สถิติธุรกิจ Business Statistics (BUA3127)

รัมภาภัค ฤกษ์วีระวัฒนา

บทที่ 1 บทนำ

- ความสำคัญของสถิติในการวิเคราะห์ข้อมูล
- แนวคิดพื้นฐานเกี่ยวกับการสถิติ
- สรุป

วัตถุประสงค์การเรียนรู้

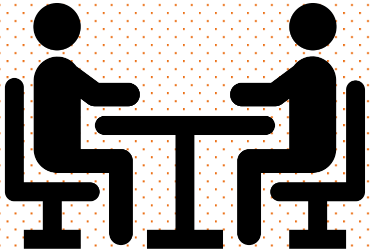
1 ได้ทราบแนวคิดและทฤษฎีทางสถิติ

2 ได้ทราบถึงประชากรและกลุ่มตัวอย่าง

3 เรียนรู้ถึงประโยชน์ของสถิติต่อธุรกิจ

บทนำ

สถิติมีบทบาทสำคัญอย่างมากในการวิเคราะห์และเข้าใจข้อมูลในหลากหลายด้านของชีวิต เริ่มตั้งแต่การตีความข้อมูลในสถานการณ์ประจำวัน ไปจนถึงการใช้สถิติในการตัดสินใจทางธุรกิจและการวิจัยทางวิทยาศาสตร์ เครื่องมือและแนวคิดทางสถิติช่วยให้เราสามารถวิเคราะห์และอธิบายข้อมูลในลักษณะที่เป็นรูปแบบ โดยเน้นการสร้างความเข้าใจและข้อสรุปที่ถูกต้อง



แนวคิดและทฤษฎีบทนำสถิติ

1. ความสำคัญของสถิติ: สถิติช่วยอธิบายข้อมูลต่างๆและมีบทบาทสำคัญในชีวิตประจำวันและในการทำงานในหลายสาขา เช่น การบริหารจัดการ, การวิเคราะห์ข้อมูลทางการแพทย์, การวิเคราะห์การตลาด, ฯลฯ

2. แนวคิดพื้นฐานเกี่ยวกับการสถิติ: คือการอธิบายเกี่ยวกับแนวคิดพื้นฐานที่ใช้ในการสถิติ เช่น ตัวแปร, การกระจาย, ค่าสถิติพื้นฐานเป็นต้น

3. การใช้สถิติในการตัดสินใจ: โดยเฉพาะในธุรกิจและองค์กร เน้นความสำคัญของการใช้ข้อมูลสถิติในการตัดสินใจอย่างมีประสิทธิภาพ

4. การทำงานกับข้อมูล: แสดงถึงขั้นตอนการเก็บรวบรวมข้อมูล, การอ้างอิงแหล่งข้อมูล, การจัดระเบียบข้อมูล, และการตรวจสอบความถูกต้องของข้อมูล

แนวคิดและทฤษฎีบทนำสถิติ

5. การแสดงข้อมูล: อธิบายเกี่ยวกับวิธีการนำเสนอข้อมูลทางสถิติให้เข้าใจง่าย เช่น การใช้กราฟ, แผนภูมิ, และตาราง

6. การวิเคราะห์ข้อมูล: อธิบายเกี่ยวกับกระบวนการในการวิเคราะห์ข้อมูลที่ได้ เช่น การคำนวณค่าสถิติต่างๆ และการตีความผลลัพธ์

7. การนำสถิติไปใช้: อธิบายเกี่ยวกับการนำความรู้สถิติไปใช้ในชีวิตประจำวันและในงานต่างๆ เพื่อการตัดสินใจและการแก้ปัญหา

ความสำคัญของสถิติในการวิเคราะห์ข้อมูล

1. การสร้างความเข้าใจ: สถิติช่วยให้เราเข้าใจและสร้างความเข้าใจเกี่ยวกับข้อมูลที่เรามีอยู่ เช่น การหาแนวโน้ม การกระจายของข้อมูลและความสัมพันธ์ระหว่างตัวแปรต่างๆ

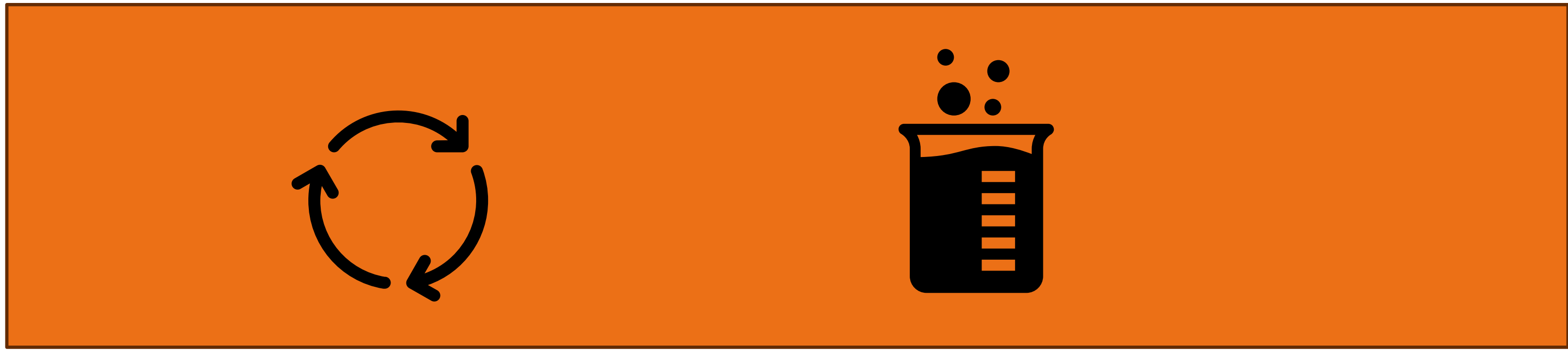
2. การตัดสินใจทางธุรกิจ: สถิติช่วยให้ผู้บริหารและนักวิเคราะห์ทางธุรกิจตัดสินใจในทิศทางที่ถูกต้อง โดยใช้ข้อมูลที่เป็นหลักฐานและคาดการณ์ที่มีความเชื่อถือได้

3. การวิเคราะห์ทางวิทยาศาสตร์: ในการวิจัยทางวิทยาศาสตร์, สถิติช่วยในการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับการทดลองและการสร้างทฤษฎีให

ความสำคัญของสถิติในการวิเคราะห์ข้อมูล

4. การวิเคราะห์ทางการแพทย์: ในการวิเคราะห์ข้อมูลทางการแพทย์, สถิติช่วยให้เราสามารถทำนายโรค, ประสิทธิภาพของการรักษา, และการสร้างนโยบายสาธารณสุข

5. การวิเคราะห์ข้อมูลทางสังคม: สถิติช่วยให้เราเข้าใจแนวโน้มและภาวะสังคมที่สำคัญ เช่น การวิเคราะห์ข้อมูลทางเศรษฐกิจ, การประเมินผลโครงการสังคม



แนวคิดพื้นฐานเกี่ยวกับการสถิติ

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

2. การกระจาย (Distribution): การกระจายของข้อมูลเป็นแนวคิดสำคัญในสถิติ เพราะมันช่วยให้เรารู้ถึงลักษณะการกระจายของข้อมูลว่าเป็นแบบไหน เช่น แบบเบี้ยวหรือแบบกระจายแบบปกติ

แนวคิดพื้นฐานเกี่ยวกับการสถิติ

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

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

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

ความหมายของสถิติ

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

2. ประการที่สอง สถิติ หมายถึงวิชาหรือศาสตร์ที่ว่าด้วยหลักการและระเบียบวิชาการทางสถิติซึ่งประกอบด้วยการเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูลและการตีความหมายข้อมูลและการตีความหมายของข้อมูล

ศัพท์ต่าง ๆ ที่ใช้วิชาสถิติ

1. ประชากร (population) หมายถึง สิ่งต่างๆ ทั้งหมดในเรื่องที่เราสนใจศึกษา ประชากรทางสถิติจะเป็นอะไรก็ได้ไม่ว่า คน สัตว์ หรือสิ่งของ

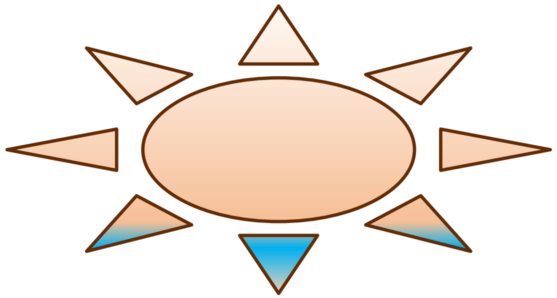
- ถ้าสนใจศึกษารายได้ของคนไทยที่จบการศึกษา ปวส. ในประเทศไทย คนไทยทั้งประเทศที่จบการศึกษา ปวส. เป็นประชากร

- ถ้าสนใจเกี่ยวกับรถยนต์เก๋งในประเทศญี่ปุ่น รถยนต์เก๋งของประเทศญี่ปุ่นทั้งหมดเป็นประชากร

ประชากรสามารถแบ่งได้เป็น 2 ประเภทคือ

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

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



ตัวอย่าง (Sample)

หมายถึง ส่วนหนึ่งหรือส่วนย่อย (Subset) ของประชากรที่ต้องการศึกษาเพื่อที่จะใช้ตัวอย่างเป็นตัวแทนของประชากรเพราะในทางปฏิบัติแล้วคงไม่สามารถนำทุกหน่วยของ ประชากรมาศึกษาได้ จึงนำเพียงส่วนหนึ่ง ของประชากรมาศึกษา เช่น สมมติว่านักศึกษาชั้นมัธยมศึกษาปีที่ 6 ในปี 2566 ทั้งหมด 200,000 คน จะต้องการศึกษาดูว่า นักศึกษาชั้นมัธยมปีที่ 6 จะเสียค่าใช้จ่ายในการศึกษาทั้งหมด ปีละเท่าไร จึงเลือกสุ่มนักศึกษาชั้นมัธยมศึกษาปีที่ 6 ในปี 2566 มา 2,000 คนมาศึกษาหาข้อมูลของค่าใช้จ่าย

ค่าสถิติ (Statistic)

หมายถึง ค่าที่เกิดขึ้นมาจากการนำข้อมูลที่ได้จากกลุ่มตัวอย่างที่เราเก็บมา คำนวณเพื่อให้ได้ค่าตามที่ต้องการ การนำข้อมูลจากกลุ่มตัวอย่างมาคำนวณหาร้อยละ ค่าเฉลี่ยเลขาคณิต ค่ามัธยฐาน

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 ข้อมูลอาจจะเป็นตัวเลข หรือไม่เป็นตัวเลขก็ได้

1. ข้อมูลเชิงประมาณ (Quantitative data) หมายถึง ข้อมูลที่เก็บได้เป็นตัวเลขที่สามารถบอก ถึง ขนาดหรือปริมาณหรือจำนวนได้ เช่น ข้อมูลที่เกี่ยวกับ อายุ จำนวนบุตร เงินเดือน ความสูง น้ำหนัก เป็น ต้น

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

ข้อมูล (Data)

ลักษณะข้อมูลทางสถิติจำแนกตามแหล่งของข้อมูลทางสถิติ ซึ่งแบ่งได้เป็น 2 ชนิด คือ

1. ข้อมูลเชิงปฐมภูมิ คือ ข้อมูลที่ได้มาจากประชากรที่เราต้องการศึกษาโดยตรงและเก็บ รวบรวมโดยการจดบันทึกหรือพิมพ์

2. ข้อมูลทุติยภูมิ คือ ข้อมูลที่ได้แหล่งข้อมูลอื่นที่ไม่ได้จากประชากรโดยตรง โดยแหล่งข้อมูลนี้ ได้เก็บข้อมูลที่สนใจไว้แล้วและเพียงไปนำมาใช้งาน เช่น ข้อมูลของสำนักงานสถิติแห่งชาติ หรือข้อมูล จากกรมวิชาการ กระทรวงเกษตรและสหกรณ์

ขั้นตอนในการปฏิบัติ 4 ขั้นตอนของวิชาสถิติ

สถิติที่มีความหมายเป็นวิชาหรือศาสตร์ การดำเนินการทางสถิติที่เกี่ยวกับข้อมูลเป็นจำนวนมาก โดยมีขั้นตอนในการปฏิบัติ 4 ขั้นตอน ดังนี้

ขั้นที่ 1 **การเก็บรวบรวมข้อมูล** คือ การรวบรวมข้อเท็จจริงจากตัวอย่างที่ผู้วิจัยสนใจในเรื่องนั้นมาข้อมูล

ขั้นที่ 2 **การนำเสนอข้อมูล** คือ การนำเอาข้อมูลที่เก็บรวบรวมมาได้แสดง ซึ่งอาจจะแสดงในรูปของกราฟหรือแผนภูมิ ตารางแจกแจงความถี่ เพื่อที่นำไปผลที่แสดงไปวิเคราะห์ และสรุปผลในขั้นต่อไป

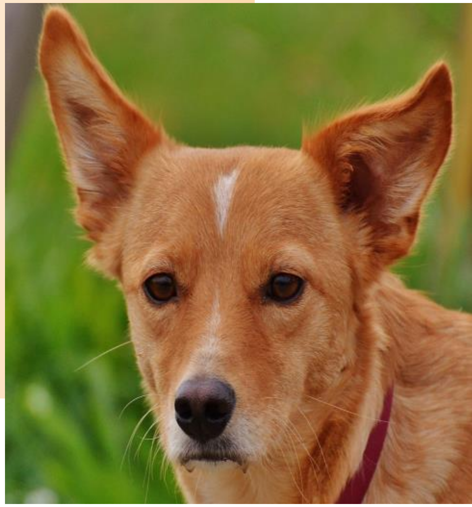
ขั้นที่ 3 **การวิเคราะห์ข้อมูล** คือ การนำข้อมูลที่เก็บรวบรวมมาได้คำนวณเพื่อให้ได้ค่าสถิติที่เรา ต้องการที่จะใช้ประโยชน์ต่อไป เช่น การหาค่าเฉลี่ย การหาค่าส่วนเบี่ยงเบนมาตรฐาน

ขั้นที่ 4 **การตีความหมายข้อมูล** คือ การนำข้อมูลที่ผ่านการวิเคราะห์มาให้ความหมาย การสรุป และอธิบาย ลักษณะของข้อมูลที่ได้วิเคราะห์

ความสำคัญของของสถิติต่อธุรกิจ

1. การวิเคราะห์ทางการตลาด: สถิติช่วยในการวิเคราะห์ข้อมูลการตลาด เช่น การศึกษาความต้องการของลูกค้า, การวิเคราะห์ความสำเร็จของแคมเปญโฆษณา, และการวิเคราะห์ผลของกิจกรรมการตลาดต่างๆ

2. การบริหารจัดการธุรกิจ: สถิติช่วยในการวิเคราะห์ข้อมูลทางธุรกิจ เช่น การวิเคราะห์รายได้และกำไร, การประเมินประสิทธิภาพของการผลิต และการประเมินความเสี่ยงทางธุรกิจ



ความสำคัญของของสถิติต่อธุรกิจ

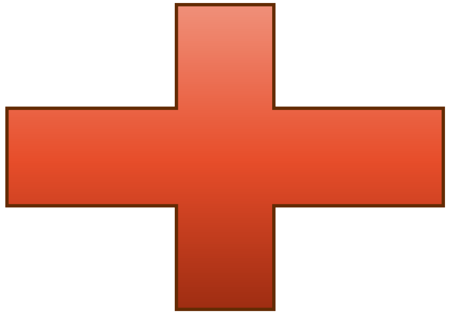
3. การตัดสินใจเชิงกลยุทธ์: สถิติช่วยในการตัดสินใจทางกลยุทธ์ เช่น การวิเคราะห์ข้อมูลเพื่อการวางแผนกลยุทธ์การตลาด, การพัฒนาผลิตภัณฑ์, และการเตรียมการต่างๆ สำหรับอุตสาหกรรม

4. การควบคุมคุณภาพ: สถิติช่วยในการวิเคราะห์และประเมินคุณภาพของผลิตภัณฑ์หรือบริการ เพื่อให้มั่นใจได้ว่ามีคุณภาพมาตรฐานและตอบสนองต่อความต้องการของลูกค้า

5. การประเมินผล: สถิติช่วยในการประเมินผลและการวิเคราะห์ข้อมูลเพื่อการปรับปรุง การใช้สถิติในการทำแบบสำรวจหรือการวิเคราะห์ข้อมูลจากลูกค้าช่วยให้ธุรกิจสามารถปรับปรุงสินค้าหรือบริการของตนได้อย่างต่อเนื่อง

ตัวอย่าง: การตัดสินใจเพื่อการวางแผนการผลิต

1. การคาดการณ์ปริมาณการขายในอนาคต
2. การประมาณค่าความเสี่ยง



ตัวอย่าง: การวิเคราะห์ข้อมูลการตลาดเพื่อวางแผนการโปรโมชั่น

1. การวิเคราะห์ข้อมูลลูกค้า

2. การประมาณค่าความน่าสนใจของโปรโมชั่น

3. การทดลองสองกลุ่ม (A/B Testing) การใช้สถิติในการทดลองสองกลุ่มเพื่อทดสอบประสิทธิภาพของแต่ละโปรโมชั่น และเลือกโปรโมชั่นที่มีผลสูงสุด

สรุป

ในบทนี้ได้ศึกษาถึงความสำคัญของสถิติในการวิเคราะห์ข้อมูลและการตัดสินใจในธุรกิจ โดยได้สรุปถึงวิธีการใช้สถิติในหลายด้านของธุรกิจ เช่น การวิเคราะห์ทางการตลาดเพื่อวางแผนการโปรโมชั่น, การตัดสินใจเชิงกลยุทธ์ในการบริหารจัดการธุรกิจ, การควบคุมคุณภาพผลิตภัณฑ์หรือบริการ, และการวางแผนการผลิต การใช้สถิติในธุรกิจช่วยให้ผู้บริหารและนักวิเคราะห์สามารถตัดสินใจอย่างมั่นใจและมีประสิทธิภาพในการจัดการและพัฒนาธุรกิจของตนอย่างเป็นระบบ

บทที่ 2 การเก็บรวบรวมข้อมูล

- การเลือกและการสำรวจข้อมูล
- การออกแบบแบบสำรวจ
- การกระจายและการจัดเก็บข้อมูล
- สรุป

วัตถุประสงค์การเรียนรู้

1 เรียนรู้ถึงการเลือกและการสำรวจข้อมูล

2 เข้าใจวิธีการรวบรวมข้อมูล

3 การกระจายและการจัดเก็บข้อมูล

4 วิเคราะห์การการกระจายของข้อมูลได้

บทนำ

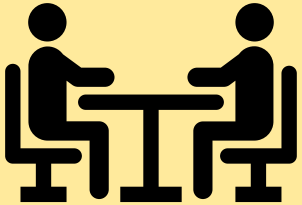
การเก็บรวบรวมข้อมูลเป็นกระบวนการที่มีความสำคัญอย่างมากในการทำงานทางธุรกิจ เนื่องจากมีบทบาทสำคัญในการเตรียมข้อมูลที่เป็นพื้นฐานสำหรับการวิเคราะห์และการตัดสินใจในระดับต่างๆ การเลือกและการสำรวจข้อมูลเป็นขั้นตอนแรกที่สำคัญในกระบวนการเก็บข้อมูล เพื่อให้ได้ข้อมูลที่เป็นประโยชน์และเชื่อถือได้

การเลือกและการสำรวจข้อมูล

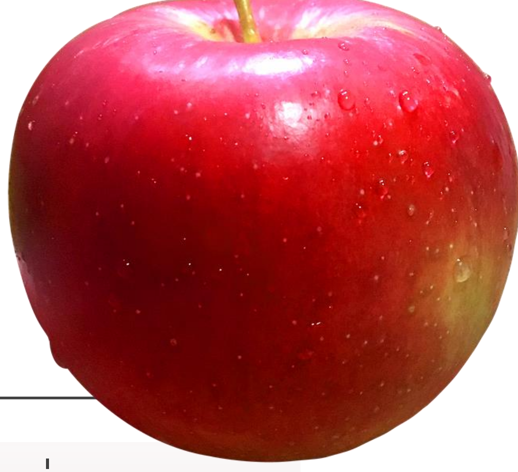
การเลือกและการสำรวจข้อมูลเป็นขั้นตอนที่สำคัญในการเก็บรวบรวมข้อมูลที่มีคุณภาพและมีประสิทธิภาพสำหรับการวิเคราะห์และการใช้ประโยชน์จากข้อมูลในองค์กร การเลือกข้อมูลเริ่มต้นจากการระบุวัตถุประสงค์และคำถามวิจัยที่ต้องการตอบ

เพื่อให้สามารถเลือกเก็บข้อมูลที่เกี่ยวข้องและมีประโยชน์ตามความต้องการ เช่น หากต้องการศึกษาความพึงพอใจของลูกค้าในผลิตภัณฑ์ ก็จะต้องระบุว่าต้องการทราบถึงปัจจัยที่มีผลต่อความพึงพอใจ และเลือกเก็บข้อมูลที่เกี่ยวข้องเช่น ความพึงพอใจในคุณภาพสินค้า การบริการหลังการขาย หรือราคาของสินค้า

การสำรวจข้อมูลเป็นขั้นตอนที่ใช้เครื่องมือและวิธีการต่างๆ เพื่อเก็บข้อมูลจากแหล่งต่างๆ อย่างเชื่อถือได้ เช่น การใช้แบบสอบถาม การสัมภาษณ์ หรือการส่งอีเมล์สอบถาม เพื่อรวบรวมข้อมูลจากกลุ่มเป้าหมาย



วิธีการเลือกข้อมูล



1. ระบุวัตถุประสงค์และคำถามวิจัย: กำหนดวัตถุประสงค์ของการเก็บข้อมูลและคำถามวิจัยที่ต้องการตอบ เพื่อให้เก็บข้อมูลที่เป็นประโยชน์และเกี่ยวข้องกับความต้องการของการวิเคราะห์หรือการตัดสินใจที่จะทำในธุรกิจ

2. กำหนดประเภทข้อมูลที่ต้องการ: ระบุประเภทข้อมูลที่ต้องการเก็บ เช่น ข้อมูลปริมาณ, ข้อมูลคุณภาพ, ข้อมูลสภาพเศรษฐกิจ, หรือข้อมูลทางสังคม

วิธีการเลือกข้อมูล

3. เลือกแหล่งข้อมูลที่เหมาะสม: ค้นหาแหล่งข้อมูลที่เหมาะสมและเชื่อถือได้ เช่น ฐานข้อมูลออนไลน์, ข้อมูลจากหน่วยงานราชการ, หรือการสำรวจข้อมูลจากผู้ใช้บริการ

4. วางแผนกระบวนการเก็บข้อมูล: วางแผนกระบวนการเก็บข้อมูลอย่างรอบคอบโดยคำนึงถึงวิธีการเก็บข้อมูล เช่น การสำรวจข้อมูลโดยใช้แบบสอบถาม, การสร้างฐานข้อมูลออนไลน์, หรือการนำเข้าข้อมูลจากแหล่งอื่น

5. ทดสอบและปรับปรุงกระบวนการ: ทดสอบและปรับปรุงกระบวนการเก็บข้อมูลเพื่อให้ได้ข้อมูลที่มีคุณภาพและเชื่อถือได้ โดยอาจต้องทดสอบการสำรวจข้อมูลกับกลุ่มเป้าหมายก่อนการนำไปใช้จริง

วิธีการรวบรวมข้อมูล

1. การสำรวจข้อมูล: การสำรวจข้อมูลโดยตรงจากแหล่งที่มา เช่น การสำรวจความคิดเห็นของลูกค้าหรือการสำรวจข้อมูลในงานวิจัย

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

3. การสืบค้นข้อมูล: การค้นหาข้อมูลจากแหล่งต่างๆ เช่น ฐานข้อมูลออนไลน์, เว็บไซต์, หรือห้องสมุด

วิธีการรวบรวมข้อมูล

4. การใช้ชุดข้อมูลอื่น: การใช้ข้อมูลที่มีอยู่แล้วจากแหล่งอื่น เช่น ข้อมูลทางการศึกษาจากกระทรวงศึกษาธิการหรือข้อมูลทางการแพทย์จากหน่วยงานที่เกี่ยวข้อง

5. การสำรวจตามสัญญาณ: การรวบรวมข้อมูลจากสัญญาณหรือเครื่องมือตรวจวัด เช่น การรวบรวมข้อมูลจากเซ็นเซอร์หรือเครื่องวัดอุณหภูมิ

6. การสำรวจการกระทำ: การรวบรวมข้อมูลโดยตรงจากการกระทำของบุคคลหรือองค์กร เช่น การเก็บข้อมูลการจ่ายเงินจากธนาคารหรือการเก็บข้อมูลการใช้บริการจากหน่วยงานต่างๆ

ประชากรและกลุ่มตัวอย่าง

"ประชากร" (population) หมายถึง กลุ่มทั้งหมดของบุคคลหรือองค์กรที่ต้องการศึกษาหรือสำรวจข้อมูล เป็นกลุ่มที่มีลักษณะที่ต้องการให้ข้อมูลและทำการวิจัย

"กลุ่มตัวอย่าง" (sample) คือ ส่วนหนึ่งของประชากรที่ถูกเลือกมาเพื่อทำการวิจัยหรือสำรวจข้อมูล มักจะเลือกมาเพียงบางส่วนเท่านั้นเนื่องจากการสำรวจทั้งหมดของประชากรอาจไม่เป็นไปตามความเป็นจริง การเลือกกลุ่มตัวอย่างที่เป็น Representative of the population จะช่วยให้สามารถทำข้อสรุปหรือนำเสนอผลลัพธ์ที่สามารถนำไปใช้งานได้ในประชากรทั้งหมดได้ถูกต้องและเชื่อถือได้ มักจะใช้**วิธีการสุ่มตัวอย่าง**ออกมาจากประชากร

การเลือกตัวอย่างแบบใช้ความน่าจะเป็น

1. การสุ่มอย่างง่าย (Simple Random Sampling)
2. การสุ่มแบบแบ่งชั้น (Stratified Random Sampling)
3. การสุ่มแบบเป็นระบบ (Systematic Sampling)
4. การสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling)

การเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็น

1. การสุ่มตัวอย่างแบบสะดวก (Convenience Sampling)
2. การสุ่มตัวอย่างแบบเจาะจง (Purposive Sampling)
3. การสุ่มตัวอย่างแบบตามโควต้า (Quota Sampling)
4. การสุ่มตัวอย่างแบบอ้างอิงด้วยบุคคลและผู้เชี่ยวชาญ (Snowball Sampling)

การออกแบบแบบสำรวจ

1. กำหนดวัตถุประสงค์ของการสำรวจ: กำหนดวัตถุประสงค์หรือคำถามวิจัยที่ต้องการตอบหรือข้อมูลที่ต้องการเก็บ เพื่อช่วยให้การสำรวจมีความมุ่งหวังและเหมาะสมต่อเนื้อหาที่ต้องการตรวจสอบ

2. เลือกวิธีการสำรวจ: เลือกวิธีการสำรวจที่เหมาะสมกับวัตถุประสงค์และลักษณะของประชากร เช่น การสำรวจด้วยแบบสอบถาม, การสัมภาษณ์, การสังเกตการณ์

3. ออกแบบคำถาม: ออกแบบคำถามให้มีความชัดเจนและเข้าใจได้ง่าย ไม่มีคำถามที่สับสนหรือกำหนดมากเกินไป เพื่อป้องกันการเข้าใจผิดและข้อมูลที่ไม่ถูกต้อง

การออกแบบแบบสำรวจ

4. วางโครงสร้างและลำดับของคำถาม: การวางโครงสร้างและลำดับของคำถามให้มีระเบียบ จากคำถามที่ง่ายไปสู่คำถามที่ยาก หรือจากคำถามที่ไม่เป็นส่วนต่อเนื่องไปสู่คำถามที่เป็นส่วนต่อเนื่อง เพื่อเพิ่มประสิทธิภาพในการทำแบบสำรวจ

5. ทดสอบและปรับปรุงแบบสำรวจ: ทดสอบแบบสำรวจกับกลุ่มเล็กของผู้ร่วมสำรวจเพื่อให้แน่ใจว่าคำถามเข้าใจได้ และไม่มีข้อผิดพลาด จากนั้นปรับปรุงและปรับแก้ตามความเหมาะสม

6. การจัดการและการสำรวจ: จัดทำและดำเนินการสำรวจตามแบบทดสอบที่ออกแบบไว้ ให้เก็บข้อมูลที่สะสมได้มาอย่างถูกต้องและครบถ้วน

การออกแบบแบบสำรวจ

7. การวิเคราะห์ข้อมูล: วิเคราะห์ข้อมูลที่ได้รับจากการสำรวจอย่างถูกต้อง และสรุปผลลัพธ์ที่ได้ให้เป็นไปตามวัตถุประสงค์ของการวิจัยหรือการสำรวจข้อมูล

การกระจายและการจัดเก็บข้อมูล

การได้เรียนรู้เกี่ยวกับการกระจายของข้อมูล เช่น การใช้ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน พิสัย และการแจกแจงของข้อมูลต่างๆ ซึ่งเป็นเครื่องมือที่ใช้ในการวิเคราะห์และอธิบายข้อมูลในรูปแบบที่ชัดเจน

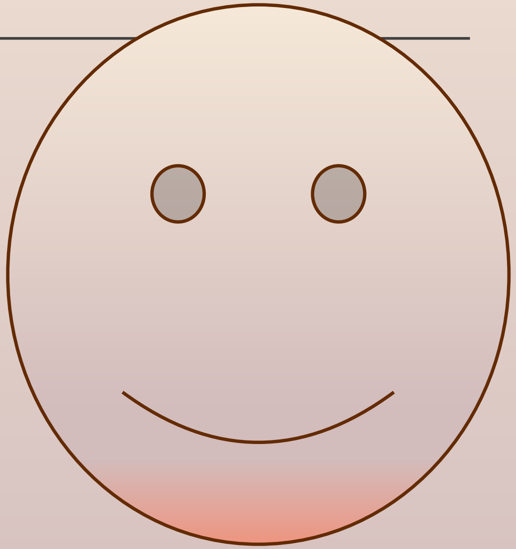
สถิติเชิงบรรยาย (Descriptive Statistics)

สถิติเชิงบรรยาย (Descriptive Statistics) เป็นเครื่องมือสำคัญในการรวมและอธิบายข้อมูลที่มีอยู่ในรูปแบบที่เข้าใจง่าย โดยมักใช้สำหรับการสรุปข้อมูลและการเข้าใจลักษณะเฉพาะของข้อมูล

ค่ากลาง (Measures of Central Tendency):

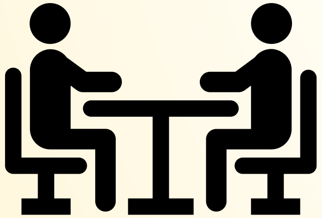
Mean (ค่าเฉลี่ย):

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



ค่ากลาง (Measures of Central Tendency):

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



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

ค่ากลาง (Measures of Central Tendency):

Mode (ฐานนิยม) คือค่าที่ปรากฏบ่อยที่สุดในชุดข้อมูล

สมมติว่ามีชุดข้อมูลดังนี้:1,2,2,3,3,3,4,5,5

ในกรณีนี้:ค่า 3 ปรากฏบ่อยที่สุด (3 ครั้ง)

ดังนั้น Mode ของชุดข้อมูลนี้คือ 3

การกระจาย (Measures of Dispersion):

