนอผลลัพธ์ที่สามารถนำไปใช้งานได้ในประชากรทั้งหมดได้ถูกต้องและเชื่อถือได้ เพื่อให้กลุ่มตัวอย่างเป็นRepresentative of the population มักจะใช้วิธีการสุ่มตัวอย่างออกมาจากประชากร แบ่งเป็นสองวิธีหลักๆ คือ1) การเลือกตัวอย่างแบบใช้ความน่าจะเป็น เช่น การสุ่มอย่างง่าย (Simple Random Sampling) และ การสุ่มแบบแบ่งชั้น (Stratified Sampling) โดยการสุ่มอย่างง่ายเป็นการเลือกตัวอย่างโดยไม่มีการกำหนดกลุ่มหรือแบ่งประชากร ในขณะที่การสุ่มแบบแบ่งชั้นจะแบ่งประชากรออกเป็นกลุ่มย่อยตามลักษณะที่สนใจ และสุ่มตัวอย่างจากกลุ่มแต่ละกลุ่มนั้นๆ เพื่อให้กลุ่มตัวอย่างมีความสัมพันธ์กับประชากรในทุกๆ ด้าน 2) การเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็น เช่น การเลือกตัวอย่างแบบตามโควต้า

การเลือกตัวอย่างแบบใช้ความน่าจะเป็น

การเลือกตัวอย่างแบบใช้ความน่าจะเป็น (probability sampling) เป็นวิธีการที่ให้โอกาสให้กับทุกตัวอย่างในประชากรที่ต้องการศึกษาหรือสำรวจข้อมูลเพื่อให้เกิดความเป็นธรรมและเชื่อถือได้ในการสรุปผลลัพธ์ (นภ

สร รัตนวุฒิขจร, 2565; ชัยณรงค์ เพียรภายลุนและคณะ,2563)

วิธีการเลือกตัวอย่างแบบใช้ความน่าจะเป็นมีหลายวิธีเช่น:

1. **การสุ่มอย่างง่าย (Simple Random Sampling)**: ในวิธีนี้ เลือกตัวอย่างจากประชากรโดยไม่มีการคำนึงถึงลักษณะหรือแบ่งกลุ่ม แต่เพียงแค่สุ่มตัวอย่างออกมาจากประชากรโดยมีโอกาสเท่าๆ กันทุกตัวอย่าง

2. **การสุ่มแบบแบ่งชั้น (Stratified Random Sampling)**: ในวิธีนี้ ประชากรจะถูกแบ่งเป็นกลุ่มย่อยตามลักษณะที่สนใจ เช่น อายุ, เพศ, หรือระดับการศึกษา และจากนั้นจะสุ่มตัวอย่างออกมาจากแต่ละกลุ่มย่อยนั้นๆ

3. **การสุ่มแบบเป็นระบบ (Systematic Sampling)**: ในวิธีนี้ กำหนดระยะห่างระหว่างตัวอย่างที่สุ่มออกมา แล้วเลือกตัวอย่างตามลำดับที่กำหนด

4. **การสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling)** จะใช้เมื่อมีการศึกษาประชากรจำนวนมาก โดยการแบ่งกลุ่มประชากรออกเป็นกลุ่มย่อย อาจแบ่งตามพื้นที่หรือกลุ่มที่มีการจัดแบ่งไว้อยู่แล้ว จากนั้นก็ทำการศึกษาตามการแบ่งกลุ่มประชากรเฉพาะกลุ่มนั้นๆ แต่การสุ่มตัวอย่างแบบแบ่งกลุ่มนี้ยังมีข้อควรคำนึงถึงความแตกต่างในการแบ่งกลุ่มด้วยเช่นกัน ไม่ว่าจะเป็นเรื่องขนาดของกลุ่มความแตกต่างกันของประชากรในแต่ละกลุ่ม ขนาดของกลุ่มตัวอย่างที่ทำการคัดเลือก

การเลือกตัวอย่างแบบใช้ความน่าจะเป็นช่วยให้ผู้วิจัยหรือผู้สำรวจข้อมูลได้รับผลการวิจัยที่เชื่อถือได้และมีความสัมพันธ์กับประชากรทั้งหมดโดยสุ่มได้อย่างมีประสิทธิภาพและไม่มีการเลือกเฉพาะบางกลุ่มเท่านั้น

การเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็น

การเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็น (non-probability sampling) คือ วิธีการที่ผู้วิจัยหรือผู้สำรวจข้อมูลเลือกตัวอย่างจากประชากรโดยไม่ให้โอกาสเท่าๆ กันให้กับทุกตัวอย่างในประชากร วิธีการเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็นมักใช้ในกรณีที่มีข้อจำกัดในการทำการวิจัยหรือสำรวจข้อมูล และไม่สามารถใช้วิธีการสุ่มได้ เช่น มีค่าใช้จ่ายสูง, ไม่สะดวกในการเข้าถึงประชากร, หรือมีเวลาจำกัด เป็นต้น

วิธีการเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็นสามารถแบ่งได้เป็นหลายรูปแบบ เช่น

1. **การสุ่มตัวอย่างแบบสะดวก (Convenience Sampling)**: เลือกตัวอย่างโดยการเลือกตามความสะดวกหรือความสมัครใจของผู้วิจัยหรือผู้สำรวจข้อมูล เช่น การสำรวจความคิดเห็นของผู้เข้าร่วมงานสัมมนา

2. **การสุ่มตัวอย่างแบบเจาะจง (Purposive Sampling)**: เลือกตัวอย่างโดยคำนึงถึงลักษณะหรือคุณสมบัติที่ต้องการศึกษา และเลือกตัวอย่างที่ตรงตามลักษณะหรือคุณสมบัตินั้น เช่น การเลือกตัวอย่างคนที่มีประสบการณ์ในงานเฉพาะ

3. **การสุ่มตัวอย่างแบบตามโควต้า (Quota Sampling)**: กำหนดโควตาให้กับกลุ่มหรือลักษณะที่สนใจ เช่น โควตาของอายุ, เพศ, หรือระดับการศึกษา แล้วเลือกตัวอย่างจากแต่ละกลุ่มให้ครบตามโควตาที่กำหนด

4. **การสุ่มตัวอย่างแบบอ้างอิงด้วยบุคคลและผู้เชี่ยวชาญ (Snowball Sampling)** : คือ การสุ่มเลือกตัวอย่างมา 1 คน จากนั้นผู้ที่ได้รับการเลือกก็ทำการเสนอหรือคัดเลือกผู้คนที่มีลักษณะใกล้เคียงต่อไป จะคล้ายกับการแนะนำปากต่อปากนั่นเอง

การเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็นมักใช้เมื่อมีข้อจำกัดหรือเงื่อนไขที่ทำให้การสุ่มไม่เป็นไปตามความเป็นจริง แต่ควรระมัดระวังในการวิเคราะห์และสรุปผลของการวิจัยหรือการสำรวจข้อมูล เนื่องจากอาจมีการสะท้อนผลลัพธ์ที่ไม่สามารถแยกแยะความเป็นจริงของประชากรได้ถูกต้อง

การออกแบบแบบสำรวจ

การออกแบบแบบสำรวจเป็นขั้นตอนสำคัญที่มีผลต่อคุณภาพของข้อมูลที่สะสมได้ ดังนั้น การออกแบบแบบสำรวจควรคำนึงถึงปัจจัยต่างๆ เพื่อให้ได้ข้อมูลที่ถูกต้องและมีประสิทธิภาพในการใช้ประโยชน์ ประกอบด้วยขั้นตอนต่างๆ ดังนี้

1. **กำหนดวัตถุประสงค์ของการสำรวจ**: กำหนดวัตถุประสงค์หรือคำถามวิจัยที่ต้องการตอบหรือข้อมูลที่ต้องการเก็บ เพื่อช่วยให้การสำรวจมีความมุ่งหวังและเหมาะสมต่อเนื้อหาที่ต้องการตรวจสอบ

2. **เลือกวิธีการสำรวจ**: เลือกวิธีการสำรวจที่เหมาะสมกับวัตถุประสงค์และลักษณะของประชากร เช่น การสำรวจด้วยแบบสอบถาม, การสัมภาษณ์, การสังเกตการณ์

3. **ออกแบบคำถาม**: ออกแบบคำถามให้มีความชัดเจนและเข้าใจได้ง่าย ไม่มีคำถามที่สับสนหรือกำหนดมากเกินไป เพื่อป้องกันการเข้าใจผิดและข้อมูลที่ไม่ถูกต้อง

4. **วางโครงสร้างและลำดับของคำถาม**: การวางโครงสร้างและลำดับของคำถามให้มีระเบียบ จากคำถามที่ง่ายไปสู่คำถามที่ยาก หรือจากคำถามที่ไม่เป็นส่วนต่อเนื่องไปสู่คำถามที่เป็นส่วนต่อเนื่อง เพื่อเพิ่มประสิทธิภาพในการทำแบบสำรวจ

5. **ทดสอบและปรับปรุงแบบสำรวจ**: ทดสอบแบบสำรวจกับกลุ่มเล็กของผู้ร่วมสำรวจเพื่อให้แน่ใจว่าคำถามเข้าใจได้ และไม่มีข้อผิดพลาด จากนั้นปรับปรุงและปรับแก้ตามความเหมาะสม

6. **การจัดการและการสำรวจ**: จัดทำและดำเนินการสำรวจตามแบบทดสอบที่ออกแบบไว้ ให้เก็บข้อมูลที่สะสมได้มาอย่างถูกต้องและครบถ้วน

7. **การวิเคราะห์ข้อมูล**: วิเคราะห์ข้อมูลที่ได้รับจากการสำรวจอย่างถูกต้อง และสรุปผลลัพธ์ที่ได้ให้เป็นไปตามวัตถุประสงค์ของการวิจัยหรือการสำรวจข้อมูล

การออกแบบแบบสำรวจอย่างถูกต้องและเหมาะสมจะช่วยให้การสำรวจมีประสิทธิภาพและนำมาใช้ประโยชน์ได้อย่างเต็มประสิทธิภาพ

ตัวอย่างการออกแบบแบบสำรวจเพื่อศึกษาความคิดเห็นของลูกค้าต่อการบริการในร้านอาหารดังนี้:

1. **วัตถุประสงค์ของการสำรวจ**: วัตถุประสงค์คือการเก็บความคิดเห็นและความพึงพอใจของลูกค้าต่อการบริการในร้านอาหาร เพื่อปรับปรุงคุณภาพและบริการให้ดียิ่งขึ้น
2. **วิธีการสำรวจ**: ใช้แบบสอบถาม (questionnaire) เป็นเครื่องมือในการสำรวจ โดยทำการกรอกข้อมูลผ่านการสอบถามลูกค้าโดยตรง
3. **ออกแบบคำถาม**: จะมีคำถามที่เกี่ยวข้องกับประสบการณ์การบริการ รวมถึงความพึงพอใจในเมนูอาหาร, บรรยากาศในร้าน, การบริการจากพนักงาน เป็นต้น
4. **วางโครงสร้างและลำดับของคำถาม**: จัดลำดับคำถามให้มีระเบียบ เริ่มจากคำถามที่ง่ายและไม่กระทบต่อความเป็นส่วนต่อเนื่องไปจนถึงคำถามที่มีความเชื่อถือได้สูงที่สุด
5. **ทดสอบและปรับปรุงแบบสำรวจ**: ทำการทดสอบแบบสำรวจกับกลุ่มเล็กของลูกค้าเพื่อตรวจสอบความเข้าใจและความถูกต้องของคำถาม และทำการปรับปรุงตามความเหมาะสม
6. **การจัดการและการสำรวจ**: ส่งแบบสำรวจให้กับลูกค้าที่มาใช้บริการและรวบรวมข้อมูลที่ได้รับ
7. **การวิเคราะห์ข้อมูล**: วิเคราะห์ข้อมูลที่ได้รับเพื่อหาข้อสรุปและแนวโน้มของความคิดเห็นของลูกค้า และจัดทำรายงานสรุปผลเพื่อนำไปใช้ในการปรับปรุงบริการในร้านอาหาร

ดังนั้น การออกแบบแบบสำรวจนี้มีเป้าหมายชัดเจนและเป็นรายละเอียด เพื่อให้ได้ข้อมูลที่มีประโยชน์และสามารถนำไปปรับปรุงการบริการในร้านอาหารได้อย่างเหมาะสม

การกระจายและการจัดเก็บข้อมูล

การกระจายและการจัดเก็บข้อมูลเป็นหัวข้อที่สำคัญในวิชาสถิติ เนื่องจากมันเป็นพื้นฐานที่ต้องเรียนรู้เพื่อทำความเข้าใจข้อมูลและนำมา

วิเคราะห์ในด้านต่างๆ ในแนววิชาสถิติ การได้เรียนรู้เกี่ยวกับการกระจายของข้อมูล เช่น การใช้ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน พิสัย และการแจกแจงของข้อมูลต่างๆ ซึ่งเป็นเครื่องมือที่ใช้ในการวิเคราะห์และอธิบายข้อมูลในรูปแบบที่ชัดเจน นอกจากนี้ การศึกษาเรื่องการจัดเก็บข้อมูลในรูปแบบต่างๆ เช่น การสุ่มตัวอย่าง, การออกแบบแบบสำรวจ, และเทคนิคทางสถิติที่เกี่ยวข้อง เพื่อให้สามารถดำเนินการวิเคราะห์ข้อมูลได้อย่างถูกต้องและมีประสิทธิภาพ การเรียนรู้เรื่องนี้จะเป็นพื้นฐานที่สำคัญในการศึกษาต่อในระดับสูงขึ้นในสาขาที่เกี่ยวข้องกับสถิติและการวิเคราะห์ข้อมูลในอนาคต

สถิติเชิงบรรยาย (Descriptive Statistics)

สถิติเชิงบรรยาย (Descriptive Statistics) เป็นเครื่องมือสำคัญในการรวมและอธิบายข้อมูลที่มีอยู่ในรูปแบบที่เข้าใจง่าย โดยมักใช้สำหรับการสรุปข้อมูลและการเข้าใจลักษณะเฉพาะของข้อมูล นี่คือสูตรที่ใช้ในสถิติเชิงบรรยาย:

1. **ค่ากลาง (Measures of Central Tendency)**:

   Mean (ค่าเฉลี่ย)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Median (ค่ามัธยฐาน) มัธยฐาน คือ ค่าที่มีตำแหน่งอยู่กึ่งกลางของข้อมูลทั้งหมด เมื่อเรียบเรียงข้อมูลจากต่าน้อยที่สุดไปหาค่าที่มากที่สุด หรือจากค่าที่มากที่สุดไปหาค่าที่น้อยที่สุด เราอาจใช้ตัวย่อ "Med" แทนค่ามัธยฐานของข้อมูล ค่ามัธยฐานอาจเป็นค่าใดค่าหนึ่งของข้อมูล หรืออาจเป็นค่าที่คำนวณขึ้นมาใหม่ ซึ่งไม่ตรงกับค่าของข้อมูลใดๆก็ได้ ค่ามัธยฐานนิยมใช้กับข้อมูล ซึ่งสูงกว่าหรือต่ำกว่าค่าอื่นมากๆ

$$Mdn = \frac{N+1}{2}$$

Mode (ฐานนิยม) คือค่าที่ปรากฏบ่อยที่สุดในชุดข้อมูล

2. **การกระจาย (Measures of Dispersion)**:

- **Range (ช่วงของข้อมูล)**: Range=Max($\boldsymbol{X}$)–Min($\boldsymbol{X}$)
- **Variance (ความแปรปรวน)**:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

- **Standard Deviation (ส่วนเบี่ยงเบนมาตรฐาน)**: $S=\sqrt{S^2}$

3. **การกระจายรูปแบบ (Shape of Distribution)**

- **ความเบ้ (Skewness)**: วัดความเบ้หงายของการกระจายของข้อมูล ถ้ามีค่ามากกว่า 0 แสดงว่ามีการเบ้หงายไปทางขวา ถ้ามีค่าน้อยกว่า 0 แสดงว่ามีการเบ้หงายไปทางซ้าย (วรางคณา เรียนสุทธิ์, 2562)
- **ความโด่ง (Kurtosis)**: วัดความแตกต่างของหางยาวระหว่างการกระจายของข้อมูลกับการกระจายเส้นกลาง ค่ามากกว่า 3 หมายถึงมีหางยาวกว่าการกระจายปกติ ค่าน้อยกว่า 3 หมายถึงมีหางสั้นกว่าการกระจายปกติ

4. **การกระจายเชิงตัวแปร (Measures of Variability)**:

- **Percentile (เปอร์เซนไทล์)**: เปอร์เซนไทล์ (Percentile) เป็นค่าที่ใช้ในการแบ่งข้อมูลออกเป็นช่วงๆ แบ่งคะแนนหรือข้อมูลทั้งหมดออกเป็น

100 ส่วน โดยใช้เปอร์เซนไทล์เป็นตัวบ่งชี้ เช่น ค่าเปอร์เซนไทล์ที่ 25 (หรือ P25) คือค่าที่แบ่งข้อมูลออกเป็นส่วนของ 25% ที่มีค่าน้อยกว่าหรือเท่ากับค่านั้น ในขณะที่ 75 เปอร์เซนไทล์ (หรือ P75) คือค่าที่แบ่งข้อมูลออกเป็นส่วนของ 75% ที่มีค่าน้อยกว่าหรือเท่ากับค่านั้น

- **เดไซล์ (Decile)** เป็นค่าที่ใช้ในการแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน โดยใช้เดไซล์เป็นตัวบ่งชี้ เช่น ค่าเดไซล์ที่ 1 (หรือ D1) คือค่าที่แบ่งข้อมูลออกเป็นส่วนของ 10% ที่มีค่าน้อยกว่าหรือเท่ากับค่านั้น ในขณะที่ ค่าเดไซล์ที่ 10 (หรือ D10) คือค่าที่แบ่งข้อมูลออกเป็นส่วนของ 90% ที่มีค่าน้อยกว่าหรือเท่ากับค่านั้น

- **ควอร์ไทล์ (Quantile)** เป็นค่าที่ใช้ในการแบ่งข้อมูลออกเป็นส่วนๆ 4 ส่วน โดยมีค่ากำหนดล่วงหน้า เช่น ควอร์ไทล์ที่ 25 (หรือ Q1) คือค่าที่แบ่งข้อมูลออกเป็นส่วนของ 25% ที่มีค่าน้อยกว่าหรือเท่ากับค่านั้น ในขณะที่ควอร์ไทล์ที่ 75 (หรือ Q3) คือค่าที่แบ่งข้อมูลออกเป็นส่วนของ 75% ที่มีค่าน้อยกว่าหรือเท่ากับค่านั้น

5. **การสรุปข้อมูล (Data Summary)**:

- **Frequency Distribution (การกระจายความถี่)**: การแบ่งข้อมูลออกเป็นช่วงๆ และนับจำนวนข้อมูลที่อยู่ในแต่ละช่วง

- **Histogram (ฮิสโตแกรม)**: กราฟแสดงการกระจายความถี่ของข้อมูล

6. **การวัดความสัมพันธ์ (Correlation)**:

- **Correlation Coefficient (สัมประสิทธิ์สหสัมพันธ์)**: วัดความสัมพันธ์ระหว่างตัวแปร มีค่าอยู่ระหว่าง -1 ถึง 1 โดยค่าใกล้ 1 แสดงถึงความสัมพันธ์แบบบวก และค่าใกล้ -1 แสดงถึงความสัมพันธ์แบบลบ

เหล่านี้เป็นสูตรพื้นฐานที่ใช้ในการวิเคราะห์และอธิบายข้อมูลในรูปแบบของสถิติเชิงบรรยาย

สรุป

ในบทนี้ ได้ศึกษาเกี่ยวกับกระบวนการที่สำคัญในการเก็บรวบรวมข้อมูลและการนำเสนอข้อมูลในสถิติ ซึ่งเริ่มต้นด้วยการเลือกและการสำรวจข้อมูล ซึ่งเป็นขั้นตอนที่สำคัญในการเริ่มต้นวิเคราะห์ข้อมูล จำเป็นต้องทราบข้อมูลที่ต้องการทราบจะอยู่ที่ไหนและมีลักษณะอย่างไร จากนั้น ทำการศึกษาถึงการออกแบบแบบสำรวจ ซึ่งเป็นกระบวนการที่เกี่ยวกับการกำหนดโครงสร้างและขั้นตอนที่เหมาะสมในการสำรวจข้อมูล เพื่อให้ได้ข้อมูลที่ถูกต้องและน่าเชื่อถือ รวมถึงการกระจายและการจัดเก็บข้อมูล ซึ่งเป็นขั้นตอนที่สำคัญในการจัดระบบข้อมูลให้เหมาะสมและเข้าถึงได้ง่าย การงรู้จักกับวิธีการกระจายข้อมูลที่เหมาะสม เพื่อให้สามารถนำข้อมูลมาวิเคราะห์ได้อย่างเต็มประสิทธิภาพ นอกจากนี้ ยังศึกษาถึงการเลือกตัวอย่าง ซึ่งเป็นกระบวนการที่สำคัญในการทำนายหรือวิเคราะห์ทั่วไป เราได้เรียนรู้วิธีการเลือกตัวอย่างที่ใช้ความน่าจะเป็นและไม่ใช้ความน่าจะเป็น เพื่อให้ได้ข้อมูลที่แท้จริงและแทนความเป็นจริงของประชากรอย่างเหมาะสม

เอกสารอ้างอิง

ชัยณรงค์ เพียรภายลุนและคณะ. (2563). การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิด้วยการเลือกลำดับที่ของชุดตัวอย่างแบบครบวงจรเมื่อจัดลำดับแบบสมบูรณ์. *วารสารวิทยาศาสตร์บูรพา*. 25(3) 1026-1034

นภสร รัตนวุฒิขจร. (2565). *การจำลองข้อมูลเพื่อประเมินประสิทธิภาพของการเลือกตัวอย่างแบบมีระบบชนิดผสม*. วิทยานิพนธ์หลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

วราพรรณ ธิยานันท์และกฤษณะ ไวยมัย. (2023). การประยุกต์ใช้ความเปลี่ยนแปลงของข้อมูลเพื่อสุ่มลดข้อมูลสำหรับการแก้ปัญหาการจัดประเภทข้อมูลที่ไม่สมดุล. *วารสารงานวิจัยและพัฒนาเชิงประยุกต์โดยสมาคมECTI*. 3 (3) 29-38

วรางคณา เรียนสุทธิ์. (2562). การเปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบไม่อิงพารามิเตอร์ส าหรับการทดสอบภาวะความเท่ากันของความแปรปรวน. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี* .21(2พฤษภาคม-สิงหาคม2562). 163-170

บทที่ 3
การแสดงข้อมูล

ในบทนี้เป็นการแสดงข้อมูลเป็นขั้นตอนสำคัญในการสื่อสารข้อมูลและการนำเสนอผลลัพธ์จากการวิเคราะห์ข้อมูลให้ผู้อื่นเข้าใจได้ง่าย โดยมาตรฐานของการแสดงข้อมูลที่ดีรวมถึงการใช้มาตรวัดของข้อมูล เพื่อสรุปข้อมูลตัวเลขอย่างถูกต้อง การใช้กราฟและแผนภูมิ เช่น กราฟแท่ง เส้น หรือวงกลม เป็นวิธีที่ดีในการแสดงแนวโน้มและความสัมพันธ์ของข้อมูล นอกจากนี้การจัดเรียงข้อมูลในรูปแบบตารางช่วยให้ข้อมูลมีการจัดเรียงและอ่านออกมาได้ง่าย โดยระบุค่าตัวเลขและข้อความที่เกี่ยวข้อง สุดท้ายการใช้ภาษาที่เข้าใจง่ายและคำอธิบายที่ชัดเจนช่วยให้ข้อมูลและผลลัพธ์จากการวิเคราะห์เข้าใจได้อย่างถูกต้องและง่าย

มาตรวัดของข้อมูล

เนื่องจากตัวแปรที่ศึกษาจะมีระดับของการวัดที่แตกต่างกัน ดังนั้นจึงต้องมีกฎเกณฑ์ที่เป็นระบบเพื่อใช้ในการวัดข้อมูล กฎเกณฑ์ดังกล่าวเรียกว่า “มาตรวัดของข้อมูล”

มาตรวัดของข้อมูลมีองค์ประกอบที่สำคัญ 3 ประการ คือ

1. ขนาด

ขนาดหมายถึง ความมากหรือน้อย

2 ความเท่ากันของช่วงคะแนน

หมายถึง ความหมายของตัวเลขแต่ละช่วงมีความหมายเหมือนกัน หรือค่าของคะแนนแต่ละช่วงมีความเท่ากัน

3 ความมีศูนย์สมบูรณ์

หมายถึง ตัวแปรมีหรือไม่มีศูนย์สมบูรณ์ เช่น ตัวแปรอัตราดอกเบี้ย 0% บ่งบอกว่า ผลตอบแทนเป็นศูนย์ แสดงว่าตัวแปรอัตราดอกเบี้ยมีศูนย์

สมบูรณ์ ตัวแปรเกรดของนักศึกษา 4 3 2 1 และ 0 ถ้านักศึกษาได้เกรด 0 เกรด 0 หมายถึงเป็นศูนย์สมมติเพราะนักศึกษาอาจจะมีคะแนนแต่หากว่าคะแนนที่ได้ไม่ถึงเกณฑ์ จึงได้ 0

ดังนั้นระดับการวัดของตัวแปร (Level of Measurement) เป็นการแบ่งค่าของตัวแปรที่วัดมาได้ออกเป็นระดับเพื่อให้สามารถเลือกใช้สถิติได้อย่างถูกต้อง โดยแบ่งเป็น 4 ระดับดังนี้ (วิภาวรรณ เล้าอรุณ, 2557)

1. การวัดระดับนามมาตรา (Nominal Scale)
2. การวัดระดับอันดับมาตรา (Ordinal Scale)
3. การวัดระดับช่วงมาตรา (Interval Scale)
4. การวัดระดับอัตราส่วนมาตรา (Ratio Scale)

การวัดระดับนามมาตรา (Nominal Scale)

เป็นระดับต่ำสุดของตัวแปร ซึ่งจัดแบ่งเป็นประเภทหรือกลุ่มโดยระบุความ แตกต่างเป็นชื่อหรือสัญลักษณ์เท่านั้น มิได้บ่งชี้ถึงความแตกต่างในคุณค่า หรือคุณภาพใดๆ ไม่สามารถจัดเรียงลำดับก่อนหลัง หรือสูงต่ำได้ แต่ละหน่วยมีความเป็นอิสระจากกันอย่าง เด็ดขาด อาจใช้ตัวเลขแทนได้ (เพศ ชาย คือ 1 เพศหญิง คือ 2)

ตัวอย่าง

เพศ : ชาย หญิง

อาชีพ : ข้าราชการ รัฐวิสาหกิจ รับจ้าง อาชีพส่วนตัว เกษตรกร อื่นๆ

การวัดระดับอันดับมาตรา (Ordinal Scale)

เป็นระดับที่สูงกว่าระดับแบ่งกลุ่มมาตราหรือนามมาตรา สามารถแบ่งเป็นกลุ่มได้ ระบุความแตกต่างโดยให้เป็นชื่อหรือสัญลักษณ์ สามารถจัดเรียงลำดับได้ และแสดงความมากน้อยได้ แต่ไม่สามารถนำผลมา บวก ลบ คูณหารกันได้

ตัวอย่าง

ชั้นยศ : นายสิบ นายดาบ นายร้อย นายพัน นายพล

ความสวยของผู้หญิง : ความสวยมาก สวยปานกลาง ความสวยพอใช้ ไม่สวย

ทัศนคติ : เห็นด้วยอย่างยิ่ง เห็นด้วย เฉยๆ ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง

สถานะทางการเงิน : ฐานะร่ำรวย ฐานะปานกลาง ฐานะยากจน

การวัดระดับช่วงมาตรา (Interval Scale)

เป็นระดับที่ความละเอียดสูงกว่าระดับแบ่งกลุ่มมาตราและอันดับมาตรา มีความแตกต่างของแต่ละหน่วย และลำดับชั้นสูงต่ำมากน้อย รวมถึงความแตกต่างระหว่างหน่วยว่ามีเท่าใด (แต่ละช่วง ความแตกต่างเท่ากัน) สามารถนำมาบวก ลบ คูณ หารได้ แต่ไม่มีจุดศูนย์แท้ (หรือศูนย์สมมติ)

ตัวอย่าง

เกรดเฉลี่ยของนักศึกษา : A, B, C, D และ F (เกรด F ไม่ได้หมายความว่าคะแนนเท่ากับศูนย์ หากแต่คะแนนไม่ถึงเกณฑ์ที่กำหนด)

อุณหภูมิ : $0^{\circ}C$, $50^{\circ}C$ (อุณหภูมิ $0^{\circ}C$ ไม่ได้หมายความว่าไม่มีความเย็นเลย)

การวัดระดับอัตราส่วนมาตรา (Ratio Scale)

เป็นระดับความละเอียดสูงสุดของการวัดตัวแปรแสดงความแตกต่างของตัวแปรระดับแบ่งกลุ่มมาตรา ระดับอันดับมาตรา และระดับช่วงมาตรา โดยมีจุดศูนย์แท้ (Absolute Zero) หรือค่าศูนย์ ตามธรรมชาติ (Natural Zero) คือความไม่มีลักษณะนั้นๆ เลย เป็นข้อมูลตัวเลขนำมาบวก ลบ คูณ หารได้

ตัวอย่าง

อายุ: 10 ปี 20 ปี 33 ปี

ความยาว: 123 เซนติเมตร 145 เซนติเมตร 234 เซนติเมตร

น้ำหนัก: 48 กิโลกรัม 56 กิโลกรัม70กิโลกรัม

จำนวนบุตร: 3 คน 4 คน 7 คน

รายจ่ายต่อเดือน: 23,000 บาท 34,000 บาท 12,000 บาท

การนำเสนอข้อมูลในรูปแบบต่างๆ

การนำเสนอข้อมูลเป็นขั้นตอนต่อจากการเก็บรวบรวมข้อมูล ซึ่งได้กระทำการตรวจสอบความ ถูกต้องเรียบร้อยแล้ว เพื่อให้ผู้ที่นำข้อมูลเหล่านี้ไปใช้งานสะดวกและรวดเร็วยิ่งขึ้น การนำเสนอข้อมูล โดยทั่วไปกระทำได้ 2 ลักษณะใหญ่ๆ คือ การนำเสนอข้อมูลอย่างไม่เป็นแบบแผน (Informal Presentation) และการนำเสนอข้อมูลอย่างเป็นแบบแผน (Formal Presentation)

1. **การนำเสนอข้อมูลอย่างไม่เป็นแบบแผน**

เป็นการนำเสนอข้อมูลที่ไม่จำเป็นต้อง มีกฎเกณฑ์อะไรมากนัก ที่ใช้กันมี 2 วิธีคือ

1.1 การนำเสนอข้อมูลในรูปแบบข้อความ ก็คือ การนำข้อมูลที่เป็นตัวเลขมาเสนอเป็นส่วน หนึ่งของข้อความ การนำเสนอแบบนี้เพื่อให้ทราบรายละเอียดที่แน่นอนเหมาะสมสำหรับการนำเสนอ ข้อมูลที่ไม่ค่อยยาวมากและไม่ซับซ้อนนัก

ตัวอย่างที่ 3.1

“รายได้ต่อหัวประชากรไทยมีแนวโน้มสูงขึ้นจาก 15,138 บาทต่อหัวต่อปีในปี 2565 เพิ่มขึ้นเป็น 23,729 บาทต่อหัวต่อปีในปี 2566 และคาดว่าเป็น 29,887 บาทต่อปีในปี 2567”

1.2 การนำเสนอข้อมูลในรูปข้อความถึงตาราง เป็นการนำเสนอข้อมูลโดยแยกตัวเลขออก จากข้อความ เพื่อให้ความสะดวกในการเปรียบเทียบและเข้าใจการนำเสนอของข้อมูลได้ง่ายขึ้น

ตัวอย่างที่ 3.2 จำนวนคนไทยที่เสียชีวิตด้วยโรคต่างๆ จากการสูบบุหรี่ต่อปีในปี 2566

โรคหัวใจ	15,000	คน
มะเร็งปอด	10,878	คน
ถุงลมโป่งพอง	6,090	คน
เส้นเลือดในสมองตีบ-แตก	2,562	คน
เส้นเลือดอื่นตีบ-แตก	1,764	คน
โรคอื่นๆ	4,830	คน
รวมทั้งสิ้น	41,124	คน

2. การนำเสนอข้อมูลอย่างเป็นแบบแผน

เป็นการนำเสนอข้อมูลที่มีกฎเกณฑ์จะต้องปฏิบัติตามมาตรฐานที่กำหนดไว้เป็นแบบอย่างการนำเสนอข้อมูลโดยวิธีนี้ที่สำคัญ ได้แก่ การนำเสนอข้อมูลในรูปแบบตาราง การนำเสนอข้อมูลในรูปแผนภูมิและแผนภาพ การนำเสนอข้อมูลในรูปกราฟ

2.1 การนำเสนอข้อมูลในรูปตาราง การนำเสนอในวิธีการนี้เหมาะสมสำหรับข้อมูลที่มี รายละเอียดเป็นจำนวนมาก เพื่อให้ผู้ใช้งานสามารถหาข้อมูลตามที่ต้องการได้รวดเร็ว ช่วยให้สะดวกใน การเปรียบเทียบ การนำเสนอเป็นตารางมีรูปแบบการนำเสนอประกอบด้วย

1) หมายเลขตาราง
2) ชื่อเรื่อง
3) ต้นขั้ว
4) หัวข้อ
5) หัวเรื่อง
6) สดมภ์
7) ตัวเรื่อง
8) หมายเหตุ
9) แหล่งที่มา

หมายเลขตาราง.....ชื่อเรื่อง.....

หัวข้อ	หัวเรื่อง		
	สดมภ์	สดมภ์	สดมภ์
ต้นขั้ว	ตัวเรื่อง	ตัวเรื่อง	ตัวเรื่อง

หมายเหตุ (ถ้ามี) .....
แหล่งที่มา .....

ตารางที่ 1 จำนวนผลผลิตข้าวและข้าวในไร่นาย ก. ปี 2560 - 2566

(หน่วย : ตัน)

ปี	ข้าว	ข้าวโพด
2560	10.5	15.2
2561	11.0	18.5
2562	12.8	10.9
2563	25.1	11.8
2564	17.5	20.1
2565	21.0	25.0
2566	18.6	18.7

หมายเหตุ จำนวนข้าวเฉพาะข้าวนาปี
แหล่งที่มา ไร่ – นา ของนาย ก.

การนำเสนอข้อมูลในรูปของตาราง สามารถจำแนกลักษณะของตารางออกเป็น 4 ชนิด คือ

ก. ตารางแสดงความถี่ (Frequency table) คือ ตารางที่แสดงตัวเรื่องเป็นความถี่ของข้อมูลนั้น เช่น

ตารางที่ 2 แสดงผลรายได้รัฐวิสาหกิจที่นำส่งกระทรวงการคลัง (ตั้งแต่ ต.ค. 65 – พ.ค. 66)

รายการ	(หน่วย : ล้านบาท)
การไฟฟ้าฝ่ายผลิต	820.00
ธนาคารกรุงไทย	590.58
การไฟฟ้านครหลวง	550.00
โรงงานยาสูบ	500.00
สำนักงานสลากกินแบ่งรัฐบาล	313.60
บริษัทบางจากปิโตรเลียม	249.89
องค์การแก้ว	1.40

ที่มา : กระทรวงการคลัง

ข. ตารางทางเดียว (One – way table) คือ ตารางที่มีการจำแนกรายการบนหัวเรื่องหรือต้นขั้วเพียงด้านเดียว หรือจำแนกเพียงลักษณะเดียวเท่านั้น เช่น

ตารางที่ 3 แสดงปริมาณน้ำนมดิบ ปี 2565 – 2566

ปี	ปริมาณ (ตัน)
2561	155,574
2562	193,895
2563	227,784
2564	293,255
2565	326,381
2566*	350,196

ที่มา : สำนักงานเศรษฐกิจการเกษตร

หมายเหตุ : *คาดการณ์

ค. ตารางสองทาง (Two - way table) คือ ตารางที่มีการจำแนกรายการบนหัวเรื่องและต้นขั้ว ทั้งสองข้าง เช่น

ตารางที่ 4 แสดงจำนวนสมาชิกสภาผู้แทนราษฎรทั่วประเทศ

ปี 2565-66

พรรคการเมือง	จำนวนสมาชิกสภาผู้แทนราษฎร	
	13 ก.ย. 35	2 ก.ย. 38
ชาติไทยแลนด์	77	92
ประชาธิปก	79	86
ความหวังใหม่	51	57
ชาติพัฒนา	60	53
ก้าวหน้า	47	23
กิจการสังคม	22	22
นำไทย	-	18
เสรีธรรม	8	11
ประชากรไทย	3	18
เอกภาพ	8	8
มวลชน	4	3

ที่มา : กระทรวงมหาดไทย

ง. ตารางหลายทาง (Multi – way table) คือ ตารางที่มีการจำแนกรายการบนหัวเรื่องหรือต้นขั้วให้ย่อยออกไปจากตารางสองทาง เช่น

ตารางที่ 5 แสดงจำนวนนักศึกษาของวิทยาลัยแห่งหนึ่ง จำแนกตามระดับละชั้นปี

ปี	จำนวนนักศึกษา	
การศึกษา	ปวช.	ปวส.

	ปีที่ 1	ปีที่ 2	ปีที่ 3	ปีที่ 1	ปีที่ 2
2562	350	318	295	350	322
2563	380	354	326	375	347
2564	400	375	342	410	372
2565	420	382	366	480	453
2566	450	378	361	550	521

ที่มา : วิทยาลัยทับทอง

2.2 การนำเสนอข้อมูลในรูปแผนภูมิและแผนภาพ

การนำเสนอแบบนี้เพื่อให้ผู้ศึกษาหรือสามารถเปรียบเทียบของข้อมูลได้ง่ายขึ้นกว่า การ นำเสนอข้อมูลเป็นเชิงตัวเลขหรือการนำเสนอข้อมูลในรูปตารางซึ่งมีหลักการนำเสนอดังนี้

1. หมายเลขแผนภูมิหรือแผนภาพ (ถ้ามีหลายแผนภูมิหรือแผนภาพ)
2. ชื่อแผนภูมิหรือแผนภาพ
3. แหล่งที่มาของแผนภูมิหรือแผนภาพ

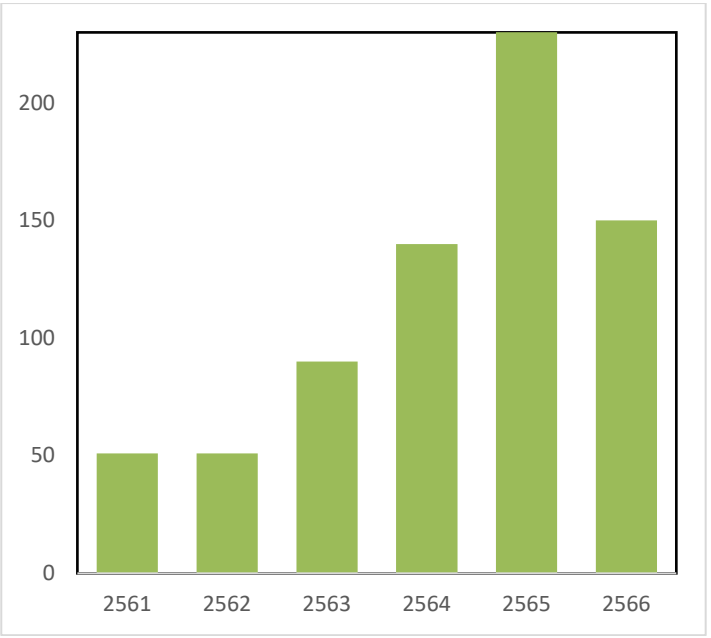
การนำเสนอข้อมูลด้วยแผนภูมิและแผนภาพ สามารถแบ่งออกเป็น 4 ประเภทดังนี้คือ แผนภูมิ แท่ง แผนภูมิกง แผนภูมิรูปภาพ และแผนที่สถิติ

ก. แผนภูมิแท่ง ประกอบด้วยแท่งสี่เหลี่ยมผืนผ้า อาจจะอยู่ในแนวตั้งหรือแนวนอนก็ได้และ เราเรียกแท่งสี่เหลี่ยมผืนผ้าว่า แท่ง (bar) ความสูงของแต่ละแท่งจะต้องได้สัดส่วนกับจำนวน หรือขนาดของข้อมูลแผนภูมิแท่งจำแนกเป็นประเภทต่างๆได้ดังนี้

1. แผนภูมิแท่งเชิงเดียว ใช้แสดงการเรียบเทียบลักษณะข้อมูลที่น่าสนใจเพียงลักษณะ เดียว เช่น จำนวนนักศึกษา จำนวนเงิน มูลค่าการนำเข้าของรถยนต์ไฟฟ้า ฯลฯ

ตัวอย่าง แผนภูมิแสดงการเปรียบเทียบยอดเงินฝากออมทรัพย์ตั้งแต่ปี 2561-2566

ยอดเงินฝาก (บาท)



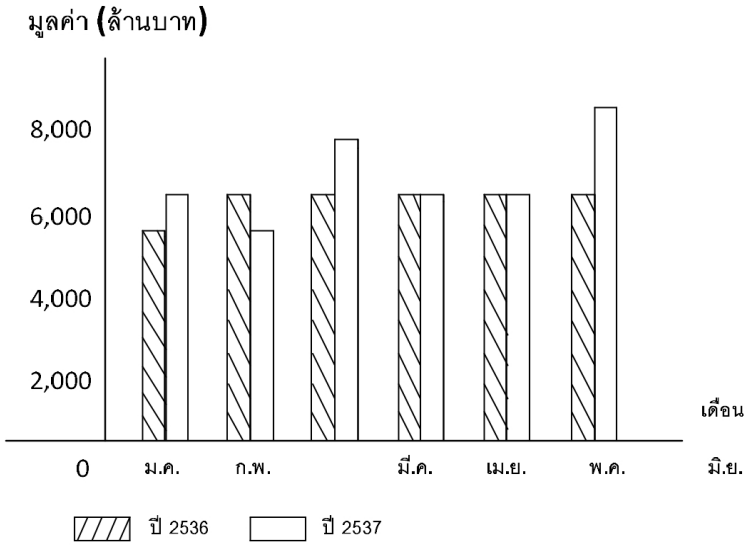
ที่มา : รายงานกิจการปี 2566 สหกรณ์ออมทรัพย์

2. แผนภูมิแท่งเชิงซ้อน ใช้แสดงการเปรียบเทียบลักษณะข้อมูลที่น่าสนใจ ตั้งแต่สอง ลักษณะหรือสองช่วงเวลาขึ้นไป

ตัวอย่าง แผนภูมิ 2 แสดงการเปรียบเทียบการส่งออกเสื้อผ้าสำเร็จรูปปี 2536 – 2537

ช่วงเดือน ม.ค. – มิ.ย.

ตัวอย่าง แผนภูมิ 2 แสดงการเปรียบเทียบการส่งออกเสื้อผ้าสำเร็จรูป 2536-37 ช่วงเดือน ม.ค. - มิ.ย.

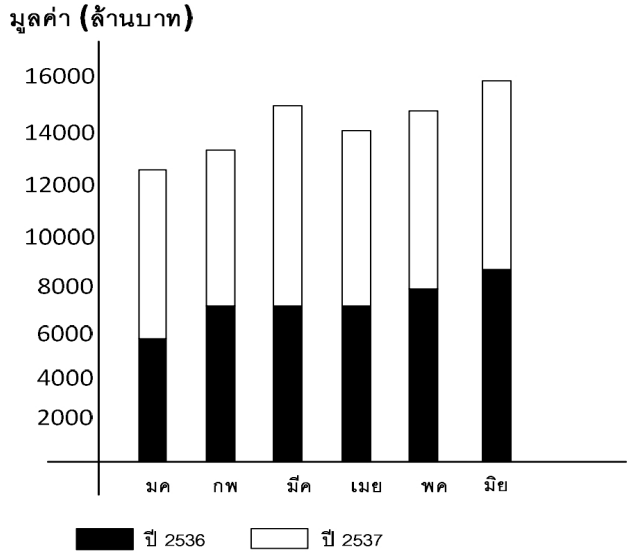


ที่มา : กรมเศรษฐกิจการพาณิชย์

3. **แผนภูมิแท่งส่วนประกอบ** ใช้แสดงรายละเอียดส่วนย่อยของข้อมูลชุดเดียวกันที่จะนำเสนอ

ตัวอย่าง แผนภูมิ 3 แสดงการเปรียบเทียบการส่งออกเสื้อผ้าสำเร็จรูปปี 2536 – 2537

ช่วงเดือน ม.ค. – มิ.ย.

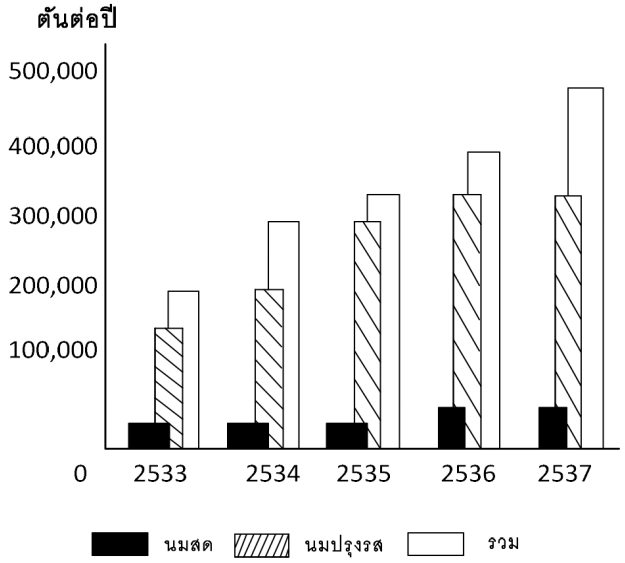


ที่มา : กรมเศรษฐกิจการพาณิชย์

4. แผนภูมิแท่งซ้อนกัน ใช้แสดงการเปรียบเทียบระหว่างแท่งสี่เหลี่ยมที่อยู่ในลักษณะ ซ้อนกันหลายๆแท่ง เพื่อให้ดูชัดเจนขึ้น

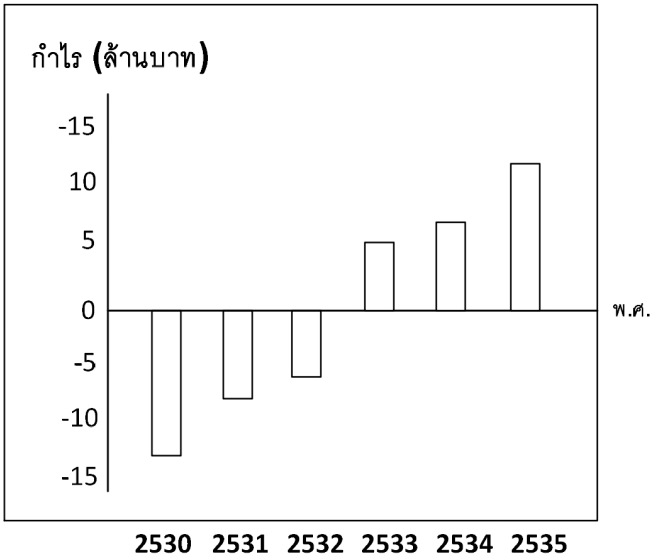
ตัวอย่าง แผนภูมิ 4 แสดงการเปรียบเทียบผลิตภัณฑ์นมพร้อมดื่มที่โรงงานต่างๆ ผลิตได้ต่อปี

ตั้งแต่ปี 2533 – 2537



5. แผนภูมิแท่งบวก – ลบ ใช้แสดงการเปรียบเทียบที่มีค่าเป็นไปได้ทั้งค่าบวกและค่าลบใช้แสดงกำไร – ขาดทุน

ตัวอย่าง แผนภูมิ 5 แสดงกำไร – ขาดทุนของบริษัท ตั้งแต่ปี พ.ศ. 2530 – 2535



ข. แผนภูมิกง ใช้แสดงการเปรียบเทียบรายละเอียดของข้อมูลชุดเดียวกัน แสดงด้วยรูป วงกลม โดย
แบ่งรูปวงกลมออกเป็นส่วนต่างๆ ที่จุดส่วนกลางตามขนาดของข้อมูล นั่นคือ

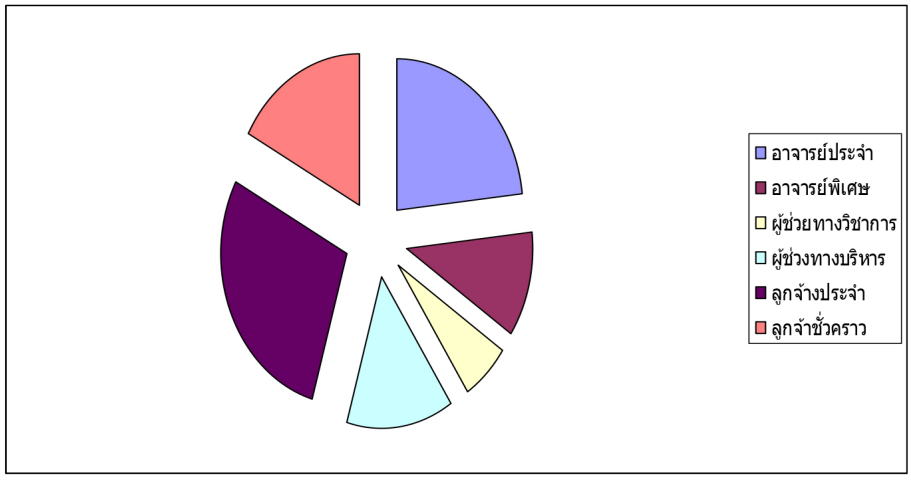
$$\text{ขนาดของข้อมูล} = \frac{\text{ข้อมูลชนิดนั้น X360}}{\text{ข้อมูลทั้งหมด}}$$

$$\text{ร้อยละของข้อมูล} = \frac{\text{ข้อมูลชนิดนั้น X100}}{\text{ข้อมูลทั้งหมด}}$$

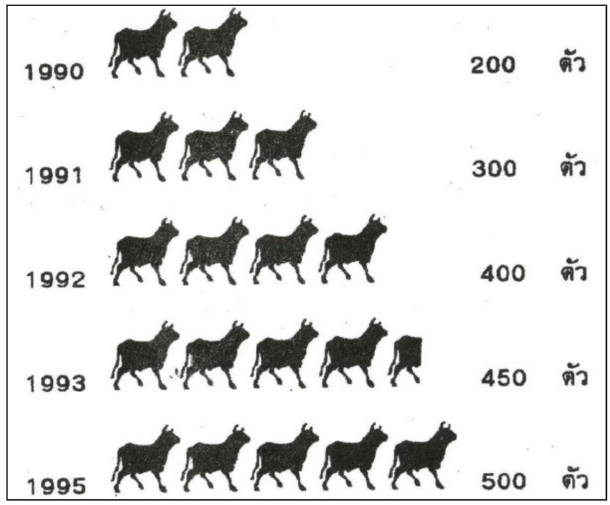
โดยทั่วไปจะเขียนแสดงว่าแต่ละส่วนคิดเป็นเปอร์เซ็นต์ของข้อมูลทั้งหมด

ตัวอย่าง ตารางแสดงการเปรียบเทียบบุคลากรของมหาวิทยาลัยเกษตรศาสตร์ มีจำนวนทั้งสิ้น 6,600 คน ในปี 2566

บุคลากร	จำนวน คน	ร้อยละ	องศา
อาจารย์ประจำ	1,542	$\frac{1542}{6600} \times 100 = 23.4$	$\frac{1542}{6600} \times 360 = 84$
อาจารย์พิเศษ	738	$\frac{738}{6600} \times 100 = 11.2$	$\frac{738}{6600} \times 360 = 40$
ผู้ช่วยทางวิชาการ	414	$\frac{414}{6600} \times 100 = 6.3$	$\frac{414}{6600} \times 360 = 23$
ผู้ช่วงทางบริหาร	896	$\frac{896}{6600} \times 100 = 13.6$	$\frac{896}{6600} \times 360 = 49$
ลูกจ้างประจำ	1,865	$\frac{1865}{6600} \times 100 = 28.3$	$\frac{1896}{6600} \times 360 = 102$
ลูกจ้าชั่วคราว	1,145	$\frac{1145}{6600} \times 100 = 17.2$	$\frac{1145}{6600} \times 360 = 62$



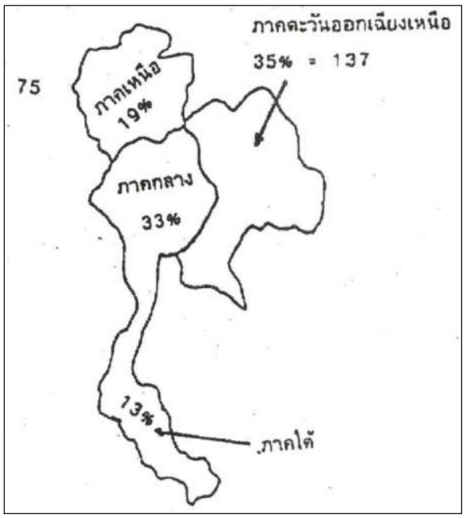
ค. **แผนภูมิรูปภาพ** แผนภูมิที่ใช้รูปภาพแทนค่าเลขจำนวนของข้อมูลที่นำมาเสนอ เช่น ภาพวัว 1 ตัว จำนวนวัวในข้อมูล 100 ตัว



แผนภูมิ แสดงจำนวนวัวในปศุสัตว์ นายวัฒนา ตั้งแต่ปี 1990 – 1995

ง. แผนที่สถิติเป็นแผนภูมิที่นำภูมิเสนอข้อมูลโดยอาศัยหลักการทางภูมิศาสตร์ เพื่อทำให้ การเปรียบเทียบข้อมูลที่อยู่ในพื้นที่ทางภูมิศาสตร์เป็นไปได้โดยง่าย และรวดเร็ว เช่น

แผนที่สถิติ แสดงการเปรียบเทียบจำนวน ส.ส. ในภาคต่างๆ ของไทย



ภาคตะวันออกเฉียงเหนือ	137 คน
ภาคกลาง	128 คน
ภาคเหนือ	75 คน
ภาคใต้	51 คน

ที่มา : มติชนสุดวัปดาห์

2.3 การนำเสนอข้อมูลในรูปกราฟเส้น

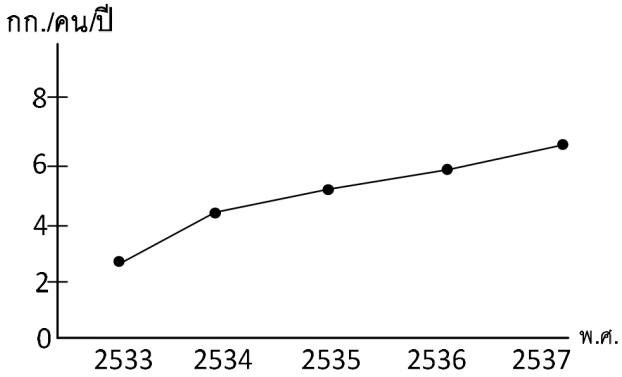
การนำเสนอข้อมูลในรูปกราฟเส้น มักนิยมใช้กับข้อมูลอนุกรมเวลา เป็นข้อมูลที่แสดงการ เปลี่ยนแปลงตามลำดับก่อนหลัง และเกิดขึ้นเป็นช่วงๆ ของเวลาหลายๆ ช่วง

การนำเสนอข้อมูลด้วยวิธีสามารถเห็นลักษณะข้อมูลได้ชัดเจน และรวดเร็วทั้งช่วยพยากรณ์
ข้อมูลที่ต้องการกราฟในอนาคตได้

การนำเสนอข้อมูลในรูปกราฟเส้นมี 4 ชนิดคือ กราฟเส้นเชิงเดียว กราฟเส้นเชิงซ้อน กราฟ เส้นเชิงประกอบ และกราฟดุล

1. **กราฟเส้นเชิงเดียว** ใช้แสดงการเปรียบเทียบลักษณะของข้อมูลที่สนใจศึกษาเพียงลักษณะ
   เดียว

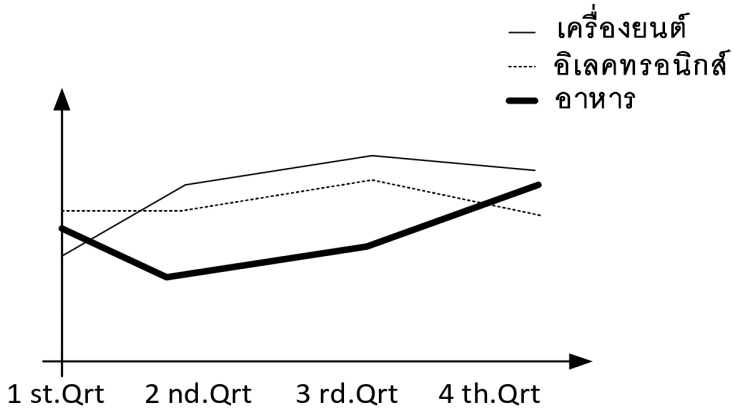
ตัวอย่าง กราฟแสดงอัตราการบริโภคนมพร้อมดื่มของประชากรไทย ตั้งแต่ปี 2533 – 2537



