บทที่ 1
ความรู้เบื้องต้นเกี่ยวกับสถิติ

BUA3127 สถิติธุรกิจ

ผส ดร มมันยา มีนคร

สถิติ ?

หมายถึง ตัวเลข หรือศาสตร์ที่ว่าด้วยข้อมูลซึ่งนำมาจัดกระทำ อาจเป็นข้อมูลซึ่งอยู่ในรูปของข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ เพื่อหาข้อสรุปที่มีประโยชน์และนำมาวิเคราะห์เพื่อใช้ในการตัดสินใจอย่างมีเหตุมีผล

ความหมายสถิติ

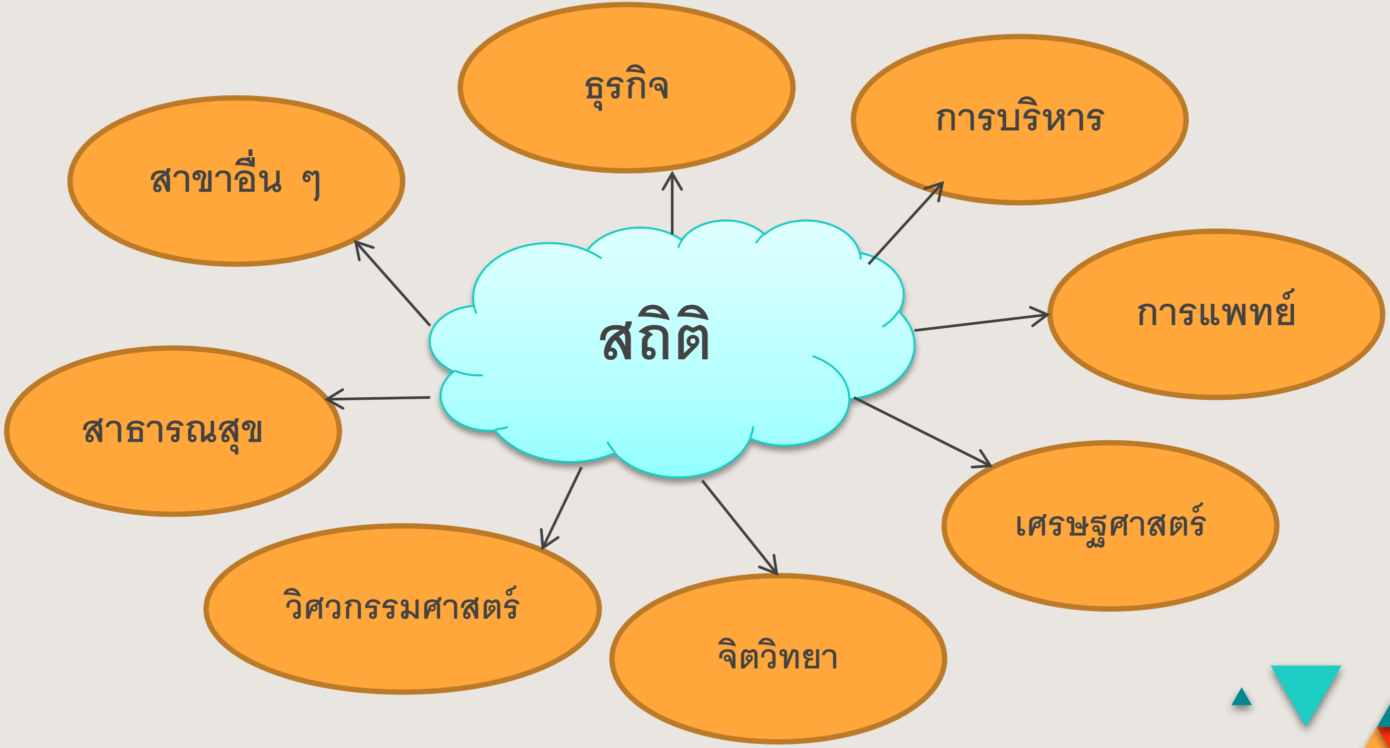
- คำว่า สถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistik มีรากศัพท์ มาจาก Stat จำแนกได้ 2 ความหมาย คือ

  ☆ สถิติหมายถึง ข้อมูลหรือตัวเลขที่แทนข้อเท็จจริงเกี่ยวกับเรื่องต่างๆที่เรา สนใจหรือที่อยู่รอบตัวเรา

  ☆ สถิติหมายถึง ศาสตร์หรือวิชาที่ว่าด้วยหลักการและวิธีการทางสถิติ ซึ่งประกอบด้วย การเก็บรวบรวมข้อมูล การจัดระเบียบข้อมูล หรือการนำเสนอข้อมูล การวิเคราะห์และการแปลความหมายข้อมูล

สถิติ หมายถึง ศาสตร์ที่ว่าด้วยการเก็บรวบรวมข้อมูล การนำเสนอข้อมูล และการวิเคราะห์ข้อมูล

สถิติสามารถนำไปใช้เป็นข้อมูลที่จำเป็นในทุกสาขา



ประเภทของสถิติศาสตร์ (Type of statistics)

สถิติศาสตร์แบ่งเป็น 2 ประเภท

สถิติเชิงพรรณนา (Descriptive statistics) คือ วิธีการทางสถิติที่ใช้ใน การพรรณาลักษณะของข้อมูลที่เก็บรวบรวมมาชุดหนึ่งซึ่งอาจเป็นข้อมูล ตัวอย่างหรือประชากรก็ได้ในรูปแบบของตาราง กราฟ หรือข้อความ แต่ไม่สามารถคาดคะเน นอกเหนือข้อมูลชุดที่มีอยู่ได้

สถิตเชิงอนุมาน (Inferential statistics) คือ วิธีการทางสถิติที่อาศัย ทฤษฎีความน่าจะเป็นในการอนุมานหรือการทำนาย ลักษณะของประชากรโดยอาศัย ข้อมูลตัวอย่าง เช่น การศึกษาความคิดเห็นของนิสิตที่มีต่อการกำหนดให้ สวมหมวกกันน๊อกในการขับขี่รถจักรยานยนต์

ประเภทของสถิติศาสตร์ (Type of statistics) (ต่อ)

สถิติเชิงพรรณนา (Descriptive statistics) เราจะใช้เครื่องมือ

- การวัดแนวโน้มเข้าสู่ส่วนกลาง

  ค่าเฉลี่ย (Mean), มัธยฐาน (Median), ฐานนิยม (Mode)

- การวัดการกระจาย

  พิสัย (Range), ความแปรปรวน (Variance), ส่วนเบี่ยงเบนมาตรฐาน (Standard Variable)

- การวัดตำแหน่งสัมพัทธ์

  เปอร์เซ็นไทล์(Percentile), ควอร์ไทล์(Quartile), เดไซล์(Decile), คะแนนมาตรฐาน (StandardScore)

- การวัดความสัมพันธ์ (Measure of Relationship)

สถิตเชิงอนุมาน (Inferential statistics)

- การประมาณค่า (Estimation)
- การทดสอบสมมติฐาน (Hypothesis Testing)
- การทำนาย (Prediction)

ข้อมูลและระดับการวัดข้อมูล

ข้อมูล (data) หมายถึง ข้อเท็จจริงต่างๆ ที่เราสนใจจะศึกษา ซึ่งอาจเป็นตัวเลขหรือมิใช่ตัวเลขก็ได้ โดยแบ่งได้เป็น 2 ประเภท คือ

ข้อมูลเชิงปริมาณ คือ ค่าที่วัดได้ออกมาเป็นตัวเลข

ข้อมูลเชิงคุณภาพ คือ ค่าที่วัดได้ไม่สามารถวัดออกมาเป็นตัวเลขได้

ระดับการวัดข้อมูล (levels of measurement)

ระดับการวัดข้อมูลแบ่งออกเป็น 4 ระดับหรือมาตรา ดังนี้

- มาตรานามบัญญัติ
- มาตราเรียงลำดับ
- มาตราอันตรภาค
- มาตราอัตราส่วน

ระดับการวัดข้อมูล

1) **มาตรานามบัญญัติ** (Nominal scale) เป็นตัวแปรที่ถูกจัดเป็นกลุ่มๆ โดยที่ตัวแปรนี้ไม่สามารถจัดลำดับก่อนหลัง หรือบอกระยะห่างได้เช่น เพศ แบ่งได้เป็น 2 กลุ่ม คือเพศชาย และเพศหญิง

2) **มาตราอันดับ** (Ordinal scale) จะมีลักษณะคล้ายกับมาตรานามบัญญัติ คือสามารถจัดเป็นกลุ่มๆ ได้และไม่สามารถบอกระยะห่างระหว่างกลุ่มได้เช่นเดียวกับมาตรานามบัญญัติ แต่มาตราจัดลำดับสามารถจัดลำดับก่อนหลังของตัวแปรได้เช่น วุฒิการศึกษา อาจแบ่งได้เป็น 3 กลุ่ม คือ ต่ำกว่าปริญญาตรี ปริญญาตรี และสูงกว่าปริญญาตรี และสามารถจัดลำดับก่อนหลังได้ว่าผู้ที่จะเรียนในระดับปริญญาตรีได้ต้องผ่านการศึกษาในระดับมัธยมมาก่อน หรือผลการประกวดนางงามที่ผลออกมาเป็น อันดับ 1, 2, 3, ... ฯลฯ

ระดับการวัดข้อมูล

3) มาตราอันตรภาค (Interval scale) เป็นการกำหนดตัวเลขให้กับข้อมูล โดยที่ช่วงห่าง 1 หน่วยจะมีปริมาณเท่ากันทำให้สามารถบอกระดับความมากหรือน้อยได้ เช่น อุณหภูมิ 33 องศาเซลเซียส และ 30 องศาเซลเซียส เราทราบได้ว่า 33 องศาเซลเซียส มากกว่า 30 องศาเซลเซียส อยู่ 3 องศาเซลเซียส แต่การวัดแบบนี้จะไม่มีคุณสมบัติที่เรียกว่า ศูนย์แท้ (Absolute zero) นั่นก็คือ คำว่า ศูนย์ไม่ได้หมายถึงไม่มี เช่น อุณหภูมิ 0 องศาเซลเซียส ไม่ได้หมายถึงไม่มีอุณหภูมิ แต่หมายถึงอุณหภูมิที่ระดับ 0 องศาเซลเซียส

4) มาตราอัตราส่วน (Ratio scale) มาตรานี้ เป็นมาตราที่มีลักษณะเหมือนกับมาตราอัตราส่วนทุกประการ แต่สิ่งที่แตกต่างกันในมาตรานี้คือ มาตรานี้เป็นมาตราที่มี ศูนย์แท้ (Absolute Zero) นั่นหมายความว่า ผลที่ได้จากการวัดในมาตรานี้หากเท่ากับศูนย์แสดงว่าไม่มีอย่างแท้จริง เช่น ตัวแปรน้ำหนัก หรือส่วนสูง 0 (ศูนย์) ของตัวแปรทั้งสองตัวนี้หมายถึงไม่มีนำหนักและไม่มีความสูงเลย

ตัวอย่างประเภทของข้อมูล

ตอนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

1. เพศ ☐ ชาย
   ☐ หญิง

2. อายุ .... ปี
   ปัจจุบันเรียนอยู่ชั้น ....

3. การใช้อินเทอร์เน็ต
   ☐ น้อยกว่า 2 ชั่วโมง/สัปดาห์
   ☐ 2-15 ชั่วโมง/สัปดาห์
   ☐ มากกว่า 15 ชั่วโมง/สัปดาห์

4. ช่องทางการเข้าถึงอินเทอร์เน็ต
   ☐ สมาร์ทโฟน ☐ คอมพิวเตอร์ ☐ แท็บเล็ต
   ☐ อื่น ๆ ระบุ ....

นามบัญญัติ

ตัวอย่างประเภทของข้อมูล

ส่วนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

1. เพศ ☐ ชาย ☐ หญิง
2. อายุ .... ปี
3. ศาสนา ☐ พุทธ ☐ คริสต์ ☐ อิสลาม ☐ อื่นๆ โปรดระบุ ....
4. ระดับการศึกษาที่จบ ☐ ต่ำกว่ามัธยมศึกษา ☐ มัธยมศึกษาตอนต้น ☐ มัธยมศึกษาตอนปลาย ☐ อนุปริญญา ☐ ปริญญาตรี ☐ สูงกว่าปริญญาตรี

นามบัญญัติ

5. รายได้เฉลี่ยต่อเดือนโดยประมาณของท่านมาจาก ☐ ☐ ☐ การขายผลผลิตทางการเกษตร ☐ .... บาท/เดือน ☐ รายได้เฉลี่ยต่อเดือน .... บาท / เดือน
6. อาชีพ

มาตราอัตราส่วน

ตัวอย่างประเภทของข้อมูล

ส่วนที่ 2 ความพึงพอใจเกี่ยวกับการจัดกิจกรรม

5 = พึงพอใจมากที่สุด 4 = พึงพอใจมาก 3 = พึงพอใจปานกลาง

2 = พึงพอใจน้อย 1 = พึงพอใจน้อยที่สุด

ส่วนที่ 3 ข้อคำถาม

รายการ	ระดับความพึงพอใจ						
	5	4	3	2	1	$\bar{X}$	S.D.
1. การประชาสัมพันธ์การจัดกิจกรรม							
2. ขั้นตอนการดำเนินกิจกรรม							
3. ระยะเวลาจัดกิจกรรม							
4. สถานที่จัดกิจกรรม							
5. เครื่องเสียงที่ใช้ในการจัดกิจกรรม							
6. บรรยากาศในการจัดกิจกรรม							
7. นักเรียนมีความตื่นตัวและกระตือรือร้นในการร่วมกิจกรรม							
8. ความเหมาะสมของการนำนักเรียนเข้าร่วมกิจกรรม							
9. การให้ความร่วมมือของนักเรียนและครู							
10. ประโยชน์ที่ได้รับจากกิจกรรม							
รวม							

มาตราอันตรภาค

ข้อคิดเห็นและข้อเสนอแนะ

....

....

....

แหล่งของข้อมูล

แหล่งปฐมภูมิ (Primary source หมายถึงข้อมูลหรือข้อเท็จจริงที่ได้จากการเก็บข้อมูลจากแหล่งข้อมูลโดยตรงเช่น ข้อมูลจาการสัมภาษณ์ การสังเกต การทดลอง การทดสอบ ข้อมูลการเข้าชั้นเรียนของนักเรียนแต่ละภาคการศึกษา แหล่งดังกล่าวเรียกว่า แหล่งปฐมภูมิ (Primary Source)

แหล่งทุติยภูมิ (Secondary source) ข้อมูลที่ผู้ใช้ไม่ได้เก็บรวบรวมเอง แต่มีผู้อื่นหรือ หน่วยงานอื่นๆ ทำการเก็บรวบรวมไว้แล้ว เช่น จากรายงาน ที่พิมพ์แล้ว หรือยังไม่ได้พิมพ์ของหน่วยงานของรัฐบาล สมาคม บริษัท สำนักงานวิจัย นักวิจัย วารสาร หนังสือพิมพ์ เป็นต้น แหล่งดังกล่าวเรียกว่า แหล่งทุติยภูมิ (Secondary Source)

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูล แบ่งเป็น 3 วิธี ดังนี้

1. การเก็บรวบรวมข้อมูลจากงานทะเบียน เช่น มหาวิทยาลัยจะมีการบันทึกหมายเหตุ รายวันในเรื่อง จำนวนอาจารย์ เจ้าหน้าที่ที่มาปฏิบัติราชการ, จำนวนอาจารย์, เจ้าหน้าที่ที่ลา, จำนวนอาจารย์ที่ไปราชการ เป็นต้น

2. การเก็บรวบรวมข้อมูลโดยการสำรวจ เป็นการเก็บรวบรวมข้อมูลจากหน่วยที่ศึกษาโดยตรง เช่น การสำรวจความคิดเห็นของประชาชนในการร่างรัฐธรรมนูญ ซึ่งหน่วยที่ศึกษาคือ ประชาชนคนไทย การเก็บรวบรวมข้อมูลโดยการสำรวจจะแบ่งออกเป็น 2 ประเภท

- การสำมะโน เก็บรวบรวมข้อมูลจากทุกๆหน่วยศึกษาของประชากร
- การสำรวจตัวอย่าง เก็บรวบรวมข้อมูลจากบางหน่วยศึกษาของประชากร

การเก็บรวบรวมข้อมูล

3. การเก็บรวบรวมข้อมูลจากการทดลอง ข้อมูลบางประเภทไม่สามารถหาได้จากการสำรวจ แต่จัดทำได้จากการทดลอง เช่น การศึกษาวิธีการปลูกพืชที่แตกต่างกัน 3 วิธี

- การนำเสนอด้วยบทความ หรือข้อความ
- การนำเสนอด้วยตาราง
- การนำเสนอด้วยกราฟ

การนำเสนอข้อมูล

การนำเสนอข้อมูลในรูปข้อความ

ตัวอย่างเช่น

ประชากรอายุตั้งแต่ 11 ปีขึ้นไปของกรุงเทพมหานครอ่านหนังสือพิมพ์รายวันฉบับหนึ่งมากที่สุด คือ ประมาณร้อยละ 38.8 รองลงมา ได้แก่ ประชากรภาคใต้ ภาคกลาง ภาคตะวันออกเฉียงเหนือ และภาคเหนือที่อ่านหนังสือพิมพ์รายวันฉบับนี้ ร้อยละ 26.0 18.2 10.8 และ 6.2 ตามลำดับ

การนำเสนอข้อมูล

การนำเสนอข้อมูลในรูปข้อความกึ่งตาราง

ตัวอย่างเช่น

ประชากรอายุตั้งแต่ 11 ปีขึ้นไปอ่านหนังสือพิมพ์รายวันฉบับหนึ่ง มีดังนี้

กรุงเทพมหานคร	ร้อยละ 38.8
ภาคใต้	ร้อยละ 26.0
ภาคกลาง	ร้อยละ 18.2
ภาคตะวันออกเฉียงเหนือ	ร้อยละ 10.8
ภาคเหนือ	ร้อยละ 6.2

การนำเสนอข้อมูล

การนำเสนอข้อมูลในรูปตาราง

4 รูปแบบ คือ

1. ตาราง**ทางเดียว** (one-way table)

ตารางที่ 1.1 ผลผลิตน้ำมันของโรงกลั่นแห่งหนึ่ง (ล้านลิตร) ในปี พ.ศ. 2547

ชนิดน้ำมัน	จำนวนที่ผลิตต่อปี (ล้านลิตร)
ดีเซล	20
เบนซิน 91	21
เบนซิน 95	25
แก๊สโซฮอล์	12
รวมทั้งหมด	78

การนำเสนอข้อมูล

2. ตาราง**สองทาง** (two –way table)

ตารางที่ 1.2 ผลผลิตน้ำมันของโรงกลั่นแห่งหนึ่ง (ล้านลิตร) ตั้งแต่ปี พ.ศ. 2544 – 2547

ชนิดน้ำมัน	พ.ศ. 2544	พ.ศ. 2545	พ.ศ. 2546	พ.ศ. 2547
ดีเซล	12	13	15	20
เบนซิน 91	13	15	19	21
เบนซิน 95	14	16	20	25
แก๊สโซฮอล์	9	10	12	12
รวมทั้งหมด	48	54	66	78

การนำเสนอข้อมูล

3. ตาราง**หลายทาง** (multi-way table)