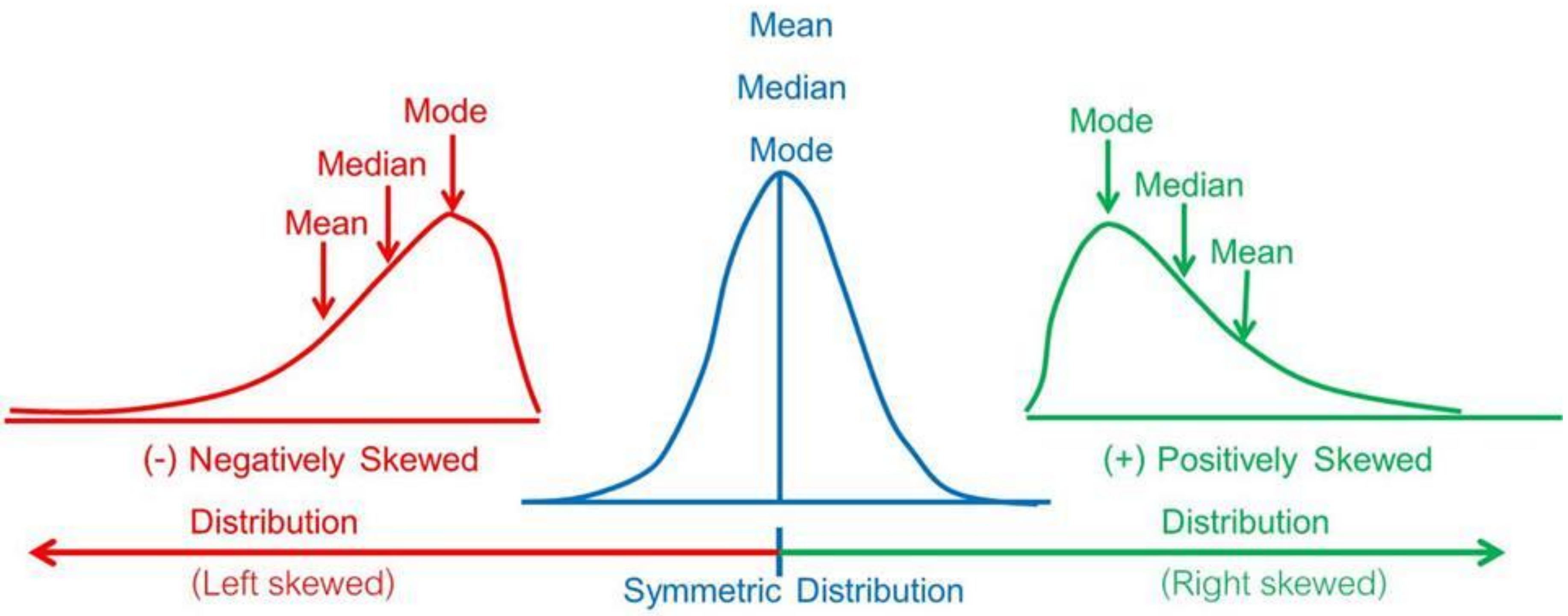
- Range (ช่วงของข้อมูล): Range=Max(x)−Min(x)
- Variance (ความแปรปรวน): $$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$
- Standard Deviation (ส่วนเบี่ยงเบนมาตรฐาน): $s = \sqrt{s^2}$

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

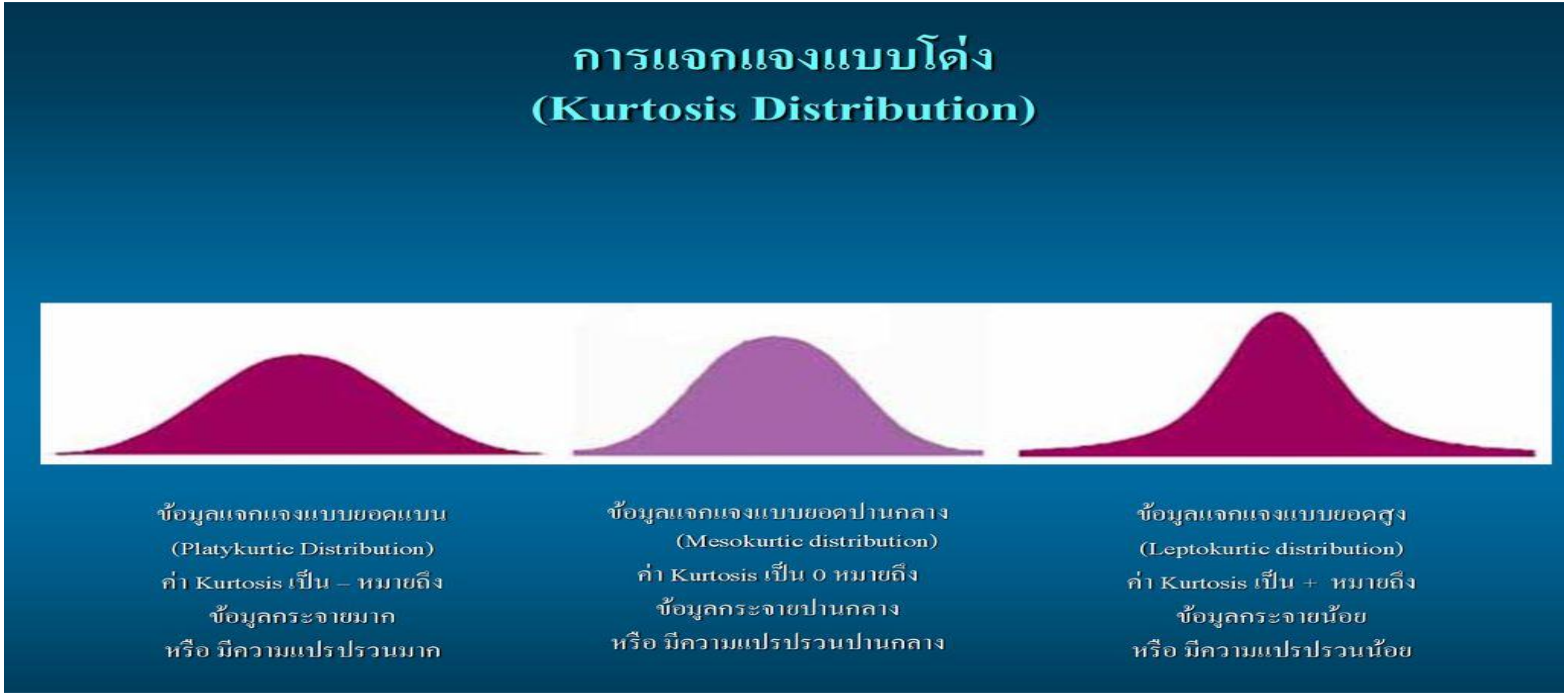
การกระจายรูปแบบ (Shape of Distribution)

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

ความเบ้ (Skewness)



ความโด่ง (Kurtosis):



การกระจายเชิงตัวแปร (Measures of Variability)

- Percentile (เปอร์เซนไทล์)
  - เดไซล์ (Decile)
  - ควอร์ไทล์ (Quantile)

Percentile (เปอร์เซนไทล์)

เปอร์เซนไทล์ (Percentile) เป็นค่าที่ใช้ในการแบ่งข้อมูลออกเป็นช่วงๆ โดยใช้เปอร์เซนไทล์เป็นตัวบ่งชี้ เช่น ค่าเปอร์เซนไทล์ที่ 25 (หรือ P25) คือค่าที่แบ่งข้อมูลออกเป็นส่วนของ 25% ที่มีค่าน้อยกว่าหรือเท่ากับค่านั้น

การคำนวณเปอร์เซนไทล์:จัดเรียงข้อมูลจากน้อยไปมากกำหนดลำดับของเปอร์เซนไทล์ที่ต้องการ เช่น 25th percentile (เปอร์เซนไทล์ที่ 25)คำนวณตำแหน่งเปอร์เซนไทล์โดยใช้สูตร

$$\left(\frac{p}{100}\right) \times (N+1)$$

โดย p คือค่าของเปอร์เซนไทล์ที่ต้องการ และ N คือจำนวนข้อมูลทั้งหมดในชุด

Percentile (เปอร์เซนไทล์)

สมมติว่ามีชุดข้อมูลดังนี้:3,1,4,1,5,9,2,6,5,3

จัดเรียงข้อมูล:1,1,2,3,3,4,5,5,6,9

ต้องการหา 25 th percentile (เปอร์เซนไทล์ที่ 25) คำนวณตำแหน่ง

$$\left(\frac{25}{100}\right) \times (10+1) = 2.75$$

ตำแหน่งที่ 2.75 คือการอินเตอร์โพลเลชันระหว่างตำแหน่งที่ 2 และ 3:

ค่าเปอร์เซนไทล์ที่ 25 =1+0. 75×(2−1)=1.75 ดังนั้น ค่าเปอร์เซนไทล์ที่ 25 คือ 1.75

จัดเรียงข้อมูล:1,1,2,3,3,4,5,5,6,9

ตำแหน่งที่ 2.75 คือการอินเตอร์โพลเลชันระหว่างตำแหน่งที่ 2 และ 3

ในชุดข้อมูลที่จัดเรียงแล้ว:ตำแหน่งที่ 2 มีค่าเป็น 1ตำแหน่งที่ 3 มีค่าเป็น 2

การอินเตอร์โพลเลชัน:ค่าเปอร์เซนไทล์ที่ 25 = ค่าตำแหน่งที่ 2 +0.75 ×(ค่าตำแหน่งที่ 3−ค่าตำแหน่งที่ 2) ค่าเปอร์เซ็นไทล์ที่ 25 = 1+ 0.75 x (2−1)

ค่าเปอร์เซ็นไทล์ที่ 25 = 1+0.75

ค่าเปอร์เซ็นไทล์ที่ 25 = 1.75

ค่าเปอร์เซนไทล์ที่ 25 = 1.75 หมายความว่า 25% ของข้อมูลทั้งหมดในชุดข้อมูลมีค่าที่น้อยกว่าหรือเท่ากับ 1.75 และ 75% ของข้อมูลมีค่ามากกว่า 1.75

ในบริบทที่ใช้ค่าเปอร์เซนไทล์ มักจะใช้ในการเปรียบเทียบตำแหน่งของข้อมูลในชุดข้อมูลหนึ่งๆ เช่น:

- ถ้าคุณมีคะแนนสอบและพบว่าคะแนนของคุณอยู่ที่เปอร์เซนไทล์ที่ 25 นั่นหมายความว่าคะแนนของคุณสูงกว่าหรือเท่ากับ 25% ของคนที่สอบทั้งหมด และต่ำกว่า 75% ของคนที่สอบทั้งหมด
- ในการศึกษาการเจริญเติบโตของเด็ก ถ้าส่วนสูงของเด็กอยู่ที่เปอร์เซนไทล์ที่ 25 นั่นหมายความว่า 25% ของเด็กที่มีอายุเท่ากันมีส่วนสูงที่น้อยกว่าหรือเท่ากับเด็กคนนี้ และ 75% ของเด็กที่มีอายุเท่ากันมีส่วนสูงมากกว่าเด็กคนนี้
- ดังนั้นค่าเปอร์เซนไทล์ที่ 25 = 1.75 แสดงให้เห็นว่าข้อมูลในชุดข้อมูลที่จัดเรียงแล้ว มี 25% ที่มีค่าไม่เกิน 1.75 และ 75% ของข้อมูลมีค่ามากกว่า 1.75.

เดไซล์ (Decile)

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

เดไซล์ (Decile)

เดไซล์ที่สำคัญประกอบด้วย:

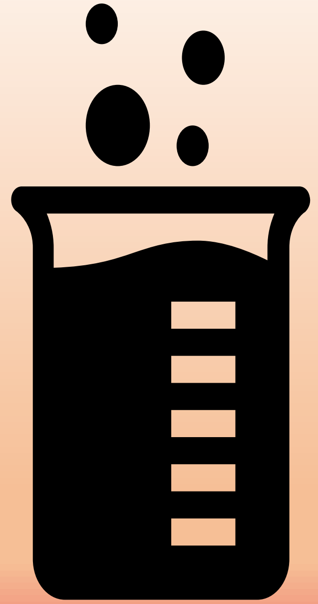
เดไซล์ที่ 1 (D1) : แบ่งชุดข้อมูลที่ตำแหน่ง 10%

เดไซล์ที่ 2 (D2) : แบ่งชุดข้อมูลที่ตำแหน่ง 20%

เดไซล์ที่ 3 (D3) : แบ่งชุดข้อมูลที่ตำแหน่ง 30%

...

เดไซล์ที่ 9 (D9) : แบ่งชุดข้อมูลที่ตำแหน่ง 90%



วิธีการคำนวณเดไซล์

จัดเรียงข้อมูลจากน้อยไปมาก

ใช้สูตรคำนวณตำแหน่งเดไซล์

$$k = \left(\frac{k}{10}\right) \times (N + 1)$$

โดย k คือเลขลำดับของเดไซล์ (1 ถึง 9) และ N คือจำนวนข้อมูลทั้งหมด

วิธีการคำนวณเดไซล์

สมมติว่ามีชุดข้อมูลดังนี้: 3,1,4,1,5,9,2,6,5,3

จัดเรียงข้อมูล:1,1,2,3,3,4,5,5,6,9

คำนวณตำแหน่ง:ตำแหน่งเดไซล์ที่ 5 = $\left(\frac{5}{10}\right) \times (10+1) = 5.5$

ตำแหน่งที่ 5.5 คือการอินเตอร์โพลเลชันระหว่างตำแหน่งที่ 5 และ 6

ค่าเดไซล์ที่ 5 = 3 +0.5 ×(4−3) = 3 +0.5 = 3.5

ดังนั้น ค่าเดไซล์ที่ 5 คือ 3.5

ค่าเดไซล์ที่ 5 = 3.5 หมายความว่า 50% ของข้อมูลทั้งหมดในชุดข้อมูลมีค่าที่น้อยกว่าหรือเท่ากับ 3.5 และ 50% ของข้อมูลมีค่ามากกว่า 3.5

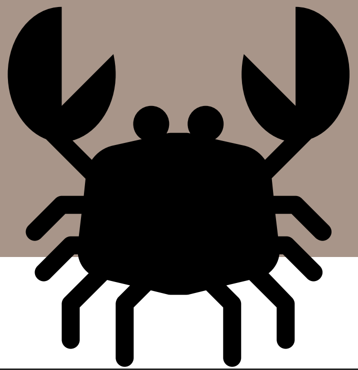
- หากคุณกำลังวิเคราะห์รายได้ของกลุ่มคนในชุมชนหนึ่งและพบว่ารายได้อยู่ที่เดไซล์ที่ 5 = 3.5 หมายความว่า 50% ของคนในชุมชนนั้นมีรายได้ต่ำกว่าหรือเท่ากับ 3.5 หน่วยเงิน และอีก 50% มีรายได้มากกว่า 3.5 หน่วยเงิน
- ในการศึกษาการเจริญเติบโตของเด็ก ถ้าความสูงของเด็กอยู่ที่เดไซล์ที่ 5 = 3.5 หมายความว่า 50% ของเด็กที่มีอายุเท่ากันมีความสูงน้อยกว่าหรือเท่ากับ 3.5 หน่วย (เช่น เซนติเมตร) และอีก 50% ของเด็กมีความสูงมากกว่า 3.5 หน่วย

ควอร์ไทล์ (Quantile)

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

การสรุปข้อมูล (Data Summary)

- Frequency Distribution (การกระจายความถี่): การแบ่งข้อมูลออกเป็นช่วงๆ และนับจำนวนข้อมูลที่อยู่ในแต่ละช่วง
- Histogram (ฮิสโตแกรม): กราฟแสดงการกระจายความถี่ของข้อมูล



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 ซึ่งเป็นกระบวนการที่เกี่ยวกับการกำหนดโครงสร้างและขั้นตอนที่เหมาะสมในการสำรวจข้อมูลเพื่อให้ได้ข้อมูลที่ถูกต้องและน่าเชื่อถือ รวมถึงการกระจายและการจัดเก็บข้อมูล

แบบฝึกหัด

ทำการบ้านเกี่ยวกับ ควอร์ไทล์ (Quantile) โดยเลือกชุดข้อมูลและตำแหน่งเอง ให้แสดงวิธีการหาตำแหน่งและค่าควอร์ไทล์ (Quantile) อย่างละเอียด ส่งในคลาสรูม ส่งก่อนเรียนครั้งหน้า

บทที่ 3 การแสดงข้อมูล

- การนำเสนอข้อมูลในรูปแบบต่างๆ
- กราฟและแผนภูมิที่เหมาะสม
- การสร้างตาราง
- การนำเสนอข้อมูลที่เข้าใจง่าย
- สรุป

วัตถุประสงค์การเรียนรู้

- นักศึกษาสามารถเรียนรู้การนำเสนอข้อมูลในรูปแบบต่างๆ
- นักศึกษาเข้าใจกราฟและแผนภูมิที่เหมาะสม
- นักศึกษาสามารถสร้างตารางข้อมูลที่เข้าใจง่าย

บทนำ

ในบทนี้เป็นการแสดงข้อมูลเป็นขั้นตอนสำคัญในการสื่อสารข้อมูลและการนำเสนอผลลัพธ์จากการวิเคราะห์ข้อมูลให้ผู้อื่นเข้าใจได้ง่าย โดยมาตรฐานของการแสดงข้อมูลที่ดีรวมถึงการใช้มาตรวัดของข้อมูลเพื่อสรุปข้อมูลตัวเลขอย่างถูกต้อง การใช้กราฟและแผนภูมิ เช่น กราฟแท่ง เส้น หรือวงกลม เป็นวิธีที่ดีในการแสดงแนวโน้มและความสัมพันธ์ของข้อมูล

มาตรวัดของข้อมูล

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

การวัดระดับนามมาตรา (Nominal Scale)

เป็นระดับต่ำสุดของตัวแปร ซึ่งจัดแบ่งเป็นประเภทหรือกลุ่มโดยระบุความ แตกต่างเป็นชื่อหรือสัญลักษณ์ เท่านั้น มิได้บ่งชี้ถึงความแตกต่างในคุณค่า หรือคุณภาพใดๆ ไม่สามารถจัดเรียงลำดับก่อนหลัง หรือสูงต่ำได้ แต่ละหน่วยมีความเป็นอิสระจากกันอย่าง เด็ดขาด อาจใช้ตัวเลขแทนได้ (เพศชาย คือ 1 เพศหญิง คือ 2)

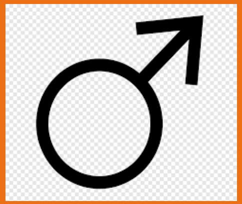
ตัวอย่าง

เพศ : ชาย หญิง

อาชีพ : ข้าราชการ รัฐวิสาหกิจ รับจ้าง อาชีพส่วนตัว เกษตรกร อื่นๆ

แบ่งเป็นกลุ่ม

ชาย



หญิง

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 ไม่เห็นด้วย อย่างยิ่ง

เครื่องหมายยศทหารบก
พลเอก
พลโท
พลตรี
พันเอก
พันโท
พันตรี
ร้อยเอก
ร้อยโท
ร้อยตรี
จ่าสิบเอก
จ่าสิบโท
จ่าสิบตรี
สิบเอก
สิบโท
สิบตรี
#เรารักกองทัพบก

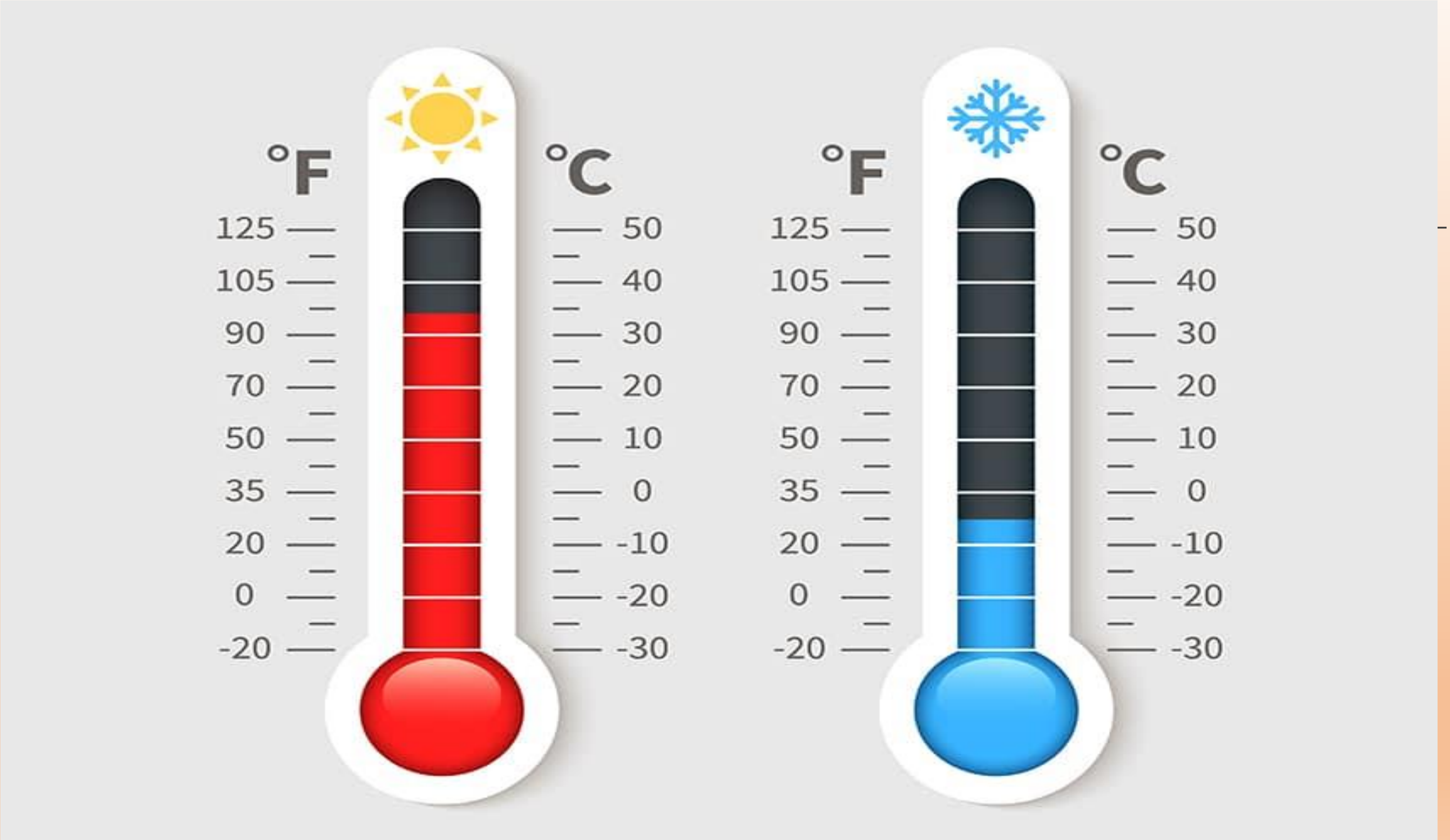
การวัดระดับช่วงมาตรา (Interval Scale)

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

ตัวอย่าง

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

อุณหภูมิ : 0oC , 50oC (อุณหภูมิ 0oC ไม่ได้หมายความว่าไม่มีความเย็นเลย)


°F
°C
125
105
90
70
50
35
20
0
-20
50
40
30
20
10
0
-10
-20
-30
°F
°C
125
105
90
70
50
35
20
0
-20
50
40
30
20
10
0
-10
-20
-30

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


1
2
3
13
14
15
4
5
6
16
17
18
7
8
9
19
20
21
10
11
12
22
23
24

การนำเสนอข้อมูลในรูปแบบต่างๆ

1. การนำเสนอข้อมูลอย่างไม่เป็นแบบแผน (Informal Presentation)

2.การนำเสนอข้อมูลอย่างเป็นแบบแผน (Formal Presentation)

การนำเสนอข้อมูลอย่างไม่เป็นแบบแผน

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

ตัวอย่างที่3.2 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

การนำเสนอข้อมูลอย่างเป็นแบบแผน

2.1 การนำเสนอข้อมูลในรูปตาราง

2.2 การนำเสนอข้อมูลในรูปแผนภูมิและแผนภาพ

2.3 การนำเสนอข้อมูลในรูปกราฟเส้น

2.1 การนำเสนอข้อมูลในรูปตาราง

หมายเลขตาราง.. ชื่อเรื่อง...

หัวข้อ	หัวเรื่อง		
	สดมภ์	สดมภ์	สดมภ์
ต้นขั้ว	ตัวเรื่อง	ตัวเรื่อง	ตัวเรื่อง

หมายเหตุ (ถ้ามี) ....

แหล่งที่มา ....

การนำเสนอข้อมูลในรูปตาราง

การนำเสนอข้อมูลในรูปของตาราง สามารถจำแนกลักษณะของตารางออกเป็น 4 ชนิด คือ

ก. ตารางแสดงความถี่ (Frequency table)

- ข. ตารางทางเดียว (One – way table)
- ค. ตารางสองทาง (Two - way table)
- ง. ตารางหลายทาง (Multi – way table)

2.2 การนำเสนอข้อมูลในรูปแผนภูมิและแผนภาพ

ก. แผนภูมิแท่ง

1. แผนภูมิแท่งเชิงเดียว
2. แผนภูมิแท่งเชิงซ้อน
3. แผนภูมิแท่งส่วนประกอบ
4. แผนภูมิแท่งซ้อนกัน
5. แผนภูมิแท่งบวก – ลบ

ข. แผนภูมิกง

ค. แผนภูมิรูปภาพ

2.3 การนำเสนอข้อมูลในรูปกราฟเส้น

1. กราฟเส้นเชิงเดียว
2. กราฟเส้นเชิงซ้อน
3. กราฟเส้นเชิงประกอบ
4. กราฟดุล

การแจกแจงความถี่

ข้อมูลที่ได้จากการเก็บรวบรวมวิธีต่างๆ อาจจะอยู่ในลักษณะที่ไม่เป็นระเบียบ จึงจำเป็นต้องจัด ข้อมูลดังกล่าวให้เป็นระเบียบและเป็นหมวดหมู่เพื่อให้ง่ายต่อการคำนวณและสะดวกต่อการวิเคราะห์ ข้อมูลขั้นต่อไป เรียกวิธีทางสถิติเช่นนี้ว่า การแจกแจงความถี่ของข้อมูล

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

ตัวอย่างที่ 3.4 โดยจัดข้อมูลเป็นหมวดหมู่

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

การแจกแจงความถี่สัมพัทธ์

ตัวอย่างที่ 3.7 จากตารางแจกแจงความถี่ที่กำหนดให้ จงหาความถี่สัมพัทธ์และร้อยละความถี่สัมพัทธ์

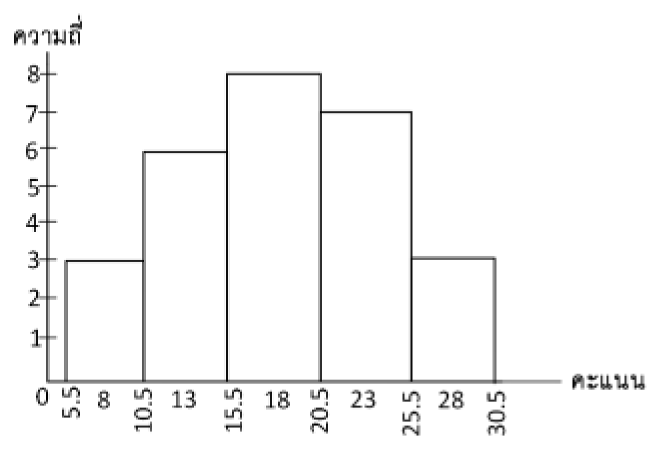
อันตรภาคชั้น	ความถี่	ความถี่สัมพัทธ์	ร้อยละความถี่สัมพัทธ์
55 – 59	2	= 0.05	5
60 – 64	14	= 0.35	35
65 – 69	8	= 0.20	20
70 – 74	12	= 0.30	30
75 - 79	4	= 0.10	10
รวม	40	1.0	100

การแจกแจงความถี่โดยใช้กราฟ

กราฟที่แสดงการแจกแจงความถี่ มี 3 แบบ คือ

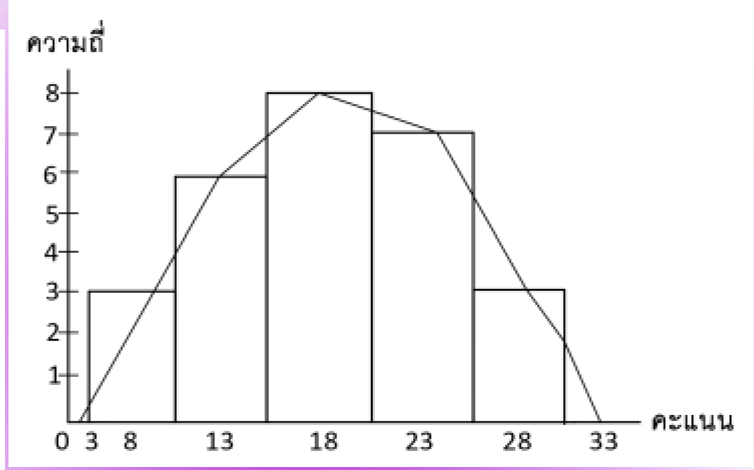
1. ฮิสโตแกรม (Histogram)
2. รูปหลายเหลี่ยมความถี่ (Frequency Polygon)
3. เส้นโค้งความถี่ (Frequency Curve)

การแจกแจงความถี่โดยใช้กราฟ



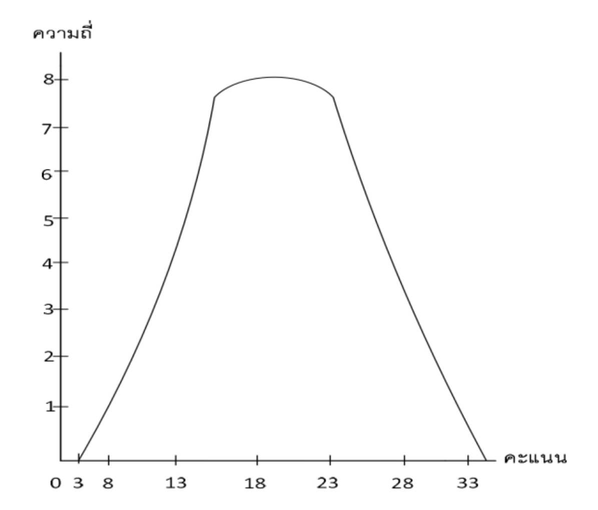
ฮิสโตแกรม

(Histogram)



รูปหลายเหลี่ยมความถี่

(Frequency Polygon)



เส้นโค้งความถี่

(Frequency Curve)

การนำเสนอข้อมูลที่เข้าใจง่าย

การนำเสนอข้อมูลที่เข้าใจง่ายเป็นการแสดงผลข้อมูลอย่างชัดเจนและเข้าใจได้ง่ายโดยไม่ซับซ้อน นี่คือบางวิธีในการนำเสนอข้อมูลที่เข้าใจง่าย:

1. ใช้กราฟและแผนภูมิที่เหมาะสม: การใช้กราฟและแผนภูมิที่เหมาะสม เช่น กราฟแท่ง กราฟเส้น และแผนภูมิวงกลม เพื่อแสดงข้อมูลในรูปแบบที่เข้าใจง่ายและชัดเจน

2. ใช้สีและขนาดต่างๆ: การใช้สีและขนาดต่างๆ ในการแสดงผลข้อมูล เพื่อเน้นความสำคัญของข้อมูลหรือสร้างการเน้นให้ข้อมูลมีความโดดเด่น

3. ใช้ตารางเรียบร้อย: การจัดเรียงข้อมูลในรูปแบบตารางที่เรียบร้อย และสะดวกต่อการอ่านและเข้าใจ

การนำเสนอข้อมูลที่เข้าใจง่าย

4. การใช้ภาษาที่เข้าใจได้: การใช้ภาษาที่เข้าใจง่าย ไม่มีคำศัพท์ทางวิชาการมากมาย และอธิบายข้อมูลอย่างชัดเจน

5. สรุปข้อมูลสำคัญ: การสรุปข้อมูลสำคัญโดยกระชับและชัดเจน เพื่อให้ผู้อ่านเข้าใจสาระสำคัญของข้อมูลได้โดยรวดเร็ว