ที่มา : ศูนย์ข้อมูลเศรษฐกิจ

2. กราฟเส้นเชิงซ้อน ใช้แสดงการเปรียบเทียบลักษณะของข้อมูลที่สนใจศึกษาตั้งแต่ 2 ลักษณะขึ้นไป สามารถเปรียบเทียบได้ทั้งข้อมูลในแต่ละลักษณะของช่วงเวลาที่ต่างกัน และ ข้อมูลที่อยู่ในช่วงเดียวกันแต่ลักษณะต่างกัน ตลอดจนแนวโน้มของข้อมูลแต่ละลักษณะอีกด้วย

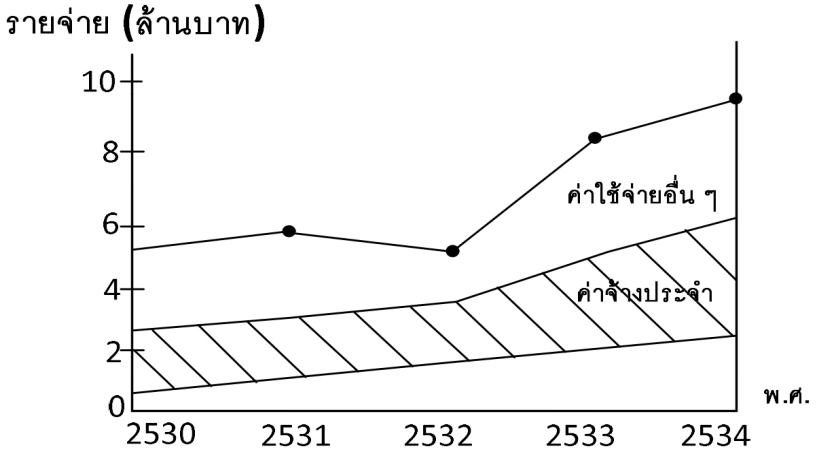
ตัวอย่าง กราฟยอดจำหน่ายสินค้า จำแนกตามหมวดสินค้าและไตรมาศ



แหล่งที่มา ห้าง DKK
หมายเหตุ ข้อมูลปี 2563

3. **กราฟเส้นเชิงประกอบ** ใช้แสดงการเปรียบเทียบรายละเอียดหรือส่วนย่อยของข้อมูลใน ช่วงเวลาต่างๆ กัน

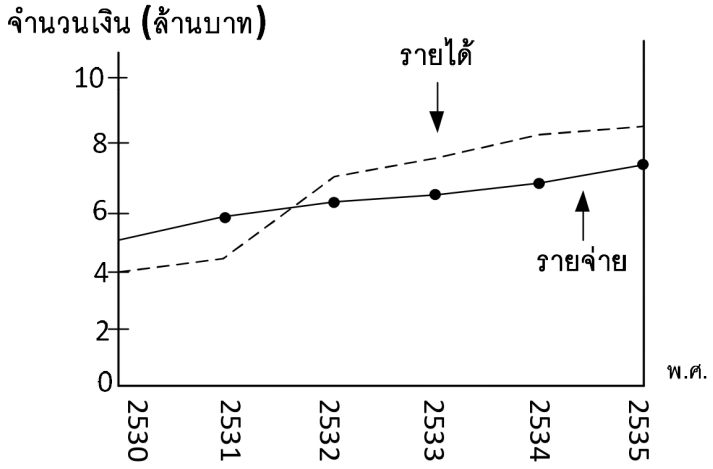
ตัวอย่าง กราฟแสดงการเปรียบเทียบรายจ่ายตามประเภทค่าใช้จ่ายของบริษัทสหซีฟูด ตั้งแต่ปี พ.ศ. 2530 -2534



ที่มา : แผนกบัญชีของบริษัทสหซีฟูด

4. **กราฟดุล** แสดงให้เห็นถึงความแตกต่างระหว่างลักษณะของข้อมูลสองลักษณะที่มีความสัมพันธ์กัน

ตัวอย่าง กราฟแสดงเปรียบเทียบรายได้และรายจ่ายของบริษัทรุ่งเรืองชัย ตั้งแต่ปี พ.ศ. 2530 – 2535



ที่มา : งบดุลของบริษัทรุ่งเรืองชัย

3.3 การแจกแจงความถี่

ข้อมูลที่ได้จากการเก็บรวบรวมวิธีต่างๆ อาจจะอยู่ในลักษณะที่ไม่เป็นระเบียบ จึงจำเป็นต้องจัด ข้อมูลดังกล่าวให้เป็นระเบียบและเป็นหมวดหมู่ เพื่อให้ง่ายต่อการคำนวณและสะดวกต่อการวิเคราะห์ ข้อมูลขั้นต่อไป เรียกวิธีทางสถิติเช่นนี้ว่า การแจกแจงความถี่ของข้อมูล (สรชัย พิศาลบุตร, 2559; สำนักการเงินการคลัง,2556)

ความถี่ (Frequency) ของค่าจากการสังเกตคือ จำนวนครั้งของค่าจากการสังเกตในข้อมูลกลุ่มหนึ่ง

การหาความถี่นิยมใช้วิธีทำรอยขีด (tally) เช่น | แทนความถี่ 1 และ ||| แทนความถี่ 5 เป็นต้น แล้วจึงสร้างตารางแจกแจงความถี่

การแจกแจงความถี่ของข้อมูลทำได้ 2 วิธี คือ

1. การแจกแจงความถี่โดยไม่จัดข้อมูลเป็นหมวดหมู่ (Ungrouped data)

2. การแจกแจงความถี่โดยจัดข้อมูลเป็นหมวดหมู่ (Grouped data)

1. การแจกแจงความถี่โดยไม่จัดข้อมูลเป็นหมวดหมู่

เป็นการนำข้อมูลทั้งหมดมาเรียงตามลำดับ จากค่ามากไปน้อย หรือน้อยไปมาก ใช้กับข้อมูลจำนวนไม่มากนัก

ตัวอย่างที่ 3.3 คะแนนสอบวิชาสถิติของนักศึกษา 30 คน ได้ดังนี้

18	17	15	21	19	22	20	15	18	16
14	15	16	17	20	21	22	18	19	15
17	19	20	21	20	22	16	15	18	17

สร้างตารางแจกแจงความถี่ได้ดังนี้

คะแนน	14	15	16	17	18	19	20	21	22
รอยขีด	|	||||	|||	||||	||||	|||	||||	|||	|||
ความถี่	1	5	3	4	4	3	4	3	3

2. การแจกแจงความถี่โดยจัดข้อมูลเป็นหมวดหมู่

ในบางคราวข้อมูลมีเป็นจำนวนมาก ถ้านำข้อมูลทำการแจกแจงความถี่โดยไม่จัดหมวดหมู่ อาจจะทำให้ตารางมีขนาดใหญ่คงจะไม่สะดวกและไม่มีประโยชน์มากนัก ดังนั้นจึงแบ่งข้อมูล ออกเป็นช่วง ซึ่งเรียกว่าอันตรภาคชั้น (Class interval) แล้วหารอยขีด จากนั้นนับรอยขีดรวมเป็นความถี่

ตัวอย่างที่ 3.4 จากตัวอย่างที่ 1.3 โดยจัดข้อมูลเป็นหมวดหมู่

คะแนน	ความถี่
10 -14	1
15-19	19
20-24	10

คำศัพท์ที่เกี่ยวกับตารางแจกแจงความถี่

1. **ขีดจำกัดล่างของอันตรภาพชั้น (Lower limit)** คือ ค่าน้อยที่สุดของแต่ละอันตรภาพชั้นนั้น เช่น 10 – 14, 15 – 19, 20 - 24 ขีดจำกัดล่างของอันตรภาคชั้นคือ 10, 15 และ 20 ตามลำาดับ
2. **ขีดจำกัดบนของอันตรภาพชั้น (Upper limit)** คือ ค่ามากที่สุดของแต่ละอันตรภาพชั้นนั้น เช่น 10 – 14, 15 – 19, 20 - 24 ขีดจำกัดบนของอันตรภาคชั้นคือ 14, 19 และ 24 ตามลำดับ
3. **ขอบล่างของอันตรภาคชั้น (Lower boundary)** คือ ค่ากึ่งกลางระหว่างขีดจำกัดล่างของ อันตรภาคชั้นกับขีดจำกัดบนของอันตรภาคชั้นต่ำกว่าที่อยู่ติดกัน เช่น

ขอบล่างของอันตรภาคชั้น 15 – 19 คือ $\frac{15 + 14}{2} = 14.5$

4. **ขอบบนของอันตรภาคชั้น (Upper boundary)** คือ ค่ากึ่งกลางระหว่างขีดจำกัดบนของ อันตรภาคชั้นกับขีดจำกัดล่างของอันตรภาคชั้นสูงกว่าที่อยู่ติดกัน เช่น

ขอบบนของอันตรภาคชั้น 15 – 19 คือ $\frac{19 + 20}{2} = 19.5$

จุดกึ่งกลางของอันตรภาคชั้น (Midpoint) คือ ค่ากึ่งกลางระหว่างขอบล่างกับขอบบนของอันตรภาคชั้นนั้น เช่น

ขอบกึ่งกลางของอันตรภาคชั้น 15 – 19 คือ $\frac{14.5 + 19.5}{2} = 17$

หรือ $\frac{15 + 19}{2} = 17$

5. ความกว้างของอันตรภาคชั้น (Class interval) คือ ผลต่างระหว่างขอบบนกับขอบล่าง ของแต่ละอันตรภาคชั้น เขียนแทนด้วย | หรือ c เช่น

ความกว้างของอันตรภาคชั้น 10 – 14 คือ 14.5 – 95 = 5

" " 15 – 19คือ 19.5 – 14.5 = 5

" " 20 – 24คือ 24.5 – 19.5 = 5

ขั้นตอนในการสร้างตารางแจกแจงความถี่ของข้อมูลที่เป็นหมวดหมู่

1. กำหนดจำนวนอันตรภาคชั้น พอสมควรประมาณ 5 – 20 ชั้น

2. กำหนดความกว้างของอันตรภาคชั้นโดยประมาณได้จากสูตร

ความกว้างของอันตรภาคชั้น (I) = (ค่าสูงสุด – ค่าต่ำสุด)/จำนวนอันตรชั้น

3. เขียนอันตรภาคชั้นแต่ละอันตรภาคชั้น เรียงตามลำดับ โดยยึดหลักว่า ค่าต่ำสุดต้องอยู่ชั้นต่ำสุด และค่าสูงสุด ต้องอยู่ในอันตรภาคชั้นสูงสุด

ตัวอย่างที่ 3.5 จงสร้างตารางแจกแจงความถี่ของข้อมูลต่อไปนี้

ข้อมูลการสูญเสียของค่าเชื้อเพลิงของรถยนต์ขนาดเล็กที่ติดในเวลา 16.00 – 19.00 น. ของ แต่ละวัน โดยศึกษาข้อมูลจากผู้ใช้รถยนต์จำนวน 85 คัน

(หน่วย : บาท)

85	86	74	104	69	56	61	74	101	74
84	98	80	64	73	69	72	103	90	79
98	104	93	83	75	94	96	92	77	93
88	78	84	87	71	90	76	84	74	92
100	79	57	104	64	83	96	79	63	87
85	69	83	86	86	80	83	94	97	81
92	71	81	64	96	91	68	84	97	64
79	85	69	88	97	78	85	69	92	100
60	65	72	84	90					

จงสร้างตารางแจกแจงความถี่ โดยให้มีจำนวนอันตรภาคชั้นเท่ากับ 7 ชั้น

วิธีทำ กำหนดอันตรภาคชั้นจำนวน 7 ชั้น

ความกว้างของอันตรภาคชั้น = $\frac{104 + 56}{2} = 6.857$ ประมาณ 7

ดังนั้นตารางแจกแจงความถี่ของข้อมูล จะมีลักษณะดังนี้

อัตราค่าสูญเสีย (บาท)	รอยขีด	ความถี่

56 – 62	||||	4
63 – 69	|||| |||| ||	12
70 – 76	|||| |||| |	11
77 – 83	|||| |||| ||||	15
84 – 90	|||| |||| |||| ||||	19
91 – 97	|||| |||| ||||	15
98 – 104	|||| ||||	9

หมายเหตุ ความกว้างของอันตรภาคชั้นออกมาเป็นทศนิยม แต่ค่าสังเกตเป็นจำนวนเต็มโดยให้ปัดเศษขึ้นเสมอ

ตัวอย่างที่ 3.6 จงสร้างตารางแจกแจงความถี่ของเกรดเฉลี่ยของนักศึกษา ปวช. 1 จำนวน 30 คน ซึ่งมีเกรดเฉลี่ยดังนี้

3.21	.26	3.93	2.35	2.67
2.58	1.98	3.69	3.40	2.36
3.33	2.24	2.91	3.23	3.37
3.42	3.64	3.26	2.96	1.81
2.41	2.10	2.15	2.94	3.15
1.92	2.81	2.49	2.76	3.36

กำหนดให้มีจำนวนอันตรภาคชั้นทั้งหมด 10 ชั้น

วิธีทำ กำหนดจำนวนอันตรภาคชั้นทั้งหมด 10 ชั้น

ความกว้างของอันตรภาคชั้น = $\frac{3.93 + 1.81}{10} = \frac{2.12}{10} = 0.21 \approx 0.22$

อัตราค่าสูญเสีย (บาท)	รอยขีด	ความถี่
1.76 - 1.97	||	2
1.98 - 2.19	|||	3
2.20 - 2.41	||||	4
2.42 - 2.63	||	2
2.64 - 2.85	|||	3

2.85 - 3.07	|||	3
3.08 - 3.29	|||||	5
3.30 - 3.51	|||||	5
3.52 - 3.73	||	2
3.74 - 3.95			1

หมายเหตุ ความกว้างของอันตรภาคชั้นเป็นทศนิยมและค่าสังเกตเป็นจำนวนเต็มให้ปัดเศษ

การแจกแจงความถี่สัมพัทธ์

ความถี่สัมพัทธ์ (Relative frequency) ของอันตรภาคชั้นใดคืออัตราส่วนระหว่างความถี่ ของอันตรภาคชั้นนั้นกับผลรวมของความถี่ทั้งหมด

ความถี่สัมพัทธ์อาจแสดงในรูปเศษส่วนหรือทศนิยมหรืออาจแสดงอยู่ในรูปร้อยละ เรียกว่า “ร้อยละของความถี่สัมพัทธ์” นั่นคือ

$$\text{ความถี่สัมพัทธ์} = \frac{\text{ความถี่ของอันตรภาคชั้นนั้น}}{\text{ผลรวมความถี่ทั้งหมด}}$$

$$\text{ร้อยละของความถี่สัมพัทธ์} = \frac{\text{ความถี่ของอันตรภาคชั้นนั้น X100}}{\text{ผลรวมความถี่ทั้งหมด}}$$

ข้อสังเกต

1. ผลรวมของความถี่สัมพัทธ์เท่ากับ 1
2. ผลรวมของร้อยละความถี่สัมพัทธ์เท่ากับ 100

ตัวอย่างที่ 3.7 จากตารางแจกแจงความถี่ที่กำหนดให้ จงหาความถี่สัมพัทธ์และร้อยละความถี่สัมพัทธ์

อันตรภาคชั้น	ความถี่	ความถี่สัมพัทธ์	ร้อยละความถี่สัมพัทธ์
55 – 59	2	= 0.05	5

60 – 64	14	= 0.35	35
65 – 69	8	= 0.20	20
70 – 74	12	= 0.30	30
75 - 79	4	= 0.10	10
รวม	40	1.0	100

การแจกแจงความถี่สะสม

ความถี่สะสม (Cumulative Frequency) ของอันตรภาคชั้นใด คือ ผลรวมของความถี่ของ อันตรภาคชั้นนั้นกับความถี่ของอันตรภาคชั้นที่ต่ำกว่าทั้งหมดหรือสูงกว่าทั้งหมดอย่างใดอย่างหนึ่ง

ตัวอย่างที่ 3.8 จากตารางแสดงความถี่กำหนด จงหาความถี่สะสม

อันตรภาคชั้น	ความถี่	ความถี่สะสม จากน้อยไปมาก	ความถี่สะสม จากมากไปน้อย
20 – 29	2	2	50
30 – 39	5	7	48
40 – 49	10	17	43
50 – 59	15	22	33
60 – 69	11	43	18
80 – 89	7	50	7
รวม	50		

โดยทั่วไปนิยม ความถี่สะสมจากน้อยไปมาก

การแจกแจงความถี่สะสมสัมพัทธ์

ความถี่สะสมสัมพัทธ์ (Relative Cumulative Frequency) ของอันตรภาคชั้นใดคือ อัตราส่วน ระหว่างความถี่สะสมของอันตรภาคชั้นนั้นกับผลรวมของความถี่ทั้งหมด

ความถี่สะสมสัมพัทธ์อาจแสดงในรูปเศษส่วนหรือทศนิยม หรือบางครั้งอาจจะแสดงอยู่ในรูป
ร้อยละ เรียกว่า “ร้อยละของความถี่สะสมสัมพัทธ์” นั่นคือ

$$\text{ความถี่สัมพัทธ์} = \frac{\text{ความถี่สะสมสัมพัทธ์}}{\text{ผลรวมความถี่}}$$

$$\text{ร้อยละของความถี่สัมพัทธ์} = \frac{\text{ความถี่สะสม } X100}{\text{ผลรวมความถี่}}$$

ตัวอย่างที่ 3.9 จากตารางแจกแจงความถี่ของคะแนนสอบวิชาคณิตศาสตร์จำนวน 80 คน เป็นดังนี้ จงสร้างตารางแจกแจงความถี่สะสม ความถี่สะสมสัมพัทธ์ และร้อยละความถี่สะสมสัมพัทธ์

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์	ร้อยละความถี่สัมพัทธ์
40 – 49	5	5	= 0.06	6
50 – 59	13	18	= 0.22	22
60 – 69	32	50	= 0.62	62
70 – 79	24	74	= 0.92	92
80 – 89	4	78	= 0.98	98
90 – 99	2	80	= 1.00	100
รวม	80			

ตัวอย่างที่ 3.10 จากตารางแจกแจงความถี่ที่กำหนดให้แสดงส่วนสูงของนักศึกษาปีที่ 1 ของวิทยาลัย แห่งหนึ่ง จำนวน 30 คน

ส่วนสูง (ซม.)	ความถี่
158 – 162	2
163 – 167	8
168 – 172	10
173 – 177	7
178 – 182	3

จงหา

1. นักศึกษาที่มีส่วนสูง 166 ซม. ตกอยู่ในอันตรภาคชั้นใด
2. มีนักศึกษาจำนวนเท่าใดที่มีส่วนสูงต่ำกว่า 168 ซม.
3. มีจำนวนนักศึกษาร้อยละเท่าไรที่มีส่วนสูงสูงกว่า 172 ซม.
4. มีนักศึกษาจำนวนเท่าใดที่มีส่วนสูงตั้งแต่ 163 ซม. ขึ้นไป แต่ไม่เกิน 177 ซม.

วิธีทำ 1. นักศึกษาที่มีส่วนสูง 166 ซม. อยู่ในช่วง 163.5 – 167.5

2. มีนักศึกษาจำนวน 2 + 8 = 10 คน ที่มีส่วนสูงต่ำกว่า 168 ซม.

3. มีนักศึกษาที่มีส่วนสูงสูงกว่า 172 ซม. อยู่ 7 + 3 = 10 คน

$$\text{คิดเป็นร้อยละ } \frac{100}{30} \times 100 = 33.33\%$$

4. มีนักศึกษาจำนวน 8 + 10 + 7 = 25 คน ที่มีส่วนสูงตั้งแต่ 163 ซม. ขึ้นไป แต่ไม่เกิน 177 ซ.ม.

การแจกแจงความถี่โดยใช้กราฟ

กราฟที่แสดงการแจกแจงความถี่ มี 3 แบบ คือ

1. ฮิสโตแกรม (Histogram)
2. รูปหลายเหลี่ยมความถี่ (Frequency Polygon)
3. เส้นโค้งความถี่ (Frequency Curve)

1. ฮิสโตแกรม เป็นกราฟแท่งรูปสี่เหลี่ยมมุมฉากวางติดกัน โดยที่ความกว้างของรูปสี่เหลี่ยม มุมฉากแทนควากว้างของอันตรภาคชั้น และพื้นที่ของรูปสี่เหลี่ยมกับมุมฉากแต่ละรูปแทนความถี่ของแต่
ละอันตรภาคชั้น

ถ้าแต่ละอันตรภาคชั้นมีความกว้างเท่ากัน ความสูงของรูปสี่เหลี่ยมมุมฉากจะแปรผันตาม ความถี่ จึงใช้ความถี่เป็นส่วนสูงของรูปสี่เหลี่ยมมุมฉากได้

วิธีการเขียนรูปฮิสโตแกรม อาจใช้ขอบล่างหรือขอบบน หรือจุดกึ่งกลางของอันตรภาคชั้นก็ได้

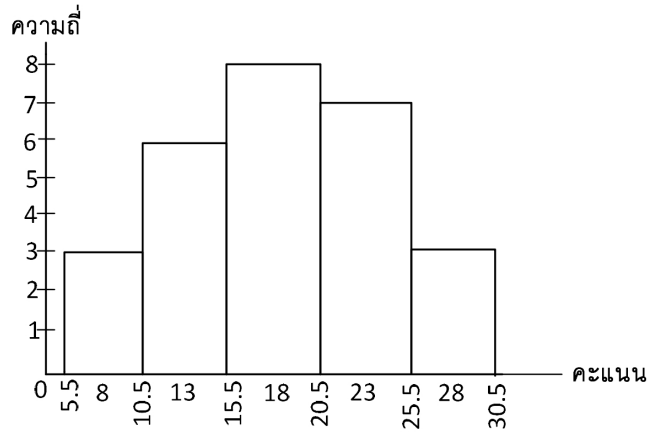
ตัวอย่างที่ 3.11 จงสร้างฮิตโตแกรมจากตารางการแจกแจงความถี่ต่อไปนี้

อันตรภาคชั้น	ความถี่
6 – 10	3
11 – 15	6
16 – 20	8
21 – 25	7
26 – 30	3

วิธีทำ

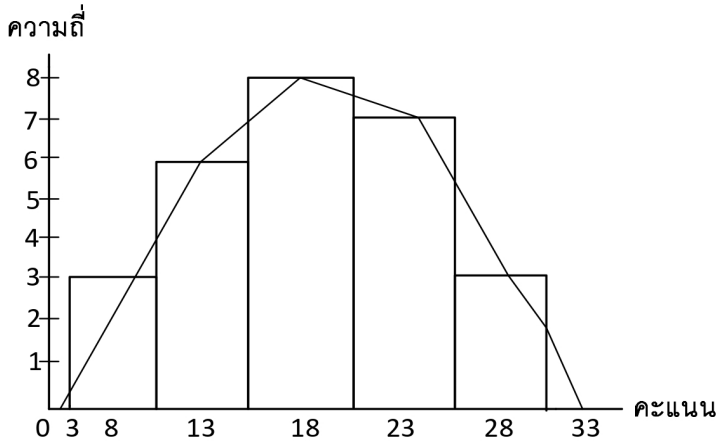
อันตรภาคชั้น	ความถี่	ขอบล่าง – ขอบบน	จุกกึ่งกลางชั้น
6 – 10	3	5.5 – 10.5	8
11 – 15	6	10.5 – 15.5	13
16 – 20	8	15.5 -20.5	18
21 – 25	7	20.5 – 25.5	23
26 – 30\	3	25.5 – 30.5	28

สร้างฮิสโตแกรมได้ดังนี้



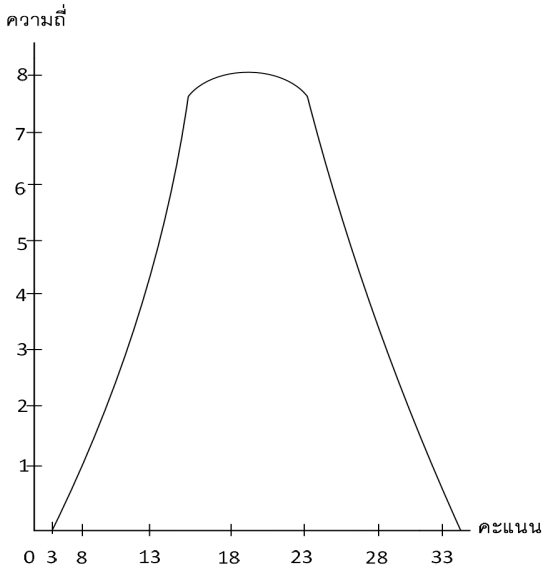
2. รูปหลายเหลี่ยมความถี่ คือ รูปสี่เหลี่ยมที่เกิดจากการต่อจุดกึ่งกลางของยอดของแท่ง สี่เหลี่ยมมุมฉากของฮิสโตแกรมด้วยเส้นของเส้นตรง

ตัวอย่างที่ 3.12 จงสร้างรูปหลายเหลี่ยมความถี่จากตารางการแจกแจงความถี่ในตัวอย่างที่ 3.11



3. เส้นโค้งความถี่ คือ เส้นโค้งที่ได้จากการปรับด้านของรูปหลายเหลี่ยมความถี่ให้เรียบขึ้น โดยการปรับ พื้นที่ใต้เส้นโค้งจะมีพื้นที่ใกล้เคียงกับพื้นที่ของรูปหลายเหลี่ยมความถี่

ตัวอย่างที่ 1.13 จงเขียนเส้นโค้งความถี่จากตารางแจกแจงความถี่ในตัวอย่างที่ 1.11

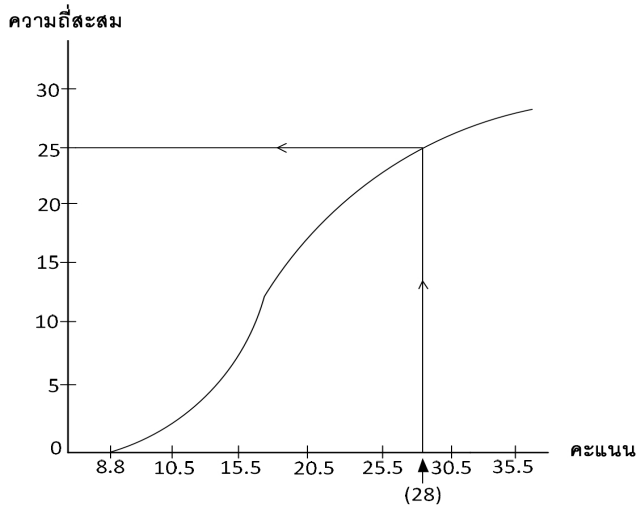


การแจกแจงความถี่สะสมโดยกราฟ

การแสดงความถี่สะสม เรียกว่า “เส้นโค้งความถี่สะสม” (Ogive Curve) เส้นโค้งความถี่สะสม เป็นกราฟแสดงความสัมพันธ์ระหว่างขอบบนของอันตรภาคชั้นกับความถี่สะสมแล้วโยงจุดนั้นด้วย เส้นตรงแล้วปรับให้เป็นเส้นโค้งเรียบ

ตัวอย่างที่ 3.13 จงสร้างเส้นโค้งความถี่สะสมจากตารางแจกแจงความถี่ที่กำหนดให้

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ขอบบน
6 – 10	3	3	10.5
11 – 15	6	9	15.5
16 – 20	8	17	20.5
21 – 25	7	24	25.5
26 – 30\	3	27	30.5
31 – 35	3	30	35.5



จากกราฟเส้นโค้งความถี่สะสม ใช้ในการหาว่ามีจำนวนข้อมูลที่มีค่าต่ำกว่าหรือสูงกว่าค่าที่กำหนด มีจำนวนมากน้อยเท่าไร เช่น อยากทราบว่ามีข้อมูลที่มีค่าต่ำกว่า 28 คะแนน อยู่จำนวนเท่าใด ก็ สามารถหาได้จากเส้นโค้งความถี่สะสม โดยลากส่วนของเส้นตรงจากแกนนอนตรงจุดที่มีค่า 28 ไปตัดเส้นโค้ง จากจุดตัดลากเส้นขนานกับแกนนอนไปตัดกับแกนตั้งที่จุดใด จุดนั้น

จะบอกจำนวนข้อมูลที่ได้ กว่า 28 คะแนน มีจำนวน 25 คน ดังรูปตัวอย่างที่ 1.14

การนำเสนอข้อมูลที่เข้าใจง่าย

การนำเสนอข้อมูลที่เข้าใจง่ายเป็นการแสดงผลข้อมูลอย่างชัดเจนและเข้าใจได้ง่ายโดยไม่ซับซ้อน นี่คือบางวิธีในการนำเสนอข้อมูลที่เข้าใจง่าย:

1. **ใช้กราฟและแผนภูมิที่เหมาะสม**: การใช้กราฟและแผนภูมิที่เหมาะสม เช่น กราฟแท่ง กราฟเส้น และแผนภูมิวงกลม เพื่อแสดงข้อมูลในรูปแบบที่เข้าใจง่ายและชัดเจน
2. **ใช้สีและขนาดต่างๆ**: การใช้สีและขนาดต่างๆ ในการแสดงผลข้อมูลเพื่อเน้นความสำคัญของข้อมูลหรือสร้างการเน้นให้ข้อมูลมีความโดดเด่น
3. **ใช้ตารางเรียบร้อย**: การจัดเรียงข้อมูลในรูปแบบตารางที่เรียบร้อยและสะดวกต่อการอ่านและเข้าใจ
4. **การใช้ภาษาที่เข้าใจได้**: การใช้ภาษาที่เข้าใจง่าย ไม่มีคำศัพท์ทางวิชาการมากมาย และอธิบายข้อมูลอย่างชัดเจน
5. **สรุปข้อมูลสำคัญ**: การสรุปข้อมูลสำคัญโดยกระชับและชัดเจนเพื่อให้ผู้อ่านเข้าใจสาระสำคัญของข้อมูลได้โดยรวดเร็ว

การนำเสนอข้อมูลที่เข้าใจง่ายช่วยให้ผู้ใช้งานหรือผู้อื่นที่ได้รับข้อมูลสามารถเข้าใจและนำไปใช้ได้โดยง่ายและสะดวก

ตัวอย่างการนำเสนอข้อมูลที่เข้าใจง่ายในทางธุรกิจ:

1. **กราฟและแผนภูมิ**: การใช้กราฟและแผนภูมิเพื่อแสดงผลยอดขายสินค้าของบริษัทในแต่ละไตรมาสของปี ในรูปแบบของกราฟเส้นหรือกราฟแท่ง เพื่อให้ผู้บริโภคและผู้บริหารเข้าใจแนวโน้มและความเปลี่ยนแปลงของยอดขายได้ง่ายขึ้น

2. **ตารางสรุปผลลัพธ์**: การสรุปผลลัพธ์ของการวิเคราะห์การตลาด เช่น การจัดอันดับสินค้ายอดนิยม หรือการรายงานผลลัพธ์ของการสำรวจความพึงพอใจลูกค้า ในรูปแบบของตารางที่เรียบง่ายและชัดเจน
3. **แผนภูมิวงกลม**: การแสดงสัดส่วนของรายได้จากแหล่งกำไรหรือกลุ่มลูกค้าต่างๆ ในรูปแบบของแผนภูมิวงกลม เพื่อให้ผู้บริหารหรือผู้ตัดสินใจเข้าใจรูปแบบการกระจายของรายได้ได้ง่ายขึ้น
4. **นำเสนอข้อความ**: การนำเสนอข้อความสรุปผลลัพธ์หรือข้อมูลสำคัญในรูปแบบของสไลด์โปรเจคเพื่อให้ผู้ใช้งานหรือผู้บริหารเข้าใจข้อมูลได้อย่างชัดเจนและอย่างถูกต้อง
5. **สรุปแผนการดำเนินงาน**: การสรุปแผนการดำเนินงานหรือโครงการในรูปแบบของตารางหรือสไลด์โปรเจค เพื่อให้ทีมงานหรือผู้เกี่ยวข้องเข้าใจและทำตามขั้นตอนได้อย่างเหมาะสม

การนำเสนอข้อมูลที่เข้าใจง่ายในทางธุรกิจช่วยให้ผู้รับข้อมูลเข้าใจและตอบสนองต่อข้อมูลได้อย่างมีประสิทธิภาพและมีประสิทธิผล

สรุป

ในบทนี้ได้เรียนรู้เกี่ยวกับการแสดงข้อมูลอย่างมีประสิทธิภาพในธุรกิจ โดยมีหลักการสำคัญทั้ง 3 ด้านคือการใช้กราฟและแผนภูมิที่เหมาะสม เพื่อแสดงผลข้อมูลอย่างชัดเจนและเข้าใจได้ง่าย การสร้างตารางที่เรียบง่ายและสะดวกต่อการอ่านและเข้าใจข้อมูล และการนำเสนอข้อมูลให้เข้าใจได้ง่าย โดยใช้ภาษาที่ไม่ซับซ้อนและการสรุปข้อมูลอย่างชัดเจน เหล่านี้เป็นเครื่องมือที่สำคัญในการสร้างการเข้าใจและการตัดสินใจที่มีความสำคัญในองค์กรและธุรกิจต่างๆ ในการจัดทำรายงานหรือการบริหารจัดการทรัพยากรต่างๆ อย่างมีประสิทธิภาพ เพื่อสามารถทำความเข้าใจและวางแผนกลยุทธ์ทางธุรกิจให้เหมาะสมกับเป้าหมายและทรัพยากรขององค์กร

แบบฝึกหัด

จากตารางแจกแจงความถี่ที่กำหนดให้แสดงน้ำหนักของนักศึกษาปีที่ 4 ของวิทยาลัย แห่งหนึ่ง จำนวน 20 คน

น้ำหนัก (กก.)	ความถี่
58 – 62	3
63 – 67	5
68 – 72	10
73 – 77	2

จงหา

1. นักศึกษาที่มีน้ำหนัก 66 กก. ตกอยู่ในอันตรภาคชั้นใด
2. มีนักศึกษาจำนวนเท่าใดที่มีน้ำหนักต่ำกว่า 68 กก.
3. มีจำนวนนักศึกษาร้อยละเท่าไรที่มีน้ำหนักหนักกว่า 72 กก.
4. มีนักศึกษาจำนวนเท่าใดที่มีน้ำหนักตั้งแต่ 63 กก. ขึ้นไป แต่ไม่เกิน 77 กก.

เอกสารอ้างอิง

วิภาวรรณ เล้าอรุณ. (2557). องค์ความรู้จากการบริการวิชาการทางสถิติ. ภาควิชาสถิติ. คณะวิทยาศาสตร์. มหาวิทยาลัยศิลปากร

บทที่ 4
การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเป็นขั้นตอนที่สำคัญในการสร้างความเข้าใจและความรู้เกี่ยวกับข้อมูลที่มีอยู่ ในบทนี้จะได้เรียนรู้เกี่ยวกับการใช้สถิติในการวิเคราะห์ข้อมูล เช่น การคำนวณค่าสถิติที่สำคัญ เช่น เฉลี่ย, ส่วนเบี่ยงเบนมาตรฐาน, ความถี่ และความสัมพันธ์ นอกจากนี้ยังเรียนรู้เกี่ยวกับวิธีการสร้างและการอ่านผลการวิเคราะห์ข้อมูล เพื่อให้เข้าใจและนำข้อมูลไปใช้ในการตัดสินใจอย่างมีประสิทธิภาพ

ประเภทของสถิติที่ใช้ในการวิเคราะห์ข้อมูล

ส่วนใหญ่แล้วสถิติที่ใช้ในการวิเคราะห์ข้อมูลมี 2 ประเภท คือ สถิติบรรยายและสถิติอ้างอิง

1. สถิติบรรยาย (Descriptive Statistics) เป็นสถิติที่ศึกษาหาคำตอบเชิงตัวเลข เพื่อบรรยายลักษณะของข้อมูล โดยสถิติประเภทนี้ไม่สามารถนำไปอ้างอิงยังประชากรได้ เป็นสถิติที่บอกลักษณะของข้อมูลเท่านั้น เช่น การแจกแจงความถี่ ร้อยละ สัดส่วน อัตราส่วน เปอร์เซ็นไทล์ การหาค่าเฉลี่ย การวัดการกระจาย การวัดการแจกแจง เช่น ความโด่ง ความเบ้

2. สถิติอ้างอิง (Inferential Statistics) เป็นสถิติที่ศึกษาเกี่ยวกับข้อมูลจากกลุ่มตัวอย่างเพื่อประมาณ คาดคะเน เปรียบเทียบ หาความสัมพันธ์ แล้วนำไปอ้างอิงและสรุปไปยังประชากรเป้าหมาย ซึ่งสถิติที่ใช้สรุปอ้างอิงนั้นจะเกี่ยวข้องกับ ทฤษฎีความน่าจะเป็นและการสุ่มตัวอย่าง การประมาณค่าของประชากร การทดสอบสมมติฐาน (ศิริชัย กาญจนวาสี และคณะ,. 2559)

การคำนวณค่าสถิติที่สำคัญ

สถิติพรรณนาหรือบรรยาย (Descriptive Statistics) ส่วนใหญ่เป็นสถิติที่ใช้กับตัวแปรตัวเดียว (Univariate Statistics) มีจุดมุ่งหมายเพื่อใช้

สรุปให้เห็นสภาพของประชากรหรือกลุ่มตัวอย่างว่าเป็นอย่างไร (สุชาติ ประสิทธิ์รัฐสินธุ์, 2555) สถิติตัวแปรตัวเดียว สามารถใช้กับตัวแปรประเภทกลุ่ม ตัวแปรประเภทอันดับ ตัวแปรประเภทช่วง ตัวแปรประเภทอัตราส่วน ประกอบด้วย การแจกแจงความถี่ อัตราส่วน สัดส่วน อัตราส่วน ควอร์ไทล์ เดไซล์ เปอร์เซ็นไทล์ การหาค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม พิสัย ส่วนเบี่ยงเบนมาตรฐาน

การแจกแจงความถี่

การแจกแจงความถี่ (Frequency Distribution) คือ การย่อ การย่อยข้อมูลที่มีมากให้เป็น กลุ่ม เป็นรายการ และผลจากการแจกแจงความถี่สามารถนำเสนอในรูปแบบที่เข้าใจได้ง่าย เช่น ข้อความบรรยาย ตาราง กราฟ แผนภูมิใช้ได้กับตัวแปรประเภทกลุ่ม ตัวแปรประเภทอันดับ ตัวแปรประเภทช่วง

การแจกแจงความถี่แบบทางเดียว

คือ การนำตัวแปรจำนวน 1 ตัวแปรมาแจกแจงจำนวนความถี่และร้อยละ ตัวอย่างดังตารางที่ 4.1 การแจกแจงความถี่ของลูกค้าชาวไทยที่มาเยี่ยมชมร้านค้า ลาบูบู้ ที่จำแนกตามรายภาค

ตารางที่ 4.1 การแจกแจงความถี่

ภาค	จำนวน	ร้อยละ
กลาง	250	32
ตะวันออกเฉียงเหนือ	200	26
ใต้	175	23
ตะวันออก	145	19
รวม	770	100

ตัวอย่างการอ่าน คือ ลูกค้าชาวไทย จำแนกตามรายภาคพบว่าส่วนใหญ่เป็นลูกค้าชาวไทยภาคกลาง 250 คน (ร้อยละ 32) รองลงมาเป็นลูกค้าชาวไทยภาคตะวันออกเฉียงเหนือ 200 คน (ร้อยละ 26) และลูกค้าชาวไทยภาคใต้

175 คน (ร้อยละ 23) และมีเพียง 145 ร้อยละ 19 เป็นลูกค้าชาวไทยภาคตะวันออก ตามลำดับ

สถิติบรรยาย

การแจกแจงความถี่แบบสองทาง

การแจกแจงความถี่แบบสองทาง คือ การนำตัวแปรจำนวน 2 ตัวแปรมาแจกแจงจำนวนความถี่และร้อยละ ตัวอย่างดังตารางที่ 4.2 จำนวนของนักศึกษาที่เข้าชมร้านค้า มอลลี่จำแนกตามชั้นปีที่ศึกษาและเพศ

ตารางที่ 4.2 การแจกแจงความถี่แบบสองทาง

ชั้นปี	ชาย	หญิง	รวม
1	64	186	250
2	60	140	200
3	55	120	175
4	35	110	145
รวม	214	556	770

อัตราส่วน สัดส่วน ร้อยละ

อัตราส่วน สัดส่วนและร้อยละใช้ได้กับ ตัวแปรประเภทกลุ่ม ตัวแปรประเภทอันดับ ตัวแปรประเภทช่วง

อัตราส่วน (Ratio) เป็นการเปรียบเทียบความถี่ ระหว่างรายการย่อยกับรายการย่อย

$$= \frac{\text{ความถี่ของรายการย่อยที่ } \mathbf{1}}{\text{ความถี่ของ รายการย่อยที่ } \mathbf{2}}$$

เช่น อัตราส่วนของนักศึกษาชายและนักศึกษาหญิง

= จำนวนของนักศึกษาชาย:จำนวนของนักศึกษาหญิง

= 100: 200 = 1:2

อัตราส่วนของเกรด

= A : B : C : D : F

= 20: 50: 75: 25: 5

สัดส่วน (Proportion) เป็นการเปรียบเทียบความถี่ ระหว่างรายการย่อยกับจำนวนทั้งหมดของตัวแปร

$$= \frac{\text{ความถี่ของรายการ}}{\text{ความถี่ทั้งหมด}}$$

เช่น สัดส่วนของลูกค้าหญิง = $\frac{\mathbf{350}}{\mathbf{700}}$ = .50

ร้อยละ (percentage) เป็นการเปรียบเทียบความถี่ของรายการย่อยกับความถี่ทั้งหมดและเทียบกับ100

$$\text{ร้อยละ} = \frac{\text{ความถี่ของรายการ}}{\text{ความถี่ทั้งหมด}} \times 100$$

เช่น ร้อยละของจำนวนลูกค้าหญิง = $\frac{\mathbf{360}}{\mathbf{700}}$ x 100 = 51

การวัดแนวโน้มเข้าสู่ส่วนกลาง

การวัดแนวโน้มเข้าสู่ส่วนกลาง (measures of central tendency) เป็นระเบียบวิธีทางสถิติในการหาค่าเพียงค่าเดียวที่จะใช้เป็นตัวแทนของข้อมูลทั้งชุด ค่าที่หาได้นี้จะทำให้สามารถทราบถึงลักษณะของข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ ค่าที่หาได้นี้จะเป็นค่ากลาง ๆ เรียกว่า ค่ากลาง (ระพินทร์ โพธิ์ศรี, 2549)
ประเภทของการวัดแนวโน้มเข้าสู่ส่วนกลาง

1. ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) ใช้กับข้อมูลค่าสังเกตที่รวบรวมมา
2. มัธยฐาน (Median) ใช้ได้กับข้อมูลค่าสังเกตที่รวบรวมมา
3. ฐานนิยม (Mode) ใช้ได้กับข้อมูลค่าสังเกตที่รวบรวมมา

ค่าเฉลี่ยเลขคณิต

ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) หมายถึง การหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลทั้งหมดการหาค่าเฉลี่ยเลขคณิตสามารถหาได้ (วุฒิ สุขเจริญ, 2561) โดย

$$\text{ค่าเฉลี่ยเลขคณิต} = \frac{\text{ผลบวกของข้อมูลทุกค่า}}{\text{จำนวนข้อมูลทั้งหมด}}$$

ค่าเฉลี่ยใช้สัญลักษณ์ $\bar{x}$ สำหรับกลุ่มตัวอย่างและใช้สัญลักษณ์ μ สำหรับประชากร

ตัวอย่าง จากการสอบถามอายุของลูกค้ากลุ่มหนึ่งเป็นดังนี้ 10, 25, 30, 42, 16, 14, 18, 17 จงหาค่าเฉลี่ยเลขคณิตของอายุลูกค้ากลุ่มนี้

$$\text{วิธีทำ} = \frac{10+25+30+42+16+14+18+17}{8}$$

$$= 21.5$$

ดังนั้นค่าเฉลี่ย อายุลูกค้ากลุ่มนี้ = 21.50 ปี

มัธยฐาน (Median)

มัธยฐาน หมายถึง ค่ากึ่งกลางของข้อมูลชุดนั้น หรือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลชุดนั้น เมื่อได้จัดเรียงค่าของข้อมูลจากน้อยที่สุด ไปหามากที่สุดหรือจากมากที่สุดไปหาน้อยที่สุด ค่ากึ่งกลางจะเป็นตัวแทนที่แสดงว่ามีข้อมูลที่มากกว่าและน้อยกว่านี้อยู่ 50%

การหาค่ามัธยฐาน สามารถหาได้ดังนี้

1. เรียงข้อมูลจากน้อยไปมาก หรือจากมากไปน้อย
2. หาตำแหน่งของมัธยฐาน จาก $= \frac{n+1}{2}$

เมื่อ n = จำนวนข้อมูลทั้งหมด

ตัวอย่าง จงหามัธยฐานของข้อมูลต่อไปนี้ 9, 12, 5, 11, 14, 6, 16, 17, 13

วิธีทำ เรียงข้อมูลที่มีค่าน้อยที่สุดไปหาข้อมูลที่มีค่ามากที่สุดคือ 5, 6, 9, 11, 12, 13, 14, 16, 17

หาตำแหน่งมัธยฐาน $= \frac{9+1}{2} = 5$

และนับเรียงจากข้อมูลตำแหน่งที่ 1 ไปถึงตำแหน่งที่ 5 มัธยฐานของข้อมูล = 12

ฐานนิยม (Mode)

ฐานนิยม หมายถึง ค่าของคะแนนที่ซ้ำกันมากที่สุดหรือค่าคะแนนที่มีความถี่สูงที่สุดในข้อมูลชุดนั้น

การหาค่าฐานนิยม สามารถหาโดยหาค่าของข้อมูลที่ซ้ำกันมากที่สุด คือฐานนิยม

ตัวอย่างจงหาฐานนิยมของข้อมูลต่อไปนี้ 3, 2, 4, 5, 6, 4, 8, 4, 9, 10

ข้อมูลที่ซ้ำกันมากที่สุดคือ 4

ฐานนิยมคือ 4

ข้อมูลบางชุดอาจมีฐานนิยม 2 ค่า เช่น 10, 14, 12, 10, 11, 15, 12, 14, 12, 10

ข้อมูลที่ซ้ำกันมากที่สุดคือ 10 กับ 12

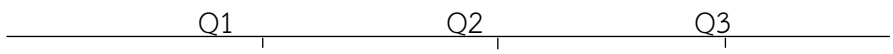
ฐานนิยม คือ 10 กับ 12

ข้อมูลบางชุดอาจจะไม่มีฐานนิยม ซึ่งได้แก่ ข้อมูลที่ไม่มีรายการซ้ำกันเลย เช่น 8, 9, 10, 11, 13, 16

ควอร์ไทล์ เดไซล์ เปอร์เซ็นไทล์

ควอร์ไทล์ เดไซล์ เปอร์เซ็นไทล์ใช้ได้กับตัวแปรประเภทอัตราส่วน

ควอร์ไทล์ (Quartile) คือ การแบ่งข้อมูลออกเป็น 4 ส่วนแต่ละส่วนมีจำนวนเท่ากันเรียกว่า ควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 ควอร์ไทล์ที่ 3 ตามลำดับ



ตัวอย่างแสดงการหาค่า
จงหาค่าควอร์ไทล์ที่ 3 ของชุดข้อมูล 11, 30, 24, 27, 26, 20, 15
วิธีทำ เรียงข้อมูลจากน้อยไปมาก จะได้ชุดข้อมูลใหม่ คือ 11, 15, 20, 24, 26, 27, 30

จากสูตร ตำแหน่งควอร์ไทล์ที่ k = $\frac{\mathbf{x_k}}{\mathbf{4}}$ x (n+1)
ควอร์ไทล์ที่ 3 = $\frac{\mathbf{3}}{\mathbf{4}}$ x (7+1) = 6

ดังนั้น ควอร์ไทล์ที่ 3 คือ ข้อมูลตัวที่ 6
ควอร์ไทล์ที่ 3 คือ ข้อมูลชุดนี้ คือ 27

การหาค่า ควอร์ไทล์ที่ไม่ใช่จำนวนเต็ม
จงหาค่าควอร์ไทล์ที่ 1 ของชุดข้อมูล 11, 30, 24, 27
วิธีทำ เรียงข้อมูลจากน้อยไปมาก จะได้ชุดข้อมูลใหม่ คือ 11, 24, 27, 30

ควอร์ไทล์ที่1 = $\frac{\mathbf{1}}{\mathbf{4}}$ x (4+1) = 1.25

ดังนั้น ควอร์ไทล์ที่ 1 คือ ข้อมูลตำแหน่งที่ 1.25 คืออยู่ระหว่าง ข้อมูลตำแหน่งที่ 1 และ 2
จึงมีวิธีหา คือ $\boldsymbol{x_{1.25}}$= $\boldsymbol{x_1}$+.25 x ($\boldsymbol{x_2 - x_1}$)

ควอร์ไทล์ที่ 1 คือ = 11+.25 x (24-11) = 14.25

การหาค่า ควอร์ไทล์ ของข้อมูลแจกแจงความถี่

ตารางที่ 10.3

ช่วง	จำนวน
21-30	2
31-40	4
41-50	4
51-60	3

จากสูตร ตำแหน่งควอร์ไทล์ที่ k = $\frac{\mathbf{x_k}}{\mathbf{4}}$ x (n)

ควอร์ไทล์ที่ k ของข้อมูลแจกแจงความถี่

$$= L + I \times \left(\frac{\frac{\mathbf{x_{k(n)}}}{\mathbf{4}} - F}{f} \right)$$

L คือ ขอบล่างของชั้นที่มีควอร์ไทล์

I คือ ความกว้างของชั้นอัตรภาค

F คือ ความถี่สะสมของชั้นที่ต่ำกว่าชั้นควอร์ไทล์

f คือ ความถี่ของชั้นควอร์ไทล์

จงหาค่าควอร์ไทล์ที่ 3 จะต้องทราบว่า ตำแหน่งควอร์ไทล์ที่ 3 อยู่ชั้นไหน

ตารางที่ 10.4

ช่วง	จำนวน	F
21-30	2	2
31-40	4	6
41-50	4	10

51-60	3	13

ตำแหน่งควอร์ไทล์ที่ 3 = $\frac{\mathbf{3}}{\mathbf{4}}$ x (13) = 9.75

หาชั้นที่ตำแหน่งควอร์ไทล์ที่ 3 อยู่ คือ 41-50

ดังนั้น ควอร์ไทล์ที่ 3 คือ = 40.5+ 10 x $\left(\frac{\frac{3}{4}(13)-6}{10}\right)$ = 44.25

เดไซล์ (Decile) คือ การแบ่งข้อมูลออกเป็น 10 ส่วนแต่ละส่วนมีจำนวนเท่ากัน D1.....D9 ตามลำดับ

จากสูตร ตำแหน่งเดไซล์ที่ k = $\frac{\mathbf{x_k}}{\mathbf{10}}$ x (n+1)

จงหาค่าเดไซล์ที่ 3 ของชุดข้อมูล 11, 30, 24, 27, 26, 20, 15

วิธีทำ เรียงข้อมูลจากน้อยไปมาก จะได้ชุดข้อมูลใหม่ คือ 11, 15, 20, 24, 26, 27, 30

ตำแหน่งเดไซล์ที่ 3 = $\frac{\mathbf{3}}{\mathbf{10}}$ x (7+1) = 2.4

ดังนั้น เดไซล์ที่ 3 อยู่ตำแหน่ง 2.4 นั้นคือระหว่างตำแหน่ง 2 และ 3

จึงมีวิธีหา คือ $\boldsymbol{x_{2.4}}$= $\boldsymbol{x_2}$+.4 x ($\boldsymbol{x_3 - x_2}$)

เดไซล์ที่ 3 = 15+.4 x (20-15) = 17

การหาค่า เดไซล์ของข้อมูลแจกแจงความถี่

ตารางที่ 10.5

ช่วง	จำนวน
21-30	2
31-40	4
41-50	4
51-60	3

จากสูตร ตำแหน่งเดไซล์ที่ k = $\frac{\mathbf{x_k}}{\mathbf{10}}$ x (n)

เดไซล์ที่ k ของข้อมูลแจกแจงความถี่

$$= L + I \times \left(\frac{\frac{\mathbf{x_k(n)}}{\mathbf{10}} - F}{f} \right)$$

L คือ ขอบล่างของชั้นที่มีเดไซล์
I คือ ความกว้างของชั้นอัตรภาค
F คือ ความถี่สะสมของชั้นที่ต่ำกว่าชั้นเดไซล์
f คือ ความถี่ของชั้นเดไซล์

จงหาค่าเดไซล์ที่ 3 จะต้องทราบว่า ตำแหน่งเดไซล์ที่ 3 อยู่ชั้นไหน
ตารางที่ 10.6

ช่วง	จำนวน	F
21-30	2	2
31-40	4	6
41-50	4	10
51-60	3	13

ตำแหน่งเดไซล์ที่ 3 = $\frac{\mathbf{3}}{\mathbf{10}}$ x (13) = 3.9

หาชั้นที่ตำแหน่งเดไซล์ที่ 3 อยู่ คือ 31-40

ดังนั้น เดไซล์ที่ 3 คือ = 30.5+ 10 x $\left(\frac{\frac{3}{10}(13) - 2}{6} \right)$ = 33.67

เปอร์เซ็นไทล์ (Percentile) คือ การแบ่งข้อมูลออกเป็น 100 ส่วนแต่ละส่วนมีจำนวนเท่ากัน P1.....P99 ตามลำดับ

จากสูตร ตำแหน่งเปอร์เซ็นไทล์ ที่ k = $\frac{\mathbf{x_k}}{\mathbf{100}}$ x (n+1)

จงหาค่าเปอร์เซ็นไทล์ ที่ 70 ของชุดข้อมูล 11, 30, 24, 27, 26, 20, 15

วิธีทำ เรียงข้อมูลจากน้อยไปมาก จะได้ชุดข้อมูลใหม่ คือ 11, 15, 20, 24, 26, 27, 30

ตำแหน่งเปอร์เซ็นไทล์ ที่ 70 = $\frac{\mathbf{70}}{\mathbf{100}}$ x (7+1) = 5.6

ดังนั้น เปอร์เซ็นไทล์ ที่ 70 อยู่ตำแหน่ง 5.6 นั้นคือระหว่างตำแหน่ง 5 และ 6 จึงมีวิธีหา คือ $\boldsymbol{x}_{\mathbf{5.6}}$= $\boldsymbol{x}_{\mathbf{5}}$+.6 x ($\boldsymbol{x}_{\mathbf{6}} - \boldsymbol{x}_{\mathbf{5}}$)
เปอร์เซ็นไทล์ ที่ 70 = 26+ .6x (27-26) = 26.6

การหาค่า เปอร์เซ็นไทล์ ของข้อมูลแจกแจงความถี่

ตารางที่ 10.7

ช่วง	จำนวน
21-30	2
31-40	4
41-50	4
51-60	3

จากสูตร ตำแหน่งเปอร์เซ็นไทล์ ที่ k = $\frac{\mathbf{x_k}}{\mathbf{100}}$ x (n)

เปอร์เซ็นไทล์ ที่ k ของข้อมูลแจกแจงความถี่ = L+ I x $\left(\frac{\frac{\mathbf{x_{k(n)}}}{\mathbf{100}} - F}{f}\right)$

L คือ ขอบล่างของชั้นที่มีเปอร์เซ็นไทล์
I คือ ความกว้างของชั้นอัตรภาค
F คือ ความถี่สะสมของชั้นที่ต่ำกว่าชั้นเปอร์เซ็นไทล์
f คือ ความถี่ของชั้นเปอร์เซ็นไทล์

จงหาค่าเปอร์เซ็นไทล์ ที่ 70 จะต้องทราบว่า ตำแหน่งเปอร์เซ็นไทล์ ที่ 70 อยู่ชั้นไหน

ตารางที่ 10.8

ช่วง	จำนวน	F
21-30	2	2
31-40	4	6
41-50	4	10
51-60	3	13

ตำแหน่งเปอร์เซ็นไทล์ ที่ 70 = $\frac{70}{100}$ x (13) = 9.1

หาชั้นที่ตำแหน่งเปอร์เซ็นไทล์ ที่ 70 อยู่ คือ 41-50

ดังนั้น เปอร์เซ็นไทล์ ที่ 70 คือ = 40.5+ 10 x $\left(\frac{\frac{70}{100}(13)-6}{10}\right)$ = 43.6

การวัดการกระจาย

การวัดการกระจาย (Measures of dispersion) คือ การวัดการกระจายของข้อมูลที่ได้จากการรวบรวมข้อมูลโดยใช้เครื่องมือการวิจัย เช่น แบบสอบถาม แบบสำรวจ การวัดที่ได้จากการวัดแนวโน้มเข้าสู่ส่วนกลางทำให้ทราบคุณลักษณะของข้อมูลที่เป็นตัวแทนกลุ่มเพียงค่าเดียวเท่านั้นแต่เนื่องจากค่าที่เป็นตัวแทนกลุ่มค่าเดียวกันไม่สามารถทำให้ทราบลักษณะของข้อมูลได้เพียงพอ (ยุทธ ไกยวรรณ์, 2551, น. 149) เช่น อายุการใช้งานของทีวี 2 ยี่ห้อ

ยี่ห้อ A (ปี) : 12 15 43 65 30 อายุการใช้งานเฉลี่ย 33 ปี

ยี่ห้อ B (ปี) : 29 32 35 37 32 อายุการใช้งานเฉลี่ย 33 ปี

จะพบว่าข้อมูลทั้ง 2 ชุดมีค่าเฉลี่ยเท่ากัน บางครั้งการพิจารณาแค่ค่าเฉลี่ยอย่างอาจจะไม่เพียงพอ จึงต้องมีการวัดการกระจายของข้อมูล ซึ่งถ้าข้อมูลชุดใดมีการกระจายน้อย จะทำให้ข้อมูลชุดนั้นน่าเชื่อถือ ดังนั้นการอธิบายข้อมูลจึงต้องมีทั้งการวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจายของข้อมูลควบคู่กันไป

สถิติที่ใช้ในการวัดการกระจาย

สถิติที่ใช้ในการวัดการกระจาย คือ พิสัย ส่วนเบี่ยงเบนมาตรฐาน

พิสัย

พิสัย (Range) คือ ผลต่างระหว่างคะแนนที่มีค่าสูงสุดและคะแนนต่ำสุดนั้นคือ ค่าสูงสุด-ค่าต่ำสุด การนำพิสัยไปใช้งานวิจัยจะใช้ควบคู่กับฐานนิยม

อายุการใช้งานทีวียี่ห้อ A (ปี) : 12 15 15 65 19 30 15 ฐานนิยม คือ 15 ปี พิสัย คือ 53

อายุการใช้งานทีวียี่ห้อ B (ปี) : 29 30 35 30 37 32 33 ฐานนิยม คือ 30 ปี พิสัย คือ 8

จะเห็นว่าอายุการใช้งานของยี่ห้อ B จะมีการกระจายน้อยกว่า แสดงว่าผู้ซื้อจะมีความเสี่ยงกว่าถ้าเลือกใช้ทีวียี่ห้อ B

ส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐานเป็นการวัดการกระจายที่นำข้อมูลทุกค่ามาคำนวณ เป็นวิธีที่นักสถิตินิยมใช้มากที่สุดและเป็นวิธีที่ใช้ในการวัดการกระจายได้ดีที่สุด สามารถไปใช้ในสถิติขั้นสูงต่อไปส่วนเบี่ยงเบนมาตรฐานใช้สัญลักษณ์ s หรือ S.D.