ตารางที่ 1.3 ผลผลิตน้ำมันของโรงกลั่น 2 แห่ง (ล้านลิตร) ตั้งแต่ปี พ.ศ. 2546 และ ปี พ.ศ. 2547

ชนิดน้ำมัน	โรงกลั่นที่ 1		โรงกลั่นที่ 2	
	พ.ศ. 2546	พ.ศ. 2547	พ.ศ. 2546	พ.ศ. 2547
ดีเซล	15	20	8	10
เบนซิน 91	19	21	12	14
เบนซิน 95	20	25	12	15
แก๊สโซฮอล์	12	12	7	9
รวมทั้งหมด	66	78	39	48

การนำเสนอข้อมูล

4. ตาราง**แจกแจงความถี่** (frequency distribution)

ตารางที่ 1.4 ตารางแจกแจงความถี่ ประกอบด้วยข้อมูล รอยขีด และความถี่จากคะแนนสอบของนักศึกษา 60 คนดังนี้

28 22 20 17 16 25 18 22 28 17 19 22 22 21 19 27 27 25 23 24

28 26 21 18 24 21 24 22 20 22 24 28 16 23 22 25 24 22 25 21

17 28 24 27 23 22 22 29 16 20 21 21 26 27 28 24 28 16 23 22

คะแนน	รอยขีด	ความถี่
16	////	4
.	.	.
.	.	.
29	/	1
รวม		60

การนำเสนอข้อมูล

4. ตารางแจกแจงความถี่ (ต่อ)

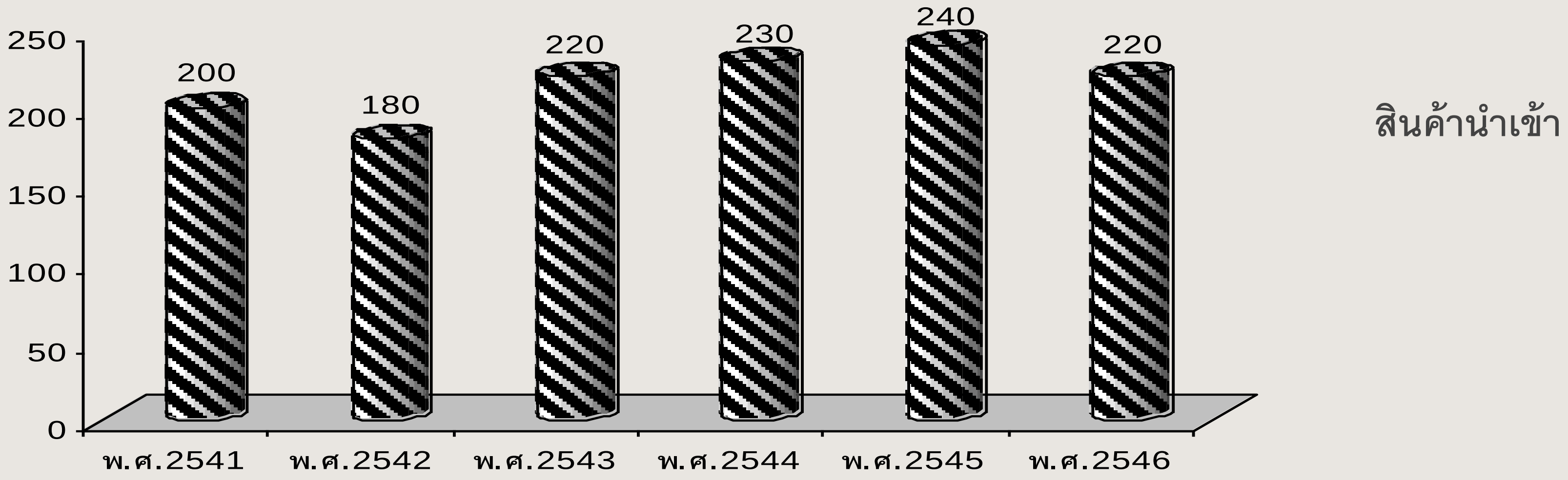
ตารางที่ 1.5 จำนวนโรงกลั่นที่ผลิตน้ำมันในประเทศไทย ปี พ.ศ. 2547 (สรุปตัวอย่าง)