สรุป

ในบทนี้ได้เรียนรู้เกี่ยวกับการแสดงข้อมูลอย่างมีประสิทธิภาพในธุรกิจ โดยมีหลักการสำคัญทั้ง 3 ด้านคือ การใช้กราฟและแผนภูมิที่เหมาะสม เพื่อแสดงผลข้อมูลอย่างชัดเจนและเข้าใจได้ง่าย การสร้างตารางที่เรียบง่ายและสะดวกต่อการอ่านและเข้าใจข้อมูล และการนำเสนอข้อมูลให้เข้าใจได้ง่าย โดยใช้ภาษาที่ไม่ซับซ้อนและการสรุปข้อมูลอย่างชัดเจน เหล่านี้เป็นเครื่องมือที่สำคัญในการสร้างการเข้าใจและการตัดสินใจที่มีความสำคัญในองค์กรและธุรกิจต่างๆ ในการจัดทำรายงานหรือการบริหารจัดการทรัพยากรต่างๆ อย่างมีประสิทธิภาพ

บทที่ 4 การวิเคราะห์ข้อมูล

- ประเภทของสถิติที่ใช้ในการวิเคราะห์ข้อมูล
- การคำนวณค่าสถิติที่สำคัญ เช่น เฉลี่ย, ส่วนเบี่ยงเบนมาตรฐาน, ความถี่, ความสัมพันธ์
- การสร้างและการอ่านผลการวิเคราะห์
- สรุป

บทนำ

การวิเคราะห์ข้อมูลเป็นขั้นตอนที่สำคัญในการสร้างความเข้าใจและความรู้เกี่ยวกับข้อมูลที่มีอยู่ ในบทนี้จะได้เรียนรู้เกี่ยวกับการใช้สถิติใน การวิเคราะห์ข้อมูล เช่น การคำนวณค่าสถิติที่สำคัญ เช่น เฉลี่ย, ส่วนเบี่ยงเบนมาตรฐาน, ความถี่ และความสัมพันธ์

ประเภทของสถิติที่ใช้ในการวิเคราะห์ข้อมูล

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

สถิติบรรยาย (Descriptive Statistics)

การแจกแจงความถี่ อัตราส่วน สัดส่วน อัตราส่วน ควอร์ไทล์ เดไซล์ เปอร์เซ็นไทล์ การหาค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม พิสัย ส่วนเบี่ยงเบนมาตรฐาน

การแจกแจงความถี่

ตัวแปรจำนวน
1 ตัวแปร

การแจกแจงความถี่แบบทางเดียว

การแจกแจงความถี่ของลูกค้าชาวไทยที่มาเยี่ยมชมร้านค้าลาบูบู้ ที่จำแนกตามรายภาค

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

การแจกแจงความถี่แบบสองทาง

ตัวแปรจำนวน
2 ตัวแปร

จำนวนของนักศึกษาที่เข้าชมร้านค้า มอลลี่ จำแนกตามชั้นปีที่ศึกษาและเพศ

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 (Ratio)

เป็นการเปรียบเทียบความถี่ ระหว่างรายการย่อยกับรายการย่อย = ความถี่ของรายการย่อยที่1/(ความถี่ของ รายการย่อยที่ 2)

เช่น **อัตราส่วนของนักศึกษาชายและนักศึกษาหญิง** = จำนวนของนักศึกษาชาย:จำนวนของนักศึกษาหญิง = 100: 200 = 1:2

อัตราส่วนของเกรด = A : B : C : D : F = 20: 50: 75: 25: 5

สัดส่วน (Proportion)

เป็นการเปรียบเทียบความถี่ ระหว่างรายการย่อยกับจำนวนทั้งหมดของตัวแปร

= ความถี่ของรายการ/(ความถี่ทั้งหมด)

เช่น สัดส่วนของนักศึกษาหญิง = 350/700 = .50

ร้อยละ (percentage)

เป็นการเปรียบเทียบความถี่ของรายการย่อยกับความถี่ทั้งหมดและเทียบกับ100

ร้อยละ =(ความถี่ของรายการ)/(ความถี่ทั้งหมด) x 100

เช่น ร้อยละของจำนวนลูกค้าหญิง = 360/700 x 100 = 51%

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ประเภทของการวัดแนวโน้มเข้าสู่ส่วนกลาง

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

ค่าเฉลี่ยเลขคณิต

ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) หมายถึง การหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลทั้งหมดการหาค่าเฉลี่ยเลขคณิตสามารถหาได้ 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

$$\text{ค่าเฉลี่ยเลขคณิต} = \frac{10+25+30+42+16+14+18+17}{8} = 21.5$$

ดังนั้นค่าเฉลี่ย อายุลูกค้ากลุ่มนี้ = 21.50 ปี

มัธยฐาน (Median)

มัธยฐาน หมายถึง ค่ากึ่งกลางของข้อมูลชุดนั้น หรือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลชุดนั้น เมื่อได้จัดเรียงค่าของข้อมูลจากน้อยที่สุด ไปหามากที่สุดหรือจากมากที่สุดไปหาน้อยที่สุด ค่ากึ่งกลางจะเป็นตัวแทนที่แสดงว่ามีข้อมูลที่มากกว่าและน้อยกว่านี้อยู่ 50%

การหาค่ามัธยฐาน สามารถหาได้ดังนี้

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

มัธยฐาน (Median)

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

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

หาตำแหน่งมัธยฐาน = (9+1)/2 = 5

และนับเรียงจากข้อมูลตำแหน่งที่ 1 ไปถึงตำแหน่งที่ 5 มัธยฐานของข้อมูล = 12

ฐานนิยม (Mode)

ฐานนิยม หมายถึง ค่าของคะแนนที่ซ้ำกันมากที่สุดหรือค่าคะแนนที่มีความถี่สูงที่สุดในข้อมูลชุดนั้น

การหารค่าฐานนิยม 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

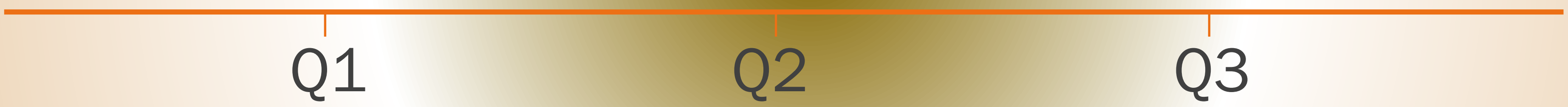
ข้อมูลที่ซ้ำกันมากที่สุดคือ 10 กับ 12

ฐานนิยม คือ 10 กับ 12

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

ควอร์ไทล์

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



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

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

จากสูตร ตำแหน่งควอร์ไทล์ที่ k = $\frac{Xk}{4} * (n+1)$

ควอร์ไทล์ที่ 3 = $\frac{3}{4} *(7+1) = 6$

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 $= \frac{1}{4} *(4+1) = 1.25$

ดังนั้น ควอร์ไทล์ที่ 1 คือ ข้อมูลตำแหน่งที่ 1.25 คืออยู่ระหว่าง ข้อมูลตำแหน่งที่ 1 และ 2

จึงมีวิธีหา คือ $x_{1.25} = x_1 + .25 * (x_2 - x_1)$

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

การหาค่า ควอร์ไทล์ ของข้อมูลแจกแจงความถี่

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

ตำแหน่งควอร์ไทล์ที่ k $= \frac{Xk}{4} * (n)$

ควอร์ไทล์ที่ k ของข้อมูลแจกแจงความถี่ จากสูตร

$$L + I * \frac{\frac{Xk(n)}{4} - F}{f}$$

L คือ ขอบล่างของชั้นที่มีควอร์ไทล์

I คือ ความกว้างของชั้นอัตรภาค

F คือ ความถี่สะสมของชั้นที่ต่ำกว่าชั้นควอร์ไทล์

f คือ ความถี่ของชั้นควอร์ไทล์

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

หาชั้นที่ตำแหน่งควอร์ไทล์ที่ 3 อยู่ คือ 41-50

$$40.5 + 10 * \frac{\frac{3(13)}{4} - 6}{4} = 49.875$$

เดไซล์

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

เปอร์เซ็นไทล์

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

การวัดการกระจาย

การวัดการกระจาย (Measures of dispersion) คือ การวัดการกระจายของข้อมูลที่ได้จากการรวบรวมข้อมูล โดยใช้เครื่องมือการวิจัย เช่น แบบสอบถาม แบบสำรวจ การวัดที่ได้จากการวัดแนวโน้มเข้าสู่ส่วนกลางทำให้ทราบคุณลักษณะของข้อมูลที่เป็นตัวแทนกลุ่มเพียงค่าเดียวเท่านั้นแต่เนื่องจากค่าที่เป็นตัวแทนกลุ่มค่าเดียวกันไม่สามารถทำให้ทราบลักษณะของข้อมูลได้เพียงพอ เช่น พัดลม

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

ส่วนเบี่ยงเบนมาตรฐานเป็นการวัดการกระจายที่นำข้อมูลทุกค่ามาคำนวณเป็นวิธีที่นักสถิตินิยมใช้มากที่สุดและเป็นวิธีที่ใช้ในการวัดการกระจายได้ดีที่สุดสามารถไปใช้ในสถิติขั้นสูงต่อไปส่วนเบี่ยงเบนมาตรฐานใช้สัญลักษณ์ S หรือ S.D.

ส่วนเบี่ยงเบนมาตรฐานของประชากร

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

ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

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

ตัวอย่างการหาค่าส่วนเบี่ยงเบนมาตรฐาน

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

หา $\bar{X} = 26$

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

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 (Inferential Statistics)

การทดสอบสมมติฐาน (Hypothesis Testing)

	True state of the world	
Decision	H_0 False	H_0 True
Reject H_0	Correct decision	Type I error
Retain H_0	Type II error	Correct decision

สถิติแบบพารามิเตอร์ (Parametric Statistics)

สถิติแบบไม่ใช้พารามิเตอร์ (Non-parametric Statistics)

การประมาณค่า

1) สถิติแบบพารามิเตอร์ (Parametric Statistics)

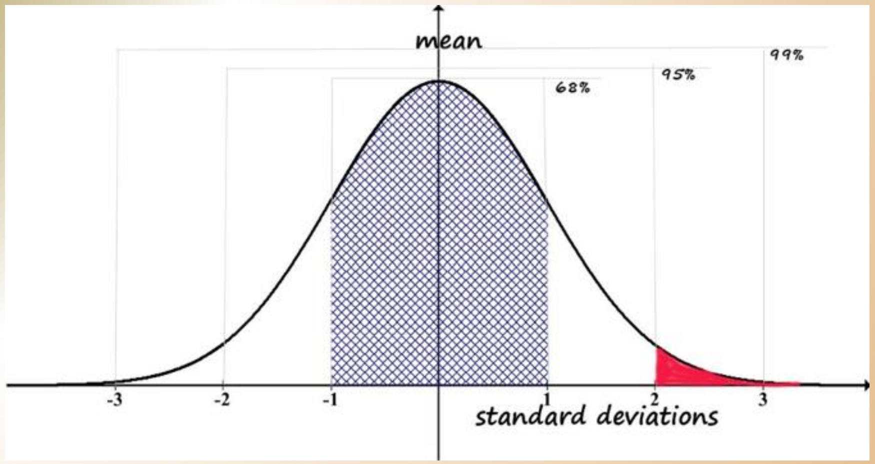
One-sample t-test: ใช้ทดสอบค่าเฉลี่ยของกลุ่มข้อมูลกับค่าคงที่

Independent t-test: ใช้ทดสอบค่าเฉลี่ยของกลุ่มข้อมูลสองกลุ่มที่ไม่เกี่ยวข้องกัน

การวิเคราะห์ความแปรปรวน (Analysis of Variance, ANOVA)

การวิเคราะห์การถดถอย (Regression Analysis)

การหาความสัมพันธ์ (Pearson Correlation)



สถิติพารามิเตอร์ (Parametric Statistics)

เป็นวิธีการทางสถิติที่จะต้องเป็นไป

ตามข้อตกลงเบื้องต้น 3 ประการ ดังนี้

(1) ข้อมูลที่เก็บรวบรวมได้จะต้องอยู่ในระดับช่วงขึ้นไป (Interval Scale)

(2) ข้อมูลที่เก็บรวบรวมได้จากกลุ่มตัวอย่างจะต้องมีการแจกแจงเป็นโค้งปกติ

(3) กลุ่มประชากรแต่ละกลุ่มที่น ามาศึกษาจะต้องมีความแปรปรวนเท่ากัน

สถิติแบบไม่ใช้พารามิเตอร์ (Nonparametric Statistics)

เป็นวิธีการทางสถิติที่ สามารถนำมาใช้ได้โดยปราศจากข้อตกลงเบื้องต้นทั้ง 3 ประการข้างต้น

2) สถิติแบบไม่ใช้พารามิเตอร์ (Non-parametric Statistics)

Chi-square test: ใช้สำหรับการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเชิงหมวดหมู่

Mann-Whitney U test: ใช้ทดสอบความแตกต่างระหว่างกลุ่มสองกลุ่มแทน t-test เมื่อข้อมูลไม่เป็นไปตามการกระจายตัวแบบปกติ

Signed-Rank test: ใช้ทดสอบความแตกต่างของกลุ่มข้อมูลสองกลุ่มที่มีความสัมพันธ์กันแทน paired t-test

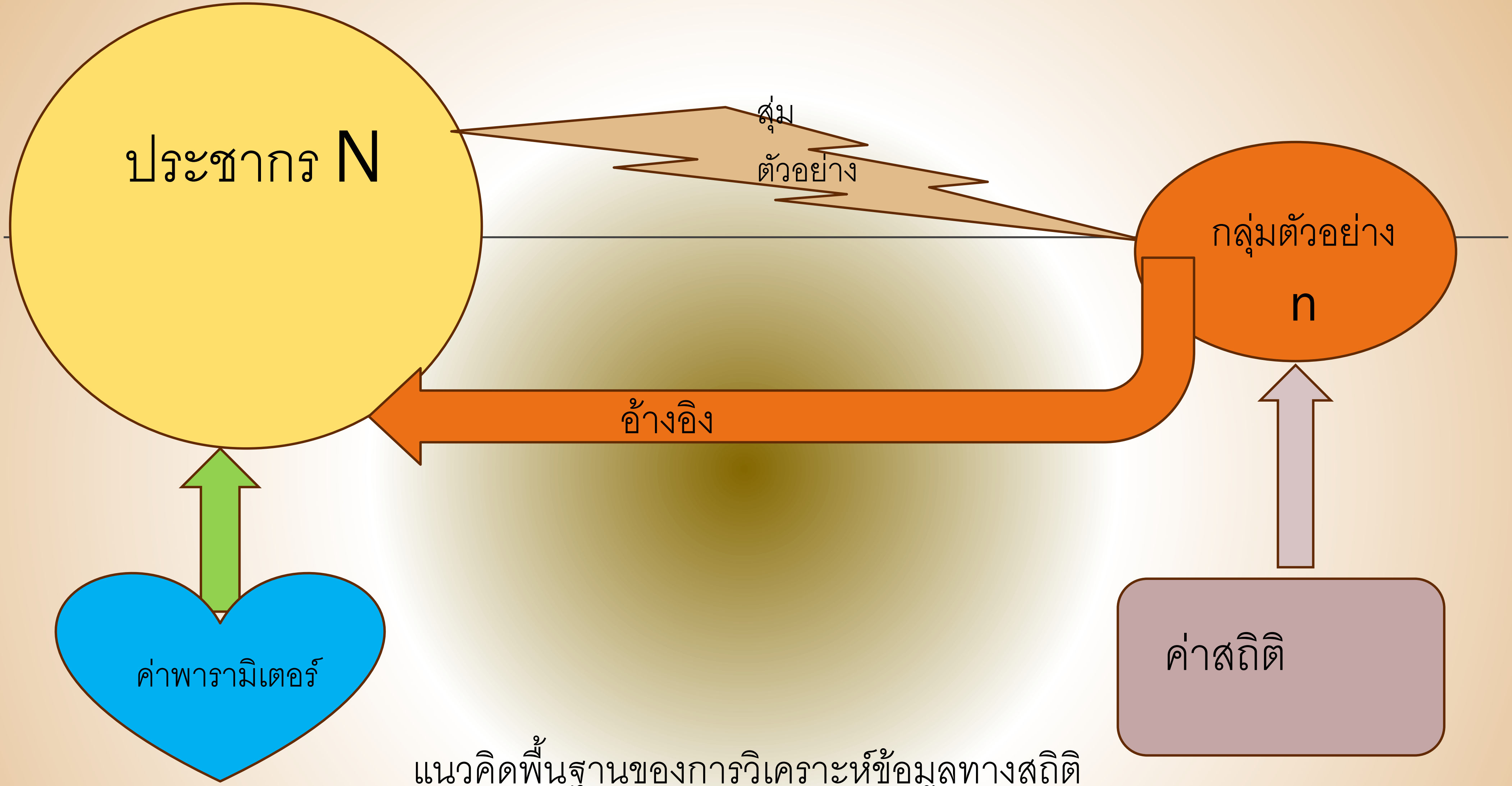
การประมาณค่า

การประมาณค่า (estimation) หมายถึงการนำเอาตัวสถิติที่ได้จากตัวอย่างไปประมาณ พารามิเตอร์ของประชากร การประมาณค่าเป็นวิธีการหนึ่งในการหาข้อสรุปเกี่ยวกับประชากร โดย การอาศัยตัวอย่างที่ได้จากการชักตัวอย่างแบบสุ่มด้วยวิธีการต่าง ๆ ที่เหมาะสมแล้วนำข้อมูลที่ได้จาก ตัวอย่างนั้นไปคำนวณหาตัวสถิติเพื่อนำไปประมาณพารามิเตอร์ที่สนใจ อันได้แก่ ค่าเฉลี่ย (μ) ความแปรปรวน (σ^2) และสัดส่วน (P) ซึ่งเป็นตัวไม่ทราบค่า ถ้าให้ φ เป็นพารามิเตอร์ใด ๆ ที่สนใจ $\hat{\varphi}$ เป็นตัวสถิติที่ใช้ประมาณพารามิเตอร์ φ เรียก $\hat{\varphi}$ ว่าตัวประมาณ (estimator) ของ φ โดยที่ $\hat{\varphi}$ จะเป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นของตัวเอง ค่า $\hat{\varphi}$ ที่คำนวณได้ เรียกว่า ค่าประมาณ (estimate) ของφ

การประมาณค่า

แบบจุด

แบบช่วง



แนวคิดพื้นฐานของการวิเคราะห์ข้อมูลทางสถิติ

สรุป

การวิเคราะห์ข้อมูลในสถิติมักใช้สถิติทั้งสถิติบรรยายและสถิติอ้างอิง เพื่อให้เข้าใจลักษณะของข้อมูลและการแจกแจง สถิติบรรยายใช้ในการอธิบายและสรุปข้อมูลตามลักษณะของข้อมูล เช่น เฉลี่ย ค่าสูงสุด ค่าต่ำสุด และการกระจายของข้อมูล ในขณะที่สถิติอ้างอิงใช้ในการทำนายหรือตรวจสอบสมมติฐานเกี่ยวกับประชากร

บทที่ 5 ตัวแปรสุ่มและการแจกแจง

- การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม
- การประมาณค่าความน่าจะเป็น
- ตัวแปรสุ่มไม่ต่อเนื่อง
- ตัวแปรสุ่มต่อเนื่อง
- สรุป

วัตถุประสงค์การเรียนรู้

- สามารถหาค่าความน่าจะเป็นของเหตุการณ์ที่สนใจเพียงบางเหตุการณ์
- สามารถหาความน่าจะเป็นของผลลัพธ์ที่มีค่าเป็นตัวเลขที่เป็นไปได้ทั้งหมดทุกกรณี
- สามารถแยกประเภทของการแจกแจงความน่าจะเป็นตัวแปรสุ่ม

บทนำ

การศึกษาเกี่ยวกับตัวแปรสุ่มและการแจกแจงเป็นส่วนสำคัญของวิชาสถิติ เนื่องจากเป็นเครื่องมือที่ช่วยให้สามารถวิเคราะห์และทำนายผลของเหตุการณ์ที่มีความไม่แน่นอนได้ โดยทั้งสองเรื่องนี้มีความเกี่ยวข้องกับ ความน่าจะเป็นที่เป็นหลักฐานทางวิทยาศาสตร์ การศึกษาเกี่ยวกับตัวแปรสุ่มทั้งหมดมีวัตถุประสงค์ในการระบุและการแก้ไขปัญหาที่เกิดขึ้นในโลกที่เต็มไปด้วยความไม่แน่นอนและการประมาณค่าความน่าจะเป็นในรูปแบบต่างๆ เป็นเครื่องมือที่มีประโยชน์ในการทำนายและการตัดสินใจทางธุรกิจ

ตัวแปรสุ่ม (Random Variable)

คือตัวแปรที่สามารถมีค่าหลายค่าได้โดยมีความน่าจะเป็นที่แต่ละค่าเป็นไปได้เท่ากับค่าความน่าจะเป็นที่กำหนดไว้ ตัวแปรสุ่มสามารถแบ่งออกเป็นสองประเภทหลักได้แก่ ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable) และตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)

1. ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable)

ลักษณะ: ตัวแปรสุ่มไม่ต่อเนื่องสามารถรับค่าเป็นจำนวนจำกัดหรือจำนวนไม่จำกัดที่สามารถนับได้ (Countable).ค่า: ค่าเป็นจำนวนเต็มหรือจำนวนที่นับได้ เช่น 0, 1, 2, 3, ...ตัวอย่าง:จำนวนการทอยลูกเต๋าให้ได้เลขหกจำนวนข้อสอบที่นักเรียนทำถูกต้องจำนวนผู้เข้าร่วมงานในแต่ละวัน

- มีค่าได้แค่บางค่าแบบจำกัด เช่น จำนวนการโยนลูกเต๋าได้ 1, 2, 3, 4, 5, หรือ 6
- ตัวอย่างของการแจกแจงความน่าจะเป็นในตัวแปรสุ่มไม่ต่อเนื่องได้แก่ การแจกแจงเบอร์นูลลี (Bernoulli Distribution) และการแจกแจงทวินาม (Binomial Distribution)

2. ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)

ลักษณะ: ตัวแปรสุ่มต่อเนื่องสามารถรับค่าใด ๆ ในช่วงหนึ่ง ๆ บนเส้นจำนวนจริง (Real number line).

ค่า: ค่าเป็นจำนวนที่ไม่จำกัดและไม่สามารถนับได้ เช่น ค่าเป็นตัวเลขที่สามารถมีค่าใด ๆ ระหว่าง 0 และ 1 เช่น 0.1, 0.01, 0.001, ...

ตัวอย่าง: เวลาในการวิ่ง 100 เมตร

- ความสูงของนักเรียนในชั้นเรียน
- อุณหภูมิในเมืองหนึ่งในแต่ละวัน

ตัวอย่างของการแจกแจงความน่าจะเป็นในตัวแปรสุ่มต่อเนื่องได้แก่ การแจกแจงปกติ (Normal Distribution) และการแจกแจงปัวซอง (Poisson Distribution)

เปรียบเทียบลักษณะการนับค่า

ตัวแปรสุ่มไม่ต่อเนื่องนับค่าได้ ส่วนตัวแปรสุ่มต่อเนื่องนับค่าไม่ได้ (เป็นช่วงค่า).

ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function - PDF): ตัวแปรสุ่มต่อเนื่องมีฟังก์ชัน PDF ซึ่งต้องใช้สำหรับการหาความน่าจะเป็นในช่วงค่าหนึ่ง.

ฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Mass Function - PMF): ตัวแปรสุ่มไม่ต่อเนื่องมีฟังก์ชัน PMF ที่ใช้สำหรับการหาความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าหนึ่งๆ.

ทั้งนี้ การระบุว่าตัวแปรสุ่มเป็นประเภทใดขึ้นอยู่กับลักษณะของข้อมูลที่ต้องการศึกษาและวิเคราะห์

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง

ถ้า X เป็นตัวแปรสุ่ม แล้วค่าของตัวแปรสุ่ม X อาจเกิดขึ้นด้วยความ น่าจะเป็นที่แตกต่างกัน

ความน่าจะเป็นของตัวแปรสุ่มเขียนแทนด้วยสัญลักษณ์

$P(X = x)$ หมายถึงค่าความน่าจะเป็นของตัวแปรสุ่ม X เมื่อ X มีค่าเท่ากับ x

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง

จากการสำรวจพบว่านักศึกษาในห้องนี้ จำนวน 50 คน

จำนวนพี่น้อง	ความถี่
0	8
1	20
2	16
3	5
4	1
รวม	50

จงหา $P(X = x)$ โดยที่ $x = 0, 1,2,3,4$

$P(X = 0) = 8/50$, $P(X = 1) = 20/50$

$P(X = 2) = 16/50$, $P(X = 3) = 5/50$

$P(X = 4) = 1/50$

$P(X = x) = \mathrm{f}(x)$

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม คือ ฟังก์ชันที่แสดงค่าความน่าจะเป็นที่จะเกิดค่าของตัวแปรสุ่ม $\boldsymbol{X}$ ในการทดลอง

$$P(X = x) = f(x)$$

$$f(0) + f(1) + f(2) + f(3) + f(4) = 1$$

จำนวนพี่น้อง	ความถี่
0	8
1	20
2	16
3	5
4	1
รวม	50

เรียกว่าตารางแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

x	0	1	2	3	4
$f(x)$	8/50	20/50	16/50	5/50	1/50

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

x	0	1	2	3	4
f(x)	8/50	20/50	16/50	5/50	1/50

$P(X < 3)$ = ความน่าจะเป็นที่นักศึกษามีพี่น้องน้อยกว่า 3 คน

$P(X > 2)$ = ความน่าจะเป็นที่นักศึกษามีพี่น้องมากกว่า 2 คน

$P2 < X < 4$) = ความน่าจะเป็นที่นักศึกษามีพี่น้องมากกว่า 2 คนและน้อยกว่า 4 คน

การประมาณค่าความน่าจะเป็น (Probability Estimation)

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

1.จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้คือ 6 หน้าของลูกเต๋า

2.จำนวนของผลลัพธ์ที่ต้องการคือ 1 หน้า (หน้าที่มีค่าเท่ากับ 3)

3.ความน่าจะเป็น = (จำนวนของผลลัพธ์ที่ต้องการ) / (จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้) = 1 / 6

4.ความน่าจะเป็น = 0.1667 หรือ 16.67%

การประมาณค่าความน่าจะเป็น (Probability Estimation)

ในกรณีที่มีลูกเต๋า 6 หน้า 2 ลูกและต้องการหาความน่าจะเป็นที่จะได้หน้าที่มีค่าบวกกันเท่ากับ 4 หลังจากการโยนลูกเต๋า

	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			

จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้คือ 6 x 6 = 36

ลูกเต๋าที่เมื่อบวกกันแล้วเท่ากับ 4 ซึ่งมีคู่ที่เป็นไปได้ดังนี้:

- ลูกเต๋าที่หน้า 1 และ 3
- ลูกเต๋าที่หน้า 2 และ 2
- ลูกเต๋าที่หน้า 3 และ 1

นับจำนวนของคู่ที่เป็นไปได้ ซึ่งในกรณีนี้มีทั้งหมด 3 คู่

ความน่าจะเป็น = (จำนวนของคู่ที่เป็นไปได้) / (จำนวนทั้งหมดของผลลัพธ์ที่เป็นไปได้) = 3 / 36

ความน่าจะเป็น = 1/12

ความน่าจะเป็น = 0.0833 หรือ 8.33%

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete Random Variable)

ตัวแปรสุ่มต่อเนื่อง (Continuous Random Variable)

ตัวแปรสุ่มไม่ต่อเนื่อง

- การแจกแจงเบอร์นูลลี (Bernoulli Distribution)
- การแจกแจงทวินาม (Binomial Random Variable)
- การแจกแจงความน่าจะเป็นแบบเอกรูป (Uniform Probability Distribution)
- การแจกแจงเรขาคณิต (Geometric Random Variable)
- การแจกแจงปัวซอง (Poisson Random Variable)
- การแจกแจงทวินามลบ (Negative Binomial Distribution)

การแจกแจงเบอร์นูลลี (Bernoulli Distribution)

การแจกแจงเบอร์นูลลี (Bernoulli Distribution) เป็นการแจกแจงของตัวแปรสุ่มที่มีเพียงสองค่าที่เป็นไปได้ คือ ความสำเร็จ (Success) กับ ความล้มเหลว (Failure) ซึ่งมักถูกแทนด้วย 1 และ 0 ตามลำดับ

outcome → Success

outcome → Failure

sample space S = {0,1}

$P(X = 1) = p$

$P(X = 0) = q$ หรือ $(1-p)$

$P(S) = 1$ ดังนั้น $P(X=x) = p^{x} \cdot (1-p)^{1-x}$

การแจกแจงเบอร์นูลลี (Bernoulli Distribution)

การแจกแจงเบอร์นูลลีถูกนำไปใช้ในสถานการณ์ที่มีเพียงสองผลลัพธ์ เช่น การโยนเหรียญ (หัวหรือก้อย), การทดสอบว่าอุปกรณ์ทำงานหรือไม่ (ทำงานหรือไม่ทำงาน), การทดสอบว่าผู้ป่วยหายจากโรคหรือไม่ เป็นต้นเพื่อให้เข้าใจได้ดีขึ้น

ลองพิจารณาตัวอย่างดังนี้:สมมติว่ามีเหรียญที่มีความน่าจะเป็นออกหัว (Success) เป็น (p = 0.7) และออกก้อย (Failure) เป็น (1- p = 0.3) ถ้าโยนเหรียญหนึ่งครั้ง ความน่าจะเป็นที่เหรียญออกหัว (X=1) คือ 0.7 ความน่าจะเป็นที่เหรียญออกก้อย (X=0) คือ 0.3

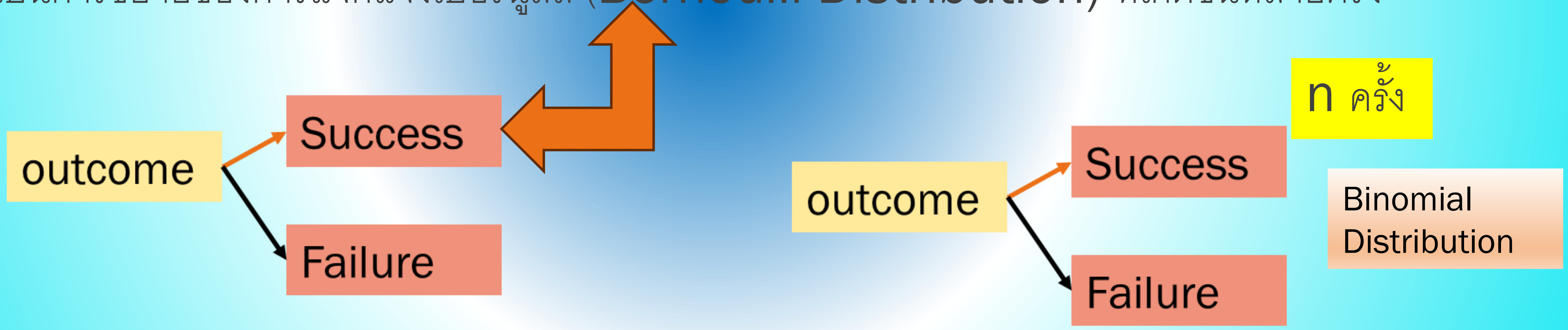
$P(X=x) = p^{x} \cdot (1-p)^{1-x}$

$P(X=1) = 0.7^{1}(0.3)^{1-1} = 0.7$

$P(X=0) = 0.7^{0}(0.3)^{1-0} = 0.3$

การแจกแจงทวินาม (Binomial Random Variable)

การแจกแจงทวินาม (Binomial Distribution) เป็นการแจกแจงของจำนวนความสำเร็จในการทดลองอิสระซ้ำๆ กัน n ครั้ง โดยแต่ละครั้งมีความน่าจะเป็นของความสำเร็จเท่ากันคือ p ซึ่งสามารถมองว่าเป็นการขยายของการแจกแจงเบอร์นูลลี (Bernoulli Distribution) ที่เกิดขึ้นหลายครั้ง