ส่วนเบี่ยงเบนมาตรฐานของประชากร

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

σ แทน ส่วนเบี่ยงเบนมาตรฐานของประชากร

x แทน ข้อมูลหรือคะแนนแต่ละตัว

μ แทน ค่าเฉลี่ยของประชากร

N แทน จำนวนข้อมูลทั้งหมด

ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

$$S.D. = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

x แทน ข้อมูล

$\bar{x}$ แทน ค่าเฉลี่ยเลขคณิต

n แทน จำนวนข้อมูลทั้งหมด

ถ้าส่วนเบี่ยงเบนมาตรฐานสูงถือว่าข้อมูลมีการกระจายมาก ดังนั้นถ้าค่าของส่วนเบี่ยงเบนมาตรฐานต่ำจะถือว่าข้อมูลมีการกระจายน้อย ข้อมูลชุดนั้นจึงน่าเชื่อถือมากกว่า

ตัวอย่างการหาค่าส่วนเบี่ยงเบนมาตรฐาน

อายุการใช้งานพัดลม ยี่ห้อ A (ปี) : 12 15 15 65 19 30

$\bar{x}$ = 26

S.D.=

$$\sqrt{\frac{(12-26)^2+(15-26)^2+(15-26)^2+(65-26)^2+(19-26)^2+(30-26)^2}{6-1}}$$

S.D. = 20.11

ความแปรปรวน

ความแปรปรวน คือ ค่าเฉลี่ยของผลรวมระหว่างผลต่างกำลังสองของค่าตัวเลขแต่ละตัวในข้อมูลชุดหนึ่ง ๆ กับค่าเฉลี่ยของข้อมูลชุดนั้น หรือค่าส่วนเบี่ยงเบนมาตรฐานยกกำลังสอง

$S.D.^2$ คือ ความแปรปรวนของกลุ่มตัวอย่าง

σ^2 คือ ความแปรปรวนของประชากร

ตัวอย่างการหาค่าความแปรปรวน

$S.D.^2 = 20.11^2$

ความแปรปรวน = 404.42

สถิติอ้างอิง

สถิติอ้างอิง (Inferential Statistics) มีการแบ่งประเภทหลัก ๆ ออกเป็นสองประเภทใหญ่ ๆ ได้แก่ สถิติแบบพารามิเตอร์ (Parametric Statistics) และสถิติแบบไม่ใช้พารามิเตอร์ (Non-parametric Statistics) โดยการเลือกใช้ขึ้นอยู่กับระดับของข้อมูล (Level of Measurement) และวัตถุประสงค์ของการวิเคราะห์ ขอสรุปประเภทหลัก ๆ และการใช้งานดังนี้

1) สถิติแบบพารามิเตอร์ (Parametric Statistics) สถิติประเภทนี้เป็นสถิติอ้างอิง (Inferential Statistics) ที่ใช้ทดสอบสมมติฐานโดยอาศัยความน่าจะเป็นโดยการศึกษาจากกลุ่มตัวอย่างเพื่อสรุปผลอ้างอิงไปยังประชากร การใช้สถิติประเภทนี้ ต้องมีการสุ่มตัวอย่างที่มีขนาดตัวอย่างที่เหมาะสมและน่าเชื่อถือ สถิติแบบพารามิเตอร์จะมีสมมติฐานเกี่ยวกับการกระจายตัวของข้อมูล เช่น **การกระจายตัวแบบปกติ** การใช้สถิติแบบพารามิเตอร์จะให้ผลลัพธ์ที่แม่นยำและมีประสิทธิภาพเมื่อข้อมูลเป็นไปตามสมมติฐานเหล่านี้ (หทัยกาญจน์ ชูตระกูล,2565)
ประเภทและการใช้งาน

1. การทดสอบสมมติฐาน (Hypothesis Testing)
    - z-test: ใช้ทดสอบค่าเฉลี่ยของประชากรเมื่อรู้ค่ามาตรฐานของประชากรและข้อมูลมีการกระจายตัวแบบปกติ
    - t-test: ใช้ทดสอบค่าเฉลี่ยของกลุ่มข้อมูลเมื่อไม่ทราบค่ามาตรฐานของประชากร
        - One-sample t-test: ใช้ทดสอบค่าเฉลี่ยของกลุ่มข้อมูลกับค่าคงที่
        - Independent t-test: ใช้ทดสอบค่าเฉลี่ยของกลุ่มข้อมูลสองกลุ่มที่ไม่เกี่ยวข้องกัน

        - Paired t-test: ใช้ทดสอบค่าเฉลี่ยของกลุ่มข้อมูลสองกลุ่มที่มีความสัมพันธ์กัน
2. การวิเคราะห์ความแปรปรวน (Analysis of Variance, ANOVA)
    - One-way ANOVA: ใช้เปรียบเทียบค่าเฉลี่ยของหลายกลุ่มที่มีปัจจัยเดียว
    - Two-way ANOVA: ใช้เปรียบเทียบค่าเฉลี่ยของหลายกลุ่มที่มีปัจจัยสองปัจจัย
3. การวิเคราะห์การถดถอย (Regression Analysis)
    - Simple Linear Regression: ใช้เมื่อศึกษาความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระหนึ่งตัวกับตัวแปรตามหนึ่งตัว
    - Multiple Linear Regression: ใช้เมื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตาม
4. การวิเคราะห์ความสัมพันธ์ (Correlation Analysis)
    - Pearson Correlation: ใช้สำหรับข้อมูลที่มีการกระจายตัวแบบปกติเพื่อวัดความสัมพันธ์เชิงเส้นระหว่างตัวแปรสองตัว

2) สถิติแบบไม่ใช้พารามิเตอร์ (Non-parametric Statistics)

สถิติแบบไม่ใช้พารามิเตอร์ไม่จำเป็นต้องมีสมมติฐานเกี่ยวกับการกระจายตัวของข้อมูล เหมาะสำหรับข้อมูลที่ไม่เป็นไปตามสมมติฐานของสถิติแบบพารามิเตอร์ เช่น ข้อมูลที่มีการกระจายตัวไม่ปกติ หรือข้อมูลที่เป็นเชิงอันดับ

ประเภทและการใช้งาน

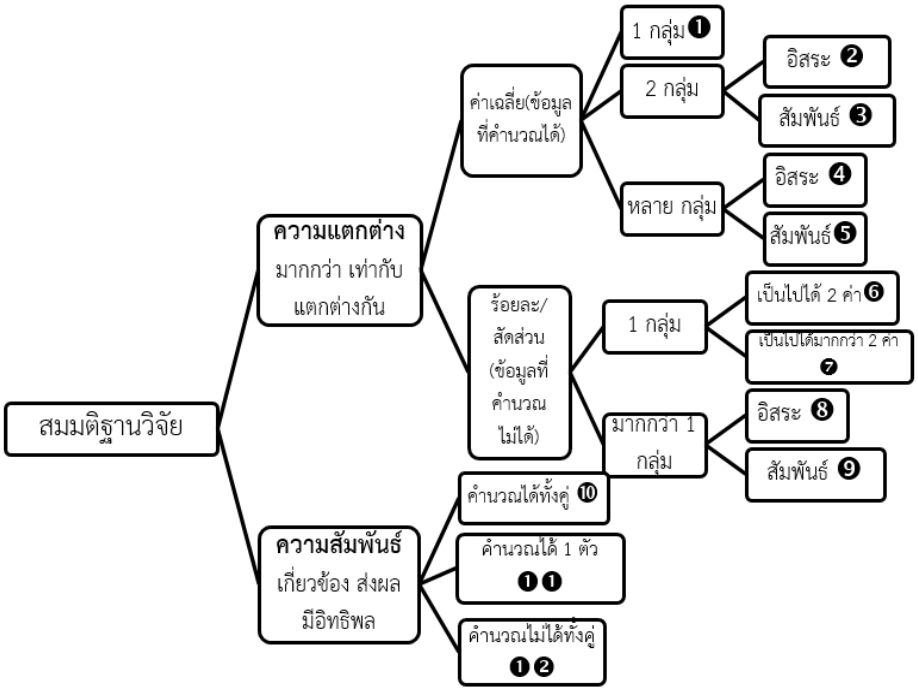
1. การทดสอบสมมติฐาน (Hypothesis Testing)
    - Chi-square test: ใช้สำหรับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเชิงหมวดหมู่
    - Mann-Whitney U test: ใช้ทดสอบความแตกต่างระหว่างกลุ่มสองกลุ่มแทน t-test เมื่อข้อมูลไม่เป็นไปตามการกระจายตัวแบบปกติ

- Wilcoxon Signed-Rank test: ใช้ทดสอบความแตกต่างของกลุ่มข้อมูลสองกลุ่มที่มีความสัมพันธ์กันแทน paired t-test

2. การวิเคราะห์ความแปรปรวน (Non-parametric ANOVA)
    - Kruskal-Wallis H test: ใช้แทน One-way ANOVA เมื่อข้อมูลไม่เป็นไปตามสมมติฐานของ ANOVA
    - Friedman test: ใช้แทน Two-way ANOVA เมื่อข้อมูลไม่เป็นไปตามสมมติฐานของ ANOVA
3. การวิเคราะห์ความสัมพันธ์ (Correlation Analysis)
    - Spearman's Rank Correlation: ใช้สำหรับข้อมูลที่ไม่ต้องมีการกระจายตัวแบบปกติ เพื่อวัดความสัมพันธ์ระหว่างตัวแปรสองตัว

การเลือกใช้สถิติแบบพารามิเตอร์หรือสถิติแบบไม่ใช้พารามิเตอร์ขึ้นอยู่กับระดับของข้อมูลและสมมติฐานเกี่ยวกับการกระจายตัวของข้อมูล หากข้อมูลเป็นไปตามสมมติฐานที่กำหนด สถิติแบบพารามิเตอร์จะให้ผลลัพธ์ที่แม่นยำกว่า แต่ถ้าข้อมูลไม่เป็นไปตามสมมติฐานเหล่านั้น สถิติแบบไม่ใช้พารามิเตอร์จะเป็นทางเลือกที่ดีกว่า

นอกจากนี้ยังมีการแบ่งประเภทของสถิติตามการทดสอบสมมติฐาน โดยแบ่งเป็น การทดสอบความแตกต่าง และการทดสอบความสัมพันธ์ (ศิริชัย พงษ์วิชัย, 2561)



ภาพที่ 4.1 สถิติที่ใช้ในการทดสอบสมมติฐาน

ที่มา: ศิริชัย พงษ์วิชัย (2561)

จากภาพที่ 4.1 มีสถิติที่ใช้ทดสอบสมมติฐาน ดังนี้

1. t-test ใช้ทดสอบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณได้ จำนวน 1 กลุ่ม

2. t-test ใช้ทดสอบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณได้ จำนวน 2 กลุ่มที่เป็นอิสระกัน

3. Paired t-test ใช้ทดสอบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณได้ จำนวน 2 กลุ่มที่สัมพันธ์กัน

4. F-test ได้จาก Anova ใช้ทดสอบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณได้ จำนวนหลายกลุ่มที่เป็นอิสระกัน

5. F-test ได้จาก Anova ใช้ทดสอบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณได้ จำนวนหลายกลุ่มที่สัมพันธ์กัน

6. Binomial test ใช้ทดสอบความแตกต่างร้อยละหรือสัดส่วนของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณไม่ได้ จำนวน 1 กลุ่มและเป็นไปได้ 2 ค่า

7. Chi-square ใช้ทดสอบความแตกต่างร้อยละหรือสัดส่วนของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณไม่ได้ จำนวน 1 กลุ่มและเป็นไปได้มากกว่า 2 ค่า

8. Chi-square ใช้ทดสอบความแตกต่างร้อยละหรือสัดส่วนของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณไม่ได้ จำนวนมากกว่า 1 กลุ่มและเป็นอิสระกัน

9. McNemar ใช้ทดสอบความแตกต่างร้อยละหรือสัดส่วนของกลุ่มตัวอย่างหรือข้อมูลที่คำนวณไม่ได้ จำนวนมากกว่า 1 กลุ่มและสัมพันธ์กัน

10. t-test (ทดสอบ) r (วัด) ใช้ทดสอบความสัมพันธ์ข้อมูลที่คำนวณได้ทั้งคู่

11. t, F-test (ทดสอบ) Eta (วัด) ใช้ทดสอบความสัมพันธ์ข้อมูลที่คำนวณได้ 1 ตัว

12. Chi-square ใช้ทดสอบความสัมพันธ์ข้อมูลที่คำนวณไม่ได้ทั้งคู่

การสร้างและการอ่านผลการวิเคราะห์

การสร้างและการอ่านผลการวิเคราะห์ข้อมูลเป็นขั้นตอนสำคัญในการทำสถิติวิเคราะห์ ซึ่งมีขั้นตอนดังนี้

1. การสร้างข้อมูล: ขั้นตอนแรกในการวิเคราะห์ข้อมูลคือการสร้างข้อมูล โดยมักจะใช้ข้อมูลจากการสำรวจ, การสัมภาษณ์, หรือการเก็บข้อมูลจากแหล่งต่างๆ ขึ้นอยู่กับวัตถุประสงค์ของการวิเคราะห์

2. การวิเคราะห์ข้อมูล: หลังจากได้รับข้อมูลแล้ว เราจะทำการวิเคราะห์ข้อมูลโดยใช้เครื่องมือทางสถิติ เช่น การหาค่าเฉลี่ย, ค่าส่วนเบี่ยงเบนมาตรฐาน, การทดสอบสมมติฐาน, และการคำนวณค่าสถิติอื่นๆ เพื่อเข้าใจและอธิบายลักษณะของข้อมูล

3. การสร้างแผนภูมิและกราฟ: การสร้างแผนภูมิและกราฟช่วยให้ข้อมูลเป็นกระแสและเข้าใจง่ายขึ้น โดยสามารถใช้แผนภูมิและกราฟต่างๆ เช่น กราฟเส้น, แผนภูมิแท่ง, แผนภูมิวงกลม เพื่อแสดงข้อมูลในรูปแบบที่สามารถอ่านได้ง่าย

4. การอ่านผลการวิเคราะห์: เมื่อได้ผลการวิเคราะห์ข้อมูลแล้ว จะต้องทำการอ่านและตีความผลการวิเคราะห์อย่างถูกต้อง เพื่อให้สามารถทำประเด็นในด้านต่างๆ ได้อย่างเหมาะสม

5. การสรุปผล: สุดท้ายคือการสรุปผลการวิเคราะห์ โดยจะสรุปออกมาเป็นข้อสรุปที่มีประโยชน์ และเชื่อถือได้ต่อผู้ที่ต้องการใช้ข้อมูลนั้นในการตัดสินใจและการวางแผนในอนาคต

ดังนั้น การสร้างและการอ่านผลการวิเคราะห์ข้อมูลเป็นขั้นตอนสำคัญที่ช่วยให้เราทำความเข้าใจและใช้ประโยชน์จากข้อมูลได้อย่างมีประสิทธิภาพ

ตัวอย่างที่ การรายงานกำไรของ บริษัทแม่น้ำสองเป็นรายเดือนของปี พ.ศ. 2567

month	profit
Jan	1230
Feb	4500
Mar	4100
Apr	8000
May	7000
Jun	5400
Jul	6200
Aug	8100
Sep	5000
Oct	6000
Nov	3540
Dec	8540

จากข้อมูลที่ได้รับ เป็นการรายงานกำไรในแต่ละเดือนของปี สามารถอ่านผลการวิเคราะห์อย่างละเอียดดังนี้

มกราคม (Jan): กำไร 1,230 บาท; กุมภาพันธ์ (Feb): กำไร 4,500 บาท
มีนาคม (Mar): กำไร 4,100 บาท; เมษายน (Apr): กำไร 8,000 บาท
พฤษภาคม (May): กำไร 7,000 บาท; มิถุนายน (Jun): กำไร 5,400 บาท
กรกฎาคม (Jul): กำไร 6,200 บาท; สิงหาคม (Aug): กำไร 8,100 บาท
กันยายน (Sep): กำไร 5,000 บาท; ตุลาคม (Oct): กำไร 6,000 บาท
พฤศจิกายน (Nov): กำไร 3,540 บาท; ธันวาคม (Dec): กำไร 8,540 บาท

จากข้อมูลนี้ สามารถสรุปได้ว่ามีการแสดงผลของกำไรในแต่ละเดือนของปี โดยรายได้สูงสุดอยู่ในเดือนเมษายนและธันวาคม ซึ่งอาจเกี่ยวข้องกับการเพิ่มขึ้นของยอดขายหรือกิจกรรมทางธุรกิจในช่วงเทศกาลหรือช่วงเวลาที่มีการลดราคาสินค้าหรือโปรโมชั่น เป็นต้น

หรือวิเคราะห์โดยสรุปดังนี้ ข้อมูลรายงานกำไรในแต่ละเดือนของปีแสดงถึงผลตอบรับทางการเงินของธุรกิจตลอดปี โดยสรุปได้ว่ากำไรสูงสุดปรากฏในเดือนเมษายนและธันวาคม เป็นไปได้ที่ความสำเร็จนี้เกิดจากยอดขายที่เพิ่มขึ้นหรือกิจกรรมโปรโมชั่นที่เป็นเดี่ยวกัน อย่างไรก็ตามความสำเร็จในเดือนนั้นๆ อาจขึ้นอยู่กับตัวปัจจัยหลายประการ เช่น การลดราคาสินค้าหรือการจัดโปรโมชั่น และอาจมีผลกระทบจากปัจจัยภายนอก เช่น การเปลี่ยนแปลงในตลาดหรือสภาพการเงินโลก

นอกจากนี้ ยังนำไปสร้างกราฟและอ่านผลการวิเคราะห์ข้อมูลหากต้องการทำการวิเคราะห์ข้อมูลนี้เพื่อให้เข้าใจดียิ่งขึ้น
ด้วยวิธีการดังนี้

1. การสร้างกราฟเส้น: เราสามารถสร้างกราฟเส้นเพื่อแสดงแนวโน้มของกำไรในแต่ละเดือนของปี โดยที่แกน X แทนเดือนและแกน Y แทนจำนวนกำไรที่ได้รับ ซึ่งจะช่วยให้เราเห็นแนวโน้มและการเปลี่ยนแปลงของกำไรตลอดระยะเวลาได้ดังภาพที่ 4.2 - 4.3

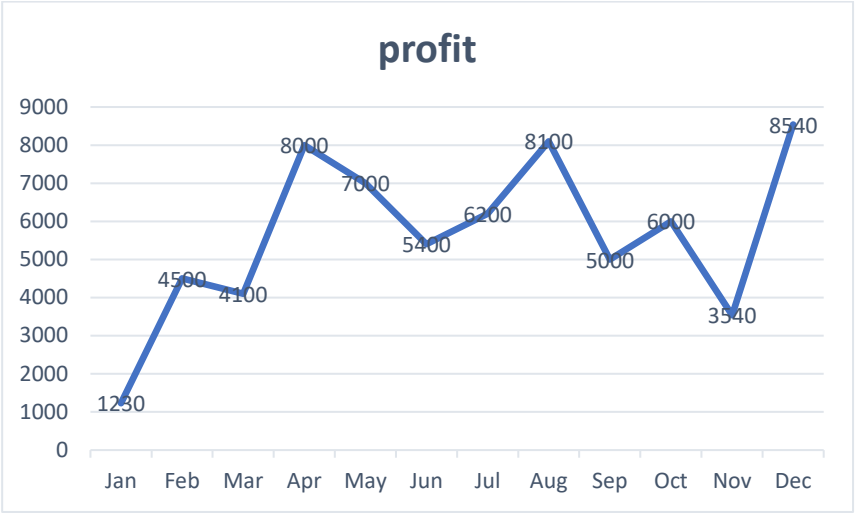
2. การคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน: เราสามารถคำนวณหาค่าเฉลี่ยของกำไรในแต่ละเดือนและค่าส่วนเบี่ยงเบนมาตรฐานเพื่อ

ดูการกระจายของข้อมูลว่ามีความแปรปรวนมากหรือน้อย และสามารถทำนายได้ว่าเดือนไหนที่คาดว่าจะมีกำไรมากหรือน้อย

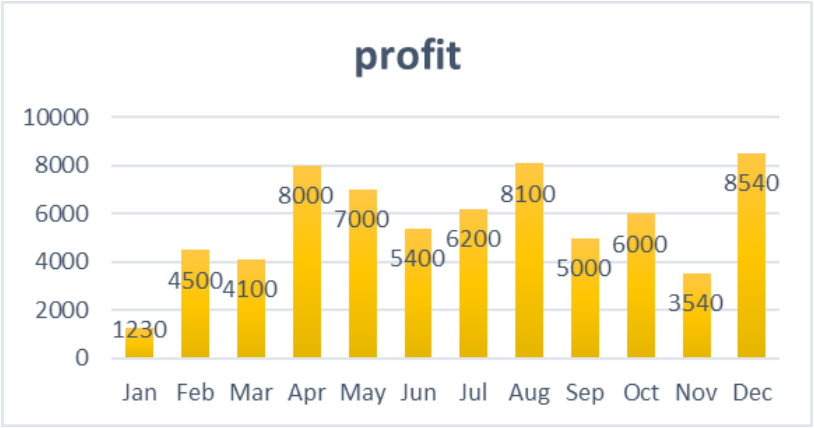
3. การสร้างกราฟแท่ง: เราสามารถสร้างกราฟแท่งเพื่อแสดงจำนวนกำไรในแต่ละเดือนของปี ซึ่งจะช่วยให้เห็นภาพรวมของการทำกำไรในแต่ละเดือนได้อย่างชัดเจนได้ดังภาพที่ 4.3

4. การทำนายแนวโน้ม: จากข้อมูลที่มี เราอาจจะสามารถทำนายแนวโน้มของกำไรในเดือนถัดไปได้โดยใช้เทคนิคการวิเคราะห์และการทำนายทางสถิติ

5. การสรุปผล: หลังจากการวิเคราะห์และการทำนาย สามารถสรุปผลว่าเดือนใดให้กำไรมากที่สุด และเดือนใดให้กำไรน้อยที่สุด และสามารถกำหนดแผนการดำเนินการในอนาคตตามความเหมาะสมได้



ภาพที่ 4.2



ภาพที่ 4.3

month	profit
Jan	1230
Feb	4500
Mar	4100
Apr	8000
May	7000
Jun	5400
Jul	6200
Aug	8100
Sep	5000
Oct	6000
Nov	3540
Dec	8540
AVE	5634.167
S.D	2138.821

จากข้อมูลที่ได้รับ สามารถวิเคราะห์ได้ดังนี้

1. ค่าเฉลี่ย (Average): ค่าเฉลี่ยของกำไรในแต่ละเดือนของปีคือ 5,634.167 บาท

2. ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation): ส่วนเบี่ยงเบนมาตรฐานของกำไรในแต่ละเดือนของปีคือ 2,138.821 บาท ซึ่งหมายถึงความแปรปรวนของข้อมูล หรือการกระจายตัวของข้อมูลอยู่ในระดับใดในทิศทางที่หนึ่ง

3. การวิเคราะห์เพิ่มเติม: ด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ได้จะเป็นประโยชน์ในการทำนายแนวโน้มของกำไรในอนาคต และทำให้เราเข้าใจความแปรปรวนของข้อมูลมากขึ้น เพื่อการวางแผนและการตัดสินใจทางธุรกิจในอนาคต

สรุป

การวิเคราะห์ข้อมูลในสถิติมักใช้สถิติทั้งสถิติบรรยายและสถิติอ้างอิง เพื่อให้เข้าใจลักษณะของข้อมูลและการแจกแจง สถิติบรรยายใช้ในการอธิบายและสรุปข้อมูลตามลักษณะของข้อมูล เช่น เฉลี่ย ค่าสูงสุด ค่าต่ำสุด และการกระจายของข้อมูล ในขณะที่สถิติอ้างอิงใช้ในการทำนายหรือตรวจสอบสมมติฐานเกี่ยวกับประชากร การสร้างและการอ่านผลการวิเคราะห์ข้อมูลช่วยให้เราเข้าใจและนำข้อมูลไปใช้ประโยชน์ได้อย่างมีประสิทธิภาพ โดยสรุปว่าการวิเคราะห์ข้อมูลเป็นกระบวนการสำคัญที่ช่วยให้เราเข้าใจและการบริหารจัดการข้อมูลได้อย่างมีประสิทธิภาพ และสามารถดำเนินการตามเป้าหมายและความต้องการขององค์กรได้อย่างเหมาะสม

แบบฝึกหัด

1. **แบบฝึกหัดที่ 1: การวิเคราะห์ข้อมูลการขายสินค้า**

ให้นักเรียนสร้างแผนภูมิแท่งเพื่อแสดงผลการขายสินค้าในร้านค้า นานา ณ เดือน ต่าง ๆ ของปี โดยใช้ข้อมูลต่อไปนี้

เดือน ยอดขาย (บาท)

มกราคม	25000
กุมภาพันธ์	35000
มีนาคม	40000
เมษายน	48000
พฤษภาคม	42000
มิถุนายน	38000

หลังจากสร้างแผนภูมิแล้ว ให้วิเคราะห์และสรุปผลว่าเดือนใดที่มีการขายสินค้ามากที่สุดและน้อยที่สุด และวิเคราะห์สาเหตุที่อาจเป็นเช่นนั้น

2. **แบบฝึกหัดที่ 2: การทำนายยอดขายในเดือนถัดไป**

ให้นักศึกษาใช้ข้อมูลยอดขายของเดือนทั้ง 6 เดือนแรก เพื่อทำการทำนายยอดขายในเดือนถัดไป โดยใช้วิธีการวิเคราะห์ข้อมูลทางสถิติ เช่น การใช้เทคนิคการคำนวณเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน และสรุปผลว่าการทำนายเป็นไปตามความเป็นจริงหรือไม่

เอกสารอ้างอิง

ทิพย์สิริ กาญจนวาสี และศิริชัย กาญจนวาสี. (2559). *วิธีวิทยาการวิจัย. กรุงเทพมหานคร*: บริษัททรีบี. การพิมพ์และตรายาง

ระพินทร์ โพธิ์ศรี.(2549). *การสร้างและคุณภาพเครื่องมือสาหรับการวิจัย*. อุตรดิตถ์: คณะครุศาสตร์มหาวิทยาลัยราชภัฏอุตรดิตถ์

วุฒิ สุขเจริญ. (2561). *วิจัยการตลาด*. กรุงเทพฯ : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

สุชาติ ประสิทธิ์รัฐสินธุ์. (2555). *ระเบียบวิธีการวิจัยทางสังคมศาสตร์*. พิมพลักษณ์, กรุงเทพฯ : ห้างหุ้นส่วนจำกัด สามลดา

หทัยกาญจน์ ชูตระกูล.(2565).*หลักสถิติเบื้องต้น*.กรุงเทพฯ : ซีเอ็ดยูเคชั่น

บทที่ 5
ตัวแปรสุ่มและการแจกแจง

การศึกษาเกี่ยวกับตัวแปรสุ่มและการแจกแจงเป็นส่วนสำคัญของวิชาสถิติ เนื่องจากเป็นเครื่องมือที่ช่วยให้สามารถวิเคราะห์และทำนายผลของเหตุการณ์ที่มีความไม่แน่นอนได้ โดยทั้งสองเรื่องนี้มีความเกี่ยวข้องกับความน่าจะเป็นที่เป็นหลักฐานทางวิทยาศาสตร์ การศึกษาเกี่ยวกับตัวแปรสุ่มทั้งหมดมีวัตถุประสงค์ในการระบุและการแก้ไขปัญหาที่เกิดขึ้นในโลกที่เต็มไปด้วยความไม่แน่นอนและการประมาณค่าความน่าจะเป็นในรูปแบบต่างๆ เป็นเครื่องมือที่มีประโยชน์ในการทำนายและการตัดสินใจทางธุรกิจ โดยทั้ง ตัวแปรสุ่มไม่ต่อเนื่อง และ ตัวแปรสุ่มต่อเนื่อง เป็นส่วนสำคัญของการศึกษาที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลในสถิติและวิทยาการจำแนกชนิดทั้งสองยังมีการใช้งานอย่างแพร่หลายในการประมาณค่าความน่าจะเป็นในสถิติวิทยาศาสตร์และในการสร้างแบบจำลองทางคณิตศาสตร์ที่มีประโยชน์ในการวิเคราะห์ข้อมูลในหลายสาขาอาชีพและวงการต่าง ๆ ทั้งในภาครัฐและเอกชน

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

ตัวแปรสุ่ม (Random Variable) คือตัวแปรที่สามารถมีค่าหลายค่าได้โดยมีความน่าจะเป็นที่แต่ละค่าเป็นไปได้เท่ากับค่าความน่าจะเป็นที่กำหนดไว้ ตัวแปรสุ่มสามารถแบ่งออกเป็นสองประเภทหลักได้แก่ ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable) และตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable) (อัชฌา อระวีพร, 2562)

- ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable) ลักษณะตัวแปรสุ่มไม่ต่อเนื่องสามารถรับค่าเป็นจำนวนจำกัดหรือจำนวนไม่จำกัดที่สามารถนับได้ (Countable) โดยมีค่าเป็นจำนวนเต็มหรือจำนวนที่นับได้ เช่น 0, 1, 2, 3, ... เช่น จำนวนการโยนลูกเต๋าได้ 1, 2, 3, 4, 5, หรือ 6 จำนวนการทอยลูกเต๋าให้ได้เลขหก จำนวนข้อสอบที่นักเรียนทำถูกต้อง จำนวน

ผู้เข้าร่วมงานในแต่ละวัน จำนวนของผลที่ออกมาจากการโยนเหรียญ, จำนวนคนที่เข้ารับการรักษาที่โรงพยาบาลในแต่ละวัน เป็นต้น

ตัวอย่างของการแจกแจงความน่าจะเป็นในตัวแปรสุ่มไม่ต่อเนื่อง ได้แก่ การแจกแจงเบอร์นูลลี (Bernoulli Distribution) และการแจกแจงทวินาม (Binomial Distribution)

2. ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable) ลักษณะของตัวแปรสุ่มต่อเนื่องสามารถรับค่าใด ๆ ในช่วงหนึ่ง ๆ บนเส้นจำนวนจริง (Real number line) โดยมีค่าเป็นจำนวนที่ไม่จำกัดและไม่สามารถนับได้ เช่น ค่าเป็นตัวเลขที่สามารถมีค่าใด ๆ ระหว่าง 0 และ 1 เช่น 0.1, 0.01, 0.001, ... เช่น อุณหภูมิในเมืองหนึ่งในแต่ละวัน, ความสูงของนักเรียนในชั้นเรียน, หรือเวลาในการวิ่ง 100 เมตร น้ำหนักของสุกร, ความสูงของคน, เวลาที่ใช้ในการทำงาน เป็นต้น

ตัวอย่างของการแจกแจงความน่าจะเป็นในตัวแปรสุ่มต่อเนื่องได้แก่ การแจกแจงปกติ (Normal Distribution) และการแจกแจงปัวซอง (Poisson Distribution)

การเลือกใช้ตัวแปรสุ่มและการแจกแจงที่เหมาะสมขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ในการวิเคราะห์หรือการทดลองที่ต้องการดำเนินการ การเข้าใจและการเลือกใช้ตัวแปรสุ่มและการแจกแจงที่ถูกต้องเป็นสิ่งสำคัญเพื่อให้การวิเคราะห์หรือการทดลองที่ดำเนินการมีความถูกต้องและเชื่อถือได้

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มเป็นส่วนสำคัญที่ใช้ในการวิเคราะห์ข้อมูลทางสถิติ เป็นแบบแจกแจงที่บอกถึงความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงหนึ่งหรือค่าบางค่าที่กำหนดไว้ หากสามารถแบ่งแยกได้ว่าเป็นการแจกแจงประเภทแบบความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Probability Distribution) หรือการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง (Continuous Probability Distribution) จะเป็นประโยชน์และนำไปปรับใช้ได้ การเลือกใช้การแจกแจงความน่าจะเป็นที่

เหมาะสมเป็นส่วนสำคัญเพื่อให้การวิเคราะห์ข้อมูลหรือการสร้างแบบจำลองทางคณิตศาสตร์มีความถูกต้องและเชื่อถือได้ และการเลือกใช้ขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ในการวิเคราะห์หรือการทดลองที่ต้องการดำเนินการ

การประมาณค่าความน่าจะเป็น

การประมาณค่าความน่าจะเป็น (Probability Estimation) เป็นกระบวนการทางสถิติที่ใช้ในการคำนวณหรือประมาณค่าของความน่าจะเป็นของเหตุการณ์หรือผลลัพธ์ที่เป็นไปได้ต่าง ๆ โดยใช้ข้อมูลหรือสถิติฐานที่มีอยู่

วิธีการประมาณค่าความน่าจะเป็นอาจแตกต่างกันไปตามลักษณะของข้อมูลและประเด็นที่ต้องการวิเคราะห์ หลายวิธีการในการประมาณค่าความน่าจะเป็นรวมถึงการใช้การแจกแจงที่เหมาะสม, การใช้ข้อมูลประวัติศาสตร์, การใช้ข้อมูลจำนวนมากเพื่อสร้างโมเดล หรือการใช้เทคนิคการประมาณค่าอื่น ๆ

ตัวอย่างของการประมาณค่าความน่าจะเป็นได้แก่:

1. **การประมาณค่าความน่าจะเป็นในการทดลอง**: การประมาณค่าความน่าจะเป็นของเหตุการณ์หรือผลลัพธ์ที่จะเกิดขึ้นในการทดลองทางวิทยาศาสตร์ โดยใช้ข้อมูลที่ได้จากการทดลอง
2. **การประมาณค่าความน่าจะเป็นในการวิเคราะห์ข้อมูล**: เช่น การใช้การแจกแจงที่เหมาะสมเพื่อประมาณค่าความน่าจะเป็นของเหตุการณ์ที่ไม่รู้จัก หรือการใช้โมเดลทางสถิติเพื่อประมาณค่าความน่าจะเป็น
3. **การประมาณค่าความน่าจะเป็นในการตัดสินใจธุรกิจ**: เช่น การประมาณค่าความน่าจะเป็นของผลกำไรหรือขาดทุนจากการดำเนินงานต่าง ๆ โดยใช้ข้อมูลการเปลี่ยนแปลงของตลาดและภาวะเศรษฐกิจ

ตัวอย่างที่ 5.1

สมมติว่ามีลูกเต๋า 6 หน้า และต้องการหาความน่าจะเป็นที่จะได้หน้าที่มีค่าเท่ากับ 3 หลังจากการโยน

x	1	2	3	4	5	6
P(x)	1/6	1/6	1/6	1/6	1/6	1/6
P(x)	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667

ลูกเต๋า การคำนวณความน่าจะเป็นในกรณีนี้คือการหาสัดส่วนของผลลัพธ์ที่ต้องการ (ได้หน้า 3) และสัดส่วนของผลลัพธ์ทั้งหมด (ทุกหน้าของลูกเต๋า)

วิธีการคำนวณ:

1. หาจำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้ ในกรณีนี้คือ 6 หน้าของลูกเต๋า
2. หาจำนวนของผลลัพธ์ที่ต้องการ ในกรณีนี้คือ หน้าที่มีค่าเท่ากับ 3
3. นำจำนวนของผลลัพธ์ที่ต้องการ หารด้วยจำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้
4. นำผลลัพธ์ที่ได้มาเปลี่ยนเป็นทศนิยม เพื่อหาค่าความน่าจะเป็น

ดังนั้นสามารถคำนวณความน่าจะเป็นในกรณีนี้ได้ดังนี้

1. จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้คือ 6 หน้าของลูกเต๋า
2. จำนวนของผลลัพธ์ที่ต้องการคือ 1 หน้า (หน้าที่มีค่าเท่ากับ 3)
3. ความน่าจะเป็น = (จำนวนของผลลัพธ์ที่ต้องการ) / (จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้) = 1 / 6
4. ความน่าจะเป็น = 0.1667 หรือ 16.67%

ดังนั้น เราสามารถประมาณค่าความน่าจะเป็นที่จะได้หน้า 3 ของลูกเต๋าได้ว่ามีโอกาสประมาณ 16.67% ในการโยกลูกเต๋าหนึ่งครั้ง

ตัวอย่างที่ 5.2

ในกรณีที่มีลูกเต๋า 6 หน้า 2 ลูกและต้องการหาความน่าจะเป็นที่จะได้หน้าที่มีค่าบวกกันเท่ากับ 4 หลังจากการโยนลูกเต๋า การคำนวณความน่าจะเป็นในกรณีนี้จะเหมือนกับการคำนวณในกรณีก่อนหน้า โดยมีขั้นตอนดังนี้:

	1,1	1,2	1,3	1,4	1,5	1,6
	2,1	2,2	2,3	2,4	2,5	2,6

	3,1	3,2	3,3	3,4	3,5	3,6
	4,1	4,2	4,3	4,4	4,5	4,6
	5,1	5,2	5,3	5,4	5,5	5,6
	6,1	6,2	6,3	6,4	6,5	6,6
P(x=4)	3 / 36	1/12	0.0833			

วิธีการคำนวณ:

1. หาจำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้ ในกรณีนี้คือ 6 หน้าของลูกเต๋าต่อลูกเต๋า ซึ่งมีทั้งหมด 6 หน้า ต่อ 2 ลูก ดังนั้นจำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้คือ 6 x 6 = 36
2. หาจำนวนของผลลัพธ์ที่ต้องการ ในกรณีนี้คือ ผลลัพธ์ที่มีผลรวมเท่ากับ 4 โดยมีสองลูกเต๋า ดังนั้นต้องหาคู่ที่เหมาะสมของหน้าของลูกเต๋าที่เมื่อบวกกันแล้วเท่ากับ 4 ซึ่งมีคู่ที่เป็นไปได้ดังนี้:
    - ลูกเต๋าที่หน้า 1 และ 3
    - ลูกเต๋าที่หน้า 2 และ 2
    - ลูกเต๋าที่หน้า 3 และ 1
3. นับจำนวนของคู่ที่เป็นไปได้ ซึ่งในกรณีนี้มีทั้งหมด 3 คู่
4. นำจำนวนของคู่ที่เป็นไปได้ หารด้วยจำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้
5. นำผลลัพธ์ที่ได้มาเปลี่ยนเป็นทศนิยม เพื่อหาค่าความน่าจะเป็น

ดังนั้น เราสามารถคำนวณความน่าจะเป็นในกรณีนี้ได้ดังนี้:

1. จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้คือ 36
2. จำนวนของคู่ที่เป็นไปได้คือ 3 คู่
3. ความน่าจะเป็น = (จำนวนของคู่ที่เป็นไปได้) / (จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้) = 3 / 36
4. ความน่าจะเป็น = 1/12
5. ความน่าจะเป็น = 0.0833 หรือ 8.33%

ดังนั้น สามารถประมาณค่าความน่าจะเป็นที่จะได้ผลรวมเท่ากับ 4 ของลูกเต๋าสองลูกได้ว่ามีโอกาสประมาณ 8.33% ในการโยกลูกเต๋าสองลูกครั้งหนึ่ง.

ตัวอย่างที่ 5.3

ในกรณีที่มีลูกเต๋า 6 หน้า 2 ลูกและโยน 3 ครั้ง และต้องการหาความน่าจะเป็นที่จะได้หน้าที่มีค่าบวกกันเท่ากับ 6 หลังจากการโยนลูกเต๋า การคำนวณความน่าจะเป็นในกรณีนี้จะเหมือนกับการคำนวณในกรณีก่อนหน้า โดยมีขั้นตอนดังนี้:

วิธีการคำนวณ:

1. หาจำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้ โดยการคูณจำนวนหน้าของลูกเต๋า (6 หน้า) กับ จำนวนลูกเต๋า (2 ลูก) และ จำนวนครั้งที่โยนลูกเต๋า (3 ครั้ง) ซึ่งได้ผลลัพธ์ทั้งหมด 6 x 6 x 3 = 108
2. หาจำนวนของผลลัพธ์ที่ต้องการ ซึ่งในกรณีนี้คือ ผลลัพธ์ที่มีผลรวมของหน้าลูกเต๋าทั้งสองลูกเท่ากับ 6
3. หาว่ามีการเกิดผลลัพธ์ที่ต้องการอยู่ในกี่ครั้ง โดยในการโยนลูกเต๋า 3 ครั้ง มีหลายวิธีที่จะได้ผลรวมเท่ากับ 6 ได้แก่:
    - ลูกเต๋าที่หน้า 1, 2, 3
    - ลูกเต๋าที่หน้า 1, 3, 2
    - ลูกเต๋าที่หน้า 2, 1, 3
    - ลูกเต๋าที่หน้า 2, 3, 1
    - ลูกเต๋าที่หน้า 3, 1, 2
    - ลูกเต๋าที่หน้า 3, 2, 1 ดังนั้น มีทั้งหมด 6 วิธีที่จะได้ผลรวมเท่ากับ 6
4. นำจำนวนของผลลัพธ์ที่ต้องการ หารด้วยจำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้
5. นำผลลัพธ์ที่ได้มาเปลี่ยนเป็นทศนิยม เพื่อหาค่าความน่าจะเป็น
ดังนั้น เราสามารถคำนวณความน่าจะเป็นในกรณีนี้ได้ดังนี้:

1. จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้คือ 108
2. จำนวนของผลลัพธ์ที่ต้องการคือ 6 ครั้ง
3. ความน่าจะเป็น = (จำนวนของผลลัพธ์ที่ต้องการ) / (จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้) = 6 / 108
4. ความน่าจะเป็น = 1/18
5. ความน่าจะเป็น = 0.0556 หรือ 5.56%

ดังนั้น เราสามารถประมาณค่าความน่าจะเป็นที่จะได้ผลรวมเท่ากับ 6 ของลูกเต๋าสองลูกจากการโยนลูกเต๋าทั้งหมด 3 ครั้งได้ว่ามีโอกาสประมาณ 5.56% ในการโยนลูกเต๋าทั้งหมด 3 ครั้งครั้งหนึ่งที่ผลรวมเท่ากับ 6 ครั้งนั้นๆ ลองคำนวณดูนะว่าผลลัพธ์เป็นยังไงบ้าง

ตัวแปรสุ่มไม่ต่อเนื่อง

ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable) เป็นตัวแปรที่มีค่าได้เฉพาะค่าจำนวนจำกัดแบบไม่ต่อเนื่อง และมักเกี่ยวข้องกับการนับหรือการเรียงลำดับของเหตุการณ์บางอย่าง โดยตัวแปรสุ่มนี้จะมีค่าเป็นไปได้เฉพาะในช่วงค่าที่แตกต่างกันไปอย่างชัดเจน (ศิริลักษณ์ สุวรรณวงศ์, 2557)

1. **การแจกแจงทวินาม (Binomial Random Variable)** เกี่ยวข้องกับการทดลองที่มีสองผลเป็นไปได้ โดยที่แต่ละผลมีความน่าจะเป็นเท่ากัน ตัวแปรสุ่มนี้บ่งชี้ถึงจำนวนครั้งที่เกิดเหตุการณ์ที่สนใจเกิดขึ้นในจำนวนทั้งหมดของการทดลอง การแจกแจงความน่าจะเป็นทวินาม (Binomial Probability Distribution) เป็นการแจกแจงที่ใช้ในกรณีที่มีการทดลองหลายครั้งและผลลัพธ์ของแต่ละครั้งเป็นสำเร็จหรือไม่สำเร็จ โดยมีสองผลลัพธ์เท่านั้นเรียกว่า "สำเร็จ" และ "ไม่สำเร็จ" โดยที่ความน่าจะเป็นที่เหตุการณ์ "สำเร็จ" เกิดขึ้นในแต่ละครั้งมีค่าคงที่ p และความน่าจะเป็นที่เหตุการณ์ "ไม่สำเร็จ" เกิดขึ้นในแต่ละครั้งมีค่าคงที่ $1-p$ โดยที่ $0 \leq p \leq 1$

สูตรความน่าจะเป็นทวินาม (Binomial Probability) สำหรับจำนวนครั้ง n ที่ทดลอง และจำนวนครั้งที่เหตุการณ์สำเร็จเกิดขึ้น k ครั้ง คือ

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

โดยที่

$P(X=k)$ คือความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าเท่ากับ k

$\binom{n}{k}$เป็นตัวบ่งชี้ทวินาม (Binomial Coefficient) หรือ "n choose k" ซึ่งคำนวณโดย$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ โดยที่ $n!$ คือค่า factorial ของ n

ตัวอย่างการคำนวณการแจกแจงทวินามสมมติว่ามีการทดลองโยนเหรียญ 5 ครั้ง โดยที่โอกาสในการได้หัวในแต่ละครั้งคือ 0.5 (หรือ 50%) และต้องการหาความน่าจะเป็นที่จะได้หัว 3 ครั้ง จะใช้สมการข้างต้นในการคำนวณ

$$P(X=3) = \binom{5}{3} \cdot (0.5)^3 \cdot (1-0.5)^{5-3}$$

$$P(X=3) = \frac{5!}{3!(5-3)!} \cdot (0.5)^3 \cdot (0.5)^2$$

$$P(X=3) = \frac{5\times4\times3!}{3!\times2!} \cdot (0.5)^3 \cdot (0.5)^2$$

$$P(X=3) = \frac{5\times4}{2\times1} \cdot (0.5)^3 \cdot (0.5)^2$$

$$P(X=3) = \frac{5\times4}{2} \cdot (0.5)^3 \cdot (0.5)^2$$

$$P(X=3) = \frac{20}{2} \cdot (0.5)^3 \cdot (0.5)^2$$

$$P(X=3) = 10 \cdot 0.125 \cdot 0.25$$

$$P(X=3) = 10 \times 0.03125$$

$$P(X=3) = 0.3125$$

ดังนั้นความน่าจะเป็นที่จะได้หัว 3 ครั้งจากการทดลองโยนเหรียญ 5 ครั้งคือ 0.3125 หรือ 31.25%

2. **การแจกแจงทวินามลบ (Negative Binomial Distribution)**

การแจกแจงทวินามลบ (Negative Binomial Distribution) เป็นการแจกแจงของตัวแปรสุ่มที่ใช้ในการนับจำนวนครั้งที่เหตุการณ์เกิดขึ้น (จำนวนที่ทำให้เหตุการณ์หนึ่งเกิดขึ้น) จนกว่าจะเกิดเหตุการณ์ที่กำหนดไว้ก่อนหน้านี้ (ตัวแปรนี้เรียกว่า "การสำเร็จ") ณ ขณะที่ทำทดลองครั้งแรกจนถึงเหตุการณ์ที่กำหนดไว้ สามารถใช้ Negative Binomial Distribution ในสถิติเพื่อระบุว่าจะใช้เวลากี่ครั้ง (หรือการทำงานกี่ครั้ง) จึงจะสำเร็จเมื่อมีการสำเร็จในจำนวนครั้งที่กำหนด โดยที่ทราบอัตราการเกิดเหตุการณ์ในแต่ละครั้งสำหรับ Negative Binomial Distribution ค่าความน่าจะเป็นของการสำเร็จคือ p ซึ่งเป็นความน่าจะเป็นของการสำเร็จในแต่ละครั้ง และ r คือจำนวนครั้งที่ต้องการให้เกิดเหตุการณ์สำเร็จ

ฟังก์ชันการความน่าจะเป็น (Probability Mass Function) สำหรับ Negative Binomial Distribution สามารถเขียนได้เป็น

$$P(X = x) = \binom{x+r-1}{x} \cdot (1-p)^r \cdot p^x$$

โดยที่ x=0,1,2,... และ $\frac{(x+r-1)}{x}$) คือผลคูณความน่าจะเป็นของการเลือก x ครั้งจาก x+r−1ครั้ง ซึ่งสามารถคำนวณได้โดยสูตรนี้

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

โดยที่ n! คือค่าแฟกทอเรียลของ n ซึ่งหมายถึงการคูณทุกจำนวนเต็มตั้งแต่ 1 ถึง n ด้วยกัน และ x! คือค่าแฟกทอเรียลของ x ซึ่งหมายถึงการคูณทุกจำนวนเต็มตั้งแต่ 1 ถึง x ด้วยกัน

สำหรับ Negative Binomial Distribution สำหรับค่าความน่าจะเป็นในการเกิดเหตุการณ์ที่สำเร็จคือ p และจำนวนครั้งที่ต้องการให้เกิดเหตุการณ์สำเร็จคือ r สำหรับตัวแปรสุ่ม X ค่าคาดหมาย (Expected Value) หรือ $E(X)$ และค่าความแปรปรวน (Variance) หรือ $Var(X)$ สามารถคำนวณได้ด้วยสูตรต่อไปนี้:

1. ค่าคาดหมาย (Expected Value):

$$E(X) = \frac{r(1-p)}{p}$$

2. ค่าความแปรปรวน (Variance):

$$Var(X) = \frac{r(1-p)}{p^2}$$

ตัวอย่างการคำนวณการแจกแจงทวินามลบ (Negative Binomial Distribution) การทดลองโยนเหรียญ และต้องการหาความน่าจะเป็นที่จะต้องโยนเหรียญอย่างน้อย 3 ครั้งจึงจะได้หัว โดยที่โอกาสในการได้หัวในแต่ละครั้งคือ 0.5 (หรือ 50%)

$P(X=k) = (\frac{\boldsymbol{k-1}}{\boldsymbol{r-1}}) \cdot p^r \cdot (1-p)^{k-r}$

$P(X=k) = (\frac{\boldsymbol{3-1}}{\boldsymbol{3-1}}) \cdot 0.5^3 \cdot (0.5)^{3-3}$

$P(X=3)=0.125$

ดังนั้นความน่าจะเป็นที่จะต้องโยนเหรียญอย่างน้อย 3 ครั้งจึงจะได้หัวคือ 0.125 หรือ 12.5%

3. **การแจกแจงความน่าจะเป็นแบบเอกรูป (Uniform Probability Distribution)** เป็นการแจกแจงที่ทุกค่าในช่วงของตัวแปรสุ่มมีความน่าจะเป็นเท่ากันทุกค่า โดยมีลักษณะเป็นเส้นตรงหรือรูปสี่เหลี่ยม การแจกแจงนี้เป็นสิ่งที่พบได้บ่อยในการแจกแจงของค่าตัวเลขที่ถูกสุ่มได้จากช่วงค่าที่กำหนดไว้ล่วงหน้า เช่น การทอยลูกเต๋าเมื่อทอยลูกเต๋าที่มีหน้าเป็นหมายเลข 1-6, ความน่าจะเป็นที่จะได้หน้าใด ๆ เท่ากันสำหรับทุกหน้า, การสุ่มหมายเลขที่จะได้จากการจับใบต่าง ๆการสุ่มหมายเลขระหว่าง 1 ถึง 10, ความน่าจะเป็นที่จะได้หมายเลขใด ๆ เท่ากันสำหรับทุกหมายเลขในช่วงที่กำหนด และการทดลองต่าง ๆ ที่มีผลลัพธ์ที่เป็นไปได้ทั้งหมดในช่วงเดียวกันโดยมีความน่าจะเป็นเท่ากันทุกค่า

สูตรสำหรับการคำนวณความน่าจะเป็นเอกรูป (Uniform Probability Distribution) ในช่วงค่าที่กำหนดไว้คือ

ถ้ามีช่วงค่าที่กำหนดไว้ระหว่าง a ถึง b และความยาวของช่วงค่า ($b-a$) และต้องการคำนวณความน่าจะเป็นที่เกิดขึ้นในช่วงค่าที่กำหนดไว้ [a,b] สำหรับค่าใดๆ ในช่วงนี้ ความน่าจะเป็นจะเท่ากับ $\frac{1}{b-a}$

สูตร: $P(X=x) = \frac{1}{b-a}$

โดยที่ $P(X=x)$ คือความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าเท่ากับ x

a คือค่าต่ำสุดในช่วงค่าที่กำหนดไว้

b คือค่าสูงสุดในช่วงค่าที่กำหนดไว้

และสำหรับค่า x ที่ไม่อยู่ในช่วงค่าที่กำหนดไว้ [a,b] ความน่าจะเป็นจะเท่ากับ 0 คือ $P(X=x)=0$

สำหรับการแจกแจงความน่าจะเป็นแบบเอกรูป (Uniform Probability Distribution) ที่มีช่วงของค่าทั้งหมดระหว่าง aและ b ค่าคาดหมาย (Expected Value) หรือ $E(x)$ และค่าความแปรปรวน (Variance) หรือ $Var(x)$ สามารถคำนวณได้ดังนี้:

1. ค่าคาดหมาย (Expected Value): $E(x)= \frac{a+b}{2}$
2. ค่าความแปรปรวน (Variance): $Var(x)= \frac{(b-a)^2}{12}$

โดยที่ a และ b คือ ขอบเขตของช่วงค่าของตัวแปรสุ่ม ซึ่งในกรณีของการแจกแจงความน่าจะเป็นแบบเอกรูปคือ ค่าต่ำสุดและค่าสูงสุดของช่วงค่าทั้งหมดที่เป็นไปได้ ตัวแปร x คือ ตัวแปรสุ่มที่มีการแจกแจงแบบเอกรูปในช่วงระหว่าง a และ b

ตัวอย่างเช่น ถ้ามีการทดลองโยนลูกเต๋า 6 หน้า และต้องการหาความน่าจะเป็นที่จะได้หน้าใดๆ คือ 1 หรือ 2 หรือ 3 หรือ 4 หรือ 5 หรือ 6 โดยที่ทุกหน้ามีโอกาสเท่ากัน ความน่าจะเป็นที่จะได้หน้าใด ๆ คือ $\frac{1}{6}$ หรือประมาณ 0.1667 หรือ 16.67%

4. **การแจกแจงเรขาคณิต (Geometric Random Variable)** เกี่ยวข้องกับจำนวนครั้งที่จำเป็นในการทดลองจนถึงครั้งแรกที่เหตุการณ์ต่าง ๆ เกิดขึ้น โดยที่แต่ละครั้งในการทดลองมีความเป็นอิสระต่อกันและมีความน่าจะเป็นเท่ากัน การแจกแจงเรขาคณิต เป็นการแจกแจงของตัวแปรสุ่มที่แสดงถึงจำนวนครั้งที่ต้องทำการทดลองจนกว่าจะได้ผลลัพธ์ครั้งแรกที่ปรากฏสำเร็จ (success) โดยมีคุณสมบัติดังนี้

- การทดลองเรื่อยๆ จนกว่าจะเกิดเหตุการณ์สำเร็จเป็นครั้งแรก
- ทุกครั้งที่ทดลองมีโอกาสสำเร็จและล้มเหลวเท่ากัน
- การทดลองแต่ละครั้งเป็นอิสระต่อกัน

สูตรที่ใช้ในการคำนวณความน่าจะเป็นของการเกิดเหตุการณ์สำเร็จในครั้งแรกในการแจกแจงเรขาคณิตคือ

$$P(X=k) = (1-p)^{k-1} \times p$$

โดยที่ p คือความน่าจะเป็นในการเกิดเหตุการณ์สำเร็จในแต่ละครั้ง และ k คือจำนวนครั้งที่ต้องทดลองจนกว่าจะได้ผลลัพธ์สำเร็จครั้งแรก

การแจกแจงเรขาคณิตมักใช้ในการวิเคราะห์และการคำนวณความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้นเมื่อมีการทดลองซ้ำๆ โดยเฉพาะในรูปแบบของการทดลองจนกว่าจะได้ผลลัพธ์ที่ต้องการ เช่น การทดลองโปรแกรมคอมพิวเตอร์จนกว่าจะเกิดข้อผิดพลาด การรอคอยจนกว่าจะมีคนตอบโทรศัพท์ หรือการทดลองทายสลากกินแบ่ง และอื่นๆ อีกมากมาย