ชนิดน้ำมัน	จำนวนโรงกลั่น (แห่ง)
ดีเซล	6
เบนซิน 91	8
เบนซิน 95	9
แก๊สโซฮอล์	2

การนำเสนอข้อมูล

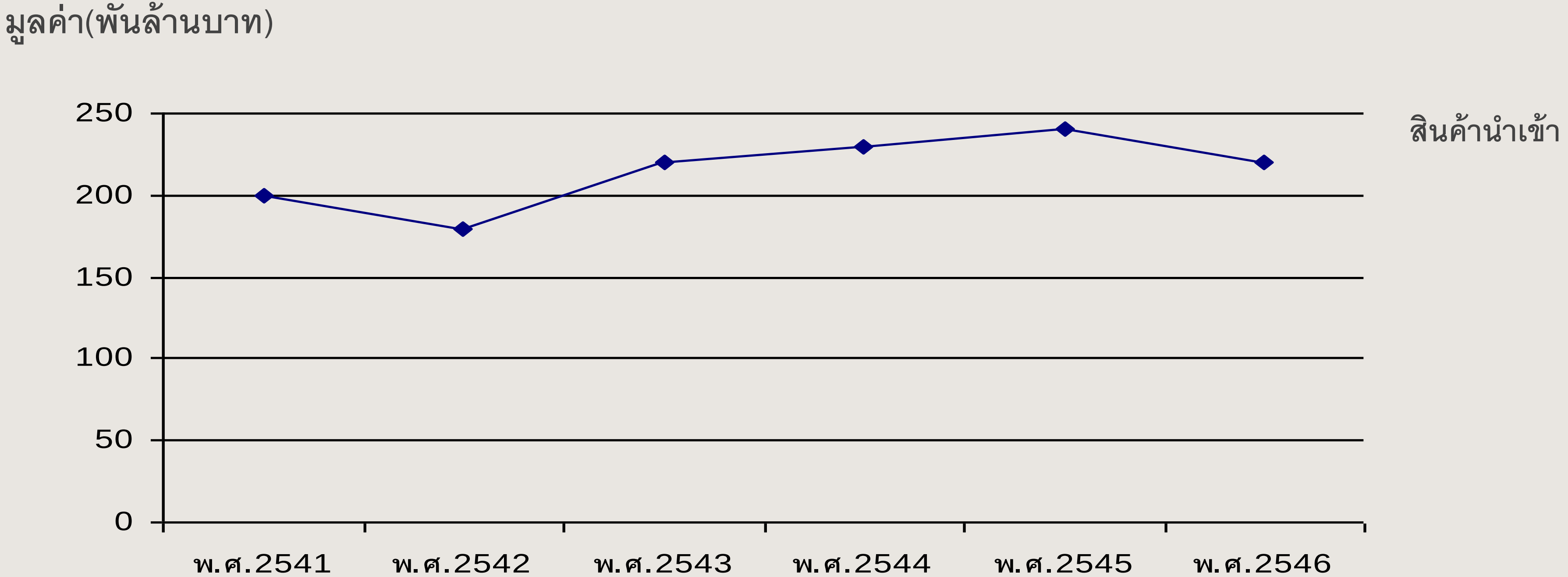
การนำเสนอข้อมูลในรูปแผนภูมิและกราฟ

แผนภูมิแท่งเชิงเดี่ยวและกราฟเส้นเชิงเดี่ยว



แผนภูมิแท่งเชิงเดี่ยว

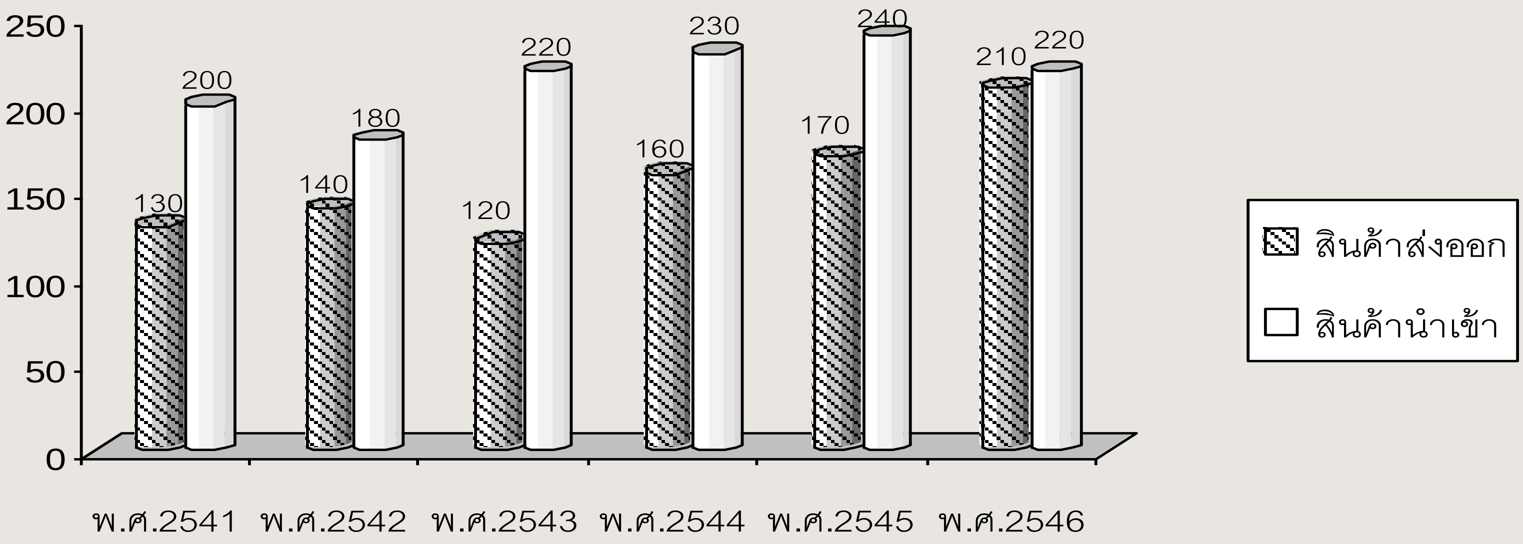
การนำเสนอข้อมูล



กราฟเส้นเชิงเดี่ยว

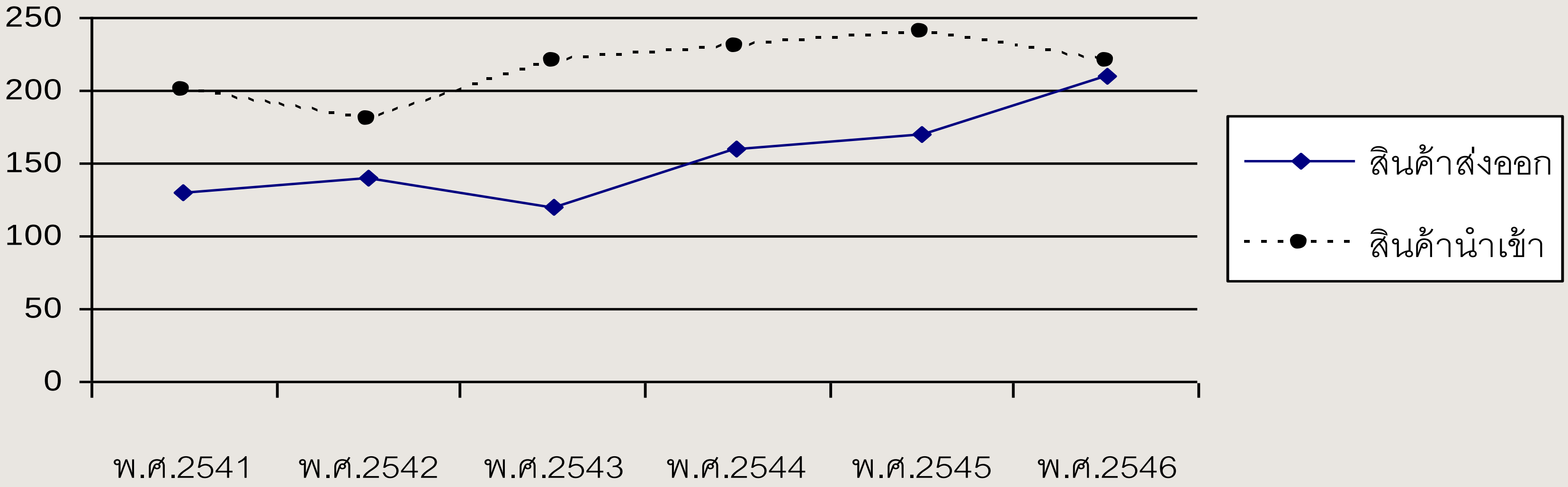
การนำเสนอข้อมูล

แผนภูมิแท่งเชิงซ้อน และกราฟเส้นเชิงซ้อน



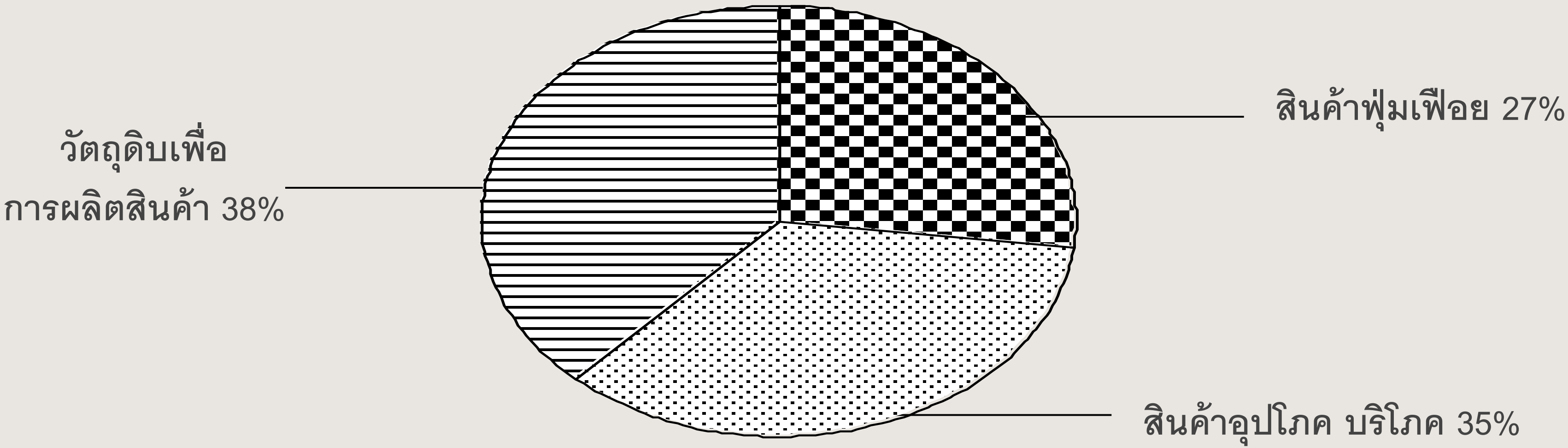
แผนภูมิแท่งเชิงซ้อน

การนำเสนอข้อมูล



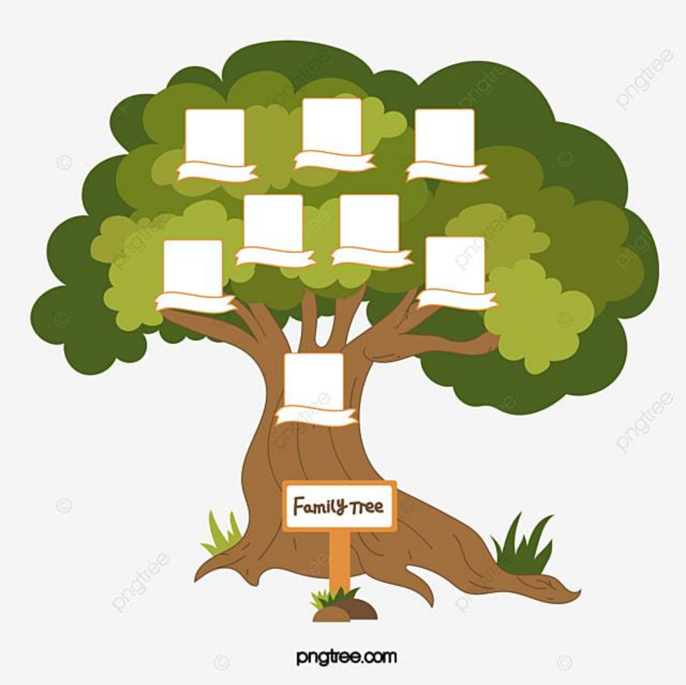
กราฟเส้นเชิงซ้อน

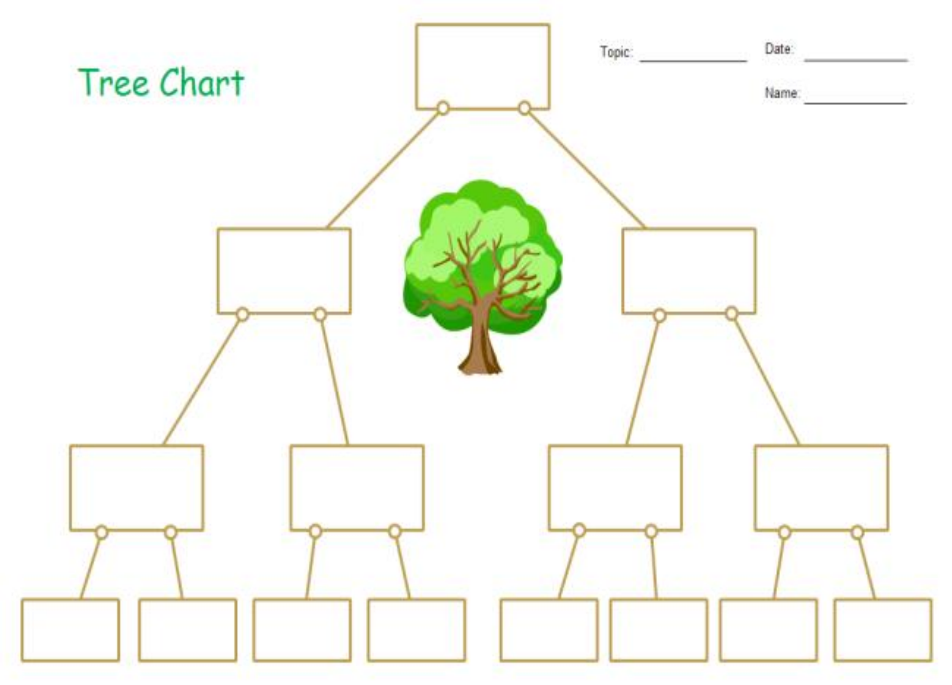
การนำเสนอข้อมูล



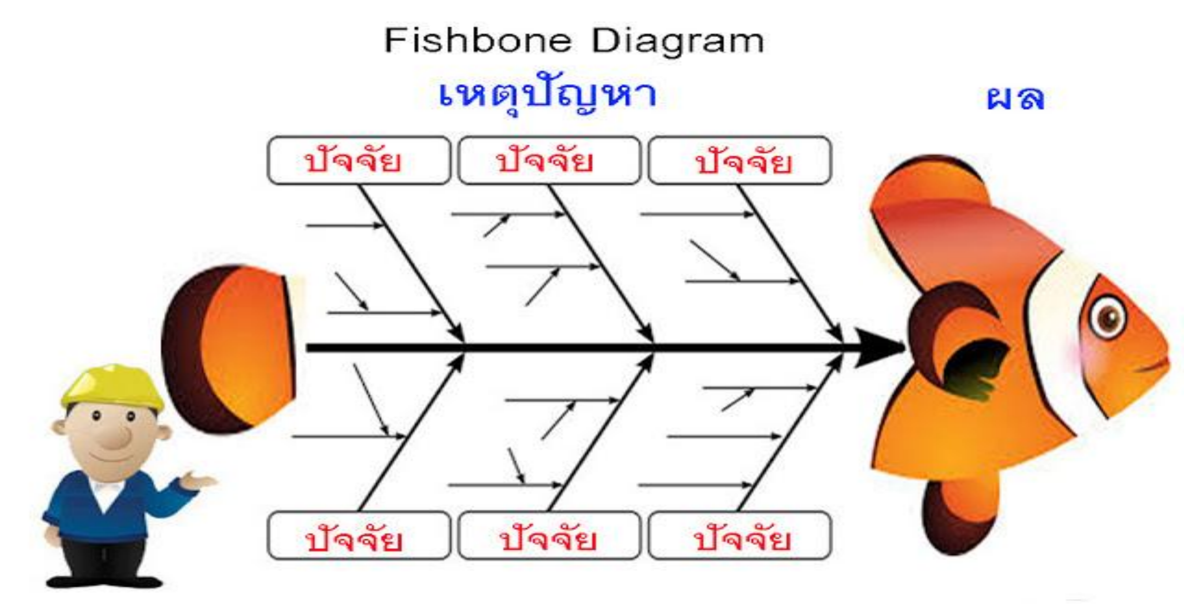
แผนภาพวงกลม

แผนภูมิต้นไม้

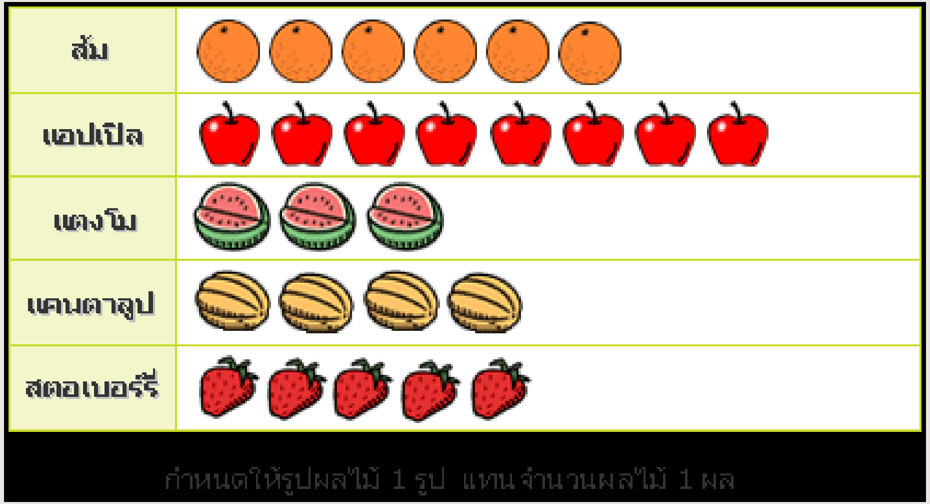




แผนภูมิก้างปลา



แผนภูมิภาพ


ส้ม
แอปเปิล
แตงโม
แคนตาลูป
สตอเบอร์รี่
กำหนดให้รูปผลไม้ 1 รูป แทนจำนวนผลไม้ 1 ผล

ประชากรและตัวอย่าง (Population and Sample)

ประชากร (Population) หมายถึง หน่วยต่างๆ ทุกหน่วยที่เราสนใจศึกษาที่ ให้ข้อมูลที่เราต้องการ เช่น การศึกษารายได้เฉลี่ยของข้าราชาการในจังหวัด ชลบุรีดังนั้นประชากรคือ ข้าราชการทุกคนในจังหวัดชลบุรี

- ประชากรที่นับได้ (Finite population)
- ประชากรที่นับไม่ได้ (Infinite population)

ประชากรและตัวอย่าง (Population and Sample)

ตัวอย่าง (Sample) หมายถึง หน่วยบางหน่วยในประชากรที่ต้องการศึกษา เช่น การศึกษารายได้เฉลี่ยของข้าราชการในจังหวัดชลบุรีตัวอย่างคือ ตัวแทนข้าราชการ 50 คน ที่ถูกสอบถามเกี่ยวกับรายได้

- ตัวอย่างสุ่ม (Random Sample) คือ ตัวอย่างที่ได้มาโดยการสุ่ม
- การเลือกตัวอย่างแบบสุ่ม (Random Sampling) คือ การได้มาซึ่งตัวอย่างที่เลือกมา
- หน่วยการเลือกตัวอย่าง (Sampling Units) คือ ทราบสมาชิกแต่ละตัวในประชากร
- กรอบตัวอย่าง (Sample Frame) คือ รายชื่อของสมาชิดทุกสมาชิกที่อยู่ในประชากร

ประชากรและตัวอย่าง (Population and sample)

ประชากรและตัวอย่าง (Population and Sample)

พารามิเตอร์ เป็นค่าตัวเลขที่คำนวณได้จากข้อมูลประชากร เพื่อใช้สรุป หรือบรรยายแสดงลักษณะของประชากร เช่น ค่าเฉลี่ยของประชากร แทนด้วยสัญลักษณ μ (มิว) ค่าส่วนเบี่ยงเบนมาตรฐานของประชากร แทนด้วยสัญลักษณ์ σ (ซิกม่า) เป็นต้น

ค่าสถิติ เป็นค่าตัวเลขที่คำนวณได้จากข้อมูลของกลุ่มตัวอย่าง เพื่อใช้สรุป หรือบรรยายแสดงลักษณะของกลุ่มตัวอย่าง เช่น ค่าเฉลี่ยของกลุ่มตัวอย่าง แทนด้วยสัญลักษณ $(\bar{x})$ ค่าส่วนเบี่ยงเบนมาตรฐานของของกลุ่มตัวอย่าง แทนด้วยสัญลักษณ์ S.D. เป็นต้น

ตารางที่ เปรียบเทียบสัญลักษณ์ของพารามิเตอร์และค่าสถิติ

ค่าที่ใช้บรรยายลักษณะ	พารามิเตอร์(ประชากร)	ค่าสถิติ (ตัวอย่าง)
ค่าเฉลี่ยเลขคณิต (Mean)	μ	$\bar{x}$
ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)	σ(ซิกม่า)	s หรือ S.D.
ความแปรปรวน (Variance)	σ^2	S^2 หรือ S.D^2
สัดส่วน (Proportion)	π (พาย) หรือ P	P
สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)	ρ (โร)	r
สัมประสิทธิ์การถดถอย (Regression Coefficient)	β (เบต้า)	b
ค่าอื่นๆ	θ (เซต้า)	$\hat{\sigma}$

ระเบียบวิธีทางสถิติ

มีอยู่ 5 ขั้นตอน ดังนี้

ขั้นที่ 1

การวางแผนงาน (planning) เป็นการกำหนดกรอบในการดำเนินการต่าง ๆ

ขั้นที่ 2

การเก็บรวบรวมข้อมูล (collecting data)

อาจเก็บจากประชากรทั้งหมดเช่นการทำสำมะโนประชากร หรือเก็บจากกลุ่มตัวอย่าง เพื่อความประหยัดของงานนั้น ๆ แล้วแต่ความเหมาะสมในแต่ละงาน

ระเบียบวิธีทางสถิติ

ขั้นที่ 3 **การนำเสนอข้อมูล**

เป็นวิธีการที่จะช่วยให้ผู้ที่สนใจงานนั้น สามารถทำความเข้าใจและนำไปใช้ได้ โดยอาจนำเสนอในรูปบทความ นำเสนอในรูปตาราง นำเสนอในรูปแผนภูมิแท่งและนำเสนอในรูปกราฟวงกลม เป็นต้น

ขั้นที่ 4 **การวิเคราะห์ (analyzing data)**

เป็นการวิเคราะห์ข้อมูลตั้งแต่ระดับพื้นฐานจนถึงการวิเคราะห์เพื่อตอบสนองวัตถุประสงค์ของงาน

ระเบียบวิธีทางสถิติ

ขั้นที่ 5

การตีความหมายและการสรุป (interpretation & conclusion)

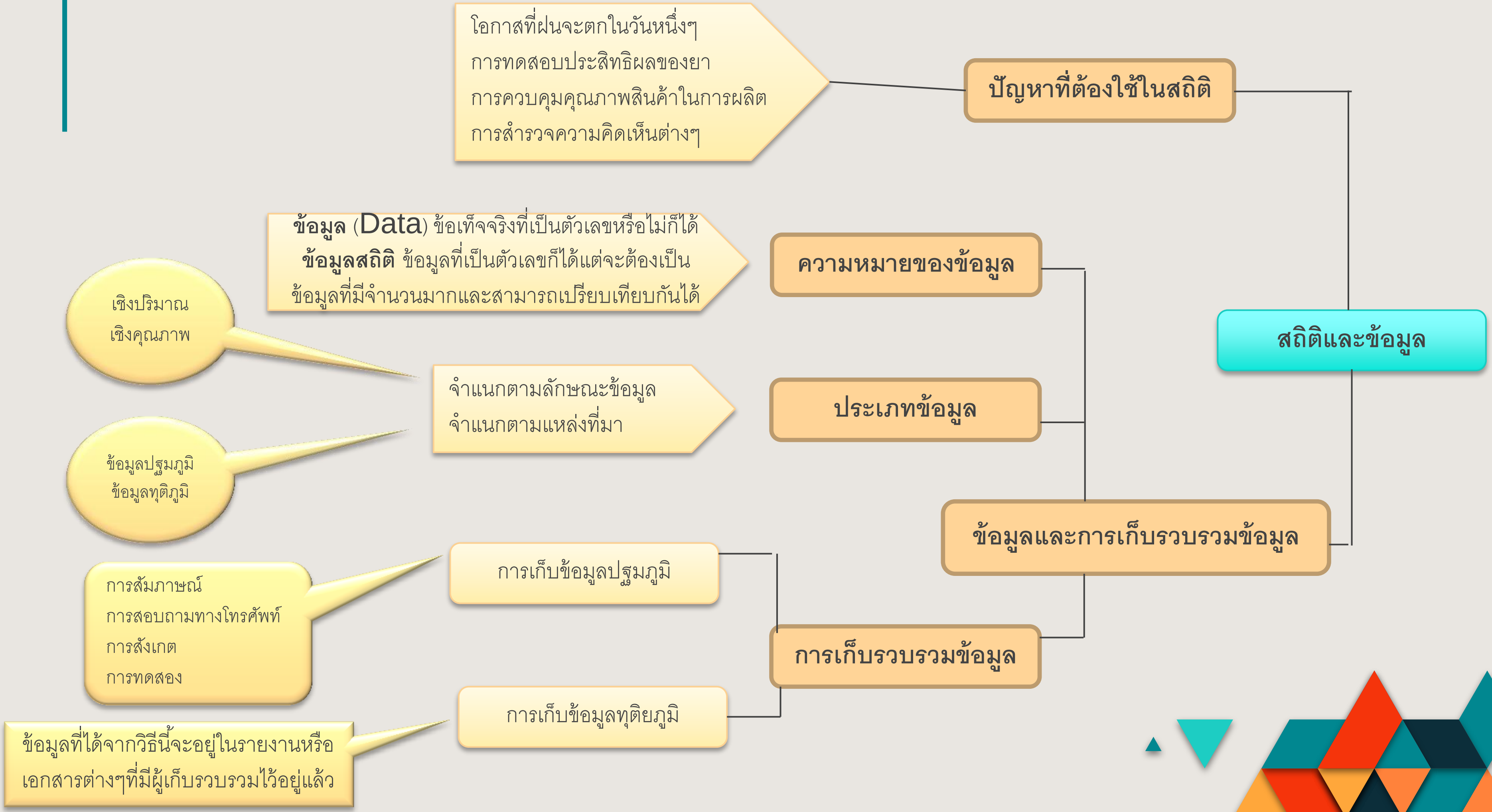
หลังจากวิเคราะห์ข้อมูลแล้วจะต้องให้ความหมายและสรุปผลเป็นภาษาที่บุคคลทั่วไปที่ไม่ใช่นักสถิติสามารถอ่านและเข้าใจได้

ประโยชน์ของสถิติ

ประโยชน์ของสถิติมิใช่เพียงแต่ใช้เป็นเครื่องมือในการช่วยตัดสินใจ และกำหนดนโยบายต่างๆ ให้เป็นไปอย่างมีประสิทธิภาพเท่านั้น เมื่อพิจารณาอีกด้านหนึ่งจะเห็นว่า สถิติเป็นเครื่องมือที่ทรงคุณประโยชน์อย่างยิ่งในการประเมินผลงานโครงการ ต่าง ๆ ที่จัดทำไปแล้วว่าได้ผลตามเป้าหมายที่วางไว้เพียงไร สมควรที่จะต้องปรับปรุงหรือแก้ไขโครงการนั้น ๆ หรือไม่อย่างไรอีกด้วย เนื่องจากสถิติมีขอบข่ายกว้างขวาง ได้รับการนำไปใช้ประโยชน์แทบทุกแขนงวิชาการ ดังนั้น นักบริหาร นักวิชาการ หรือแม้แต่สามัญชนทั่วไป จึงควรมีความรู้ทางสถิติตามสมควร หรือตามความจำเป็น กล่าวคือ อย่างน้อยก็สามารถอ่านข้อมูลจากตาราง จากแผนภูมิ หรือจากแผนภาพต่าง ๆ ให้เข้าใจได้ถูกต้อง ประโยชน์ของสถิติสรุปได้

ประโยชน์ของสถิติ

1. ช่วยในการ **วิเคราะห์ข้อมูล** เพื่อให้เข้าใจสถานการณ์จริง
2. ใช้ในการ **ตัดสินใจ** อย่างมีเหตุผล
3. สนับสนุนการวางแผน การคาดการณ์ และการวิจัย
4. ใช้ในการ **เปรียบเทียบข้อมูล** หรือดูแนวโน้มการเปลี่ยนแปลง


สถิติและข้อมูล
ปัญหาที่ต้องใช้ในสถิติ
โอกาสที่ฝนจะตกในวันหนึ่งๆ
การทดสอบประสิทธิผลของยา
การควบคุมคุณภาพสินค้าในการผลิต
การสำรวจความคิดเห็นต่างๆ
ข้อมูลและการเก็บรวบรวมข้อมูล
ความหมายของข้อมูล
ข้อมูล (Data) ข้อเท็จจริงที่เป็นตัวเลขหรือไม่ก็ได้
ข้อมูลสถิติ ข้อมูลที่เป็นตัวเลขก็ได้แต่จะต้องเป็นข้อมูลที่มีจำนวนมากและสามารถเปรียบเทียบกันได้
ประเภทข้อมูล
จำแนกตามลักษณะข้อมูล
จำแนกตามแหล่งที่มา
เชิงปริมาณ
เชิงคุณภาพ
ข้อมูลปฐมภูมิ
ข้อมูลทุติภูมิ
การเก็บรวบรวมข้อมูล
การเก็บข้อมูลปฐมภูมิ
การสัมภาษณ์
การสอบถามทางโทรศัพท์
การสังเกต
การทดสอง
การเก็บข้อมูลทุติยภูมิ
ข้อมูลที่ได้จากวิธีนี้จะอยู่ในรายงานหรือเอกสารต่างๆที่มีผู้เก็บรวบรวมไว้อยู่แล้ว

สถิติและข้อมูล

- **ความหมายของสถิติ**
  - ตัวเลขที่ใช้บรรยายเหตุการณ์หรือข้อเท็จจริงของเรื่องต่างๆที่เราต้องศึกษา
  - ศาสตร์หรือวิชาที่ว่าด้วยหลักการและระเบียบวิธีการทางสถิติเรียกว่า **"สถิติศาสตร์"**
  - ค่าที่คำนวณขึ้นมาจากตัวอย่างเพื่อแสดงถึงจุดลักษณะบางอย่างของข้อมูลชุดนั้น
- **กระบวนการแก้ปัญหาทางสถิติ**
  - การสร้างคำถามทางสถิติ
  - การเก็บรวบรวมข้อมูล
  - การวิเคราะห์ข้อมูล
  - การแปลความหมายข้อมูล
- **ประเภทของสถิติ**
  - สถิติเชิงพรรณา
    - สถิติที่ใช้อธิบายคุณลักษณะต่างๆ ของสิ่งที่ต้องการศึกษาในกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ
  - สถิติอ้างอิง
    - สถิติที่ใช้อธิบายคุณลักษณะต่างๆ ของสิ่งที่ต้องการศึกษาในกลุ่มใดกลุ่มหนึ่งและสามารถอ้างอิงไปยังกลุ่มอื่นๆได้
- **ประโยชน์สถิติ**
  - ด้านการพัฒนาประเทศ
  - ด้านธุรกิจ
  - ด้านการเกษตร
  - ด้านการศึกษา

Assignment

โรงเรียนประถมศึกษาแห่งหนึ่งในปีการศึกษา 2559 มีจำนวนนักเรียนดังนี้ โดยการนำเสนอข้อมูลเป็นกราฟแท่ง และกราฟเส้น

ชั้น	ชาย	หญิง	รวม
ปีที่ 1	95	65	160
ปีที่ 2	43	79	122
ปีที่ 3	64	81	145
ปีที่ 4	48	70	118
ปีที่ 5	84	72	156
ปีที่ 6	69	85	154

Assignment

ในการบันทึกของเจ้าหน้าที่หอสมุดปรากฏว่ามีจำนวนนักศึกษามาใช้บริการในหอสมุดตั้งแต่วันจันทร์ถึงวันศุกร์ดังนี้

วัน	จันทร์	อังคาร	พุธ	พฤหัสบดี	ศุกร์
จำนวนคน	224	305	333	193	250

ให้นำเสนอข้อมูลด้วย **กราฟแท่ง กราฟเส้น และกราฟวงกลม**

THANKS!