การแจกแจงทวินาม (Binomial Random Variable)

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

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

k คือ จำนวนครั้งความสำเร็จ

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

การแจกแจงทวินามถูกนำไปใช้ในสถานการณ์ที่ต้องการนับจำนวนความสำเร็จในชุดของการทดลองซ้ำ ๆ เช่น การโยนเหรียญหลายครั้ง การผลิตสินค้าที่มีโอกาสผ่านมาตรฐานหรือไม่ การสำรวจประชามติที่ตอบว่าใช่หรือไม่ เป็นต้น

การแจกแจงทวินาม (Binomial Random Variable)

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

มีเหรียญที่มีความน่าจะเป็นออกหัว (Success) เป็น $p=0.5$ และทำการโยนเหรียญ $n=10$ ครั้ง ถ้าต้องการหาความน่าจะเป็นที่จะได้หัวจำนวน $k=6$ ครั้ง

จำนวนความเป็นไปได้ทั้งหมดที่โยนเหรียญ 10 ครั้งได้หัว 6 ครั้งคือ $\binom{10}{6}$

ความน่าจะเป็นของแต่ละความเป็นไปได้คือ $(0.5)^6(1-0.5)^{10-6}$

ดังนั้น ความน่าจะเป็นที่ได้หัว 6 ครั้งใน 10 ครั้งคือ $P(X=6)=\binom{10}{6}(0.5)^6(0.5)^4 = \binom{10}{6}(0.5)^{10}$

$P(X=6)=\frac{210}{1024} \approx 0.205$ ดังนั้น ความน่าจะเป็นที่ได้หัว 6 ครั้งใน 10 ครั้งคือประมาณ 0.205 หรือ 20.5%

การแจกแจงความน่าจะเป็นแบบเอกรูป (Uniform Probability Distribution)

การแจกแจงความน่าจะเป็นแบบเอกรูป (Uniform Probability Distribution) เป็นการแจกแจงที่ทุกค่าของตัวแปรสุ่มมีความน่าจะเป็นเกิดขึ้นเท่ากัน การแจกแจงแบบเอกรูปสามารถแบ่งออกเป็นสองประเภทหลัก ๆ คือ การแจกแจงแบบเอกรูปไม่ต่อเนื่อง (Discrete Uniform Distribution) และการแจกแจงแบบเอกรูปต่อเนื่อง (Continuous Uniform Distribution)

การแจกแจงแบบเอกรูปไม่ต่อเนื่อง (Discrete Uniform Distribution)

การแจกแจงแบบเอกรูปไม่ต่อเนื่องหมายถึงการที่ค่าของตัวแปรสุ่มมีจำนวนจำกัดและทุกค่ามีความน่าจะเป็นเกิดขึ้นเท่ากัน เช่น การทอยลูกเต๋าที่มี 6 หน้า

ตัวแปรสุ่ม X สามารถมีค่า $x_1, x_2, \dots, x_n$

ความน่าจะเป็นของแต่ละค่าคือ $\frac{1}{n}$

$P(X=x_i) = \frac{1}{n}$ **ตัวอย่าง**: การทอยลูกเต๋าหนึ่งลูก ความน่าจะเป็นที่ได้แต่ละหน้า (1 ถึง 6) เท่ากับ $\frac{1}{6}$

การแจกแจงเรขาคณิต (Geometric Random Variable)

การแจกแจงเรขาคณิต (Geometric Distribution) เป็นการแจกแจงของตัวแปรสุ่มที่นับจำนวนการทดลองที่ล้มเหลวก่อนที่จะเกิดความสำเร็จครั้งแรก การทดลองแต่ละครั้งเป็นการทดลองแบบเบอร์นูลลี (Bernoulli Trial) ที่มีเพียงสองผลลัพธ์ คือ ความสำเร็จ (Success) และ ความล้มเหลว (Failure)

ตัวแปรสุ่ม X แทนจำนวนครั้งของการทดลองที่ล้มเหลวก่อนความสำเร็จครั้งแรก

ความน่าจะเป็นของความสำเร็จในแต่ละครั้งของการทดลองคือ p

ความน่าจะเป็นของความล้มเหลวในแต่ละครั้งของการทดลองคือ 1−p

$P(X=k)=(1-p)^k p$

เมื่อ k เป็นจำนวนเต็มที่ไม่ลบ ($k = 0,1,2,\ldots$)

การแจกแจงเรขาคณิต (Geometric Random Variable)

การแจกแจงเรขาคณิตถูกนำไปใช้ในสถานการณ์ที่ต้องการนับจำนวนครั้งที่เกิดความล้มเหลวก่อนที่จะเกิดความสำเร็จ เช่น:

- การโยนเหรียญซ้ำ ๆ จนกว่าจะออกหัว
- การทดสอบผลิตภัณฑ์จนกว่าจะได้ผลิตภัณฑ์ที่มีคุณภาพ
- การสุ่มเลือกตัวอย่างจนกว่าจะพบตัวอย่างที่ตรงตามเงื่อนไขที่ต้องการ

การแจกแจงเรขาคณิต (Geometric Random Variable)

สมมติว่ามีเหรียญที่มีความน่าจะเป็นออกหัว (Success) เป็น $p=0.3$ และความน่าจะเป็นออกก้อย (Failure) คือ $1-p=0.7$ ต้องการหาความน่าจะเป็นที่จะต้องโยนเหรียญ k ครั้งก่อนที่จะออกหัวครั้งแรก

$P(X=k)=(1-p)^k p$
เมื่อ k เป็นจำนวนเต็มที่ไม่ลบ
($k = 0,1,2,\ldots$)

ความน่าจะเป็นที่จะต้องโยนเหรียญ 2 ครั้งก่อนที่จะออกหัวครั้งแรก($X=2$):

$$P(X=2)=(1-0.3)^2\times 0.3 = 0.7^2\times 0.3 = 0.49\times 0.3 = 0.147$$

ดังนั้น ความน่าจะเป็นที่จะต้องโยนเหรียญ 2 ครั้งก่อนที่จะออกหัวครั้งแรกคือ 0.147 หรือ 14.7%

การแจกแจงปัวซอง (Poisson Random Variable)

การแจกแจงปัวซอง (Poisson Distribution) เป็นการแจกแจงความน่าจะเป็นที่ใช้ในการอธิบายจำนวนครั้งที่เหตุการณ์เกิดขึ้นในช่วงเวลาหรือพื้นที่ที่กำหนด โดยมีสมมติฐานว่าเหตุการณ์เกิดขึ้นอย่างเป็นอิสระจากกันและมีอัตราการเกิดเหตุการณ์คงที่

- ตัวแปรสุ่ม X แทนจำนวนครั้งที่เหตุการณ์เกิดขึ้นในช่วงเวลาหรือพื้นที่ที่กำหนด
- ค่าเฉลี่ย (Mean) หรืออัตราการเกิดเหตุการณ์ต่อหน่วยเวลา/พื้นที่คือ λ
- ตัวแปรสุ่ม X สามารถมีค่าจำนวนเต็มที่ไม่ลบ (0, 1, 2, ...)

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$P(X = x) = f(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

การแจกแจงปัวซอง (Poisson Random Variable)

$$P(X = x) = f(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \dots$$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

k คือจำนวนครั้งที่เหตุการณ์เกิดขึ้น (k = 0, 1, 2, ...)

λ คือค่าเฉลี่ยหรืออัตราการเกิดเหตุการณ์

e คือค่าคงที่ทางคณิตศาสตร์ (ประมาณ 2.71828)

การแจกแจงปัวซองถูกนำไปใช้ในสถานการณ์ที่ต้องการวิเคราะห์จำนวนครั้งที่เหตุการณ์เกิดขึ้นในช่วงเวลาหรือพื้นที่ที่กำหนด เช่น

* จำนวนการโทรเข้าสายด่วนในหนึ่งชั่วโมง* จำนวนรถที่ผ่านทางแยกในหนึ่งนาที *จำนวนข้อผิดพลาดในหน้านิตยสาร

การแจกแจงปัวซอง (Poisson Random Variable)

สมมติว่าอัตราการโทรเข้าศูนย์บริการลูกค้าคือ 5 สายต่อชั่วโมง (λ=5) ต้องการหาความน่าจะเป็นที่จะมีการโทรเข้า 3 สายในหนึ่งชั่วโมง

$$P(X = x) = f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0, 1, 2, \ldots$$

$$P(X=3) = \frac{5^3 e^{-5}}{3!}$$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$P(X=3) = \frac{125 \cdot 0.0067}{6} \approx \frac{0.8375}{6} \approx 0.1396$$

ดังนั้น ความน่าจะเป็นที่จะมีการโทรเข้า 3 สายในหนึ่งชั่วโมงคือประมาณ 0.1396 หรือ 13.96%

การแจกแจงทวินามลบ (Negative Binomial Distribution)

การแจกแจงทวินามลบ (Negative Binomial Distribution) เป็นการแจกแจงของตัวแปรสุ่มที่นับจำนวนครั้งของความล้มเหลวก่อนที่จะเกิดความสำเร็จครั้งที่ r ในการทดลองแบบเบอร์นูลลี (Bernoulli Trials) โดยแต่ละครั้งมีความน่าจะเป็นของความสำเร็จคงที่ที่ p

- ตัวแปรสุ่ม X แทนจำนวนครั้งของความล้มเหลวก่อนที่จะเกิดความสำเร็จครั้งที่ r
- ความน่าจะเป็นของความสำเร็จในแต่ละครั้งของการทดลองคือp
- ความน่าจะเป็นของความล้มเหลวในแต่ละครั้งของการทดลองคือ 1−p
- r เป็นจำนวนความสำเร็จที่ต้องการ

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

การแจกแจงทวินามลบ (Negative Binomial Distribution)

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

k คือจำนวนความล้มเหลวก่อนที่จะเกิดความสำเร็จครั้งที่ r

$\binom{k+r-1}{k}$ คือ สัญลักษณ์ทวินาม (Binomial Coefficient) ซึ่งคำนวณได้จาก

$$\binom{k+r-1}{k}=\frac{(k+r-1)!}{k!(r-1)!}$$

การแจกแจงทวินามลบถูกนำไปใช้ในสถานการณ์ที่ต้องการวิเคราะห์จำนวนครั้งของความล้มเหลวก่อนที่จะเกิดความสำเร็จครั้งที่ r เช่น

- การทดสอบผลิตภัณฑ์จนกว่าจะได้ผลิตภัณฑ์ที่มีคุณภาพครบ r ชิ้น
- การทอยลูกเต๋าจนกว่าจะได้ผลลัพธ์ที่ต้องการครบr ครั้ง

การแจกแจงทวินามลบ (Negative Binomial Distribution)

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

สมมติว่ามีเหรียญที่มีความน่าจะเป็นออกหัว (Success) เป็น p=0.3 และความน่าจะเป็นออกก้อย (Failure) คือ 1−p=0.7 ต้องการหาความน่าจะเป็นที่จะต้องโยนเหรียญ k ครั้งก่อนที่จะออกหัวครั้งที่ 3 (r=3)

$$P(X=5) = \binom{5+3-1}{5} 0.3^3 0.7^3 = \approx 0.09529569$$

ดังนั้น ความน่าจะเป็นที่จะต้องโยนเหรียญ 5 ครั้งก่อนที่จะออกหัวครั้งที่ 3 คือประมาณ 0.0953 หรือ 9.53%

ตัวแปรสุ่มต่อเนื่อง

- การแจกแจงแบบเอกรูปต่อเนื่อง (Continuous Uniform Distribution)
- การแจกแจงที (t-distribution)
- การแจกแจงไคสแควร์ (Chi-Square(χ^2) Test)
- การแจกแจงเอฟ (F - distribution)
- การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

การแจกแจงแบบเอกรูปต่อเนื่อง (Continuous Uniform Distribution)

การแจกแจงแบบเอกรูปต่อเนื่องหมายถึงการที่ตัวแปรสุ่มสามารถมีค่าที่ต่อเนื่องภายในช่วงหนึ่งและทุกค่าภายในช่วงนั้นมีความน่าจะเป็นเกิดขึ้นเท่ากัน

ตัวแปรสุ่ม X มีค่าต่อเนื่องในช่วง $[a,b]$ ความน่าจะเป็นในช่วง $[a,b]$ มีค่าเท่ากัน **สำหรับค่าใดๆ ในช่วงนี้ ความน่าจะเป็นจะเท่ากับ** $\frac{1}{b-a}$ **ดังนั้น** $P(X=x) = \frac{1}{b-a}$, a$\leq x \leq$ b เมื่อ a< b

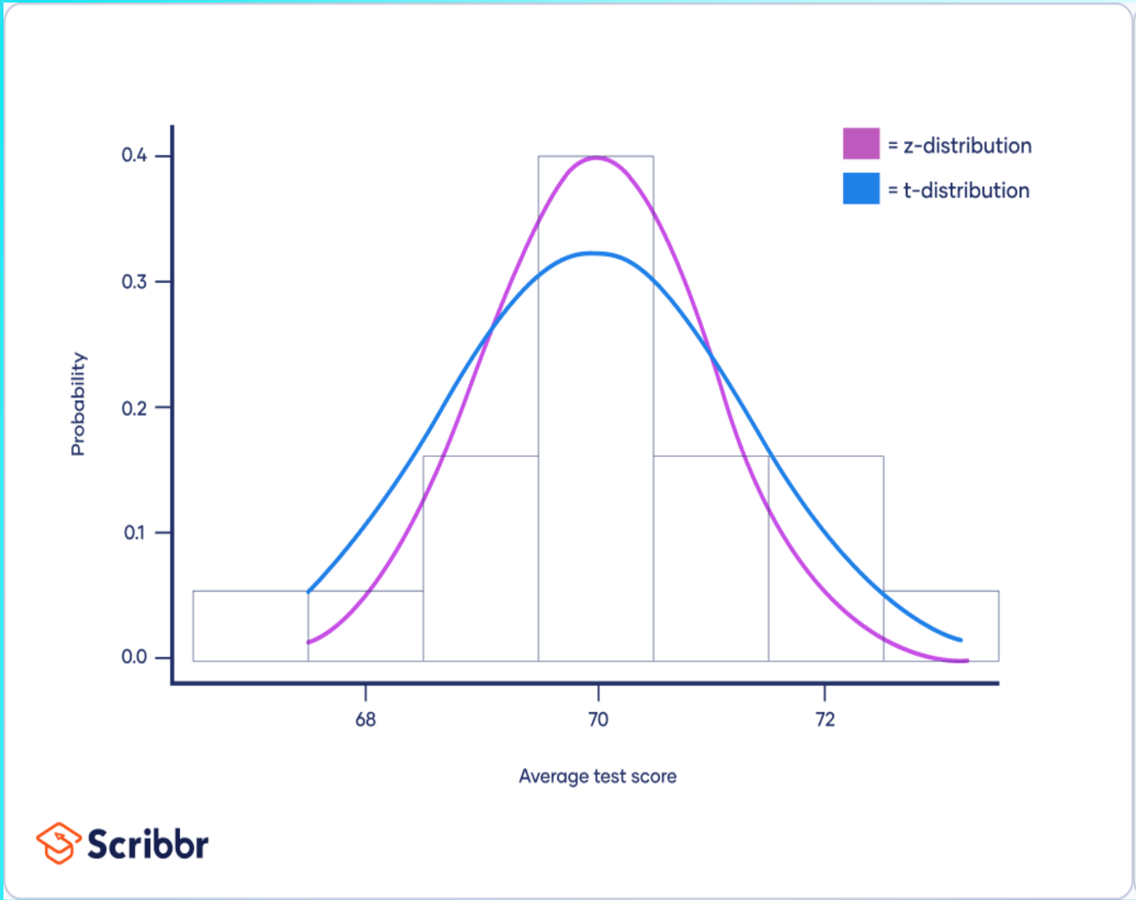
ตัวแปรสุ่ม x มีการแจกแจงแบบเอกรูปต่อเนื่อง ในช่วง -1 ถึง 1 = $\frac{1}{1-(-1}$ $= \frac{1}{2}$, -1$\leq x \leq 1$

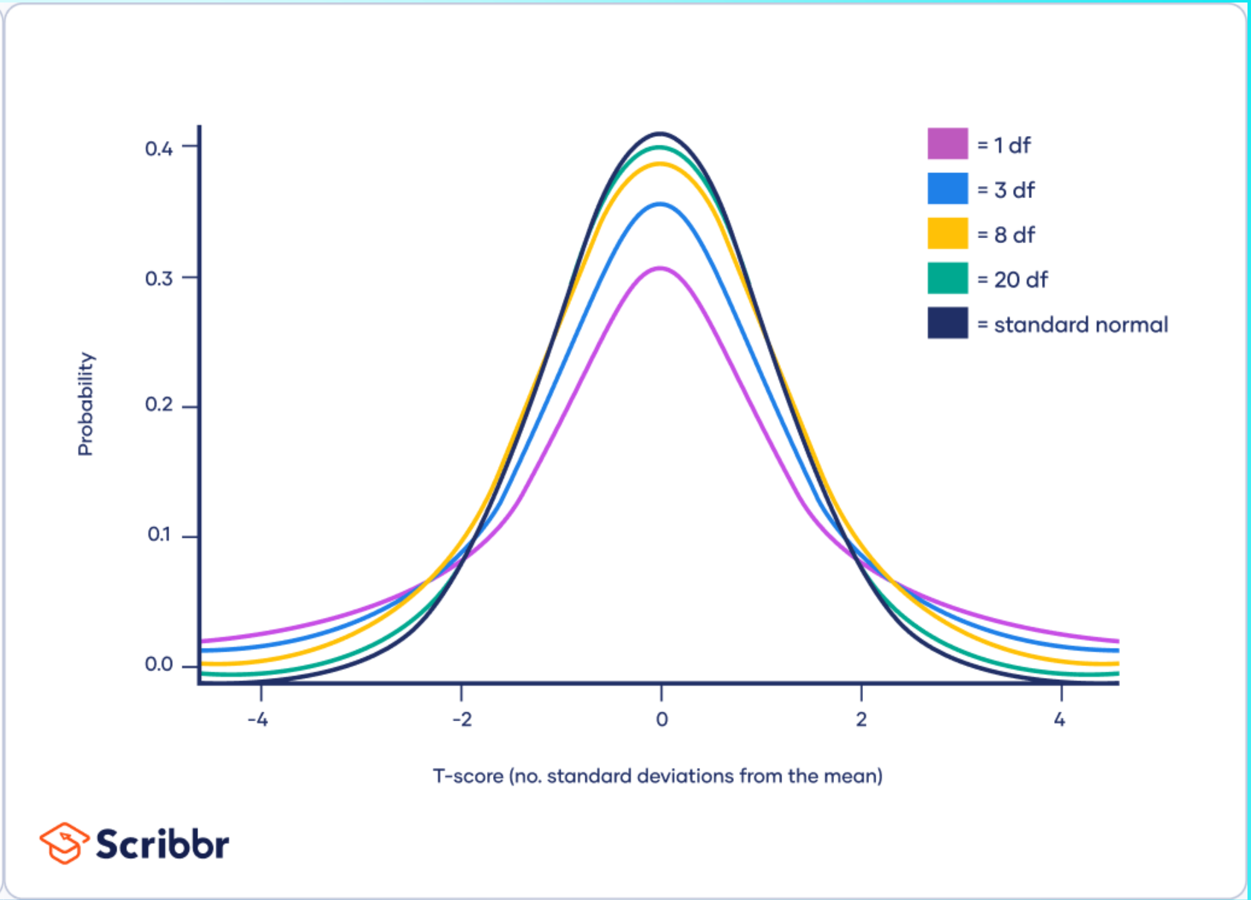
ถ้า x มีค่าอื่นจะเท่า 0

การแจกแจงที (t-distribution)

การแจกแจงที (t-distribution) หรือที่เรียกอย่างเป็นทางการว่า Student's t-distribution เป็นการแจกแจงความน่าจะเป็นที่ใช้บ่อยในสถิติ โดยเฉพาะในการทดสอบสมมติฐานและการประมาณค่าช่วงความเชื่อมั่น (confidence intervals) และการทดสอบทีใช้ในการเปรียบเทียบค่าเฉลี่ยของสองกลุ่มตัวอย่าง (หนึ่งกลุ่มหรือสองกลุ่ม)

- สำหรับค่าเฉลี่ยของประชากรเมื่อขนาดตัวอย่างมีขนาดเล็ก
- ความแปรปรวนของประชากรไม่เป็นที่รู้จัก





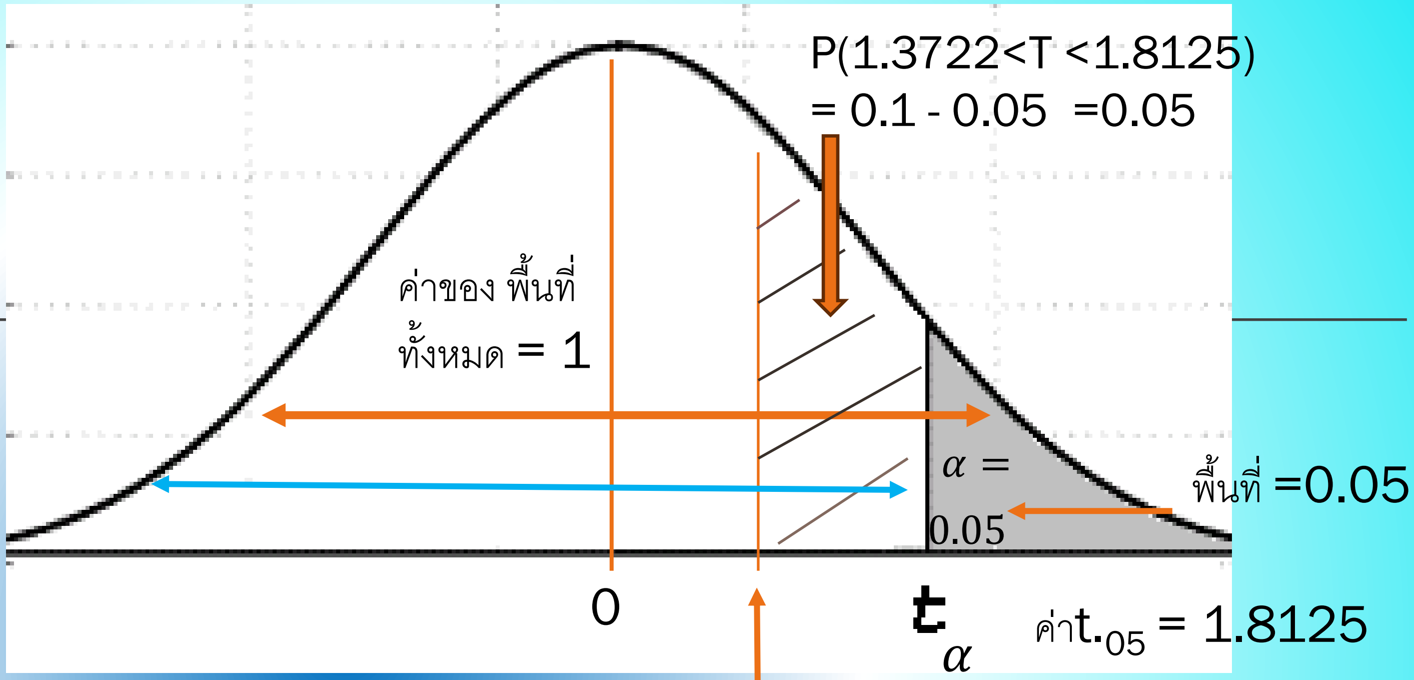
As the degrees of freedom (total number of observations minus 1) increases, the t-distribution will get closer and closer to matching the standard normal distribution

ลักษณะของการแจกแจงที่การแจกแจงที่มีลักษณะคล้ายกับการแจกแจงปกติ (Normal Distribution) แต่มีความแบนกว่าที่หาง การแจกแจงที่จะเข้าใกล้การแจกแจงปกติมากขึ้นเมื่อจำนวนองศาอิสระ (degrees of freedom, ν) เพิ่มขึ้น

การอ่านตาราง t

ให้หาตัวแปรสุ่ม T ที่มีการแจกแจง t โดยที่ df = 10

1. P(T > 1.8125)
2. P(T < 1.8125)
3. P(1.8125 <T <1.3722)



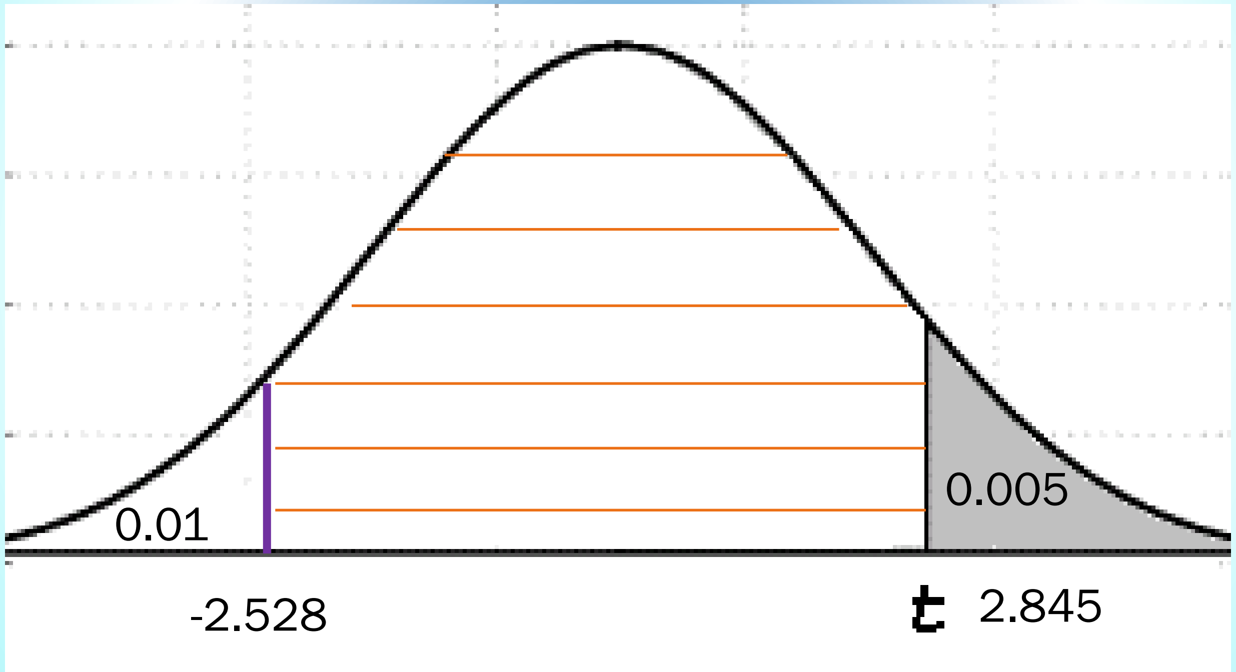
ค่าt .1 = 1.3722

เปิดตาราง t ที่ df =10 จะพบว่า 1.8125 = 0.05

1. P(T > 1.8125) = 0.05 พื้นที่สีเทา
2. P(T ≤ 1.8125) = 1-0.05= 0.95 พื้นที่ลูกศรสีฟ้า
3. P(1.3722< T <1.8125) = 0.1 – 0. 05 =0.05 พื้นที่แรเงา

α=0.05 df =10 เปิดตาราง t (1- tail) = 1.8125
α=0.01 df =20 เปิดตาราง t (1- tail) = 2.5280
α=0.025 df =10 เปิดตาราง t (1- tail) = 2.2281

P (-2.528 < t < 2.845) df = 20



P (-2.528 < t < 2.845)
= 1 – (0.01+0.005)
=0.985

ศึกษาเพิ่มเติม

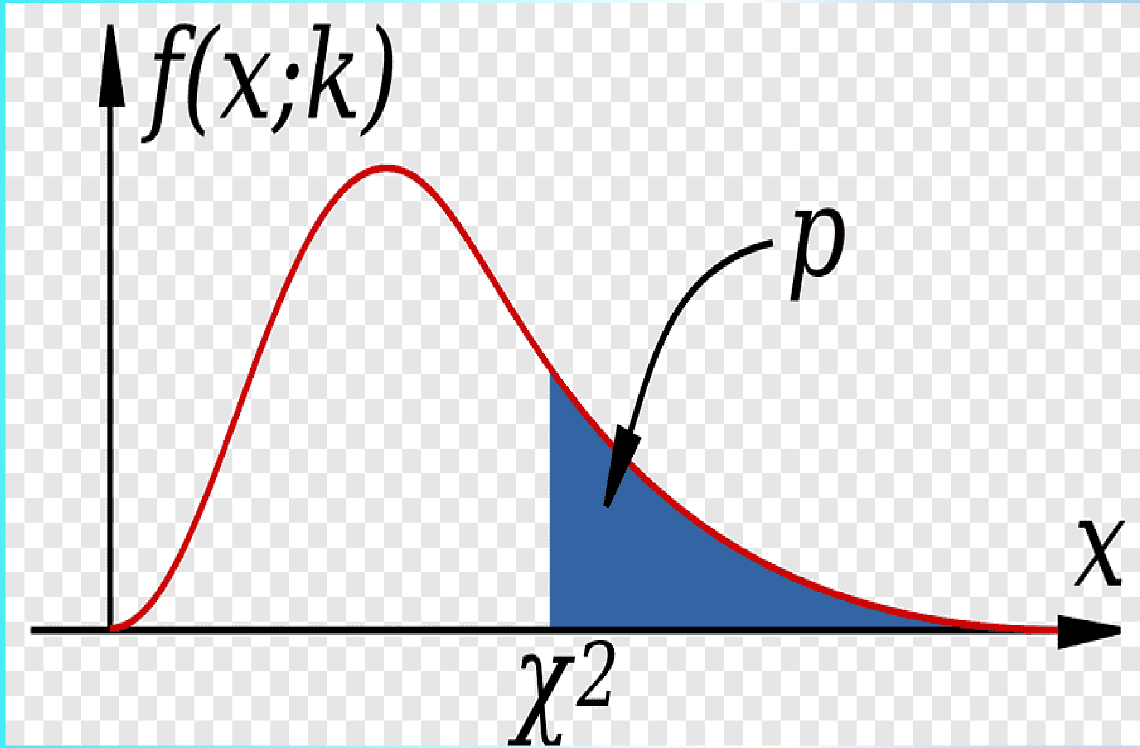
https://www.youtube.com/watch?v=OFeSvAAdZw0

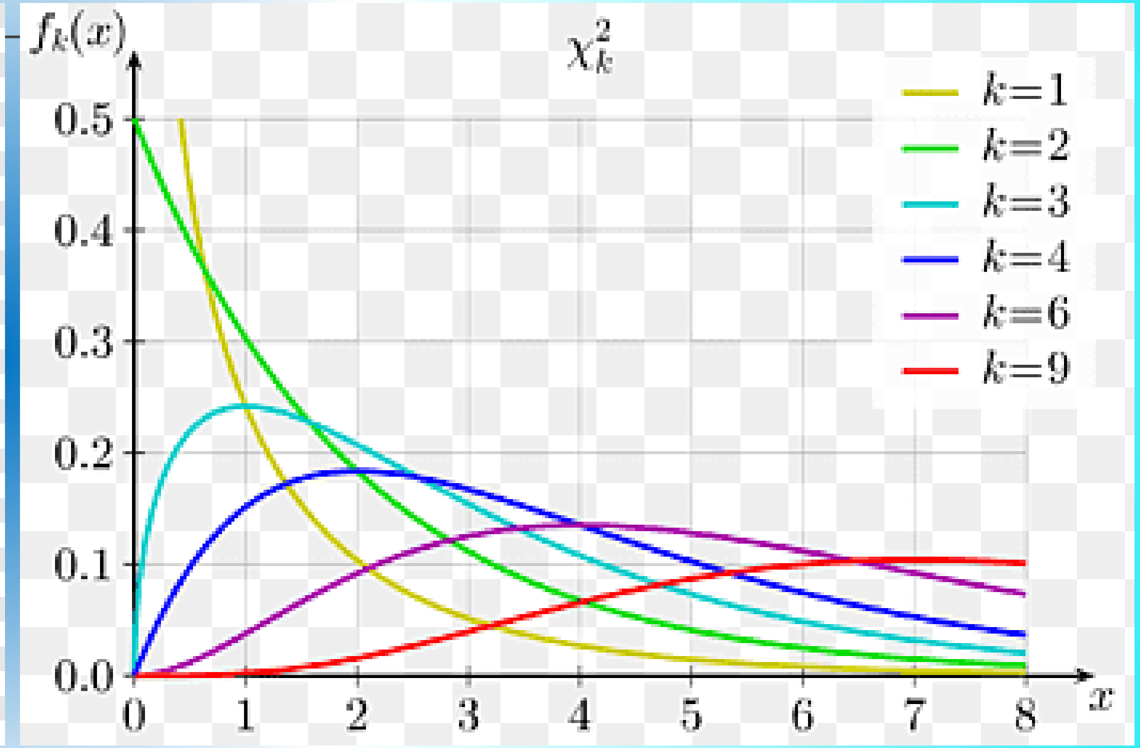
การแจกแจงไคสแควร์ (Chi-Square(χ^2) Test)