ตัวอย่างการคำนวณความน่าจะเป็นของการเกิดเหตุการณ์สำเร็จในครั้งแรก (success on the first trial) จากการแจกแจงเรขาคณิต โดยให้ค่า p=0.3 ซึ่งหมายถึงโอกาสในการเกิดเหตุการณ์สำเร็จในแต่ละครั้งเท่ากับ 0.3 ต้องการหาความน่าจะเป็นที่เหตุการณ์จะเกิดขึ้นในครั้งแรก (k = 1) ซึ่งสามารถคำนวณได้จากสมการแจกแจงเรขาคณิต:

$$P(X=1) = (1-p)^{1-1} \times p$$

$$P(X=1) = (1-0.3)^{1-1} \times 0.3$$

$P(X=1) = 0.70 \times 0.3$

$P(X=1) = 1 \times 0.3$

$P(X=1) = 0.3$

ดังนั้นความน่าจะเป็นที่เหตุการณ์จะเกิดขึ้นในครั้งแรก (success on the first trial) คือ 0.3 หรือ 30%

5. **การแจกแจงเบอร์นูลลี (Bernoulli Random Variable)** เป็นการแจกแจงที่เกี่ยวข้องกับการทดลองที่มีผลลัพธ์เป็นไปได้เพียงสองประเภท (สำเร็จหรือล้มเหลว) โดยที่ความน่าจะเป็นของผลลัพธ์แต่ละประเภทมีค่าที่แตกต่างกัน ในกรณีที่ตัวแปรสุ่มมีสองค่าเท่านั้น เช่น สำหรับการทดลองทอยเหรียญที่มีแค่สองผลลัพธ์คือ "สำเร็จ" หรือ "ไม่สำเร็จ" โดยที่ความน่าจะเป็นที่เกิดเหตุการณ์ "สำเร็จ" คือ p และความน่าจะเป็นที่เกิดเหตุการณ์ "ไม่สำเร็จ" คือ $1-p$ โดยที่ $0 \leq p \leq 1$ เป็นความน่าจะเป็นที่เหตุการณ์ "สำเร็จ" เกิดขึ้น และ $1-p$ เป็นความน่าจะเป็นที่เหตุการณ์ "ไม่สำเร็จ" เกิดขึ้น

สูตร: $P(X=x) = p^x \cdot (1-p)^{1-x}$

โดยที่

- $P(X=x)$ คือความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าเท่ากับ x ซึ่งในกรณีของแบบเบอร์นูลลี ค่า x จะเป็น 0 หรือ 1 เท่านั้น
- p คือความน่าจะเป็นที่เหตุการณ์ "สำเร็จ" เกิดขึ้น โดยที่ $0 \leq p \leq 1$
- $1-p$ คือความน่าจะเป็นที่เหตุการณ์ "ไม่สำเร็จ" เกิดขึ้น
- x มีค่าเท่ากับ 0 หรือ 1 เพื่อแทนเหตุการณ์ "ไม่สำเร็จ" หรือ "สำเร็จ" ตามลำดับ

6. **การแจกแจงปัวซอง (Poisson Random Variable)** เกี่ยวข้องกับการนับจำนวนเหตุการณ์ที่เกิดขึ้นในช่วงเวลาหรือพื้นที่ที่กำหนดไว้ล่วงหน้า โดยที่อัตราการเกิดเหตุการณ์เฉลี่ยในหนึ่งหน่วยเวลาหรือพื้นที่เป็นค่าคงที่ การแจกแจงความน่าจะเป็นปัวซอง (Poisson Probability Distribution) เป็น

การแจกแจงที่ใช้ในการนับจำนวนการเกิดเหตุการณ์ในช่วงเวลาหนึ่ง โดยที่อยู่ในช่วงเวลานั้นเหตุการณ์แต่ละครั้งเกิดอิสระต่อกันและมีอัตราการเกิดเหตุการณ์เฉลี่ยในหนึ่งหน่วยเวลาเท่ากับ λ โดยที่ λ เป็นค่าคงที่บวก สูตรความน่าจะเป็นปัวซอง (Poisson Probability) สำหรับจำนวนการเกิดเหตุการณ์ k ในช่วงเวลาหนึ่ง คือ

$$P(X=k) = \frac{e^{-\lambda} \cdot \lambda^{k}}{k!} \quad , k = 0,1,2,...$$

โดยที่

$P(X=k)$ คือความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าเท่ากับ k

e คือค่าคงที่ของ Euler ($e \approx 2.71828$)

λ คืออัตราการเกิดเหตุการณ์เฉลี่ยในหนึ่งหน่วยเวลา

k คือจำนวนการเกิดเหตุการณ์ที่นับได้ในช่วงเวลาหนึ่ง

$k!$ คือค่า factorial ของ k

Example ถ้าจำนวนเด็กในภาคกลางที่เป็นโรคเลือด ในรอบ 10 ปีที่ผ่าน มาโดยเฉลี่ยเป็น 4 คน จงหาเปอร์เซ็นต์ที่เด็กในภาคกลางจะเป็นโรคเลือด 6คน วิธีทำ ให้ x เป็นจำนวนเด็กที่ที่เป็นโรคเลือด $X \sim P(4,4)$ จากตารางแสดงค่าของ $e^{-\lambda}$ เมื่อ $\lambda = 4$,$e^{-\lambda} = 0.01832$

$P(X = 6) = \frac{e^{-4} \cdot \lambda^{6}}{6!}$ 0.1042 นั้นคือเปอร์เซ็นต์ที่เด็กภาคกลางเป็นโรคเลือด 6 คน $= 10.42$

ตัวแปรสุ่มต่อเนื่อง

ตัวแปรสุ่มต่อเนื่อง Continuous Random Variable ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable) เป็นตัวแปรที่สามารถมีค่าได้ในช่วงค่าที่ไม่มีจำนวนค่ามากๆ โดยตัวแปรเหล่านี้มักเกี่ยวข้องกับข้อมูลที่

สามารถวัดได้เป็นเลขทศนิยมหรือจำนวนที่มีความต่อเนื่อง เช่น น้ำหนัก, ความสูง, อุณหภูมิ, เวลา, ราคาหุ้น เป็นต้น

ตัวอย่างของตัวแปรสุ่มต่อเนื่องได้แก่ การแจกแจงปกติ (Normal Distribution) การแจกแจงแบบเอ็กซ์โปเนนเชียล (Exponential Distribution) การแจกแจงแบบเบต้า (Beta Distribution) การแจกแจงแบบแกมมา (Gamma Distribution) การแจกแจงแบบแปรผันต่อเนื่อง (Continuous Uniform Distribution) การแจกแจงแบบล็อกนอร์มอล (Log-Normal Distribution) เป็นต้น

1. การแจกแจงปกติ (Normal Distribution): การแจกแจงที่มีลักษณะทรงกราบกระจายแบบกราฟกระจายเดี่ยวที่มีรูปร่างคล้ายกราฟกระจายที่เป็นรูปโค้งโค้ง (bell-shaped curve) โดยมักจะมีค่าสูงสุดที่ค่าเฉลี่ยและกระจายตัวคล้ายกัน ใช้ในการวิเคราะห์และระบุลักษณะการกระจายของข้อมูลในหลายสาขาวิชา เช่น ในการวิเคราะห์การทดลองทางการแพทย์, การวิเคราะห์ข้อมูลการเงิน, การวิเคราะห์สถิติทางเศรษฐศาสตร์ เป็นต้น

การแจกแจงปกติ (Normal Distribution) เป็นหนึ่งในการแจกแจงที่สำคัญที่สุดในสถิติ มักใช้ในการแสดงการกระจายของข้อมูลที่มีลักษณะเป็นกระจายเส้นตรง โดยมีลักษณะเด่นดังนี้

- มีค่าคาดหมาย (Mean) และค่าความแปรปรวน (Variance) ที่กำหนดรูปร่างของการแจกแจง
- มีรูปร่างคล้ายกระจายเบลล์ (bell-shaped curve) หรือรูปร่างที่เรียบสวยและมักเป็นไปตามแบบกำลังสอง
- การกระจายตัวของข้อมูลจะมีค่ามากที่สุดที่ค่าคาดหมายและลดลงเรื่อย ๆ ไปทางสองข้าง
- หากกระจายของข้อมูลมีค่าคาดหมายและค่าความแปรปรวนเท่ากับ 0 จะได้การแจกแจงที่เป็นเส้นตรง

สูตรการของการแจกแจงปกติคือ

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

โดยที่ μ คือค่าคาดหมาย (mean) และ σ^2 คือค่าความแปรปรวน (variance)

การคำนวณค่าของการแจกแจงปกตินั้นมักใช้ตารางค่าZ (z-table) หรือโปรแกรมคำนวณทางสถิติ เนื่องจากคำนวณโดยตรงในสมการข้างต้นอาจจะทำได้ยาก ดังนั้นการแจกแจงปกตินั้นเป็นที่นิยมใช้เป็นการแจกแจงที่ใช้ในการวิเคราะห์ข้อมูลและการทดลองทางสถิติในชีวิตประจำวันและงานวิจัยทางวิทยาศาสตร์

ตารางค่าZ เป็นตารางที่ใช้ในการหาค่าของฟังก์ชันการความน่าจะเป็นสำหรับการแจกแจงปกติมาตรฐาน (standard normal distribution) โดยมีการระบุค่าของพารามิเตอร์ Z ซึ่งเป็นระยะห่างระหว่างค่าของตัวแปรสุ่มและค่าคาดหมายในหน่วยของความแปรปรวนของการแจกแจงปกติมาตรฐาน ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ 1

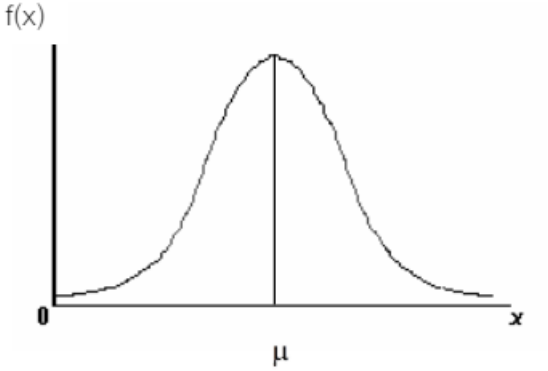
การใช้ z-table นั้นสามารถหาค่าความน่าจะเป็นในการมองเห็นต่าง ๆ ของการแจกแจงปกติมาตรฐานได้ เช่น ความน่าจะเป็นที่ต่ำกว่าค่าที่กำหนด (cumulative probability) หรือการหาค่าที่ตรงกับความน่าจะเป็นที่กำหนด (percentile) (หทัยกาญจน์ ชูตระกูล, 2565)

ตัวอย่างของการใช้ z-table คือการหาค่าความน่าจะเป็นที่ต่ำกว่าค่า Z = 1.96 ซึ่งแทนค่าความน่าจะเป็นสะสม (cumulative probability) ของการแจกแจงปกติมาตรฐานที่มีค่า Z น้อยกว่าหรือเท่ากับ 1.96 โดยจาก z-table เราสามารถหาค่านี้ได้เป็น 0.9750 หรือ 97.50%

การใช้ z-table เป็นเครื่องมือที่สำคัญในการวิเคราะห์ข้อมูลที่มีการแจกแจงปกติมาตรฐาน โดยมักนำมาใช้ในการทดสอบสมมติฐาน (hypothesis

testing) และการวิเคราะห์ทางสถิติในหลาย ๆ สาขาวิชาที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลและการทดลองทางสถิติ

โดยทั่วไปจะใช้สัญลักษณ์ X ~ N (μ,σ^2) แทน ตัวแปรสุ่ม X มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ μ และ σ^2 ถ้าเขียนกราฟของฟังก์ชั่นความหนาแน่นของ X เมื่อ X ~ N (μ,σ^2) จะ ได้กราฟลักษณะ ดังนี้



จะเห็นว่า ลักษณะของกราฟเป็นรูประฆังคว่ำ และโดยทั่วไปนิยมเรียกกราฟของ f(x) ของการแจกแจงแบบปกติว่า โค้งปกติ (normal curve)

ค่า Z ในการคำนวณสำหรับการใช้ z-table

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

โดยที่

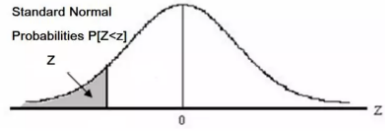
- X คือค่าของตัวแปรที่ต้องการคำนวณ
- μ คือค่าคาดหมาย (mean) ของการแจกแจง
- σ คือค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ของการแจกแจง

เมื่อคำนวณค่า Z แล้ว คุณสามารถใช้ค่า Z นี้เพื่อหาค่าความน่าจะเป็น (probability) ที่เกี่ยวข้องกับค่า X โดยใช้ z-table หรือฟังก์ชันความน่าจะเป็นของการแจกแจงปกติมาตรฐาน (standard normal distribution)

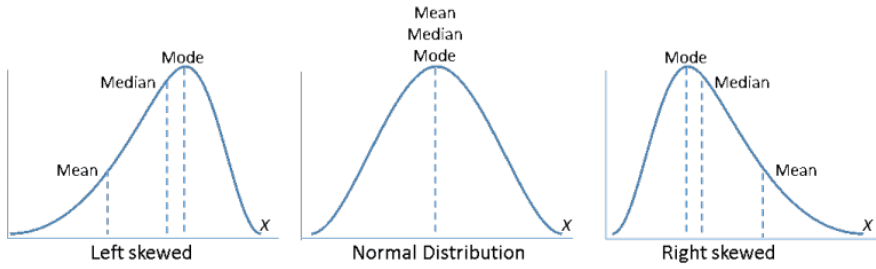
ตัวอย่างการคำนวณค่าพื้นที่ใต้กราฟ Z:

1. กำหนดให้ X=1.96 , μ=0 , σ=1
2. คำนวณค่า Z: $Z = \frac{1.96-0}{1}$ =1.96
3. ใช้ z-table เพื่อหาค่าความน่าจะเป็นที่มากกว่าหรือเท่ากับ 1.96 ซึ่งจะได้ค่าประมาณ 0.9750 หรือ 97.50%
4. ดังนั้น, พื้นที่ใต้กราฟของการแจกแจงปกติมาตรฐานที่มีค่า Z ไม่เกิน 1.96 คือ 0.9750 หรือ 97.50%

ตารางสถิติ

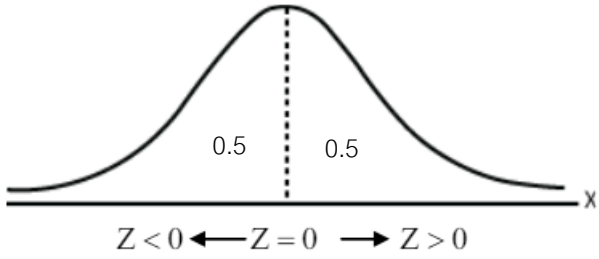


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

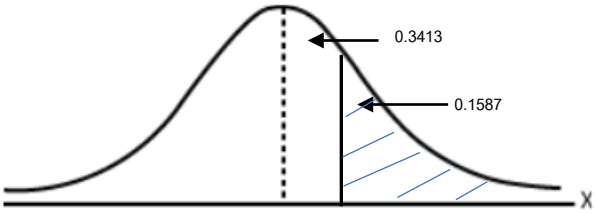


ในการแจกแจงแบบเบ้ในเชิงบวกหรือเบ้ขวา ค่าเฉลี่ยมักจะมากกว่าค่ามัธยฐาน ซึ่งมากกว่าค่าฐานนิยมเสมอ (ค่าเฉลี่ย > ค่ามัธยฐาน > ฐานนิยม) เป็นเพราะค่าผิดปกติทางด้านขวาของการกระจาย (ตัวเลขที่มากกว่า) ดึงค่าเฉลี่ยขึ้น ตัวอย่างเช่น การกระจายของรายรับ หากต้องรวบรวมรายรับต่อปีของทุกคนในประเทศ A อาจพบว่าคนส่วนใหญ่มีรายได้ค่อนข้างน้อย ในขณะที่คนจำนวนน้อยมีรายรับมากกว่ามาก ทำให้การกระจายของข้อมูลส่วนใหญ่กระจุกตัวอยู่ทางซ้าย (รายได้ต่ำ) โดยมีหางยาวเหยียดออกไปทางขวา (รายรับสูง)

ในการแจกแจงแบบเบ้เชิงลบหรือเบ้ซ้าย ค่าเฉลี่ยมักจะน้อยกว่าค่ามัธยฐาน ซึ่งน้อยกว่าค่าฐานนิยมเสมอ ค่าเฉลี่ย < ค่ามัธยฐาน < ฐานนิยม) เพราะค่าผิดปกติทางด้านซ้ายของการกระจาย (ตัวเลขที่น้อยกว่า) ดึงค่าเฉลี่ยลง ตัวอย่างเช่น คะแนนการสอบ นักศึกษาส่วนใหญ่จะทำคะแนนได้เกือบ 100% โดยมีคะแนนต่ำกว่าอยู่เล็กน้อย สิ่งนี้สร้างการกระจายที่เบ้ไปทางซ้าย ข้อมูลส่วนใหญ่จะกระจุกตัวอยู่ทางขวา (คะแนนสูงกว่า) โดยมีหางยาวไปทางซ้าย (คะแนนต่ำกว่า)



พื้นที่ทั้งหมดทางขวามือ = 0.5 เท่ากับ พื้นที่ทั้งหมดทางซ้ายมือ = 0.5



ดังนั้น $P(Z \leq 1) = 0.5 + P(0 < Z < 1) = 0.5 + 0.3413 = 0.8413$

และ $P(Z \geq 1) = 0.5 - P(0 < Z < 1) = 0.5 - 0.3413 = 0.1587$

ตัวอย่าง ถ้า X มีการแจกแจงแบบปกติ ที่มี μ = 2 และ σ = 2 จงหาความน่าจะเป็น ที่ X จะมีค่าระหว่าง 3 และ 5 นั้นคือ X ~ N (2,4) ต้องการหา $P(3 < X < 5)$

เมื่อ $X1 = \frac{3-2}{2}$ = 0.5

เมื่อ $X1 = \frac{5-2}{2}$ = 1.5

นั้นคือ $P(3 < X < 5) = P(0.5 < Z < 1.5)$

คือพื้นที่ระหว่าง Z = 0.5 และ Z = 1.5

0.9332 - 0.6915 = 0.2417

2. การแจกแจงที (t-distribution) เป็นการแจกแจงที่ใช้ในการประมาณค่าของค่าเฉลี่ยของตัวแปรสุ่มต่อเนื่องเมื่อขนาดของกลุ่มตัวอย่างเล็ก ๆ (n) และค่าของความแปรปรวนของประชากรไม่รู้จัก มันจะมีรูปร่างคล้ายกับการแจกแจงปกติ แต่มีความแปรปรวนที่มากกว่าเล็กน้อย t-distribution มักถูกใช้ในการสร้างสถิติที่เกี่ยวกับค่าเฉลี่ยของกลุ่มตัวอย่างที่มีขนาดเล็ก โดยเฉพาะในการสร้างค่าความเชื่อมั่น (confidence intervals) หรือการทดสอบสมมติฐาน (hypothesis testing) ที่มีตัวอย่างน้อย

ในการคำนวณค่าที่เกี่ยวข้องกับการแจกแจงที (t-distribution) จำเป็นต้องใช้ค่าระดับความเชื่อมั่น (confidence level) และขนาดของ

ตัวอย่าง (sample size) เพื่อคำนวณหาค่า t-score หรือค่า t-critical ที่ใช้ในการตัดสินใจในการทดสอบสมมติฐาน

สูตรสำหรับการคำนวณค่า t-score หรือค่า t-critical คือ

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

โดยที่

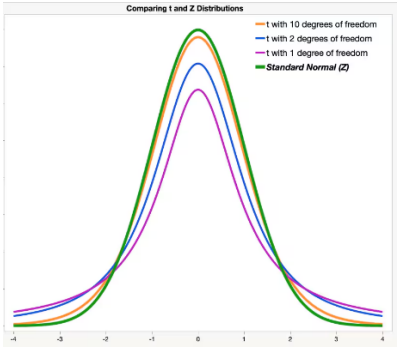
$\bar{X}$ คือค่าเฉลี่ยของตัวอย่าง

μ คือค่าคาดหมายของประชากร

s คือค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

n คือขนาดของตัวอย่าง

นอกจากนี้ t-distribution ยังมีคุณสมบัติที่มีประโยชน์ในการประมาณค่าความแปรปรวนของประชากรเมื่อขนาดของกลุ่มตัวอย่างมีขนาดน้อย โดยเฉพาะเมื่อการประมาณค่าความแปรปรวนโดยใช้ข้อมูลจากกลุ่มตัวอย่างมีขนาดน้อยและตัวอย่างมีค่าความแปรปรวนที่แตกต่างจากค่าความแปรปรวนของประชากรจริงๆ



3. การแจกแจงไคสแควร์ (Chi-Square Distribution) เป็นการแจกแจงที่ใช้ในการประมาณค่าความแปรปรวนของตัวแปรสุ่มต่อเนื่อง โดยที่ข้อมูลที่ใช้ในการประมาณนั้นเป็นข้อมูลที่ไม่ต่อเนื่อง (discrete data) และมักจะมีลักษณะเป็นการแจกแจงที่มีความแปรปรวนแบบเชี่ยวชาญ (skewed distribution) โดยมีค่าของความแปรปรวนขึ้นอยู่กับจำนวนอิสระ (degrees of freedom) Chi-Square Distribution มักถูกใช้ในการทำสถิติทางการแพทย์, การวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการทดลองทางวิทยาศาสตร์, การวิเคราะห์การทดสอบความเที่ยงตรงของข้อมูล, การทดสอบสมมติฐานทางสถิติ เป็นต้น การคำนวณค่า Chi-Square สามารถทำได้โดยใช้สูตรที่เหมาะสมกับความต้องการ โดยมักใช้โปรแกรมทางสถิติหรือเครื่องมือการคำนวณสถิติในการดำเนินการนี้

4. การแจกแจงเอฟ (F-distribution) เป็นการแจกแจงที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับความแปรปรวนของสองกลุ่มตัวอย่าง โดยการทดสอบว่าความแปรปรวนของสองกลุ่มนั้นมีความแตกต่างกันหรือไม่ หรือใช้ในการตรวจสอบความแตกต่างของความแปรปรวนระหว่างกลุ่มตัวอย่างมากกว่าสองกลุ่ม การแจกแจงเอฟมักถูกใช้ในการทดสอบสมมติฐานในการวิเคราะห์การเชื่อมโยงระหว่างตัวแปร หรือในการทดสอบความแปรปรวนระหว่างกลุ่มตัวอย่าง โดยเฉพาะในการทดสอบสมมติฐานเกี่ยวกับความแตกต่างของความแปรปรวนระหว่างกลุ่มที่มีขนาดต่างกัน สำหรับการคำนวณค่าเอฟ, มักจะใช้ซอฟต์แวร์สถิติหรือเครื่องมือทางสถิติที่มีอยู่ เนื่องจากการคำนวณค่าเอฟจำเป็นต้องใช้การคำนวณที่ซับซ้อนและต้องใช้ค่าความแตกต่างระหว่างความแปรปรวนของกลุ่มตัวอย่างที่ทดสอบ

5. การแจกแจงแบบเอ็กซ์โปเนนเชียล (Exponential Distribution): การแจกแจงที่มีลักษณะทรงกราบกระจายที่น้อยลงเรื่อย ๆ ตามแนวแกน x เมื่อ x เพิ่มขึ้น ใช้ในการวิเคราะห์เวลาที่เกิดเหตุการณ์ เช่น เวลาที่รอรถโดยสารถึงสถานี, เวลาที่รอในการบริการของลูกค้า หรือเวลาการส่งมอบสินค้า

การแจกแจงแบบเบต้า (Beta Distribution):
การแจกแจงที่มีลักษณะการกระจายแบบทวิภาค คือ มีรูปร่างคล้ายกราฟที่มีลักษณะของการแกว่ง (skewness) ไปทางขวาหรือทางซ้ายของแกน x ใช้ในการวิเคราะห์ความน่าจะเป็นในการเกิดเหตุการณ์ที่มีความแปรปรวนเชิงความเสี่ยง เช่น ความเสี่ยงในการลงทุนทางการเงิน, การวางแผนการผลิตสินค้า, หรือการวิเคราะห์ความน่าจะเป็นในการเปิดตัวสินค้าใหม่

6. การแจกแจงแบบแกมมา (Gamma Distribution): การแจกแจงที่มีลักษณะการกระจายที่มีทั้งการแกว่งและการแตกต่างของความหนาแน่นของค่าในช่วงต่าง ๆ ของแกน x ใช้ในการวิเคราะห์เวลาที่เกิดเหตุการณ์ที่มีลักษณะการเรียงต่อเนื่อง เช่น เวลาที่รอการเดินทางระหว่างเมือง, เวลาที่รอการจัดส่งสินค้า เป็นต้น

7. การแจกแจงแบบแปรผันต่อเนื่อง (Continuous Uniform Distribution): การแจกแจงที่มีลักษณะการกระจายที่มีความหนาแน่นคงที่ในช่วงค่าที่กำหนดไว้ล่วงหน้า ใช้ในการวิเคราะห์ข้อมูลที่มีความแปรปรวนต่ำและมีค่าเท่าๆ กันในช่วงที่กำหนด เช่น การกระจายของราคาสินค้า, การกระจายของระยะเวลาที่ใช้ในการซื้อสินค้าออนไลน์

8. การแจกแจงแบบล็อก-นอร์มอล (Log-Normal Distribution): การแจกแจงที่มีลักษณะการกระจายที่ผลลัพธ์ของการนำล็อกฐาน e ของตัวแปรสุ่มแบบปกติ มีการกระจายแบบปกติ ใช้ในการวิเคราะห์ข้อมูลที่มีการกระจายที่ผิดปกติและมีการแตกต่างของการกระจายระหว่างกลุ่มข้อมูล ตัวอย่างเช่น การกระจายของรายได้ของประชากร, การกระจายของขนาดของโรคในประชากร เป็นต้น

สรุป

การศึกษาเกี่ยวกับตัวแปรสุ่มและการแจกแจงเป็นส่วนสำคัญของวิชาสถิติ เพราะช่วยให้สามารถวิเคราะห์และทำนายผลของเหตุการณ์ที่มีความไม่แน่นอนได้ ตัวแปรสุ่มแบ่งออกเป็นสองประเภทหลัก: ตัวแปรสุ่มไม่ต่อเนื่อง

(เช่น จำนวนการโยนลูกเต๋า) และตัวแปรสุ่มต่อเนื่อง (เช่น ความยาว, น้ำหนัก) ตัวอย่างของการแจกแจงในตัวแปรสุ่มไม่ต่อเนื่อง ได้แก่ การแจกแจงเบอร์นูลลี และการแจกแจงทวินาม ส่วนตัวอย่างของการแจกแจงในตัวแปรสุ่มต่อเนื่อง ได้แก่ การแจกแจงปกติและการแจกแจงเอ็กซ์โปเนนเชียล การเลือกใช้ตัวแปรสุ่มและการแจกแจงที่เหมาะสมขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ในการวิเคราะห์ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มช่วยให้การวิเคราะห์ข้อมูลทางสถิติมีความถูกต้องและเชื่อถือได้ การประมาณค่าความน่าจะเป็น (Probability Estimation) เป็นกระบวนการทางสถิติที่ใช้ในการคำนวณหรือประมาณค่าของความน่าจะเป็นของเหตุการณ์หรือผลลัพธ์ที่เป็นไปได้ โดยใช้ข้อมูลหรือสถิติฐานที่มีอยู่ ตัวอย่างของการประมาณค่าความน่าจะเป็น ได้แก่ การประมาณค่าความน่าจะเป็นในการทดลองทางวิทยาศาสตร์ การใช้การแจกแจงที่เหมาะสมในการวิเคราะห์ข้อมูล และการประมาณค่าความน่าจะเป็นในการตัดสินใจธุรกิจ การเลือกใช้วิธีการที่เหมาะสมเป็นสิ่งสำคัญเพื่อให้การวิเคราะห์ข้อมูลหรือการสร้างแบบจำลองทางคณิตศาสตร์มีความถูกต้องและเชื่อถือได้

VDO ศึกษาประกอบ

1. https://media.stou.ac.th/view_video.php?act=&vid=23641
2. https://media.stou.ac.th/view_video.php?act=&vid=23642
3. https://media.stou.ac.th/view_video.php?act=&vid=23643

เอกสารอ้างอิง

ศิริลักษณ์ สุวรรณวงศ์. (2557). *สถิติศาสตร์*. กรุงเทพฯ : ซีเอ็ดยูเคชั่น

อัชฌา อระวีพร. (2562). *ความน่าจะเป็น*. กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย

หทัยกาญจน์ ชูตระกูล.(2565).*หลักสถิติเบื้องต้น*.กรุงเทพฯ : ซีเอ็ดยูเคชั่น

บทที่ 6
การประมาณค่า

ในการวิเคราะห์ข้อมูลสถิติ การประมาณค่าเป็นกระบวนการสำคัญที่ช่วยในการสรุปและตัดสินใจจากข้อมูลตัวอย่างที่มีอยู่ โดยทั่วไป การประมาณค่าสามารถแบ่งออกได้เป็นสองวิธีหลัก ๆ คือ การประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation) การประมาณค่าแบบจุดคือการเลือกค่าตัวเลขหนึ่งค่าจากข้อมูลตัวอย่างมาเป็นตัวแทนของค่าพารามิเตอร์ที่ต้องการในประชากร ขณะที่การประมาณค่าแบบช่วงจะให้ช่วงของค่าที่เป็นไปได้สำหรับพารามิเตอร์นั้น โดยมีระดับความเชื่อมั่นที่กำหนด การใช้ทั้งสองวิธีนี้ร่วมกันช่วยเพิ่มความแม่นยำและความน่าเชื่อถือในการวิเคราะห์ข้อมูลและการตัดสินใจในงานวิจัยต่าง ๆ

การประมาณค่า

การประมาณค่าเป็นกระบวนการที่สำคัญในสถิติและวิทยาการข้อมูล ซึ่งใช้เพื่อคาดการณ์หรือประเมินค่าพารามิเตอร์ของประชากรจากข้อมูลตัวอย่างที่มีอยู่ กระบวนการนี้สามารถแบ่งออกเป็นสองวิธีหลัก ได้แก่ การประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation) การประมาณค่าแบบจุดเกี่ยวข้องกับการเลือกค่าตัวเลขหนึ่งค่าเพื่อเป็นตัวแทนของค่าพารามิเตอร์ที่ต้องการในประชากร ขณะที่การประมาณค่าแบบช่วงให้ช่วงของค่าที่เป็นไปได้สำหรับพารามิเตอร์นั้น พร้อมกับระดับความเชื่อมั่นที่กำหนดไว้ วิธีการทั้งสองนี้ช่วยให้นักวิจัยสามารถวิเคราะห์และตีความข้อมูลได้อย่างมีประสิทธิภาพมากขึ้น และเป็นพื้นฐานในการตัดสินใจที่มีความถูกต้องในงานวิจัยและการประยุกต์ใช้งานในด้านต่าง ๆ

การประมาณค่ามีประโยชน์อย่างมากในหลายด้านของการวิจัยและการประยุกต์ใช้ในสาขาต่าง ๆ ดังนี้:

1. **การตัดสินใจและการวางแผน**: การประมาณค่าช่วยให้นักวิจัยและผู้ประกอบการสามารถทำการตัดสินใจที่มีข้อมูลรองรับ โดยใช้ค่าประมาณที่ได้จากตัวอย่างเพื่อคาดการณ์พฤติกรรมของประชากร ซึ่งช่วยในการวางแผนกลยุทธ์และการบริหารจัดการอย่างมีประสิทธิภาพ

2. **การวิเคราะห์ข้อมูลและการสรุปผล**: การประมาณค่าช่วยในการสรุปข้อมูลที่มีอยู่ในรูปแบบที่เข้าใจง่ายขึ้น ทำให้นักวิจัยสามารถตีความข้อมูลและนำเสนอผลลัพธ์ที่ชัดเจนและเชื่อถือได้

3. **การทำนายและคาดการณ์**: การประมาณค่าช่วยในการทำนายแนวโน้มและพฤติกรรมในอนาคต เช่น การคาดการณ์ยอดขาย การประเมินความเสี่ยง และการวิเคราะห์ความต้องการของลูกค้า ซึ่งเป็นพื้นฐานในการพัฒนากลยุทธ์และการตัดสินใจ

4. **การตรวจสอบสมมติฐาน**: การประมาณค่าช่วยในการตรวจสอบสมมติฐานทางสถิติ โดยให้ข้อมูลที่จำเป็นในการทดสอบความเป็นไปได้ของสมมติฐานที่ตั้งขึ้น เช่น การทดสอบความแตกต่างระหว่างกลุ่มตัวอย่างหรือการทดสอบความสัมพันธ์ระหว่างตัวแปร

5. **การประเมินประสิทธิภาพและการปรับปรุง**: การประมาณค่าช่วยในการประเมินประสิทธิภาพของกระบวนการหรือระบบ และเป็นเครื่องมือในการหาข้อบกพร่องหรือจุดที่ต้องปรับปรุงเพื่อเพิ่มประสิทธิภาพและคุณภาพของผลลัพธ์

6. **การจัดการทรัพยากร**: การประมาณค่าช่วยในการประเมินปริมาณทรัพยากรที่จำเป็นต้องใช้ เช่น งบประมาณ แรงงาน หรือวัตถุดิบ ซึ่งช่วยให้สามารถจัดสรรทรัพยากรได้อย่างเหมาะสมและมีประสิทธิภาพ

7. **การสนับสนุนการวิจัยและพัฒนา**: การประมาณค่าช่วยให้นักวิจัยสามารถประเมินผลกระทบของการเปลี่ยนแปลงหรือการปรับปรุงต่าง ๆ และช่วยในการตัดสินใจในการพัฒนาโครงการใหม่ ๆ หรือการปรับปรุงกระบวนการที่มีอยู่

การใช้การประมาณค่าอย่างถูกต้องและเหมาะสมจะช่วยเพิ่มความน่าเชื่อถือและความแม่นยำในการวิเคราะห์ข้อมูลและการตัดสินใจในทุกด้านของการวิจัยและการปฏิบัติงาน (ศิริลักษณ์ สุวรรณวงศ์, 2557)

การประมาณค่าแบบจุด (Point Estimation)

การประมาณค่าแบบจุด (Point Estimation) เป็นวิธีการในสถิติที่ใช้เพื่อคาดการณ์ค่าพารามิเตอร์ของประชากรโดยอาศัยข้อมูลจากตัวอย่างที่มีอยู่ การประมาณค่าแบบจุดนี้จะให้ค่าตัวเลขหนึ่งค่าเป็นตัวแทนของค่าพารามิเตอร์ที่ต้องการในประชากร ตัวอย่างเช่น การคาดการณ์ค่าเฉลี่ย (Mean), ค่ามัธยฐาน (Median), หรือค่าความแปรปรวน (Variance) ของประชากร โดยการใช้ตัวอย่างที่มี

ตัวอย่างของการประมาณค่าแบบจุด

1. **ค่าเฉลี่ยตัวอย่าง (Sample Mean, $\bar{x}$)**: การคำนวณค่าเฉลี่ยจากข้อมูลตัวอย่าง เพื่อประมาณค่าเฉลี่ยประชากร (μ) ตัวอย่างเช่น หากมีข้อมูลตัวอย่างของคะแนนสอบของนักเรียน 5 คน คือ 85, 90, 78, 92 และ 88 คะแนน ค่าเฉลี่ยตัวอย่าง ($\bar{x}$) จะคำนวณได้ดังนี้

$$\bar{x} = \frac{85+90+78+92+88}{5} = 86.6$$

ดังนั้น 86.6 คือการประมาณค่าแบบจุดของค่าเฉลี่ยประชากร

2. **สัดส่วนตัวอย่าง (Sample Proportion, $\hat{p}$)**: การคำนวณสัดส่วนจากข้อมูลตัวอย่างเพื่อประมาณสัดส่วนประชากร (p) เช่น ในการสำรวจความคิดเห็นของผู้คน 100 คนเกี่ยวกับผลิตภัณฑ์หนึ่ง หาก 60 คนตอบว่าชอบผลิตภัณฑ์นั้น สัดส่วนตัวอย่าง ($\hat{p}$) จะคำนวณได้ดังนี้:

$\hat{p}$ =60/100=0.6

ดังนั้น 0.6 คือการประมาณค่าแบบจุดของสัดส่วนประชากร

คุณสมบัติของตัวประมาณที่ดี

ตัวประมาณที่ดีควรมีคุณสมบัติสำคัญดังนี้

1. **ไม่มีอคติ (Unbiased)**: ตัวประมาณควรมีค่าคาดหวังเท่ากับค่าพารามิเตอร์ที่ต้องการ เช่น ค่าเฉลี่ยตัวอย่าง ($\bar{x}$) เป็นตัวประมาณที่ไม่มีอคติของค่าเฉลี่ยประชากร (μ) เพราะค่าคาดหวังของ $\bar{x}$ เท่ากับ μ

2. **มีประสิทธิภาพ (Efficient)**: ตัวประมาณควรมีความแปรปรวนน้อยที่สุดเมื่อเทียบกับตัวประมาณอื่นๆ ที่ไม่มีอคติ เช่น ค่าเฉลี่ยตัวอย่างมักมีความแปรปรวนน้อยกว่าค่ามัธยฐานตัวอย่าง

3. **มีความสม่ำเสมอ (Consistent)**: ตัวประมาณควรมีแนวโน้มที่จะเข้าใกล้ค่าพารามิเตอร์ที่แท้จริงเมื่อขนาดตัวอย่างเพิ่มขึ้น

การใช้การประมาณค่าแบบจุดในงานวิจัย

การประมาณค่าแบบจุดเป็นเครื่องมือที่สำคัญในงานวิจัยและการวิเคราะห์ข้อมูล โดยช่วยให้เราสามารถคาดการณ์ค่าพารามิเตอร์ที่สนใจได้อย่างรวดเร็วและง่ายดาย ตัวอย่างเช่น ในการวิจัยด้านการตลาด บางครั้งอาจต้องการประมาณค่าเฉลี่ยการใช้จ่ายของลูกค้าในแต่ละเดือน หรือในงานวิจัยทางการแพทย์ เราอาจต้องการประมาณค่าเฉลี่ยระดับน้ำตาลในเลือดของกลุ่มผู้ป่วย

การใช้การประมาณค่าแบบจุดร่วมกับการประมาณค่าแบบช่วง (Interval Estimation) สามารถเพิ่มความน่าเชื่อถือในการวิเคราะห์ข้อมูลและการตัดสินใจได้อย่างมาก เนื่องจากการประมาณค่าแบบช่วงจะให้ข้อมูลเพิ่มเติมเกี่ยวกับความไม่แน่นอนและความแปรปรวนของค่าที่ประมาณได้

การประมาณค่าแบบช่วง

การประมาณค่าแบบช่วง (Interval Estimation) เป็นวิธีการในสถิติที่ใช้เพื่อคาดการณ์ค่าพารามิเตอร์ของประชากรโดยการกำหนดช่วงของค่าที่มี

ความเป็นไปได้ว่าจะครอบคลุมค่าพารามิเตอร์นั้น พร้อมกับระดับความเชื่อมั่นที่กำหนด วิธีนี้จะให้ข้อมูลที่มีความไม่แน่นอนและความเชื่อมั่น ซึ่งมากกว่าการประมาณค่าแบบจุด (Point Estimation) ที่ให้เพียงค่าตัวเลขเดียว (หทัยกาญจน์ ชูตระกูล, 2565)

การประมาณค่าแบบช่วงจะให้ช่วงของค่าที่คาดว่าจะครอบคลุมค่าพารามิเตอร์ที่สนใจ โดยทั่วไป ช่วงนี้จะเรียกว่า **ช่วงความเชื่อมั่น (Confidence Interval, CI)** ช่วงความเชื่อมั่นจะประกอบด้วยขอบล่างและขอบบน ซึ่งเป็นค่าที่สร้างขึ้นจากข้อมูลตัวอย่างและระดับความเชื่อมั่นที่เลือก (เช่น 95% หรือ 99%)

ตัวอย่างการประมาณค่าแบบช่วง

ตัวอย่างการหาช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยหรือการประมาณค่าเฉลี่ยประชากร (μ)

กรณีกลุ่มตัวอย่างมากกว่าหรือเท่ากับ 30

การประมาณค่าเฉลี่ยประชากรด้วยการประมาณค่าแบบช่วง (confidence interval) เป็นเทคนิคทางสถิติที่ใช้ในการประเมินค่าเฉลี่ยของประชากร โดยอาศัยค่าจากตัวอย่างที่สุ่มมาจากประชากรนั้นๆ ในกรณีที่ขนาดตัวอย่าง (n) มีค่าเท่ากับหรือมากกว่า 30 จะสามารถใช้การแจกแจงปกติ (Normal Distribution) ในการคำนวณช่วงความเชื่อมั่นได้ โดย:

1) ตัวอย่างที่สุ่มมาเป็นการสุ่มอย่างแท้จริงจากประชากร

2) ขนาดตัวอย่างมีค่าเท่ากับหรือมากกว่า 30 (n >= 30) ซึ่งเพียงพอที่จะใช้การแจกแจงปกติ

ขั้นตอนการคำนวณช่วงความเชื่อมั่น

1) หาค่าเฉลี่ยของตัวอย่าง (Sample Mean, $\bar{x}$)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

2) หาค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง (Sample Standard Deviation, s)

$$S= \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$

3) เลือกระดับความเชื่อมั่น (Confidence Level, CL) เช่น 95% หรือ 99%

4) หาค่าที่สอดคล้องกับระดับความเชื่อมั่นในตาราง Z (Z-value)

- สำหรับ CL = 95%, Z=1.96
- สำหรับ CL = 99%, Z=2.576

5) คำนวณค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย (Standard Error, SE)

$$SE = \frac{s}{\sqrt{n}}$$

6) คำนวณช่วงความเชื่อมั่น (Confidence Interval, CI)

$$CI= \bar{x} \pm Z \times SE$$

ตัวอย่าง

สมมติว่ามีข้อมูลตัวอย่างขนาด n=50 โดยมีค่าเฉลี่ยของตัวอย่าง $\bar{x}$=100 และค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง s=15และต้องการหาช่วงความเชื่อมั่นที่ระดับ 95%

ค่าเฉลี่ยของตัวอย่าง $\bar{x}$ =100, ค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง: s=15, ระดับความเชื่อมั่น: 95%, ค่า Z สำหรับระดับความเชื่อมั่น 95%: Z=1.96, คำนวณค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย = SE = $\frac{15}{\sqrt{50}}$ $\approx$2.12

คำนวณช่วงความเชื่อมั่น CI= 100 ± 1.96 × 2.12 = 100± 4.15

CI= [95.85,104.15]

ดังนั้นช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยประชากรจะอยู่ระหว่าง 95.85 ถึง 104.15

กรณีกลุ่มตัวอย่างน้อยกว่า 30

การประมาณค่าเฉลี่ยของประชากรที่ไม่สามารถหาค่าความแปรปรวนของประชากรได้และขนาดกลุ่มตัวอย่างน้อยกว่า 30 เพราะค่าใช้จ่ายในการรวบรวมข้อมูลไม่เอื้ออำนวยที่จะหาขนาดกลุ่มตัวอย่างใหญ่เท่าที่ต้องการได้ จึงต้องพิจารณาว่าประชากรชุดที่เกี่ยวข้องจะต้องมีรูปร่างโดยประมาณเป็นรูประฆัง ช่วงความเชื่อมั่นจะหาได้โดยอาศัยการแจกแจงความน่าจะเป็นของสถิติ t (ภรณี เจริญภักตร์และคณะ, 2555)

สมมติว่าต้องการประมาณช่วงความเชื่อมั่น 95% สำหรับค่าเฉลี่ยน้ำหนักของนักเรียนในโรงเรียนจากข้อมูลตัวอย่าง 10 คน มีค่าเฉลี่ย ($\bar{x}$) เท่ากับ 60.2 กิโลกรัม และส่วนเบี่ยงเบนมาตรฐานตัวอย่าง (s) เท่ากับ 6.31 กิโลกรัม

1. คำนวณค่า t ที่สอดคล้องกับระดับความเชื่อมั่น 95% และ df (degrees of freedom) เท่ากับ 9 (n-1): ค่า $t_{0.025,9}$ คือ 2.262
2. ใช้สูตรช่วงความเชื่อมั่นสำหรับค่าเฉลี่ย:

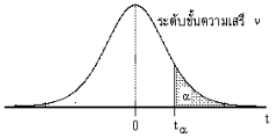
$$\bar{x} \pm t_{\alpha/2,\, df} \cdot \frac{s}{\sqrt{n}}$$

$$60.2 \pm 2.262 \cdot \frac{6.31}{\sqrt{10}}$$

$$60.2 \pm 4.52$$

ดังนั้น ช่วงความเชื่อมั่น 95% สำหรับค่าเฉลี่ยประชากรจะอยู่ระหว่าง 55.68 ถึง 64.72 กิโลกรัม ซึ่งหมายความว่าเรามั่นใจ 95% ว่าค่าเฉลี่ยจริงของน้ำหนักนักเรียนทั้งโรงเรียนจะอยู่ในช่วงนี้

ตารางที่ 4. ตารางค่าวิกฤตของการแจกแจง t
เมื่อกำหนดระดับขั้นความเสรี ν และค่า α
ค่าจากตารางคือ t_α คือค่าที่ทำให้ $P(t_\alpha < T < \infty) = \alpha$



ν	α				
	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845

ตัวอย่างช่วงความเชื่อมั่นสำหรับสัดส่วน

การสำรวจความคิดเห็นของผู้คน 200 คนเกี่ยวกับผลิตภัณฑ์หนึ่ง และพบว่า 150 คนชอบผลิตภัณฑ์นี้

1. คำนวณสัดส่วนตัวอย่าง ($\hat{p}$):

$$\hat{p} = \frac{150}{200} = 0.75$$

2. คำนวณช่วงความเชื่อมั่น 95% สำหรับสัดส่วน (p) โดยใช้สูตร:

$$\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ค่า $Z_{0.025}$ คือ 1.96

$$0.75 \pm 1.96 \cdot \sqrt{\frac{0.75(0.25)}{200}}$$

$$0.75 \pm 0.061$$

ดังนั้น ช่วงความเชื่อมั่น 95% สำหรับสัดส่วนประชากรที่ชอบผลิตภัณฑ์นี้จะอยู่ระหว่าง 0.689 ถึง 0.811 หรือ 68.9% ถึง 81.1%

ประโยชน์ของการประมาณค่าแบบช่วง

1. เพิ่มความน่าเชื่อถือ: การให้ช่วงของค่าที่เป็นไปได้พร้อมกับระดับความเชื่อมั่นทำให้การประมาณค่ามีความน่าเชื่อถือมากขึ้น เพราะแสดงถึงความไม่แน่นอนที่มีในข้อมูล

2. การวางแผนและการตัดสินใจ: ช่วงความเชื่อมั่นช่วยให้ผู้วางแผนและผู้ตัดสินใจมีข้อมูลที่ครบถ้วนมากขึ้น สามารถประเมินความเสี่ยงและความเป็นไปได้ได้ดียิ่งขึ้น

3. การทดสอบสมมติฐาน: ช่วงความเชื่อมั่นสามารถใช้ในการทดสอบสมมติฐานทางสถิติ โดยการตรวจสอบว่าค่าพารามิเตอร์ที่ตั้งสมมติฐานไว้ตกอยู่ในช่วงความเชื่อมั่นหรือไม่

การประมาณค่าแบบช่วงเป็นวิธีที่สำคัญและทรงพลังใน การคาดการณ์ค่าพารามิเตอร์ของประชากรจากข้อมูลตัวอย่าง โดยให้ช่วงของค่าที่เป็นไปได้พร้อมกับระดับความเชื่อมั่น ช่วยให้นักวิจัยและผู้ตัดสินใจสามารถวิเคราะห์และตีความข้อมูลได้อย่างมีประสิทธิภาพมากขึ้น และทำให้การตัดสินใจมีข้อมูลรองรับที่ชัดเจนและเชื่อถือได้

สรุป

การประมาณค่าเป็นกระบวนการสำคัญที่ช่วยในการสรุปและตัดสินใจจากข้อมูลตัวอย่างที่มีอยู่ โดยทั่วไป การประมาณค่าสามารถแบ่งออกได้เป็นสองวิธีหลัก ๆ คือ การประมาณค่าแบบจุดและการประมาณค่าแบบช่วง การประมาณค่าแบบจุดคือการเลือกค่าตัวเลขหนึ่งค่าจากข้อมูลตัวอย่างมาเป็นตัวแทนของค่าพารามิเตอร์ที่ต้องการในประชากร ขณะที่การประมาณค่าแบบช่วงจะให้ช่วงของค่าที่เป็นไปได้สำหรับพารามิเตอร์นั้น โดยมีระดับความเชื่อมั่นที่กำหนด การใช้ทั้งสองวิธีนี้ร่วมกันช่วยเพิ่มความแม่นยำและความน่าเชื่อถือใน การวิเคราะห์ข้อมูลและการตัดสินใจในงานวิจัยต่าง ๆ นอกจากนี้ การประมาณค่ายังช่วยให้เราสามารถประเมินความไม่แน่นอนของ

ข้อมูล ทำให้สามารถวางแผนและตัดสินใจได้ดีขึ้นในสถานการณ์ที่มีความเสี่ยงและไม่แน่นอน

แบบฝึกหัด

แบบฝึกหัดที่ 1: การประมาณค่าแบบจุด (Point Estimation)

จากข้อมูลที่ได้ทำการสำรวจประชากรผู้ใช้บริการห้องสมุดในมหาวิทยาลัยแห่งหนึ่ง พบว่ามีการเก็บข้อมูลจำนวนชั่วโมงที่นักศึกษาใช้บริการห้องสมุดในแต่ละสัปดาห์ ดังนี้: 5, 7, 8, 6, 9, 4, 7, 8, 5, 6

1.1 คำนวณค่าเฉลี่ย (mean) ของจำนวนชั่วโมงที่นักศึกษาใช้บริการห้องสมุดในแต่ละสัปดาห์ 1.2 คำนวณส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ของจำนวนชั่วโมงที่นักศึกษาใช้บริการห้องสมุดในแต่ละสัปดาห์

แบบฝึกหัดที่ 2: การประมาณค่าแบบช่วง (Interval Estimation)

จากการสำรวจค่าใช้จ่ายเฉลี่ยในการซื้อหนังสือเรียนของนักศึกษามหาวิทยาลัย พบว่ามีค่าเฉลี่ยอยู่ที่ 2000 บาท และส่วนเบี่ยงเบนมาตรฐานอยู่ที่ 300 บาท จากกลุ่มตัวอย่าง 50 คน

2.1 กำหนดช่วงความเชื่อมั่น 95% (ใช้ค่า Z = 1.96) สำหรับค่าใช้จ่ายเฉลี่ยในการซื้อหนังสือเรียนของนักศึกษาในประชากร 2.2 กำหนดช่วงความเชื่อมั่น 99% (ใช้ค่า Z = 2.576) สำหรับค่าใช้จ่ายเฉลี่ยในการซื้อหนังสือเรียนของนักศึกษาในประชากร

เอกสารอ้างอิง

ภรณี เจริญภักตร์และคณะ. (2555). *ความน่าจะเป็นและสถิติ*. กรุงเทพฯ : โรง พิมพ์ห้างหุ้นส่วนจำกัดพิทักษ์การพิมพ์

ศิริลักษณ์ สุวรรณวงศ์. (2557). *สถิติศาสตร์*.กรุงเทพฯ : ซีเอ็ดยูเคชั่น

หทัยกาญจน์ ชูตระกูล.(2565).*หลักสถิติเบื้องต้น*.กรุงเทพฯ : ซีเอ็ดยูเคชั่น

เฉลยแบบฝึกหัดที่ 1:

1.1 ค่าเฉลี่ย (mean) = (5 + 7 + 8 + 6 + 9 + 4 + 7 + 8 + 5 + 6) / 10 = 6.5

1.2 ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) คำนวณได้จากสูตร

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

เฉลยแบบฝึกหัดที่ 2:

2.1 ช่วงความเชื่อมั่น 95% สำหรับค่าใช้จ่ายเฉลี่ยในการซื้อหนังสือเรียน:

$$\bar{x} \pm Z \times \frac{\sigma}{\sqrt{n}} = 2000 \pm 1.96 \times \frac{300}{\sqrt{50}} = 2000 \pm 83.16$$

ดังนั้นช่วงความเชื่อมั่นคือ [1916.84, 2083.16] บาท

2.2 ช่วงความเชื่อมั่น 99% สำหรับค่าใช้จ่ายเฉลี่ยในการซื้อหนังสือเรียน:

$$\bar{x} \pm Z \times \frac{\sigma}{\sqrt{n}} = 2000 \pm 2.576 \times \frac{300}{\sqrt{50}} = 2000 \pm 109.22$$

ดังนั้นช่วงความเชื่อมั่นคือ [1890.78, 2109.22] บาท

บทที่ 7
การทดสอบสมมติฐาน

การทดสอบสมมติฐานเป็นกระบวนการทางสถิติที่สำคัญในการวิเคราะห์ข้อมูลและการตัดสินใจในหลากหลายสถานการณ์ เช่น การวิจัยทางการแพทย์ การพัฒนาผลิตภัณฑ์ใหม่ หรือการปรับปรุงกระบวนการทางธุรกิจ โดยมีวัตถุประสงค์เพื่อทำให้สามารถตัดสินใจเกี่ยวกับประเด็นหรือสรุปผลลัพธ์ที่มีการสนับสนุนหรือไม่สนับสนุนสมมติฐานนั้น ๆ อย่างมั่นใจและมีเหตุผล ในบทนี้เป็นการสร้างพื้นฐานเบื้องต้นเกี่ยวกับการทดสอบสมมติฐาน โดยนำเสนอขั้นตอนหลักๆ ของกระบวนการนี้ รวมถึงความสำคัญของการตีความผลลัพธ์ที่ได้จากการทดสอบ เพื่อให้ผู้อ่านเข้าใจและนำไปใช้ในการวิเคราะห์ข้อมูลและการตัดสินใจในงานที่เกี่ยวข้องกับสถิติวิทยาได้อย่างมีประสิทธิภาพและแม่นยำ ตั้งแต่การตั้งสมมติฐาน การเลือกวิธีการทดสอบที่เหมาะสม จนถึงการวิเคราะห์ผลลัพธ์อย่างถูกต้อง

สมมติฐาน

สมมติฐาน (Hypothesis) คือการสร้างข้อความหรือการประมาณค่าที่เกี่ยวข้องกับปัญหาหรือสถานการณ์ที่ต้องการศึกษาหรือวิจัย เพื่อที่จะทำการทดสอบหรือตรวจสอบความถูกต้องของข้อความนั้นๆ โดยใช้ข้อมูลที่มีอยู่

สมมติฐานทางสถิติ: เป็นการกำหนดค่าพารามิเตอร์ของประชากรหรือข้อมูลที่มีความเกี่ยวข้องกับปัญหาวิจัย เพื่อที่จะทำการทดสอบว่าค่านั้นๆ มีค่าเฉลี่ยหรือการกระจายที่แตกต่างกันอย่างมีนัยสำคัญหรือไม่ เช่น "ค่าเฉลี่ยของความสูงของชาวบ้านในเขตนี้มีค่าเท่ากับ 170 เซนติเมตร" หรือ "ความผันผวนของราคาหุ้นบริษัท XYZ มีค่ามาตรฐานเท่ากับ 5%

ประเภทของสมมติฐาน

สมมติฐานสามารถแบ่งออกเป็นหลายประเภทตามลักษณะและลักษณะของข้อมูลที่กำลังทดสอบ บางประเภทที่พบบ่อย ๆ ได้แก่:

1. สมมติฐานว่างหรือกลาง (Null Hypothesis, $H0$): เป็นสมมติฐานที่ระบุว่าไม่มีความแตกต่างหรือไม่มีผลกระทบของตัวแปรอิสระต่อตัวแปรตาม โดยทั่วไปจะเขียนแทนด้วย $H0$ เช่น $H0$:μ=50 หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่กำลังศึกษามีค่าเท่ากับ 50

2. สมมติฐานทดสอบ (Alternative Hypothesis, $H1$ หรือ Ha): เป็นสมมติฐานที่กำหนดความแตกต่างหรือผลกระทบของตัวแปรอิสระต่อตัวแปรตาม ซึ่งเป็นส่วนที่สนใจในการวิจัย เช่น $H1$: μ>50 หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่กำลังศึกษามีค่ามากกว่า 50

3. สมมติฐานสองทาง (Two-Tailed Hypothesis): เป็นสมมติฐานที่กำหนดความแตกต่างหรือผลกระทบของตัวแปรอิสระต่อตัวแปรตามทั้งสองทิศทาง เช่น $H1$: μ≠50 หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่กำลังศึกษาไม่เท่ากับ 50

4. สมมติฐานทางเดียว (One-Tailed Hypothesis) หมายถึงสมมติฐานที่กำหนดค่าของตัวแปรตามที่สนใจให้มากกว่าหรือน้อยกว่าค่าที่กำหนดไว้ โดยมีทิศทางเดียวเท่านั้น สามารถแบ่งสมมติฐานทางเดียวเป็นสองประเภทย่อยได้ดังนี้

สมมติฐานทางเดียวแบบบวก (One-Tailed Hypothesis - Positive): เป็นสมมติฐานที่กำหนดว่าค่าของตัวแปรที่สนใจมากกว่าหรือมากกว่าค่าที่กำหนด ในทิศทางเดียวเท่านั้น เช่น H 1 : μ > 50

หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่กำลังศึกษามีค่ามากกว่า 50

สมมติฐานทางเดียวแบบลบ (One-Tailed Hypothesis - Negative): เป็นสมมติฐานที่กำหนดว่าค่าของตัวแปรที่สนใจน้อยกว่าหรือน้อยกว่าค่าที่กำหนด ในทิศทางเดียวเท่านั้น เช่น $H1$: μ < 50

หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่กำลังศึกษามีค่าน้อยกว่า 50

การทดสอบสมมติฐานทางสถิติ

การทดสอบสมมติฐานทางสถิติเป็นกระบวนการที่สำคัญในการวิเคราะห์ข้อมูลและตัดสินใจในหลากหลายสถานการณ์ เช่น การวิจัยทางการแพทย์ การพัฒนาผลิตภัณฑ์ หรือการปรับปรุงกระบวนการทางธุรกิจ การทดสอบสมมติฐานมีเป้าหมายเพื่อทำให้เราสามารถตัดสินใจเกี่ยวกับประเด็นหรือสรุปผลลัพธ์ที่มีการสนับสนุนหรือไม่สนับสนุนสมมติฐานนั้น ๆ อย่างมั่นใจและมีเหตุผลอย่างเป็นรูปธรรม

ขั้นตอนการทดสอบสมมติฐานทางสถิติปกติมีดังนี้:

1. **ตั้งสมมติฐาน (Formulate Hypotheses)**:

I. สมมติฐานที่เราต้องการทดสอบมักจะแบ่งเป็นสมมติฐานที่เรียกว่าสมมติฐานว่าง (null hypothesis, $H0$) และสมมติฐานที่เรียกว่าสมมติฐานทดสอบ (alternative hypothesis, $H1$ หรือ Ha).