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik

Please keep this slide for attribution

บทที่ 2

การแจกแจงความถี่และการวิเคราะห์แนวโน้มเข้าสู่ส่วนกลาง

ผศ ดร มนันยา มีนคร
แขนงการจัดการธุรกิจบริการ
มหาวิทยาลัยราชภัฏสวนสุนันทา

เนื้อหา.....บทเรียน 2

1. การแจกแจงความถี่
2. ค่าเฉลี่ย (average)
3. มัธยฐาน (median)
4. ฐานนิยม (mode)

ความหมายของการแจกแจงความถี่

การแจกแจงความถี่ หมายถึง การเรียงลำดับข้อมูลดิบที่รวบรวมมาได้โดยจัดเป็นหมวดหมู่แล้วหาจำนวนของข้อมูลในแต่ละหมู่ จำนวนในแต่ละหมู่นี้เรียกว่า “ความถี่ จะเขียนแทนสัญลักษณ์ f “

การแจกแจงความถี่ทำได้ 2 แบบ ดังต่อไปนี้

1. การแจกแจงความถี่ด้วยตาราง เรียกว่า ตารางแจกแจงความถี่
2. การแจกแจงความถี่ด้วยกราฟหรือแผนภูมิ

ตารางแจกแจงความถี่ (Frequency Distribution Table)

ลักษณะของตารางแจกแจงความถี่โดยทั่วไปจะประกอบไปด้วย คะแนนหรือข้อมูล รอยคะแนนหรือรอยขีดและความถี่ ดังตัวอย่าง

ตัวอย่าง คะแนนสอบวิชาคณิตศาสตร์ ของนักเรียนวิชาเอกคณิต ชั้นปี 2

คะแนน	รอยคะแนน	ความถี่
31	/	1
32	//	2
33	///	3
34	////	4

ตารางแจกแจงความถี่ (Frequency Distribution Table)

ถ้าข้อมูลมีจำนวนมากไม่สะดวกที่จะจัดตาราง ก็อาจจะจัดตารางแจกแจงความถี่ได้อีกวิธีหนึ่งโดยการจัดคะแนนที่ใกล้เคียงกันเป็นหมู่หรือเป็นชั้น (groupn or class) ช่วงกว้างของคะแนนแต่ละชั้นเรียกว่า อันตราภาคชั้น (class interval)

ตารางแจกแจงความถี่ (Frequency Distribution Table)

ตัวอย่าง คะแนนทดสอบวิชาคณิตศาสตร์ครั้งที่ 1 ของนักศึกษา 50 คน

คะแนน	รอยขีด	ความถี่
10-19	//	2
20-29	//// ///	8
30-39	//// ////	9
40-49	//// //// ////	14
50-59	//// ///	8
60-69	//// /	6
70-79	///	3
		50

ตารางแจกแจงความถี่ (Frequency Distribution Table)

โดยการสร้างตารางแจกแจงความถี่มีส่วนประกอบดังนี้

1. อัตรภาคชั้น (class intervals) คือ ช่องกว้างของคะแนนในแต่ละชั้น

   = ขีดกำจัดบน-ขีดจำกัดล่าง

2. พิสัย (range) คือ คะแนนสูงสุด - คะแนนต่ำสุด

3. ความกว้างชั้น (class width) คือ ความกว้างข้อมูลในแต่ละชั้นซึ่งหาได้จากพิสัยหารด้วยจำนวนชั้นทั้งหมดของตารางแจกแจงความถี่หรืออัตรภาคชั้น (สัญลักษณ์ I)

4. คะแนนทุกๆชั้นจะมีขีดจำกัดล่าง (Low limit) และขีดจำกัดบน (upper limit) ขีดจำกัดทั้งสองนี้เรียกว่า **ขีดกำจัดที่แท้จริง**

ตารางแจกแจงความถี่ (Frequency Distribution Table)

5. ขอบขน(+0.5)และ ขอบล่าง (- 0.5)

อันตรภาคชั้น	ขอบบน	ขอบล่าง
41 - 50	50.5	40.5
51 - 60	60.5	50.5

$$\text{ขอบล่าง} = \frac{\text{ค่าต่ำสูงสุดของอันตรภาคชั้นนั้น} + \text{ค่าสูงสุดของอันตรภาคชั้นที่ต่ำกว่าหนึ่งชั้น}}{2}$$

$$\text{ขอบบน} = \frac{\text{ค่าสูงสุดของอันตรภาคชั้นนั้น} + \text{ค่าต่ำสุดของอันตรภาคชั้นที่สูงกว่าหนึ่งชั้น}}{2}$$

ตารางแจกแจงความถี่ (Frequency Distribution Table)

6. จุดกึ่งกลางชั้น (mid midpoint) คือ คะแนนที่อยู่ในตำแหน่งกลางของแต่ละชั้น เช่น ระหว่าง 10-19 จุดกึ่งกลางชั้น 14.5 และจุดกลางชั้นของคะแนน 20-29 คือ 24.5 เป็นต้น หรหรือ เราอาจจะหาจุดกึ่งกลางชั้นได้จากการแบ่งครั้งผลรวมของขีดจำกำกัดล่างกับขีดกำจัดบนของคะแนนในแต่ละชั้น เช่น

	10-19	20-29	30-39	40-49
9.5	19.5	29.5	39.5	49.5

$$\frac{9.5+19.5}{2} = 14.5$$

$$\frac{19.5+29.5}{2} = 24.5$$

หรืออาจใช้คะแนนตัวบนกับตัวล่างแทนขีดจำกัดที่แท้จริงก็ได้เช่น

$$\frac{10+19}{2} = 14.5$$

$$\frac{20+29}{2} = 24.5$$

ตารางแจกแจงความถี่ (Frequency Distribution Table)

จุดกึ่งกลางชั้น (mid midpoint)

$$\text{จุดกึ่งกลางชั้น} = \frac{\text{ขอบบน} + \text{ขอบล่าง}}{2}$$

7. นับคะแนน คือ รอยขีด (mark)

8. ความถี่ (Frequency) คือ ตัวเลขที่แทนรอยขีด

ตารางแจกแจงความถี่ (Frequency Distribution Table)

9. ความถี่สะสม (cumulative frequency) เป็นผลรวมของความถี่ของค่านั้นหรืออันตรภาคชั้นนั้นกับอันตรภาคชั้นที่มีช่วงคะแนนต่ำกว่าทั้งหมด สัญลักษณ์ F

10. ความถี่สัมพัทธ์(relative frequency) เป็นอัตราส่วนของความถี่ของแต่ละชั้นข้อมูลกับความถี่รวม หรือจำนวนข้อมูลทั้งหมด

11. ความถี่สะสมสัมพัทธ์(relative cumulative frequency) เป็นผลรวมสะสมของความถี่สัมพัทธ์สะสมจากชั้นที่มีค่าของข้อมูลต่ำสุดไปยังค่าของข้อมูลสูงสุด

ตารางแจกแจงความถี่ (Frequency Distribution Table)

ตัวอย่าง ตารางคะแนนสอบวิชาคณิตศาสตร์ครั้งที่ 1 ของนักศึกษา

คะแนน	ความถี่	ขีดกำจัดที่แท้จริง	จุดกึ่งกลางชั้น
10-19	2	9.5-19.5	14.5
20-29	8	19.5-29.5	24.5
30-39	9	29.5-39.5	34.5
40-49	14		
50-59	8		
60-69	6		
70-79	3		
	N=50		

ตารางแจกแจงความถี่ ความถี่สะสม ความถี่สัมพัทธ์ และความถี่สะสมสัมพัทธ์ จำแนกตามช่วงอายุของผู้เข้าร่วมสัมมนาทางวิชาการ

อายุ	ความถี่ (f)	ความถี่สะสม (F)	ความถี่สัมพัทธ์	ความถี่สะสมสัมพัทธ์
20 - 32	10	10	0.250	0.250
33 - 45	6	16	0.150	0.400
46 - 58	9	25	0.225	0.625
59 - 71	4	29	0.100	0.725
72 - 84	8	37	0.200	0.925
85 - 98	3	40	0.075	1.000
รวม	40		1.000	

10+6 = 16

10+6+9 = 25

10/40 = 0.250

6/40 = 0.150

10/40 = 0.250

16/40 = 0.400

การวัดแนวโน้มเข้าสู่ส่วนกลาง

การวัดแนวโน้มเข้าสู่ส่วนกลาง เป็นระเบียบวิธีการทางสถิติในสถิติเชิงพรรณา ที่ใช้ในการหาค่าคะแนนกลาง หรือค่าเฉลี่ย เพื่อใช้เป็นตัวแทนแสดงขนาดและลักษณะของข้อมูลแต่ละชุด ดังนั้นประโยชน์จากการวัดแนวโน้มเข้าสู่ส่วนกลางคือ ทำให้ได้ตัวแทนของข้อมูลซึ่งเป็นตัวเลขจำนวนเดียวที่แทนคะแนนทั้งหมดในข้อมูลแต่ละชุดมารายงาน โดยมาไม่จำเป็นต้องนำข้อมูลทั้งชุดมาพิจารณา การวัดมี 3 วิธีคือ

1. มัชฌิมเลขคณิต
2. มัธยฐาน
3. ฐานนิยม

การวัดแนวโน้มเข้าสู่ส่วนกลาง

การคำนวณหาค่ากลางทั้งสามชนิดนี้โดยทั่วไปแบ่งออกได้เป็น 2 กรณีใหญ่ คือ

- 1.การหาค่ากลางของข้อมูลที่ไม่ได้แจกแจงความถี่ (Ungrouped data) ซึ่งค่าที่ได้เป็นค่ากลางที่ถูกต้องแน่นอนของข้อมูลชุดนั้น
- 2.การหาค่าลางของข้อมูลที่แจกแจงความถี่แล้ว (Grouped Data) ซึ่งค่าที่ได้เป็นค่ากลางโดยประมาณของข้อมูลชุดนั้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ค่าเฉลี่ยเลขคณิต (arithmetic mean) เป็นค่าที่หาได้จากการหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลที่มี

ค่าเฉลี่ยเลขคณิตกรณี ข้อมูลไม่แจกแจงความถี่ โดยการหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูล นั่นคือ ถ้าให้เป็นข้อมูลขนาด จากประชากร และเป็นข้อมูลขนาดจากกลุ่มตัวอย่าง จะได้

ค่าเฉลี่ยเลขคณิต (arithmetic mean)

ค่าเฉลี่ยเลขคณิตของประชากร คือ $\mu = \frac{\sum_{i=1}^{N} x}{N}$

ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง คือ $\bar{x} = \frac{\sum_{i=1}^{n} x}{N}$

ค่าเฉลี่ยเลขคณิต (arithmetic mean)

<u>ตัวอย่าง</u> จงหาค่าเฉลี่ยของข้อมูลที่กำหนดให้ต่อไปนี้ 22 23 24 24 25 25 21 23 15 20

วิธีทำ $\bar{x} = \frac{\sum x}{N} = \frac{22+23+24+24+25+25+21+23+15+20}{10}$

$= \frac{222}{10}$

$= 22.2$

นั่นคือ ค่าเฉลี่ยของข้อมูลนี้เท่ากับ 22.2 คะแนน

ค่าเฉลี่ยเลขคณิต (arithmetic mean)

การหาค่าเฉลี่ยในกรณีที่ข้อมูลแจกแจงความถี่

ค่าเฉลี่ยเลขคณิตของประชากรคือ $\mu = \dfrac{\sum_{i=1}^{k} f\, x_i}{N}$

ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่างคือ $\bar{X} = \dfrac{\sum_{i=1}^{k} f\, x_i}{N}$

• ตัวอย่าง จงหา $\bar{x}$ 11 12 13 14 15

x (อันตรภาคชั้น)	f (หรือ N ก็ได้)	fx
15	3	45
14	2	28
13	5	65
12	3	36
11	2	22
Σ	15	196

ค่าเฉลี่ยเลขคณิต (arithmetic mean)

$$\therefore \overline{X} = \frac{\Sigma f x}{N}$$

$$= \frac{196}{15}$$

$$= 13.07$$ #

<u>ตัวอย่าง</u> จงหามัชฌิมเลขคณิตของคะแนนจากตารางแจกแจงความถี่ที่กำหนดให้ต่อไปนี้

คะแนน	ความถี่ (f)
10-19	2
20-29	8
30-39	9
40-49	14
50-59	8
60-69	6
70-79	3
รวม	50

วิธีทำ

10+19 / 2 = 14.5

คะแนน	ความถี่ (f)	ค่ากึ่งกลาง x_i	ผลรวมคะแนนในชั้น $f\,x_i$
10-19	2	14.5	29
20-29	8	24.5	196
30-39	9	34.5	310
40-49	14	44.5	623
50-59	8	54.5	436
60-69	6	64.5	387
70-79	3	74.5	223.5
รวม	50		2,205

14.5*2 = 29.0

ค่าเฉลี่ยเลขคณิต (arithmetic mean)

วิธีทำ คำนวณค่าเฉลี่ยได้

$$\bar{X} = \frac{\sum fx}{N}$$

$$= \frac{2,205}{50}$$

$$= 44.1 \ \#$$

นั่นคือ ค่าเฉลี่ยของข้อมูลชุดนี้ คือ 44.1

ตัวอย่าง ผลการสอบวิชา ภาษาอังกฤษของนักศึกษาชั้นปีที่ **1** ได้คะแนนตามตารางแจกแจงความถี่ที่กำหนดให้ จงคำนวณหามัชฌิมเลขคณิต ของคะแนนวิชาภาษาอังกฤษของนักศึกษา **40** คน

คะแนน (X)	ความถี่ (f)	fx
30	1	30
31	1	31
32	3	96
33	5	165
34	7	238
35	9	315
36	6	216
37	5	185
38	2	76
39	1	39
	N=40	1,391

ค่าเฉลี่ยเลขคณิต (arithmetic mean)

วิธีทำ คำนวณค่าเฉลี่ยได้

$$\bar{X} = \frac{\sum fx}{N}$$

$$= \frac{1,391}{40}$$

$$= 34.775 \ \#$$

นั่นคือ ค่าเฉลี่ยของคะแนนวิชาภาษาอังกฤษของนักศึกษา 40 คน = 34.775

การวัดแนวโน้มเข้าสู่ส่วนกลาง

มัธยฐาน (Median) ค่าที่มีตำแหน่งอยู่กึ่งกลางของข้อมูลทั้งหมด ข้อมูลจะเรียบเรียงจากค่าน้อยที่สุดไปหาค่าที่มากที่สุด หรือจากค่าที่มากที่สุดไปหาค่าที่น้อยที่สุด เราอาจใช้ตัวย่อ "Med" หรือ Md แทนค่ามัธยฐานของข้อมูล

มัธยฐาน (Median)

มัธยฐาน (Median : Med) โดยวิธีการหาได้ 2 วิธี ดังนี้

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่ คือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งชุด เมื่อเรียงจากน้อย ⇄ มาก

ถ้าข้อมูลเป็นจำนวนคี่ มัธยฐาน คือ ค่ากลางของข้อมูลตัว

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

Mdn คือ ค่ามัธยฐาน

N คือ จำนวนข้อมูลทั้งหมด

ตัวอย่าง จงหามัธยฐานของข้อมูลที่กำหนดให้ 1, 3, 2, 2, 5, 3, 4, 4, 3

วิธีทำ จัดเรียงข้อมูลใหม่ได้ดังนี้ 1, 2, 2, 3, 3, 3, 4, 4, 5

เนื่องจากข้อมูลเป็นจำนวนคี่ ดังนั้น ตำแหน่งกึ่งกลาง คือ $\frac{N+1}{2} = \frac{9+1}{2} = 5$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ **3** #

มัธยฐาน (Median)

ตัวอย่าง ข้อมูลที่ไม่ได้แจกแจงความถี่

10 , 12 , 13 , 17 , 18 $= 13$

10 , 12 , 13 , 17 , 18 , 20 $= \frac{13+17}{2}$

$= 15$ #

กรณีข้อมูลที่แจกแจงความถี่แล้ว

1. หาความถี่สะสม
2. หาตำแหน่งของมัธยฐาน Med = $\frac{N}{2}$
3. สมการ

$$Me = L + I\frac{\left[\frac{N}{2} - \sum F\right]}{f}$$

เมื่อ

L ขอบเขตล่างของชั้นที่มัธยฐานอยู่

$N/2$ ตำแหน่งของมัธยฐาน

I ความกว้างของอัตรภาคชั้น

F ผลรวมของความถี่ของทุกชั้นที่มีข้อมูลต่ำกว่าชั้นที่มีมัธยฐานอยู่

f ความถี่ของชั้นที่มีมัธยฐานอยู่

ตัวอย่าง จงหามัธยฐานของข้อมูลที่กำหนดให้ต่อไปนี้

คะแนน	ความถี่	ความถี่สะสม
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N=50	

มัธยฐาน (Median)

วิธีทำ เนื่องจาก $\frac{N}{2} = \frac{50}{2} = 25$ ดังนั้นมัธยฐานอยู่ในชั้น 40-49

และได้ว่า L = 39.5 , I = 10 , F = 19 , f = 14

เข้าสูตรสมการ $Me = L + I \left[\frac{\frac{N}{2} - \sum F}{f} \right]$

คะแนน	ความถี่	ความถี่สะสม
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N=50	

มัธยฐาน (Median)

จะได้

$$\text{Med} = 39.5 + 10\left[\frac{25 - 19}{14}\right]$$

$$= 43.8\#$$

นั่นคือ มัธยฐานของข้อมูลชุดนี้คือ 43.8 #

ตัวอย่าง จงหาค่ามัธยฐานของข้อมูลในตารางดังต่อไปนี้

วิธีทำ 1. หาค่าความถี่สะสมของข้อมูล

ข้อมูล	ความถี่ (f)	ความถี่สะสม(F)
11-15	2	2
16-20	8	10
21-25	19	29
26-30	10	39
31-35	1	40
รวม	n=40	

ข้อมูล	ความถี่ (f)	ความถี่สะสม(F)
11-15	2	2
16-20	8	F = 10
21-25	f= 19	29
26-30	10	39
31-35	1	40
รวม	n=40	

หาตำแหน่งของมัธยฐาน $\frac{40}{2} = 20$

มัธยฐานจะอยู่ระหว่างค่าของข้อมูล 21 และ 25

หาค่ามัธยฐานโดยใช้สูตร ดังนี้

$$\mathbf{Me} = \mathbf{L} + \mathbf{i}\left(\frac{\frac{\mathbf{n}}{\mathbf{2}} - \sum F}{\mathbf{f}}\right)$$

$$= 20.5 + 5\left(\frac{20 - 10}{19}\right)$$

$$= 20.5 + 5\,(0.526)$$

$$= 23.13 \quad \#$$

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ฐานนิยม (Mode : Mo) คือ ค่าของข้อมูลที่ปรากฏย่อยครั้งที่สุดหรือมีความถี่สูงสุดทั้งนี้ข้อมูลชุดหนึ่งอาจไม่มีฐานนิยมเลยก็ได้ แต่ถ้ามีก็อาจมีได้มากกว่าหนึ่งค่า โดยวิธีการหาได้ 2 วิธี ดังนี้

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่ ฐานนิยม คือ ค่าของข้อมูลที่มีความถี่สูงสุด

ตัวอย่าง จงหาฐานนิยมของข้อมูลต่อไปนี้

ก. 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12 และ 18

Mo = 9

ข. 3, 5, 8, 10, 12, 15, 16

ข้อมูลชุดนี้ไม่มี Mo เนื่องจากไม่มีข้อมูลค่าใดมีความถี่สูงสุด

ค. 2, 3, 4, 4, 4, 5, 5, 7, 7, 7 และ 9

Mo = 4 และ 7

กรณีข้อมูลที่แจกแจงความถี่แล้ว

คำนวณหาฐานนิยมได้จากสูตร ต่อไปนี้

$$Mo = L + I\left[\frac{\Delta 1}{\Delta 1 + \Delta 2}\right]$$

เมื่อ คือ L ขอบเขตล่างของชั้นที่ฐานนิยมอยู่

$\Delta 1$ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลต่ำกว่า

$\Delta 2$ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลสูงกว่า

I ความกว้างของชั้น

ตัวอย่าง หาฐานนิยมของคะแนนนักศึกษา 150 คน

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 - 29.5	6	10
29.5 – 34.5	$\Delta 1$ 12	22
34.5 – 39.5	16	38
39.5 – 44.5	12 $\Delta 2$	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

ฐานนิยม (Mode : Mo)