การแจกแจงไคสแควร์เป็นการแจกแจงความน่าจะเป็นของผลรวมของตัวแปรสุ่มปกติยกกำลังสองอิสระ k ตัว โดยมีลักษณะเป็นพารามิเตอร์เดียวคือจำนวนองศาอิสระ (degrees of freedom, k)

เส้นกราฟ มีลักษณะไม่สมมาตร มีเฉพาะค่าบวก ถ้าค่า degrees of freedom, kเพิ่มขึ้นจะลดความเบ้ลง

การแจกแจงไคสแควร์ (Chi-Square(χ^2) Test)





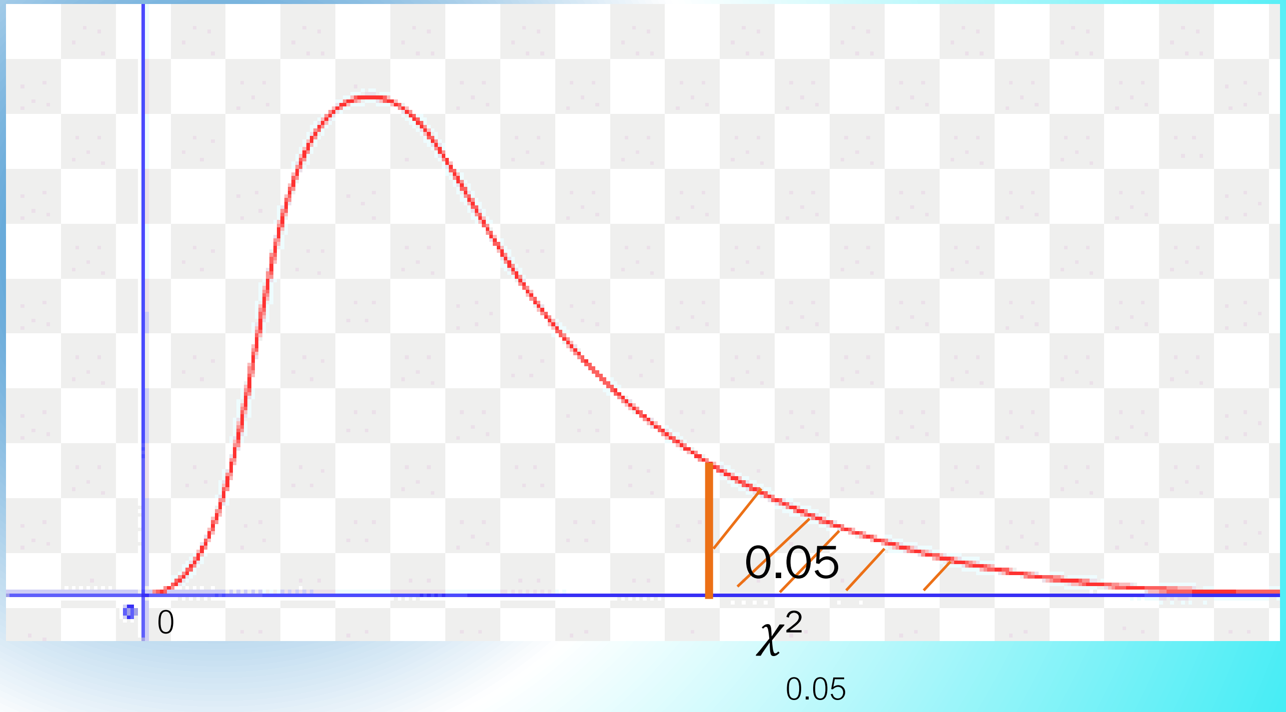
ระดับขององศาอิสระ(df) มีค่ามากขึ้นจะทำให้เส้นกราฟไคสแควร์จะเข้าใกล้เส้นกราฟ Z มากขึ้น

การแจกแจงไคสแควร์ (Chi-Square(χ^2) Test)

ค่าไคสแควร์ $\chi^2_{0.05}$

จะเท่ากับพื้นที่ใต้โค้ง

0.05 ที่แรเงา ด้านขวา

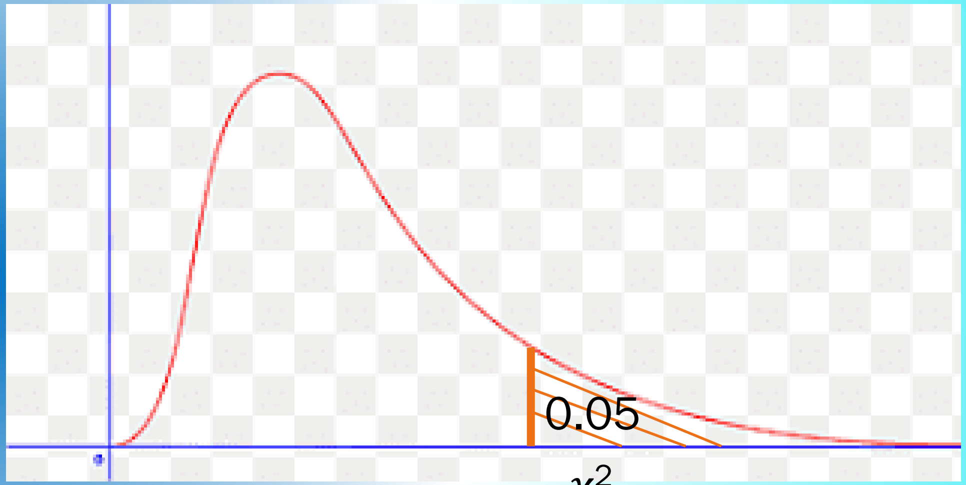


การแจกแจงไคสแควร์ (Chi-Square(χ^2) Test)

$\chi^2_{0.05,17}$ =? เปิดตารางไคสแควร์ที่ df =17

และ α = 0.05 จะได้ค่า

ไคสแควร์ = 27.5871



$\chi^2_{0.05,17}$ =27.5871

หาค่า P(χ^2 >3.49); df = 8

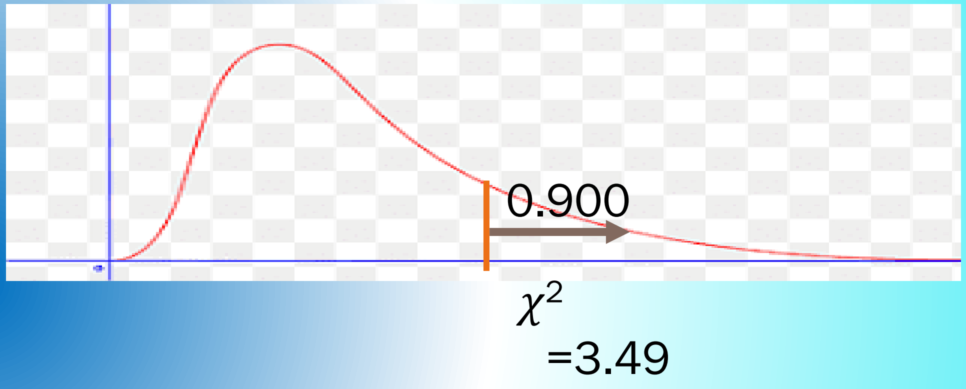
จะได้ ค่า α = 0.900

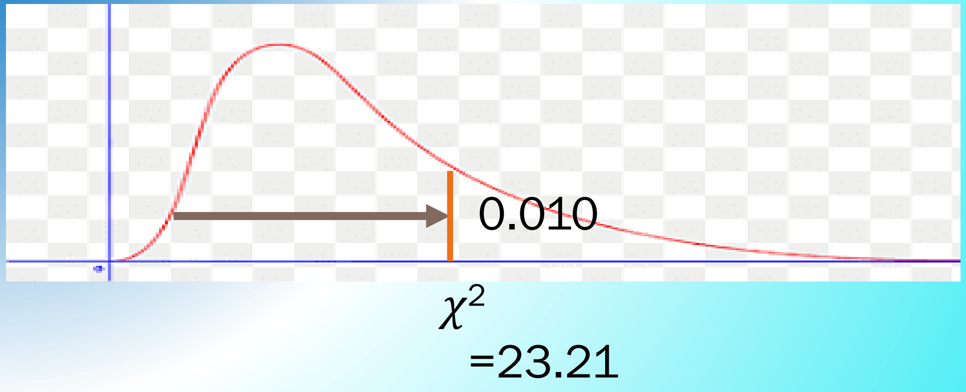
ค่าความน่าจะเป็นที่ χ^2 >3.49 ที่องศาเสรี 8

จะเท่ากับ 0.9

หาค่า P(χ^2 <23.21); df = 10

= 1- 0.010 = 0.99

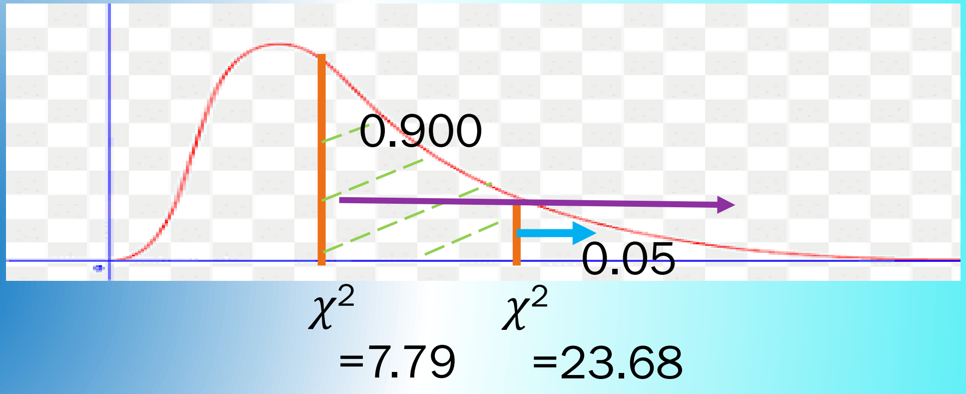




การแจกแจงไคสแควร์ (Chi-Square(χ^2) Test)

หาค่า $P(7.79<\chi^2 <23.68)$; df = 14

$0.90-0.05 = 0.85$



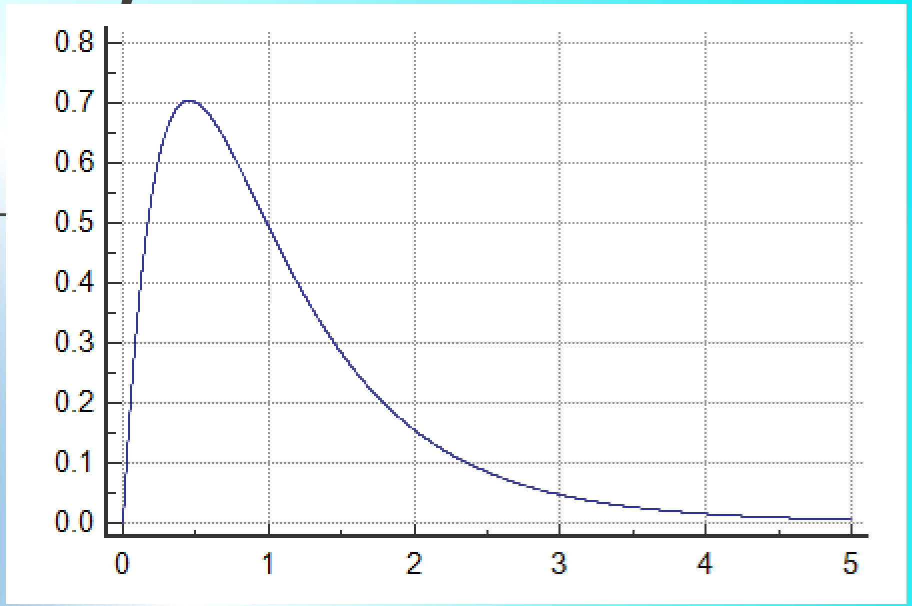
การแจกแจงไคสแควร์ (Chi-Square(χ2) Test)

https://www.youtube.com/watch?v=P8s5RPj4S0w

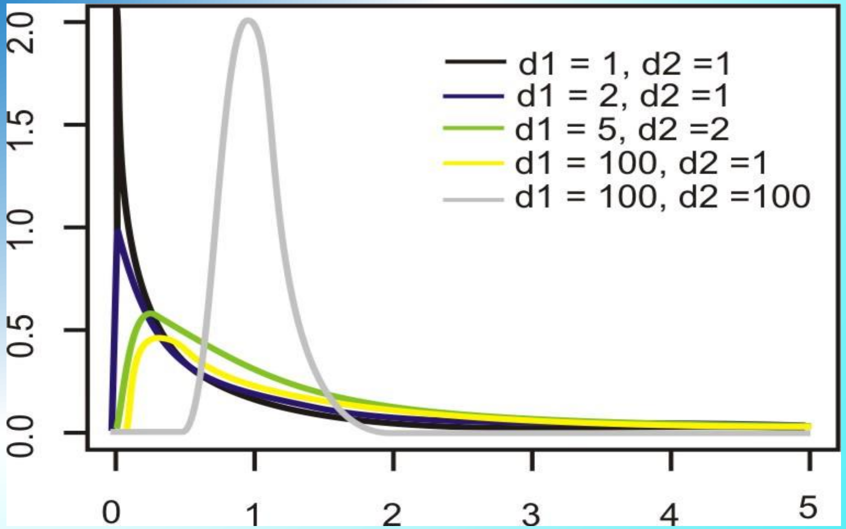
การแจกแจงเอฟ (F - distribution)

ถ้า U และ V เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน โดยที่ $U \sim \chi^2_{(n)}$ และ $V \sim \chi^2_{(m)}$ จะได้ว่า

$F = \dfrac{\frac{U}{n}}{\frac{V}{m}}$ เป็นตัวแปรสุ่มที่มีการแจกแจงแบบ F มี $df_1 = n$ และ $df_2 = m$



- กราฟของฟังก์ชั่นมีลักษณะเบ้ขวา
- พื้นที่ใต้กราฟรวมกันทั้งหมด มีค่าเท่ากับ 1
- $F \geq 0$
- $P(F = f_0) = 0$



การแจกแจงเอฟ (F - distribution)

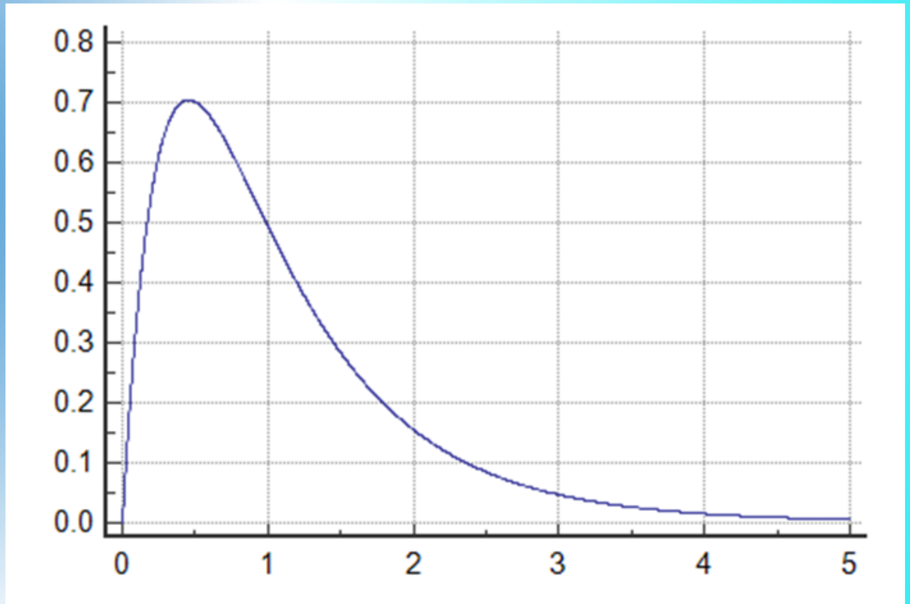
เป็นตัวแปรสุ่ม มีองศาความเสรีเท่ากับ df1 , df2 จงคำนวณค่า

ก. $F_{.10}$ เมื่อ $df_1 = 10$ และ $df_2 = 12$

ข. $F_{.01}$ เมื่อ $df_1 = 20$ และ $df_2 = 30$

เปิดตาราง $F_{.10,10,12} = 2.1878$

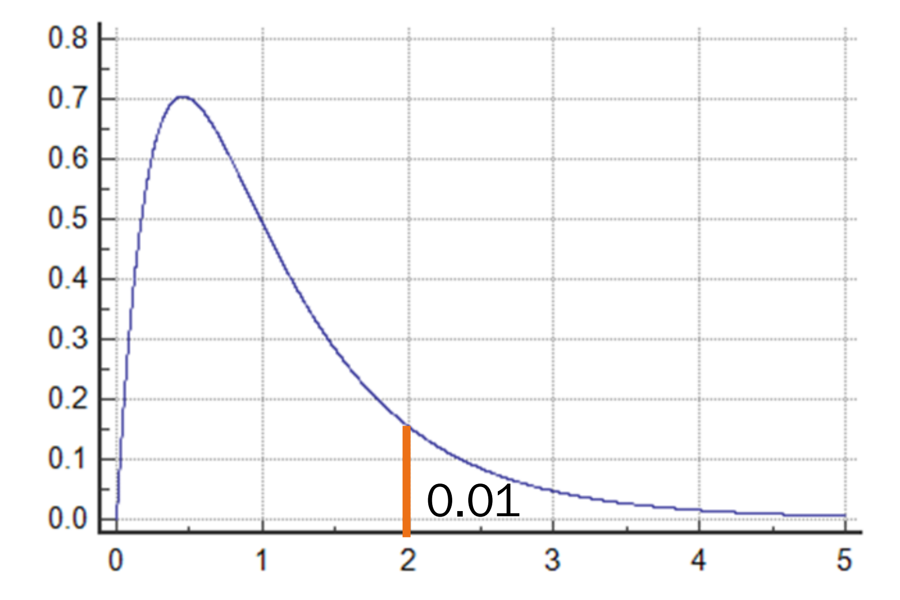
เปิดตาราง $F_{.01, 20, 30} = 2.5487$



การแจกแจงเอฟ (F - distribution)

P(f > 5.61) df_1=7 df_2=9

= 0.01



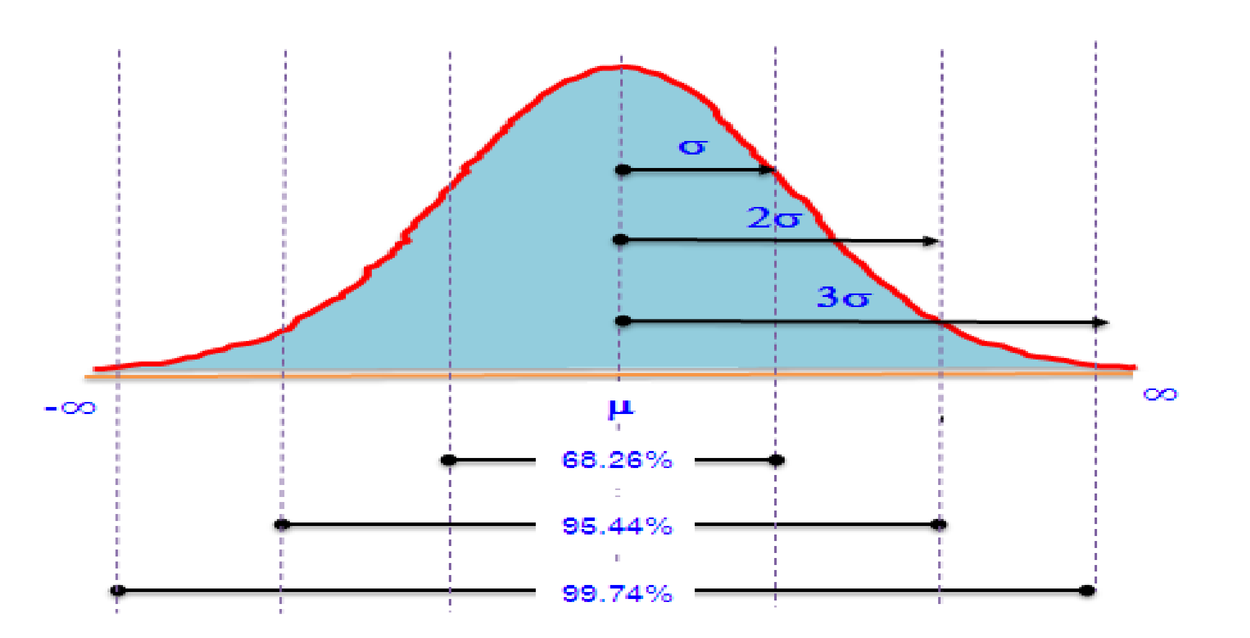
$f_{0.01,7,9}$

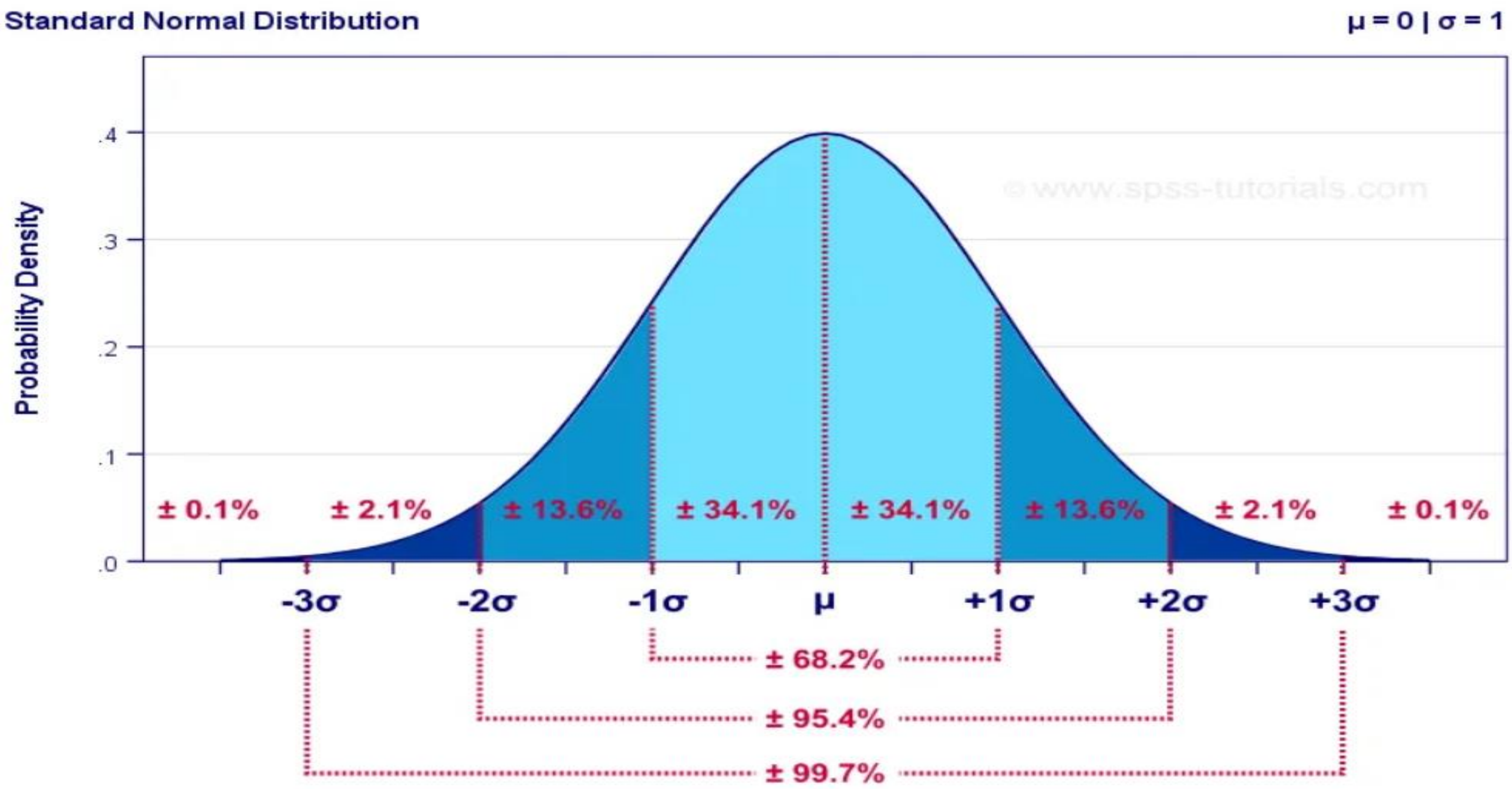
การแจกแจงเอฟ (F - distribution)

https://youtu.be/Be03IsagZvU?t=11

การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

การแจกแจงความน่าจะเป็นแบบปกติมีลักษณะเป็นรูประฆัง (bell-shaped curve) ซึ่งสมมาตร และมีจุดสูงสุดอยู่ตรงกลาง ค่าเฉลี่ย (mean, μ) และส่วนเบี่ยงเบนมาตรฐาน (standard deviation, σ) เป็นพารามิเตอร์ที่กำหนดลักษณะของการแจกแจง



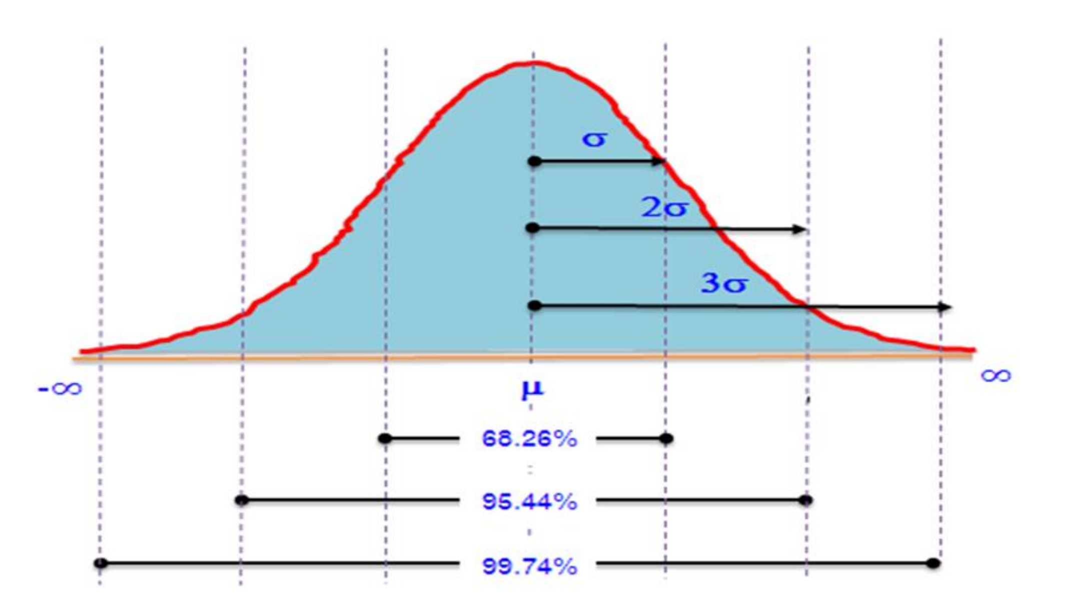


Normal Probability Distribution Example

การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

คุณสมบัติของการแจกแจงปกติสมมาตร:

- การแจกแจงปกติสมมาตรกับค่าเฉลี่ย μ ซึ่งหมายความว่าครึ่งหนึ่งของข้อมูลอยู่ทางด้านซ้ายของ μ และอีกครึ่งหนึ่งอยู่ทางด้านขวา
- ความสูงสุด : ความหนาแน่นสูงสุดเกิดขึ้นที่ค่าเฉลี่ย μ
- ค่าเฉลี่ย มัธยฐาน และฐานนิยม: ในการแจกแจงปกติ ค่าเฉลี่ย (μ), มัธยฐาน (median), และฐานนิยม (mode) จะมีค่าเท่ากัน
- กฎ 68-95-99.7:
  - ประมาณ 68% ของข้อมูลอยู่ในช่วง $\mu \pm \sigma$μ±σ
  - ประมาณ 95% ของข้อมูลอยู่ในช่วง $\mu \pm 2\sigma$μ±2σ
  - ประมาณ 99.7% ของข้อมูลอยู่ในช่วง $\mu \pm 3\sigma$μ±3σ



การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

การแจกแจงปกติมาตรฐานเป็นกรณีพิเศษของการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งเขียนเป็น Z∼N(0,1), X~ normal(μ ,σ^2)

การแปลงคะแนนเป็นคะแนนมาตรฐาน (Z-score)การแปลงคะแนนเป็นคะแนนมาตรฐานช่วยให้สามารถเปรียบเทียบค่าต่าง ๆ จากการแจกแจงปกติที่แตกต่างกันได้โดยใช้การแจกแจงปกติมาตรฐานสูตรในการแปลงคะแนนเป็นคะแนนมาตรฐานคือ

X คือค่าของตัวแปรสุ่ม

μ คือค่าเฉลี่ย

σ คือส่วนเบี่ยงเบนมาตรฐาน

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

การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

สมมติว่ามีการแจกแจงปกติของน้ำหนักของนักเรียนในโรงเรียนหนึ่ง โดยมีค่าเฉลี่ย $\mu=70$ กิโลกรัม และส่วนเบี่ยงเบนมาตรฐาน $\sigma=10$ กิโลกรัมการคำนวณความน่าจะเป็น ต้องการหาความน่าจะเป็นที่นักเรียนคนหนึ่งจะมีน้ำหนักระหว่าง 60 ถึง 80 กิโลกรัม

แปลงค่าเป็น Z-score

:สำหรับ 60 กิโลกรัม:$Z=\dfrac{60-70}{10}=-1$

สำหรับ 80 กิโลกรัม: $Z=\dfrac{80-70}{10}=1$

ใช้ตาราง Z หรือฟังก์ชันในโปรแกรมสถิติ:ความน่าจะเป็นที่ Z อยู่ระหว่าง -1 ถึง 1 จากตาราง Z คือประมาณ 0.6826

ดังนั้น ความน่าจะเป็นที่นักเรียนคนหนึ่งจะมีน้ำหนักระหว่าง 60 ถึง 80 กิโลกรัมคือ 68.26%

การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

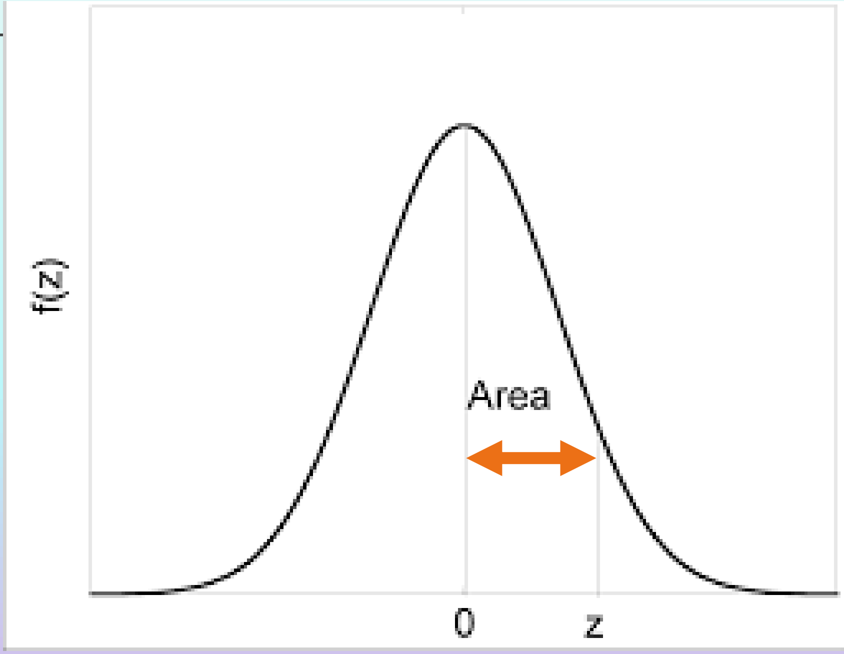
การเปิดตาราง Z ระบุค่า Z ที่ต้องการหา:

ค่า Z ที่ต้องการหาอาจมีทั้งค่าบวกและค่าลบ

แยกส่วนทศนิยมและส่วนเศษส่วน ของค่า Z=1.96

ส่วนทศนิยม: 1.9 ส่วนเศษส่วน: 0.06

ค้นหาค่าความน่าจะเป็นในตาราง = 0.4750



พื้นที่ใต้โค้งทั้งหมด= 1 และ

ด้านซ้ายและขวามีค่า =0.5

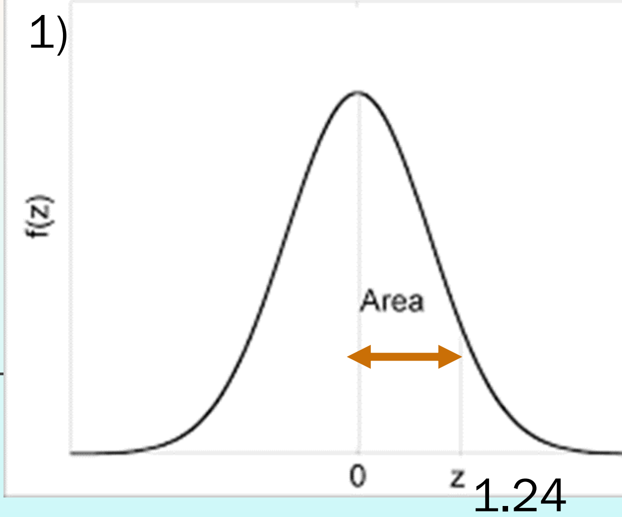
การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

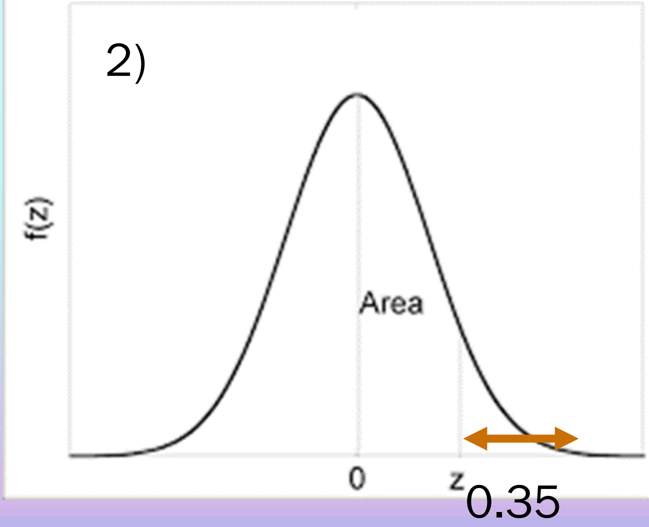
จงหา 1) P(0 <Z<1.24) = 0.3925

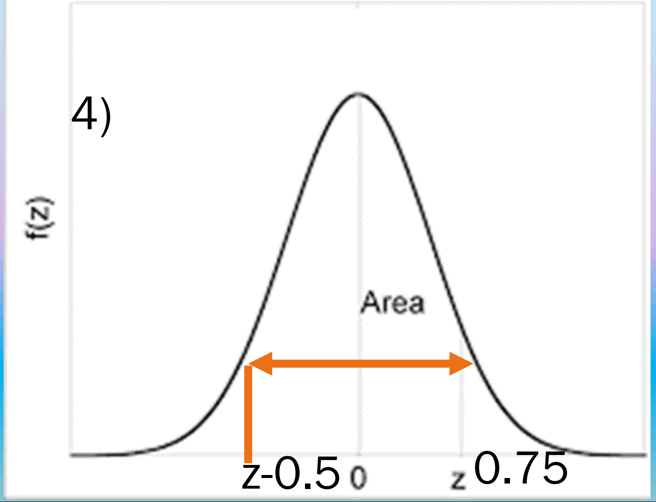
2) P(Z>0.35) = 0.5-0.1368 = 0.3632

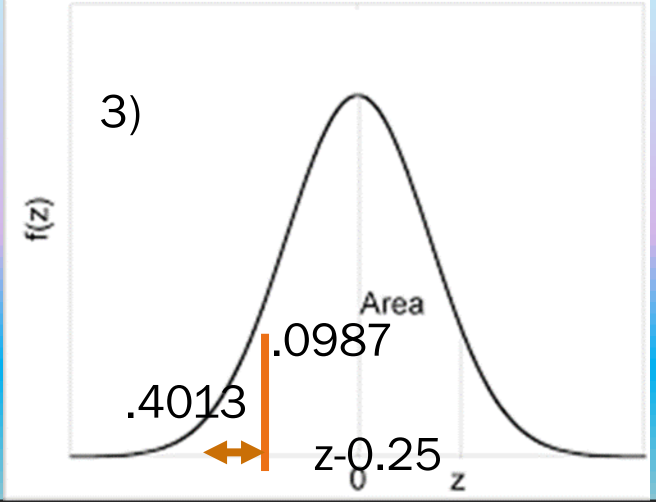
3) P(Z<-0.25) = 0.5- 0.0987 = 0.4013

4) P(-0.5<Z<0.75) = 0.1915+ 0.2734= 0.4649









-0.25

การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

https://www.youtube.com/watch?v=5Lrun372LJg

https://www.youtube.com/watch?v=xgzS5tvB_XY

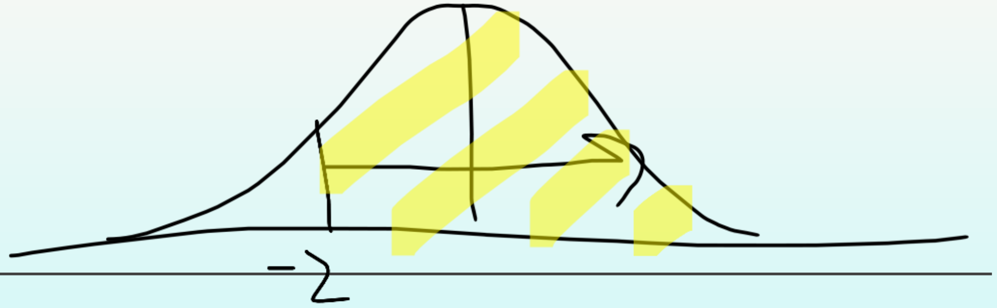
Z คือ ระดับความมั่นใจที่กำหนด หรือระดับนัยสำคัญทางสถิติ แบบ 2-tail

เช่น Z ที่ระดับนัยสำคัญ 0.10 เท่ากับ 1.65 (ความเชื่อมั่น 90%) Z = 1.65

Z ที่ระดับนัยสำคัญ 0.05 เท่ากับ 1.96 (ความเชื่อมั่น 95%) Z = 1.96

Z ที่ระดับนัยสำคัญ 0.01 เท่ากับ 2.58 (ความเชื่อมั่น 99%) Z = 2.58.

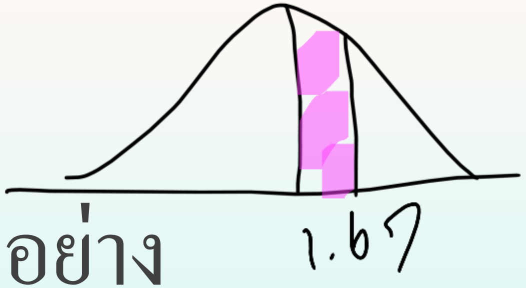
ตัวอย่าง

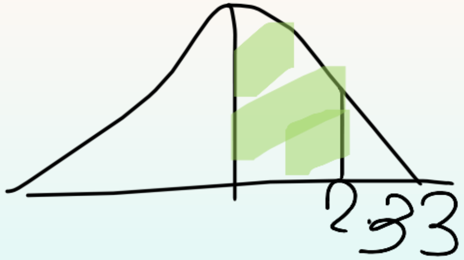


เครื่องจักรของโรงงานแห่งหนึ่งมีอายุการใช้งานโดยเฉลี่ย 5 ปี ส่วน เบี่ยงเบนมาตรฐาน 1ปี ถ้าอายุการใช้งานของเครื่องจักรมีการแจกแจงแบบปกติ จงหาความน่าจะเป็นที่เครื่องจักรนี้จะมีอายุการใช้งานมากกวา 3 ปี X~ normal(μ ,σ^2), μ=5, σ=1

ต้องการหา P(X>3) ต้องเปลี่ยน X เป็น ค่า Z $= \frac{X-\mu}{\sigma} = \frac{3-5}{1}$ =-2

เปิดตาราง Z เพื่อหา P(Z>-2) = 0.5+0.4772 = .9772





ตัวอย่าง

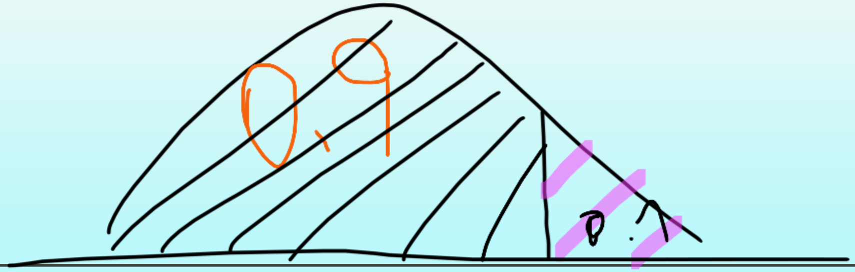
คะแนนสอบของนิสิตในวิชาสถิติมีการแจกแจงปกติด้วยคะแนน เฉลี่ย 60 คะแนนค่าเบี่ยงเบนมาตรฐานเป็น 15 คะแนน อยากทราบว่ามีนิสิตกี่ เปอร์เซ็นต์ที่สอบได้คะแนนตั้งแต่ 85-95 คะแนน X~ normal(μ ,σ^2), μ=60, σ=15

ต้องการหา P(85<X<95) ต้องแปลงX เป็นค่าZ $\left(\frac{85-60}{15} < \frac{X-\mu}{\sigma} < \frac{95-60}{15}\right)$

=(1.67< Z< 2.33)= 0.4901- 0.4525= .0376 ประมาณ 4%

นำพื้นที่ส่วนสีเขียวออกจากพื้นที่ส่วนสีชมพูจะได้พื้นที่สีเหลือง

ตัวอย่าง



ถ้าอาจารย์ผู้สอนวิชาสถิติ ให้Aแก่นักศึกษาที่สอบได้คะแนนสูงสุด 10%ของ ห้อง อยากทราบว่านักศึกษาจะต้องใช้คะแนนอย่างน้อยเท่าใดจึงจะได้เกรด A , วิธีทำ ให้Xเป็นคะแนนสอบวิชาสถิติ ดังนั้น $X \sim normal\ (60,15^2)$ $\mu=60$, $\sigma=15$ ให้ a เป็นคะแนนต่ำสุดของการให้เกรด

หา $P(X \geq a) = 0.10$, $P(Z \geq \frac{a-60}{15}) = 0.10$, $P(Z < \frac{a-60}{15}) = 0.90 = 1.28$,

$\frac{a-60}{15} = 1.28$, $a = 1.28(15) + 60 = 79.2$คะแนน

ต้องได้คะแนนอย่างน้อย 79.2 คะแนนจึงจะได้เกรด A

z	0	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	0.7704	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

ตัวอย่าง

การผลิตไอศครีมขาย โดยกำหนดให้มีน้ำตาลในไอศครีมแต่ละกล่องเป็น18- 22% โดยที่บริษัทมีโรงงานผลิตไอศครีม 2 โรงงานคือ โรงงาน A มีน้ำตาลในไอศครีมเฉลี่ยต่อกล่อง 20%ค่าเบี่ยงเบนมาตรฐาน 2% โรงงานB มีน้ำตาลในไอศกรีมเฉลี่ยต่อกล่อง19%ค่าเบี่ยงเบนมาตรฐาน1%

ถ้าปริมาณน้ำตาลในไอศครีมแต่ละกล่องมีการแจกแจงแบบปกติ อยากทราบว่า โรงงานใดสามารถทำไอศกรีมได้ตามที่ กำหนดดีกว่าอีกโรงงานหนึ่ง

โรงงาน A : X~normal (μ=20%, σ^2=4 %)

โรงงาน B : X~normal (μ=19%, σ^2=1 %)

การเปรียบเทียบ $P(18\% \leq X \leq 22\%)$ ของโรงงาน A และ B

โรงงาน A , $P(18\% \leq X \leq 22\%) = P(\frac{18-20}{2} \leq Z \leq \frac{22-20}{2}) = P{-1} \leq Z \leq 1)$

$0.3413 + 0.3413 = .6826$

โรงงาน B , $P(18\% \leq X \leq 22\%) = P(\frac{18-19}{1} \leq Z \leq \frac{22-19}{1}) = P{-1} \leq Z \leq 3)$

$0.3413 + 0.4987 = .84$,โรงงาน A 68.26% โรงงาน B 84%

โจทย์ที่ 1:

บริษัท A ผลิตสินค้าที่มีน้ำหนักเฉลี่ย 500 กรัม และมีส่วนเบี่ยงเบนมาตรฐาน 10 กรัม น้ำหนักของสินค้าจะมีการแจกแจงแบบปกติ

1.1 หากสุ่มหยิบสินค้าหนึ่งชิ้นมาจากสายการผลิต จะมีความน่าจะเป็นเท่าใดที่สินค้าชิ้นนั้นจะมีน้ำหนักน้อยกว่า 490 กรัม?

1.2 ความน่าจะเป็นที่สินค้าชิ้นนั้นจะมีน้ำหนักอยู่ระหว่าง 495 ถึง 505 กรัมคือเท่าใด?

โจทย์ที่ 2:

บริษัท B ได้รับผลตอบแทนจากการลงทุน (ROI) เฉลี่ยต่อเดือนเท่ากับ 2% และมีส่วนเบี่ยงเบนมาตรฐาน 0.5% ROI ของบริษัท B มีการแจกแจงแบบปกติ

2.1 ความน่าจะเป็นที่ ROI ในเดือนถัดไปจะน้อยกว่า 1.5% คือเท่าใด?

2.2 หากบริษัท B ต้องการให้มีความมั่นใจ 95% ว่า ROI จะไม่ต่ำกว่าระดับที่กำหนด ควรกำหนด ROI ขั้นต่ำไว้ที่เท่าใด?

โจทย์ที่ 3:

เวลาที่ใช้ในการผลิตสินค้าหนึ่งชิ้นของบริษัท C มีค่าเฉลี่ย 30 นาที และส่วนเบี่ยงเบนมาตรฐาน 4 นาที เวลาที่ใช้ในการผลิตมีการแจกแจงแบบปกติ

3.1 ความน่าจะเป็นที่เวลาผลิตสินค้าหนึ่งชิ้นจะน้อยกว่า 25 นาทีคือเท่าใด?

3.2 หากเวลาผลิตอยู่ระหว่าง 28 ถึง 35 นาที ความน่าจะเป็นที่จะเกิดเหตุการณ์นี้คือเท่าใด?

โจทย์ที่ 4:

บริษัท D ผลิตสินค้าที่มีค่าเฉลี่ยการใช้งานอยู่ที่ 1000 ชั่วโมง และมีส่วนเบี่ยงเบนมาตรฐาน 100 ชั่วโมง อายุการใช้งานของสินค้าจะมีการแจกแจงแบบปกติ

4.1 ความน่าจะเป็นที่สินค้าหนึ่งชิ้นจะมีอายุการใช้งานมากกว่า 1100 ชั่วโมงคือเท่าใด?

4.2 หากบริษัท D ต้องการโฆษณาว่าสินค้าของตนมีอายุการใช้งานน้อยกว่า x ชั่วโมงไม่เกิน 5% ควรระบุค่า x ไว้ที่เท่าใด?

การแก้โจทย์เหล่านี้สามารถทำได้โดยการแปลงค่าตัวแปรต้นฉบับเป็นค่า Z-score และใช้ตาราง Z หรือโปรแกรมสถิติคำนวณค่าความน่าจะเป็น

ดังนี้:แปลงค่า X เป็น Z-score:

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

เมื่อX คือค่าที่ต้องการ, μ คือค่าเฉลี่ย, และ σ คือส่วนเบี่ยงเบนมาตรฐาน

ใช้ตาราง Z คำนวณค่าความน่าจะเป็น

โจทย์ 1.1= 0.1587 , 1.2 =0.3830

โจทย์ 2.1= 0.1587 , 2.2 =2.98

โจทย์ 3.1= 0.1056 , 3.2 =0.5859

โจทย์ 1.1= 0.1587 , 1.2 =0.3830

บทที่ 6 การประมาณค่า

- การประมาณค่าแบบจุด
- การประมาณค่าแบบช่วง
- สรุป

วัตถุประสงค์การเรียนรู้

- **อธิบายแนวคิดพื้นฐานของการประมาณค่า** ได้ทั้งการประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation)
- **คำนวณการประมาณค่าแบบจุด** ของค่าเฉลี่ย ประชากร (μ) และสัดส่วนประชากร (p) ได้จากข้อมูลตัวอย่าง
- **สร้างช่วงความเชื่อมั่น (Confidence Interval)** สำหรับค่าเฉลี่ยและสัดส่วนประชากร ภายใต้เงื่อนไขที่เหมาะสม (กรณีทราบและไม่ทราบค่าเบี่ยงเบนมาตรฐานประชากร)

บทนำ

การประมาณค่าเป็นกระบวนการสำคัญที่ช่วยในการสรุปและตัดสินใจจากข้อมูลตัวอย่างที่มีอยู่ โดยทั่วไปการประมาณค่าสามารถแบ่งออกได้เป็นสองวิธีหลัก ๆ คือ การประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation) การประมาณค่าแบบจุดคือการเลือกค่าตัวเลขหนึ่งค่าจากข้อมูลตัวอย่างมาเป็นตัวแทนของค่าพารามิเตอร์ที่ต้องการในประชากร

ประโยชน์ของการประมาณค่า

1. การตัดสินใจและการวางแผน
2. การวิเคราะห์ข้อมูลและการสรุปผล
3. การทำนายและคาดการณ์
4. การสนับสนุนการวิจัยและพัฒนา

การประมาณค่าแบบจุด (Point Estimation)

- ค่าเฉลี่ยตัวอย่าง (Sample Mean, $\bar{x}$)
- สัดส่วนตัวอย่าง (Sample Proportion, $\hat{p}$)

ค่าเฉลี่ยตัวอย่าง (Sample Mean, $\bar{x}$)

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

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

สัดส่วนตัวอย่าง (Sample Proportion, $\hat{p}$)

การคำนวณสัดส่วนจากขอมูลตัวอย่างเพื่อประมาณสัดส่วนประชากร (p) เช่น ในการสำรวจความคิดเห็นของผู้คน 100 คนเกี่ยวกับผลิตภัณฑ์หนึ่ง หาก 60 คนตอบว่าชอบผลิตภัณฑ์นั้น สัดส่วนตัวอย่าง ($\hat{p}$) จะคำนวณได้ดังนี้

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

ดังนั้น 0.6 คือการประมาณค่าแบบจุดของสัดส่วนประชากร

การประมาณค่าแบบช่วง

การประมาณค่าแบบช่วง (Interval Estimation) เป็นวิธีการในสถิติที่ใช้เพื่อคาดการณ์ค่าพารามิเตอร์ของประชากรโดยการกำหนดช่วงของค่าที่มีความเป็นไปได้ว่าจะครอบคลุมค่าพารามิเตอร์นั้น พร้อมกับระดับความเชื่อมั่นที่กำหนด

ตัวอย่างการประมาณค่าแบบช่วง

การหาช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยหรือการประมาณค่าเฉลี่ยประชากร (μ)

- กรณีกลุ่มตัวอย่างมากกว่าหรือเท่ากับ 30
- กรณีกลุ่มตัวอย่างน้อยกว่า 30

กรณีกลุ่มตัวอย่างมากกว่าหรือเท่ากับ 30

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%

ขั้นตอนการคำนวณช่วงความเชื่อมั่น

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

1.96 is used because the 95% confidence interval has only 2.5% on each side. The probability for a z score below -1.96 is 2.5%, and similarly for a z score above $+1.96$; added together this is 5%. 1.64 would be correct for a 90% confidence interval, as the two sides (5% each) add up to 10%

ตัวอย่าง

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

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

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

CI= [95.85,104.15] ช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยประชากรจะอยู่ระหว่าง 95.85 ถึง 104.15

กรณีกลุ่มตัวอย่างน้อยกว่า 30

การประมาณค่าเฉลี่ยของประชากรที่ไม่สามารถหาค่าความแปรปรวนของประชากรได้และขนาดกลุ่มตัวอย่างน้อยกว่า 30 เพราะค่าใช้จ่ายในการรวบรวมข้อมูลไม่เอื้ออำนวยที่จะหาขนาดกลุ่มตัวอย่างใหญ่เท่าที่ต้องการได้ จึงต้องพิจารณาว่าประชากรชุดที่เกี่ยวข้องจะต้องมีรูปร่างโดยประมาณเป็นรูประฆัง ช่วงความเชื่อมั่นจะหาได้โดยอาศัยการแจกแจงความน่าจะเป็นของสถิติ t

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

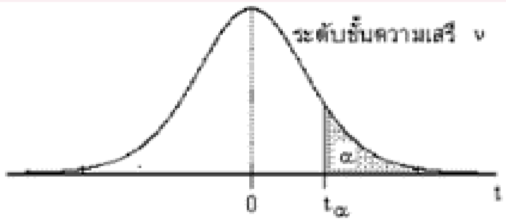
$$60.2 \pm 2.262 \cdot \frac{6.31}{\sqrt{10}} = 60.2 \pm 4.52$$

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

ตารางที่ 4. **ตารางค่าวิกฤตของการแจกแจง t**

เมื่อกำหนดระดับขั้นความเสรี v และค่า α

ค่าจากตารางคือ t_α คือค่าที่ทำให้ $P(t_\alpha < T < \infty) = \alpha$



v	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 ± 0.061

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

คุณสมบัติของตัวประมาณที่ดี

1.ไม่มีอคติ (Unbiased): ตัวประมาณควรมีค่าคาดหวังเท่ากับค่าพารามิเตอร์ที่ต้องการ $\bar{x}$ เท่ากับ μ

2.มีประสิทธิภาพ (Efficient): ตัวประมาณควรมีความแปรปรวนน้อยที่สุดเมื่อเทียบกับตัวประมาณอื่นๆ

3.มีความสม่ำเสมอ (Consistent): ตัวประมาณควรมีแนวโน้มที่จะเข้าใกล้ค่าพารามิเตอร์ที่แท้จริงเมื่อขนาดตัวอย่างเพิ่มขึ้น

สรุป

การประมาณค่าเป็นกระบวนการสำคัญที่ช่วยในการสรุปและตัดสินใจจากข้อมูลตัวอย่างที่มีอยู่ โดยทั่วไป การประมาณค่าสามารถแบ่งออกได้เป็นสองวิธีหลัก ๆ คือ การประมาณค่าแบบจุดและการประมาณค่าแบบช่วง การประมาณค่าแบบจุดคือการเลือกค่าตัวเลขหนึ่งค่าจากข้อมูลตัวอย่างมาเป็นตัวแทนของค่าพารามิเตอร์ที่ต้องการในประชากร ขณะที่การประมาณค่าแบบช่วงจะให้ช่วงของค่าที่เป็นไปได้สำหรับพารามิเตอร์นั้น โดยมีระดับ ความเชื่อมั่นที่กำหนด

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

บทที่ 7 การทดสอบสมมติฐาน

- การทดสอบสมมติฐานทางสถิติ
- การตีความผลการทดสอบ
- สรุป

วัตถุประสงค์การเรียนรู้

- อธิบายแนวคิดและความสำคัญของการทดสอบสมมติฐาน ได้อย่างถูกต้อง
- **ระบุและตั้งสมมติฐานทางสถิติ** ได้ทั้งสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1)
- **อธิบายความหมายของความคลาดเคลื่อน** (Type I Error และ Type II Error) และระดับนัยสำคัญ (α)

บทนำ

การทดสอบสมมติฐานเป็นกระบวนการทางสถิติที่สำคัญในการวิเคราะห์ข้อมูลและการตัดสินใจในหลากหลายสถานการณ์ เช่น การวิจัยทางการแพทย์ การพัฒนาผลิตภัณฑ์ใหม่ หรือการปรับปรุงกระบวนการทางธุรกิจ โดยมีวัตถุประสงค์เพื่อทำให้สามารถตัดสินใจเกี่ยวกับประเด็นหรือสรุปผลลัพธ์ที่มีการสนับสนุนหรือไม่สนับสนุนสมมติฐานนั้น ๆ อย่างมั่นใจและมีเหตุผล

สมมติฐาน

สมมติฐาน (Hypothesis) คือการสร้างข้อความหรือการประมาณค่าที่เกี่ยวข้องกับปัญหาหรือสถานการณ์ที่ต้องการศึกษาหรือวิจัย เพื่อที่จะทำการทดสอบหรือตรวจสอบความถูกต้องของข้อความนั้นๆ โดยใช้ข้อมูลที่มีอยู่

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

ความผิดพลาดในการตัดสินใจ

เป็นความผิดพลาดที่เกิดขึ้นจากการตัดสินใจยอมรับหรือปฏิเสธสมมติฐานที่ผิด มี 2 ประเภท ดังนี้

1 ความผิดพลาดประเภทที่ 1 โอกาสที่จะเกิดความผิดพลาดประเภท 1 หรือความน่าจะเป็นที่จะปฏิเสธสมมติฐานหลักจริง

เรียกว่า ระดับความมีนัยสำคัญ ของการทดสอบ โดยปกติมักกำหนด $\alpha = 0.01, 0.05$ หรือ 0.10 โดยที่ $\alpha = 0.10$ หมายความว่า ในการทดลอง 100 ครั้ง จะมี 10 ครั้ง ที่เกิดความผิดพลาดประเภทที่ 1 ขึ้น

2 ความผิดพลาดประเภทที่ 2 โอกาสที่จะเกิดความผิดพลาดประเภท 2 หรือความน่าจะเป็นที่จะยอมรับสมมติฐานหลักโดยที่สมมติฐานหลักไม่เป็นจริง

ความผิดพลาดในการตัดสินใจ

การตัดสินใจสำหรับสมมติฐานหลัก	ความจริง จริง	สมมติฐานหลักเป็น ไม่จริง
ยอมรับ	ตัดสินใจถูก (1 - α)	ตัดสินใจผิด (Type II Error : β)
ปฏิเสธ	ตัดสินใจผิด (Type I Error: α)	ตัดสินใจถูก (1 - β)

https://statisticapp.weebly.com/

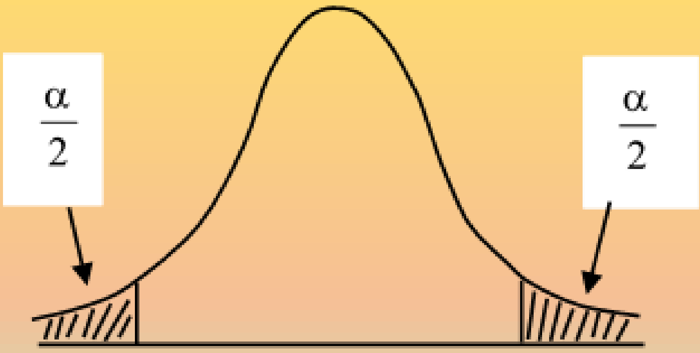
ประเภทของสมมติฐาน

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

2. สมมติฐานทดสอบ (Alternative Hypothesis, $H1$ หรือ Ha): เป็นสมมติฐานที่กำหนดความแตกต่างหรือผลกระทบของตัวแปรอิสระต่อตัวแปรตาม ซึ่งเป็นส่วนที่สนใจในการวิจัย เช่น $H1 : \mu \neq 50$ หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่กำลังศึกษามีค่ามากกว่า 50

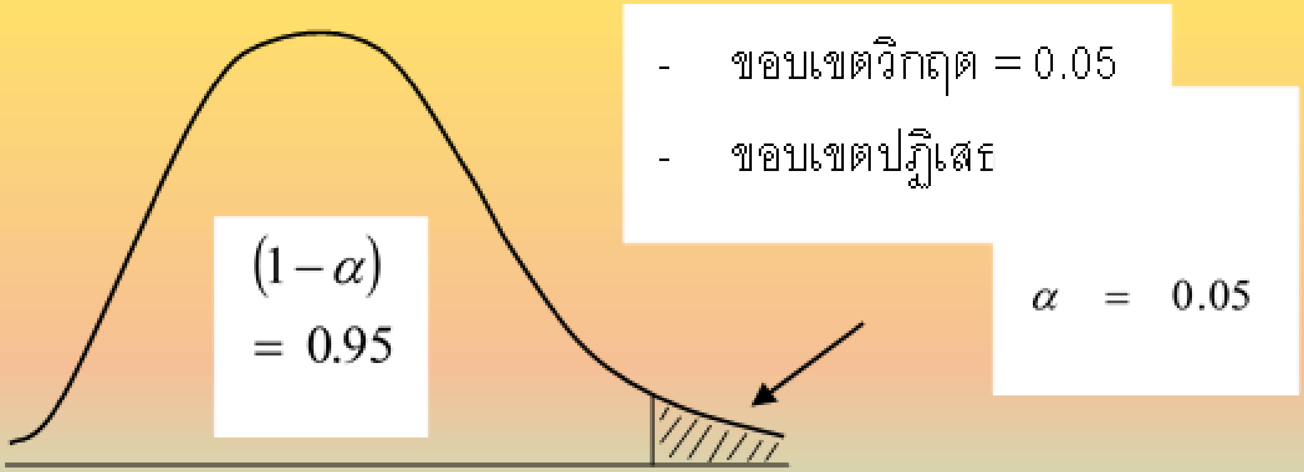
ประเภทของสมมติฐาน

3.สมมติฐานสองทาง (Two-Tailed Hypothesis): เป็นสมมติฐานที่กำหนดความแตกต่างหรือผลกระทบของตัวแปรอิสระต่อตัวแปรตามทั้งสองทิศทาง เช่น H1 : $\mu \neq 50$ หมายความว่า ค่าเฉลี่ยของกลุ่มหรือประชากรที่กำลังศึกษาไม่เท่ากับ 50



ประเภทของสมมติฐาน

4. สมมติฐานทางเดียว (One-Tailed Hypothesis) หมายถึงสมมติฐานที่กำหนดค่าของตัวแปรตามที่สนใจให้มากกว่าหรือน้อยกว่าค่าที่กำหนดไว้ โดยมีทิศทางเดียวเท่านั้น



ขั้นตอนการทดสอบสมมติฐานทางสถิติ

1. ตั้งสมมติฐาน (Formulate Hypotheses)

สมมติฐานว่าง (null hypothesis, H_0) และสมมติฐานที่เรียกว่าสมมติฐานทดสอบ (alternative hypothesis, H_1 หรือ H_a).