II. สมมติฐานว่าง (null hypothesis) ส่วนใหญ่จะเป็นการสร้างสมมติฐานว่าไม่มีผลกระทบหรือไม่มีความแตกต่าง เช่น $H0 : \mu=50$ หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่สนใจมีค่าเท่ากับ 50.

III. สมมติฐานทดสอบ (alternative hypothesis) จะเป็นการระบุค่าที่เราสนใจว่ามีผลกระทบหรือความแตกต่าง เช่น $H1 : \mu > 50$ หมายความว่าค่าเฉลี่ยของกลุ่มหรือประชากรที่สนใจมีค่ามากกว่า 50

2. **เลือกสถิติที่ต้องการทดสอบ (Select Test Statistic)**:

การเลือกสถิติที่ต้องการทดสอบจะขึ้นอยู่กับลักษณะของข้อมูล และสมมติฐานที่ต้องการทดสอบ เช่น หากต้องการทดสอบค่าเฉลี่ยของกลุ่มตัวอย่างสามารถใช้ t-test หรือ z-test ได้ และหากต้องการทดสอบความแตกต่างของสัดส่วนสองกลุ่มสามารถใช้ chi-square test หรือ Fisher's exact test ได้ การเลือกใช้สถิติที่ใช้ทดสอบจะขึ้นอยู่กับวัตถุประสงค์และเงื่อนไขของสถิติที่เหมาะสมกับข้อมูล

3. **กำหนดระดับความสำคัญ (Set Significance Level)**:

ระดับความสำคัญ (significance level) บ่งบอกถึงความเสี่ยงที่ยอมรับในการทำนิยามผิดพลาดเพื่อปฏิเสธสมมติฐานที่เสนอ ระดับความสำคัญที่ใช้มักเป็น 0.05 หรือ 0.01.

4. **เก็บข้อมูลและคำนวณค่าสถิติ (Collect Data and Calculate Test Statistic)**:

นำข้อมูลที่ได้มาจากการสำรวจหรือการทดลอง และใช้วิธีการทางสถิติที่เลือกไว้เพื่อคำนวณค่าสถิติที่เกี่ยวข้อง เช่น t-value, z-value, chi-square statistic, หรือ F-value.

5. **ตีความผลการทดสอบ (Interpret Results)**:

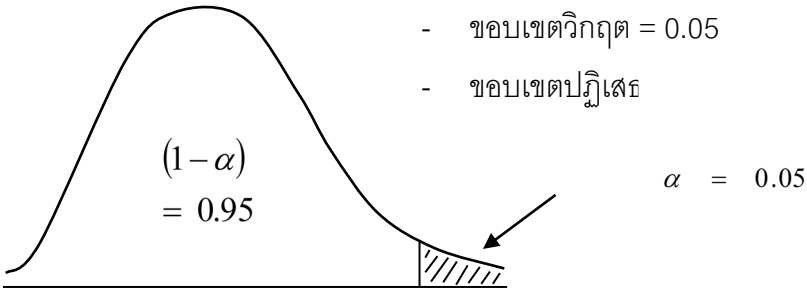
ตีความผลลัพธ์ที่ได้จากการทดสอบโดยเปรียบเทียบค่าสถิติที่คำนวณได้กับค่าที่คาดหวัง และใช้ค่า p-value เพื่อตัดสินใจว่าจะปฏิเสธหรือยอมรับสมมติฐานว่าง (null hypothesis) โดยปกติถ้า p-value น้อยกว่าระดับความสำคัญที่กำหนด (significance level) จะปฏิเสธสมมติฐานว่าง.

6. **สรุปผลลัพธ์ (Conclusion)**:

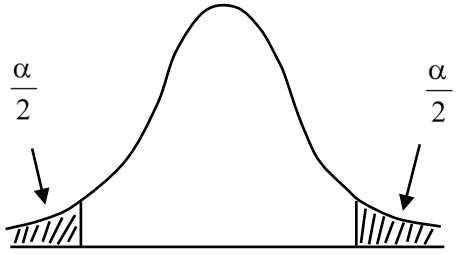
สรุปผลลัพธ์ของการทดสอบสมมติฐานโดยรวมและแสดงความเชื่อมั่นในการสรุป และอธิบายผลกระทบหรือความสำคัญของผลลัพธ์ต่องานหรือวิชาที่เกี่ยวข้อง

การตีความผลการทดสอบ

การทดสอบสมมติฐาน คือการหาข้อสรุปว่าสมมติฐานที่สนใจนั้นคือสมมติฐานว่าง (H_0) จะถูกยอมรับหรือปฏิเสธ ถ้าถูกยอมรับ นั่นก็แปลว่าสมมติฐานทดสอบที่ตั้งไว้เป็นจริง หากว่าสมมติฐานทดสอบถูกปฏิเสธแปลว่าสมมติฐานนั้นไม่เป็นจริง ดังนั้นผู้วิจัยจึงต้องมีการกำหนดค่าระดับนัยสำคัญเพื่อที่จะกำหนดเขตวิกฤตในการยอมรับหรือปฏิเสธสมมติฐาน ตามรูปภาพที่ 7.1 (สมมติฐานแบบมีทิศทางเดียว) และ รูปภาพที่ 7.2 (สมมติฐานแบบมีสองทิศทาง)



ภาพที่ 7.1 : ขอบเขตวิกฤตที่ระดับนัยสำคัญ .05 แบบมีทิศทางเดียว



ภาพที่ 7.2 : ขอบเขตวิกฤตที่ระดับนัยสำคัญแบบมีสองทิศทาง

หากค่าสถิติที่คำนวณได้มีค่ามากกว่าค่าวิกฤตที่ระดับนัยสำคัญ .05 หมายความว่า สมมติฐานที่ทดสอบมีนัยสำคัญทางสถิติ

สำหรับการวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติให้พิจารณาจากค่า P-value หรือค่า sig น้อยกว่าค่า α ที่ระบุไว้หมายความว่า สมมติฐานที่ทดสอบมีนัยสำคัญทางสถิติ

One-sample t-test

การทดสอบนี้ใช้เพื่อเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างหนึ่งกลุ่มกับค่าเฉลี่ยที่เป็นค่าคงที่ที่กำหนดขึ้น (เช่น ค่าเฉลี่ยของประชากรที่คาดหวัง) (ปรีดาภรณ์ กาญจนสำราญวงศ์,2560) โดยมีขั้นตอนดังนี้

1. **กำหนดสมมติฐาน**

H_0 : ค่าเฉลี่ยของกลุ่มตัวอย่างเท่ากับค่าคงที่ (เช่น $\mu=\mu_0$)

H_1: ค่าเฉลี่ยของกลุ่มตัวอย่างไม่เท่ากับค่าคงที่ (เช่น $\mu \neq \mu_0$)

2. **คำนวณค่า t**

สูตรการคำนวณค่า t สำหรับ One-sample t-test คือ

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

โดยที่ $\bar{x}$ คือ ค่าเฉลี่ยของกลุ่มตัวอย่าง

μ_0 คือ ค่าคงที่ที่เราต้องการเปรียบเทียบ

s คือ ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

n คือ จำนวนตัวอย่าง

3. **เปรียบเทียบกับค่า t วิกฤต**

เมื่อคำนวณค่า t ได้แล้ว ให้เปรียบเทียบกับค่า t วิกฤตจากตาราง t distribution ที่ระดับความเชื่อมั่นที่กำหนด (เช่น 0.05) และ df = n-1

หาก $|t|$ ที่คำนวณได้ มากกว่าค่า t วิกฤต ปฏิเสธสมมติฐาน H_0 และสรุปว่าค่าเฉลี่ยของกลุ่มตัวอย่างแตกต่างจากค่าคงที่

ตัวอย่างการใช้ One-sample t-test

สมมติว่าต้องการทดสอบว่านักเรียนกลุ่มหนึ่งมีค่าเฉลี่ยคะแนนสอบเท่ากับ 75 หรือไม่ โดยมีข้อมูลดังนี้

ค่าเฉลี่ยของกลุ่มตัวอย่าง ($\bar{x}$) = 78

ค่าคงที่ที่เราต้องการเปรียบเทียบ (μ_0) = 75

ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (s) = 10

จำนวนตัวอย่าง (n) = 30

คำนวณค่า $t = \frac{78-75}{\frac{10}{\sqrt{30}}} = \frac{3}{1.83} \approx 1.64$

ถ้าค่า t วิกฤตที่ระดับความเชื่อมั่น 0.05 สำหรับ df=29 คือ 2.045

เนื่องจาก $|1.64| < 2.045$ จึงไม่สามารถปฏิเสธสมมติฐาน H_0 ได้ ซึ่งหมายความว่าค่าเฉลี่ยของนักเรียนกลุ่มนี้ไม่แตกต่างจาก 75 อย่างมีนัยสำคัญ

ตารางการแจกแจง t

df	0.1 0.2	0.05 0.1	0.025 0.05	0.02 0.04	0.015 0.03	0.01 0.02	0.005 0.01	0.0025 0.005	0.0005 0.001	One-tail Two-tail
1	3.0777	6.3137	12.7062	15.8945	21.2051	31.8210	63.6559	127.3211	636.5776	
2	1.8856	2.9200	4.3027	4.8487	5.6428	6.9645	9.9250	14.0892	31.5998	
3	1.6377	2.3534	3.1824	3.4819	3.8961	4.5407	5.8408	7.4532	12.9244	
4	1.5332	2.1318	2.7765	2.9985	3.2976	3.7469	4.6041	5.5975	8.6101	
23	1.3195	1.7139	2.0687	2.1770	2.3132	2.4999	2.8073	3.1040	3.7676	
24	1.3178	1.7109	2.0639	2.1715	2.3069	2.4922	2.7970	3.0905	3.7454	
25	1.3163	1.7081	2.0595	2.1666	2.3011	2.4851	2.7874	3.0782	3.7251	
26	1.3150	1.7056	2.0555	2.1620	2.2958	2.4786	2.7787	3.0669	3.7067	
27	1.3137	1.7033	2.0518	2.1578	2.2909	2.4727	2.7707	3.0565	3.6895	
28	1.3125	1.7011	2.0484	2.1539	2.2864	2.4671	2.7633	3.0470	3.6739	
29	1.3114	1.6991	2.0452	2.1503	2.2822	2.4620	2.7564	3.0380	3.6595	
30	1.3104	1.6973	2.0423	2.1470	2.2783	2.4573	2.7500	3.0298	3.6460	
31	1.3095	1.6955	2.0395	2.1438	2.2746	2.4528	2.7440	3.0221	3.6335	

Independent t-test

การทดสอบนี้ใช้เมื่อมีการสุ่มตัวอย่างสองกลุ่มที่เป็นอิสระจากกัน และต้องการตรวจสอบว่าค่าเฉลี่ยของสองกลุ่มนั้นแตกต่างกันหรือไม่ โดยมีขั้นตอนดังนี้

1. **กำหนดสมมติฐาน**

   H0: ค่าเฉลี่ยของกลุ่มตัวอย่างทั้งสองกลุ่มเท่ากัน ($\mu 1 = \mu 2$)

   H1: ค่าเฉลี่ยของกลุ่มตัวอย่างทั้งสองกลุ่มไม่เท่ากัน ($\mu 1 \neq \mu 2$)

2. **คำนวณค่า t**

   สูตรการคำนวณค่า t สำหรับ Independent t-test คือ

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

3. **เปรียบเทียบกับค่า t วิกฤต**

เมื่อคำนวณค่า t ได้แล้ว ให้เปรียบเทียบกับค่า t วิกฤตจากตาราง t distribution ที่ระดับความเชื่อมั่นที่กำหนด (เช่น 0.05) และ df = n-1

หาก |t| มากกว่าค่า t วิกฤต ปฏิเสธสมมติฐาน H_0 และสรุปว่าค่าเฉลี่ยของกลุ่มตัวอย่างสองกลุ่มแตกต่างกันอย่างมีนัยสำคัญ

ตัวอย่างเรื่องความสามารถในการวิ่งระยะทางของเพศชายและเพศหญิง โดยสมมติฐานว่าไม่มีความแตกต่างในความสามารถในการวิ่งระยะทางระหว่างเพศชายและเพศหญิง

สมมติฐานว่าง: $H_0 : \mu_1=\mu_2$) และ

สมมติฐานทดสอบว่ามีความแตกต่าง

สมมติฐานทดสอบ: $H_1:\mu_1\neq\mu_2$

โดยที่ μ_1 แทนค่าเฉลี่ยของเวลาวิ่งระยะทางของเพศชายและ μ_2 แทนค่าเฉลี่ยของเวลาวิ่งระยะทางของเพศหญิง

สามารถใช้ t-test สำหรับตารางค่าสรุปเพื่อทดสอบสมมติฐานนี้ โดยทำการสำรวจข้อมูลการวิ่งระยะทางของกลุ่มชายและหญิง และคำนวณค่า t-statistic และ p-value เพื่อทำการตัดสินใจว่าจะปฏิเสธหรือยอมรับสมมติฐานว่าง ดังนี้:

เพศ	จำนวน	เวลาวิ่งระยะทาง (นาที)
ชาย	20	15, 18, 17, 16, 20, 19, 22, 21, 18, 16, 17, 19, 20, 21, 22, 20, 18, 19, 20, 21
หญิง	20	16, 17, 15, 14, 18, 17, 20, 19, 16, 15, 17, 18, 19, 20, 21, 19, 17, 18, 19, 20

หลังจากนั้น เราคำนวณหาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของแต่ละกลุ่ม และคำนวณค่า t-statistic และ p-value จากการใช้ t-test เพื่อทดสอบสมมติฐาน:

- ค่าเฉลี่ยของเวลาวิ่งระยะทางของเพศชาย ($\bar{\mathbf{X}}_1$): 19.05 นาที
- ค่าเฉลี่ยของเวลาวิ่งระยะทางของเพศหญิง ($\bar{\mathbf{X}}_2$): 17.45 นาที

- ค่าส่วนเบี่ยงเบนมาตรฐานของเวลาวิ่งระยะทางของเพศชาย (S_1): 1.87 นาที
- ค่าส่วนเบี่ยงเบนมาตรฐานของเวลาวิ่งระยะทางของเพศหญิง (S_2): 1.66 นาที

นำค่าเหล่านี้มาคำนวณค่า t-statistic และ p-value โดยใช้สูตรที่เหมาะสม และหาก p-value น้อยกว่าระดับนัยสำคัญที่กำหนดไว้หรือหากค่า t-statistic มากกว่าค่า t ที่เปิดได้จากตาราง จะปฏิเสธสมมติฐานว่าง แสดงว่ามีความแตกต่างในเวลาวิ่งระยะทางระหว่างเพศชายและเพศหญิง

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{19.05 - 17.45}{\sqrt{\frac{1.87^2}{20} + \frac{1.66^2}{20}}}$$

ดังนั้นค่า t-statistic ที่ได้คือ 2.862

เปิดตาราง t ที่ degrees of freedom (d.f.) = n -1 = 20-1 = 19 ระดับนัยสำคัญ .05 ได้ 2.0930 เมื่อ t-statistic ที่ได้คำนวณได้คือ 2.862 มากกว่าค่าที่เปิดจากตาราง 2.0930 แสดงว่าจึงตีความว่ามีความแตกต่างในเวลาวิ่งระยะทางระหว่างเพศชายและเพศหญิง

ตารางการแจกแจง t

	0.1	0.05	0.025	0.02	0.015	0.01	0.005	0.0025	0.0005	One-tail
df	0.2	0.1	0.05	0.04	0.03	0.02	0.01	0.005	0.001	Two-tail
1	3.0777	6.3137	12.7062	15.8945	21.2051	31.8210	63.6559	127.3211	636.5776	
2	1.8856	2.9200	4.3027	4.8487	5.6428	6.9645	9.9250	14.0892	31.5998	
3	1.6377	2.3534	3.1824	3.4819	3.8961	4.5407	5.8408	7.4532	12.9244	
4	1.5332	2.1318	2.7765	2.9985	3.2976	3.7469	4.6041	5.5975	8.6101	
5	1.4759	2.0150	2.5706	2.7565	3.0029	3.3649	4.0321	4.7733	6.8685	
6	1.4398	1.9432	2.4469	2.6122	2.8289	3.1427	3.7074	4.3168	5.9587	
7	1.4149	1.8946	2.3646	2.5168	2.7146	2.9979	3.4995	4.0294	5.4081	
8	1.3968	1.8595	2.3060	2.4490	2.6338	2.8965	3.3554	3.8325	5.0414	
9	1.3830	1.8331	2.2622	2.3984	2.5738	2.8214	3.2498	3.6896	4.7809	
10	1.3722	1.8125	2.2281	2.3593	2.5275	2.7638	3.1693	3.5814	4.5868	
11	1.3634	1.7959	2.2010	2.3281	2.4907	2.7181	3.1058	3.4966	4.4369	
12	1.3562	1.7823	2.1788	2.3027	2.4607	2.6810	3.0545	3.4284	4.3178	
13	1.3502	1.7709	2.1604	2.2816	2.4358	2.6503	3.0123	3.3725	4.2209	
14	1.3450	1.7613	2.1448	2.2638	2.4149	2.6245	2.9768	3.3257	4.1403	
15	1.3406	1.7531	2.1315	2.2485	2.3970	2.6025	2.9467	3.2860	4.0728	
16	1.3368	1.7459	2.1199	2.2354	2.3815	2.5835	2.9208	3.2520	4.0149	
17	1.3334	1.7396	2.1098	2.2238	2.3681	2.5669	2.8982	3.2224	3.9651	
18	1.3304	1.7341	2.1009	2.2137	2.3562	2.5524	2.8784	3.1966	3.9217	
19	1.3277	1.7291	2.0930	2.2047	2.3457	2.5395	2.8609	3.1737	3.8833	

Paired t-test

การทดสอบนี้ใช้เมื่อมีการวัดค่าเฉลี่ยจากกลุ่มตัวอย่างเดียวกันหรือคู่ข้อมูลที่มีความสัมพันธ์กัน (เช่น ก่อนและหลังการทดลอง) และต้องการตรวจสอบว่าค่าเฉลี่ยของความแตกต่างระหว่างคู่ข้อมูลมีความแตกต่างกันหรือไม่ โดยมีขั้นตอนดังนี้

1. **กำหนดสมมติฐาน**

   H_0: ค่าเฉลี่ยของความแตกต่างระหว่างคู่ข้อมูลเท่ากับ 0 ($\mu_D=0$)

   H_1: ค่าเฉลี่ยของความแตกต่างระหว่างคู่ข้อมูลไม่เท่ากับ 0 ($\mu_D\neq0$)

2. **คำนวณค่า t**

   สูตรการคำนวณค่า t สำหรับ Paired t-test คือ

$$t = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}}$$

โดยที่

$\overline{D}$ คือ ค่าเฉลี่ยของความแตกต่างระหว่างคู่ข้อมูล

S_D คือ ส่วนเบี่ยงเบนมาตรฐานของความแตกต่างระหว่างคู่ข้อมูล

n คือ จำนวนคู่ข้อมูล

3. **เปรียบเทียบกับค่า t วิกฤต**

เมื่อคำนวณค่า t ได้แล้ว ให้เปรียบเทียบกับค่า t วิกฤตจากตาราง t distribution ที่ระดับความเชื่อมั่นที่กำหนด (เช่น 0.05)

หาก $|t|$ มากกว่าค่า t วิกฤต เราปฏิเสธสมมติฐาน H0 และสรุปว่าค่าเฉลี่ยของความแตกต่างระหว่างคู่ข้อมูลมีความแตกต่างกันอย่างมีนัยสำคัญ

ตัวอย่างการใช้ Paired t-test

สมมติว่าต้องการทดสอบว่านักเรียนกลุ่มหนึ่งมีคะแนนสอบก่อนและหลังการเรียนรู้วิชาใหม่แตกต่างกันหรือไม่ เรามีข้อมูลดังนี้

คะแนนสอบก่อนการเรียนรู้ [78, 82, 88, 90, 75]

คะแนนสอบหลังการเรียนรู้ [85, 87, 90, 95, 80]

คำนวณความแตกต่างระหว่างคะแนนสอบก่อนและหลังการเรียนรู้

ความแตกต่าง: [7, 5, 2, 5, 5]

ค่าเฉลี่ยของความแตกต่าง $\overline{D} = \frac{7+5+2+5+5}{5} = \frac{24}{5} = 4.8$

ส่วนเบี่ยงเบนมาตรฐานของความแตกต่าง (S_D)

$$s_D = \sqrt{\frac{\sum(D_i - \bar{D})^2}{n-1}}$$

S_D =

$$\sqrt{\frac{(7-4.8)^2+(5-4.8)^2+(2-4.8)^2+(5-4.8)^2+(5-4.8)^2}{5-1}}$$

$$S_D = \sqrt{\frac{(\mathbf{2.2})^2+(\mathbf{0.2})^2+(\mathbf{-2.8})^2+(\mathbf{0.2})^2+(\mathbf{0.28})^2}{\mathbf{4}}}$$

$$S_D = \sqrt{\frac{\mathbf{4.84+0.04+7.84+0.04+0.04}}{\mathbf{4}}}$$

$$S_D = \sqrt{\frac{\mathbf{12.8}}{\mathbf{4}}} = \sqrt{\mathbf{3.2}} \approx 1.79$$

คำนวณค่า t

$$t = \frac{\mathbf{4.8}}{\frac{\mathbf{1.79}}{\sqrt{\mathbf{5}}}} = \frac{\mathbf{4.8}}{\mathbf{0.8}} = 6$$

ถ้าค่า t วิกฤตที่ระดับความเชื่อมั่น 0.05 สำหรับ df=4 คือ 2.776
เนื่องจาก 6>2.7766 จึงปฏิเสธสมมติฐาน H_0 ซึ่งหมายความว่าค่าเฉลี่ยของคะแนนสอบก่อนและหลังการเรียนรู้แตกต่างกันอย่างมีนัยสำคัญ

Paired t-test เป็นวิธีการทดสอบความแตกต่างของค่าเฉลี่ยของกลุ่มตัวอย่างสองกลุ่มที่มีความสัมพันธ์กัน โดยการตรวจสอบค่าเฉลี่ยของความแตกต่างระหว่างคู่ข้อมูล และเป็นเครื่องมือสำคัญในสถิติอ้างอิงที่ช่วยในการวิเคราะห์และตัดสินใจเกี่ยวกับข้อมูล

สรุป

สมมติฐานทางสถิติคือข้อสันนิษฐานเกี่ยวกับประชากรที่ต้องการทดสอบ โดยแบ่งออกเป็นสองประเภทหลัก คือ สมมติฐานศูนย์ (null hypothesis, H_0) ซึ่งระบุว่าไม่มีความแตกต่างหรือความสัมพันธ์ และสมมติฐานทางเลือก (alternative hypothesis, H_1) ซึ่งระบุว่ามีความแตกต่างหรือความสัมพันธ์ การทดสอบสมมติฐานทางสถิติมีหลายวิธี รวมถึง One-sample t-test, Independent t-test, และ Paired t-test โดย One-

sample t-test ใช้ทดสอบค่าเฉลี่ยของกลุ่มตัวอย่างหนึ่งกับค่าคงที่, Independent t-test ใช้ทดสอบความแตกต่างของค่าเฉลี่ยระหว่างสองกลุ่มที่เป็นอิสระจากกัน, และ Paired t-test ใช้ทดสอบความแตกต่างของค่าเฉลี่ยระหว่างสองกลุ่มที่มีความสัมพันธ์กัน เช่น การวัดก่อนและหลังในกลุ่มตัวอย่างเดียวกัน ทั้งสามวิธีนี้ช่วยในการวิเคราะห์และตัดสินใจเกี่ยวกับข้อมูลโดยใช้หลักการของสถิติอ้างอิง (Inferential Statistics)

แบบฝึกหัด: การทดสอบสมมติฐานด้วย One-sample t-test

สถานการณ์:

บริษัทผลิตขนมหวานแห่งหนึ่งอ้างว่าคุกกี้ที่ผลิตมีน้ำหนักเฉลี่ย 50 กรัม คุณเป็นนักวิเคราะห์ข้อมูลที่ได้รับมอบหมายให้ตรวจสอบว่าค่าเฉลี่ยน้ำหนักของคุกกี้ที่บริษัทอ้างว่าน้ำหนัก 50 กรัมเป็นจริงหรือไม่ โดยคุณสุ่มตัวอย่างคุกกี้มา 30 ชิ้น แล้วคำนวณค่าน้ำหนักเฉลี่ยได้ 48.5 กรัม และส่วนเบี่ยงเบนมาตรฐานของน้ำหนักคุกกี้ในกลุ่มตัวอย่างคือ 2.5 กรัม

คำถาม:

1. กำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1)
2. คำนวณค่า t สถิติสำหรับการทดสอบนี้
3. เปรียบเทียบค่า t ที่ได้กับค่า t วิกฤตที่ระดับความเชื่อมั่น 0.05 โดยใช้ข้อมูลจากตาราง t distribution (df = n-1)
4. สรุปผลการทดสอบสมมติฐานว่าคุณจะปฏิเสธหรือไม่ปฏิเสธสมมติฐานศูนย์

แนวทางการทำ:

1. กำหนดสมมติฐาน:

   H0: ค่าเฉลี่ยน้ำหนักของคุกกี้เท่ากับ 50 กรัม (μ=50)

   H1: ค่าเฉลี่ยน้ำหนักของคุกกี้ไม่เท่ากับ 50 กรัม ($\mu \neq 50$)
2. คำนวณค่า t

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$ =48.5 กรัม ค่าเฉลี่ยของกลุ่มตัวอย่าง

μ_0 =50 กรัม คือ ค่าคงที่ที่ต้องการทดสอบ

s=2.5 กรัม ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

n=30 จำนวนตัวอย่าง

3. เปรียบเทียบกับค่า t วิกฤต:

ดูค่าจากตาราง t distribution ที่ระดับความเชื่อมั่น 0.05 สำหรับ df=30−1=29

สรุปผลการทดสอบ

- หาก |t| มากกว่าค่า t วิกฤต ให้ปฏิเสธสมมติฐาน H_0
- หาก |t| น้อยกว่าหรือเท่ากับค่า t วิกฤต ให้ไม่ปฏิเสธสมมติฐาน H_0

เอกสารอ้างอิง

ปรีดาภรณ์ กาญจนสำราญวงศ์.(2560). *หลักสถิติเบื้องต้น*.นนทบุรี. ไอดีซี พรีเมียร์

บทที่ 8
การวิเคราะห์ความแปรปรวน

การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เป็นวิธีการทางสถิติที่ใช้ในการเปรียบเทียบค่าเฉลี่ยของกลุ่มข้อมูลหลายกลุ่ม เพื่อพิจารณาว่าค่าเฉลี่ยของกลุ่มใดแตกต่างกันอย่างมีนัยสำคัญหรือไม่ การวิเคราะห์นี้มีความสำคัญอย่างยิ่งในงานวิจัยที่ต้องการตรวจสอบความแตกต่างระหว่างกลุ่มต่าง ๆ เช่น การเปรียบเทียบประสิทธิภาพของยาหลายชนิด การวิเคราะห์ผลการเรียนของนักเรียนในห้องเรียนต่าง ๆ หรือการเปรียบเทียบคุณภาพของผลิตภัณฑ์จากแหล่งที่มาต่าง ๆ การวิเคราะห์การแจกแจงความแปรปรวน เป็นขั้นตอนที่ใช้ในการประเมินว่าข้อมูลแปรปรวนในกลุ่มต่าง ๆ นั้นมีการแจกแจงอย่างไร และการแปรปรวนนั้นมีผลต่อค่าเฉลี่ยของกลุ่มอย่างไรบ้าง โดยเครื่องมือทางสถิติเช่น ANOVA จะช่วยในการตรวจสอบและวิเคราะห์การแจกแจงนี้ นอกจากนี้ การใช้เครื่องมือสถิติยังครอบคลุมถึงการทดสอบสมมติฐาน การคำนวณค่าพารามิเตอร์ต่าง ๆ และการประเมินความน่าเชื่อถือของผลการวิเคราะห์ เพื่อให้ได้ผลการวิจัยที่มีความแม่นยำและเป็นไปตามหลักวิชาการ ด้วยการใช้เครื่องมือสถิติที่เหมาะสม นักวิจัยสามารถระบุแหล่งที่มาของความแปรปรวนในข้อมูลได้อย่างถูกต้อง และสามารถสรุปผลการวิเคราะห์ได้อย่างชัดเจนและมีเหตุผล ทำให้การตัดสินใจและการดำเนินการตามผลการวิจัยนั้นมีความน่าเชื่อถือและเป็นไปอย่างมีประสิทธิภาพ

การวิเคราะห์การแจกแจงความแปรปรวน

การวิเคราะห์การแจกแจงความแปรปรวน (Variance Analysis) เป็นกระบวนการที่ใช้เพื่อศึกษาและทำความเข้าใจการกระจายตัวของข้อมูล

ภายในกลุ่มต่าง ๆ รวมถึงการตรวจสอบว่าแหล่งที่มาของความแปรปรวนในข้อมูลมีอะไรบ้าง และความแปรปรวนนั้นส่งผลต่อค่าเฉลี่ยของกลุ่มอย่างไร การวิเคราะห์นี้มักใช้ในบริบทของการทดสอบสมมติฐานและการเปรียบเทียบกลุ่มหลายกลุ่ม โดยเฉพาะในกรณีที่มีการทดลองที่เกี่ยวข้องกับการสุ่มตัวอย่างและการวัดผลหลายกลุ่ม (ปรีดาภรณ์ กาญจนสำราญวงศ์,2560)

เงื่อนไขของการวิเคราะห์ความแปรปรวน

1) การสุ่มตัวอย่างแต่ละชุดจะต้องสุ่มอย่างเป็นอิสระกัน
2) สุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติ
3) ค่าแปรปรวนของประชากรแต่ละกลุ่มต้องเท่ากัน $\sigma_1^2 = \sigma_2^2 ... = \sigma_k^2$

ขั้นตอนในการวิเคราะห์การแจกแจงความแปรปรวน

1. **กำหนดสมมติฐาน**:

• สมมติฐานว่าง (Null Hypothesis, H0): ค่าเฉลี่ยของทุกกลุ่มไม่มีความแตกต่างกัน

• สมมติฐานทางเลือก (Alternative Hypothesis, H1): มีค่าเฉลี่ยของกลุ่มใดกลุ่มหนึ่งหรือมากกว่านั้นแตกต่างกัน

2. **คำนวณความแปรปรวน**:

• ความแปรปรวนภายในกลุ่ม (Within-group variance): วัดความแปรปรวนของข้อมูลภายในแต่ละกลุ่ม

• ความแปรปรวนนอกกลุ่ม (Between-group variance): วัดความแปรปรวนระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ

3. **ใช้สถิติ F-test**:

• คำนวณค่า F ซึ่งเป็นอัตราส่วนระหว่างความแปรปรวนนอกกลุ่มกับความแปรปรวนภายในกลุ่ม

• เปรียบเทียบค่า F ที่คำนวณได้กับค่าวิกฤต (Critical Value) ที่ระดับความเชื่อมั่นที่กำหนด (เช่น 0.05 หรือ 0.01)

4. **การตัดสินใจ**:

• หากค่า F ที่คำนวณได้มากกว่าค่าวิกฤต แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ ปฏิเสธสมมติฐานหลัก

• หากค่า F ที่คำนวณได้น้อยกว่าหรือเท่ากับค่าวิกฤต แสดงว่าไม่มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานหลัก ยอมรับสมมติฐานหลัก

การใช้เครื่องมือสถิติในการวิเคราะห์

การวิเคราะห์การแจกแจงความแปรปรวนสามารถทำได้ด้วยเครื่องมือสถิติต่าง ๆ เช่น

โปรแกรมคอมพิวเตอร์: เช่น SPSS, R, SAS หรือ Python ที่มีแพ็กเกจสำหรับการวิเคราะห์ ANOVA

การคำนวณด้วยตนเอง: สำหรับการคำนวณเบื้องต้นหรือในกรณีที่มีจำนวนข้อมูลไม่มาก สามารถใช้สูตรและตารางสถิติเพื่อคำนวณค่า F และค่าชี้ขาด

ประโยชน์ของการวิเคราะห์การแจกแจงความแปรปรวน

การประเมินความแตกต่าง: ช่วยให้สามารถประเมินว่ามีความแตกต่างระหว่างกลุ่มหรือไม่

การเข้าใจแหล่งที่มาของความแปรปรวน: ทำให้นักวิจัยเข้าใจว่าความแปรปรวนเกิดจากอะไร

การตัดสินใจที่มีข้อมูลรองรับ: ผลการวิเคราะห์ที่ได้สามารถใช้ในการตัดสินใจที่มีความมั่นใจและมีหลักฐานรองรับ

การวิเคราะห์การแจกแจงความแปรปรวนจึงเป็นเครื่องมือที่สำคัญในการวิจัยและการวิเคราะห์ข้อมูลทางสถิติ ช่วยให้นักวิจัยสามารถตรวจสอบและตีความผลการทดลองได้อย่างถูกต้องและแม่นยำ

กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	กลุ่มk
X_{11}	X_{21}	X_{31} ...	X_{k1}
X_{12}	X_{22}	X_{32} ...	X_{k2}
X_{13}	X_{23}	X_{33} ...	X_{k3}

	...	...	...		
	X_{1n}	X_{2n}	X_{3n} ...	X_{kn}	
รวม	T_1	T_2	T_3 ...	T_k	T
เฉลี่ย	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_k$	$\bar{x}$

X_{ij} แทนข้อมูลลำดับที่ i กลุ่มที่ j

i = 1,2,3,...,n และ j=1,2,3,...,k

T_j แทนผลรวมของข้อมูลภายในกลุ่มที่ j

T แทนผลรวมข้อมูลทั้งหมด

$\bar{x}_j$ แทนค่าเฉลี่ยของข้อมูลกลุ่มที่ j

$\bar{x}$ แทนค่าเฉลี่ยของข้อมูลทั้งหมด

k แทนจำนวนกลุ่ม

N แทนจำนวนข้อมูลทั้งหมด เท่ากับ $n_1+n_2+n_3+...+n_k$

เนื่องจากการวิเคราะห์ความแปรปรวนทางเดียวเป็นการศึกษาอิทธิพลของตัวแปรหรือปัจจัยเดียว (วุฒิ สุขเจริญ, 2561) ที่มีผลทำ ให้ค่าสังเกตแตกต่างกันนั้น คือข้อมูลมีความแตกต่างเนื่องจากกลุ่มที่แตกต่างกัน จึงวิเคราะห์ด้วยการแบ่งความแปรปรวนของข้อมูลเป็นดังนี้

ความแปรปรวนระหว่างกลุ่ม (Between Groups Sum of Square) เขียนแทนด้วย สัญลักษณ SS between, SSB เป็นการพิจารณา ความแปรปรวนที่เกิดจากการที่ค่าเฉลี่ยของตัวอย่างในแต่ละกลุ่ม แตกต่างจากค่าเฉลี่ยรวม คือ

SSB, SS between $= \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2$

ความแปรปรวนภายในกลุ่ม (Within Group Sum of Square) เขียนแทนด้วย สัญลักษณ์ SS error, SSE เป็นการพิจารณาความแปรปรวนที่เกิดขึ้นภายในกลุ่มแต่ละกลุ่มซึ่งไม่ทราบสาเหตุ ว่าเป็นความแปรปรวนที่เกิดจากสาเหตุใด ในบางทีจึงเรียกว่าความคลาดเคลื่อน (Error Sum of Square) คือ

SSE, SS error = $\sum_j^k \sum_i^n (x_{ji} - \bar{x}_i)^2$

ความแปรปรวนรวม (Total Sum of Square) เขียนแทนด้วยสัญลักษณ์ SST เป็ น การพิจารณาความแปรปรวนที่เกิดจากค่าสังเกตแต่ละค่าแตกต่างจากค่าเฉลี่ยรวม คือ

SST = $\sum_j^k \sum_i^n (x_{ji} - \bar{x})^2$ และ SST = SSB + SSE

ตารางการวิเคราะห์ความแปรปรวนทางเดียว

ตารางการวิเคราะห์ความแปรปรวนทางเดียว (One-Way ANOVA)

Source of Variation	DF	Sum Square	Mean Square	F-test
ระหว่างกลุ่ม	k-1	SS between	MS b = $\frac{\text{SS between}}{\text{k}-1}$	$\frac{\text{MSb}}{\text{MSe}}$
ภายในกลุ่ม(error)	N-k	SS error	MS e = $\frac{\text{SS error}}{\text{N}-\text{k}}$	
Total	N-1	SS total		

การคำนวณความแปรปรวนและการใช้ค่าวิกฤต (Critical Value) เป็นส่วนสำคัญของการทดสอบ ANOVA โดยการใช้ตาราง F-distribution เพื่อเปรียบเทียบกับค่า F ที่คำนวณได้ ต่อไปนี้เป็นตัวอย่างการคำนวณโดยการเปิดตารางเพื่อใช้ค่าวิกฤต

ตัวอย่างการคำนวณ ANOVA ด้วยการเปิดตารางใช้ค่าวิกฤต

สมมติว่าต้องการเปรียบเทียบผลการทดสอบของนักเรียนในสามกลุ่มเรียน (A, B, และ C) โดยมีคะแนนดังนี้:

กลุ่ม A: 85, 90, 88, 86, 91

กลุ่ม B: 78, 82, 80, 76, 79

กลุ่ม C: 92, 95, 94, 90, 93

	กลุ่ม A	กลุ่ม B	กลุ่ม C	
	85	78	92	
	90	82	95	
	88	80	94	
	86	76	90	
	91	79	93	
รวม	440	395	464	1,299
เฉลี่ย	88	79	92.8	86.6

ขั้นตอนการคำนวณ ANOVA

1. **คำนวณค่าเฉลี่ยของแต่ละกลุ่มและค่าเฉลี่ยรวม**

ค่าเฉลี่ยของกลุ่ม A: (85+90+88+86+91)/5=88

ค่าเฉลี่ยของกลุ่ม B: (78+82+80+76+79)/5=79

ค่าเฉลี่ยของกลุ่ม C: (92+95+94+90+93)/5=92.8

ค่าเฉลี่ยรวม: (88+79+92.8)/3=86.6

2. **คำนวณความแปรปรวนนอกกลุ่ม (Between-group variance)**

$$SS_{between} = n \times ((\overline{X}_A - \overline{X}_{total})^2 + (\overline{X}_B - \overline{X}_{total})^2 + (\overline{X}_C - \overline{X}_{total})^2)$$

$=5 \times ((88-86.6)^2+(79-86.6)^2+(92.8-86.6)^2)$

$=5 \times (1.96+57.76+38.44)$

$=5 \times 98.16=490.8$

$df_{between} = k-1 = 3-1 = 2$

$MS_{between} = SS_{between}/df_{between} = 490.8/2=245.4$

3. **คำนวณความแปรปรวนภายในกลุ่ม (Within-group variance)**

$SS_{within} = \sum(X_{ji} - \overline{X}i)^2$

$= \sum ((85-88)^2+(90-88)^2+...+(93-92.8)^2)$

กลุ่ม A: $(85-88)^2+(90-88)^2+(88-88)^2+(86-88)^2+(91-88)^2$

$=9+4+0+4+9=26$

กลุ่ม B: $(78-79)^2+(82-79)^2+(80-79)^2+(76-79)^2+(79-79)^2$

$=1+9+1+9+0=20$

กลุ่ม C: $(92-92.8)^2+(95-92.8)^2+(94-92.8)^2+(90-92.8)^2+(93-92.8)^2$

$=0.64+4.84+1.44+7.84+0.04=14.8$

$SS_{within}=26+20+14.8 = 60.8$

$df_{within}= N-k =15-3 =12$

$MS_{within}=SS_{within}/df_{within}=60.8/12 =5.067$

4. **คำนวณค่า F**

$$F=\frac{\mathrm{MS_{between}}}{\mathrm{MS_{within}}}$$

$$= \frac{\mathbf{245.4}}{\mathbf{5.067}} = 48.41$$

ใช้ค่าวิกฤตจากตาราง F-distribution

ที่ระดับความเชื่อมั่น α=0.05 และ $df_{between}$=2, df_{within}=12, ค่าวิกฤต $F_{critical}$ จากตารางคือประมาณ 3.89

5. **การตัดสินใจ**

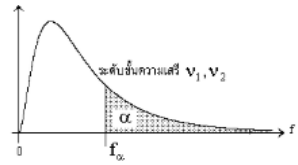
เนื่องจาก F = 48.41 > $F_{critical}$ = 3.89 จึงปฏิเสธสมมติฐานว่าง (null hypothesis) และสรุปได้ว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของกลุ่มเรียนทั้งสาม

ตารางที่ 6. ตารางค่าวิกฤตของการแจกแจง F

เมื่อกำหนดองศาความอิสระ ν_1, ν_2 และค่า α =0.05

ค่าจากตารางคือ $f_{0.05}$ คือค่าที่ทำให้ $P(f_{0.05} < F < \infty) = \alpha = 0.05$

$f_{0.05,(\nu_1,\nu_2)}$



	ν_1								
ν_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49

การวิเคราะห์ความแปรปรวนสองทาง (Two-way ANOVA)

การวิเคราะห์ความแปรปรวนสองทาง (Two-way ANOVA) เป็นการทดสอบทางสถิติที่ใช้เพื่อวิเคราะห์ผลกระทบของตัวแปรต้นสองตัวที่มีลักษณะเป็นหมวดหมู่ (Nominal) ต่อตัวแปรตามที่มีลักษณะเป็นเชิงปริมาณ (Continuous) และสามารถทดสอบการโต้ตอบระหว่างตัวแปรต้นสองตัวได้ด้วย (ภรณี เจริญภักตร์และคณะ, 2555)

วัตถุประสงค์

1. ผลหลัก: ตรวจสอบผลกระทบของตัวแปรแต่ละตัวที่มีต่อตัวแปรตาม
2. ผลการโต้ตอบ: ตรวจสอบว่าตัวแปรสองตัวมีการโต้ตอบกันในผลกระทบที่มีต่อตัวแปรตามหรือไม่

ขั้นตอนในการทำ Two-way ANOVA

1. **ตั้งสมมติฐาน**

สมมติฐานศูนย์ (H_0):

1. ไม่มีผลกระทบที่มีนัยสำคัญของตัวแปร A ต่อตัวแปรตาม
2. ไม่มีผลกระทบที่มีนัยสำคัญของตัวแปร B ต่อตัวแปรตาม
3. ไม่มีผลการโต้ตอบระหว่างตัวแปร A และตัวแปร B ต่อตัวแปรตาม

สมมติฐานทางเลือก (H_A):

1. มีผลกระทบที่มีนัยสำคัญของตัวแปร A ต่อตัวแปรตาม
2. มีผลกระทบที่มีนัยสำคัญของตัวแปร B ต่อตัวแปรตาม
3. มีผลการโต้ตอบระหว่างตัวแปร A และตัวแปร B ต่อตัวแปรตาม

2**. ข้อสมมติฐาน**

- ความเป็นอิสระของค่าสังเกต
- การกระจายตัวตามปกติของประชากรในแต่ละกลุ่ม
- ความเท่าเทียมของความแปรปรวน (Homogeneity of variances) ในแต่ละกลุ่ม

3. **การเก็บรวบรวมและเตรียมข้อมูล:**

1. การเก็บข้อมูลตัวแปรตามภายใต้การรวมตัวของตัวแปรสองตัว

2. การตรวจสอบให้แน่ใจว่าข้อมูลถูกจัดเรียงอย่างถูกต้อง โดยทั่วไปในรูปแบบตารางที่แถวแทนการสังเกตและคอลัมน์แทนตัวแปรต่าง ๆ และตัวแปรตาม

4. **ดำเนินการวิเคราะห์:**

1.ใช้ซอฟต์แวร์สถิติหรือภาษาการเขียนโปรแกรม (เช่น R, Python, SPSS) เพื่อทำการวิเคราะห์ Two-way ANOVA

2.คำนวณผลรวมของกำลังสองสำหรับแต่ละผลหลักและผลการโต้ตอบ, องศาเสรี, ค่าเฉลี่ยกำลังสอง, สถิติ F และค่า p-value

5. **การแปลผลลัพธ์:**

เปรียบเทียบค่า p-value กับระดับนัยสำคัญ (โดยปกติคือ 0.05) เพื่อพิจารณาว่าจะยอมรับหรือปฏิเสธสมมติฐานศูนย์

6. **การทดสอบหลังการวิเคราะห์ (Post-hoc Tests):**

ถ้าพบผลหลักหรือผลการโต้ตอบที่มีนัยสำคัญ ทำการทดสอบ Post-hoc (เช่น Tukey's HSD) เพื่อหาว่ากลุ่มใดมีความแตกต่างกัน
ตาราง ANOVA โดยทั่วไปจะประกอบด้วย

ผลรวมของกำลังสอง (Sum of Squares, SS): ความแปรปรวนที่เกิดจากแต่ละปัจจัยและผลการโต้ตอบ

องศาเสรี (Degrees of Freedom, df): จำนวนระดับของแต่ละปัจจัยลบด้วยหนึ่ง

ค่าเฉลี่ยกำลังสอง (Mean Square, MS): SS หารด้วย df ที่สอดคล้องกัน

ค่า F (F-value): อัตราส่วนของ MS ของปัจจัยต่อ MS ของส่วนเหลือ

ตารางการวิเคราะห์ความแปรปรวนสองทาง

ตารางการวิเคราะห์ความแปรปรวนสองทาง (Two-Way ANOVA)

Source of Variation	DF	Sum Square	Mean Square	F-test
ปัจจัยที่ 1	$n_{f1}-1$	SS f1	MS f1= $\frac{\text{SS f1}}{n_{f1}-1}$	F f1 = $\frac{\text{MS f1}}{\text{MSe}}$
ปัจจัยที่ 2	$n_{f2}-1$	SS f2	MS f2= $\frac{\text{SS f1}}{n_{f2}-1}$	F f2 = $\frac{\text{MS f2}}{\text{MSe}}$
interaction f_1*f_2	$(n_{f1}-1)$ * $(n_{f2}-1)$	SS f1*f2	MS f_1*f_2 = $\frac{\text{SS f1} * \text{f2}}{(n_{f1}-1)(n_{f2}-1)}$	F f_1*f_2 = $\frac{\text{MS } f_1 * f_2}{\text{MSe}}$
ภายในกลุ่ม (error)	$N-(n_{f1}* n_{f2})$	SS error	MS e = $\frac{\text{SS error}}{N-(n_{f1}* n_{f2})}$	
Total	N-1	SS total		

สมมติว่ากำลังศึกษาผลของสองปัจจัยที่มีต่อน้ำหนักของสัตว์ทดลอง

Factor A (ชนิดอาหาร): อาหารชนิด A1 และ A2

Factor B (ประเภทการออกกำลังกาย): การออกกำลังกายแบบ B1 และ B2

ประเภทออกกำลังกาย / ชนิดอาหาร	B1	B2
A1	23	25
A2	27	30
A1	22	24

A2	28	29

คำถามในการทดลอง

ชนิดอาหาร (A1 และ A2) มีผลต่อน้ำหนักของสัตว์ทดลองหรือไม่?

ประเภทการออกกำลังกาย (B1 และ B2) มีผลต่อน้ำหนักของสัตว์ทดลองหรือไม่?

อิทธิพลร่วมระหว่างชนิดอาหารและประเภทการออกกำลังกายมีผลต่อน้ำหนักของสัตว์ทดลองหรือไม่?

การตั้งสมมติฐาน:

สมมติฐานศูนย์ (Null Hypothesis, H0)

ชนิดอาหาร (Factor A) ไม่มีผลต่อน้ำหนักของสัตว์ทดลอง

ประเภทการออกกำลังกาย (Factor B) ไม่มีผลต่อน้ำหนักของสัตว์ทดลอง

ไม่มีอิทธิพลร่วมระหว่างชนิดอาหารและประเภทการออกกำลังกายที่มีผลต่อน้ำหนักของสัตว์ทดลอง

สมมติฐานทางเลือก (Alternative Hypothesis, H1)

ชนิดอาหาร (Factor A) มีผลต่อน้ำหนักของสัตว์ทดลอง

ประเภทการออกกำลังกาย (Factor B) มีผลต่อน้ำหนักของสัตว์ทดลอง

มีอิทธิพลร่วมระหว่างชนิดอาหารและประเภทการออกกำลังกายที่มีผลต่อน้ำหนักของสัตว์ทดลอง

ขั้นตอนที่ 1: การคำนวณค่าเฉลี่ย

1) ค่าเฉลี่ยรวมทั้งหมด (Grand Mean)

$$\bar{Y}_{..}=\frac{\sum Y_{ij}}{N} = \frac{23+25+27+30+22+24+28+29}{8}=\frac{208}{8}=26$$

2) ค่าเฉลี่ยของแต่ละกลุ่ม Factor A

$$\bar{Y}_{A1}=\frac{23+25+22+24}{4}=\frac{94}{4}=23.5$$

$$\bar{Y}_{A2}=\frac{27+30+28+29}{4}=\frac{114}{4}=28.5$$

3) ค่าเฉลี่ยของแต่ละกลุ่ม Factor B

$\bar{Y}_{B1} = \frac{23+27+22+28}{4} = \frac{100}{4} = 25$

$\bar{Y}_{B2} = \frac{25+30+24+29}{4} = \frac{104}{4} = 27$

4) ค่าเฉลี่ยของแต่ละกลุ่มการรวมกันของ Factor A และ B

$\bar{Y}_{A1B1} = \frac{23+22}{2} = \frac{45}{2} = 22.5$

$\bar{Y}_{A1B2} = \frac{25+24}{2} = \frac{49}{2} = 24.5$

$\bar{Y}_{A2B1} = \frac{27+28}{2} = \frac{55}{2} = 27.5$

$\bar{Y}_{A2B2} = \frac{30+29}{2} = \frac{59}{2} = 29.5$

ขั้นตอนที่ 2: คำนวณค่า Sum of Squares (SS)

1) Total Sum of Squares (SST)

$SST = \sum(Y_{ij} - \bar{Y}..)^2 = (23-26)^2 + (25-26)^2 + (27-26)^2 + (30-26)^2 + (22-26)^2 + (24-26)^2 + (28-26)^2 + (29-26)^2$

SST = 9+1+1+16+16+4+4+9 = **60**

2) Sum of Squares for Factor A (SSA)

$SS_A = \sum n_i(\bar{Y}_i - \bar{Y}..)^2 = 4(23.5-26)^2 + 4(28.5-26)^2$

$= 4(-2.5)^2 + 4(2.5)^2 = 4(6.25) + 4(6.25) = 25+25 =$ **50**

3) Sum of Squares for Factor B (SSB)

$SS_B = \sum n_j(\bar{Y}_j - \bar{Y}..)^2 = 4(25-26)^2 + 4(27-26)^2$

$= 4(-1)^2 + 4(1)^2 = 4(1) + 4(1) = 4+4 =$ **8**

4) Sum of Squares for Interaction (SSAB)

$SS_{AB} = \sum n_{ij}(\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}..)^2$

$= 2(22.5-23.5-25+26)^2 + 2(24.5-23.5-27+26)^2 + 2(27.5-28.5-25+26)^2$

$+ 2(29.5-28.5-27+26)^2$

= 2(0)2+2(0)2+2(0)2+2(0)2 = **0**

5) Sum of Squares for Error (SSE)

SSE = SST−SS_A−SS_B−SS_{AB} =60−50−8−0 =2

ขั้นตอนที่ 3: คำนวณค่า F-value

1) **Degrees of Freedom (df)**

$df_A = n_{f1}-1 = 2-1 = 1$

$df_B = n_{f2}-1 = 2-1 = 1$

$df_{AB} = (n_{f1}-1)(n_{f2}-1) = 1\times1 = 1$

$df_E = N - n_{f1}n_{f2} = 8-4 = 4$

$df_{Total} = N-1 = 8-1 = 7$

2) Mean Squares (MS)

$MS_A = \frac{SS_A}{df_A} = \frac{50}{1} = 50$

$MS_B = \frac{SS_B}{df_B} = \frac{8}{1} = 8$

$MS_{AB} = \frac{SS_{AB}}{df_{AB}} = \frac{0}{1} = 0$

$MS_E = \frac{SS_E}{df_E} = \frac{2}{4} = 0.5$

3) F-value

$F_A = \frac{MS_A}{MS_E} = \frac{50}{0.5} = 100$

$F_B = \frac{MS_B}{MS_E} = \frac{8}{0.5} = 16$

$F_{AB} = \frac{MS_{AB}}{MS_E} = \frac{0}{0.5} = 0$

ขั้นตอนที่ 4: หาค่า Critical Value จากตาราง F-distribution

จากตาราง F-distribution สำหรับระดับนัยสำคัญ α=0.05และองศาเสรี df1=1 และ df2=4 , ค่า Critical F-value สำหรับ α=0.05 คือ 7.71

โดยที่ F_A, df1 = c-1 = 2-1 =1, df2 = N-(c-1)(r-1) = 8-(2)(2) = 4

โดยที่ F_B, df1 = r-1 = 2-1 =1, df2 = N-(c-1)(r-1) = 8-(2)(2) = 4

โดยที่ F_{AB}, df1 = (c-1)(r-1) = 1x1=1, df2 = N-(c-1)(r-1) = 8-(2)(2) = 4

การแปลผลลัพธ์

ค่า F_A=100 มากกว่าค่า Critical F-value (7.71) ดังนั้นปฏิเสธสมมติฐานศูนย์สำหรับ Factor A

ค่า F_B=16 มากกว่าค่า Critical F-value (7.71) ดังนั้นเราปฏิเสธสมมติฐานศูนย์สำหรับ Factor B

ค่า F_{AB}=0 น้อยกว่าค่า Critical F-value (7.71) ดังนั้นยอมรับสมมติฐานศูนย์สำหรับอิทธิพลร่วมระหว่าง Factor A และ B

สรุปคือ Factor A และ Factor B มีผลกระทบที่มีนัยสำคัญต่อตัวแปรตามหรือน้ำหนักของสัตว์ทดลอง แต่ไม่มีอิทธิพลร่วมระหว่างชนิดอาหารและประเภทการออกกำลังกายที่มีผลต่อน้ำหนักของสัตว์ทดลอง

การทดสอบ Post-hoc ด้วย Tukey's HSD

การทดสอบ Post-hoc ด้วย Tukey's HSD (Honestly Significant Difference) หลังจากที่ทำการทดสอบ Two-way ANOVA และพบว่ามีผลต่างที่มีนัยสำคัญระหว่างระดับของปัจจัย จึงจำเป็นต้องทำการทดสอบ Post-hoc เพื่อระบุว่าระดับใดของปัจจัยที่มีความแตกต่างกันบ้าง การทดสอบ Tukey's HSD (Honestly Significant Difference) เป็นหนึ่งในวิธีที่ใช้กันอย่างแพร่หลาย

ประเภทออกกำลังกาย / ชนิดอาหาร	B1	B2
A1	23	25
A2	27	30
A1	22	24
A2	28	29

ขั้นตอนการคำนวณ Tukey's HSD

1. คำนวณค่าเฉลี่ยสำหรับแต่ละกลุ่ม

ค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้

$$\bar{Y}_{A1B1} = \frac{23+22}{2} = \frac{45}{2} = 22.5$$

$$\bar{Y}_{A1B2} = \frac{25+24}{2} = \frac{49}{2} = 24.5$$

$\bar{Y}_{A2B1} = \frac{27+28}{2} = \frac{55}{2} = 27.5$

$\bar{Y}_{A2B2} = \frac{30+29}{2} = \frac{59}{2} = 29.5$

2. คำนวณค่า Mean Square Error (MSE)

MSE มาจากผลการคำนวณ Two-way ANOVA ก่อนหน้านี้

$MS_E = \frac{SS_E}{df_E} = \frac{2}{4} = 0.5$

3. คำนวณค่า Standard Error (SE)

$SE = \sqrt{\frac{MS_E}{n}} = \sqrt{\frac{0.5}{2}} = \sqrt{0.25} = 0.5$; n = 2

4. คำนวณค่า HSD

HSD=q*SE ค่า q เป็นค่า Critical Value จากตาราง Studentized Range Distribution ที่ระดับนัยสำคัญ α=0.05 และจำนวนกลุ่มเปรียบเทียบ (k) สำหรับข้อมูลนี้ k=4 ค่า q สำหรับ k=4และ df=4 จากตาราง q-distribution ประมาณ 4.197 ดังนั้น HSD=4.197*0.5=2.0985

5. เปรียบเทียบค่าเฉลี่ยของแต่ละกลุ่ม

ค่าเฉลี่ยของแต่ละกลุ่มที่แตกต่างกันมากกว่า 2.0985 จะถือว่ามีความแตกต่างกันอย่างมีนัยสำคัญ