<u>วิธีทำ</u> เนื่องจากความถี่สูงสุด = 16 ดังนั้นฐานนิยมอยู่ในชั้น 34.5 – 39.5

จะได้ว่า $L = 34.5$, $l = 5$, $\Delta 1 = 16 - 12 = 4$, $\Delta 2 = 16-12 = 4$

แทนค่าสมการ

$$Mo = 34.5 + 5 \left[\frac{4}{4+4} \right]$$

$$= 37$$

นั่นคือ ฐานนิยมของข้อมูลชุดนี้คือ 37

ตัวอย่าง จงคำนวณหาฐานนิยมจากข้อมูลดังต่อไปนี้

คะแนน	ความถี่
10-19	2
20-29	8
30-39	9
40-49	14
50-59	8
60-69	6
70-79	3
	N=50

$\Delta 1$

$\Delta 2$

ฐานนิยม (Mode : Mo)

วิธีทำ เนื่องจากความถี่สูงสุด = 14 ดังนั้นฐานนิยมอยู่ในชั้น 40-49

จะได้ว่า $L = 39.5, I = 10$

$\Delta 1 = 14 - 9 = 5$

$\Delta 2 = 14-8 = 6$

แทนค่าสมการ $Mo = 39.5 + 10 \left(\frac{5}{5+6} \right)$

$= 44$

นั่นคือ ฐานนิยมของข้อมูลชุดนี้คือ 44

Assignment

ตัวอย่างข้อมูลที่มีการแจกแจงความถี่จงหาค่ามัธยฐานของคะแนนสอบวิชาสถิติ ของนักศึกษา 76 คน

คะแนน	ความถี่	F
40-49	3	76
30-39	9	73
20-29	25	64
10-19	37	39
0-9	2	2

Assignment

ตัวอย่าง จงหาฐานนิยม

คะแนน	ความถี่
5-7	5
8-10	7
11-13	15
14-16	9
17-19	4
	40

Assignment

ตัวอย่าง จงหามัธยฐาน

คะแนน	ความถี่
30-39	3
40-49	5
50-59	12
60-69	7
70-79	2
80-89	1

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Storyset**

บทที่ 3

การวัดตำแหน่งที่ของข้อมูล

ผศ ดร มนันยา มีนคร
แขนงการจัดการธุรกิจบริการ
มหาวิทยาลัยราชภัฏสวนสุนันทา

การเรียนการสอน...............บทที่ 3

1. การหาควอร์ไทล์
2. การหาเดไซล์
3. การหาเปอร์เซ็นต์ไทล์
4. คำถามทบทวน

การวัดตำแหน่งที่ของข้อมูล

ควอร์ไทล์ (quartile : Q) เป็นการเปรียบเทียบตำแหน่งของข้อมูลใดข้อมูลหนึ่งกับข้อมูลทั้งหมด โดยถือว่ามีข้อมูลทั้งหมด 4 ส่วน หรือตำแหน่ง ที่บอกให้ทราบว่ามีข้อมูลกี่ส่วนจาก 4 ส่วนที่มีค่าน้อยกว่าหรือเท่ากับข้อมูล ณ ตำแหน่งควอร์ไทล์นั้น ซึ่งอาจเป็นควอร์ไทล์ที่ 1 (Q_1) ควอร์ไทล์ที่ 2 (Q_2)และควอร์ไทล์ที่ 3 (Q_3) ตัวอย่างเช่น A สอบได้คะแนนวิชาภาษาอังกฤษ 55คะแนน ซึ่งตรงกับตำแหน่งควอร์ไทล์ที่ 3 เขียนแทนได้ว่า $Q_3 = 55$ หมายความว่า มีนักเรียน 3 ส่วน จาก 4 ส่วน ที่ได้คะแนนน้อยกว่าหรือเท่ากับ 55 คะแนน

การวัดตำแหน่งที่ของข้อมูล

เดไซล์ (decile : D) เป็นการเปรียบเทียบตำแหน่งของข้อมูลใด ข้อมูลหนึ่งกับข้อมูลทั้งหมด โดยถือว่ามีข้อมูลทั้งหมด 10 ส่วน หรือตำแหน่ง ที่บอกให้ทราบว่ามีข้อมูลกี่ส่วนจาก 10 ส่วนที่มีค่าน้อยกว่าหรือเท่ากับข้อมูล ณตำแหน่ง เดไซล์นั้น ซึ่งอาจเป็นเดไซล์ที่ 1 (D_1) เดไซล์ที่ 2 (D_2) เดไซล์ที่ 3 (D_3) ... เดไซล์ที่ 9 (D_9) ตัวอย่างเช่น B สอบได้คะแนนวิชาคณิตศาสตร์พื้นฐาน 70 คะแนน ซึ่งตรงกับตำแหน่งเดไซล์ที่ 5 เขียนแทนได้ว่า D_5 = 70 หมายความว่า มีนักเรียน 5 ส่วน จาก 10 ส่วน ที่ได้คะแนนน้อยกว่าหรือ เท่ากับ 70

การวัดตำแหน่งที่ของข้อมูล

เปอร์เซ็นไทล์ (P) หมายถึง ตำแหน่งที่แสดงให้ทราบว่ามีจำนวนร้อยละเท่าไรของจำนวนคะแนนที่มีค่าต่ำกว่าคะแนน ณ ตำแหน่งนั้น โดยการแบ่งคะแนนทั้งหมดออกเป็น 100 ส่วน เพื่อใช้ในการเปรียบเทียบข้อมูล สามารถแบ่งออกเป็น 2 ส่วนได้แก่ คะแนนที่ไม่ได้จัดหมวดหมู่ และคะแนนที่จัดหมวดหมู่

การหาค่าคะแนนที่ไม่ได้จัดหมวดหมู่

เช่น C สอบได้คะแนนวิชาสถิติ 80 คะแนน ซึ่งตรงกับตำแหน่ง เปอร์เซ็นไทล์ที่ 30 เขียนแทนได้ว่า $P30 = 80$ หมายความว่า มีนักเรียน 30 ส่วน จาก 100 ส่วน ที่ได้คะแนนน้อยกว่าหรือเท่ากับ 80

การวัดตำแหน่งที่ของข้อมูล

การหาตำแหน่งควอร์ไทล์ เดไซล์และเปอร์เซ็นไทล์สามารถทำได้ 2 กรณี คือ

1. ข้อมูลไม่ได้แจกแจงความถี่
2. ข้อมูลแจกแจงความถี่

1. การหาค่าของตำแหน่งควอร์ไทล์ เดไซล์และเปอร์เซ็นไทล์

★ กรณีข้อมูลไม่แจกแจงความถี่ มีขั้นตอน ดังนี้

1) เรียงลำดับข้อมูลจากน้อยไปมาก

2) หาตำแหน่งควอร์ไทล์ เดไซล์และเปอร์เซ็นไทล์จาก

ตำแหน่งควอร์ไทล์ที่ r	ตำแหน่งของ $Q_r = \frac{r(N+1)}{4}$
ตำแหน่งเดไซล์ที่ r	ตำแหน่งของ $D_r = \frac{r(N+1)}{10}$
ตำแหน่งเปอร์เซ็นไทล์ที่ r	ตำแหน่งของ $P_r = \frac{r(N+1)}{100}$

3) หาควอร์ไทล์ เดไซล์และเปอร์เซ็นไทล์โดยนับตำแหน่งที่ของ ข้อมูลที่เรียงลำดับไว้ได้ หากตำแหน่งที่ได้ไม่เป็นจำนวนเต็มให้เทียบ บัญญัติไตรยางศ์ระหว่างความแตกต่างของตำแหน่งที่ของข้อมูลและคะแนนที่แตกต่างกัน

ตัวอย่าง คะแนนการสอบของนักเรียน 9 คน เป็นดังนี้

34 , 8 , 6 , 22 , 38 , 2 , 40 , 18 , 30 จงหา P30 และ D5

วิธีทำ ขั้นที่ 1 เรียงข้อมูลจากน้อยไปหามาก

ตำแหน่งของข้อมูล	1	2	3	4	5	6	7	8	9
ข้อมูล	2	6	8	18	22	30	34	38	40

ขั้นที่ 2 หาตำแหน่งของ P30 และ D5

ตำแหน่งของ $P_r = \frac{r(N+1)}{100}$

$$P_{30} = \frac{30\,(9+1)}{100}$$

$$= 3$$

และ

ตำแหน่งของ $D_r = \frac{r(N+1)}{10}$

$$D_5 = \frac{5(9+1)}{10}$$

$$= 5$$

ขั้นที่ 3 หาค่าของ P30 และ D5 จะได้

$$P_{30} = 8$$

$$D_5 = 22$$

ตำแหน่งของข้อมูล	1	2	3	4	5	6	7	8	9
ข้อมูล	2	6	8	18	22	30	34	38	40

ตัวอย่าง ถ้าข้อมูลเป็นดังนี้

52 , 42 , 57 , 53 , 44 , 39 , 33 , 35

จงหา 1. Q_1 2. D_7 3. P_{50}

วิธีทำ ขั้นที่ 1 เรียงข้อมูลจากน้อยไปหามาก (N = 8) ได้ดังนี้

ตำแหน่งของข้อมูล	1	2	3	4	5	6	7	8
ข้อมูล	33	35	39	42	44	52	53	57

ขั้นที่ 2 หาตำแหน่งของ 1. Q_1 2. D_7 3. P_{50}

ตำแหน่งของ $Q_r = \frac{r(N+1)}{4}$

ตำแหน่งของ $D_r = \frac{r(N+1)}{10}$

$$Q_1 = \frac{1\,(8+1)}{4} = 2.25$$

$$D_7 = \frac{7\,(8+1)}{10} = 6.3$$

ตำแหน่งของ $P_r = \frac{r(N+1)}{100}$

$$P_{50} = \frac{50\,(8+1)}{100}$$

$$= 4.5$$

ตำแหน่งของข้อมูล	1	2	3	4	5	6	7	8
ข้อมูล	33	35	39	42	44	52	53	57

39-35 = 4

2.25-2 (ตำแหน่ง2) = 0.25

ขั้นที่ 3 หาค่าของ Q_1 = 2.25 มีค่าอยู่ระหว่าง 35 กับ 39

ตำแหน่งต่างกัน 1 ข้อมูลต่างกัน 4

ตำแหน่งต่างกัน 0.25 ข้อมูลต่างกันจะได้ $\frac{4 \times 0.25}{1} = 1$

ดังนั้น Q1 = 35 + 1 = 36

มาจากตำแหน่งที่ 2

ตำแหน่งของข้อมูล	1	2	3	4	5	6	7	8
ข้อมูล	33	35	39	42	44	52	53	57

6.3-6 (ตำแหน่ง6) = 0.3

ค่าของ D7= 6.3 มีค่าอยู่ระหว่าง 52 กับ 53

ตำแหน่งต่างกัน 1 ข้อมูลต่างกัน 1

ตำแหน่งต่างกัน 0.3 ข้อมูลต่างกัน $\frac{1 \times 0.3}{1} = 0.3$

ดังนั้น D_7 = 52 + 0.3 = 52.3

ค่าของ P50 มีค่าอยู่ระหว่าง 42 กับ 44

ตำแหน่งต่างกัน 1 ข้อมูลต่างกัน 2

ตำแหน่งต่างกัน 0.5 ข้อมูลต่างกัน $\frac{2 \times 0.5}{1} = 1$

ดังนั้น $P_{50} = 42 + 1 = 43$

2. การหาค่าของตำแหน่งควอร์ไทล์ เดไซล์และเปอร์เซ็นไทล์

☆ กรณีข้อมูลแจกแจงความถี่ มีขั้นตอน ดังนี้

1) สร้างตารางแจกแจงความถี่สะสม (F)

2) หาตำแหน่งควอร์ไทล์ เดไซล์และเปอร์เซ็นไทล์จาก

ตำแหน่งควอร์ไทล์ที่ r $Q_r = \frac{rN}{4}$

ตำแหน่งเดไซล์ที่ r $D_r = \frac{rN}{10}$

ตำแหน่งเปอร์เซ็นไทล์ที่ r $P_r = \frac{rN}{100}$

3) หาค่าของควอไทล์ เดไซล์และเปอร์เซ็นไทล์ที่ r จากสูตร

3. หาคำตอบโดยใช้สูตร

$$Q_k = L + I\left(\frac{\frac{kN}{4} - \sum f_L}{f_{me}}\right), \quad D_k = L + I\left(\frac{\frac{kN}{10} - \sum f_L}{f_{me}}\right) \quad \text{และ} \quad P_k = L + I\left(\frac{\frac{kN}{100} - \sum f_L}{f_{me}}\right)$$

เมื่อ L คือ ขอบล่างของชั้นควอไทล์ เดไซล์ เปอร์เซนต์ไทล์

I คือ ความกว้างของอันตรภาคชั้น

$\sum f_L$ คือ ความถี่สะสมของชั้นที่ต่ำกว่าชั้นควอไทล์ เดไซล์ เปอร์เซนต์ไ

f_{me} คือ ความถี่ของชั้นชั้นควอไทล์ เดไซล์ เปอร์เซนต์ไทล์

ตัวอย่าง จงหา D4 จากตารางแจกแจงความถี่ ต่อไปนี้

อันตรภาคชั้น	11 – 15	16 – 20	21 – 25	26 ~~36~~ – 30	31 – 35	36 – 40	41 – 45
ความถี่	5	11	15	18	12	10	9

วิธีทำ ขั้นที่ 1 สร้างตารางความถี่สะสม

คะแนน	**ความถี่ (f)**	**ความถี่สะสม ~~(f)~~ F**
11 – 15	5	5
16 – 20	11	16
21 – 25	15	31
26 – 30	18	49
31 – 35	12	61
36 – 40	10	71
41 – 45	9	80
รวม	**N = 80**	

ขั้นที่ 2 หาตำแหน่งของ D_4 โดยใช้สูตร

ตำแหน่งของ $D_4 = \frac{4}{10}(80)$

$= 32$

คะแนน	ความถี่ (f)	ความถี่สะสม (f)
11 – 15	5	5
16 – 20	11	16
21 – 25	15	31
36 – 30	18	49
31 – 35	12	61
36 – 40	10	71
41 – 45	9	80
รวม	**N = 80**	

ขั้นที่ 3 หาอันตรภาคชั้น D_4 อยู่ ซึ่งตรงกับอันตรภาคชั้น 26 - 30

$$D_4 = L + I \left[\frac{\frac{rN}{10} - \sum F}{f} \right]$$

ในที่นี้ I = 5 , f = 18 , F = 31 และ L = 25.5

$$D_4 = 25.5 + 5\left[\frac{80 - 31}{18}\right]$$

$$= 25.78$$

ดังนั้น D_4 มีค่าเท่ากับ 25.78

คะแนน	ความถี่ (f)	ความถี่สะสม (f)
11 – 15	5	5
16 – 20	11	16
21 – 25	15	31 ← $\sum F$
26 – 30	18	49
31 – 35	12	61
36 – 40	10	71
41 – 45	9	80
รวม	N = 80	

เปอร์เซ็นต์ไทล์

1. **การหาค่าคะแนนที่จัดหมวดหมู่**

สูตรในการหาค่า

$$P = L_1 + i \frac{(F_n - F_1)}{(F_2 - F_1)}$$

กำหนดให้

P = คะแนน ณ ตำแหน่งเปอร์เซ็นต์ไทล์ที่ต้องการ

L_1 = ขีดจำกัดล่างชั้นที่มี F_n อยู่

i = อันตรภาคชั้น

F_n = ตำแหน่งที่ต้องการ = PxN/100

F_1 = ความถี่สะสมชั้นที่อยู่ถัดจากชั้น F_n ไปทางคะแนนน้อย

F_2 = ความถี่สะสมชั้น F_n

เปอร์เซ็นต์ไทล์

1. **ตัวอย่าง** จงคำนวณหาคะแนนในตำแหน่ง P_{25} คะแนนของนักเรียน 50 คนดังนี้

คะแนน	ความถี่
10-19	2
20-29	8
30-39	9
40-49	14
50-59	8
60-69	6
70-79	3
	N = 50

เปอร์เซ็นต์ไทล์

วิธีทำ สร้างตารางแจกแจงความถี่สะสมจากคะแนนน้อยไปหาคะแนนมากดังนี้

คะแนน	f (ความถี่)	F (ความถี่สะสม)
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N = 50	

เปอร์เซ็นต์ไทล์

วิธีทำ คำนวณหาตำแหน่งที่ต้องการ (F_n)

P_{25} ของคะแนนชุดนี้ตรงกับตำแหน่ง PxN/100 = 25x50/100 = 12.5

หาชั้นของคะแนนในตารางความถี่สะสมที่มีตำแหน่งที่ต้องการ (F_n = 12.5) อยู่ในชั้นที่มีคะแนนระหว่าง 30-39 ดังนั้น

L_1 = 29.5

i = 19.5-9.5 = 10

F_1 = 10

F_2 = 19

เปอร์เซ็นต์ไทล์

วิธีทำ คำนวณหาคะแนนในตำแหน่งเปอร์เซ็นต์ไทล์

$$P = L_1 + i \frac{(F_n - F_1)}{(F_2 - F_1)}$$

$$P_{25} = 29.5 + 10 \frac{(12.5 - 10)}{19 - 10}$$

$$= 29.5 + 10 \frac{(2.5)}{9}$$

$$= \frac{29.5 + 25}{9}$$

$$= 29.5 + 2.78 = 32.28 \#$$

ดังนั้น คะแนนในตำแหน่ง P_{25} ของคะแนนชุดนี้คือ 32.28 **(หมายความว่ามีนักเรียนร้อยละ 25 ของนักเรียน 50 คนที่ได้คะแนนต่ำกว่า 32.28 คะแนน)**

ควอร์ไทล์

- การคำนวณหาควอร์ไทล์ทำได้เช่นเดียวกับการคำนวณหาเปอร์เซ็นต์ไทล์ต่างกันที่การคำนวณเดไซล์ต้องแบ่งหมู่ออกเป็น 10 ส่วน แต่ควอร์ไทล์แบ่งหมู่ออกเป็น 4 ส่วน และเปอร์เซ็นต์ไทล์แบ่งหมู่ออกเป็น 100 ส่วน
- จากตารางคะแนนนักเรียน 50 คน จงคำนวณหา D_9 และ Q_3

คะแนน	f (ความถี่)	F (ความถี่สะสม)
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N = 50	

เดไซล์

วิธีทำ การหาค่าคะแนนในตำแหน่ง D_9

D_9 ของคะแนนชุดนี้ตรงกับตำแหน่ง DxN/10 = 9x50/10 = 45

ตำแหน่งที่ 45 อยู่ในชั้นระหว่าง 60-69 ดังนั้น

$F_n = 45$; $L_1 = 59.5$; $i = 10$

$F_1 = 41$; $F_2 = 47$

คะแนน	f (ความถี่)	F (ความถี่สะสม)
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N = 50	

เดไซล์

$$P = L_1 + i \frac{(F_n - F_1)}{(F_2 - F_1)}$$

- วิธีทำ
  - $D_9 = 59.5 + 10 \frac{(45 - 41)}{(47 - 41)}$

    $= 59.5 + 10\,(4/6)$

    $= 59.5 + 6.7$

    $= 66.2$

คะแนนในตำแหน่ง D_9 ของข้อมูลคือ 66.2

ควอร์ไทล์

วิธีทำ การหาค่าคะแนนในตำแหน่ง Q_3

Q_3 ของคะแนนชุดนี้ตรงกับตำแหน่ง $QxN/4 = 3x50/4 = 37.5$

ตำแหน่งที่ 37.5 อยู่ในชั้นระหว่าง 50-59 ดังนั้น

$F_n = 37.5$; $L_1 = 49.5$; $i = 10$

$F_1 = 33$; $F_2 = 41$

ควอร์ไทล์

วิธีทำ

$$Q_3 = 49.5 + 10\,\frac{(37.5 - 33)}{(41 - 33)}$$

$$= 49.5 + 10\,(4.5/8)$$

$$= 49.5 + 5.625$$

$$= 55.125\ \#$$

คะแนนในตำแหน่ง Q_3 ของข้อมูลคือ 55.125

Assignment

1. จากคะแนน 24, 25, 29, 30, 35, 37, 40, 42, 46 และ 49 จงหา P_{75} , Q_{25} , $D_{4.2}$
2. จงหา P_{30} จากคะแนนดังตารางนี้
3. จงหา Q_2 , D_4 , P_{25} จากคะแนนดังตารางนี้
4. จงหาคะแนน 42 ตรงกับตำแหน่ง P, D และ Q เท่าใด จากคะแนนดังตารางนี้