$H_0 : \mu$</=50

$H_1 : \mu > 50$

ขั้นตอนการทดสอบสมมติฐานทางสถิติ

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

ขั้นตอนการทดสอบสมมติฐานทางสถิติ

3. กำหนดระดับความสำคัญ (Set Significance Level):

ระดับความสำคัญ (significance level) บ่งบอกถึงความเสี่ยงที่ยอมรับในการทำนิยามผิดพลาดเพื่อปฏิเสธสมมติฐานที่เสนอ ระดับความสำคัญที่ใช้มักเป็น 0.05 หรือ 0.01.

ขั้นตอนการทดสอบสมมติฐานทางสถิติ

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

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

ขั้นตอนการทดสอบสมมติฐานทางสถิติ

5. ตีความผลการทดสอบ (Interpret Results):

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

6. สรุปผลลัพธ์ (Conclusion):

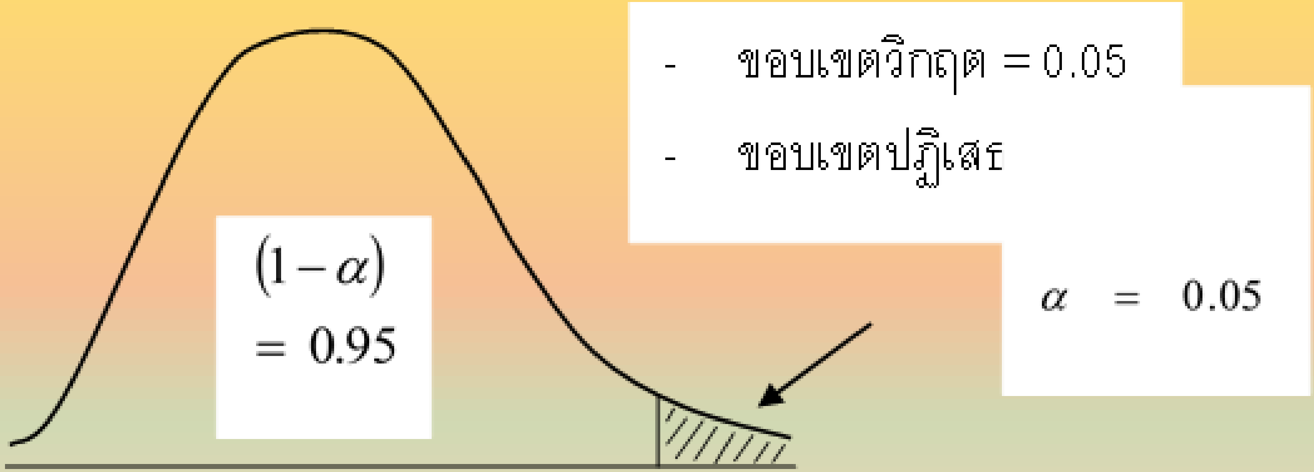
สรุปผลลัพธ์ของการทดสอบสมมติฐานโดยรวมและแสดงความเชื่อมั่นในการสรุป และอธิบายผลกระทบหรือความสำคัญของผลลัพธ์ต่องานหรือวิชาที่เกี่ยวข้อง

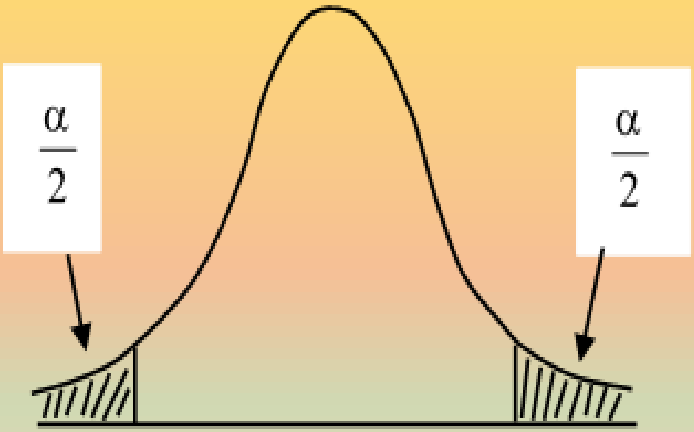
การตีความผลการทดสอบ

หากค่าสถิติที่คำนวณได้มีค่ามากกว่าค่าวิกฤตที่ระดับนัยสำคัญ .05 หมายความว่า สมมติฐานที่ทดสอบมีนัยสำคัญทางสถิติ

สำหรับการวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติให้พิจารณาจากค่า P-value หรือค่า sig น้อยกว่าค่านัยสำคัญทางสถิติ ที่ระบุไว้หมายความว่า สมมติฐานที่ทดสอบมีนัยสำคัญทางสถิติ

การทดสอบสมมติฐาน คือการหาข้อสรุปว่า**สมมติฐานที่สนใจนั้นคือ สมมติฐานว่าง (H_0) จะถูกยอมรับหรือปฏิเสธ ถ้าถูกยอมรับ** นั่นก็แปลว่า สมมติฐานทดสอบที่ตั้งไว้เป็นจริง **หากว่าสมมติฐานทดสอบถูกปฏิเสธแปลว่าสมมติฐานนั้นไม่เป็นจริง**





One-sample t-test

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

การทดสอบนี้ใช้เพื่อเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างหนึ่งกลุ่มกับค่าเฉลี่ยที่เป็นค่าคงที่ที่กำหนดขึ้น (เช่น ค่าเฉลี่ยของประชากรที่คาดหวัง) $H_0 : \mu=\mu_0$, $H_1 : \mu\neq\mu_0$

สมมติว่าต้องการทดสอบว่านักเรียนกลุ่มหนึ่งมีค่าเฉลี่ยคะแนนสอบเท่ากับ 75 หรือไม่ โดยมีข้อมูลดังนี้

ค่าเฉลี่ยของกลุ่มตัวอย่าง ($\bar{x}$) = 78 , ค่าคงที่ที่ต้องการเปรียบเทียบ ($\mu 0$) = 75, ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (s) = 10 , จำนวนตัวอย่าง (n) = 30

H0 : $\mu=75$, H1 : $\mu\neq75$, $t = \frac{78-75}{\frac{10}{\sqrt{30}}} = \frac{3}{1.83} \approx 1.64$

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

ถ้าค่า t วิกฤตที่ระดับความเชื่อมั่น 0.05 สำหรับ df=29 คือ 2.0452

เนื่องจาก $|1.64|<2.0452$ จึงไม่สามารถปฏิเสธสมมติฐาน H_0 ได้ ซึ่งหมายความว่าค่าเฉลี่ยของนักเรียนกลุ่มนี้ไม่แตกต่างจาก 75 อย่างมีนัยสำคัญ

ตาราง t

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

บทที่ 8 การวิเคราะห์ความแปรปรวน

- การวิเคราะห์การแจกแจงความแปรปรวน
- การใช้เครื่องมือสถิติในการวิเคราะห์ข้อมูลแปรปรวน
- สรุป

วัตถุประสงค์การเรียนรู้

- อธิบายความหมายและหลักการของการวิเคราะห์ความแปรปรวน (ANOVA) ได้อย่างถูกต้อง
- ระบุวัตถุประสงค์ของการใช้ ANOVA ในการเปรียบเทียบค่าเฉลี่ยของกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป
- อธิบายสมมติฐานของ ANOVA ได้แก่ ความเป็นอิสระของข้อมูล ความเป็นปกติ (Normality) และความเท่ากันของความแปรปรวน (Homogeneity of Variance)
- คำนวณค่า F-test และจัดทำตาราง ANOVA ได้จากข้อมูลตัวอย่าง
- แปลความหมายของค่า F และค่า p-value เพื่อสรุปผลการทดสอบสมมติฐาน

บทนำ

การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เป็นวิธีการทางสถิติที่ใช้ในการเปรียบเทียบค่าเฉลี่ยของกลุ่มข้อมูลหลายกลุ่ม เพื่อพิจารณาว่าค่าเฉลี่ยของกลุ่มใดแตกต่างกันอย่างมีนัยสำคัญหรือไม่ การวิเคราะห์นี้มีความสำคัญอย่างยิ่งในงานวิจัยที่ต้องการตรวจสอบความแตกต่างระหว่างกลุ่มต่าง ๆ เช่น การเปรียบเทียบประสิทธิภาพของยาหลายชนิด การวิเคราะห์ผลการเรียนของนักเรียนในห้องเรียนต่าง ๆ หรือ การเปรียบเทียบคุณภาพของผลิตภัณฑ์จากแหล่งที่มาต่าง ๆ

เงื่อนไขของการวิเคราะห์ความแปรปรวน

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

ขั้นตอนในการวิเคราะห์การแจกแจงความแปรปรวน

1.กำหนดสมมติฐาน

- สมมติฐานว่าง (Null Hypothesis, H0): ค่าเฉลี่ยของทุกกลุ่มไม่มีความแตกต่างกัน
- สมมติฐานทางเลือก (Alternative Hypothesis, H1): มีค่าเฉลี่ยของกลุ่มใดกลุ่มหนึ่งหรือมากกว่านั้นแตกต่างกัน

ขั้นตอนในการวิเคราะห์การแจกแจงความแปรปรวน

2.คำนวณความแปรปรวน

- ความแปรปรวนภายในกลุ่ม (Within-group variance): วัด ความแปรปรวนของข้อมูลภายในแต่ละกลุ่ม
- ความแปรปรวนนอกกลุ่ม (Between-group variance): วัด ความแปรปรวนระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ

ขั้นตอนในการวิเคราะห์การแจกแจงความแปรปรวน

3.ใช้สถิติ F-test:

- คำนวณค่า F ซึ่งเป็นอัตราส่วนระหว่างความแปรปรวนนอกกลุ่มกับความแปรปรวนภายในกลุ่ม
- เปรียบเทียบค่า F ที่คำนวณได้กับค่าวิกฤต (Critical Value) ที่ระดับความเชื่อมั่นที่กำหนด (เช่น 0.05 หรือ 0.01)

ขั้นตอนในการวิเคราะห์การแจกแจงความแปรปรวน

4.การตัดสินใจ

- หากค่า F ที่คำนวณได้มากกว่าค่าวิกฤต แสดงว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ ปฏิเสธสมมติฐานหลัก
- หากค่า F ที่คำนวณได้น้อยกว่าหรือเท่ากับค่าวิกฤต แสดงว่าไม่มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานหลัก ยอมรับสมมติฐานหลัก

การคำนวณ

	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 ji 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 n1+n2+n3 +…+nk

ตารางการวิเคราะห์ความแปรปรวนทางเดียว (One-Way ANOVA)

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

สูตรคำนวณ

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

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

SST $= \sum_{j}^{k} \sum_{i}^{n} (x_{ji} - \bar{x})^2$

SST = SSB + SSE

การเปรียบเทียบกับค่าวิกฤต

ใช้ค่าวิกฤตจากตาราง F-distribution

ที่ระดับความเชื่อมั่น α=0.05 และ $df_{between}$, df_{within}, เพื่อหาค่าวิกฤต $F_{critical}$

ถ้าค่า F ที่คำนวณได้ > $F_{critical}$ จึงปฏิเสธสมมติฐานว่าง (null hypothesis)

ตัวอย่าง

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

1) คำนวณค่าเฉลี่ย

	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

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)

กลุ่ม A: $\bar{X}_A = 88.0$

กลุ่ม B: $\bar{X}_B = 79.0$

กลุ่ม C: $\bar{X}_C = 92.8$

$$SS_B = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2$$

n_i = จำนวนตัวอย่างในกลุ่ม i

$\bar{X}_i$ = ค่าเฉลี่ยของกลุ่ม i

A

$$\bar{X}_A - \bar{X} = 88.0 - 86.6 = 1.4$$

$$(\bar{X}_A - \bar{X})^2 = 1.4^2 = 1.96$$

$$n_A(\bar{X}_A - \bar{X})^2 = 5 * 1.96 = 9.8$$

B

$$\bar{X}_B - \bar{X} = 79.0 - 86.6 = -7.6$$

$$(\bar{X}_B - \bar{X})^2 = (-7.6)^2 = 57.76$$

$$n_B(\bar{X}_B - \bar{X})^2 = 5 * 57.76 = 288.8$$

C

$$\bar{X}_C - \bar{X} = 92.8 - 86.6 = 6.2$$

$$(\bar{X}_C - \bar{X})^2 = 6.2^2 = 38.44$$

$$n_C(\bar{X}_C - \bar{X})^2 = 5 * 38.44 = 192.2$$

$$SS_B = 9.8 + 288.8 + 192.2 = 490.8$$

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

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

3.คำนวณความแปรปรวนภายในกลุ่ม (Within-group variance)

$SS_{within} = \sum (X_{ji} - \bar{X}_i)^2 = \sum ((85-88)^2+(90-88)^2+\ldots+(93-92.8)^2)$

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

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

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

$= 60.8$

$df_{within} = N - k$

$= 15-3 = 12$

$\mathrm{MSE}/MS_{within} =$

SS_{within}/df_{within}

$= 60.8/12 = 5.067$

4) คำนวณค่า F

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

$$= \frac{245.4}{5.067} = \boxed{48.43}$$

5)ใช้ค่าวิกฤตจากตาราง F-distribution

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

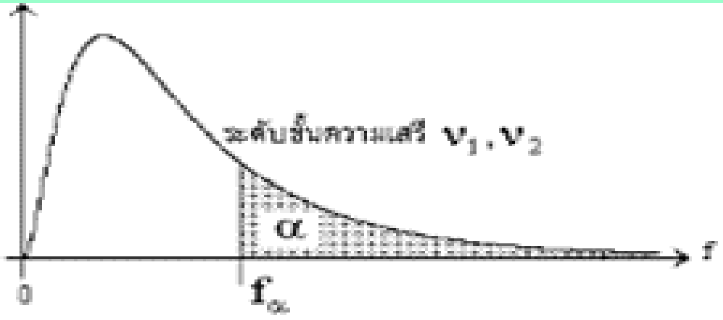
การตัดสินใจ

เนื่องจาก F = 48.43 > $F_{critical}$ = 3.89 จึงปฏิเสธสมมติฐานว่าง (null hypothesis) และสรุปได้ว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของกลุ่มเรียนทั้งสาม

ตารางที่ 6. ตารางค่าวิกฤตของการแจกแจง F

เมื่อกำหนดองศาความอิสระ v_1, v_2 และค่า $\alpha = 0.05$

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



$f_{0.05,(v_1,v_2)}$

v_2	v_1 = 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

การทดสอบ Post-hoc ด้วย Tukey's HSD

สำหรับโจทย์นี้มีคะแนนของนักเรียนในสามกลุ่ม (A, B, และ C) ดังนี้

กลุ่ม A: 85, 90, 88, 86, 91

กลุ่ม B: 78, 82, 80, 76, 79

กลุ่ม C: 92, 95, 94, 90, 93

ขั้นตอนการคำนวณ Tukey's HSD

1) คำนวณค่าเฉลี่ยของแต่ละกลุ่ม

$$\bar{A} = \frac{85+90+88+86+91}{5} = \frac{440}{5} = 88$$

$$\bar{B} = \frac{78+82+80+76+79}{5} = \frac{395}{5} = 79$$

$$\bar{C} = \frac{92+95+94+90+93}{5} = \frac{464}{5} = 92.8$$

ขั้นตอนการคำนวณ Tukey's HSD

2)คำนวณค่า Mean Square Error (MSE)

5.067 จากการคำนวณ F-test ด้านบน

3) คำนวณค่า Standard Error (SE)

$$SE=\sqrt{\frac{MSE}{n}} \quad ; n = 5$$

$$SE=\sqrt{\frac{5.067}{5}} \quad = 1.0067$$

คำนวณค่า HSD

ค่า q จากตาราง Studentized Range Distribution สำหรับ $\alpha=0.05$ $k=3$ และ $df=12$ ประมาณ 3.77

$HSD=q \cdot SE= 3.77 \cdot 1.0067 =3.80$

เปรียบเทียบค่าเฉลี่ยของแต่ละกลุ่ม

$|\bar{A} - \bar{B}| = |88-79| = 9 > 3.80$ (แตกต่างอย่างมีนัยสำคัญ)

$|\bar{A} - \bar{C}| = |88-92.8| = 4.8 > 3.80$ (แตกต่างอย่างมีนัยสำคัญ)

$|\bar{B} - \bar{C}| = |79-92.8| = 13.8 > 3.80$ (แตกต่างอย่างมีนัยสำคัญ)

สรุปผลการทดสอบ Post-hoc ด้วย Tukey's HSD: การทดสอบพบว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของทุกกลุ่ม:

กลุ่ม A และ B มีความแตกต่างกันอย่างมีนัยสำคัญ

กลุ่ม A และ C มีความแตกต่างกันอย่างมีนัยสำคัญ

กลุ่ม B และ C มีความแตกต่างกันอย่างมีนัยสำคัญ

สรุป

การวิเคราะห์ความแปรปรวน (ANOVA) เป็นเครื่องมือสถิติที่สำคัญและมีประสิทธิภาพในการเปรียบเทียบค่าเฉลี่ยของกลุ่มหลายกลุ่ม เพื่อพิจารณาว่ามีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ การวิเคราะห์การแจกแจงความแปรปรวนช่วยให้เข้าใจการกระจายตัวของข้อมูลภายในและระหว่างกลุ่ม

บทที่ 9 สถิตินอนพาราเมตริก

- การคำนวณและการตีความสถิติพาราเมตริก
- สรุป

วัตถุประสงค์การเรียนรู้

- **อธิบายความหมายและความสำคัญของสถิตินอนพาราเมตริก** และเปรียบเทียบกับสถิติพาราเมตริกได้
- **ระบุเงื่อนไขและสถานการณ์ที่เหมาะสมในการใช้สถิตินอนพาราเมตริก**
- **เลือกใช้สถิตินอนพาราเมตริกที่เหมาะสม** สำหรับปัญหาต่าง ๆ

บทนำ

สถิตินอนพาราเมตริก (Non-parametric statistics) เป็นสาขาหนึ่งของสถิติที่มีความสำคัญในการวิเคราะห์ข้อมูลที่ไม่สามารถประยุกต์ใช้กับ การกระจายตัวที่กำหนดไว้ล่วงหน้า เช่น การกระจายตัวปกติ ซึ่งแตกต่างจากสถิติพาราเมตริกที่ต้องอาศัยสมมติฐานเกี่ยวกับการกระจายตัวของข้อมูล สถิตินอนพาราเมตริกมักถูกนำมาใช้ในกรณีที่ข้อมูลมีขนาดเล็ก หรือข้อมูลที่ไม่เป็นเชิงเส้น ไม่สมมาตร หรือมีลักษณะเป็นอันดับ

ประเภทสถิตินอนพาราเมตริก

1.**การทดสอบอันดับ (Rank Tests)** : การทดสอบวิลคอกซัน (Wilcoxon Rank-Sum Test) ใช้เปรียบเทียบค่ากลางของสองกลุ่มที่ไม่เป็นอิสระต่อกัน , การทดสอบแมนน์-วิตนีย์ (Mann-Whitney U Test) ใช้เปรียบเทียบค่ากลางของสองกลุ่มอิสระ , การทดสอบเครื่องหมาย (Sign Test) จัดอยู่ในประเภทการทดสอบอันดับ (Rank Tests) ในกลุ่มสถิตินอนพาราเมตริก โดยเป็นการทดสอบที่ใช้เพื่อเปรียบเทียบค่ากลางของกลุ่มข้อมูลสองกลุ่ม 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

การคำนวณสถิตินอนพาราเมตริกสำหรับการทดสอบกลุ่มตัวอย่างไม่เกิน 2 กลุ่มใช้ในกรณีที่ข้อมูลไม่เป็นไปตามสมมติฐานของการกระจายตัวปกติ หรือเมื่อข้อมูลมีขนาดเล็กและไม่สามารถใช้สถิติพาราเมตริกได้ การทดสอบนอนพาราเมตริกที่นิยมใช้สำหรับกลุ่มตัวอย่างไม่เกิน 2 กลุ่ม

- การทดสอบไค-สแควร์
- การทดสอบแมนน์-วิตนีย์ (Mann-Whitney U Test)
- การทดสอบสัญญาณ

การทดสอบไค-สแควร์

การทดสอบไค-สแควร์ (Chi-Square Test) เป็นการทดสอบทางสถิติที่ใช้ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรหมวดหมู่ (categorical variables) สองตัว หรือตรวจสอบการกระจายของข้อมูล เพื่อประเมินว่าข้อมูลนั้นตรงตามการคาดการณ์ที่กำหนดหรือไม่ โดยทั่วไปมีสองประเภทหลักคือ การทดสอบไค-สแควร์สำหรับความเป็นอิสระ (Chi-Square Test for Independence) และการทดสอบไค-สแควร์สำหรับความเหมาะสม (Chi-Square Test for Goodness of Fit)

เงื่อนไขการใช้การทดสอบไค-สแควร์

- ข้อมูลที่ใช้ในการทดสอบต้องเป็นตัวแปรหมวดหมู่ (categorical variables)
- ข้อมูลในแต่ละหมวดหมู่ต้องเป็นอิสระจากกัน
- ค่าคาดหมาย (expected frequency) ในแต่ละเซลล์ของตารางนับ (contingency table) อย่างน้อย 5 ค่าขึ้นไป
- ข้อมูลควรถูกเก็บรวบรวมจากการสุ่มตัวอย่าง (random sampling)

- การทดสอบไค-สแควร์สำหรับความเป็นอิสระใช้ตรวจสอบว่าตัวแปรสองตัวเป็นอิสระจากกันหรือไม่

เช่น การศึกษาความสัมพันธ์ระหว่างเพศกับการเลือกวิชาที่เรียน

- การทดสอบไค-สแควร์สำหรับความเหมาะสมใช้ตรวจสอบว่าการแจกแจงของตัวแปรหนึ่งตัวเป็นไปตามการแจกแจงที่คาดหวังหรือไม่

เช่น การตรวจสอบว่าการทอยลูกเต๋าเป็นไปตามการแจกแจงแบบสม่ำเสมอหรือไม่

คำนวณค่าไค-สแควร์โดยใช้สูตร

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

โดยที่ O_i คือค่าที่สังเกตได้ (observed value) และ E_i คือค่าที่คาดหวัง (expected value)

การทดสอบความเป็นอิสระของตัวแปร 2 ชุด

แสดงจำนวนพนักงานแผนกต่างๆ แยกตามประเภทหนังสือพิมพ์ที่ชอบอ่าน

มีตัวแปรเกี่ยวข้อง 2 ชุดเป็นข้อมูลที่แยกตามแผนก (ตัวแปรชุด 1) และแยกตามประเภทหนังสือพิมพ์ที่ชอบอ่าน (ตัวแปรชุด 2) ผู้วิเคราะห์สามารถใช้ข้อมูลเช่น ตัวอย่างนี้ทำ**การทดสอบไค-สแควร์** หาข้อสรุปได้ว่าตัวแปรทั้งสองชุดมีความเกี่ยวข้องหรือมีอิทธิพลต่อกันหรือไม่ หรือการอยู่ต่างแผนกกันเป็นเหตุให้พนักงานชอบอ่านหนังสือต่างฉบับกันหรือไม่เช่นนี้เป็นต้น

สมมติฐาน

H_0 : ตัวแปรทั้งสองชุดมีความเป็นอิสระจากกัน

H_1 : ตัวแปรทั้งสองชุดไม่เป็นอิสระต่อกัน

จำนวนพนักงาน ในแต่ละแผนกที่ชอบอ่านหนังสือพิมพ์แต่ละประเภทควรเป็นสัดส่วนเดียวกัน

ข้อมูลชุด O

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

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

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

แทนค่าจากตารางทั้งสองในสูตร

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

$$\chi_s^2 = \frac{(183.0-175.7)^2}{175.7} + \frac{(107.0-107.8)^2}{107.8} + \frac{(60.0-66.5)^2}{66.5}$$

$$+ \frac{(54.0-50.2)^2}{50.2} + \frac{(27.0-30.8)^2}{30.8} + \frac{(19.0-19.0)^2}{19.0}$$

$$+ \frac{(14.0-25.1)^2}{25.1} + \frac{(20.0-15.4)^2}{15.4} + \frac{(16.0-9.5)^2}{9.5}$$

$$= 12.431$$

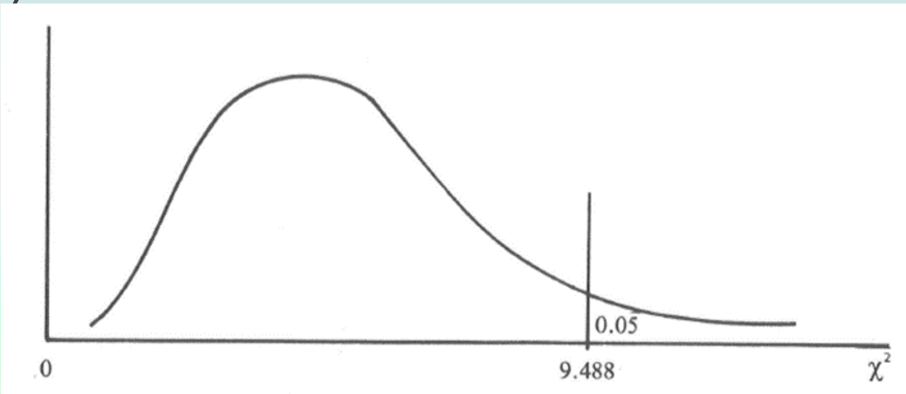
พนักงานแผนก	ประเภทหนังสือพิมพ์ที่ชอบอ่าน			
	ก	ข	ค	รวม
บุคคล	175.7	107.8	66.5	350.0
บัญชี	50.2	30.8	19.0	100.0
ตลาด	25.1	15.4	9.5	50.0
รวม	251.0	154.0	95.0	500

นำไปเปรียบเทียบกับค่าวิกฤต

สมมุติให้การทดสอบทำด้วยระดับนัยสำคัญ 0.05 ($\alpha = 0.05$) และปัญหาข้อนี้ข้อมูลอยู่ในลักษณะตารางที่มี 3 แถว ($r = 3$) และ 3 คอลัมน์ ($k = 3$) ซึ่งข้อมูลลักษณะนี้ค่า df คำนวณได้ดังนี้

$$df = (r - 1)(k - 1)$$
$$= (3 - 1)(3 - 1)$$
$$= 4$$

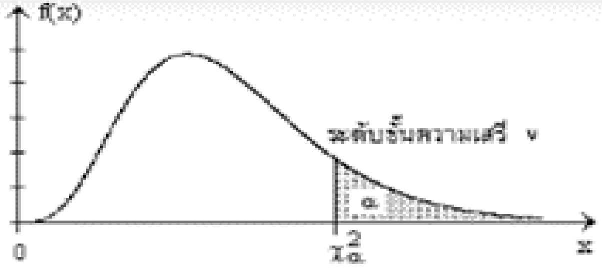
ค่า χ^2 ที่ใช้เป็นค่าตัดสินคือค่า $\chi^2_{(0.05,4)} = 9.488$



ตารางที่ 5. ตารางค่าวิกฤตของการแจกแจงไคสแควร์ χ^2

เมื่อระดับขั้นความเสรีเท่ากับ v

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



v	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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

การตัดสินใจสำหรับปัญหานี้จึงเขียนได้ว่าจะไม่ยอมรับ H_0 ถ้า χ^2_S 12.431 > 9.488 ตัวแปรทั้งสองชุดไม่เป็นอิสระต่อกัน

การทดสอบแมนน์-วิตนีย์ (Mann-Whitney U Test)

ใช้เพื่อเปรียบเทียบค่ากลางของสองกลุ่มอิสระ โดยมี สมมติฐาน (Null hypothesis, $H0$): การแจกแจงของสองกลุ่มเท่ากัน และ $H1$: 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

ข้อตกลงเบื้องต้น

1. การกระจายของข้อมูลไม่จำเป็นต้องเป็นแบบโค้งปกติ

2. กลุ่มตัวอย่างทั้งสองกลุ่มต้องเป็นอิสระกัน

3. ข้อมูลที่นำมาใช้ทดสอบอยู่ในระดับมาตรจัดอันดับ โดยสามารถนำข้อมูลทั้งสอง กลุ่มมาจัดอันดับร่วมกันได้

ตัวอย่างการคำนวณ Mann-Whitney U Test

สมมติว่ามีข้อมูลสองกลุ่มอิสระ ,กลุ่ม 1: [5, 7, 8, 6, 9], กลุ่ม 2: [6, 9, 7, 10, 8]

สมมติฐาน, H_0: การแจกแจงของสองกลุ่มเท่ากัน, H_1: การแจกแจงของสองกลุ่มไม่เท่ากัน

○ 5 ได้อันดับที่ 1
○ 6 ทั้งสองค่าได้อันดับที่ 2 และ 3
ดังนั้นอันดับเฉลี่ยจะเป็น (2 + 3) / 2 = 2.5

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]

2. แยกอันดับของแต่ละกลุ่ม: กลุ่ม 1: [1, 4.5, 6.5, 2.5, 8.5], กลุ่ม 2: [2.5, 8.5, 4.5, 10, 6.5]

3. คำนวณอันดับรวม: $R_1 = 1 + 4.5 + 6.5 + 2.5 + 8.5 = 23$, $R_2 = 2.5 + 8.5 + 4.5 + 10 + 6.5 = 32$

ตัวอย่างการคำนวณ Mann-Whitney U Test

4. คำนวณค่า U

$$U_1 = 5 \times 5 + \frac{5(5+1)}{2} - 23 = 25+15-23 = 17$$

$$U_2 = 5 \times 5 + \frac{5(5+1)}{2} - 32 = 25+15-32 = 8$$

ค่า U ที่ใช้คือค่าน้อยที่สุดระหว่าง U_1 และ U_2, ดังนั้น $U = 8$

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

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

$R_1 = 23$, $R_2 = 32$

$n_1 = 5$, $n_2 = 5$

อาณาเขตวิกฤตและการสรุปผล

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

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 $\frac{\alpha}{2}$

กรณีที่ใช้ค่า U หาก $9 \leq n_1 \leq 20$ หรือ $9 \leq n_2 \leq 20$

จะปฏิเสธสมมติฐาน H_0 เมื่อค่า U ที่คำนวณได้น้อยกว่าหรือเท่ากับ ค่าวิกฤตของ U จากตาราง

เปรียบเทียบค่าที่ได้กับตาราง Mann-Whitney U

n1	n2	α=0.05 (two-tailed)
5	5	2

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

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

สรุปผลการทดสอบ

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

การทดสอบสัญญาณ

การทดสอบสัญญาณ (Sign Test) เป็นการทดสอบนอนพาราเมตริกที่ใช้สำหรับการเปรียบเทียบค่ามัธยฐานของสองกลุ่มที่เกี่ยวข้องกัน โดยเฉพาะในกรณีที่ข้อมูลไม่มีการกระจายตัวที่ทราบ หรือมีการกระจายตัวที่ไม่เป็นปกติ

ขั้นตอนการทดสอบสัญญาณ (Sign Test)

1.กำหนดสมมติฐาน

สมมติฐานที่ศึกษา (Null Hypothesis, H0): ค่ามัธยฐานของ ความแตกต่างระหว่างคู่ข้อมูลเท่ากับศูนย์

สมมติฐานที่เชื่อ (Alternative Hypothesis, H1): ค่ามัธยฐานของความแตกต่างระหว่างคู่ข้อมูลไม่เท่ากับศูนย์

การคำนวณ

- หาความแตกต่างของคู่ข้อมูลแต่ละคู่ (เช่น ความแตกต่างระหว่างการวัดก่อนและหลังการทดลอง)
- แทนความแตกต่างที่เป็นบวกด้วยเครื่องหมาย "+" และความแตกต่างที่เป็นลบด้วยเครื่องหมาย "-" โดยไม่นับค่าความแตกต่างที่เป็นศูนย์
- นับจำนวนครั้งที่เครื่องหมาย "+" และ "-" ปรากฏ

การวิเคราะห์

- จำนวนครั้งที่เครื่องหมาย "+" และ "-" ปรากฏจะถูกใช้เพื่อคำนวณค่าสถิติการทดสอบ
- ถ้าไม่มีความแตกต่างอย่างมีนัยสำคัญ จะคาดหวังว่าเครื่องหมาย "+" และ "-" จะปรากฏใกล้เคียงกัน
- ใช้ตารางการแจกแจงทวินาม (Binomial Distribution) เพื่อหาค่าความน่าจะเป็น (p-value) สำหรับจำนวนครั้งที่เครื่องหมาย "+" และ "-" ปรากฏ

ตัวอย่างการทดสอบสัญญาณ ด้วยการใช้การแจกแจงทวินาม (Binomial Distribution)

ข้อมูลที่ใช้ น้ำหนักของแมว เป็น กก.