6. ทดสอบคู่ของค่าเฉลี่ยแต่ละกลุ่ม

1) ระหว่าง $\bar{Y}_{A1B1}$ กับ $\bar{Y}_{A1B2}$ = |22.5−24.5|=2 (ไม่แตกต่างกันอย่างมีนัยสำคัญ) นำ 2 ไปเทียบกับ 2.0985 พบว่าน้อยกว่าจึงไม่พบความแตกต่างกันอย่างมีนัยสำคัญ

2) ระหว่าง $\bar{Y}_{A1B1}$ กับ $\bar{Y}_{A2B1}$ = |22.5−27.5|=5 (แตกต่างกันอย่างมีนัยสำคัญ) นำ 5 ไปเทียบกับ 2.0985 พบว่ามากกว่าจึงพบความแตกต่างกันอย่างมีนัยสำคัญ

3) ระหว่าง $\bar{Y}_{A1B1}$ กับ $\bar{Y}_{A2B2}$ = |22.5−29.5|=7 (แตกต่างกันอย่างมีนัยสำคัญ) นำ 7 ไปเทียบกับ 2.0985 พบว่ามากกว่าจึงพบความแตกต่างกันอย่างมีนัยสำคัญ

4) ระหว่าง $\bar{Y}_{A1B2}$ กับ $\bar{Y}_{A2B1}$ = |24.5−27.5|=3 (แตกต่างกันอย่างมีนัยสำคัญ) นำ 3 ไปเทียบกับ 2.0985 พบว่ามากกว่าจึงพบความแตกต่างกันอย่างมีนัยสำคัญ

5) ระหว่าง $\bar{Y}_{A1B2}$ กับ $\bar{Y}_{A2B2}$ = |24.5−29.5|=5 (แตกต่างกันอย่างมีนัยสำคัญ) นำ 5 ไปเทียบกับ 2.0985 พบว่ามากกว่าจึงพบความแตกต่างกันอย่างมีนัยสำคัญ

6) ระหว่าง $\bar{Y}_{A2B1}$ กับ $\bar{Y}_{A2B2}$ = |27.5−29.5|=2 (ไม่แตกต่างกันอย่างมีนัยสำคัญ) นำ 2 ไปเทียบกับ 2.0985 พบว่าน้อยกว่าจึงไม่พบความแตกต่างกันอย่างมีนัยสำคัญ

จากผลการคำนวณ

$\bar{Y}_{A1B1}$ แตกต่างกับ $\bar{Y}_{A2B1}$ และ $\bar{Y}_{A2B2}$

$\bar{Y}_{A1B2}$ แตกต่างกับ $\bar{Y}_{A2B1}$ และ $\bar{Y}_{A2B2}$

$\bar{Y}_{A2B1}$ และ $\bar{Y}_{A2B2}$ ไม่แตกต่างกันอย่างมีนัยสำคัญ

การทดสอบนี้แสดงให้เห็นว่ากลุ่มต่าง ๆ ของ Factor A และ B มีความแตกต่างกันอย่างมีนัยสำคัญ ยกเว้นคู่ระหว่าง $\bar{Y}_{A2B1}$ และ $\bar{Y}_{A2B2}$

การใช้เครื่องมือสถิติในการวิเคราะห์ข้อมูลแปรปรวน

การใช้เครื่องมือสถิติในการวิเคราะห์ข้อมูลแปรปรวนเป็นกระบวนการที่สำคัญสำหรับการตรวจสอบความแตกต่างของค่าเฉลี่ยในกลุ่มข้อมูลหลายกลุ่ม เครื่องมือสถิติต่าง ๆ ช่วยให้นักวิจัยสามารถวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพและแม่นยำ ซึ่งสามารถแบ่งออกเป็นขั้นตอนหลัก ๆ ดังนี้

1. การเก็บรวบรวมและเตรียมข้อมูล

ก่อนการวิเคราะห์ จำเป็นต้องรวบรวมข้อมูลและจัดเตรียมในรูปแบบที่เหมาะสม ซึ่งรวมถึงการตรวจสอบความครบถ้วนของข้อมูลและการจัดการกับค่าผิดพลาด (Outliers) หรือข้อมูลที่หายไป

2. การเลือกเครื่องมือสถิติ

มีเครื่องมือสถิติหลายประเภทที่สามารถใช้ในการวิเคราะห์ข้อมูลแปรปรวนได้ โดยเครื่องมือที่นิยมใช้ได้แก่:

- ANOVA (Analysis of Variance): ใช้สำหรับเปรียบเทียบค่าเฉลี่ยของกลุ่มมากกว่าสองกลุ่ม
- MANOVA (Multivariate Analysis of Variance): ใช้เมื่อมีตัวแปรตามหลายตัว
- ANCOVA (Analysis of Covariance): ใช้ในการวิเคราะห์ความแปรปรวนเมื่อมีการควบคุมตัวแปรร่วม
- Repeated Measures ANOVA: ใช้เมื่อมีการวัดผลซ้ำในกลุ่มเดียวกัน

3. การใช้ซอฟต์แวร์สถิติ

ซอฟต์แวร์สถิติช่วยให้การวิเคราะห์ข้อมูลแปรปรวนทำได้ง่ายขึ้น มีหลายโปรแกรมที่นิยมใช้ เช่น:

- SPSS (Statistical Package for the Social Sciences): ใช้งานง่าย มีอินเตอร์เฟซที่เป็นมิตรกับผู้ใช้
- R: เป็นโปรแกรมโอเพ่นซอร์ส มีความยืดหยุ่นสูงและมีแพ็กเกจเฉพาะสำหรับการวิเคราะห์ ANOVA
- SAS (Statistical Analysis System): ใช้ในการวิเคราะห์ข้อมูลทางธุรกิจและวิจัย
- Python: ใช้ร่วมกับไลบรารีสถิติเช่น SciPy, Pandas, และ Statsmodels

4. การดำเนินการวิเคราะห์

ตัวอย่างการวิเคราะห์ข้อมูลแปรปรวนด้วยซอฟต์แวร์ R:

```R
# โหลดแพ็กเกจที่จำเป็น
install.packages("ggplot2")
library(ggplot2)

# สร้างข้อมูลตัวอย่าง
data <- data.frame(
  group = factor(rep(c("A", "B", "C"), each = 10)),
  value = c(rnorm(10, mean = 5, sd = 1),
            rnorm(10, mean = 6, sd = 1),
            rnorm(10, mean = 7, sd = 1))
)

# การวิเคราะห์ ANOVA
anova_result <- aov(value ~ group, data = data)
summary(anova_result)
```

5. การตีความผลลัพธ์

ผลการวิเคราะห์ ANOVA จะให้ค่าความน่าจะเป็น (p-value) ซึ่งใช้ในการตัดสินใจเกี่ยวกับสมมติฐาน:

- หาก p-value < 0.05: ปฏิเสธสมมติฐานหลัก แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของกลุ่ม
- หาก p-value ≥ 0.05: ยอมรับสมมติฐานหลัก แสดงว่าไม่มีความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของกลุ่ม

6. การทำการวิเคราะห์เชิงลึกเพิ่มเติม

หากผลลัพธ์แสดงว่ามีความแตกต่างระหว่างกลุ่ม จำเป็นต้องทำการวิเคราะห์เชิงลึกเพิ่มเติม เช่นการทดสอบ Post-hoc (เช่น Tukey's HSD) เพื่อระบุว่ากลุ่มใดมีความแตกต่างกันบ้าง

การใช้เครื่องมือสถิติในการวิเคราะห์ข้อมูลแปรปรวนจึงเป็นกระบวนการที่ครอบคลุมตั้งแต่การเก็บรวบรวมข้อมูล การเลือกเครื่องมือและ

ซอฟต์แวร์ที่เหมาะสม ไปจนถึงการตีความและสรุปผลการวิเคราะห์ ซึ่งช่วยให้การวิจัยและการตัดสินใจมีความแม่นยำและเชื่อถือได้

สรุป

การวิเคราะห์ความแปรปรวน (ANOVA) เป็นเครื่องมือสถิติที่สำคัญและมีประสิทธิภาพในการเปรียบเทียบค่าเฉลี่ยของกลุ่มหลายกลุ่ม เพื่อพิจารณาว่ามีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ การวิเคราะห์การแจกแจงความแปรปรวนช่วยให้เข้าใจการกระจายตัวของข้อมูลภายในและระหว่างกลุ่ม ซึ่งมีประโยชน์อย่างยิ่งในหลายสาขาวิชาทั้งในงานวิจัยและการประยุกต์ใช้ในอุตสาหกรรม การใช้เครื่องมือสถิติในการวิเคราะห์ข้อมูลแปรปรวนไม่เพียงแต่ทำให้กระบวนการวิเคราะห์มีความแม่นยำและเชื่อถือได้มากขึ้น แต่ยังช่วยลดความซับซ้อนและเวลาในการคำนวณด้วย การใช้ซอฟต์แวร์สถิติเช่น SPSS, R, SAS, และ Python ช่วยให้การวิเคราะห์ข้อมูลเป็นไปอย่างรวดเร็วและมีประสิทธิภาพ การคำนวณด้วยตนเองและการใช้ค่าวิกฤตจากตาราง F-distribution เป็นขั้นตอนที่สำคัญสำหรับการทำความเข้าใจพื้นฐานของการวิเคราะห์ ANOVA นักวิจัยสามารถใช้วิธีนี้ในการตรวจสอบและตีความผลการทดลองได้อย่างถูกต้องและชัดเจน ท้ายที่สุด การวิเคราะห์ ANOVA ช่วยให้เราสามารถทำการตัดสินใจที่มีข้อมูลรองรับและมีหลักฐานทางสถิติที่น่าเชื่อถือ ซึ่งเป็นส่วนสำคัญในการพัฒนาความรู้และการตัดสินใจในหลายสาขา โดยการเข้าใจและใช้เครื่องมือสถิตินี้อย่างถูกต้อง นักวิจัยและผู้ปฏิบัติงานจะสามารถสรุปผลการวิจัยได้อย่างมีประสิทธิภาพและน่าเชื่อถือ

แบบฝึกหัด การวิเคราะห์ความแปรปรวน (ANOVA)

ข้อ 1: การคำนวณค่าเฉลี่ยและความแปรปรวนภายในกลุ่ม

นักเรียนสามกลุ่มได้คะแนนดังนี้:

- กลุ่ม A: 70, 75, 80, 65, 90

- กลุ่ม B: 85, 88, 90, 83, 87
- กลุ่ม C: 78, 74, 80, 82, 76

1.1 คำนวณค่าเฉลี่ยของแต่ละกลุ่ม

1.2 คำนวณความแปรปรวนภายในกลุ่ม (SS within) สำหรับแต่ละกลุ่ม

1.3 รวมค่าความแปรปรวนภายในกลุ่มทั้งหมด

ข้อ 2: การคำนวณค่า F และการใช้ค่าวิกฤต

จากข้อมูลในข้อ 1 และสมมติว่าค่าเฉลี่ยรวมของทั้งสามกลุ่มคือ 80:

2.1 คำนวณความแปรปรวนนอกกลุ่ม (SS between)

2.2 คำนวณค่า MS between และ MS within

2.3 คำนวณค่า F

2.4 ใช้ค่าวิกฤตจากตาราง F-distribution ที่ระดับความเชื่อมั่น 0.05 และ $df_{between}$=2, df_{within}=12เพื่อตัดสินใจว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างกลุ่มหรือไม่

เอกสารอ้างอิง

ปรีดาภรณ์ กาญจนสำราญวงศ์.(2560). *หลักสถิติเบื้องต้น*.นนทบุรี. ไอดีซี พรีเมียร์

ภรณี เจริญภักตร์และคณะ. (2555). *ความน่าจะเป็นและสถิติ*. กรุงเทพฯ : โรง พิมพ์ห้างหุ้นส่วนจำกัดพิทักษ์การพิมพ์

วุฒิ สุขเจริญ. (2561). *วิจัยการตลาด*. กรุงเทพฯ : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

บทที่ 9
สถิตินอนพาราเมตริก

สถิตินอนพาราเมตริก (Non-parametric statistics) เป็นสาขาหนึ่งของสถิติที่มีความสำคัญในการวิเคราะห์ข้อมูลที่ไม่สามารถประยุกต์ใช้กับการกระจายตัวที่กำหนดไว้ล่วงหน้า เช่น การกระจายตัวปกติ ซึ่งแตกต่างจากสถิติพาราเมตริกที่ต้องอาศัยสมมติฐานเกี่ยวกับการกระจายตัวของข้อมูล สถิตินอนพาราเมตริกมักถูกนำมาใช้ในกรณีที่ข้อมูลมีขนาดเล็ก หรือข้อมูลที่ไม่เป็นเชิงเส้น ไม่สมมาตร หรือมีลักษณะเป็นอันดับ การวิเคราะห์ด้วยวิธีนอนพาราเมตริกช่วยให้สามารถประเมินและสรุปข้อมูลได้อย่างมีความยืดหยุ่นและไม่ถูกจำกัดด้วยข้อสมมติฐานที่เข้มงวด ทำให้เป็นเครื่องมือที่มีประโยชน์สำหรับการวิจัยและการวิเคราะห์ข้อมูลในหลายๆ สาขา

ประเภทสถิตินอนพาราเมตริก

สถิตินอนพาราเมตริก (Non-parametric statistics) เป็นเครื่องมือทางสถิติที่ใช้เมื่อข้อมูลไม่เป็นไปตามสมมติฐานของการกระจายตัวที่กำหนด หรือเมื่อไม่สามารถระบุสมมติฐานเกี่ยวกับรูปแบบการกระจายตัวของข้อมูลได้ สถิตินอนพาราเมตริกมีความยืดหยุ่นและเหมาะสมกับข้อมูลที่ไม่เป็นไปตามการกระจายตัวปกติหรือข้อมูลที่มีขนาดเล็ก

ประเภทของสถิตินอนพาราเมตริก

1. การทดสอบอันดับ (Rank Tests)

- การทดสอบวิลคอกซัน (Wilcoxon Rank-Sum Test) ใช้เปรียบเทียบค่ากลางของสองกลุ่มที่ไม่เป็นอิสระต่อกัน
- การทดสอบแมนน์-วิตนีย์ (Mann-Whitney U Test) ใช้เปรียบเทียบค่ากลางของสองกลุ่มอิสระ
- การทดสอบเครื่องหมาย (Sign Test) จัดอยู่ในประเภทการทดสอบอันดับ (Rank Tests) ในกลุ่มสถิตินอนพาราเมตริก โดยเป็นการทดสอบที่ใช้

เพื่อเปรียบเทียบค่ากลางของกลุ่มข้อมูลสองกลุ่ม หรือใช้เพื่อตรวจสอบว่าค่ามัธยฐานของกลุ่มข้อมูลหนึ่งแตกต่างจากค่าที่กำหนดหรือไม่ โดยไม่ต้องอาศัยสมมติฐานเกี่ยวกับการกระจายตัวของข้อมูล

2. การทดสอบเชิงบูรณาการ (Distribution-Free Tests)

• การทดสอบไค-สแควร์ (Chi-Square Test) ใช้ทดสอบความเป็นอิสระหรือความพอดีของข้อมูลกับการกระจายที่คาดการณ์ไว้

• การทดสอบการเปลี่ยนแปลงแมคนีมาร์ (McNemar's Test) ซึ่งเป็นการทดสอบที่ไม่พึ่งพาการกระจายของข้อมูลที่เฉพาะเจาะจง และใช้สำหรับการวิเคราะห์ความเปลี่ยนแปลงของข้อมูลในตารางจำแนกประเภท (contingency table) ที่มีข้อมูลจับคู่กัน (paired data) เช่น ข้อมูลก่อนและหลังการรักษาในกลุ่มเดียวกัน

• การทดสอบฟิชเชอร์ (Fisher’s Exact Test) ใช้สำหรับข้อมูลในตารางจำแนกประเภทขนาดเล็ก

3. การทดสอบความสัมพันธ์ (Correlation Tests)

• สหสัมพันธ์อันดับของสเปียร์แมน (Spearman's Rank Correlation) ใช้สำหรับวัดความสัมพันธ์ระหว่างสองตัวแปรที่เป็นลำดับ

• สหสัมพันธ์อันดับของเคนดัลล์ (Kendall’s Tau) ใช้สำหรับวัดความสัมพันธ์ระหว่างตัวแปรที่ไม่เป็นเชิงเส้น

การคำนวณและการตีความสถิตินอนพาราเมตริก

การคำนวณสถิตินอนพาราเมตริก มักง่ายกว่าการคำนวณสถิติพาราเมตริกเพราะไม่ต้องอาศัยสมมติฐานการกระจายตัวของข้อมูล ตัวอย่างการคำนวณเช่น:

• วิลคอกซัน เทสต์: เรียงลำดับข้อมูลจากน้อยไปมาก และคำนวณอันดับรวมของแต่ละกลุ่ม จากนั้นหาค่าสถิติ Z และเปรียบเทียบกับค่าที่คำนวณได้จากตารางวิลคอกซัน

• สหสัมพันธ์ของสเปียร์แมน: เปลี่ยนค่าข้อมูลเป็นอันดับ และคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอันดับ

การตีความสถิติพาราเมตริก ต้องระวังในเรื่องของขนาดตัวอย่างและสมมติฐานพื้นฐาน เพราะถ้าสมมติฐานไม่เป็นไปตามที่กำหนด ผลลัพธ์ที่ได้อาจไม่แม่นยำ

โดยสรุป สถิตินอนพาราเมตริกเป็นเครื่องมือที่มีประโยชน์สำหรับการวิเคราะห์ข้อมูลที่ไม่สามารถวิเคราะห์ด้วยสถิติพาราเมตริกได้ ช่วยให้สามารถทำการวิเคราะห์และตีความข้อมูลได้อย่างถูกต้องแม้ในกรณีที่ข้อมูลมีข้อจำกัด

การคำนวณสถิตินอนพาราเมตริก

การคำนวณสถิตินอนพาราเมตริกสำหรับการทดสอบกลุ่มตัวอย่างไม่เกิน 2 กลุ่มใช้ในกรณีที่ข้อมูลไม่เป็นไปตามสมมติฐานของการกระจายตัวปกติ หรือเมื่อข้อมูลมีขนาดเล็กและไม่สามารถใช้สถิติพาราเมตริกได้ การทดสอบนอนพาราเมตริกที่นิยมใช้สำหรับกลุ่มตัวอย่างไม่เกิน 2 กลุ่มมีดังนี้

การทดสอบไค-สแควร์

ในการทดสอบไค-สแควร์ทุกรูปแบบจะมีข้อมูลสองชุด ชุดหนึ่งคือข้อมูลที่ได้จากการทดสอบจากตัวอย่างหรือข้อมูลที่เก็บรวบรวมได้จริง เรียกว่าข้อมูลที่ได้จากการสังเกตการณ์ การทดสอบจะเกิดขึ้นก็ต่อเมื่อมีข้อสงสัยที่ต้องการข้อสรุปจากข้อมูลที่ได้มาจากการสังเกตนี้ ข้อสงสัยที่เกิดขึ้นอาจมีได้หลายรูปแบบอย่างไรก็ตามขั้นตอนการทดสอบข้อมูลสงสัยจะถูกตั้งขึ้นเป็นสมมติฐานที่ต้องทำการทดสอบ (H_0) และสมมุติฐานทางเลือกอื่น (H_1) (สำรวม จงเจริญ, 2563) ตัวอย่างเช่น

การทดสอบความเป็นอิสระของข้อมูลสองชุด

H_0 : ข้อมูลสองชุดมีความเป็นอิสระจากกัน

H_1 : ข้อมูลสองชุดไม่เป็นอิสระจากกัน

การเปรียบเทียบค่าสัดส่วนของตัวอย่าง k กลุ่ม

H_0 : $\pi_1 = \pi_2 = \pi_3 = \ldots\ldots\ldots\ldots = \pi_k$

H_1 : ค่าสัดส่วนไม่เท่ากับทุกกลุ่ม

การทดสอบรูปแบบการแจกแจงของข้อมูล

H_0 : ข้อมูลมีการแจกแจงแบบปกติ

H_1 : ข้อมูลมีการแจกแจงไม่เป็นแบบปกติ

จากการตั้งสมมุติฐานในลักษณะที่แสดงตัวอย่างข้างต้น ผู้ทำการทดสอบจะสร้างข้อมูลขึ้นอีกชุดหนึ่งที่มีลักษณะตรงตามที่ระบุไว้ใน H_0 กล่าวคือ ถ้า H_0 เป็นความจริงข้อมูลควรมีลักษณะเป็นอย่างไรผู้ทำการทดสอบจะสร้างข้อมูลให้มีลักษณะอย่างนั้น ข้อมูลชุดที่สองนี้เรียกกว่าข้อมูลที่คาดว่าควรจะเป็น ในขั้นตอนการทดสอบจะทำการเปรียบเทียบความคล้ายคลึงของข้อมูลทั้งสองชุดว่ามีความแตกต่างกันมากน้อยเพียงใดถ้าไม่มีความแตกต่างกันหรือค่าแตกต่างมีน้อยจะยอมรับ H_0 ถ้ามีความแตกต่างกันมากจะยอมรับ H_1 หรืออีกนัยหนึ่งคือ ถ้าค่าแตกต่างเท่ากับศูนย์จะยอมรับ H_0 แต่ถ้าค่าแตกต่างกันมากกว่าศูนย์อย่างมีนัยสำคัญจะยอมรับ H_1

การปรับค่าแตกต่างของข้อมูลทั้งสองชุดให้เป็นค่า X^2 คำนวณได้จากสูตร

$$X_s^2 = \sum \frac{(O-E)^2}{n-1}$$

เมื่อ O คือ ข้อมูลชุดที่ได้จากการสังเกตการณ์

E คือ ข้อมูลชุดที่สร้างขึ้น

ค่า X_s^2 ที่คำนวณได้นี้จะเป็นค่าที่ชี้ว่าค่าแตกต่างของข้อมูลทั้งสองชุดมีนัยสำคัญหรือไม่ โดยนำไปเปรียบเทียบถึงค่า X^2 จากตารางที่ใช้เป็นค่า

ตัดสิน ซึ่งขั้นตอนการหาค่า X_S^2 และค่า X^2 นี้ ในการทดสอบแต่ละรูปแบบแม้จะมีพื้นฐานแนวคิวกันแต่รายละเอียดในการคำนวณแตกต่างกัน ดังอธิบายต่อไป

การทดสอบความเป็นอิสระของตัวแปร 2 ชุด

ข้อมูลที่ใช้ในการทดสอบความเป็นอิสระของตัวแปร คือข้อที่มีมากกว่าหนึ่งกลุ่มและแต่ละกลุ่มแยกเป็นประเภทได้ชัดเจน (ปรีดาภรณ์ กาญจนสำราญวงศ์,2560) ตัวอย่างเช่น ในการสำรวจพฤติกรรมการอ่านหนังสือพิมพ์ของพนักงานในบริษัทผู้ทำการสำรวจได้สอบถามพนักงานในแผนกต่างๆ 500 คน ว่าในเวลาว่างชอบอ่านหนังสือฉบับใด คำตอบที่ได้รับบันทึกได้ตาราง 9.1

ตาราง 9.1 แสดงจำนวนพนักงานแผนกต่างๆ แยกตามประเภทหนังสือพิมพ์ที่ชอบอ่าน

พนักงานแผนก	ประเภทหนังสือพิมพ์ที่ชอบอ่าน			
	ก	ข	ค	รวม
บุคคล	183	107	60	350
บัญชี	54	27	19	100
ตลาด	14	20	16	50
รวม	251	154	95	500

จากข้อมูลชุดที่เก็บรวบรวมได้ในตาราง 9.1 มีตัวแปรเกี่ยวข้อง 2 ชุดเป็นข้อมูลที่แยกตามแผนก (ตัวแปรชุด 1) และแยกตามประเภทหนังสือพิมพ์ที่ชอบอ่าน (ตัวแปรชุด 2) ผู้วิเคราะห์สามารถใช้ข้อมูลเช่น ตัวอย่างนี้ทำการทดสอบไค-สแควร์ หาข้อสรุปได้ว่าตัวแปรทั้งสองชุดมีความเกี่ยวข้องหรือ

มีอิทธิพลต่อกันหรือไม่ หรือการอยู่ต่างแผนกกันเป็นเหตุให้พนักงานชอบอ่านหนังสือต่างฉบับกันหรือไม่เช่นนี้เป็นต้น

สำหรับขั้นตอนการทดสอบทั่วไปเริ่มต้นด้วยการตั้งสมมติฐานดังนี้

H_0 : ตัวแปรทั้งสองชุดมีความเป็นอิสระจากกัน

H_1 : ตัวแปรทั้งสองชุดไม่เป็นอิสระต่อกัน

ในขั้นตอนการตั้งสมมติฐานนี้ ผู้ทดสอบจะต้องจัดทำข้อมูลชุดที่คาดว่าจะเป็นถ้า H_0 ถูกต้อง (ชุดE) เพื่อที่สามารถนำไปเปรียบเทียบกับข้อมูลที่สังเกตการณ์ได้ (ชุด O) ตัวอย่างจากข้อมูลในตาราง 9.1 เป็นข้อมูลชุดที่สังเกตการณ์ได้ซึ่งมี 3 แถว (r = 3) และ 3 คอลัมน์(K- 3) ข้อมูลชุดที่คาดว่าจะเป็นถ้าH_0 ถูกต้อง ก็จะต้องมี 3 แถว และ 3 คอลัมน์ เช่นกัน

วิธีการจัดทำข้อมูลชุดที่คาดว่าจะเป็นถ้า H_0 ถูกต้อง มีหลักการอยู่ว่าถ้าตัวแปรชุด 1 และตัวแปรชุดที่ 2 ไม่มีความสัมพันธ์กันจริง (การทำงานอยู่ต่างแผนกกันไม่มีผลต่อพฤติกรมในการอ่าน) จำนวนพนักงาน ในแต่ละแผนกที่ชอบอ่านหนังสือพิมพ์แต่ละประเภทควรเป็นสัดส่วนเดียวกัน นั่นคือจากพนักงานทั้งหมด 500 คน (โดยไม่คำนึงถึงว่าจะทำงานอยู่ในแผนกใด) มีผู้ชอบอ่านหนังสือพิมพ์ ก. 251 คนหนังสือพิมพ์ ข.154 คน และหนังสือพิมพ์ ค. 95 คน

สัดส่วนของพนักงานที่ชอบอ่านหนังสือ ก. = $\frac{251}{500}$

สัดส่วนของพนักงานที่ชอบอ่านหนังสือ ข. = $\frac{154}{500}$

สัดส่วนของพนักงานที่ชอบอ่านหนังสือ ค. = $\frac{95}{500}$

ดังนั้นในแผนกบุคคลซึ่งมีพนักงาน 350 คน ถ้า H_0 เป็นจริง จำนวนผู้ที่ชอบอ่านหนังสือพิมพ์ทั้งสามฉบับควรมีดังนี้

ชอบอ่านหนังสือ ก. = $\frac{251}{500}$ x 350 = 175.7 คน

ชอบอ่านหนังสือ ข. = $\frac{154}{500}$ x 350 = 107.8 คน

ชอบอ่านหนังสือ ค. = $\frac{95}{500}$ x 350 = 66.5 คน

การคำนวณหาจำนวนผู้ชอบอ่านหนังสือทั้งสามฉบับ สำหรับแผนกบัญชีและแผนกตลาดคำนวณในทำนองเดียวกัน ข้อมูลชุดที่คาดว่าจะเป็นถ้า H_0 ถูกต้อง จึงมีรูปแบบดังแสดงในตาราง 9.2

ตาราง 9.2 แสดงจำนวนพนักงานแผนกต่างๆ แยกตามประเภทหนังสือพิมพ์ที่ชอบอ่านถ้า H_0 เป็นจริง (ข้อมูลชุด E)

พนักงานแผนก	ประเภทหนังสือพิมพ์ที่ชอบอ่าน			
	ก	ข	ค	รวม
บุคคล	175.7	107.8	66.5	350.0
บัญชี	50.2	30.8	19.0	100.0
ตลาด	25.1	15.4	9.5	50.0
รวม	251.0	154.0	95.0	500

จากตาราง 9.1 และตาราง 9.2 คือข้อมูลชุดที่ได้จาการสังเกตการณ์ (O) และข้อมูลชุดที่คาดว่าจะ เป็นถ้า H_0 เป็นจริง (E) ผู้ทดสอบจะเปรียบเทียบตารางทั้งสองว่ามีความคล้ายคลึงกันมากน้อยเพียงใด ถ้าเหมือนกันหรือคล้ายคลึงกันจนสรุปได้ว่าไม่มีความแตกต่างกันอย่างมีนัยสำคัญแล้ว ข้อมูลชุด O จะเป็นข้อมูลที่ยืนยันได้ว่า H_0 เป็นจริงด้วย นั่นคือ ยอมรับ H_0 ว่าเป็นจริง แต่ถ้าข้อมูลทั้งสองชุดมีความแตกต่างกันอย่างมีนัยสำคัญหมายความว่าข้อมูลชุด O จะเป็นข้อมูลที่ชี้ว่า H_0 ไม่เป็นจริงและยอมรับ H_1

ในการเปรียบเทียบข้อมูลทั้งสองชุดจะดูที่ความแตกต่างของค่าแต่ละค่า และปรับเป็นค่า χ^2 ค่า χ^2 ที่คำนวณได้นี้จะเป็นค่า χ_s^2 ที่ใช้ตัดสินต่อไป การคำนวณทำได้ดังนี้

จากสูตร $\chi_s^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

เมื่อ r = จำนวนแถวในตาราง

K = จำนวนคอลัมน์ในตาราง

O_{ij} = ข้อมูลชุด O

E_{ij} = ข้อมูลชุด E

แทนค่าจากตารางทั้งสองในสูตร

$$\chi_s^2 = \frac{(183.0-175.7)^2}{175.7} + \frac{(107.0-107.8)^2}{107.8} + \frac{(60.0-66.5)^2}{66.5}$$
$$+ \frac{(54.0-50.2)^2}{50.2} + \frac{(27.0-30.8)^2}{30.8} + \frac{(19.0-19.0)^2}{19.0}$$
$$+ \frac{(14.0-25.1)^2}{25.1} + \frac{(20.0-15.4)^2}{15.4} + \frac{(16.0-9.5)^2}{9.5}$$
$$= 12.431$$

ค่า χ_s^2 ที่คำนวณได้ชี้ให้เห็นว่าข้อมูลชุด O และชุด E มีความแตกต่างกันระดับหนึ่ง เพราะ $\chi_s^2 > 0$ แต่ค่า χ_s^2 มากกว่า 0 อย่างมีนัยสำคัญหรือไม่ ต้องนำไปเปรียบเทียบกับค่าวิกฤต การกำหนดค่าวิกฤตทำได้ดังนี้

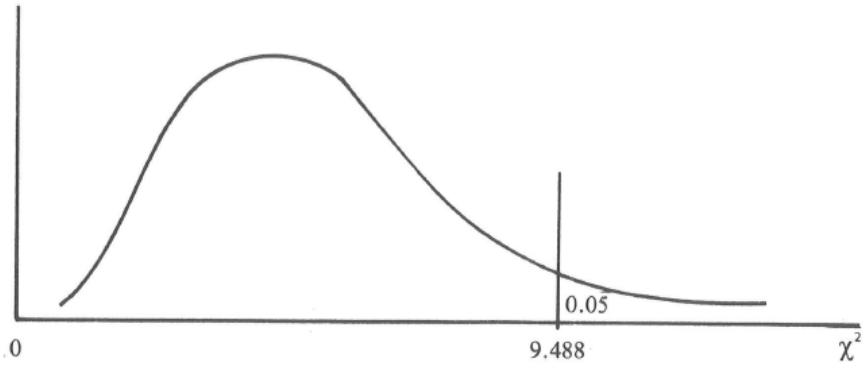
สมมุติให้การทดสอบทำด้วยระดับนัยสำคัญ 0.05 (α = 0.05) และปัญหาข้อนี้ข้อมูลอยู่ในลักษณะตารางที่มี 3 แถว (r = 3) และ 3 คอลัมน์ (k = 3) ซึ่งข้อมูลลักษณะนี้ค่า df คำนวณได้ดังนี้

$$df = (r - 1)(k - 1)$$

$= \quad (3 - 1)(3 - 1)$

$= \quad 4$

ค่า χ^2 ที่ใช้เป็นค่าตัดสินคือค่า $\chi^2_{(0.05,4)} = 9.488$ (จากตารางไค-สแควร์) ดูภาพ 9.1 ประกอบ



ภาพ 9.1 แสดงค่าวิกฤต $\chi^2_{(0.05,4)}$ จากตางรางไค-สแควร์

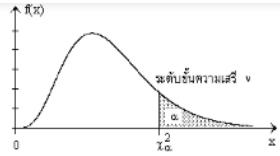
ค่า $\chi^2_{(0.05,4)}$ คือค่าที่ใช้เป็นเป็นเกณฑ์ในการตัดสินใจ กล่าวคือ ถ้า χ^2_s มีค่าน้อยกว่า $\chi^2_{(0.05,4)}$ จะถือว่าค่า χ^2_s มีค่าเท่ากับศูนย์คือต่างจากศูนย์อย่างไม่มีนัยสำคัญ แต่ถ้า χ^2_s มีค่ามากกว่า $\chi^2_{(0.05,4)}$ จะสรุปว่าค่า χ^2_s มากกว่าศูนย์อย่างมีนัยสำคัญ การตัดสินใจสำหรับปัญหานี้จึงเขียนได้ว่าจะไม่ยอมรับ H_0 ถ้า $\chi^2_s > 9.488$

ในขั้นตอนการสรุปผลการทดสอบ โดยเหตุที่ค่า $\chi^2_s = 12.431$ ซึ่งมีค่ามากกว่า 9.488 ปัญหาข้อนี้จึงไม่ยอมรับ H_0 ด้วยระดับนัยสำคัญ 0.05 และสรุปว่าการทำงานต่างแผนกกันมีผลทำให้พนักงานนิยมอ่าน หนังสือพิมพ์ต่างกัน หรือการทำงานต่างแผนกมีผลต่อพฤติกรรมการอ่านหนังสือพิมพ์

ตารางที่ 5. ตารางค่าวิกฤตของการแจกแจงไคสแควร์ χ^2

เมื่อระดับขั้นความเสรีเท่ากับ ν

χ^2_α คือค่าที่ทำให้ $P(\chi^2 > \chi^2_\alpha) = \alpha$



ν	α									
	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.0^4393	0.0^3257	0.0^3982	0.0^2393	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275

การทดสอบแมนน์-วิตนีย์ (Mann-Whitney U Test)

ใช้เพื่อเปรียบเทียบค่ากลางของสองกลุ่มอิสระ โดยมี สมมติฐาน (Null hypothesis, H_0): การแจกแจงของสองกลุ่มเท่ากัน และ H_1: การแจกแจงของสองกลุ่มไม่เท่ากัน

วิธีการคำนวณ

- รวมข้อมูลจากทั้งสองกลุ่มเข้าด้วยกันและจัดอันดับ (rank) ข้อมูล
- คำนวณอันดับรวม (sum of ranks) ของแต่ละกลุ่ม
- ใช้สูตรในการคำนวณค่า U

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

ที่นี่ n_1 และ n_2 คือขนาดของกลุ่ม 1 และกลุ่ม 2 ตามลำดับ, R_1 และ R_2 คืออันดับรวมของแต่ละกลุ่ม

- ค่า U ที่ใช้คือค่าน้อยที่สุดระหว่าง U_1 และ U_2

ตัวอย่างการคำนวณ Mann-Whitney U Test

สมมติว่ามีข้อมูลสองกลุ่มอิสระ

กลุ่ม 1: [5, 7, 8, 6, 9]

กลุ่ม 2: [6, 9, 7, 10, 8]

1. รวมข้อมูลจากทั้งสองกลุ่มและจัดอันดับ

   ข้อมูลทั้งหมด: [5, 6, 6, 7, 7, 8, 8, 9, 9, 10]

   อันดับ: [1, 2.5, 2.5, 4.5, 4.5, 6.5, 6.5, 8.5, 8.5, 10]

อันดับ (ranks) ถูกจัดตามค่าของข้อมูลจากน้อยไปมาก โดยค่าที่เท่ากันจะได้รับการจัดอันดับเฉลี่ย (average ranks) ร่วมกัน ดังนั้นเราจะเริ่มจากการรวมข้อมูลจากทั้งสองกลุ่ม และจัดเรียงตามค่าที่เพิ่มขึ้น จากนั้นจึงกำหนดอันดับให้แต่ละค่า

ขั้นตอนการจัดอันดับ

1) รวมข้อมูลจากกลุ่ม 1 และกลุ่ม 2
   - ข้อมูลทั้งหมด: [5, 6, 7, 8, 9, 6, 9, 7, 10, 8]

2) จัดเรียงข้อมูลทั้งหมดจากน้อยไปมาก:
    - ข้อมูลเรียง: [5, 6, 6, 7, 7, 8, 8, 9, 9, 10]
3) กำหนดอันดับให้แต่ละค่า:
    - 5 ได้อันดับที่ 1
    - 6 ทั้งสองค่าได้อันดับที่ 2 และ 3 ดังนั้นอันดับเฉลี่ยจะเป็น (2 + 3) / 2 = 2.5
    - 7 ทั้งสองค่าได้อันดับที่ 4 และ 5 ดังนั้นอันดับเฉลี่ยจะเป็น (4 + 5) / 2 = 4.5
    - 8 ทั้งสองค่าได้อันดับที่ 6 และ 7 ดังนั้นอันดับเฉลี่ยจะเป็น (6 + 7) / 2 = 6.5
    - 9 ทั้งสองค่าได้อันดับที่ 8 และ 9 ดังนั้นอันดับเฉลี่ยจะเป็น (8 + 9) / 2 = 8.5
    - 10 ได้อันดับที่ 10

ดังนั้นอันดับจะเป็นดังนี้

- ข้อมูลทั้งหมด: [5, 6, 6, 7, 7, 8, 8, 9, 9, 10]
- อันดับ: [1, 2.5, 2.5, 4.5, 4.5, 6.5, 6.5, 8.5, 8.5, 10]

สรุปอันดับที่จัดได้

- 5 ได้อันดับที่ 1
- 6 ได้อันดับเฉลี่ย 2.5
- 6 ได้อันดับเฉลี่ย 2.5
- 7 ได้อันดับเฉลี่ย 4.5
- 7 ได้อันดับเฉลี่ย 4.5
- 8 ได้อันดับเฉลี่ย 6.5
- 8 ได้อันดับเฉลี่ย 6.5
- 9 ได้อันดับเฉลี่ย 8.5
- 9 ได้อันดับเฉลี่ย 8.5
- 10 ได้อันดับที่ 10

2. แยกอันดับของแต่ละกลุ่ม:

กลุ่ม 1: [1, 4.5, 6.5, 2.5, 8.5]

กลุ่ม 2: [2.5, 8.5, 4.5, 10, 6.5]

การแยกอันดับของแต่ละกลุ่ม

หลังจากที่ได้จัดอันดับข้อมูลรวมกันแล้ว ขั้นตอนต่อไปคือการแยกอันดับตามแต่ละกลุ่มตามค่าของข้อมูลในแต่ละกลุ่มเดิม

ข้อมูลสองกลุ่มอิสระ

กลุ่ม 1: [5, 7, 8, 6, 9]

กลุ่ม 2: [6, 9, 7, 10, 8]

อันดับที่จัดไว้สำหรับข้อมูลรวม

- ข้อมูลทั้งหมด: [5, 6, 6, 7, 7, 8, 8, 9, 9, 10]
- อันดับ: [1, 2.5, 2.5, 4.5, 4.5, 6.5, 6.5, 8.5, 8.5, 10]

แยกอันดับของแต่ละกลุ่ม

จากนั้นทำการแยกอันดับตามค่าของข้อมูลในแต่ละกลุ่มเดิม ดังนี้

กลุ่ม 1:

- ค่า 5 ได้อันดับที่ 1
- ค่า 7 ได้อันดับเฉลี่ยที่ 4.5
- ค่า 8 ได้อันดับเฉลี่ยที่ 6.5
- ค่า 6 ได้อันดับเฉลี่ยที่ 2.5
- ค่า 9 ได้อันดับเฉลี่ยที่ 8.5

ดังนั้นอันดับของ **กลุ่ม 1** คือ:

- **กลุ่ม 1**: [1, 4.5, 6.5, 2.5, 8.5]

กลุ่ม 2:

- ค่า 6 ได้อันดับเฉลี่ยที่ 2.5
- ค่า 9 ได้อันดับเฉลี่ยที่ 8.5
- ค่า 7 ได้อันดับเฉลี่ยที่ 4.5
- ค่า 10 ได้อันดับที่ 10

- ค่า 8 ได้อันดับเฉลี่ยที่ 6.5

ดังนั้นอันดับของ **กลุ่ม 2** คือ:

- **กลุ่ม 2**: [2.5, 8.5, 4.5, 10, 6.5]

สรุปการแยกอันดับของแต่ละกลุ่ม:

- **กลุ่ม 1**: [1, 4.5, 6.5, 2.5, 8.5]
- **กลุ่ม 2**: [2.5, 8.5, 4.5, 10, 6.5]

3. คำนวณอันดับรวม:
    - R_1 = 1 + 4.5 + 6.5 + 2.5 + 8.5 = 23
    - R_2 = 2.5 + 8.5 + 4.5 + 10 + 6.5 = 32
4. คำนวณค่า U
    - U_1=5 x 5+$\frac{5(5+1)}{2}$−23 = 25+15−23= 17
    - U_2=5 x 5+$\frac{5(5+1)}{2}$−32 = 25+15−32= 8
    - ค่า U ที่ใช้คือค่าน้อยที่สุดระหว่าง U_1 และ U_2, ดังนั้น U=8

การตัดสินใจ:

- ค่าน้อยที่สุดของ U คือ 8
- เพื่อการตัดสินใจว่าค่าที่ได้มีนัยสำคัญทางสถิติหรือไม่ จำเป็นต้องเปรียบเทียบค่า U กับค่าความน่าจะเป็นวิกฤต (Critical Value) จากตาราง Mann-Whitney U Test ที่ระดับนัยสำคัญ (เช่น 0.05)
- หากค่า U น้อยกว่าหรือเท่ากับค่า Critical Value ในตารางที่ระดับนัยสำคัญที่กำหนด เราจะปฏิเสธสมมติฐาน H_0 และสรุปว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างสองกลุ่ม

หมายเหตุ: ค่าความน่าจะเป็นวิกฤตสำหรับการเปรียบเทียบนี้ขึ้นอยู่กับขนาดของตัวอย่าง (n1 และ n2) และระดับนัยสำคัญที่เลือกใช้ (เช่น 0.05)

ค่าตารางสำหรับ Mann-Whitney U Test สามารถดูได้จากแหล่งอ้างอิงที่เกี่ยวข้องกับการทดสอบนี้ เช่น ตารางในหนังสือสถิติหรือเครื่องมือออนไลน์ที่ใช้ในการคำนวณค่าพี (p-value)

เปรียบเทียบค่าที่ได้กับตาราง Mann-Whitney U

สำหรับการเปรียบเทียบค่าที่ได้กับตาราง Mann-Whitney U จำเป็นต้องทราบค่าจำนวนข้อมูลในแต่ละกลุ่ม (n1 และ n2) และค่าระดับนัยสำคัญ (α) ที่กำหนดไว้ล่วงหน้า เช่น α=0.05 จากนั้นจึงใช้ตาราง Mann-Whitney U เพื่อดูค่า U-critical

กรณีที่ใช้ค่า U หาก$n_1 \leq 8$ และ$n_2 \leq 8$ (ปุญญพัฒน์, ไชยเมล์,2555)

กรณีที่เป็นการทดสอบทางเดียว จะปฏิเสธสมมติฐาน H_0 เมื่อค่าความน่าจะเป็นที่เปิดได้จากตาราง (ตามค่า n_1 , n_2 และ U ที่คำนวณได้) มีน้อยกว่าหรือเท่ากับระดับนัยสำคัญ (ค่าα) ที่กำหนด

กรณีที่เป็นการทดสอบสองทาง จะปฏิเสธสมมติฐาน H_0 เมื่อค่าความน่าจะเป็นที่เปิดได้จากตาราง (ตามค่า n_1 , n_2 และ U ที่คำนวณได้) มีน้อยกว่าหรือเท่ากับ $\frac{\alpha}{2}$

กรณีที่ใ ช้ค่า U หาก $9 \leq n_1 \leq 20$ หรือ $9 \leq n_2 \leq 20$

จะปฏิเสธสมมติฐาน H_0 เมื่อค่า U ที่คำนวณได้น้อยกว่าหรือเท่ากับค่าวิกฤตของ U จากตาราง

ข้อมูลที่มี

- n1=5
- n2=5
- ค่า U ที่คำนวณได้: U=8
- ระดับนัยสำคัญ α=0.05 สำหรับการทดสอบสองด้าน (two-tailed test)

ตาราง Mann-Whitney U (ตารางบางส่วน)

n1	n2	α=0.05 (two-tailed)
5	5	2

ในกรณีนี้ ค่าที่กำหนดไว้ในตาราง Mann-Whitney U สำหรับ n1=5 และ n2=5 ที่ระดับนัยสำคัญ 0.05 (สองด้าน) คือ 2

การเปรียบเทียบ

- ค่า U ที่คำนวณได้: U=8
- ค่า U-critical ที่ตารางกำหนด: U critical=2

สรุปผลการทดสอบ

เนื่องจาก U ที่คำนวณได้ (8) มากกว่า U critical (2) จึงไม่สามารถปฏิเสธสมมติฐานศูนย์ (H_0) ได้

คำสรุป

ไม่มีหลักฐานเพียงพอที่จะสรุปว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างค่ากลางของสองกลุ่ม ดังนั้น จึงยอมรับสมมติฐานศูนย์ H_0ว่าไม่มีความแตกต่างระหว่างสองกลุ่มที่ระดับนัยสำคัญ 0.05

Critical Values of the Mann-Whitney U
(Two-Tailed Testing)

n_2	α	n_1																	
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	3	3
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	.05	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	.01	--	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
	.05	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48

การทดสอบสัญญาณ

การทดสอบสัญญาณ (Sign Test) เป็นการทดสอบนอนพาราเมตริกที่ใช้สำหรับการเปรียบเทียบค่ามัธยฐานของสองกลุ่มที่เกี่ยวข้องกัน โดยเฉพาะในกรณีที่ข้อมูลไม่มีการกระจายตัวที่ทราบ หรือมีการกระจายตัวที่ไม่เป็นปกติ การทดสอบนี้เป็นเครื่องมือที่ง่ายและทรงพลังในการตรวจสอบว่าค่ามัธยฐานของความแตกต่างระหว่างค่าคู่มีความแตกต่างจากศูนย์หรือไม่

ขั้นตอนการทดสอบสัญญาณ (Sign Test)

1. **กำหนดสมมติฐาน**

สมมติฐานที่ศึกษา (Null Hypothesis, H_0): ค่ามัธยฐานของ ความแตกต่างระหว่างคู่ข้อมูลเท่ากับศูนย์

สมมติฐานที่เชื่อ (Alternative Hypothesis, H_1): ค่ามัธยฐานของความแตกต่างระหว่างคู่ข้อมูลไม่เท่ากับศูนย์

2. **การคำนวณ**
    - หาความแตกต่างของคู่ข้อมูลแต่ละคู่ (เช่น ความแตกต่างระหว่างการวัดก่อนและหลังการทดลอง)
    - แทนความแตกต่างที่เป็นบวกด้วยเครื่องหมาย "+" และความแตกต่างที่เป็นลบด้วยเครื่องหมาย "-" โดยไม่นับค่าความแตกต่างที่เป็นศูนย์
    - นับจำนวนครั้งที่เครื่องหมาย "+" และ "-" ปรากฏ
3. **การวิเคราะห์**
    - จำนวนครั้งที่เครื่องหมาย "+" และ "-" ปรากฏจะถูกใช้เพื่อคำนวณค่าสถิติการทดสอบ
    - ถ้าไม่มีความแตกต่างอย่างมีนัยสำคัญ จะคาดหวังว่าเครื่องหมาย "+" และ "-" จะปรากฏใกล้เคียงกัน
    - ใช้ตารางการแจกแจงทวินาม (Binomial Distribution) เพื่อหาค่าความน่าจะเป็น (p-value) สำหรับจำนวนครั้งที่เครื่องหมาย "+" และ "-" ปรากฏ

ตัวอย่างการทดสอบสัญญาณ ด้วยการใช้การแจกแจงทวินาม (Binomial Distribution)

ข้อมูลที่ใช้: น้ำหนักแมว (ก.ก.)

ก่อนการรักษา: [4, 5, 6, 7, 8, 9, 5, 6, 7, 8]

หลังการรักษา: [5, 6, 5, 8, 7, 10, 4, 7, 6, 9]

ความแตกต่างของค่าคู่:

ความแตกต่าง: [1, 1, -1, 1, -1, 1, -1, 1, -1, 1]

แทนความแตกต่างด้วยเครื่องหมาย: [+, +, -, +, -, +, -, +, -, +]

การนับจำนวนเครื่องหมาย:

จำนวนครั้งที่เป็น "+": 6

จำนวนครั้งที่เป็น "-": 4

การวิเคราะห์ผลโดยใช้การแจกแจงทวินาม:

ภายใต้สมมติฐาน H_0 (ไม่มีความแตกต่าง) เครื่องหมาย "+" และ "-" มีโอกาสเกิดขึ้นเท่ากันที่ 0.5

การทดสอบนี้เป็นการแจกแจงทวินาม Binomial(n=10,p=0.5)

ขั้นตอนการทดสอบสมมติฐาน

H_0: p=0.5 (ไม่มีความแตกต่าง)

H_1: p≠0.5 (มีความแตกต่าง)

α=0.05 (สองด้าน)

คำนวณค่า test statistic

ค่าที่น้อยที่สุดของจำนวนครั้งที่เป็นเครื่องหมายบวกหรือลบ คือ 4

หาค่าวิกฤต (Critical Value) จากตารางการแจกแจงทวินาม

n=10 และ p=0.5

ตารางการแจกแจงทวินาม

k	$P(X \leq k)$	$P(X \geq k)$
0	0.00098	1.0000
1	0.01074	0.99902
2	0.05469	0.98926
3	0.17188	0.94531
4	0.37695	0.82812
5	0.62305	0.62305
6	0.82812	0.37695
7	0.94531	0.17188
8	0.98926	0.05469
9	0.99902	0.01074
10	1.0000	0.00098

จากตาราง

สำหรับ $P(X \leq k_1)$ ใกล้เคียง 0.025 พบว่า $k_1=2$ ($P(X \leq 2) = 0.05469$)

สำหรับ $P(X \geq k_2)$ ใกล้เคียง 0.025 พบว่า $k_2=8$ ($P(X \geq 8) = 0.05469$)

การตัดสินใจ

ค่า test statistic = 4 , จากตาราง ค่าความน่าจะเป็นสะสม (critical value) คือ $k_1=2$ และ $k_2=8$, ถ้า test statistic ≤ 2 หรือ test statistic ≥ 8, เราจะปฏิเสธสมมติฐานเริ่มต้น H_0 ในกรณีนี้ test statistic = 4 ซึ่งไม่อยู่ในเขตการปฏิเสธสมมติฐานเริ่มต้น ดังนั้น จึงไม่สามารถปฏิเสธสมมติฐานเริ่มต้น H_0 ได้ ซึ่งหมายความว่าไม่มีหลักฐานเพียงพอที่จะกล่าวว่าน้ำหนักของแมวก่อนและหลังการรักษามีความแตกต่างกันอย่างมีนัยสำคัญ

การทดสอบสัญญาณ (Sign Test) โดยการใช้ตารางสามารถทำได้โดยใช้ตารางค่าความน่าจะเป็น (Critical Values Table) ที่แสดงค่าวิกฤตสำหรับการทดสอบที่ระดับนัยสำคัญต่างๆ เช่น 0.05 หรือ 0.01 วิธีนี้จะช่วยให้ตัดสินใจได้ว่าความแตกต่างที่พบมีนัยสำคัญหรือไม่

การทดสอบการเปลี่ยนแปลงแมคนีมาร์

การทดสอบการเปลี่ยนแปลงแมคนีมาร์ (McNemar's Test) จัดอยู่ในประเภทของ **การทดสอบเชิงบูรณาการ (Distribution-Free Tests)** ซึ่งเป็นการทดสอบที่ไม่พึ่งพาการกระจายของข้อมูลที่เฉพาะเจาะจง และใช้สำหรับการวิเคราะห์ความเปลี่ยนแปลงของข้อมูลในตารางจำแนกประเภท (contingency table) ที่มีข้อมูลจับคู่กัน (paired data) เช่น ข้อมูลก่อนและหลังการรักษาในกลุ่มเดียวกัน McNemar's Test จัดอยู่ในประเภทของการทดสอบเชิงบูรณาการ (Distribution-Free Tests) เช่นเดียวกับการทดสอบไค-สแควร์ (Chi-Square Test) และการทดสอบฟิชเชอร์ (Fisher’s Exact Test)

การทดสอบแมคนีมาร์ (McNemar's Test) เป็นการทดสอบทางสถิติที่ใช้สำหรับเปรียบเทียบสัดส่วนของผลลัพธ์ที่เปลี่ยนแปลงในข้อมูลที่จับคู่กัน เช่น การเปลี่ยนแปลงพฤติกรรมก่อนและหลังการรักษาในกลุ่มเดียวกัน หรือการทดสอบผลของสองวิธีการที่แตกต่างกันในกลุ่มตัวอย่างเดียวกัน

ขั้นตอนการทดสอบแมคนีมาร์

1. **สร้างตาราง 2x2**: สร้างตารางความถี่ที่แสดงจำนวนของคู่ของตัวอย่างที่เปลี่ยนแปลงไปในแต่ละกรณี โดยแบ่งออกเป็น 4 กลุ่มตามตารางต่อไปนี้

	ผลลัพธ์หลัง เป็นบวก (Positive)	ผลลัพธ์หลัง เป็นลบ (Negative)
ผลลัพธ์ก่อน เป็นบวก (Positive)	a	b
ผลลัพธ์ก่อน เป็นลบ (Negative)	c	d

2. **คำนวณสถิติแมคนีมาร์**: ใช้สูตรต่อไปนี้ในการคำนวณสถิติแมคนีมาร์

$$\chi^2 = \frac{(b-c)^2}{b+c}$$

โดยที่ b และ c เป็นค่าความถี่จากตาราง 2x2

3. **ตรวจสอบระดับนัยสำคัญ**: เปรียบเทียบค่าสถิติที่คำนวณได้กับค่า χ^2จากตารางการแจกแจงไค-สแควร์ (Chi-Square distribution) ที่มีอิสระระดับ 1 (df = 1) และระดับนัยสำคัญ (α) ที่กำหนด เช่น 0.05

ตัวอย่างการคำนวณแมคนีมาร์

สมมติว่ามีการศึกษาพฤติกรรมการสูบบุหรี่ก่อนและหลังการรักษาในกลุ่มตัวอย่าง 50 คน โดยมีผลลัพธ์ตามตารางดังนี้

	สูบบุหรี่หลังการรักษา (Yes)	ไม่สูบบุหรี่หลังการรักษา (No)
สูบบุหรี่ก่อนการรักษา (Yes)	30	10
ไม่สูบบุหรี่ก่อนการรักษา (No)	5	5

จากตารางนี้

a=30, b=10, c=5, d=5

4. **คำนวณสถิติแมคนีมาร์**

$$\chi^2 = \frac{(b-c)^2}{b+c} = \frac{(10-5)^2}{10+5} = \frac{25}{15} \approx 1.67$$

5. **เปรียบเทียบค่า**ที่คำนวณได้ (χ^2=1.67) กับค่า χ^2จากตารางไค-สแควร์ที่มีอิสระระดับ 1 (df = 1) ที่ระดับนัยสำคัญ 0.05 ซึ่งมีค่า $\chi^2_{0.05,1}$=3.841

เนื่องจากค่า χ^2=1.67น้อยกว่าค่า $\chi^2_{critical}$=3.841, ดังนั้นจึงไม่สามารถปฏิเสธสมมติฐานศูนย์ได้

6. **สรุปผล**

สมมติฐานศูนย์ (Null Hypothesis, H_0): ไม่มีความแตกต่างอย่างมีนัยสำคัญระหว่างผลลัพธ์ก่อนและหลังการรักษา

สมมติฐานทางเลือก (Alternative Hypothesis, H_a): มีความแตกต่างอย่างมีนัยสำคัญระหว่างผลลัพธ์ก่อนและหลังการรักษา

เนื่องจากค่า χ^2=1.67 น้อยกว่าค่า $\chi^2_{critical}$=3.841 จึงไม่สามารถปฏิเสธสมมติฐานศูนย์ H_0 ได้ ดังนั้น ไม่มีหลักฐานเพียงพอที่จะสรุปว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างผลลัพธ์ก่อนและหลังการรักษา

หมายเหตุ หาก b+c มีค่าน้อยกว่า 25 ควรใช้การทดสอบแมคนีมาร์แบบมีการปรับค่าคอนติเนินซี (Continuity Correction) โดยใช้สูตร

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c}$$

การทดสอบลำดับที่วิลคอกซอน

การทดสอบลำดับที่วิลคอกซอน (Wilcoxon Signed-Rank Test) เป็นสถิตินอนพาราเมตริกที่ใช้สำหรับเปรียบเทียบค่ามัธยฐานของสองกลุ่มที่ไม่เป็นอิสระต่อกัน (paired samples) โดยเฉพาะเมื่อข้อมูลไม่เป็นไปตามการแจกแจงปกติ การทดสอบนี้ใช้แทนการทดสอบที (t-test) สำหรับข้อมูลที่จับคู่กันหรือข้อมูลที่ซ้ำกัน (repeated measures) (อรุณ จิรวัฒน์กุล, 2558)