คะแนน	f
20-24	2
25-29	3
30-34	5
35-39	9
40-44	10
45-49	4
50-54	4
55-59	3
	N = 40

Thanks

บทที่ 4 การวัดการกระจาย (ต่อ)

ผศ ดร มนันยา มีนคร
แขนงการจัดการธุรกิจบริการ
มหาวิทยาลัยราชภัฏสวนสุนันทา

การวัดการกระจายสัมพัทธ์ (relative Variation)

- การวัดการกระจายสัมพัทธ์ คือ การหาค่าเพื่อเปรียบเทียบการกระจายระหว่างข้อมูลมากกว่าหนึ่งชุด โดยใช้อัตราส่วน การเปรียบเทียบการกระจายของข้อมูลระหว่างชุดที่นิยมใช้มี 2 ชนิดคือ

- 1. สัมประสิทธิ์ของพิสัย(coefficient of range)
- 2.สัมประสิทธิ์ของการแปรผัน(coefficient of variation)

1. สัมประสิทธิ์ของพิสัย(coefficient of range) คือ อัตราส่วนระหว่างผลต่างของค่าสูงสุดและค่าต่ำสุด กับผลบวกของค่าสูงสุดและต่ำสุดของข้อมูลชุดนั้น

- หาได้จากสูตร C.R = $\dfrac{x_{max} - x_{min}}{x_{max} + x_{min}}$

2. สัมประสิทธิ์ของการแปรผัน(coefficient of variation) ตัวย่อ(C.V.) อัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น

หาได้จากสูตร

$$C.V = \frac{S.D}{\bar{X}} \times 100$$

ตัวอย่าง จงเปรียบเทียบการกระจายของราคาข้าวเปลือกและราคาข้าวสาร จากข้อมูลดังต่อไปนี้

ราคาข้าวเปลือก (บาท/กิโลกรัม)	**22**	**25**	**23**	**24**	**26**	**21**	**20**
ราคาข้าวสาร (บาท/กิโลกรัม)	65	68	62	64	67	60	63

วิธีทำ หาสัมประสิทธิ์ของพิสัย

ราคาข้าวเปลือก (บาท/กิโลกรัม)

จาก C.R = $\dfrac{x_{max} - x_{min}}{x_{max} + x_{min}}$

จะได้ C.R = $\dfrac{26-20}{26+20}$

$= \dfrac{6}{46}$

$= 0.13$

ราคาข้าวสาร (บาท/กิโลกรัม)

จาก C.R = $\dfrac{x_{max} - x_{min}}{x_{max} + x_{min}}$

C.R = $\dfrac{68-60}{68+60}$

$= \dfrac{8}{128}$

$= 0.062$

ดังนั้น สัมประสิทธิ์ของพิสัยข้าวเปลือก 0.13 และ สัมประสิทธิ์ของพิสัยข้าวสาร 0.062 (บาท/กิโลกรัม)

ตัวอย่าง จงหาสัมประสิทธิ์ของการแปรผันของการสอบ นักศึกษาทั้ง2ครั้ง

	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตราฐาน
สอบครั้งที่ 1	60	6
สอบครั้งที่2	700	7

วิธีทำ หาสัมประสิทธิ์ของการแปรผัน $= \frac{S}{\bar{X}}$

สัมประสิทธิ์ของการแปรผันของการสอบครั้งที่ 1 $= \frac{6}{60}$

$= \frac{1}{10}$

$= 0.1$

ดังนั้น สัมประสิทธิ์ของการแปรผันของการสอบครั้งที่ 1 เท่ากับ 10%

สัมประสิทธิ์ของการแปรผันของการสอบครั้งที่ 2 $= \frac{7}{700}$

$= \frac{1}{100}$

$= 0.01$

ดังนั้น สัมประสิทธิ์ของการแปรผันของการสอบครั้งที่ 2 เท่ากับ 1%

จะเห็นได้ว่าสัมประสิทธิ์ของการแปรผันของการสอบครั้งที่1 สูงกว่าครั้งที่ 2 แสดงว่าการสอบครั้งที่1 มีการกระจายคะแนนสอบมากกว่าครั้งที่2 นั่นคือ คะแนนการสอบครั้งที่2 ดีกว่าครั้งที่1

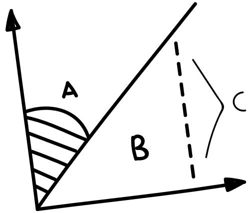
assignment

พิจารณาข้อมูลต่อไปนี้ แสดงวิธีทำโดยกำหนดตารางแจกแจงความถี่

คะแนน	21-23	24-26	27-29	30-32	33-35	36-38
ความถี่	2	8	10	17	5	2

1. พิสัย มีค่าเท่ากับเท่าใด 17
2. ส่วนเบี่ยงเบนควอร์ไทล์มีค่าเท่ากับเท่าใด 3.3
3. ส่วนเบี่ยงเบนเฉลี่ย มีค่าเท่ากับเท่าใด 0.94
4. ส่วนเบี่ยงเบนมำตรฐำน มีค่าเท่ากับเท่าใด 6.53

04



$(-3\sqrt{2}) - 4(3)(-3M+2)$

THANK you

$\frac{3 \sin 4/8}{\sqrt{3.2.4+2}}$

การวิเคราะห์สหสัมพันธ์และการถดถอย

ผศ ดร มนันยา มีนคร
คณะวิทยาการจัดการ

การวิเคราะห์สหสัมพันธ์

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัวโดยทั่วไปจะศึกษาใน 3 ลักษณะ คือ

1. มีความสัมพันธ์กันหรือไม่
2. มีความสัมพันธ์กันมากน้อยเพียงใด
3. มีความสัมพันธ์กันในทิศทางใด

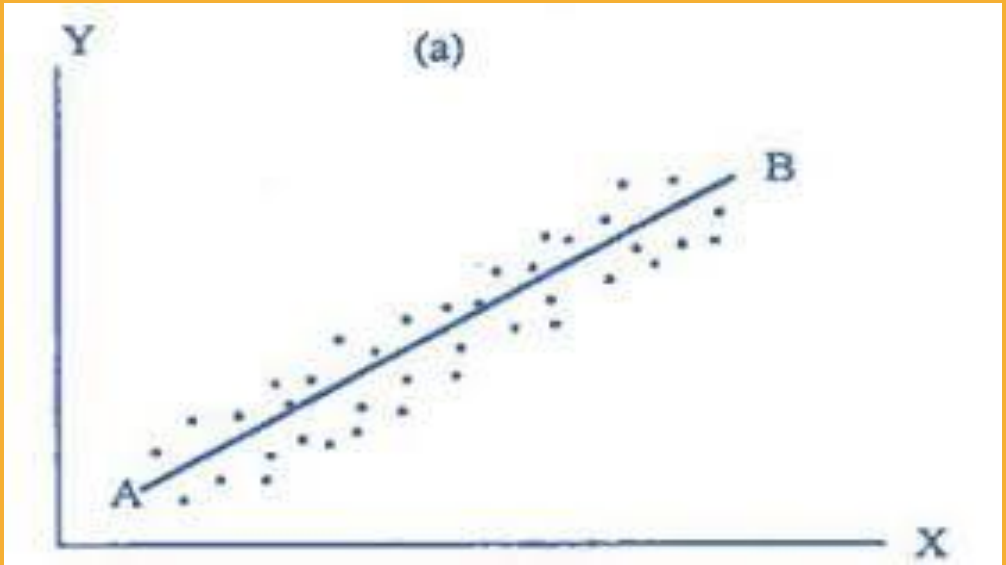
เช่น เช่นความสูงกับน้ำหนักของคน มีความสัมพันธ์กันมากหรือน้อยและมีความสัมพันธ์ในทิศทางเดียวกันหรือตรงกันข้าม

วิธีการตรวจสอบลักษณะความสัมพันธ์

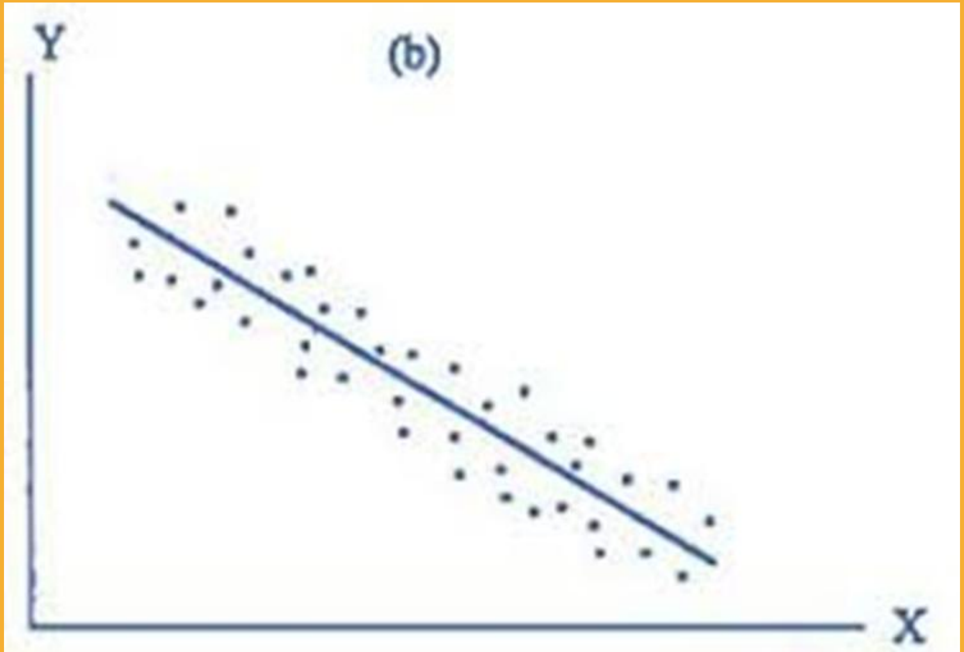
การตรวจสอบลักษณะความสัมพันธ์ของตัวแปรสามารถทำได้หลายวิธีดังนี้

1. แผนภาพการกระจาย (scatter diagram)
2. ค่าสัมประสิทธิ์สหสัมพันธ์(correlation coefficient)

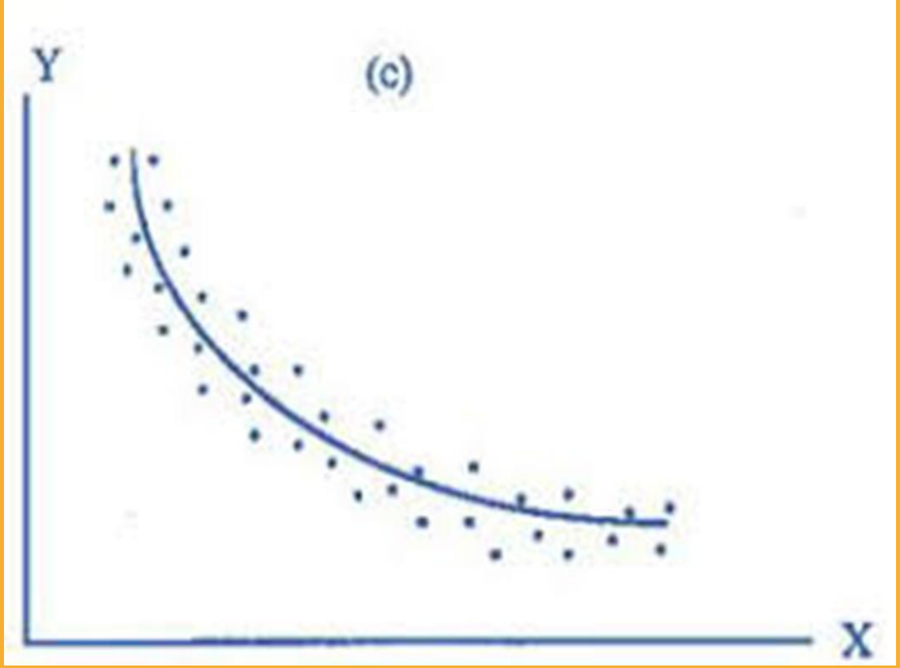
แผนภาพการกระจาย



ในรู ป (a) แนวโน้มของจุดชี้ขึ้นด้านขวาตามแนวเส้นตรง เมื่อ x มีค่ามากy มีค่ามากเมื่อx มีค่าน้อย y มีค่าน้อย เรียกว่ามีความสัมพันธ์เชิงเส้นเชิงบวก



ในรูป (b) แนวโน้มของจุดชี้ลงด้านขวาตามแนวเส้นตรง เมื่อ x มีค่ามากy มีค่าน้อย เมื่อx มีค่าน้อย y มีค่ามาก เรียกว่ามีความสัมพนั ธ์เชิงเส้นเชิงลบ



ในรูป (c) แนวโน้มของจุดชี้ลงด้านขวาตามแนวเส้นโค้ง เรียกว่ามีความสัมพันธ์ไม่เชิงเส้นเชิงลบ

ค่าสัมประสิทธิ์สหสัมพันธ์

สำหรับตัวสถิติที่ใช้วัดค่าสหสัมพันธ์อย่างง่ายว่ามีความสัมพันมากหรือน้อยเพียงใดคือสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ซึ่งในกรณีของสหสัมพันธ์อย่างง่ายตัว

สถิตินี้เรียกว่าสัมประสิทธิ์สหสัมพันธ์ อย่างง่าย (Simple Correlation Coefficient) เขียนแทนด้วย

สัญลักษณ์ ρ หรือ r_{xy} ในกรณีที่เป็นค่าพารามิเตอร์และ r หรือ r_{xy} ในกรณีที่เป็นค่าสถิติโดยที่ ρ และ r จะไม่มีหน่วยและมีค่าตั้งแต่-1 ถึง 1

สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันโพรดักโมเมนต์

เป็นการหาความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร คือตัวแปร X และตัวแปร Y ที่ไม่ได้สนใจว่าตัวใดเป็นเหตุและตัวใดเป็นผลซึ่งเขียนแทนด้วย r_{xy}

สูตรที่ 1 คำนวณโดยใช้คะแนนมาตรฐาน

$$r_{xy} = \frac{\sum Z_X Z_y}{n}$$

เมื่อ r_{xy} **แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y**

Z_X **แทน คะแนนมาตรฐานแต่ละค่าของตัวแปร X**

Z_y **แทน คะแนนมาตรฐานแต่ละค่าของตัวแปร Y**

n **แทน จำนวนคู่ของข้อมูล**

สูตรที่ 2 คำนวณโดยใช้คะแนนที่เป็นค่าเบี่ยงเบนจากค่าเฉลี่ย

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left(n\sum X^2 - (\sum X)^2\right)\left(n\sum Y^2 - (\sum Y)^2\right)}}$$

เมื่อ r_{xy} แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y

n แทน จำนวนคู่ของข้อมูล

$\sum XY$ แทน ผลรวมทั้งหมดของผลคูณของตัวแปร XกับตัวแปรY

$\sum X$ แทน ผลรวมของตัวแปร X

$\sum Y$ แทน ผลรวมของตัวแปร Y

$\sum X^2$ แทน ผลรวมของตัวแปร X แต่ละค่ายกกำลังสอง

$\sum Y^2$ แทน ผลรวมของตัวแปร Y แต่ละค่ายกกำลังสอง

สูตรที่2 เป็นสูตรที่นิยมใช้มากที่สุด เพราะข้อมูลที่เก็บรวบรวมมาได้ส่วนใหญ่อยู่ในรูปคะแนนดิบ

ตัวอย่าง จงหาความสัมพันธ์ระหว่างคะแนนสอบวิชาคณิตศาสตร์กับคะแนนวิชาวิทยาศาสตร์ของ นศ จำนวน 6 คนดังตาราง

คนที่	คะแนนวิชาคณิตศาสตร์ (X)	คะแนนวิชาวิทยาศาสตร์ (Y)
1	9	5
2	10	9
3	8	5
4	7	8
5	4	6
6	8	4

วิธีทำ

คนที่	X	Y	X^2	Y^2	XY
1	9	5	81	25	45
2	10	9	100	81	90
3	8	5	64	25	40
4	7	8	49	64	56
5	4	6	16	36	24
6	8	4	64	16	32
รวม	46	37	374	247	287

จากตารางจะได้ $n = 6$, $\sum X = 46$, $\sum Y = 37$, $\sum X^2 = 374$, $\sum Y^2 = 247$, $\sum XY = 287$

แทนค่าในสูตร

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left(n\sum X^2 - (\sum X)^2\right)\left(n\sum Y^2 - (\sum Y)^2\right)}}$$

$$= \frac{6(287) - (46)(37)}{\sqrt{\left(6(374) - (46)^2\right)\left(6(247) - (37)^2\right)}}$$

$$= \frac{20}{\sqrt{14464}}$$

$$= 0.17$$

สัมประสิทธิ์สหสัมพันธ์แบบอันดับที่ของสเปียร์แมนหรือเรียกอีกอย่างหนึ่งว่า rank correlation เป็นวิธีการทางสถิติในการศึกษาความสัมพันธ์ของ 2 ตัวแปร จากกลุ่มตัวอย่างเดียวกัน โดยข้อมูลที่ได้จากตัวแปรทั้งสองอยู่ในมาตราเรียงอันดับ(ordinal scale) สูตรที่ใช้คือ

$$\rho = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

ρ แทน สัมประสิทธิ์สหสัมพันธ์

D แทน ผลต่างอันดับที่ของข้อมูลแต่ละคู่

n แทน จำนวนคู่ของข้อมูล

ตัวอย่าง ผลการเรียงอันดับความตั้งใจเรียนของ นิสิต ปริญญาเอก จำนวน 6 คน โดยอาจารย์ที่ปรึกษา และคณบดีได้ผลการจัดอันดับดังนี้

นิสิตคนที่	ผลการจัดอันดับที่	
	ที่ปรึกษา(X)	คณบดี(Y)
1	6	5
2	1	2
3	4	3
4	3	4
5	5	1
6	2	6

วิธีทำ

นิสิตคนที่	X	Y	D=X-Y	$D^2=(X-Y)^2$
1	6	5	1	1
2	1	2	-1	1
3	4	3	1	1
4	3	4	-1	1
5	5	1	4	16
6	2	6	-4	16
				36