- 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 Ho (ไม่มีความแตกต่าง) เครื่องหมาย "+" และ "-" มีโอกาสเกิดขึ้นเท่ากันที่ 0.5
- การทดสอบนี้เป็นการแจกแจงทวินาม Binomial(n=10,p=0.5)

ขั้นตอนการทดสอบสมมติฐาน

H0: p=0.5 (ไม่มีความแตกต่าง)

H1: p≠0.5 (มีความแตกต่าง)

α=0.05 (สองด้าน)

คำนวณค่า test statistic คือ ค่าที่น้อยที่สุดของจำนวนครั้งที่เป็นเครื่องหมายบวกหรือลบ คือ 4

หาค่าวิกฤต (Critical Value) จากตารางการแจกแจงทวินาม $n=10$ และ $p=0.5$

k	P(X ≤ k)	P(X ≥ 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 $P(X \le k_1)$ ใกล้เคียง 0.025 พบว่า $k_1=2$ ($P(X \le 2) = 0.05469$)

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

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

การทดสอบการเปลี่ยนแปลงแมคนีมาร์

บทที่ 10 การวิเคราะห์การถดถอยและสหสัมพันธ์

- การทดสอบสหสัมพันธ์
- การวิเคราะห์การถดถอยและการใช้โมเดลทางสถิติ
- สรุป

วัตถุประสงค์การเรียนรู้

- **อธิบายความหมายและความแตกต่างระหว่างการวิเคราะห์สหสัมพันธ์และการวิเคราะห์การถดถอย** ได้อย่างถูกต้อง
- **คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)** และแปลความหมายความสัมพันธ์ระหว่างตัวแปรได้
- **สร้างสมการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression)** และตีความค่าสัมประสิทธิ์การถดถอย (Regression Coefficients) ได้

บทนำ

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)

สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

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

- ค่า 1 หมายถึงความสัมพันธ์เชิงบวกที่สมบูรณ์ (เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรจะเพิ่มขึ้นตาม)
- ค่า -1 หมายถึงความสัมพันธ์เชิงลบที่สมบูรณ์ (เมื่อค่าของตัวแปรหนึ่งเพิ่มขึ้น ค่าของอีกตัวแปรจะลดลง)
- ค่า 0 หมายถึงไม่มีความสัมพันธ์ระหว่างตัวแปรทั้งสอง

สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

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

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

ตัวอย่างการคำนวณสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

1. หาค่าเฉลี่ย $\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$, $\bar{Y} = \frac{2+3+4+5+6}{5} = \frac{20}{5} = 4$

2. หาผลต่าง ระหว่างค่าของแต่ละตัวแปรกับค่าเฉลี่ยและนำมาคูณ

X	Y	$Xi - \bar{x}$	$Yi - \bar{Y}$	$(Xi - \bar{x})(Yi - \bar{Y})$
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

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

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

สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

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

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

5. หาผลรวม ของกำลังสองของผลต่างของแต่ละตัวแปร

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

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

6. นำผลรวม ที่ได้มาคูณกันและหาค่ารากที่สอง: $r = \frac{10}{\sqrt{10*10}} = 10/10 = 1$

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

จัดลำดับ ค่าของตัวแปร X และ Y

หาค่าความแตกต่างของลำดับ (d_i) สำหรับแต่ละคู่ข้อมูล

หาค่ากำลังสองของความแตกต่างของลำดับ (d_i^2)

นำค่าที่ได้จากข้อ 3 มาหาผลรวม

คำนวณ โดยใช้สูตร

สมมติว่ามีข้อมูลเกี่ยวกับตัวแปรสองตัว คือ 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 (d_i) และค่ากำลังสองของความแตกต่างของลำดับ (d_i^2)

X	Rank X	Y	Rank Y	d_i	d_i^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 **di²**

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

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)

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

Y คือค่าของตัวแปรตาม

X คือค่าของตัวแปรอิสระ

a คือค่าคงที่ (Intercept)

b คือสัมประสิทธิ์การถดถอย (Slope)

ϵ คือความคลาดเคลื่อน (Error Term)

การวิเคราะห์การถดถอยช่วยในการทำนายค่าอนาคตของตัวแปรตามเมื่อรู้ค่าของตัวแปรอิสระ และสามารถใช้ในการทดสอบสมมติฐานเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร นอกจากนี้ยังช่วยในการทำความเข้าใจถึง ความแรงและทิศทางของความสัมพันธ์ 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 YY เมื่อ XX เท่ากับ 0

b คือสัมประสิทธิ์การถดถอย (Slope) ซึ่งแสดงถึงการเปลี่ยนแปลงของ Y เมื่อ X เพิ่มขึ้น 1 หน่วย

X คือค่าของตัวแปรอิสระ

ขั้นตอนการคำนวณ

หาค่าเฉลี่ย ของตัวแปร X และ Y:

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

หาค่าสัมประสิทธิ์การถดถอย (b)

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

หาค่าคงที่ (a)

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

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

บทที่ 11 การพยากรณ์และการตัดสินใจทางธุรกิจ

- การใช้สถิติในการพยากรณ์และการวางแผนทางธุรกิจ
- การใช้ข้อมูลสถิติในการตัดสินใจทางธุรกิจ
- สรุป

วัตถุประสงค์การเรียนรู้

- **อธิบายความหมาย ความสำคัญ และประเภทของการพยากรณ์** ที่ใช้ในงานธุรกิจได้
- **จำแนกวิธีการพยากรณ์** ทั้งเชิงคุณภาพ (Qualitative) และเชิงปริมาณ (Quantitative) เช่น การถัวเฉลี่ยเคลื่อนที่ (Moving Average), การปรับเรียบเอ็กซ์โพเนนเชียล (Exponential Smoothing), และการวิเคราะห์แนวโน้ม (Trend Analysis)

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 (การวางแผนการผลิตและการขาย) การเงิน (การวางแผนการลงทุน) และการเมือง (การทำนายผลการเลือกตั้ง) (วุฒิ สุขเจริญ,2561)

การพยากรณ์สามารถแบ่งเป็นหลายประเภทตามลักษณะของข้อมูลและวิธีการทำนาย นี่คือบางประเภท และวิธีการพยากรณ์ที่ใช้กันบ่อย:

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 ได้แก่ ค่าแนวโน้ม ค่าวัดอิทธิพลของฤดูกาล ค่าวัดวัฏจักร ค่าวัดเหตุการณ์ที่ผิดปกติ ค่าวัดจะได้จากวิธีการเฉลี่ยแบบธรรมดา (simple average) แบบเคลื่อนที่ (moving average) การพยากรณ์แบบนี้จะมีการเก็บข้อมูลความต้องการ สินค้า หรือยอดขายระยะเวลาหนึ่งสม่ำเสมอโดยมีช่วงห่างใน การเก็บข้อมูลที่เท่า ๆ กัน โดยทั่วไป ความต้องการสินค้าจะมากหรือน้อยนั้นเกิดจากอิทธิพล 4 ประการ ได้แก่ แนวโน้ม (Trend) ฤดูกาล (Seasonal) วัฏจักร (Cycle) และความผิดปกติหรือความไม่แน่นอน

การพยากรณ์สามารถแบ่งเป็นหลายประเภทตามลักษณะของข้อมูลและวิธีการทำนาย นี่คือบางประเภทและวิธีการพยากรณ์ที่ใช้กันบ่อย:

1.3 วิธีปรับให้เรียบ (Smoothing Method) ค่าพยากรณ์ในอนาคตเป็นค่าที่ได้จากค่าสังเกตในอดีตโดยการให้น้ำหนัก (weight) กับค่าสังเกตแบบต่างๆ หากให้น้ำหนักกับค่าสังเกตเท่าๆกันจะเรียกว่า วิธีเฉลี่ยเคลื่อนที่ (moving average method)

2. การศึกษาความสัมพันธ์ระหว่างปัจจัยหรือตัวแปรต่างๆที่เป็นเหตุและผลเนื่องจากเหตุการณ์ต่างๆ 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

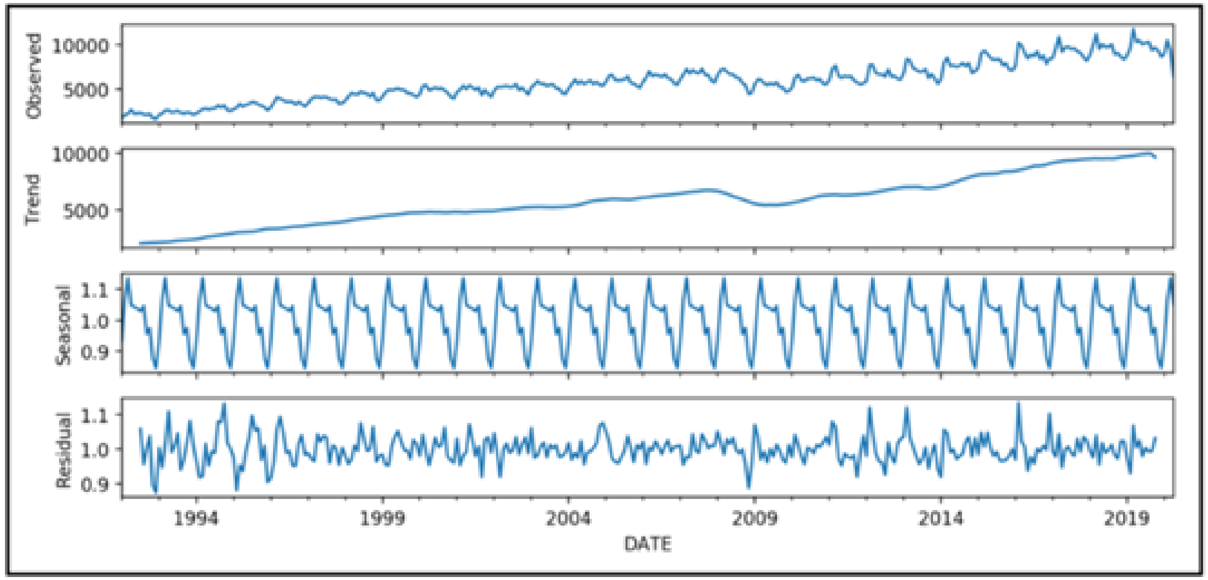
2.ฤดูกาล (Seasonal): การเปลี่ยนแปลงของความต้องการหรือยอดขายที่มีลักษณะเกิดซ้ำซ้อนในระยะเวลาที่กำหนด เช่น การเพิ่มขึ้นของยอดขายในช่วงเทศกาลหรือเทศกาลวันหยุด เราสามารถใช้การแยกส่วนฤดูกาลออกจากข้อมูลเพื่อทำนายผลลัพธ์ที่ถูกต้องในอนาคต

การใช้สถิติในการพยากรณ์และการวางแผนทางธุรกิจ

3. จักร (Cycle): การเปลี่ยนแปลงของความต้องการหรือยอดขายที่มีลักษณะซ้ำซ้อนแต่ไม่สม่ำเสมอในระยะเวลายาวๆ เช่น การเพิ่มขึ้นของยอดขายในรอบของเศรษฐกิจ การรับรู้และการปรับปรุงวัฏจักรสามารถช่วยให้เราตระหนักถึงภาวะทางเศรษฐกิจและวางแผนการดำเนินธุรกิจในอนาคตได้

4. ความผิดปกติหรือความไม่แน่นอน (Irregular component): ความผิดปกติหรือความไม่แน่นอนอาจเกิดขึ้นจากปัจจัยที่ไม่คาดคิด เช่น สภาวะภูมิอากาศ, ภาวะเศรษฐกิจโลก, หรือเหตุการณ์ไม่คาดคิดอื่นๆ การวิเคราะห์และการจัดการกับความไม่แน่นอนอาจช่วยให้เราสามารถปรับแนวทางการวางแผนและการบริหารจัดการในสถานการณ์ที่ซับซ้อนและเปลี่ยนแปลงได้เร็วขึ้น

อนุกรมเวลาหมายถึงกลุ่มของค่าสังเกตที่เก็บรวบรวมมาตามเวลาอย่างต่อเนื่อง(ภูมิฐาน รังคกุลนุวัฒน์,2556) ช่วงเวลาจะเท่ากันหรือไม่เท่ากันก็ได้ จะใช้สัญลักษณ์ yt แทนอนุกรมเวลา y1,y2,...yn ที่เก็บมาใน n ช่วงเวลา วิธีการพยากรณ์จะใช้ช่วงเวลาห่างกันเท่ากัน เช่น ปี ครึ่งปี ไตรมาส เดือน เป็นต้น ตัวอย่างของอนุกรมเวลาคือ ยอดขายสินค้ารายปีของบริษัทมอลลี่ จำนวนนักท่องเที่ยวรายปีที่เดินทางเข้ามาประเทศไทย หากต้องการศึกษาอิทธิพลของฤดูกาลต้องเก็บข้อมูลเป็นรายไตรมาสหรือรายเดือน



แหล่งที่มา : Ivan Despot,2024

ส่วนประกอบของอนุกรมเวลา

ส่วนประกอบของอนุกรมเวลาประเภทการแยกส่วนแบบบวก (Additive decomposition) ข้อมูลชุดเวลา yt ถูกพิจารณาเป็นผลรวมของส่วนประกอบต่าง ๆ ดังนี้: แนวโน้ม (Tt), วัฏจักร (Ct), ฤดูกาล (St), และส่วนข้อผิดพลาดหรือข้อผิดปกติ (It) โดยมีสมการเป็นดังนี้:

$$yt=Tt+Ct+St+It$$

โดยที่:

Tt แทนส่วนแนวโน้ม (Trend) ซึ่งบ่งบอกถึงทิศทางหรือแนวโน้มของชุดเวลาในระยะยาว

Ct แทนส่วนวัฏจักร (Cycle) ซึ่งบังคับปริมาณที่มีการเปลี่ยนแปลงเป็นระยะหรือลำดับเวลาที่ไม่ได้เกี่ยวข้องกับฤดูกาล

St แทนส่วนฤดูกาล (Seasonal) ซึ่งแสดงแนวโน้มที่เป็นรูปแบบแบบซ้ำซ้อนในช่วงเวลาที่เฉพาะเจาะจง เช่น รายวัน รายสัปดาห์ หรือรายปี

It แทนส่วนข้อผิดพลาดหรือข้อผิดปกติ (Irregular or error term) ซึ่งบังคับปริมาณที่มีการเปลี่ยนแปลงสุ่มหรือเสียงในข้อมูลที่ไม่สามารถอธิบายได้โดยส่วนอื่น

โดยการแยกส่วนเวลาชุดข้อมูลเป็นส่วนประกอบนี้ จะทำให้เราสามารถวิเคราะห์และเข้าใจรูปแบบและโครงสร้างข้อมูลได้อย่างชัดเจนขึ้น และสามารถทำนายได้ด้วยความแม่นยำมากขึ้นโดยการจำแนกแต่ละส่วนแยกกันในการประมาณค่า

		รูปแบบบวก			
t	Y_t =	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) เป็นวิธีการทำนาย หรือประมาณค่าในอนาคตโดยใช้ค่าเฉลี่ยของข้อมูลในช่วงเวลาหนึ่ง 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 $\boldsymbol{n}$=3

สำหรับช่วงเวลาแรก [10, 15, 20] ค่าเฉลี่ย MA1= $\frac{\mathbf{10+15+20}}{\mathbf{3}}$= $\frac{\mathbf{45}}{\mathbf{3}}$=15

สำหรับช่วงเวลาถัดไป [15, 20, 25] ค่าเฉลี่ย MA2= $\frac{\mathbf{15+20+25}}{\mathbf{3}}$= $\frac{\mathbf{60}}{\mathbf{3}}$=20

สำหรับช่วงเวลาถัดไป [20, 25, 30] ค่าเฉลี่ย MA3= $\frac{\mathbf{20+25+30}}{\mathbf{3}}$= $\frac{\mathbf{75}}{\mathbf{3}}$=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)**: เลือกทางเลือกที่คิดว่าเหมาะสมที่สุดตามการวิเคราะห์เชิงคณิคศาสตร์อย่างใดอย่างหนึ่ง

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:

สมาร์ทโฟน A: ราคา 10,000 บาท, หน่วยความจำ 128GB, กล้อง 12MP

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 การตัดสินใจเลือกเครื่องจักรจะง่ายมากเพราะข้อมูลทั้งหมดที่จำเป็นใน การตัดสินใจมีความแน่นอน โดยจะเลือกใช้เครื่องจักร B เนื่องจากสามารถผลิตได้มากกว่าและค่าใช้จ่ายต่อหน่วยผลิตต่ำกว่า

ตัวอย่างของการตัดสินใจภายใต้ความแน่นอน

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

ลักษณะสำคัญ

แนวทางที่รอบคอบ ที่ป้องกันการตัดสินใจ "ผิด"

เน้นต้นทุนโอกาส มากกว่าผลลัพธ์ที่เป็นตัวเลขสัมบูรณ์

มีประโยชน์เมื่อ คต้องการลดความเป็นไปได้ของความผิดหวังอย่างรุนแรง

แตกต่างจากเกณฑ์มินิแมกซ์ ซึ่งมุ่งลดการสูญเสียสูงสุดมากกว่าการลดความเสียใจ

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 การตัดสินใจเชิงนี้มักใช้เครื่องมือทางคณิตศาสตร์และสถิติ เพื่อวิเคราะห์ความเสี่ยงและผลตอบแทนที่เป็นไปได้

ขั้นตอนในการตัดสินใจภายใต้ความเสี่ยง

ระบุทางเลือก (Identify Alternatives): กำหนดทางเลือกต่าง ๆ ที่เป็นไปได้ในการตัดสินใจ

ระบุสถานการณ์ที่เป็นไปได้ (Identify Possible Outcomes): กำหนดผลลัพธ์ที่เป็นไปได้สำหรับแต่ละทางเลือก

กำหนดความน่าจะเป็น (Assign Probabilities): กำหนดความน่าจะเป็นของแต่ละผลลัพธ์

คำนวณค่าคาดหวัง (Calculate Expected Value): คำนวณมูลค่าคาดหวัง (Expected Value) ของแต่ละทางเลือกโดยใช้สูตร

เลือกทางเลือกที่ดีที่สุด (Select the Best Alternative): เลือกทางเลือกที่มีค่าคาดหวังดีที่สุด

สูตรการคำนวณค่าคาดหวัง (Expected Value, EV)

$$EV = \sum (Pi \times Xi)$$

โดย:

- Pi คือ ความน่าจะเป็นของผลลัพธ์ที่ i
- Xi 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✕300,000)+(0.5✕200,000)+(0.2✕100,000

EV(A)=90,000+100,000+20,000

EV(A)=210,000

โครงการ B: EV(B)=(0.3✕400,000)+(0.5✕100,000)+(0.2✕50,000)

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 ทำให้ การตัดสินใจมีพื้นฐานเชิงเหตุผลมากขึ้นแม้จะมีความเสี่ยงที่เกี่ยวข้อง

แนวทางการตัดสินใจภายใต้ความเสี่ยงยังรวมถึงการใช้เกณฑ์แบบ Bayesian (Bayesian Criterion) ซึ่งเป็นการนำทฤษฎีเบย์สมาใช้ในการตัดสินใจ โดยการคำนวณค่าคาดหวังแบบ Bayesian จะพิจารณาความน่าจะเป็นตามเงื่อนไข (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)=(0.4×300,000)+(0.4×200,000)+(0.2×100,000) EV(A)=120,000+80,000+20,000

EV(A)=220,000

โครงการ B : EV(B)=(0.4×400,000)+(0.4×100,000)+(0.2×50,000) 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

บทที่ 12 ดัชนีและการประยุกต์ใช้

- ดัชนี
- การประยุกต์ใช้สถิติในสาขาต่างๆ เช่น การวิเคราะห์การตลาด, การวิเคราะห์การเงิน
- สรุป

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

เลขดัชนีในทางสถิติ

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 เพื่อช่วยในการวางแผนและตัดสินใจได้อย่างมีประสิทธิภาพมา

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

วิธีการคำนวณ CPI

CPI คำนวณโดยการรวบรวมราคาสินค้าและบริการที่ครัวเรือนทั่วไปบริโภค แล้วนำมาคำนวณดัชนีด้วยสูตร Laspeyres Index เป็นส่วนใหญ่ สูตรการคำนวณ CPI ดังนี้

$$\text{Laspeyres Index} = \left(\frac{\sum(p_t \times q_0)}{\sum(p_0 \times q_0)}\right) \times 100$$

- p_t = 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

ตัวอย่างการใช้งาน

- **รัฐบาล**: ใช้ CPI ในการวางแผนนโยบายเศรษฐกิจและการปรับปรุงค่าจ้าง
- **ธุรกิจ**: ใช้ในการตัดสินใจด้านการกำหนดราคาและการวางแผนการเงิน
- **บุคคลทั่วไป**: ใช้ในการปรับการบริโภคและการวางแผนการเงินส่วนบุคคล

ความสำคัญของ CPI

CPI เป็นตัวชี้วัดที่สำคัญและใช้กันอย่างแพร่หลายเพื่อทำความเข้าใจการเปลี่ยนแปลงของราคาสินค้าและบริการที่ครัวเรือนทั่วไปบริโภค และมีบทบาทสำคัญในการวางแผนและตัดสินใจทั้งในระดับบุคคลและระดับนโยบาย

ตัวอย่างการคำนวณดัชนีราคาผู้บริโภค (CPI) แบบละเอียด

ข้อมูลพื้นฐาน

สมมติว่ามีข้อมูลราคาสินค้าและบริการสองประเภทในปีฐานและปีปัจจุบัน รวมถึงปริมาณการบริโภคในปีฐาน ดังนี้

สินค้า/บริการ	ราคาปีฐาน (p_0)	ราคาปีปัจจุบัน (p_t)	ปริมาณปีฐาน (q_0)
ข้าวสาร	10	12	100
เสื้อผ้า	20	25	50

ขั้นตอนการคำนวณ CPI

1. คำนวณค่าผลรวมของ $p_0 \times q_0$

$\sum(p_0+q_0)$ = (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:

$$CPI = \left(\frac{\sum(p_t+q_0)}{\sum(p_0+q_0)}\right) \times 100$$

$$CPI = \left(\frac{2450}{2000}\right) \times 100 = 1.225 \times 100 = 122.5$$

การตีความผลลัพธ์

CPI = 122.5 หมายความว่า ราคาสินค้าและบริการในปีปัจจุบันเพิ่มขึ้น 22.5% เมื่อเทียบกับปีฐาน

เพิ่มเติม: การคำนวณอัตราเงินเฟ้อ

หากเราต้องการคำนวณอัตราเงินเฟ้อจากข้อมูล CPI สามารถทำได้โดยใช้สูตรดังนี้:

$$\text{อัตราเงินเฟ้อ}=\frac{\textbf{CPI}\ \text{ในปีปัจจุบัน}-\textbf{CPI}\ \text{ในปีที่ผ่านมา}}{\textbf{CPI}\ \text{ในปีที่ผ่านมา}}\times 100$$

สมมติว่า CPI ในปีที่ผ่านมาเป็น 115

$$\text{อัตราเงินเฟ้อ}=\frac{122.5-115}{115}\times 100 = 7.5115\times 100 \approx 6.52\%$$

ดังนั้น อัตราเงินเฟ้ออยู่ที่ประมาณ 6.52%

สรุป

การคำนวณ CPI และอัตราเงินเฟ้อช่วยให้สามารถเข้าใจ การเปลี่ยนแปลงของระดับราคาสินค้าและบริการในช่วงเวลาที่แตกต่างกันได้ ซึ่งเป็นข้อมูลสำคัญสำหรับการวางแผนทางเศรษฐกิจและการเงิน

ดัชนีราคาผู้ผลิต

ดัชนีราคาผู้ผลิต (Producer Price Index - PPI) เป็นตัวชี้วัดที่สำคัญในการวัดการเปลี่ยนแปลงของระดับราคาของสินค้าและบริการที่ผู้ผลิตขาย โดย PPI เน้นที่ราคาที่ผู้ผลิตได้รับแทนที่จะเป็นราคาที่ผู้บริโภคจ่าย เหมือนในกรณีของดัชนีราคาผู้บริโภค (CPI) PPI เป็นตัวชี้วัดที่ช่วยในการประเมินสภาพเศรษฐกิจและภาวะเงินเฟ้อในระดับต้นน้ำของห่วงโซ่อุปทาน

วัตถุประสงค์ของ PPI

- **วัดการเปลี่ยนแปลงของราคา**: ใช้ในการติดตามและวัดการเปลี่ยนแปลงของราคาสินค้าและบริการที่ผู้ผลิตขาย
- **ชี้วัดภาวะเงินเฟ้อในระดับต้นน้ำ**: ใช้เพื่อประเมินอัตราเงินเฟ้อที่เกิดขึ้นก่อนที่จะส่งต่อไปยังผู้บริโภค
- **วางแผนและตัดสินใจทางเศรษฐกิจ**: หน่วยงานรัฐบาลและธุรกิจใช้ PPI ในการวางแผนนโยบายเศรษฐกิจและการตัดสินใจด้านการเงิน

วิธีการคำนวณ PPI

PPI คำนวณโดยการรวบรวมราคาสินค้าและบริการที่ผู้ผลิตขายในตลาดต่าง ๆ จากนั้นนำมาคำนวณดัชนีด้วยสูตร Laspeyres Index เป็นส่วนใหญ่ สูตร การคำนวณ PPI ดังนี้

$$\mathrm{PPI} = \left(\frac{\sum (p_t \times q_0)}{\sum (p_0 \times q_0)} \right) \times 100$$

- 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 กลุ่มสินค้าที่สำคัญได้แก่ **สินค้าการเกษตร**: ราคาสินค้าทางการเกษตรที่ผู้ผลิตขาย **สินค้าก่อสร้าง**: ราคาวัสดุก่อสร้างและอุปกรณ์ **สินค้าพลังงาน**: ราคาน้ำมัน, ก๊าซ, และพลังงานอื่นๆ **สินค้าอุตสาหกรรม**: ราคาวัตถุดิบและสินค้าที่ผลิตในโรงงาน

การตีความ PPI

- PPI **เพิ่มขึ้น**: แสดงถึงการเพิ่มขึ้นของราคาสินค้าและบริการที่ผู้ผลิตขาย ซึ่งอาจเป็นสัญญาณของภาวะเงินเฟ้อ
- PPI **ลดลง**: แสดงถึงการลดลงของราคาสินค้าและบริการที่ผู้ผลิตขาย ซึ่งอาจเป็นสัญญาณของภาวะเงินฝืด

ดัชนีปริมาณการผลิต

ดัชนีปริมาณการผลิต (Production Index) เป็นตัวชี้วัดที่ใช้ในการวัดการเปลี่ยนแปลงของปริมาณการผลิตในภาคอุตสาหกรรมหรือภาคเศรษฐกิจต่าง ๆ ในช่วงเวลาที่แตกต่างกัน โดยดัชนีนี้ช่วยให้เราทราบถึงการเติบโตหรือหดตัวของการผลิต ซึ่งมีผลต่อสภาพเศรษฐกิจโดยรวม

วัตถุประสงค์ของดัชนีปริมาณการผลิต

วัดการเปลี่ยนแปลงของการผลิต: ช่วยในการติดตามและวิเคราะห์การเปลี่ยนแปลงของปริมาณการผลิตในอุตสาหกรรมต่างๆ

ประเมินประสิทธิภาพทางเศรษฐกิจ: ช่วยในการประเมินภาวะเศรษฐกิจว่ามีการเติบโตหรือหดตัว

สนับสนุนการวางแผนและนโยบาย: ใช้ข้อมูลในการวางแผนและกำหนดนโยบายทางเศรษฐกิจและการผลิต

วิธีการคำนวณดัชนีปริมาณการผลิต

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 $q_0 \times p_0$**

$$\sum(q_{0+}\ p_0) = (1000 \times 500{,}000) + (2000 \times 100{,}000) = 500{,}000{,}000 + 200{,}000{,}000 = 700{,}000{,}000$$

2. **คำนวณค่าผลรวมของ $q_t \times p_0$**

$$\sum(q_{t+}\ p_0) = \sum(q_t \times p_0) = (1200 \times 500{,}000) + (2500 \times 100{,}000) = 600{,}000{,}000 + 250{,}000{,}000 = 850{,}000{,}000$$

3. **คำนวณดัชนีปริมาณการผลิต**

$$\text{ดัชนีปริมาณการผลิต} = \left(\frac{850{,}000{,}000}{700{,}000{,}000}\right) \times 100 = 1.2143 \times 100 = 121.43$$

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

การวิเคราะห์การตลาด (Marketing Analysis)

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

การวิเคราะห์การเงิน (Financial Analysis)

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

การประยุกต์ใช้ในสาขาอื่นๆ

การวิเคราะห์ผลการเรียน (Student Performance Analysis)

1. **การวิเคราะห์เชิงพรรณนา (Descriptive Statistics)**: ใช้ในการสรุปผลการเรียน เช่น ค่าเฉลี่ย, ค่ามัธยฐาน, ค่าเบี่ยงเบนมาตรฐาน

การวิเคราะห์การถดถอย (Regression Analysis): ใช้ในการวิเคราะห์ปัจจัยที่มีผลต่อผลการเรียน

2. **การประเมินคุณภาพการศึกษา (Educational Assessment)**

การวิเคราะห์เชิงสถิติ (Statistical Analysis): ใช้ในการประเมินและเปรียบเทียบคุณภาพการศึกษาของโรงเรียนหรือโปรแกรมการศึกษาต่างๆ

การประยุกต์ใช้ในสาขาอื่นๆ

วิทยาศาสตร์และการวิจัย (Scientific Research)

1. การออกแบบการทดลอง (Experimental Design)

การวิเคราะห์เชิงสถิติ (Statistical Analysis): ใช้ในการออกแบบและวิเคราะห์การทดลองเพื่อตอบคำถามวิจัยอย่างมีประสิทธิภาพ

2. การวิเคราะห์ข้อมูล (Data Analysis)

การทดสอบสมมติฐาน (Hypothesis Testing): ใช้ในการทดสอบสมมติฐานต่างๆ จากผลการทดลอง

การวิเคราะห์การถดถอย (Regression Analysis): ใช้ในการศึกษาความสัมพันธ์ระหว่างตัวแปรต่างๆ

สรุป

ดัชนีเป็นเครื่องมือทางสถิติที่สำคัญสำหรับการวัดและวิเคราะห์แนวโน้มหรือการเปลี่ยนแปลงของปรากฏการณ์ต่างๆ เช่น ตลาด การเงิน หรือเศรษฐกิจ โดยมีประโยชน์ในการตัดสินใจทางธุรกิจ การวิเคราะห์พฤติกรรมผู้บริโภค และการจัดการการเงินอย่างมีประสิทธิภาพ ดัชนีช่วยให้สามารถคาดการณ์อนาคต ประเมินสภาพเศรษฐกิจ และเปรียบเทียบผลลัพธ์ในช่วงเวลาต่างๆ ได้อย่างแม่นยำ ทำให้การบริหารจัดการในด้านต่างๆ เป็นไปอย่างมีระบบและมีประสิทธิภาพมากขึ้น

เอกสารอ่านเพิ่มเติม

Title Lorem Ipsum

Q1	Q2	Q3	Q4
Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum
Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum
Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum	Lorem ipsum et tula lorem ipsum et lorem ipsum