ขั้นตอนการทดสอบวิลคอกซอน

1. **คำนวณความแตกต่างของค่าคู่ที่จับกัน**: คำนวณความแตกต่างระหว่างค่าคู่ (D) เช่น $D_i=X_{i1}-X_{i2}$ โดยที่ X_{i1} และ X_{i2} เป็นค่าคู่ที่จับกันสำหรับกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ
2. **ตัดค่าความแตกต่างที่เป็นศูนย์**: ถ้าความแตกต่างเป็นศูนย์ ให้ตัดค่านั้นออกไป
3. **จัดอันดับค่าความแตกต่าง**: จัดอันดับค่าความแตกต่างที่เหลือโดยไม่สนใจเครื่องหมาย
4. **กำหนดเครื่องหมายให้กับอันดับ**: กำหนดเครื่องหมาย (+ หรือ -) ให้กับอันดับตามเครื่องหมายของค่าความแตกต่างเดิม
5. **คำนวณผลรวมของอันดับบวกและอันดับลบ**: คำนวณผลรวมของอันดับบวก (T^+) และอันดับลบ (T^-)
6. **หาค่าต่ำสุดของอันดับรวม**: คำนวณค่าที่น้อยที่สุดระหว่าง T^+ และ T^- ซึ่งเป็นสถิติวิลคอกซอน (W)
7. **เปรียบเทียบกับค่าที่คาดหวัง**: เปรียบเทียบกับค่าที่คาดหวังตามตารางวิลคอกซอน

ตัวอย่างการคำนวณ

สมมติว่ามีการศึกษาความแตกต่างในความสูงของต้นไม้ก่อนและหลังการใช้ปุ๋ยในกลุ่มตัวอย่าง 8 ต้นไม้ดังนี้:

ความสูงก่อน: [58, 60, 65, 59, 61, 60, 62, 63]

ความสูงหลัง: [62, 61, 67, 60, 63, 59, 64, 65]

1. คำนวณความแตกต่าง

ความสูงก่อน	ความสูงหลัง	ความแตกต่าง(D)
58	62	-4
60	61	-1
65	67	-2
59	60	-1
61	63	-2

60	59	1
62	64	-2
63	65	-2

ในการจัดอันดับความแตกต่างโดยไม่สนใจเครื่องหมายสำหรับการทดสอบวิลคอกซอน (Wilcoxon Signed-Rank Test) จะดำเนินการตามขั้นตอนดังนี้

1. **เรียงลำดับความแตกต่างที่ไม่สนใจเครื่องหมาย**: เรียงค่าความแตกต่างในลำดับจากน้อยไปมาก
2. **กำหนดอันดับ**: กำหนดอันดับให้กับค่าความแตกต่างที่เรียงลำดับแล้ว
3. **จัดการกับค่าที่ซ้ำกัน**: หากมีค่าที่ซ้ำกัน ให้ใช้อันดับเฉลี่ย (average rank)

ขั้นตอนการคำนวณอันดับ

ข้อมูลความแตกต่างโดยไม่สนใจเครื่องหมาย:

$$|D| = [4,1,2,1,2,1,2,2]$$

1. **เรียงลำดับจากน้อยไปมาก**:

$$[1,1,1,2,2,2,2,4]$$

2. **จัดการกับค่าที่ซ้ำกัน**:
   - ค่าที่เป็น 1 มี 3 ค่า: อันดับที่ 1, 2, 3 ดังนั้นเฉลี่ย = (1+2+3)/3 = 2
   - ค่าที่เป็น 2 มี 4 ค่า: อันดับที่ 4, 5, 6, 7 ดังนั้นเฉลี่ย = (4+5+6+7)/4 = 5.5

สรุปการจัดอันดับ

ความแตกต่าง	**อันดับ (Rank)**
1	2
1	2
1	2
2	5.5

ความแตกต่าง	อันดับ (Rank)
2	5.5
2	5.5
2	5.5
4	8

ตอนนี้จะใช้ตารางข้างบนเพื่อจัดอันดับความแตกต่างในตารางหลัก

ความสูงก่อน	ความสูงหลัง	ความแตกต่าง(D)	อันดับ
58	62	-4	8
60	61	-1	2
65	67	-2	5.5
59	60	-1	2
61	63	-2	5.5
60	59	1	2
62	64	-2	5.5
63	65	-2	5.5

คำนวณผลรวมของอันดับบวกและลบ

1. **อันดับบวก (T^+)**
    - อันดับบวก = 2 (จากความแตกต่างที่มีค่าเป็นบวก)
2. **อันดับลบ (T^-)**
    - อันดับลบ = 8 + 2 + 5.5 + 2 + 5.5 + 5.5 + 5.5 = 34

ค่าต่ำสุดของอันดับรวม (W)

$$W = \min(T^+, T^-) = \min(2, 34) = 2$$

เปรียบเทียบกับค่าที่คาดหวัง

เปรียบเทียบค่า W ที่คำนวณได้กับค่า W ในตารางวิลคอกซอนที่ระดับนัยสำคัญที่กำหนด เช่น 0.05 เพื่อสรุปผลการทดสอบ

สรุปผลการทดสอบ

- **สมมติฐานศูนย์ (Null Hypothesis, H_0)**: ไม่มีความแตกต่างในค่ามัธยฐานระหว่างสองกลุ่มที่จับคู่กัน
- **สมมติฐานทางเลือก (Alternative Hypothesis, H_a)**: มีความแตกต่างในค่ามัธยฐานระหว่างสองกลุ่มที่จับคู่กัน

ถ้าค่า W ที่คำนวณได้มีค่าน้อยกว่าหรือเท่ากับค่าที่คาดหวังในตารางวิลคอกซอน จะปฏิเสธสมมติฐานศูนย์ H_0 และสรุปว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างสองกลุ่มที่จับคู่กัน

จากตัวอย่าง **ดูค่าที่คาดหวังจากตารางวิลคอกซอน**: สำหรับจำนวนตัวอย่าง n=8 (เพราะมี 8 คู่ข้อมูล) และระดับนัยสำคัญ 0.05

ในกรณีนี้ต้องดูค่าจากตารางวิลคอกซอนที่ n=8และระดับนัยสำคัญ 0.05

จากตารางวิลคอกซอน (สำหรับตัวอย่างขนาดเล็ก):

Critical Values of the Wilcoxon Signed Ranks Test

n	Two-Tailed Test		One-Tailed Test	
	α = .05	α = .01	α = .05	α = .01
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15

สำหรับ n=8 และ α=0.05 ค่า W ที่คาดหวังคือ 3

เปรียบเทียบค่า W ที่คำนวณได้กับค่าที่คาดหวัง

- ค่า W ที่คำนวณได้คือ 2
- ค่า W ที่คาดหวังจากตารางที่ α=0.05 คือ 3

การสรุปผลการทดสอบ

- ถ้าค่า W ที่คำนวณได้ ≤ ค่า W ที่คาดหวังจากตารางวิลคอกซอนที่ α=0.05
    - ในกรณีนี้ 2≤3 แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติ
    - ดังนั้นจะ **ปฏิเสธสมมติฐานศูนย์ (H_0)** และสรุปว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างสองกลุ่มที่จับคู่กัน
- ถ้าค่า W ที่คำนวณได้ > ค่า W ที่คาดหวังจากตารางวิลคอกซอนที่ α=0.05
    - จะไม่ปฏิเสธสมมติฐานศูนย์ (H_0) และสรุปว่าไม่มีความแตกต่างอย่างมีนัยสำคัญระหว่างสองกลุ่มที่จับคู่กัน

สรุปผล

ในตัวอย่างนี้ ค่า W ที่คำนวณได้ (2) น้อยกว่าค่า W ที่คาดหวังจากตาราง (3) ที่ระดับนัยสำคัญ 0.05 ดังนั้นจึง **ปฏิเสธสมมติฐานศูนย์ (H0)** และสรุปว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างสองกลุ่มที่จับคู่กัน

สรุป

สถิตินอนพาราเมตริกเป็นเครื่องมือสถิติที่ใช้ในการวิเคราะห์ข้อมูลที่ไม่ต้องพึ่งพาสมมติฐานเกี่ยวกับการกระจายตัวของข้อมูล เช่น การกระจายตัวแบบปกติ เหมาะสมสำหรับข้อมูลที่มีขนาดเล็ก, ข้อมูลที่ไม่ได้เป็นแบบจำลองเชิงเส้น หรือข้อมูลที่มีการกระจายตัวที่ไม่ทราบรูปแบบ การทดสอบนอนพาราเมตริกที่สำคัญได้แก่ การทดสอบ Mann-Whitney U Test ซึ่งใช้สำหรับเปรียบเทียบความแตกต่างระหว่างสองกลุ่มอิสระ และการทดสอบ Wilcoxon Signed-Rank Test สำหรับกลุ่มตัวอย่างเดียวกันที่ทดสอบสองครั้ง วิธีการเหล่านี้ช่วยให้สามารถวิเคราะห์ข้อมูลได้แม้ในกรณีที่ข้อมูลไม่เป็นไปตามสมมติฐานทางสถิติแบบพาราเมตริก ทำให้การทดสอบมีความยืดหยุ่นและใช้งานได้หลากหลายสถานการณ์มากขึ้น

ศึกษาประกอบ

https://www.youtube.com/watch?v=neLNnw1_Q8M&list=PL51ZCXHa8bMlNE368mMJraGGYQ3mBFNtq&index=4

https://youtu.be/neLNnw1_Q8M?list=PL51ZCXHa8bMlNE368mMJraGGYQ3mBFNtq&t=36

https://youtu.be/zealFsQxJzo?list=PL51ZCXHa8bMlNE368mMJraGGYQ3mBFNtq&t=9

https://youtu.be/neLNnw1_Q8M?list=PL51ZCXHa8bMlNE368mMJraGGYQ3mBFNtq&t=353

https://youtu.be/iMwhJi6klIM?t=326

แบบฝึกหัด

แบบฝึกหัดที่ 1: การทดสอบ Mann-Whitney U Test

ในบริษัทแห่งหนึ่งมีการสำรวจความพึงพอใจของพนักงานในแผนก A และแผนก B ต่อระบบการจัดการใหม่ โดยคะแนนความพึงพอใจมีค่าตั้งแต่ 1 ถึง 10 มีการเก็บข้อมูลดังนี้:

- แผนก A: 7, 8, 5, 6, 7, 8, 9
- แผนก B: 6, 5, 4, 5, 6, 5, 4

1.1 ใช้การทดสอบ Mann-Whitney U Test เพื่อตรวจสอบว่าความพึงพอใจของพนักงานในแผนก A และแผนก B มีความแตกต่างกันหรือไม่ (สมมติระดับนัยสำคัญที่ 0.05)

แบบฝึกหัดที่ 2: การทดสอบไคสแควร์ (Chi-Square Test)

ในงานวิจัยเรื่องการเลือกเครื่องดื่มในกลุ่มวัยรุ่น มีการเก็บข้อมูลจำนวนวัยรุ่นที่เลือกเครื่องดื่มแต่ละประเภทดังนี้:

- ชา: 30 คน
- กาแฟ: 25 คน
- น้ำผลไม้: 20 คน

- น้ำอัดลม: 15 คน

2.1 ใช้การทดสอบไคสแควร์ (Chi-Square Test) เพื่อตรวจสอบว่าวัยรุ่นมีการเลือกเครื่องดื่มแต่ละประเภทอย่างเท่าเทียมกันหรือไม่ (สมมติระดับนัยสำคัญที่ 0.05)

เอกสารอ้างอิง

ปรีดาภรณ์ กาญจนสำราญวงศ์.(2560). *หลักสถิติเบื้องต้น*.นนทบุรี. ไ อ ดี ซี พรีเมียร์

สำรวม จงเจริญ. (2563). *การวิเคราะห์เชิงสถิติแบบไม่ใช้พารามิเตอร์*. กรุงเทพฯ : โอ.เอส. พริ้นติ้ง เฮ้าส์

อรุณ จิรวัฒน์กุล. (2558). *ประสิทธิภาพอำนาจการทดสอบแบบนอนพารา เมตริก*. วารสารวิชาการสาธารณสุข. 24(5) .820-821

ปุญญพัฒน์, ไชยเมล์.(2555).สถิติและการวิเคราะห์ข้อมูลทางสุขภาพ.เข้าถึง 2 2 มิ ถุ น า ย น 2 5 6 7 จ า ก http://hsmi2.psu.ac.th/edu/upload/forum/lec_567730_lesson_09_nonparametric_statistics.pdf

เฉลยแบบฝึกหัดที่ 1:

1.1 ใช้ Mann-Whitney U Test คำนวณ U-statistic และเปรียบเทียบกับค่าจากตารางเพื่อดูว่ามีนัยสำคัญหรือไม่ สำหรับ n_a=7 และ n_b=7 ที่ระดับนัยสำคัญ α=0.05, ค่า critical value จากตาราง Mann-Whitney คือ 8

ถ้า U≤ critical value, ปฏิเสธสมมติฐานว่าไม่มีความแตกต่าง

ในที่นี้ U=7.5≤8 , ดังนั้นเราปฏิเสธสมมติฐานที่ไม่มีความแตกต่าง

เฉลยแบบฝึกหัดที่ 2:

2.1 ใช้สูตรการทดสอบไคสแควร์ดังนี้:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

โดยที่ Oi คือค่าจริงที่สังเกตได้ และ Ei คือค่าคาดหมาย ซึ่งคำนวณได้จากการสมมติว่าการเลือกเครื่องดื่มแต่ละประเภทมีโอกาสเท่ากัน

1. คำนวณค่าคาดหมาย (Expected Value) สำหรับแต่ละประเภทของเครื่องดื่ม:

$$E_i = \frac{\text{จำนวนวัยรุ่นทั้งหมด}}{\text{จำนวนประเภทของเครื่องดื่ม}} = \frac{30+25+20+15}{4} = 22.5$$

2. คำนวณค่าไคสแควร์ (Chi-Square):

$$\chi^2 = \frac{(30-22.5)^2}{22.5} + \frac{(25-22.5)^2}{22.5} + \frac{(20-22.5)^2}{22.5} + \frac{(15-22.5)^2}{22.5}$$

3. เปรียบเทียบค่า χ^2 ที่คำนวณได้กับค่า χ^2 จากตารางที่ระดับความนัยสำคัญ 0.05 โดยมีอิสระที่ (df) = 3 เพื่อดูว่ามีนัยสำคัญหรือไม่

บทที่ 10

การวิเคราะห์การถดถอยและสหสัมพันธ์

การวิเคราะห์การถดถอยและการทดสอบสหสัมพันธ์เป็นเครื่องมือทางสถิติที่สำคัญในการวิจัยและวิเคราะห์ข้อมูล โดยการทดสอบสหสัมพันธ์ใช้ในการวัดความสัมพันธ์ระหว่างตัวแปรสองตัวว่าเกี่ยวข้องกันอย่างไรและมีความสัมพันธ์ในระดับใด ส่วนการวิเคราะห์การถดถอยใช้ในการสร้างโมเดลเพื่อทำนายหรืออธิบายความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ ซึ่งทั้งสองเครื่องมือนี้มีบทบาทสำคัญในการทำความเข้าใจและตัดสินใจในงานวิจัยด้านต่างๆ เช่น เศรษฐศาสตร์ การตลาด และวิทยาศาสตร์สุขภาพ

การทดสอบสหสัมพันธ์

การทดสอบสหสัมพันธ์ (Correlation Testing) เป็นกระบวนการทางสถิติที่ใช้ในการตรวจสอบความสัมพันธ์ระหว่างตัวแปรสองตัวว่ามีความเกี่ยวข้องกันอย่างไรและในระดับใด ความสัมพันธ์นี้สามารถเป็นได้ทั้งเชิงบวก (positive correlation) และเชิงลบ (negative correlation) หรือไม่มีความสัมพันธ์เลย (no correlation) โดยการวัดความสัมพันธ์ระหว่างตัวแปรสามารถทำได้ด้วยการคำนวณสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient)

หนึ่งในเครื่องมือที่นิยมใช้ในการวัดความสัมพันธ์คือ สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation Coefficient) ซึ่งคำนวณค่าความสัมพันธ์ระหว่างตัวแปรสองตัวในลักษณะเชิงเส้น (linear relationship) โดยมีค่าตั้งแต่ -1 ถึง 1:

- ค่า 1 หมายถึงความสัมพันธ์เชิงบวกที่สมบูรณ์ (เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรจะเพิ่มขึ้นตาม)
- ค่า -1 หมายถึงความสัมพันธ์เชิงลบที่สมบูรณ์ (เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรจะลดลง)

- ค่า 0 หมายถึงไม่มีความสัมพันธ์ระหว่างตัวแปรทั้งสอง

นอกจากสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน ยังมีเครื่องมืออื่นๆ เช่น สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน (Spearman's Rank Correlation) ที่ใช้สำหรับข้อมูลที่ไม่เป็นเชิงเส้นหรือข้อมูลอันดับ (ordinal data) และสัมประสิทธิ์สหสัมพันธ์เคนดัลล์ (Kendall's Tau) ที่ใช้สำหรับข้อมูลที่มีลำดับ

การทดสอบสหสัมพันธ์มีความสำคัญในการวิจัยและการวิเคราะห์ข้อมูล เนื่องจากช่วยให้นักวิจัยสามารถระบุได้ว่าตัวแปรใดมีความสัมพันธ์กันและในระดับใด ซึ่งสามารถนำไปใช้ในการวางแผนการทดลอง การวิเคราะห์เพิ่มเติม และการสร้างโมเดลทางสถิติที่มีประสิทธิภาพมากขึ้น

สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation Coefficient) เป็นเครื่องมือทางสถิติที่ใช้วัดความสัมพันธ์เชิงเส้นระหว่างตัวแปรสองตัว โดยค่าสัมประสิทธิ์นี้แสดงถึงความแรงและทิศทางของความสัมพันธ์ ค่านี้มีค่าอยู่ในช่วงตั้งแต่ -1 ถึง 1 โดยมีความหมายดังนี้:

- ค่า 1 หมายถึงความสัมพันธ์เชิงบวกที่สมบูรณ์ (เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรก็เพิ่มขึ้นตามในอัตราที่สมบูรณ์)
- ค่า -1 หมายถึงความสัมพันธ์เชิงลบที่สมบูรณ์ (เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรจะลดลงในอัตราที่สมบูรณ์)
- ค่า 0 หมายถึงไม่มีความสัมพันธ์เชิงเส้นระหว่างตัวแปรทั้งสอง

การคำนวณสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันใช้สูตรดังนี้

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

โดยที่

- r คือค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

- Xi และ Yi คือค่าของตัวแปร X และ Y ตามลำดับ
- $\bar{X}$ และ $\bar{Y}$ คือค่าเฉลี่ยของตัวแปร X และ Y ตามลำดับ

การแปลความหมาย

r ใกล้ 1 แสดงว่ามีความสัมพันธ์เชิงบวกที่แข็งแรง

r ใกล้ -1 แสดงว่ามีความสัมพันธ์เชิงลบที่แข็งแรง

r ใกล้ 0 แสดงว่าไม่มีความสัมพันธ์เชิงเส้นระหว่างตัวแปรทั้งสอง

ข้อจำกัด

- ความสัมพันธ์ที่พบอาจไม่ได้หมายถึงความเป็นเหตุเป็นผล (causation)
- การทดสอบนี้เหมาะกับความสัมพันธ์เชิงเส้นเท่านั้น ถ้าความสัมพันธ์เป็นแบบไม่เชิงเส้น (non-linear) อาจต้องใช้การทดสอบอื่นๆ เช่น สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน (Spearman's Rank Correlation)

การใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สันเป็นวิธีที่มีประสิทธิภาพในการวิเคราะห์และทำความเข้าใจความสัมพันธ์ระหว่างตัวแปรในหลายสาขาวิชา ไม่ว่าจะเป็นการวิจัยทางวิทยาศาสตร์ เศรษฐศาสตร์ หรือสังคมศาสตร์

ตัวอย่างการคำนวณสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

สมมติว่ามีข้อมูลเกี่ยวกับตัวแปรสองตัว คือ X และ Y ดังนี้:

X	Y
1	2
2	3
3	4
4	5
5	6

คำนวณสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (r) สำหรับข้อมูลชุดนี้

ขั้นตอนการคำนวณ

1. หาค่าเฉลี่ย ของ X และ Y:

$$\bar{\mathbf{X}} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$

$$\bar{\mathrm{Y}} = \frac{2+3+4+5+6}{5} = \frac{20}{5} = 4$$

2. หาผลต่าง ระหว่างค่าของแต่ละตัวแปรกับค่าเฉลี่ย

X	Y	$Xi - \bar{\mathbf{X}}$	$Yi - \bar{\mathrm{Y}}$
1	2	1 - 3 = -2	2 - 4 = -2
2	3	2 - 3 = -1	3 - 4 = -1
3	4	3 - 3 = 0	4 - 4 = 0
4	5	4 - 3 = 1	5 - 4 = 1
5	6	5 - 3 = 2	6 - 4 = 2

3. คูณผลต่าง ของตัวแปร X และ Y สำหรับแต่ละคู่ข้อมูล:

X	Y	$Xi - \bar{\mathbf{X}}$	$Yi - \bar{\mathrm{Y}}$	$(Xi - \bar{\mathbf{X}})(Yi - \bar{\mathrm{Y}})$
1	2	-2	-2	4
2	3	-1	-1	1
3	4	0	0	0
4	5	1	1	1
5	6	2	2	4

4. หา $\sum(Xi - \bar{\mathbf{X}})(Yi - \bar{\mathrm{Y}}) = 4+1+0+1+4=10$

5. หาผลรวม ของกำลังสองของผลต่างของแต่ละตัวแปร

$\sum(Xi - \bar{\mathbf{X}})^2 = (-2)^2+(-1)^2+0^2+1^2+2^2 = 4+1+0+1+4=10$

$\sum(Yi - \bar{\mathrm{Y}})^2 = (-2)^2+(-1)^2+0^2+1^2+2^2 = 4+1+0+1+4=10$

6. **นำผลรวม** ที่ได้มาคูณกันและหาค่ารากที่สอง:
   $\sqrt{10 \times 10}$= $\sqrt{100}$ =10
7. นำของผลคูณจากข้อ 4 หารด้วยค่าที่ได้จากข้อ 6:
   r= $\frac{\mathbf{10}}{\mathbf{10}}$ = 1

ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (r) สำหรับข้อมูลชุดนี้คือ 1 ซึ่งหมายถึงตัวแปร X และ Y มีความสัมพันธ์เชิงบวกที่สมบูรณ์ ค่าของ X และ Y เพิ่มขึ้นตามกันในอัตราที่สมบูรณ์

สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน

สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน (Spearman's Rank Correlation Coefficient) เป็นเครื่องมือทางสถิติที่ใช้วัดความสัมพันธ์ระหว่างตัวแปรสองตัวโดยอิงตามลำดับ (rank) ของข้อมูลแทนค่าตัวเลขจริง จึงสามารถใช้วัดความสัมพันธ์ที่ไม่เชิงเส้น (non-linear relationships) ได้และเหมาะสำหรับข้อมูลที่มีลำดับหรือข้อมูลที่ไม่ปกติ (non-parametric data)

สูตรคำนวณ

สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน (ρ หรือ r_s) สามารถคำนวณได้จากสูตร

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

โดยที่

- d_i คือความแตกต่างระหว่างลำดับของคู่ข้อมูลที่สอดคล้องกัน (Rank difference)

- n คือจำนวนคู่ข้อมูล

ขั้นตอนการคำนวณ

1. จัดลำดับ ค่าของตัวแปร X และ Y
2. หาค่าความแตกต่างของลำดับ (d_i) สำหรับแต่ละคู่ข้อมูล
3. หาค่ากำลังสองของความแตกต่างของลำดับ (d_i^2)
4. นำค่าที่ได้จากข้อ 3 มาหาผลรวม
5. คำนวณ โดยใช้สูตร

สมมติว่ามีข้อมูลเกี่ยวกับตัวแปรสองตัวคือ X และ Y ดังนี้

X	Y
1	10
2	8
3	6
4	4
5	2

1. จัดลำดับ ค่าของ X และ Y

X	Rank X	Y	Rank Y
1	1	10	5
2	2	8	4
3	3	6	3
4	4	4	2
5	5	2	1

2. หาค่าความแตกต่างของลำดับ (di) และค่ากำลังสองของความแตกต่างของลำดับ (di^2)

X	Rank X	Y	Rank Y	di	di^2
1	1	10	5	1 - 5 = -4	16
2	2	8	4	2 - 4 = -2	4
3	3	6	3	3 - 3 = 0	0
4	4	4	2	4 - 2 = 2	4
5	5	2	1	5 - 1 = 4	16

3. หาผลรวม ของ di^2:

$\sum di^2 = 16+4+0+4+16 = 40$

4. คำนวณ โดยใช้สูตร:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6\times 40}{5(5^2-1)} = 1 - \frac{240}{5\times 24} = 1 - \frac{240}{120} = 1 - 2 = -1$$

สรุป คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน (ρ หรือ r_s) สำหรับข้อมูลชุดนี้คือ -1 ซึ่งหมายถึงตัวแปร X และ Y มีความสัมพันธ์เชิงลบที่สมบูรณ์ เมื่อค่าของ X เพิ่มขึ้น ค่าของ Y จะลดลงในอัตราที่สมบูรณ์

การใช้สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนเป็นวิธีที่มีประสิทธิภาพในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร โดยเฉพาะเมื่อข้อมูลไม่ได้เป็นเชิงเส้นหรือไม่เป็นแบบปกติ

การวิเคราะห์การถดถอยและการใช้โมเดลทางสถิติ

การวิเคราะห์การถดถอย (Regression Analysis) เป็นกระบวนการทางสถิติที่ใช้ในการประมาณหรือทำนายความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Variable) กับตัวแปรอิสระ (Independent Variables) หลายตัว เครื่องมือการถดถอยที่ใช้กันบ่อยได้แก่

- การถดถอยเชิงเส้น (Linear Regression)

- การถดถอยโลจิสติก (Logistic Regression)
- การถดถอยพหุคูณ (Multiple Regression)

โมเดลการถดถอยเชิงเส้นมีสมการพื้นฐานคือ $Y=a+bX+\epsilon$ โดยที่

Y คือค่าของตัวแปรตาม

X คือค่าของตัวแปรอิสระ

a คือค่าคงที่ (Intercept)

b คือสัมประสิทธิ์การถดถอย (Slope)

ϵ คือความคลาดเคลื่อน (Error Term)

การวิเคราะห์การถดถอยช่วยในการทำนายค่าอนาคตของตัวแปรตามเมื่อรู้ค่าของตัวแปรอิสระ และสามารถใช้ในการทดสอบสมมติฐานเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร นอกจากนี้ยังช่วยในการทำความเข้าใจถึงความแรงและทิศทางของความสัมพันธ์ และสามารถนำไปใช้ในการสร้างโมเดลทางสถิติที่ซับซ้อนขึ้นได้

การวิเคราะห์การถดถอย (Regression Analysis) สำหรับ 2 ตัวแปร

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีทางสถิติที่ใช้ในการตรวจสอบและสร้างแบบจำลองความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Variable) และตัวแปรอิสระ (Independent Variable) สำหรับกรณีของการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) จะมีตัวแปรอิสระเพียงตัวเดียว และมีสมการของเส้นตรงที่ใช้ทำนายค่าของตัวแปรตาม (สุพล ดุรงค์วัฒนา, 2558)

สมการการถดถอยเชิงเส้นอย่างง่าย

สมการการถดถอยเชิงเส้นอย่างง่ายมีรูปแบบดังนี้

$$Y=a+bX$$

โดยที่

- Y คือค่าที่ทำนายของตัวแปรตาม

- a คือค่าคงที่ (Intercept) ซึ่งเป็นค่าของ Y Y เมื่อ X X เท่ากับ 0
- b คือสัมประสิทธิ์การถดถอย (Slope) ซึ่งแสดงถึงการเปลี่ยนแปลงของ Y เมื่อ X เพิ่มขึ้น 1 หน่วย
- X คือค่าของตัวแปรอิสระ

ขั้นตอนการคำนวณ

1. หาค่าเฉลี่ย ของตัวแปร X และ Y:

$$\bar{X} = \frac{\sum Xi}{n} \quad , \quad \bar{Y} = \frac{\sum Yi}{n}$$

2. หาค่าสัมประสิทธิ์การถดถอย (b)

$$b = \frac{\sum(Xi-\bar{x})(Yi-\bar{Y})}{\sum(Xi-\bar{x})^2}$$

3. หาค่าคงที่ (a)

$$a = \bar{Y} - b\,\bar{X}$$

4. สร้างสมการการถดถอย

$$Y = a + bX$$

สมมติว่ามีข้อมูลเกี่ยวกับตัวแปร X และ Y ดังนี้

X	Y
1	2
2	3
3	5
4	4
5	6

1. หาค่าเฉลี่ย ของตัวแปร X และ Y:

$\bar{X}$ = (1+2+3+4+5)/5 = 15/5 = 3

$\bar{Y}$ = (2+3+5+4+6)/5 = 20/5 = 4

2. หาค่าสัมประสิทธิ์การถดถอย (b)

$$b = \frac{(1-3)(2-4)+(2-3)(3-4)+(3-3)(5-4)+(4-3)(4-4)+(5-3)(6-4)}{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}$$

$$b = \frac{4+1+0+0+4}{4+1+0+1+4}$$

b= 9/10 =0.9

3. หาค่าคงที่ (a)

$a = \bar{Y} - b\,\bar{X}$

a= 4–(0.9×3)

a= 4–2.7

a= 1.3

4. สร้างสมการการถดถอย

Y=1.3+0.9X

สมการการถดถอยเชิงเส้นอย่างง่ายที่ได้คือ Y=1.3+0.9X ซึ่งหมายความว่าเมื่อค่า X เพิ่มขึ้น 1 หน่วย ค่า Y จะเพิ่มขึ้น 0.9 หน่วย นอกจากนี้ค่าคงที่ 1.3 หมายถึงค่า Y เมื่อ X เท่ากับ

สรุป

การวิเคราะห์การถดถอยและสหสัมพันธ์เป็นเครื่องมือทางสถิติที่สำคัญในการวิจัยและวิเคราะห์ข้อมูล โดยการทดสอบสหสัมพันธ์ใช้ในการวัดและแสดงระดับความสัมพันธ์ระหว่างตัวแปรสองตัวว่ามีทิศทางและความแรงของความสัมพันธ์อย่างไร เช่น สัมประสิทธิ์สหสัมพันธ์ของเพียร์สันและสเปียร์แมน ส่วนการวิเคราะห์การถดถอยใช้ในการสร้างแบบจำลองเพื่อทำนายค่าของตัวแปรตามจากค่าของตัวแปรอิสระ เช่น การถดถอยเชิงเส้นอย่างง่ายที่ช่วยแสดงความสัมพันธ์เชิงเส้นระหว่างตัวแปร ซึ่งทั้งสองวิธีมีบทบาทสำคัญในการทำความเข้าใจพฤติกรรมของข้อมูลและสนับสนุนการตัดสินใจในหลากหลายสาขาวิชา

แบบฝึกหัดการวิเคราะห์การถดถอยและสหสัมพันธ์

แบบฝึกหัดที่ 1: การทดสอบสหสัมพันธ์ของเพียร์สัน

ชุดข้อมูล

X	Y
1	2
2	4
3	6
4	8
5	10

1. คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (r) ระหว่างตัวแปร X และ Y
2. อธิบายความหมายของค่าสัมประสิทธิ์ที่ได้

แบบฝึกหัดที่ 2: การทดสอบสหสัมพันธ์ของสเปียร์แมน

ชุดข้อมูล

A	B
10	1
20	2
30	3
40	4
50	5

1. จัดลำดับค่าของตัวแปร A และ B
2. คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน (ρ) ระหว่างตัวแปร A และ B
3. อธิบายความหมายของค่าสัมประสิทธิ์ที่ได้

แบบฝึกหัดที่ 3: การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

ชุดข้อมูล

X	Y
2	3
3	5
5	8
7	11
9	14

1. คำนวณค่าสัมประสิทธิ์การถดถอย (b) และค่าคงที่ (a) สำหรับสมการการถดถอยเชิงเส้นอย่างง่าย
2. เขียนสมการการถดถอยเชิงเส้นอย่างง่ายจากค่าที่คำนวณได้
3. ใช้สมการที่ได้ทำนายค่า Y เมื่อ X เท่ากับ 6

เอกสารอ้างอิง

สุพล ดุรงค์วัฒนา. (2558). Regression Models : Analytical-based Approach. กรุงเทพฯ. แดเน็กซ์อินเตอร์คอร์ปอเรชั่น

เฉลยสำหรับแบบฝึกหัด

แบบฝึกหัดที่ 1: การทดสอบสหสัมพันธ์ของเพียร์สัน

1. ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (r) คือ 1
2. ความหมาย: ตัวแปร X และ Y มีความสัมพันธ์เชิงบวกที่สมบูรณ์แบบ เมื่อค่าของ X เพิ่มขึ้น ค่า Y ก็เพิ่มขึ้นในอัตราที่สมบูรณ์

แบบฝึกหัดที่ 2: การทดสอบสหสัมพันธ์ของสเปียร์แมน

1. (ลำดับของ A และ B จะเหมือนกัน เนื่องจากข้อมูลเป็นลำดับที่เพิ่มขึ้นตรงกัน)
2. ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน (ρ) คือ 1
3. ความหมาย: ตัวแปร A และ B มีความสัมพันธ์เชิงบวกที่สมบูรณ์แบบ เมื่อค่าของ A เพิ่มขึ้น ค่า B ก็เพิ่มขึ้นในอัตราที่สมบูรณ์

แบบฝึกหัดที่ 3: การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

1. ค่าสัมประสิทธิ์การถดถอย (b) คือ 1.5 และค่าคงที่ (a) คือ 0.5
2. สมการการถดถอยเชิงเส้นอย่างง่าย: Y=0.5+1.5X Y=0.5+1.5X
3. ค่า Y เมื่อ X เท่ากับ 6 คือ 9.5

บทที่ 11
การพยากรณ์และการตัดสินใจทางธุรกิจ

การใช้สถิติในการพยากรณ์และการตัดสินใจทางธุรกิจเป็นกระบวนการสำคัญที่ช่วยให้ธุรกิจสามารถวางแผนและตัดสินใจในการดำเนินกิจการได้อย่างมั่นคงและมีประสิทธิภาพ เมื่อใช้สถิติในการพยากรณ์ เช่น การวิเคราะห์แนวโน้มของตลาดหรือการพยากรณ์ยอดขาย เราสามารถใช้ข้อมูลที่มีอยู่เพื่อทำนายผลลัพธ์ในอนาคตได้ ส่วนการใช้ข้อมูลสถิติในการตัดสินใจทางธุรกิจช่วยให้เราทราบถึงข้อมูลที่สำคัญและการเชื่อมโยงระหว่างตัวแปรต่าง ๆ ที่ส่งผลต่อกิจการ เช่น การวิเคราะห์ข้อมูลลูกค้าเพื่อปรับปรุงผลิตภัณฑ์หรือบริการ เราสามารถตัดสินใจในการวางกลยุทธ์ใหม่โดยอิงตามข้อมูลที่ได้รับจากการวิเคราะห์ดังกล่าว การใช้สถิติในทั้งการพยากรณ์และการตัดสินใจนี้เป็นเครื่องมือที่มีประสิทธิภาพในการสนับสนุนการเติบโตและความสำเร็จของธุรกิจในระยะยาว

สถิติการพยากรณ์

สถิติการพยากรณ์เป็นกระบวนการที่ใช้ข้อมูลที่มีอยู่ในอดีตเพื่อทำนายผลลัพธ์หรือเหตุการณ์ในอนาคต มันช่วยให้เราสามารถวิเคราะห์แนวโน้ม และทำนายสิ่งต่าง ๆ ได้ เช่น ยอดขายสินค้าในอนาคต อัตราการเติบโตของตลาด หรือแม้กระทั่งสภาวะอากาศในวันพรุ่งนี้ การใช้สถิติในการพยากรณ์มักจะใช้เครื่องมือและเทคนิคต่าง ๆ เช่น การวิเคราะห์แนวโน้ม (trend analysis), การสร้างโมเดล (modeling), และการใช้สูตรทางสถิติ เช่น การใช้เทคนิคเชิงเส้น (linear regression) หรือเทคนิคที่ซับซ้อนขึ้นเช่น การใช้โมเดลทางการเรียนรู้ของเครื่อง (machine learning models) สำหรับประเภทข้อมูลที่ซับซ้อนมากขึ้น เช่น ข้อมูลที่มีการเปลี่ยนแปลงตลอดเวลา การพยากรณ์ดังกล่าวมีประโยชน์ในหลายด้านของชีวิต เช่น ในธุรกิจ (การ

วางแผนการผลิตและการขาย) การเงิน (การวางแผนการลงทุน) และการเมือง (การทำนายผลการเลือกตั้ง) (วุฒิ สุขเจริญ,2561)

การพยากรณ์สามารถแบ่งเป็นหลายประเภทตามลักษณะของข้อมูลและวิธีการทำนาย นี่คือบางประเภทและวิธีการพยากรณ์ที่ใช้กันบ่อย:

1. **การพยากรณ์เชิงปริมาณ (Quantitative Forecasting)**: วิธีการนี้ใช้ข้อมูลเชิงปริมาณที่เก็บรวบรวมมาอย่างต่อเนื่องเพื่อทำนายผลลัพธ์ โดยการวิเคราะห์แนวโน้มของข้อมูลตามเวลา หรือตามตัวแปรอื่น ๆ ที่เกี่ยวข้อง เช่น การใช้เทคนิคการเชิงเส้นในการทำนายยอดขายสินค้าในอนาคตโดยใช้ข้อมูลยอดขายในช่วงเวลาที่ผ่านมาเป็นพื้นฐาน แบ่งเป็น 3 กลุ่ม คือการวิเคราะห์อนุกรมเวลา การศึกษาความสัมพันธ์ระหว่างตัวแปร และการตรวจสอบการดำเนินงานมีหลักกว้างๆดังนี้

1.1 วิธีง่าย (Naive Method): ค่าพยากรณ์ในอนาคตจะมีค่าเป็นสัดส่วนของค่าสังเกตล่าสุด ซึ่งสัดส่วนอย่างไรนั้นผู้พยากรณ์จะเป็นผู้กำหนดขึ้น

1.2 วิธีแยกส่วนประกอบ (Decomposition Method) ค่าพยากรณ์ในอนาคตจะได้จากการรวมค่าวัดส่วนประกอบของอนุกรมเวลา ได้แก่ค่าแนวโน้ม ค่าวัดอิทธิพลของฤดูกาล ค่าวัดวัฏจักร ค่าวัดเหตุการณ์ที่ผิดปกติ ค่าวัดจะได้จากวิธีการเฉลี่ยแบบธรรมดา (simple average) แบบเคลื่อนที่ (moving average) การพยากรณ์แบบนี้จะมีการเก็บข้อมูลความต้องการสินค้า หรือยอดขายระยะเวลาหนึ่งสม่ำเสมอโดยมีช่วงห่างใน การเก็บข้อมูลที่เท่า ๆ กัน โดยทั่วไป ความต้องการสินค้าจะมากหรือน้อยนั้นเกิดจากอิทธิพล 4 ประการ ได้แก่ แนวโน้ม (Trend) ฤดูกาล (Seasonal) วัฏจักร (Cycle) และความผิดปกติหรือความไม่แน่นอน

1.3 วิธีปรับให้เรียบ (Smoothing Method) ค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากค่าสังเกตในอดีตโดยการให้น้ำหนัก (weight) กับค่าสังเกตแบบต่างๆ หากให้น้ำหนักกับค่าสังเกตเท่าๆกันจะเรียกว่า วิธีเฉลี่ยเคลื่อนที่ (moving average method)

2. **การศึกษาความสัมพันธ์ระหว่างปัจจัย**หรือตัวแปรต่างๆที่เป็นเหตุและผลเนื่องจากเหตุการณ์ต่างๆ เช่น การวิเคราะห์การถดถอยอย่างง่ายคือความผันแปรของตัวแปรหนึ่งจะขึ้นกับความผันแปรของอีกตัวแปรหนึ่ง

การใช้สถิติในการพยากรณ์และการวางแผนทางธุรกิจ

ในการพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting) โดยการเก็บข้อมูลความต้องการสินค้าหรือยอดขายระยะเวลาหนึ่งๆ สม่ำเสมอโดยมีช่วงห่างในการเก็บข้อมูลที่เท่า ๆ กัน มักจะพิจารณาความต้องการดังกล่าวในแง่ของแนวโน้ม (Trend), ฤดูกาล (Seasonal), วัฏจักร (Cycle), และความผิดปกติหรือความไม่แน่นอน (กรินทร์ กาญทนานนท์, 2561)โดยมีลักษณะดังนี้

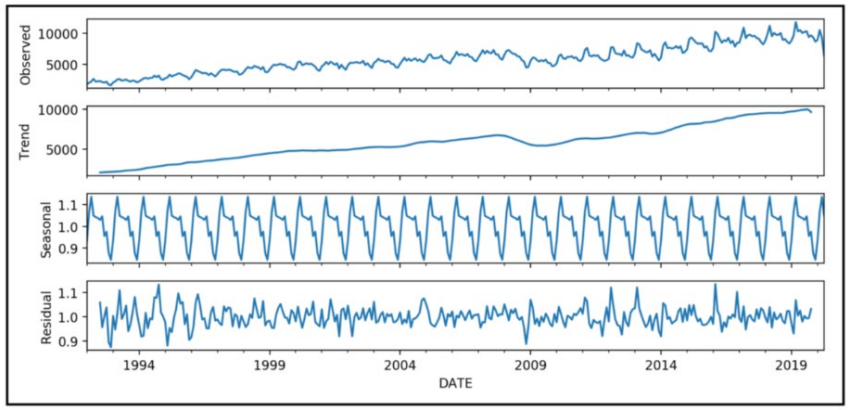
1. **แนวโน้ม (Trend)**: การเปลี่ยนแปลงของความต้องการหรือยอดขายที่มีลักษณะเรียงตามแนวโน้มที่ชัดเจน เช่น เพิ่มขึ้นหรือลดลงตามเวลา การตรวจสอบและพยากรณ์แนวโน้มสามารถช่วยให้เราระบุแนวโน้มการเปลี่ยนแปลงในอนาคต เพื่อปรับการผลิตหรือการส่งเสริมการขายให้เหมาะสม

2. **ฤดูกาล (Seasonal)**: การเปลี่ยนแปลงของความต้องการหรือยอดขายที่มีลักษณะเกิดซ้ำซ้อนในระยะเวลาที่กำหนด เช่น การเพิ่มขึ้นของยอดขายในช่วงเทศกาลหรือเทศกาลวันหยุด เราสามารถใช้การแยกส่วนฤดูกาลออกจากข้อมูลเพื่อทำนายผลลัพธ์ที่ถูกต้องในอนาคต

3. **วัฏจักร (Cycle)**: การเปลี่ยนแปลงของความต้องการหรือยอดขายที่มีลักษณะซ้ำซ้อนแต่ไม่สม่ำเสมอในระยะเวลายาวๆ เช่น การเพิ่มขึ้นของยอดขายในรอบของเศรษฐกิจ การรับรู้และการปรับปรุงวัฏจักรสามารถช่วยให้เราตระหนักถึงภาวะทางเศรษฐกิจและวางแผนการดำเนินธุรกิจในอนาคตได้

4. **ความผิดปกติหรือความไม่แน่นอน (Irregular component)**: ความผิดปกติหรือความไม่แน่นอนอาจเกิดขึ้นจากปัจจัยที่ไม่คาดคิด เช่น สภาวะภูมิอากาศ, ภาวะเศรษฐกิจโลก, หรือเหตุการณ์ไม่คาดคิดอื่นๆ การวิเคราะห์และการจัดการกับความไม่แน่นอนอาจช่วยให้เราสามารถปรับแนวทางการวางแผนและการบริหารจัดการในสถานการณ์ที่ซับซ้อนและเปลี่ยนแปลงได้เร็วขึ้น

อนุกรมเวลาหมายถึงกลุ่มของค่าสังเกตที่เก็บรวบรวมมาตามเวลาอย่างต่อเนื่อง(ภูมิฐาน รังคกุลนุวัฒน์,2556) ช่วงเวลาจะเท่ากันหรือไม่เท่ากันก็ได้ จะใช้สัญลักษณ์ y_t แทนอนุกรมเวลา $y_1, y_2, \ldots y_n$ ที่เก็บมาใน n ช่วงเวลา วิธีการพยากรณ์จะใช้ช่วงเวลาห่างกันเท่ากัน เช่น ปี ครึ่งปี ไตรมาส เดือน เป็นต้น ตัวอย่างของอนุกรมเวลาคือ ยอดขายสินค้ารายปีของบริษัทมอลลี่ จำนวนนักท่องเที่ยวรายปีที่เดินทางเข้ามาประเทศไทย หากต้องการศึกษาอิทธิพลของฤดูกาลต้องเก็บข้อมูลเป็นรายไตรมาสหรือรายเดือน



แหล่งที่มา : Ivan Despot,2024

ส่วนประกอบของอนุกรมเวลา

ส่วนประกอบของอนุกรมเวลาประเภทการแยกส่วนแบบบวก (Additive decomposition) ข้อมูลชุดเวลา y_t ถูกพิจารณาเป็นผลรวมของส่วนประกอบต่าง ๆ ดังนี้: แนวโน้ม (T_t), วัฏจักร (C_t), ฤดูกาล (S_t), และส่วนข้อผิดพลาดหรือข้อผิดปกติ (I_t) โดยมีสมการเป็นดังนี้:

$$y_t = T_t + C_t + S_t + I_t$$

โดยที่:

- T_t แทนส่วนแนวโน้ม (Trend) ซึ่งบ่งบอกถึงทิศทางหรือแนวโน้มของชุดเวลาในระยะยาว
- C_t แทนส่วนวัฏจักร (Cycle) ซึ่งบังคับปริมาณที่มีการเปลี่ยนแปลงเป็นระยะหรือลำดับเวลาที่ไม่ได้เกี่ยวข้องกับฤดูกาล
- S_t แทนส่วนฤดูกาล (Seasonal) ซึ่งแสดงแนวโน้มที่เป็นรูปแบบแบบซ้ำซ้อนในช่วงเวลาที่เฉพาะเจาะจง เช่น รายวัน รายสัปดาห์ หรือรายปี
- I_t แทนส่วนข้อผิดพลาดหรือข้อผิดปกติ (Irregular or error term) ซึ่งบังคับปริมาณที่มีการเปลี่ยนแปลงสุ่มหรือเสียงในข้อมูลที่ไม่สามารถอธิบายได้โดยส่วนอื่น

โดยการแยกส่วนเวลาชุดข้อมูลเป็นส่วนประกอบนี้ จะทำให้เราสามารถวิเคราะห์และเข้าใจรูปแบบและโครงสร้างข้อมูลได้อย่างชัดเจนขึ้น และสามารถทำนายได้ด้วยความแม่นยำมากขึ้นโดยการจำแนกแต่ละส่วนแยกกันในการประมาณค่า

t	Y_t =	รูปแบบบวก			
		T_t +	S_t +	C_t +	I_t
1	6.750	5.000	1.5	.000	.25
2	5.750	5.125	0.0	.125	.50
3	3.500	5.250	-1.5	.250	-.50
4	5.375	5.375	0.0	.250	-.25
5	7.625	5.500	1.5	.125	.50
6	5.375	5.625	0.0	.000	-.25
7	3.375	5.750	-1.5	-.125	-.75
8	5.875	5.875	1.5	-.250	.25
9	6.625	6.000	0.0	-.375	-.50
10	5.875	6.125	-1.5	-.500	.25
11	3.500	6.250	1.5	-.500	-.75

วิธีง่าย

การพยากรณ์วิธีง่ายหรือ Naive Method เป็นวิธีการพยากรณ์ที่ง่ายและเรียบง่ายที่สุด โดยใช้ค่าข้อมูลล่าสุดเป็นค่าพยากรณ์สำหรับช่วงเวลาถัดไป ซึ่งไม่คำนึงถึงแนวโน้มหรือลักษณะของข้อมูลในอดีต วิธีการนี้มักถูกใช้ในสถานการณ์ที่ข้อมูลมีความไม่เสม่นคลาดหรือไม่มีลักษณะที่สามารถนำมาใช้ในการคาดการณ์ได้ เช่น ในการทำนายยอดขายของสินค้าที่มีความเสถียรและไม่มีการเปลี่ยนแปลงมากในช่วงเวลาที่สนใจ

วิธีการคำนวณของ Naive Method คือการใช้ค่าข้อมูลล่าสุด y_{t-1}เป็นค่าพยากรณ์สำหรับช่วงเวลาถัดไป y_t ดังนี้:

พยากรณ์ y_t=y_{t-1}

โดยที่:

- y_t คือ ค่าข้อมูลในเวลา t
- y_{t-1} คือ ค่าข้อมูลในเวลา $t-1$

ในแบบนี้เราเพียงแค่ใช้ค่าข้อมูลล่าสุดเป็นพยากรณ์สำหรับช่วงเวลาถัดไป โดยไม่คำนึงถึงแนวโน้มหรือลักษณะของข้อมูลในอดีต เป็นวิธีที่ง่ายและไม่ซับซ้อน แต่อาจจะไม่เหมาะสมในสถานการณ์ที่ข้อมูลมีความซับซ้อนหรือมีการเปลี่ยนแปลงอย่างมากในช่วงเวลาที่สนใจ

ตัวอย่างการคำนวณ Naive Method ด้วยชุดข้อมูลที่มีค่าเป็น [10, 15, 20, 25, 30] และต้องการคำนวณพยากรณ์สำหรับช่วงเวลาถัดไป:

1. ให้ y_{t-1}=30 เนื่องจากข้อมูลล่าสุดคือ 30
2. คำนวณค่าพยากรณ์ y_t : =y_{t-1} = 30

ดังนั้น ค่าพยากรณ์สำหรับช่วงเวลาถัดไป y_t จะเป็น 30 โดยใช้วิธีการ Naive Method ซึ่งเป็นการใช้ค่าข้อมูลล่าสุดเป็นค่าพยากรณ์ต่อไปทุกครั้ง

วิธีเฉลี่ยเคลื่อนที่

วิธีเฉลี่ยเคลื่อนที่ (Moving Average Method) เป็นวิธีการทำนายหรือประมาณค่าในอนาคตโดยใช้ค่าเฉลี่ยของข้อมูลในช่วงเวลาหนึ่ง ๆ เพื่อคำนวณค่าพยากรณ์สำหรับช่วงเวลาถัดไป โดยมักนำมาใช้เมื่อข้อมูลมีความแปรปรวนต่ำหรือเปลี่ยนแปลงไม่มาก

วิธีการคำนวณของ Moving Average Method คือการใช้ค่าเฉลี่ยของข้อมูลในช่วงเวลาที่กำหนดมาเป็นค่าพยากรณ์สำหรับช่วงเวลาถัดไป โดยมีขั้นตอนดังนี้:

1. เลือกช่วงเวลาที่จะใช้ในการคำนวณค่าเฉลี่ย เช่น n ช่วงเวลา
2. คำนวณค่าเฉลี่ยของข้อมูลในช่วงเวลานั้น ๆ และใช้เป็นค่าพยากรณ์สำหรับช่วงเวลาถัดไป

สำหรับตัวอย่างการคำนวณ Moving Average Method ด้วยชุดข้อมูลที่มีค่าเป็น [10, 15, 20, 25, 30] และต้องการคำนวณพยากรณ์สำหรับช่วงเวลาถัดไปโดยใช้ช่วงเวลา n=3 จะเป็นดังนี้:

1. คำนวณ Moving Average สำหรับช่วงเวลา n=3
    - สำหรับช่วงเวลาแรก [10, 15, 20] ค่าเฉลี่ย MA1= $\frac{10+15+20}{3}$= $\frac{45}{3}$=15
    - สำหรับช่วงเวลาถัดไป [15, 20, 25] ค่าเฉลี่ย MA2= $\frac{15+20+25}{3}$= $\frac{60}{3}$=20
    - สำหรับช่วงเวลาถัดไป [20, 25, 30] ค่าเฉลี่ย MA3= $\frac{20+25+30}{3}$= $\frac{75}{3}$=25
2. ใช้ค่า Moving Average เป็นค่าพยากรณ์สำหรับช่วงเวลาถัดไป:
    - พยากรณ์สำหรับช่วงเวลาถัดไป MA4 =25

ดังนั้น ค่าพยากรณ์สำหรับช่วงเวลาถัดไปจะเป็น 25 โดยใช้ Moving Average Method ซึ่งเป็นการใช้ค่าเฉลี่ยของข้อมูลในช่วงเวลาที่กำหนดเป็นค่าพยากรณ์ต่อไป

การตัดสินใจทางธุรกิจ

การตัดสินใจทางธุรกิจ (Business Decision Making) เป็นกระบวนการที่สำคัญสำหรับการดำเนินงานขององค์กรหรือบริษัท การตัดสินใจที่ดีสามารถนำไปสู่ความสำเร็จและการเติบโต ในขณะที่การตัดสินใจที่ไม่ดีอาจส่งผลกระทบเชิงลบ ดังนั้น การทำความเข้าใจและมีทักษะในการตัดสินใจจึงมีความสำคัญอย่างยิ่ง (รัชนี ภูวพัฒนะพันธุ์และมาริสา แก้วสุวรรณ , 2562)

ขั้นตอนในการตัดสินใจทางธุรกิจ

1. **การระบุปัญหา (Problem Identification)**: ขั้นแรกคือการระบุว่าปัญหาหรือโอกาสที่ต้องการตัดสินใจคืออะไร ต้องการเป้าหมายอะไร

2. **การเก็บรวบรวมข้อมูล (Information Gathering)**: รวบรวมข้อมูลที่เกี่ยวข้องกับปัญหาหรือโอกาสนั้น ๆ เพื่อให้มีข้อมูลเพียงพอในการวิเคราะห์

3. **การวิเคราะห์หาผลลัพธ์แต่ละทางเลือก (Output Analysis)**: พิจารณาผลลัพธ์ของแต่ละทางเลือก

4. **การวิเคราะห์ทางเลือก (Alternative Analysis):** วิเคราะห์ทางเลือกต่าง ๆ ที่เป็นไปได้โดยพิจารณาข้อดีข้อเสียของแต่ละทางเลือกจากผลลัพธ์ของแต่ละทางเลือก

5. **การเลือกทางเลือกที่ดีที่สุด (Choosing the Best Alternative)**: เลือกทางเลือกที่คิดว่าเหมาะสมที่สุดตามการวิเคราะห์เชิงคณิตศาสตร์อย่างใดอย่างหนึ่ง

6. **การนำมาใช้ (Implementation)**: ดำเนินการตามทางเลือกที่เลือกไว้และทำการตัดสินใจ

ประเภทของการตัดสินใจ

1. การตัดสินใจภายใต้ความแน่นอน (Decision Making Under Certainty)
2.การตัดสินใจภายใต้ความไม่แน่นอน (Decision Making Under Uncertainty)
3. การตัดสินใจภายใต้ความเสี่ยง (Decision Making Under Risk)

การตัดสินใจภายใต้ความแน่นอน

การตัดสินใจภายใต้ความแน่นอน (Decision making under certainty) เป็นกระบวนการตัดสินใจที่ผู้ตัดสินใจทราบข้อมูลทั้งหมดเกี่ยวกับทางเลือกและผลลัพธ์ของแต่ละทางเลือกอย่างชัดเจน ไม่มีความไม่แน่นอนหรือความเสี่ยงที่เกี่ยวข้อง

ตัวอย่างของการตัดสินใจภายใต้ความแน่นอน

ตัวอย่าง 1: การเลือกซื้อสินค้าในร้านค้าปลีก

สมมติว่าต้องการซื้อสมาร์ทโฟนรุ่นใหม่ มีข้อมูลทั้งหมดเกี่ยวกับราคาและคุณสมบัติของแต่ละรุ่นที่คุณสนใจอยู่ ดังนั้นการตัดสินใจเลือกซื้อจะขึ้นอยู่กับข้อมูลที่แน่นอน เช่น:

1. สมาร์ทโฟน A: ราคา 10,000 บาท, หน่วยความจำ 128GB, กล้อง 12MP
2. สมาร์ทโฟน B: ราคา 12,000 บาท, หน่วยความจำ 256GB, กล้อง 16MP

หากให้ความสำคัญกับหน่วยความจำมากกว่า เลือกสมาร์ทโฟน B แต่ถ้าต้องการประหยัดเงินจะเลือกสมาร์ทโฟน A

ตัวอย่าง 2: การจัดสรรทรัพยากรในโรงงาน

ในโรงงานผลิตสินค้าหนึ่ง สมมติว่ามีเครื่องจักร 2 เครื่องที่สามารถผลิตสินค้าเดียวกันได้ โดยที่:

- เครื่องจักร A ผลิตสินค้าได้ 100 หน่วยต่อวัน ค่าใช้จ่าย 5,000 บาทต่อวัน
- เครื่องจักร B ผลิตสินค้าได้ 150 หน่วยต่อวัน ค่าใช้จ่าย 7,500 บาทต่อวัน

หากเป้าหมายคือการผลิตสินค้าให้ได้มากที่สุดด้วยต้นทุนที่ต่ำที่สุด การตัดสินใจเลือกเครื่องจักรจะง่ายมากเพราะข้อมูลทั้งหมดที่จำเป็นในการตัดสินใจมีความแน่นอน โดยจะเลือกใช้เครื่องจักร B เนื่องจากสามารถผลิตได้มากกว่าและค่าใช้จ่ายต่อหน่วยผลิตต่ำกว่า