นิสิตจำนวน 6 คน ดังนั้น n = 6

แทนค่าในสูตร

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

$$= 1 - \frac{6(36)}{6(6^2-1)}$$

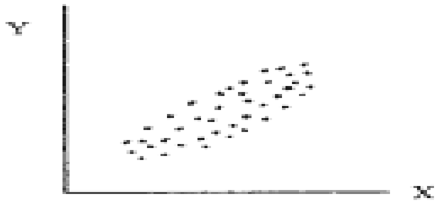
$$= 1 - \frac{216}{210}$$

$$= 1 - 1.03$$

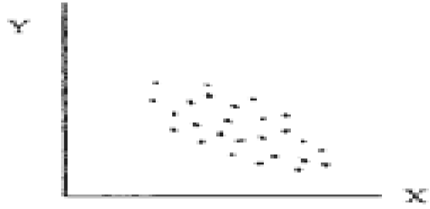
$$= -0.03$$

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

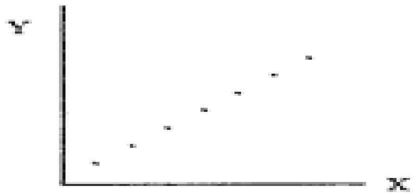
หลักการของการวิเคราะห์การถดถอยคือ การใช้ข้อมูลเมื่อกำหนดค่าตัวแปรอิสระ(X) และทำให้เกิดค่าตัวแปรตาม (Y) ในอดีตมีการนำมาสร้างสมการเชิงเส้นที่เหมาะสมที่สุด เพื่อพยากรณ์ค่า Y ในอนาคต



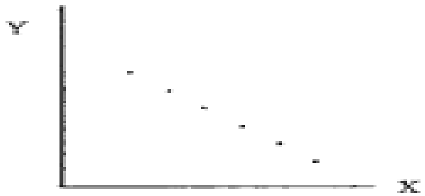
มีความสัมพันธ์ทางบวก
(r เป็นบวก)



มีความสัมพันธ์ทางลบ
(r เป็นลบ)



มีความสัมพันธ์ทางบวกอย่างสมบูรณ์
(r = +1.00)

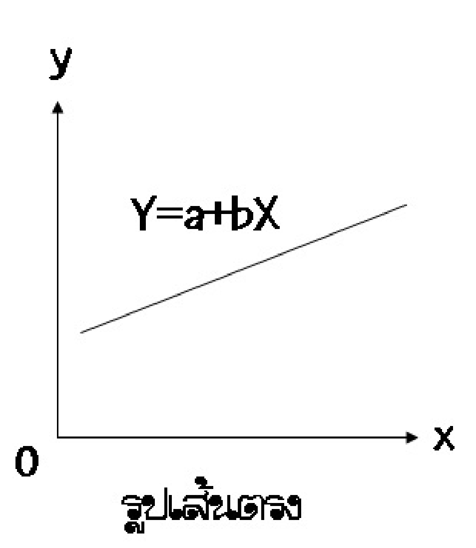


มีความสัมพันธ์ทางลบอย่างสมบูรณ์
(r = -1.00)

ภาพของการกระจาย

สมการถดถอยเชิงเส้นอย่างง่าย

สมการเสนตรงทางคณิตศาสตร์เพื่อแสดงความสัมพันธ์ระหว่าง X และ Y



b เป็นความชันของเส้นตรงหรือค่าของ yที่เปลี่ยนไป เมื่อ x เปลี่ยนไปหนึ่งหน่วย a เป็นระยะตัดแกน y และเป็นค่าคงตัวที่ต้องการหา

สมการถดถอยเชิงเส้นอย่างง่าย

สูตรสมการถดถอยเชิงเส้นอย่างง่าย

$$b = \frac{n\sum xy - \Sigma x \Sigma y}{n\sum x^2 - (\Sigma x)^2} \longrightarrow 1$$

$$a = \bar{y}_t - b\bar{x}_t \longrightarrow 2$$

สมการถดถอยเชิงเส้นอย่างง่าย

สูตรสมการถดถอยเชิงเส้นอย่างง่าย

$$\hat{Y}_t = a + bx_t \longrightarrow 3$$

$\hat{Y}_t$ คือ ค่าแนวโน้มที่แปรตามตัวแปรเวลาที่เราต้องการหา

a คือ ระยะทางจากจุดกำเนิดถึงจุดตัดบนแกน y

b คือ ความชันของสมการเส้นตรง

x_t คือ ตัวแปรเวลา ณ เวลาที่ t

ตัวอย่าง จากข้อมูลในตารางเป็นคะแนนสอบกลางภาควิชาสถิติและวิชาอังกฤษของนักศึกษาแผนกวิชาคอมพิวเตอร์วิทยาลัยแห่งหนึ่งจำนวน 10 คน

คะแนนสอบวิชาสถิติ	75	55	71	72	80	95	95	98	65	82
คะแนนสอบวิชาอังกฤษ	80	65	78	35	48	85	98	98	66	83

1. จงสร้างสมการการถดถอยที่แสดงความสัมพันธ์ระหว่างคะแนนสอบวิชาสถิติกับวิชาภาษาอังกฤษ
2. ถ้านักศึกษาคนหนึ่งสอบวิชาสถิติได้ 70 คะแนน จะสอบวิชาภาษาอังกฤษได้ประมาณกี่คะแนน

คะแนนสอบวิชาสถิติ (X)	คะแนนสอบวิชาภาษาอังกฤษ(Y)	X^2	Y^2	XY
75	80	5,625	6,400	6,000
55	65	3,025	4,225	3,575
71	78	5,041	6,084	5,538
72	35	5,184	1,225	2,520
80	48	6,400	2,304	3,840
95	85	9,025	7,225	8,075
95	98	9,025	9,604	9,310
98	98	9,604	9,604	9,604
65	66	4,225	4,356	4,290
82	83	6,724	6,889	6,806
$\sum$=788	$\sum$=736	$\sum$=63,878	$\sum$=5,7916	$\sum$=59,558

$$b = \frac{n\sum xy - \Sigma x \Sigma y}{n\sum x^2 - (\Sigma x)^2} \longrightarrow 1$$

$$b = \frac{10(59{,}558) - 788(736)}{10(63{,}878) - (788)^2}$$

$$b = \frac{15{,}612}{17{,}836}$$

$$b = 0.8753$$

จากนั้นหา =

$\bar{x} = \dfrac{\Sigma X}{n}$ และ $\bar{y} = \dfrac{\Sigma Y}{n}$

$\bar{x} = \dfrac{788}{10}$ $(y)^{-} = \dfrac{736}{10}$

$\bar{x} = 78.8$ และ $\bar{y} = 73.6$

$$a = \bar{y}_t - b\bar{x}_t \longrightarrow 2$$

$$a = 73.6 - 0.8753\ (78.8)$$

$$a = 73.6 - 68.97364$$

1. สมการการถดถอยที่แสดงความสัมพันธ์ระหว่างคะแนนสอบวิชาสถิติกับวิชาภาษาอังกฤษ คือ

$$a = 4.6264$$

$$\hat{Y}_t = a + bx_t \longrightarrow 3$$

$$\hat{y} = 4.6264 + 0.8753\ X$$

2. นักศึกษาคนหนึ่งสอบวิชาสถิติได้ 70 คะแนน X = 70

$$\hat{y} = 4.6264 + 0.8753\ X$$

$$\hat{y} = 4.6264 + 0.8753(70)$$

$$= 4.6264 + 61.261$$

$$= 65.8974$$

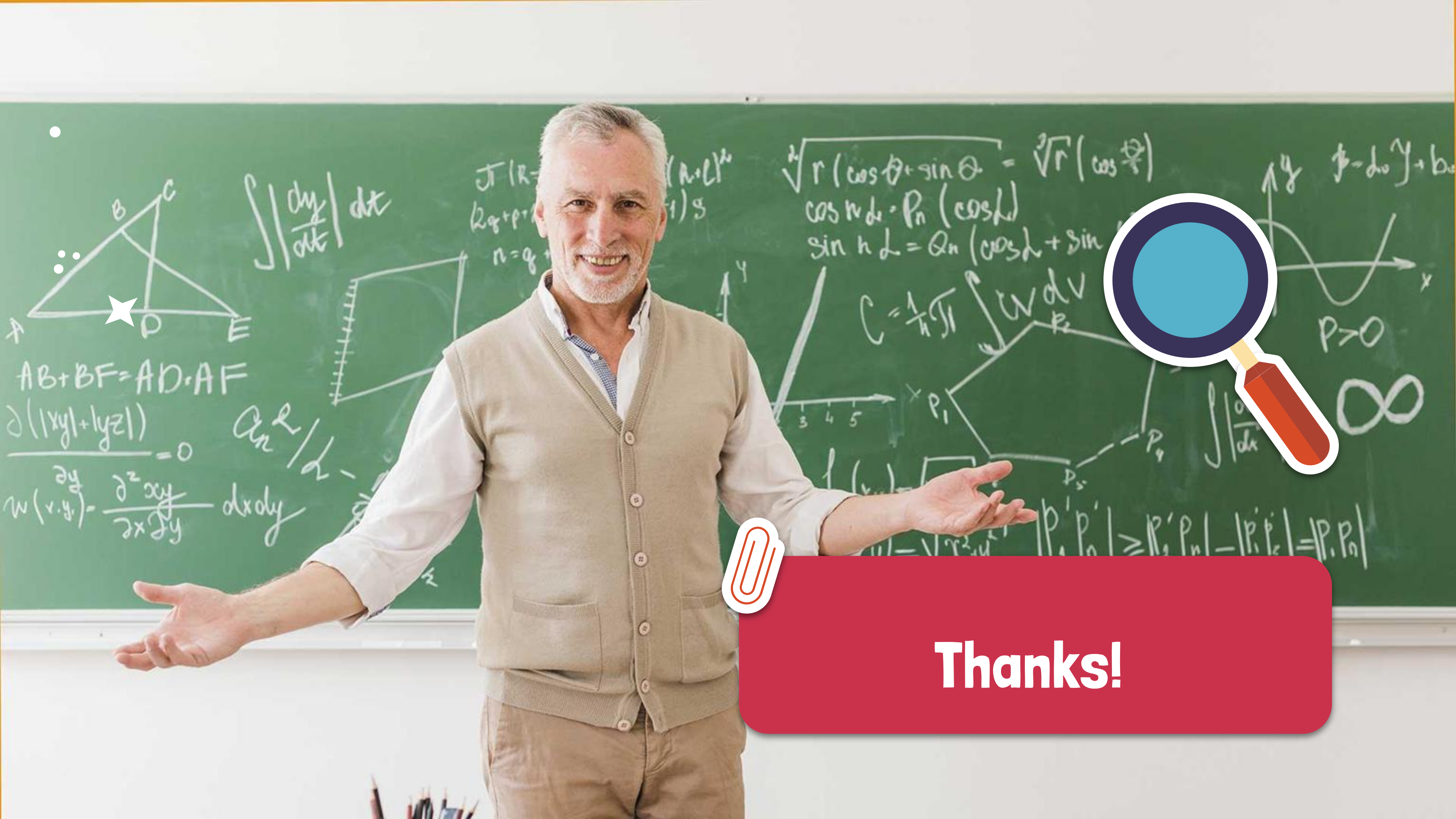
ดังนั้น นักศึกษาคนหนึ่งที่สอบวิชาสถิติได้ 70 คะแนน คาดว่าจะสอบวิชาภาษาอังกฤษได้ประมาณ 66 คะแนน

assignment

จากคะแนนสอบของนักศึกษาเอกคณิตศาสตร์ที่สอบวิชาแคลคูลัส 1 และแคลคูลัส 2 จำนวน 12 คน เป็นตารางด้านล่างอยากทราบว่าคะแนนสอบทั้ง 2 วิชามีความสัมพันธ์กันเพียงใด

คนที่	แคลคูลัส1	แคลคูลัส2
1	51	74
2	68	70
3	72	88
4	97	93
5	55	67
6	73	73
7	95	99
8	74	73
9	20	33
10	91	91
11	74	80
12	80	86

คำตอบ = 0.935


Thanks!

บทที่ 6 การวิเคราะห์สหสัมพันธ์และการถดถอย

ผศ ดร มนันยา มีนคร
คณะวิทยาการจัดการ

การวิเคราะห์สหสัมพันธ์

ค่าสัมประสิทธิ์สหสัมพันธ์ เป็นตัวเลขตัวเดียวที่บอกให้ทราบว่า 2 สิ่งมีความสัมพันธ์กันหรือไม่ และมีความสัมพันธ์กันมากน้อยเพียงใด ซึ่งเป็นรากฐานของการทำนาย เช่น ถ้าเราอยากทราบว่า คนที่ได้คะแนนคนที่ได้วิชาสถิติสูง จะทำงานได้บัญชีได้ดี เราก็จะสามารถใช้คะแนนวิชาสถิติทำนายการปฏิบัติงานด้านการบัญชีในภายหลังได้ ถ้าการทำนายเป็นไปอย่างถูกต้องมาก ก็กล่าวได้ว่ามีสหสัมพันธ์ระหว่างคะแนนวิชาสถิติและความสำเร็จในการทำงานบัญชี เป็นบวก ซึ่งหาได้โดยการคำนวนหาค่าสัมประสิทธ์ของคะแนนของคนหลายๆคน คะแนนของคนเดียว จะคำนวนหาค่าสัมประสิทธ์สหสัมพันธ์ไม่ได้และการที่เรารู้ค่าสหสัมพันธ์และใช้ทำนายได้ เป็นประโยชน์มากทั้งในด้านการคัดเลือกหรือการจัดคนให้เกมาะสมกับงานโดยใช้การทดสอบ ในเรื่องอื่นๆก็เช่นเดียวกัน

การวิเคราะห์สหสัมพันธ์

เราสามารถทำนายลักษณะของตัวแปรหนึ่ง จากอีกตัวแปรหนึ่งได้ คสามมสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรนี้ เรียกว่า “สหสัมพันธ์อย่างง่าย” ถ้าหากความสัมพันธ์นั้นมีตัวแปรตั้งแต่ 3 ตัวขึ้นไป เรียกว่า สหสัมพันธ์เชิงซ้อน นักสถิติใช้สัญลักษณ์ r แสดงค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายของกลุ่มตัวอย่าง และใช้สัญลักษณ์ $\rho\ (rho)$แสดงค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายของประชากร

การวิเคราะห์สหสัมพันธ์

ในการศึกษาสหสัมพันธ์อย่างง่าย เราใช้ค่าสถิติ r ประมาณค่าพารามิเตอร์ ρ เพราะในทางปฏิบัติเราไม่สามารถหาค่าสังเกตทั้งหมดจากประชากรได้ r จึงเป็นค่าแสดงความสัมพันธ์ระหว่างข้อมูล 2 ชุด และ r จะมีค่าอยู่ระหว่าง -1 และ +1

ถ้า r มีค่า +1 แสดงว่าข้อมูลทั้งสองชุดมีความสัมพันธ์กันอย่างสมบูรณ์ในทิศทางเดียวกัน หรือค่าสหสัมพันธ์เป็นบวกสมบูรณ์ (X เพิ่ม Y เพิ่ม)

ถ้า r มีค่า -1 แสดงว่าข้อมูลทั้งสองชุดมีความสัมพันธ์กันอย่างสมบูรณ์ในทิศทางตรงกันข้าม หรือค่าสหสัมพันธ์เป็นลบสมบูรณ์ (X เพิ่ม Y ลด)

การวิเคราะห์สหสัมพันธ์

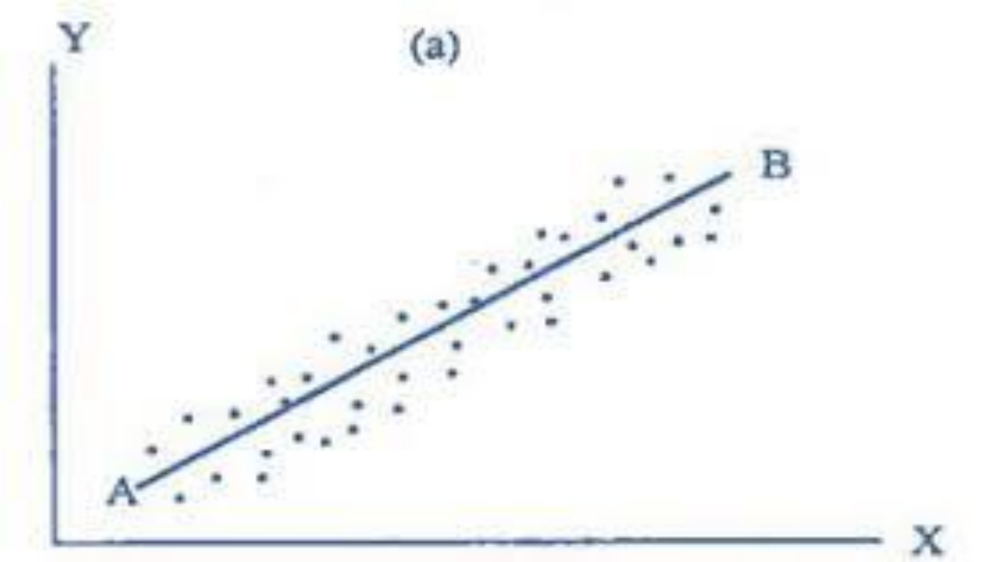
ถ้า $r = 0$ แสดงว่าข้อมูลทั้งสองชุดไม่มีความสัมพันธ์กันในกรณีที่ $r = \pm 1$ เราสามารถใช้ข้อมูล x ไปทำนายข้อมูล y ในเส้นถดถอยที่คำนวณอย่างถูกต้อง 100%

ถ้า r มีค่าเข้าใกล้ 0 แสดงว่าข้อมูล x และ y ไม่มีความสัมพันธ์กันเลย หรือมีก็น้อยมาก จึงไม่สามารถใช้ข้อมูล x ไปทำนายข้อมูล y ได้

วิธีการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายนิยมทำดังนี้

1. พิจารณาลักษณะการกระจายของข้อมูลจากแผนภาพการกระจาย
2. โดยการใช้สูตรคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์

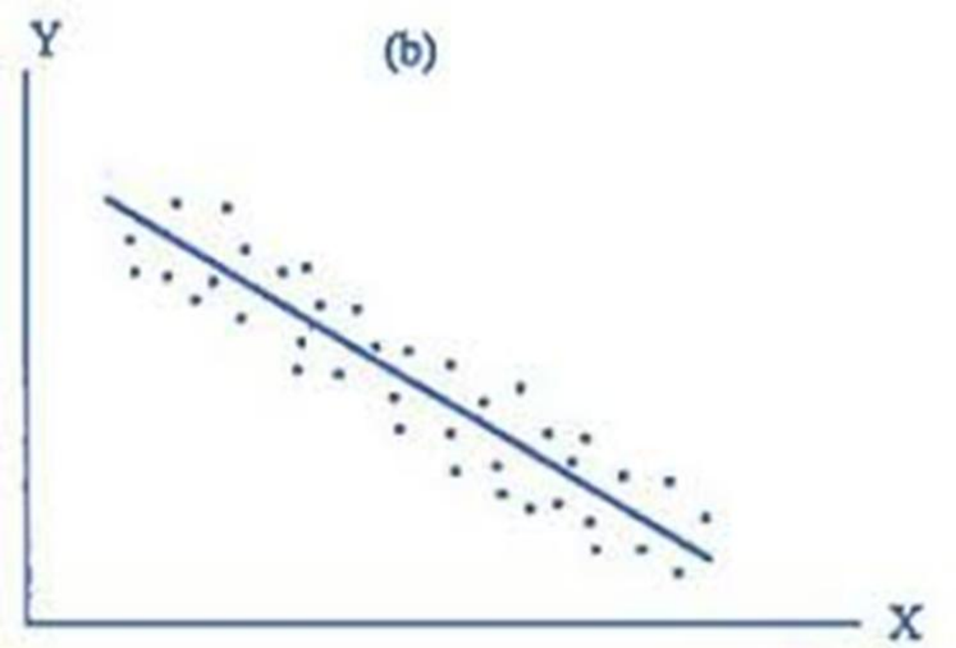
แผนภาพการกระจาย



ในรู ป (a) แนวโน้มของจุดชี้ขึ้นด้านขวาตามแนวเส้นตรง เมื่อ x มีค่ามากy มีค่ามากและ เมื่อx มีค่าน้อย y มีค่าน้อย