ตัวอย่าง 3: การเลือกโครงการลงทุน

บริษัทหนึ่งมีงบประมาณสำหรับลงทุนในโครงการใหม่ 2 โครงการ โดยที่ข้อมูลทางการเงินของแต่ละโครงการมีดังนี้:

- โครงการ A: ลงทุน 1,000,000 บาท ผลตอบแทนคาดหวัง 200,000 บาทต่อปี
- โครงการ B: ลงทุน 1,000,000 บาท ผลตอบแทนคาดหวัง 250,000 บาทต่อปี

ในกรณีนี้ หากพิจารณาเฉพาะผลตอบแทนคาดหวัง บริษัทจะเลือกลงทุนในโครงการ B เพราะมีผลตอบแทนสูงกว่า

การตัดสินใจภายใต้ความแน่นอนเป็นการตัดสินใจที่มีข้อมูลครบถ้วนและชัดเจนเกี่ยวกับผลลัพธ์ที่เป็นไปได้ของแต่ละทางเลือก โดยไม่มีความเสี่ยงหรือความไม่แน่นอนเข้ามาเกี่ยวข้อง การตัดสินใจในลักษณะนี้จะง่ายกว่าและสามารถคาดการณ์ผลลัพธ์ได้แม่นยำมากกว่า

การตัดสินใจภายใต้ความไม่แน่นอน

การตัดสินใจภายใต้ความไม่แน่นอน (Decision making under uncertainty) เกิดขึ้นเมื่อผู้ตัดสินใจไม่มีข้อมูลครบถ้วนหรือไม่สามารถคาดการณ์ผลลัพธ์ของการตัดสินใจได้อย่างแน่นอน ในสถานการณ์เช่นนี้มักใช้เกณฑ์การตัดสินใจต่าง ๆ เพื่อช่วยในการตัดสินใจ รวมถึงเกณฑ์ Maximax, Maximin และ Minimax Regret

เกณฑ์การตัดสินใจ

Maximax Criteria: เลือกทางเลือกที่มีผลลัพธ์ที่ดีที่สุดจากผลลัพธ์ที่ดีที่สุดในทุกทางเลือก มักเหมาะสำหรับผู้ที่มีแนวโน้มจะเสี่ยง

Maximin Criteria: เลือกทางเลือกที่มีผลลัพธ์ที่ดีที่สุดจากผลลัพธ์ที่แย่ที่สุดในทุกทางเลือก มักเหมาะสำหรับผู้ที่หลีกเลี่ยงความเสี่ยง

Minimax Regret Criteria: เลือกทางเลือกที่มีค่าเสียใจ (regret) สูงสุดต่ำที่สุด โดยคำนึงถึงการเปรียบเทียบกับผลลัพธ์ที่ดีที่สุดที่เป็นไปได้ในแต่ละสถานการณ์

ตัวอย่าง

สมมติว่ามีการตัดสินใจลงทุนในโครงการ 3 โครงการ (A, B, และ C) โดยมีผลลัพธ์ที่เป็นไปได้ภายใต้สถานการณ์เศรษฐกิจที่แตกต่างกัน 3 แบบ (เศรษฐกิจดี, เศรษฐกิจปานกลาง, เศรษฐกิจแย่)

โครงการ	เศรษฐกิจดี	เศรษฐกิจปานกลาง	เศรษฐกิจแย่
A	300,000	200,000	100,000
B	400,000	100,000	50,000
C	500,000	300,000	-100,000

การใช้ Maximax Criteria

หากอยู่ในรูปแบบของรายรับ รายได้ กำไร เลือกทางเลือกที่มีผลลัพธ์ที่ดีที่สุดจากผลลัพธ์ที่ดีที่สุดในทุกทางเลือก:

A: สูงสุด = 300,000

B: สูงสุด = 400,000

C: สูงสุด = 500,000

ดังนั้นเลือกโครงการ C (500,000)

การใช้ Maximin Criteria

หากอยู่ในรูปของต้นทุนหรือค่าใช้จ่ายเลือกทางเลือกที่มีผลลัพธ์ที่ดีที่สุดจากผลลัพธ์ที่แย่ที่สุดในทุกทางเลือก:

A: ต่ำสุด = 100,000

B: ต่ำสุด = 50,000

C: ต่ำสุด = -100,000

ดังนั้นเลือกโครงการ A (100,000)

การใช้ Minimax Regret Criteria

คำนวณค่าเสียใจ (regret) สำหรับแต่ละทางเลือกในแต่ละสถานการณ์:

การใช้ Minimax Regret Criteria เพื่อตัดสินใจภายใต้ความไม่แน่นอนต้องคำนวณตารางค่าเสียใจ (Opportunity Loss Table) สำหรับแต่ละทางเลือกในแต่ละสถานการณ์ ค่าเสียใจคำนวณจากการเปรียบเทียบผลลัพธ์ของแต่ละทางเลือกกับผลลัพธ์ที่ดีที่สุดในสถานการณ์นั้น ๆ แล้วหาค่าเสียใจสูงสุดในแต่ละทางเลือก

ตารางผลลัพธ์

โครงการ	**เศรษฐกิจดี**	**เศรษฐกิจปานกลาง**	**เศรษฐกิจแย่**
A	300,000	200,000	100,000
B	400,000	100,000	50,000
C	500,000	300,000	-100,000

ขั้นตอนในการคำนวณ Minimax Regret Criteria

1. สร้างตาราง Opportunity Loss Table

คำนวณค่าเสียใจ (regret) คล้ายกับค่าเสียโอกาส สำหรับแต่ละทางเลือกในแต่ละสถานการณ์ โดยใช้ผลลัพธ์ที่ดีที่สุดในแต่ละสถานการณ์เป็นฐาน

สถานการณ์	**เศรษฐกิจดี**	**เศรษฐกิจปานกลาง**	**เศรษฐกิจแย่**
ผลลัพธ์ที่ดีที่สุด	500,000	300,000	100,000

คำนวณค่าเสียใจโดยการลบผลลัพธ์ของแต่ละโครงการจากผลลัพธ์ที่ดีที่สุดในแต่ละสถานการณ์:

โครงการ	**เศรษฐกิจดี (500,000-...)**	**เศรษฐกิจปานกลาง (300,000-...)**	**เศรษฐกิจแย่ (100,000-...)**
A	500,000 - 300,000 = 200,000	300,000 - 200,000 = 100,000	100,000 - 100,000 = 0
B	500,000 - 400,000 = 100,000	300,000 - 100,000 = 200,000	100,000 - 50,000 = 50,000
C	500,000 - 500,000 = 0	300,000 - 300,000 = 0	100,000 - (-100,000) = 200,000

2. สร้างตาราง Opportunity Loss Table

โครงการ	**เศรษฐกิจดี**	**เศรษฐกิจปานกลาง**	**เศรษฐกิจแย่**
A	200,000	100,000	0
B	100,000	200,000	50,000
C	0	0	200,000

3. หาค่าเสียใจสูงสุดสำหรับแต่ละโครงการ

โครงการ	ค่าเสียใจสูงสุด
A	200,000
B	200,000
C	200,000

4. เลือกโครงการที่มีค่าเสียใจสูงสุดต่ำที่สุด (Minimax Regret)

เนื่องจากค่าเสียใจสูงสุดสำหรับทุกโครงการเท่ากันที่ 200,000 การตัดสินใจจะพิจารณาเลือกโครงการใดก็ได้ตามเกณฑ์ Minimax Regret เนื่องจากทุกโครงการมีค่าเสียใจสูงสุดเท่ากัน

สรุป ในตัวอย่างนี้ ทุกโครงการ (A, B, และ C) มีค่าเสียใจสูงสุดเท่ากันที่ 200,000 ดังนั้นการตัดสินใจตามเกณฑ์ Minimax Regret จะสามารถเลือกโครงการใดก็ได้เพราะไม่มีโครงการใดที่มีค่าเสียใจสูงสุดน้อยกว่าโครงการอื่น

การตัดสินใจภายใต้ความเสี่ยง

การตัดสินใจภายใต้ความเสี่ยง (Decision Making Under Risk) เกิดขึ้นเมื่อผู้ตัดสินใจมีข้อมูลเกี่ยวกับความน่าจะเป็นของผลลัพธ์ต่าง ๆ แต่ละทางเลือก การตัดสินใจเชิงนี้มักใช้เครื่องมือทางคณิตศาสตร์และสถิติเพื่อวิเคราะห์ความเสี่ยงและผลตอบแทนที่เป็นไปได้

ขั้นตอนในการตัดสินใจภายใต้ความเสี่ยง

1. ระบุทางเลือก (Identify Alternatives): กำหนดทางเลือกต่าง ๆ ที่เป็นไปได้ในการตัดสินใจ

2. ระบุสถานการณ์ที่เป็นไปได้ (Identify Possible Outcomes): กำหนดผลลัพธ์ที่เป็นไปได้สำหรับแต่ละทางเลือก
3. กำหนดความน่าจะเป็น (Assign Probabilities): กำหนดความน่าจะเป็นของแต่ละผลลัพธ์
4. คำนวณค่าคาดหวัง (Calculate Expected Value): คำนวณมูลค่าคาดหวัง (Expected Value) ของแต่ละทางเลือกโดยใช้สูตร
5. เลือกทางเลือกที่ดีที่สุด (Select the Best Alternative): เลือกทางเลือกที่มีค่าคาดหวังดีที่สุด

สูตรการคำนวณค่าคาดหวัง (Expected Value, EV)

$$EV=\sum (P_i \times X_i)$$

โดย:

- P_i คือ ความน่าจะเป็นของผลลัพธ์ที่ i
- X_i คือ มูลค่าของผลลัพธ์ที่ i

ตัวอย่างการตัดสินใจภายใต้ความเสี่ยง

สมมติว่าบริษัทหนึ่งกำลังพิจารณาโครงการลงทุน 2 โครงการ (A และ B) แต่ละโครงการมีผลลัพธ์ที่เป็นไปได้ 3 สถานการณ์ (เศรษฐกิจดี, เศรษฐกิจปานกลาง, เศรษฐกิจแย่) พร้อมกับความน่าจะเป็นของแต่ละสถานการณ์

ข้อมูล

โครงการ A:

เศรษฐกิจดี: ผลลัพธ์ = 300,000 บาท, ความน่าจะเป็น = 0.3

เศรษฐกิจปานกลาง: ผลลัพธ์ = 200,000 บาท, ความน่าจะเป็น = 0.5

เศรษฐกิจแย่: ผลลัพธ์ = 100,000 บาท, ความน่าจะเป็น = 0.2

โครงการ B:

เศรษฐกิจดี: ผลลัพธ์ = 400,000 บาท, ความน่าจะเป็น = 0.3

เศรษฐกิจปานกลาง: ผลลัพธ์ = 100,000 บาท, ความน่าจะเป็น = 0.5

เศรษฐกิจแย่: ผลลัพธ์ = 50,000 บาท, ความน่าจะเป็น = 0.2

คำนวณค่าคาดหวัง (Expected Value, EV)

โครงการ A: $EV(A)=(0.3\times300,000)+(0.5\times200,000)+(0.2\times100,000$

$EV(A)=90,000+100,000+20,000$

$EV(A)=210,000$

โครงการ B: $EV(B)=(0.3\times400,000)+(0.5\times100,000)+(0.2\times50,000)$

$EV(B)=120,000+50,000+10,000$

$EV(B)=180,000$

การตัดสินใจ

เปรียบเทียบค่าคาดหวังของแต่ละโครงการ:

- ค่าคาดหวังของโครงการ A = 210,000 บาท
- ค่าคาดหวังของโครงการ B = 180,000 บาท

เลือกโครงการ A เพราะมีค่าคาดหวังสูงกว่าคือ 210,000 บาท

สรุป การตัดสินใจภายใต้ความเสี่ยงใช้การคำนวณค่าคาดหวังเพื่อเลือกทางเลือกที่ดีที่สุดโดยพิจารณาความน่าจะเป็นของผลลัพธ์ต่าง ๆ ทำให้การตัดสินใจมีพื้นฐานเชิงเหตุผลมากขึ้นแม้จะมีความเสี่ยงที่เกี่ยวข้อง

แนวทางการตัดสินใจภายใต้ความเสี่ยงยังรวมถึงการใช้เกณฑ์แบบ Bayesian (Bayesian Criterion) ซึ่งเป็นการนำทฤษฎีเบย์สมาใช้ในการตัดสินใจ โดยการคำนวณค่าคาดหวังแบบ Bayesian จะพิจารณาความน่าจะ

เป็นตามเงื่อนไข (Posterior Probabilities) ซึ่งอาจปรับเปลี่ยนตามข้อมูลใหม่หรือหลักฐานเพิ่มเติมที่ได้รับ

ขั้นตอนในการใช้ Bayesian Criterion

- กำหนดความน่าจะเป็นล่วงหน้า (Prior Probabilities): กำหนดความน่าจะเป็นเบื้องต้นสำหรับสถานการณ์ต่าง ๆ
- รวบรวมข้อมูลใหม่ (Gather New Information): รวบรวมข้อมูลหรือหลักฐานใหม่ที่เกี่ยวข้อง
- คำนวณความน่าจะเป็นตามเงื่อนไข (Calculate Posterior Probabilities): ใช้ทฤษฎีเบย์สในการปรับปรุงความน่าจะเป็นตามข้อมูลใหม่
- คำนวณค่าคาดหวังแบบ Bayesian (Calculate Bayesian Expected Value): คำนวณค่าคาดหวังสำหรับแต่ละทางเลือกโดยใช้ความน่าจะเป็นตามเงื่อนไข
- เลือกทางเลือกที่ดีที่สุด (Select the Best Alternative): เลือกทางเลือกที่มีค่าคาดหวังแบบ Bayesian สูงที่สุด

สูตรทฤษฎีเบย์ส

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

P(H|E) คือ ความน่าจะเป็นของเหตุการณ์ H เมื่อมีข้อมูลหรือหลักฐาน E
P(E|H) คือ ความน่าจะเป็นของข้อมูลหรือหลักฐาน E เมื่อเหตุการณ์ H เกิดขึ้น
P(H) คือ ความน่าจะเป็นล่วงหน้าของเหตุการณ์ H
P(E) คือ ความน่าจะเป็นของข้อมูลหรือหลักฐาน E

ตัวอย่างการใช้ Bayesian Criterion

สมมติว่าบริษัทหนึ่งกำลังพิจารณาโครงการลงทุน 2 โครงการ (A และ B) โดยมีสถานการณ์เศรษฐกิจที่แตกต่างกัน 3 แบบ (เศรษฐกิจดี, เศรษฐกิจปานกลาง , เศรษฐกิจแย่) และมีข้อมูลใหม่ที่ได้มาเกี่ยวกับแนวโน้มเศรษฐกิจ

ข้อมูลเบื้องต้น (Prior Probabilities)

- เศรษฐกิจดี: ความน่าจะเป็น = 0.3
- เศรษฐกิจปานกลาง: ความน่าจะเป็น = 0.5
- เศรษฐกิจแย่: ความน่าจะเป็น = 0.2

ข้อมูลใหม่ (New Information)

สมมติว่ามีข้อมูลใหม่ว่าแนวโน้มเศรษฐกิจมีการเปลี่ยนแปลง ความน่าจะเป็นตามเงื่อนไข (Posterior Probabilities) ที่ได้จากการใช้ทฤษฎีเบย์สคือ:

- เศรษฐกิจดี: ความน่าจะเป็น = 0.4
- เศรษฐกิจปานกลาง: ความน่าจะเป็น = 0.4
- เศรษฐกิจแย่: ความน่าจะเป็น = 0.2

ผลลัพธ์ของโครงการ

โครงการ	เศรษฐกิจดี	เศรษฐกิจปานกลาง	เศรษฐกิจแย่
A	300,000	200,000	100,000
B	400,000	100,000	50,000

คำนวณค่าคาดหวังแบบ Bayesian (Bayesian Expected Value)

โครงการ A: $EV(A)=(0.4\times300,000)+(0.4\times200,000)+(0.2\times100,000)$

$EV(A)=120,000+80,000+20,000$

$EV(A)=220,000$

โครงการ B: $EV(B)=(0.4\times400,000)+(0.4\times100,000)+(0.2\times50,000)$

$EV(B)=160,000+40,000+10,000$

$EV(B)=210,000$

การตัดสินใจ

เปรียบเทียบค่าคาดหวังแบบ Bayesian ของแต่ละโครงการ:

- ค่าคาดหวังของโครงการ A = 220,000 บาท
- ค่าคาดหวังของโครงการ B = 210,000 บาท

เลือกโครงการ A เพราะมีค่าคาดหวังแบบ Bayesian สูงกว่าคือ 220,000 บาท **สรุป** การตัดสินใจภายใต้ความเสี่ยงโดยใช้ Bayesian Criterion ช่วยให้ผู้ตัดสินใจสามารถใช้ข้อมูลใหม่และปรับปรุงความน่าจะเป็นเพื่อให้การตัดสินใจมีความแม่นยำมากขึ้น โดยคำนึงถึงข้อมูลและสถานการณ์ที่เปลี่ยนแปลง

สรุป

การใช้สถิติในการพยากรณ์และการตัดสินใจทางธุรกิจเป็นกระบวนการสำคัญที่ช่วยให้ธุรกิจสามารถวางแผนและดำเนินกิจการได้อย่างมั่นคงและมีประสิทธิภาพ โดยใช้สถิติในการพยากรณ์ เช่น การวิเคราะห์แนวโน้มตลาดหรือการพยากรณ์ยอดขาย ธุรกิจสามารถใช้ข้อมูลที่มีอยู่เพื่อทำนายผลลัพธ์ในอนาคตได้ การใช้ข้อมูลสถิติในการตัดสินใจทางธุรกิจช่วยให้ทราบถึงข้อมูลสำคัญและการเชื่อมโยงระหว่างตัวแปรต่าง ๆ ที่ส่งผลต่อกิจการ เช่น การวิเคราะห์ข้อมูลลูกค้าเพื่อปรับปรุงผลิตภัณฑ์หรือบริการ ทำให้สามารถวางกลยุทธ์ใหม่โดยอิงตามข้อมูลที่ได้รับจากการวิเคราะห์ สรุปคือ การใช้สถิติทั้งในด้านการพยากรณ์และการตัดสินใจเป็นเครื่องมือที่มีประสิทธิภาพในการสนับสนุนการเติบโตและความสำเร็จของธุรกิจในระยะยาว

อ่านประกอบ

https://elcls.ssru.ac.th/piyamas_kl/pluginfile.php/79/mod_resource/content/1/Chapter%203%20%E0%B8%9A%E0%B8%97%E0%B8%97%E0%B8%B5%E0%B9%88%203%20Forecast.pdf

เอกสารอ้างอิง

กรินทร์ กาญทนานนท์. (2561). *การพยากรณ์ทางสถิติ*. กรุงเทพฯ. : ซีเอ็ดยูเคชั่น

ภูมิฐาน รังคกุลนุวัฒน์. (2556). *การวิเคราะห์อนุกรมเวลาสำหรับเศรษฐศาสตร์และธุรกิจ*. กรุงเทพฯ. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

รัชนี ภูวพัฒนะพันธุ์และมาริสา แก้วสุวรรณ. (2562). *สถิติธุรกิจและการวิเคราะห์เชิงปริมาณ*. ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง.

วุฒิ สุขเจริญ. (2561). *วิจัยการตลาด*. กรุงเทพฯ : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

บทที่ 12
ดัชนีและการประยุกต์ใช้

ดัชนีเป็นเครื่องมือทางสถิติที่ใช้ในการวัดและบ่งชี้ถึงแนวโน้มหรือสภาพการเปลี่ยนแปลงของปรากฏการณ์หนึ่งๆ ในระยะเวลาที่กำหนด โดยมักจะเป็นตัวเลขที่สะท้อนสภาพจริงของสิ่งที่ต้องการวัด เช่น ดัชนีราคา, ดัชนีผู้บริโภค, และดัชนีผู้ผลิต ซึ่งสามารถนำมาใช้ในการวิเคราะห์แนวโน้มของตลาด การเงิน หรือการเศรษฐกิจโดยรวมได้ การประยุกต์ใช้สถิติในสาขาต่างๆ เช่น การวิเคราะห์การตลาดและการวิเคราะห์การเงิน เป็นการนำเอาข้อมูลสถิติมาใช้ในการวิเคราะห์และทำนายผลลัพธ์ในสถานการณ์ต่างๆ เพื่อช่วยในการตัดสินใจทางธุรกิจ การวิเคราะห์การตลาดทำให้เราเข้าใจพฤติกรรมของผู้บริโภค และเสนอแนวทางในการพัฒนาผลิตภัณฑ์หรือบริการให้ตอบสนองต่อความต้องการของตลาดอย่างมีประสิทธิภาพ ในขณะเดียวกัน การวิเคราะห์การเงินช่วยให้เราเข้าใจสภาพการเงินขององค์กรหรือบุคคล ทำนายแนวโน้มการเงินในอนาคต และจัดการทรัพยากรการเงินให้มีประสิทธิภาพในการบริหารจัดการ

เลขดัชนี

เลขดัชนีสามารถหมายถึงหลายสิ่งขึ้นอยู่กับบริบท ต่อไปนี้คือประเภทของเลขดัชนีที่พบบ่อยและคำอธิบายที่เกี่ยวข้อง (ดวงใจ วิสกุลและคณะ, 2552)

- **เลขดัชนีในทางคณิตศาสตร์**: เลขดัชนี หรือเลขยกกำลัง (Exponent) เป็นการแสดงการคูณจำนวนตัวเลขด้วยตัวมันเองหลายครั้ง เช่น $2^3 = 2 \times 2 \times 2 = 8$
- **เลขดัชนีในทางสถิติ:** ดัชนี (Index) ใช้ในการเปรียบเทียบค่าต่างๆ ในช่วงเวลาต่างๆ หรือระหว่างกลุ่มต่างๆ เช่น ดัชนีราคาผู้บริโภค (Consumer

Price Index: CPI) ซึ่งใช้วัดการเปลี่ยนแปลงของราคาสินค้าและบริการในช่วงเวลาที่ต่างกัน

• **เลขดัชนีในทางเศรษฐศาสตร์และการเงิน**: ดัชนีตลาดหุ้น (Stock Market Index) เช่น ดัชนี S&P 500 หรือดัชนี SET ของตลาดหลักทรัพย์แห่งประเทศไทย ใช้ในการวัดผลการดำเนินงานของกลุ่มหุ้นบางกลุ่มหรือทั้งตลาด หรือ ดัชนีชี้วัดเศรษฐกิจ (Economic Indicator Index) เช่น ดัชนีการผลิตอุตสาหกรรม (Industrial Production Index) หรือดัชนีความเชื่อมั่นผู้บริโภค (Consumer Confidence Index)

• **เลขดัชนีในทางเอกสาร**: เลขดัชนีหนังสือ (ISBN - International Standard Book Number) เป็นระบบเลขที่ใช้ในการระบุหนังสือเฉพาะในระดับสากล

เลขดัชนีในทางสถิติ

เลขดัชนีในทางสถิติ (Index Number) เป็นเครื่องมือที่ใช้ในการวัดการเปลี่ยนแปลงของปริมาณหรือมูลค่าของปรากฏการณ์ทางเศรษฐกิจหรือสังคมในช่วงเวลาที่แตกต่างกัน เลขดัชนีสามารถแสดงในรูปแบบของอัตราส่วนหรือร้อยละ ซึ่งทำให้ง่ายต่อการเปรียบเทียบข้อมูลระหว่างช่วงเวลาต่าง ๆ หรือระหว่างกลุ่มต่าง ๆ ต่อไปนี้คือประเภทของเลขดัชนีในทางสถิติที่พบได้บ่อย

1. ดัชนีราคาผู้บริโภค (Consumer Price Index - CPI)

• ความหมาย: วัดการเปลี่ยนแปลงของระดับราคาของตะกร้าสินค้าและบริการที่ครัวเรือนบริโภค

• การคำนวณ: มักใช้สูตรลาสแปร์ (Laspeyres Index) ซึ่งใช้ราคาของสินค้าปีฐานในการเปรียบเทียบ

2. ดัชนีราคาผู้ผลิต (Producer Price Index - PPI)

• ความหมาย: วัดการเปลี่ยนแปลงของระดับราคาของสินค้าที่ผู้ผลิตจำหน่าย

• การคำนวณ: ใช้วิธีการคำนวณคล้ายกับ CPI แต่ใช้ราคาที่ผู้ผลิตได้รับ

3. ดัชนีปริมาณการผลิต (Production Index)

- ความหมาย: วัดการเปลี่ยนแปลงของปริมาณการผลิตในภาคอุตสาหกรรมหรือภาคเกษตรกรรม
- การคำนวณ: ใช้สูตรดัชนีปริมาณเพื่อวัดการเปลี่ยนแปลงของปริมาณที่ผลิตได้

4. ดัชนีมูลค่าการค้าระหว่างประเทศ (International Trade Value Index)

- ความหมาย: วัดการเปลี่ยนแปลงของมูลค่าการส่งออกหรือการนำเข้าสินค้า
- การคำนวณ: ใช้สูตรที่เปรียบเทียบมูลค่าการค้าในช่วงเวลาต่าง ๆ

5. ดัชนีราคาสินค้าเกษตร (Agricultural Price Index)

- ความหมาย: วัดการเปลี่ยนแปลงของระดับราคาของสินค้าทางการเกษตร
- การคำนวณ: มักใช้สูตรลาสแปร์ (Laspeyres Index) หรือ พาส์เชอ (Paasche Index)

วิธีการคำนวณดัชนี

วิธีการคำนวณดัชนี (Index Calculation Methods) มีหลายวิธีในการคำนวณดัชนีในทางสถิติ ได้แก่

1. **ดัชนีแบบ Laspeyres**:

$$\text{Laspeyres Index} = \left(\frac{\sum (p_t \times q_0)}{\sum (p_0 \times q_0)} \right) \times 100$$

- p_t = ราคาสินค้าในปีปัจจุบัน
- p_0 = ราคาสินค้าในปีฐาน
- q_0 = ปริมาณสินค้าที่บริโภคในปีฐาน

2. **ดัชนีแบบ Paasche**:

$$\text{Paasche Index} = \left(\frac{\sum(p_t \times q_t)}{\sum(p_0 \times q_t)} \right) \times 100$$

- q_t= ปริมาณสินค้าที่บริโภคในปีปัจจุบัน

3. **ดัชนีแบบ Fisher**:

$$\text{Fisher Index} = \sqrt{\text{Laspeyres Index} \times \text{Paasche Index}}$$

เป็นดัชนีที่ผสมผสานระหว่างดัชนีแบบ Laspeyres และ Paasche เพื่อให้ได้ค่าที่สมดุลมากขึ้น

ประโยชน์ของเลขดัชนี

- ช่วยในการวิเคราะห์แนวโน้มการเปลี่ยนแปลงของราคาและปริมาณ
- เป็นเครื่องมือในการตัดสินใจทางเศรษฐกิจและนโยบายสาธารณะ
- ใช้ในการคำนวณอัตราเงินเฟ้อและการปรับค่าใช้จ่ายในสัญญาทางการเงิน

เลขดัชนีในทางสถิติเป็นเครื่องมือที่มีประโยชน์อย่างมากในการวิเคราะห์และทำความเข้าใจการเปลี่ยนแปลงในเศรษฐกิจและสังคม เพื่อช่วยในการวางแผนและตัดสินใจได้อย่างมีประสิทธิภาพมากขึ้น

ดัชนีราคาผู้บริโภค

ดัชนีราคาผู้บริโภค (Consumer Price Index - CPI) เป็นเครื่องมือสำคัญในการวัดการเปลี่ยนแปลงของระดับราคาสินค้าและบริการที่ครัวเรือนบริโภคในช่วงเวลาหนึ่ง โดยมีการเปรียบเทียบกับช่วงเวลาอ้างอิงหรือปีฐาน CPI เป็นตัวชี้วัดสำคัญในการวิเคราะห์อัตราเงินเฟ้อและใช้ในนโยบายทางเศรษฐกิจและการเงิน

วัตถุประสงค์ของ CPI

- **วัดอัตราเงินเฟ้อ**: CPI เป็นตัวชี้วัดหลักที่ใช้ในการวัดการเปลี่ยนแปลงของราคา และเป็นการสะท้อนถึงอัตราเงินเฟ้อหรือภาวะเงินฝืด

• **ปรับค่าจ้างและรายได้**: ใช้ CPI ในการปรับปรุงค่าจ้าง, เงินบำนาญ, และสัญญาทางการเงินเพื่อรักษากำลังซื้อ

• **การวางนโยบายเศรษฐกิจ**: หน่วยงานรัฐบาลใช้ CPI ในการวางนโยบายการเงินและการคลัง รวมถึงการกำหนดอัตราดอกเบี้ย

วิธีการคำนวณ CPI

CPI คำนวณโดยการรวบรวมราคาสินค้าและบริการที่ครัวเรือนทั่วไปบริโภค แล้วนำมาคำนวณดัชนีด้วยสูตร Laspeyres Index เป็นส่วนใหญ่ สูตรการคำนวณ CPI ดังนี้

$$\text{Laspeyres Index} = \left(\frac{\sum (p_t \times q_0)}{\sum (p_0 \times q_0)} \right) \times 100$$

- pt = ราคาของสินค้าหรือบริการในปีปัจจุบัน
- p_0 = ราคาของสินค้าหรือบริการในปีฐาน
- q_0 = ปริมาณสินค้าหรือบริการที่บริโภคในปีฐาน

การเก็บรวบรวมข้อมูล

การคำนวณ CPI ต้องมีการเก็บรวบรวมข้อมูลจากหลากหลายแหล่ง เช่น

• **ร้านค้าและห้างสรรพสินค้า**: รวบรวมข้อมูลราคาของสินค้าอุปโภคบริโภค

• **บริการสาธารณะและเอกชน**: รวบรวมข้อมูลราคาบริการ เช่น ค่าไฟฟ้า, ค่าน้ำประปา, ค่าเช่าบ้าน

• **การสำรวจครัวเรือน**: ใช้แบบสำรวจการบริโภคของครัวเรือนเพื่อกำหนดน้ำหนักของสินค้าและบริการต่าง ๆ ในตะกร้า CPI

โครงสร้างของตะกร้า CPI

ตะกร้า CPI ประกอบด้วยกลุ่มสินค้าที่ครัวเรือนทั่วไปบริโภคเป็นประจำ กลุ่มสินค้าที่สำคัญ ได้แก่: อาหารและเครื่องดื่ม เสื้อผ้าและรองเท้า ที่อยู่อาศัย พาหนะและการขนส่ง การรักษาพยาบาล การบันเทิงและการพักผ่อน

การตีความ CPI

- **CPI เพิ่มขึ้น**: แสดงถึงการเพิ่มขึ้นของราคาสินค้าและบริการ ซึ่งอาจนำไปสู่อัตราเงินเฟ้อ
- **CPI ลดลง**: แสดงถึงการลดลงของราคาสินค้าและบริการ ซึ่งอาจนำไปสู่ภาวะเงินฝืด

ตัวอย่างการใช้งาน

- **รัฐบาล**: ใช้ CPI ในการวางแผนนโยบายเศรษฐกิจและการปรับปรุงค่าจ้าง
- **ธุรกิจ**: ใช้ในการตัดสินใจด้านการกำหนดราคาและการวางแผนการเงิน
- **บุคคลทั่วไป**: ใช้ในการปรับการบริโภคและการวางแผนการเงินส่วนบุคคล

ความสำคัญของ CPI

CPI เป็นตัวชี้วัดที่สำคัญและใช้กันอย่างแพร่หลายเพื่อทำความเข้าใจการเปลี่ยนแปลงของราคาสินค้าและบริการที่ครัวเรือนทั่วไปบริโภค และมีบทบาทสำคัญในการวางแผนและตัดสินใจทั้งในระดับบุคคลและระดับนโยบาย

ตัวอย่างการคำนวณดัชนีราคาผู้บริโภค (CPI) แบบละเอียด

ข้อมูลพื้นฐาน

สมมติว่ามีข้อมูลราคาสินค้าและบริการสองประเภทในปีฐานและปีปัจจุบัน รวมถึงปริมาณการบริโภคในปีฐาน ดังนี้

สินค้า/บริการ	ราคาปีฐาน ($\boldsymbol{p}_0$)	ราคาปีปัจจุบัน ($\boldsymbol{p}t$)	ปริมาณปีฐาน ($\boldsymbol{q}_0$)
ข้าวสาร	10	12	100
เสื้อผ้า	20	25	50

ขั้นตอนการคำนวณ CPI

1. **คำนวณค่าผลรวมของ $\boldsymbol{p}_0 \times \boldsymbol{q}_0$**

   $\sum(\boldsymbol{p}_0+\boldsymbol{q}_0)$ = (10×100) + (20×50) = 1000+1000 =2000

คำนวณค่าผลรวมของ $p_t \times q_0$

$\sum(p_t + q_0)$ = (12×100)+(25×50) =1200+1250 =2450

คำนวณ CPI:

CPI = $\left(\frac{\sum(p_t+q_0)}{\sum(p_0+q_0)}\right)$ ×100

CPI= $(\frac{2450}{2000})$×100 = 1.225×100 =122.5

การตีความผลลัพธ์

CPI = 122.5 หมายความว่า ราคาสินค้าและบริการในปีปัจจุบันเพิ่มขึ้น 22.5% เมื่อเทียบกับปีฐาน

เพิ่มเติม: การคำนวณอัตราเงินเฟ้อ

หากเราต้องการคำนวณอัตราเงินเฟ้อจากข้อมูล CPI สามารถทำได้โดยใช้สูตรดังนี้:

อัตราเงินเฟ้อ=$\frac{\text{CPI ในปีปัจจุบัน} - \text{CPI ในปีที่ผ่านมา}}{\text{CPI ในปีที่ผ่านมา}}$×100

สมมติว่า CPI ในปีที่ผ่านมาเป็น 115

อัตราเงินเฟ้อ=$\frac{122.5-115}{115}$×100 =7.5115×100 ≈ 6.52%

ดังนั้น อัตราเงินเฟ้ออยู่ที่ประมาณ 6.52% ค่ะ

สรุป

การคำนวณ CPI และอัตราเงินเฟ้อช่วยให้สามารถเข้าใจการเปลี่ยนแปลงของระดับราคาสินค้าและบริการในช่วงเวลาที่แตกต่างกันได้ ซึ่งเป็นข้อมูลสำคัญสำหรับการวางแผนทางเศรษฐกิจและการเงิน

ดัชนีราคาผู้ผลิต

ดัชนีราคาผู้ผลิต (Producer Price Index - PPI) เป็นตัวชี้วัดที่สำคัญในการวัดการเปลี่ยนแปลงของระดับราคาของสินค้าและบริการที่ผู้ผลิตขาย โดย PPI เน้นที่ราคาที่ผู้ผลิตได้รับแทนที่จะเป็นราคาที่ผู้บริโภคจ่าย เหมือนใน

กรณีของดัชนีราคาผู้บริโภค (CPI) PPI เป็นตัวชี้วัดที่ช่วยในการประเมินสภาพเศรษฐกิจและภาวะเงินเฟ้อในระดับต้นน้ำของห่วงโซ่อุปทาน

วัตถุประสงค์ของ PPI

- **วัดการเปลี่ยนแปลงของราคา**: ใช้ในการติดตามและวัดการเปลี่ยนแปลงของราคาสินค้าและบริการที่ผู้ผลิตขาย
- **ชี้วัดภาวะเงินเฟ้อในระดับต้นน้ำ**: ใช้เพื่อประเมินอัตราเงินเฟ้อที่เกิดขึ้นก่อนที่จะส่งต่อไปยังผู้บริโภค
- **วางแผนและตัดสินใจทางเศรษฐกิจ**: หน่วยงานรัฐบาลและธุรกิจใช้ PPI ในการวางแผนนโยบายเศรษฐกิจและการตัดสินใจด้านการเงิน

วิธีการคำนวณ PPI

PPI คำนวณโดยการรวบรวมราคาสินค้าและบริการที่ผู้ผลิตขายในตลาดต่างๆ จากนั้นนำมาคำนวณดัชนีด้วยสูตร Laspeyres Index เป็นส่วนใหญ่ สูตร การคำนวณ PPI ดังนี้

$$\text{PPI} = \left(\frac{\sum (p_t \times q_0)}{\sum (p_0 \times q_0)} \right) \times 100$$

- p_t = ราคาของสินค้าหรือบริการในปีปัจจุบัน
- p_0 = ราคาของสินค้าหรือบริการในปีฐาน
- q_0 = ปริมาณสินค้าหรือบริการที่ขายในปีฐาน

การเก็บรวบรวมข้อมูล

การคำนวณ PPI ต้องมีการเก็บรวบรวมข้อมูลจากหลากหลายแหล่ง เช่น

- **โรงงานและผู้ผลิต**: รวบรวมข้อมูลราคาของสินค้าที่ผลิตในโรงงาน
- **การสำรวจธุรกิจ**: ใช้แบบสำรวจเพื่อเก็บข้อมูลราคาขายส่งจากธุรกิจต่างๆ
- **ตลาดสินค้าโภคภัณฑ์**: รวบรวมข้อมูลราคาสินค้าโภคภัณฑ์ เช่น น้ำมัน, โลหะ, และวัตถุดิบต่างๆ

โครงสร้างของตะกร้า PPI

ตะกร้า PPI ประกอบด้วยกลุ่มสินค้าที่ผลิตและขายโดยผู้ผลิตในตลาดต่างๆ กลุ่มสินค้าที่สำคัญ ได้แก่ **สินค้าการเกษตร**: ราคาสินค้าทางการเกษตรที่ผู้ผลิตขาย **สินค้าก่อสร้าง**: ราคาวัสดุก่อสร้างและอุปกรณ์ **สินค้าพลังงาน**: ราคาน้ำมัน, ก๊าซ, และพลังงานอื่นๆ **สินค้าอุตสาหกรรม**: ราคาวัตถุดิบและสินค้าที่ผลิตในโรงงาน

การตีความ PPI

- **PPI เพิ่มขึ้น**: แสดงถึงการเพิ่มขึ้นของราคาสินค้าและบริการที่ผู้ผลิตขาย ซึ่งอาจเป็นสัญญาณของภาวะเงินเฟ้อ
- **PPI ลดลง**: แสดงถึงการลดลงของราคาสินค้าและบริการที่ผู้ผลิตขาย ซึ่งอาจเป็นสัญญาณของภาวะเงินฝืด

ดัชนีปริมาณการผลิต

ดัชนีปริมาณการผลิต (Production Index) เป็นตัวชี้วัดที่ใช้ในการวัดการเปลี่ยนแปลงของปริมาณการผลิตในภาคอุตสาหกรรมหรือภาคเศรษฐกิจต่างๆ ในช่วงเวลาที่แตกต่างกัน โดยดัชนีนี้ช่วยให้เราทราบถึงการเติบโตหรือหดตัวของการผลิต ซึ่งมีผลต่อสภาพเศรษฐกิจโดยรวม

วัตถุประสงค์ของดัชนีปริมาณการผลิต

วัดการเปลี่ยนแปลงของการผลิต: ช่วยในการติดตามและวิเคราะห์การเปลี่ยนแปลงของปริมาณการผลิตในอุตสาหกรรมต่างๆ

ประเมินประสิทธิภาพทางเศรษฐกิจ: ช่วยในการประเมินภาวะเศรษฐกิจว่ามีการเติบโตหรือหดตัว

สนับสนุนการวางแผนและนโยบาย: ใช้ข้อมูลในการวางแผนและกำหนดนโยบายทางเศรษฐกิจและการผลิต

วิธีการคำนวณดัชนีปริมาณการผลิต

การคำนวณดัชนีปริมาณการผลิตมักใช้วิธีการคำนวณที่คล้ายกับการคำนวณดัชนีราคาผู้บริโภค (CPI) โดยมีการเก็บรวบรวมข้อมูลปริมาณการผลิต

จากแหล่งต่างๆ แล้วนำมาคำนวณดัชนีด้วยสูตร Laspeyres Index หรือสูตรอื่นที่เหมาะสม

สูตร Laspeyres Index สำหรับดัชนีปริมาณการผลิต

$$\text{ดัชนีปริมาณการผลิต} = \left(\frac{\sum(q_t \times p_0)}{\sum(q_0 \times p_0)} \right) \times 100$$

q_t = ปริมาณการผลิตในปีปัจจุบัน

q_0 = ปริมาณการผลิตในปีฐาน

p_0 = ราคาสินค้าในปีฐาน (ใช้เป็นน้ำหนัก)

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลสำหรับการคำนวณดัชนีปริมาณการผลิตมาจากแหล่งต่างๆ เช่น

โรงงานและอุตสาหกรรม: ข้อมูลการผลิตจากโรงงานและภาคอุตสาหกรรมต่างๆ

การสำรวจภาคการผลิต: ใช้แบบสำรวจในการรวบรวมข้อมูลปริมาณการผลิต

หน่วยงานสถิติ: ข้อมูลจากหน่วยงานสถิติที่เก็บรวบรวมข้อมูลเศรษฐกิจและการผลิต

โครงสร้างของดัชนีปริมาณการผลิต

ดัชนีปริมาณการผลิตประกอบด้วยกลุ่มอุตสาหกรรมหลักต่างๆ เช่น

การผลิตสินค้าอุตสาหกรรม: รวมถึงการผลิตเครื่องจักร, อิเล็กทรอนิกส์, ยานยนต์

การผลิตสินค้าเกษตร: รวมถึงการผลิตพืช, ปศุสัตว์, ผลิตภัณฑ์จากเกษตรกรรม

การผลิตสินค้าโภคภัณฑ์: รวมถึงการผลิตเหล็ก, ปิโตรเลียม, วัตถุดิบอื่นๆ

ตัวอย่างการคำนวณดัชนีปริมาณการผลิต

สมมติว่ามีข้อมูลการผลิตสองประเภทในปีฐานและปีปัจจุบัน ดังนี้

ตัวอย่างการคำนวณดัชนีปริมาณการผลิต

สมมติว่าเรามีข้อมูลการผลิตสองประเภทในปีฐานและปีปัจจุบัน ดังนี้:

สินค้า/บริการ	ปริมาณปีฐาน (q_0)	ปริมาณปีปัจจุบัน (q_t)	ราคาปีฐาน (p_0)
รถยนต์	1000	1200	500,000
อิเล็กทรอนิกส์	2000	2500	100,000

1. **คำนวณค่าผลรวมของ $q_0 \times p_0$**

$$\sum(q_0 + p_0) =$$

(1000×500,000)+(2000×100,000)=500,000,000+200,000,000=700,000,000

2. **คำนวณค่าผลรวมของ $q_t \times p_0$**

$$\sum(q_t + p_0) =$$

$\sum(q_t \times p_0)$=(1200×500,000)+(2500×100,000)=600,000,000+250,000,000=850,000,000

3. **คำนวณดัชนีปริมาณการผลิต**

ดัชนีปริมาณการผลิต=$\left(\frac{850,000,000}{700,000,000}\right)$ ×100 =1.2143×100 =121.43

การตีความผลลัพธ์

ดัชนีปริมาณการผลิต = 121.43 หมายความว่า ปริมาณการผลิตในปีปัจจุบันเพิ่มขึ้น 21.43% เมื่อเทียบกับปีฐาน

ดัชนีปริมาณการผลิตเป็นตัวชี้วัดที่สำคัญในการวิเคราะห์ การเปลี่ยนแปลงของปริมาณการผลิตในภาคอุตสาหกรรมและภาคเศรษฐกิจต่างๆ ช่วยให้เราทราบถึงการเติบโตหรือหดตัวของการผลิต ซึ่งมีผลกระทบต่อสภาพเศรษฐกิจโดยรวมและการตัดสินใจในด้านธุรกิจและนโยบายเศรษฐกิจ

การประยุกต์ใช้สถิติในสาขาต่างๆ

สถิติเป็นเครื่องมือที่มีประโยชน์และมีการประยุกต์ใช้อย่างกว้างขวางในหลายสาขาวิชา เช่น การวิเคราะห์การตลาดและการวิเคราะห์การเงิน โดยการใช้สถิติเพื่อวิเคราะห์ข้อมูลและช่วยในการตัดสินใจอย่างมีเหตุผลและมีประสิทธิภาพ มีรายละเอียดพร้อมดังนี้

การวิเคราะห์การตลาด (Marketing Analysis)

1. **การสำรวจตลาด**

การวิเคราะห์เชิงพรรณนา (Descriptive Statistics): ใช้สถิติพื้นฐาน เช่น ค่าเฉลี่ย, มัธยฐาน, การกระจายตัว เพื่อสรุปข้อมูลจากแบบสำรวจ

การทดสอบสมมติฐาน (Hypothesis Testing): ใช้ทดสอบความแตกต่างระหว่างกลุ่มลูกค้าหรือเปรียบเทียบความพึงพอใจระหว่างผลิตภัณฑ์ต่างๆ

การวิเคราะห์การถดถอย (Regression Analysis): ใช้เพื่อหาความสัมพันธ์ระหว่างตัวแปรต่างๆ เช่น ราคากับยอดขาย

2. **การแบ่งกลุ่มลูกค้า (Customer Segmentation)**

การวิเคราะห์กลุ่ม (Cluster Analysis): ใช้ในการแบ่งกลุ่มลูกค้าตามลักษณะทางประชากรและพฤติกรรมการซื้อ

การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis - PCA): ใช้ลดความซับซ้อนของข้อมูลหลายมิติให้เป็นมิติที่น้อยลง

3. **การวิเคราะห์ความพึงพอใจของลูกค้า (Customer Satisfaction Analysis)**

การวิเคราะห์ปัจจัย (Factor Analysis): ใช้ในการระบุปัจจัยหลักที่มีผลต่อความพึงพอใจของลูกค้า

การวิเคราะห์การถดถอยพหุ (Multiple Regression Analysis): ใช้ในการประเมินผลกระทบของหลายตัวแปรต่อความพึงพอใจของลูกค้า

4. **การวิเคราะห์ผลตอบแทนจากการโฆษณา (Advertising ROI Analysis)**

การวิเคราะห์ความสัมพันธ์ (Correlation Analysis): ใช้ตรวจสอบความสัมพันธ์ระหว่างการใช้จ่ายด้านการโฆษณาและยอดขาย

การวิเคราะห์การถดถอย (Regression Analysis): ใช้ประเมินผลกระทบของการใช้จ่ายโฆษณาต่อยอดขายและผลตอบแทน

การวิเคราะห์การเงิน (Financial Analysis)

1. **การวิเคราะห์งบการเงิน (Financial Statement Analysis)**

การวิเคราะห์เชิงพรรณนา (Descriptive Statistics): ใช้ในการสรุปข้อมูลทางการเงิน เช่น อัตราส่วนทางการเงิน (Financial Ratios) ค่าเฉลี่ยของรายได้, ค่าเบี่ยงเบนมาตรฐาน

การทดสอบสมมติฐาน (Hypothesis Testing): ใช้ในการเปรียบเทียบผลการดำเนินงานระหว่างบริษัทต่างๆ

2. **การวิเคราะห์ความเสี่ยงและการจัดการพอร์ตโฟลิโอ (Risk Analysis and Portfolio Management)**

การวิเคราะห์การกระจาย (Variance and Standard Deviation): ใช้ในการวัดความเสี่ยงของพอร์ตโฟลิโอ

การวิเคราะห์การถดถอย (Regression Analysis): ใช้ในการหาความสัมพันธ์ระหว่างผลตอบแทนของสินทรัพย์และปัจจัยทางเศรษฐกิจต่างๆ

การวิเคราะห์พอร์ตโฟลิโอ (Markowitz Portfolio Theory): ใช้ในการหาพอร์ตโฟลิโอที่มีความเสี่ยงต่ำสุดหรือผลตอบแทนสูงสุด

3. **การพยากรณ์ทางการเงิน (Financial Forecasting)**

การวิเคราะห์อนุกรมเวลา (Time Series Analysis): ใช้ในการพยากรณ์รายได้, ค่าใช้จ่าย, และกำไรในอนาคต

การวิเคราะห์การถดถอยพหุ (Multiple Regression Analysis): ใช้ในการพยากรณ์ผลประกอบการทางการเงินโดยใช้ตัวแปรหลายตัว

4. **การวิเคราะห์การลงทุน (Investment Analysis)**

การวิเคราะห์มูลค่าปัจจุบันสุทธิ (Net Present Value - NPV): ใช้ในการประเมินความน่าลงทุนของโครงการ

การวิเคราะห์อัตราผลตอบแทนภายใน (Internal Rate of Return - IRR): ใช้ในการประเมินผลตอบแทนของการลงทุน

การวิเคราะห์ความเสี่ยง (Risk Analysis): ใช้ในการประเมินความเสี่ยงของการลงทุนในสินทรัพย์ต่างๆ

การประยุกต์ใช้ในสาขาอื่นๆ

การแพทย์และสาธารณสุข (Medical and Public Health Analysis)

1. **การทดลองทางคลินิก (Clinical Trials)**

การสุ่มตัวอย่าง (Random Sampling): ใช้ในการสุ่มเลือกผู้เข้าร่วมการทดลอง

การวิเคราะห์เชิงสถิติ (Statistical Significance Testing): ใช้ในการทดสอบผลการรักษาเพื่อดูว่ามีความแตกต่างอย่างมีนัยสำคัญหรือไม่

2. **การระบาดวิทยา (Epidemiology)**

การวิเคราะห์การกระจาย (Incidence and Prevalence Rates): ใช้ในการวัดการแพร่กระจายของโรค

การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis): ใช้ในการประเมินความเสี่ยงของการเกิดโรค

การศึกษา (Education Analysis)

1. **การวิเคราะห์ผลการเรียน (Student Performance Analysis)**

การวิเคราะห์เชิงพรรณนา (Descriptive Statistics): ใช้ในการสรุปผลการเรียน เช่น ค่าเฉลี่ย, ค่ามัธยฐาน, ค่าเบี่ยงเบนมาตรฐาน

การวิเคราะห์การถดถอย (Regression Analysis): ใช้ในการวิเคราะห์ปัจจัยที่มีผลต่อผลการเรียน

2. **การประเมินคุณภาพการศึกษา (Educational Assessment)**

การวิเคราะห์เชิงสถิติ (Statistical Analysis): ใช้ในการประเมินและเปรียบเทียบคุณภาพการศึกษาของโรงเรียนหรือโปรแกรมการศึกษาต่างๆ

วิทยาศาสตร์และการวิจัย (Scientific Research)

1. **การออกแบบการทดลอง (Experimental Design)**

การวิเคราะห์เชิงสถิติ (Statistical Analysis): ใช้ในการออกแบบและวิเคราะห์การทดลองเพื่อตอบคำถามวิจัยอย่างมีประสิทธิภาพ

2. **การวิเคราะห์ข้อมูล (Data Analysis)**

การทดสอบสมมติฐาน (Hypothesis Testing): ใช้ในการทดสอบสมมติฐานต่างๆ จากผลการทดลอง

การวิเคราะห์การถดถอย (Regression Analysis): ใช้ในการศึกษาความสัมพันธ์ระหว่างตัวแปรต่างๆ

สถิติมีการประยุกต์ใช้ในหลายสาขาวิชาและเทคนิคสถิติที่ใช้ในแต่ละสาขาจะแตกต่างกันตามลักษณะของข้อมูลและวัตถุประสงค์การวิเคราะห์ ทั้งนี้ การใช้สถิติที่เหมาะสมช่วยให้การตัดสินใจและการวิเคราะห์ข้อมูลมีความแม่นยำและมีประสิทธิภาพมากขึ้น

สรุป

ดัชนีเป็นเครื่องมือทางสถิติที่สำคัญสำหรับการวัดและวิเคราะห์แนวโน้มหรือการเปลี่ยนแปลงของปรากฏการณ์ต่างๆ เช่น ตลาด การเงิน หรือเศรษฐกิจ โดยมีประโยชน์ในการตัดสินใจทางธุรกิจ การวิเคราะห์พฤติกรรมผู้บริโภค และการจัดการการเงินอย่างมีประสิทธิภาพ ดัชนีช่วยให้สามารถคาดการณ์อนาคต ประเมินสภาพเศรษฐกิจ และเปรียบเทียบผลลัพธ์ในช่วงเวลาต่างๆ ได้อย่างแม่นยำ ทำให้การบริหารจัดการในด้านต่างๆ เป็นไปอย่างมีระบบและมีประสิทธิภาพมากขึ้น

เอกสารอ่านเพิ่มเติม

https://www.rpubs.com/somsak_ch/Price_index

แบบฝึกหัด

แบบฝึกหัดที่ 1: การคำนวณดัชนีราคา

สมมติว่าในปีฐาน (ปี 2020) ราคาของสินค้าสามประเภทคือ:

- ข้าว: 10 บาทต่อกิโลกรัม
- นม: 20 บาทต่อขวด
- น้ำมันพืช: 30 บาทต่อขวด

ในปีถัดมา (ปี 2021) ราคาของสินค้าทั้งสามเพิ่มขึ้นเป็น:

- ข้าว: 12 บาทต่อกิโลกรัม
- นม: 22 บาทต่อขวด
- น้ำมันพืช: 35 บาทต่อขวด

จงคำนวณดัชนีราคาของปี 2021 โดยใช้ปี 2020 เป็นปีฐาน

แบบฝึกหัดที่ 2: การวิเคราะห์ดัชนีผู้บริโภค

ดัชนีผู้บริโภค (CPI) ของประเทศ A ในเดือนมกราคม, กุมภาพันธ์ และมีนาคม ปี 2023 มีค่าเท่ากับ 100, 102 และ 105 ตามลำดับ

- จงคำนวณอัตราการเปลี่ยนแปลงของดัชนีผู้บริโภคในแต่ละเดือน
- จงวิเคราะห์แนวโน้มของดัชนีผู้บริโภคในช่วงสามเดือนนี้ว่ามีแนวโน้มเป็นอย่างไร

เอกสารอ้างอิง

ดวงใจ วิสกุลและคณะ. (2552). สถิติธุรกิจ. กรุงเทพฯ. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

เฉลยแบบฝึกหัดที่ 1: การคำนวณดัชนีราคา

1. ดัชนีราคาของแต่ละสินค้าในปี 2021:
   - ข้าว: (12 / 10) * 100 = 120

- นม: (22 / 20) * 100 = 110
- น้ำมันพืช: (35 / 30) * 100 = 116.67

2. ดัชนีราคาเฉลี่ย:
    - (120 + 110 + 116.67) / 3 = 115.56

เฉลยแบบฝึกหัดที่ 2: การวิเคราะห์ดัชนีผู้บริโภค

1. อัตราการเปลี่ยนแปลงของดัชนีผู้บริโภคในแต่ละเดือน:
    - กุมภาพันธ์: ((102 - 100) / 100) * 100 = 2%
    - มีนาคม: ((105 - 102) / 102) * 100 = 2.94%
2. แนวโน้มของดัชนีผู้บริโภค:
    - แนวโน้มดัชนีผู้บริโภคในช่วงสามเดือนนี้เพิ่มขึ้นอย่างต่อเนื่องจาก 100 ในเดือนมกราคม เป็น 105 ในเดือนมีนาคม แสดงถึงการเพิ่มขึ้นของราคาโดยรวมในช่วงเวลานี้

บรรณานุกรม

กรินทร์ กาญทนานนท์. (2561). *การพยากรณ์ทางสถิติ*. กรุงเทพฯ. : ซีเอ็ดยูเคชั่น

ชัยณรงค์ เพียรภายลุนและคณะ. (2563). การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิด้วยการเลือกลำดับที่ของชุดตัวอย่างแบบครบวงจรเมื่อจัดลำดับแบบสมบูรณ์. *วารสารวิทยาศาสตร์บูรพา*. 25(3) 1026-1034

ดวงใจ วิสกุลและคณะ. (2552). สถิติธุรกิจ. กรุงเทพฯ. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

นภสร รัตนวุฒิขจร. (2565). *การจำลองข้อมูลเพื่อประเมินประสิทธิภาพของการเลือกตัวอย่างแบบมีระบบชนิดผสม*. วิทยานิพนธ์หลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

วราพรรณ ธิยานันท์และกฤษณะ ไวยมัย. (2023). การประยุกต์ใช้ความเปลี่ยนแปลงของข้อมูลเพื่อสุ่มลดข้อมูลสำหรับการแก้ปัญหาการจัดประเภทข้อมูลที่ไม่สมดุล. *วารสารงานวิจัยและพัฒนาเชิงประยุกต์โดยสมาคมECTI*. 3 (3) 29-38

ปรีดาภรณ์ กาญจนสำราญวงศ์.(2560). *หลักสถิติเบื้องต้น*.นนทบุรี. ไอดีซีพรีเมียร์

ภูมิฐาน รังคกุลนุวัฒน์. (2556). *การวิเคราะห์อนุกรมเวลาสำหรับเศรษฐศาสตร์และธุรกิจ*. กรุงเทพฯ. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

รัชนี ภูวพัฒนะพันธุ์และมาริสา แก้วสุวรรณ. (2562). *สถิติธุรกิจและการวิเคราะห์เชิงปริมาณ*. ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง.

วรางคณา เรียนสุทธิ์. (2562). การเปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบไม่อิงพารามิเตอร์ส าหรับการทดสอบภาวะความเท่ากันของความแปรปรวน. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี* .21(2พฤษภาคม-สิงหาคม2562). 163-170

วุฒิ สุขเจริญ. (2561). *วิจัยการตลาด*. กรุงเทพฯ : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

สุพล ดุรงค์วัฒนา. (2558). Regression Models : Analytical-based Approach. กรุงเทพฯ. แดเน็กซ์อินเตอร์คอร์ปอเรชั่น