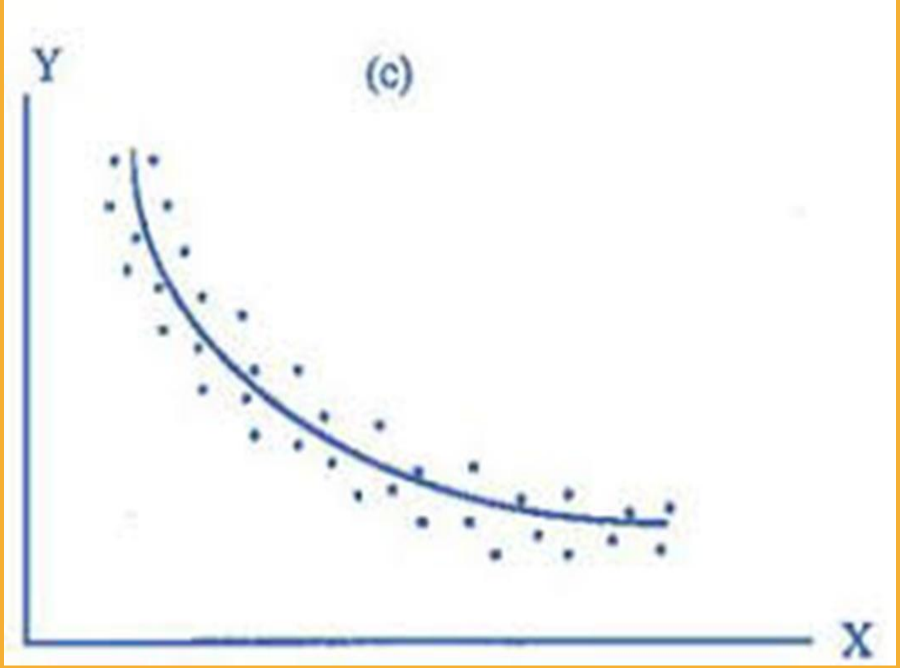
เรียกว่ามีความสัมพันธ์เชิงเส้นเชิงบวก

แผนภาพการกระจาย



ในรูป (b) แนวโน้มของจุดชี้ลงด้านขวาตามแนวเส้นตรงเมื่อ x มีค่ามากy มีค่าน้อย
เมื่อx มีค่าน้อย y มีค่ามาก
เรียกว่ามีความสัมพันธ์เชิงเส้นเชิงลบ

แผนภาพการกระจาย



ในรูป (C) แนวโน้มของจุดชี้ลงด้านขวาตามแนวเส้นโค้ง เรียกว่ามีความสัมพันธ์ไม่เชิงเส้นเชิงลบ

การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

วิธีนี้ใช้ในการคำนวณหาค่าเป็นการหาสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรสองตัวที่ใช้กันมากและรู้จักกันดีที่สุด ทั้งยังเป็นรากฐานของสหสัมพันธ์แบบอื่นๆบางแบบเขียนแทนสัญลักษณ์ r หรือ r_{xy}

การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

สูตรที่ 1 คำนวณโดยใช้คะแนนมาตรฐาน

$$r_{xy} = \frac{\sum Z_X Z_y}{n}$$

เมื่อ r_{xy} แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y

Z_X แทน คะแนนมาตรฐานแต่ละค่าของตัวแปร X

Zy แทน คะแนนมาตรฐานแต่ละค่าของตัวแปร Y

n แทน จำนวนคู่ของข้อมูล

สูตรที่ 2 คำนวณโดยใช้คะแนนที่เป็นค่าเบี่ยงเบนจากค่าเฉลี่ย

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left(n\sum X^2 - (\sum X)^2\right)\left(n\sum Y^2 - (\sum Y)^2\right)}}$$

เมื่อ r_{xy} แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y

n แทน จำนวนคู่ของข้อมูล

$\sum XY$ แทน ผลรวมทั้งหมดของผลคูณของตัวแปร XกับตัวแปรY

$\sum X$ แทน ผลรวมของตัวแปร X

$\sum Y$ แทน ผลรวมของตัวแปร Y

$\sum X^2$ แทน ผลรวมของตัวแปร X แต่ละค่ายกกำลังสอง

$\sum Y^2$ แทน ผลรวมของตัวแปร Y แต่ละค่ายกกำลังสอง

สูตรที่2 เป็นสูตรที่นิยมใช้มากที่สุด เพราะข้อมูลที่เก็บรวบรวมมาได้ส่วนใหญ่อยู่ในรูปคะแนนดิบ

ตัวอย่าง จงหาความสัมพันธ์ระหว่างคะแนนสอบวิชาคณิตศาสตร์กับคะแนนวิชาวิทยาศาสตร์ของ นศ จำนวน 6 คนดังตาราง

คนที่	คะแนนวิชาคณิตศาสตร์ (X)	คะแนนวิชาวิทยาศาสตร์ (Y)
1	9	5
2	10	9
3	8	5
4	7	8
5	4	6
6	8	4

วิธีทำ

คนที่	X	Y	X^2	Y^2	XY
1	9	5	81	25	45
2	10	9	100	81	90
3	8	5	64	25	40
4	7	8	49	64	56
5	4	6	16	36	24
6	8	4	64	16	32
รวม	46	37	374	247	287

จากตารางจะได้ $n = 6$, $\sum X = 46$, $\sum Y = 37$, $\sum X^2 = 374$, $\sum Y^2 = 247$, $\sum XY = 287$

แทนค่าสูตร

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left(n\sum X^2 - (\sum X)^2\right)\left(n\sum Y^2 - (\sum Y)^2\right)}}$$

$$= \frac{6(287) - (46)(37)}{\sqrt{\left(6(374) - (46)^2\right)\left(6(247) - (37)^2\right)}}$$

$$= \frac{20}{\sqrt{14464}}$$

$$= 0.17$$

ตัวอย่าง จงหาความสัมพันธ์จากคะแนนของนักเรียน 100 คนที่กำหนดให้ต่อไปนี้

คนที่	X	Y	XY	X^2	Y^2
1	14	19	266	196	361
2	11	11	121	121	121
3	6	5	30	36	25
4	17	14	238	289	196
5	8	10	80	64	100
6	13	16	208	169	256
7	6	14	84	36	196
8	9	10	90	81	100
9	16	15	240	256	225
10	12	14	168	144	196
	112	128	1,525	1,392	1,776

แทนค่าสูตร

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left(n\sum X^2 - (\sum X)^2\right)\left(n\sum Y^2 - (\sum Y)^2\right)}}$$

$$= \frac{10(1{,}525) - (112)(128)}{\sqrt{\left(10(1{,}392) - (112)^2\right)\left(10(1{,}776) - (128)^2\right)}}$$

$$= \frac{914}{\sqrt{(1{,}376)(1{,}376)}}$$

$$= \frac{914}{1{,}376}$$

$$= 0.66$$

สัมประสิทธิ์สหสัมพันธ์แบบอันดับที่ของสเปียร์แมน หรือเรียกอีกอย่างหนึ่งว่า rank correlation เป็นวิธีการทางสถิติในการศึกษาความสัมพันธ์ของ 2 ตัวแปร จากกลุ่มตัวอย่างเดียวกัน โดยข้อมูลที่ได้จากตัวแปรทั้งสองอยู่ในมาตราเรียงอันดับ(ordinal scale) ซึ่งเขียนแทนสัญลักษณ์ ρ และคำนวณหาได้จากสูตร

$$\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

ρ แทน **สัมประสิทธิ์สหสัมพันธ์**

D แทน **ผลต่างอันดับที่ของข้อมูลแต่ละคู่**

n แทน **จำนวนคู่ของข้อมูล**

ตัวอย่าง ผลการเรียงอันดับความตั้งใจเรียนของ นิสิต ปริญญาเอก จำนวน 6 คน โดยอาจารย์ที่ปรึกษา และคณบดีได้ผลการจัดอันดับดังนี้

นักศึกษาคนที่	ผลการจัดอันดับที่	
	ที่ปรึกษา(X)	คณบดี(Y)
1	6	5
2	1	2
3	4	3
4	3	4
5	5	1
6	2	6

วิธีทำ

นักศึกษาคนที่	X	Y	D=X-Y	$D^2=(X-Y)^2$
1	6	5	1	1
2	1	2	-1	1
3	4	3	1	1
4	3	4	-1	1
5	5	1	4	16
6	2	6	-4	16
				36

นิสิตจำนวน 6 คน ดังนั้น n = 6

แทนค่าในสูตร

$$\rho = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6(36)}{6(6^2 - 1)}$$

$$= 1 - \frac{216}{210}$$

$$= 1 - 1.03$$

$$= -0.03$$

ตัวอย่าง ผลการสอบวิชาคณิตและวิชาวิทยาศาสตร์ของนักศึกษา 10 คนปรากฏดังนี้

คนที่	1	2	3	4	5	6	7	8	9	10
คะแนนวิชาคณิต	10	17	12	19	11	14	15	16	13	18
คะแนนวิชาวิทย	11	15	13	19	12	16	14	18	10	17

จงหาค่าสหสัมพันธ์ระหว่างอันดับที่ของผลการสอบวิชาทั้งสอง

วิธีทำ ก่อนอื่นต้องจัดอันดับคะแนนทั้งสองวิชาดังตาราง

คนที่	คณิตR_x	วิทย์ R_y	D= R_x-R_y	D^2
1	10	9	1	1
2	3	5	-2	4
3	8	7	1	1
4	1	1	0	0
5	9	8	1	1
6	6	4	2	4
7	5	6	-1	1
8	4	2	2	4
9	7	10	-3	9
10	2	3	-1	1

จากตารางจะได้ $\sum D = 0$

$\sum D^2 = 26$

$$\rho = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

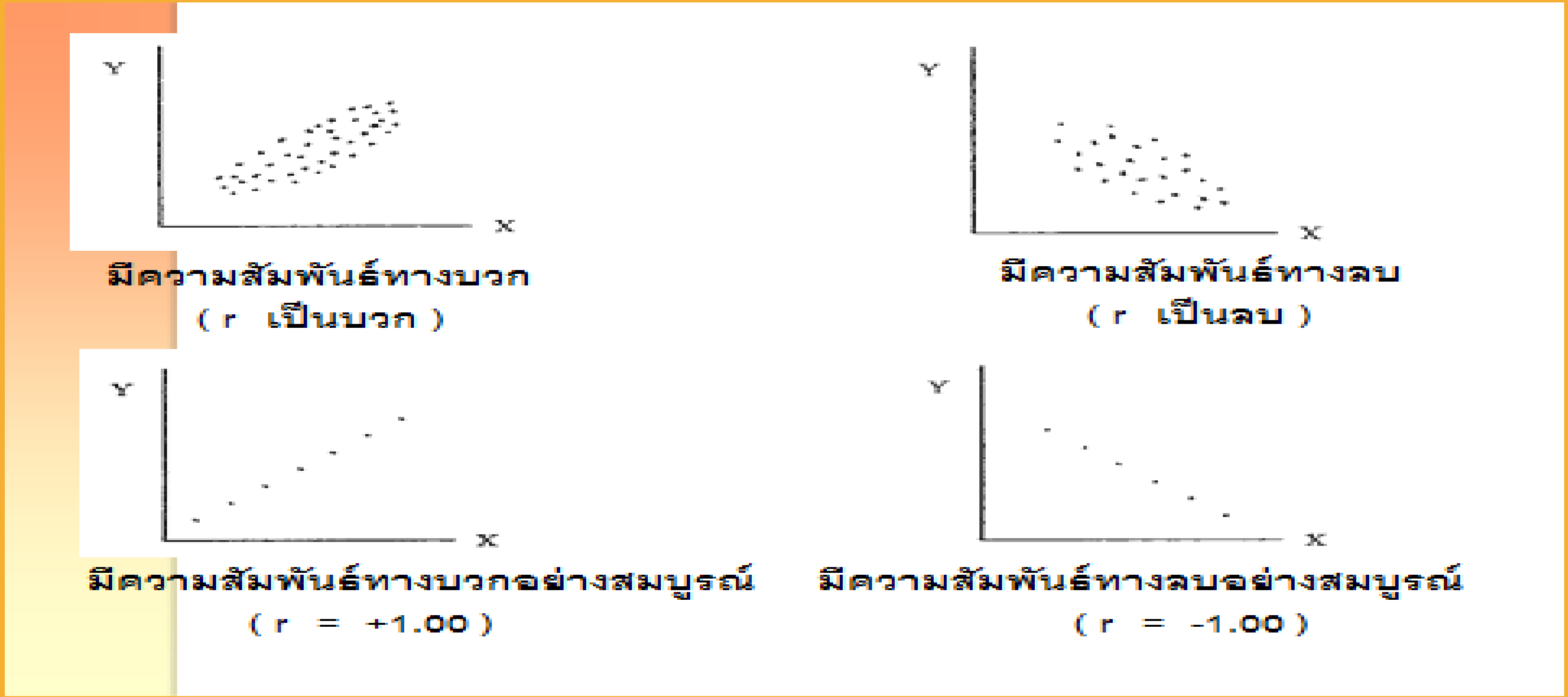
$$= 1 - \frac{6(26)}{10(100-1)}$$

$$= 1 - 0.1576$$

$$= 0.84$$

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

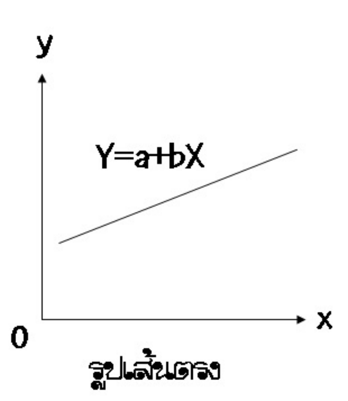
การวิเคราะห์การถดถอย เป็นวิธีการหาสมการที่สามารถใช้ในการพยากรณ์ หรือคาดคะเนเกี่ยวกับตัวแปรที่ต้องการศึกษา โดยอาศัยความรู้เกี่ยวกับค่าของตัวแปรที่เกี่ยวข้องนั่นคือ จากค่าของตัวแปรที่ต้องการศึกษาซึ่งจะเรียกว่า ตัวแปรตาม กับค่าของตัวแปรที่เกี่ยวข้อง ซึ่งจะเรียกว่า ตัวแปรอิสระ เราจะหาสมการซึ่งเพียงพอในการกำหนดความสัมพันธ์ระหว่างตัวแปรทั้งสอง สมการนี้ เรียกว่า **สมการเส้นถดถอย**



ภาพของการกระจาย

สมการถดถอยเชิงเส้นอย่างง่าย

การวิเคราะห์ความสัมพันธ์อย่างง่ายเป็นการประมาณค่า และพยากรณ์ค่าของตัวแปรตัวหนึ่ง โดยใช้ค่าของอีกตัวแปรหนึ่งเป็นตัวพยากรณ์ ในการวิเคราะห์ถดถอยของตัวแปร 2ตัว ซึ่งมีความสัมพันธ์ในแบบเส้นตรงจะต้องสร้างสมการเส้นถดถอยเพื่อใช้ในการพยากรณ์ โดยอาศัยรูปแบบของสมการเส้นตรงในทางคณิตศาสตร์ $y = a+bx$



เมื่อ a และ b เป็นค่าคงที่ สมการถดถอยของกลุ่มตัวอย่าง

สมการถดถอยเชิงเส้นอย่างง่าย

สูตรสมการถดถอยเชิงเส้นอย่างง่าย

$$b = \frac{n\sum xy - \Sigma x \Sigma y}{n\sum x^2 - (\Sigma x)^2} \longrightarrow 1$$

$$a = \bar{y}_t - b\bar{x}_t \longrightarrow 2$$

สมการถดถอยเชิงเส้นอย่างง่าย

สูตรสมการถดถอยเชิงเส้นอย่างง่าย

$$\hat{Y}_t = a + bx_t \longrightarrow 3$$

$\hat{Y}_t$ คือ ค่าแนวโน้มที่แปรตามตัวแปรเวลาที่เราต้องการหา

a คือ ระยะทางจากจุดกำเนิดถึงจุดตัดบนแกน y

b คือ ความชันของสมการเส้นตรง

x_t คือ ตัวแปรเวลา ณ เวลาที่ t

ตัวอย่าง จากข้อมูลในตารางเป็นคะแนนสอบกลางภาควิชาสถิติและวิชาอังกฤษของนักศึกษาแผนกวิชาคอมพิวเตอร์วิทยาลัยแห่งหนึ่งจำนวน 10 คน

คะแนนสอบวิชาสถิติ	75	55	71	72	80	95	95	98	65	82
คะแนนสอบวิชาอังกฤษ	80	65	78	35	48	85	98	98	66	83

1. จงสร้างสมการการถดถอยที่แสดงความสัมพันธ์ระหว่างคะแนนสอบวิชาสถิติกับวิชาภาษาอังกฤษ
2. ถ้านักศึกษาคนหนึ่งสอบวิชาสถิติได้ 70 คะแนน จะสอบวิชาภาษาอังกฤษได้ประมาณกี่คะแนน

คะแนนสอบวิชาสถิติ (X)	คะแนนสอบวิชาภาษาอังกฤษ(Y)	X^2	Y^2	XY
75	80	5,625	6,400	6,000
55	65	3,025	4,225	3,575
71	78	5,041	6,084	5,538
72	35	5,184	1,225	2,520
80	48	6,400	2,304	3,840
95	85	9,025	7,225	8,075
95	98	9,025	9,604	9,310
98	98	9,604	9,604	9,604
65	66	4,225	4,356	4,290
82	83	6,724	6,889	6,806
$\sum$=788	$\sum$=736	$\sum$=63,878	$\sum$=5,7916	$\sum$=59,558

$$b = \frac{n\sum xy - \Sigma x \Sigma y}{n\sum x^2 - (\Sigma x)^2}$$

1

$$b = \frac{10(59{,}558) - 788(736)}{10(63{,}878) - (788)^2}$$

$$b = \frac{15{,}612}{17{,}836}$$

b = 0.8753

จากนั้นหา =

$\bar{x} = \frac{\Sigma X}{n}$ และ $\bar{y} = \frac{\Sigma Y}{n}$

$\bar{x} = \frac{788}{10}$ $(y)^{-} = \frac{736}{10}$

$\bar{x} = 78.8$ และ $\bar{y} = 73.6$

$$a = \bar{y}_t - b\bar{x}_t \longrightarrow \boxed{2}$$

$$a = 73.6 - 0.8753\ (78.8)$$

$$a = 73.6 - 68.97364$$

1. สมการการถดถอยที่แสดงความสัมพันธ์ระหว่างคะแนนสอบวิชาสถิติกับวิชาภาษาอังกฤษ คือ

$$a = 4.6264$$

$$\hat{Y}_t = a + bx_t \longrightarrow \boxed{3}$$

$$\hat{y} = 4.6264 + 0.8753\ X$$

2. นักศึกษาคนหนึ่งสอบวิชาสถิติได้ 70 คะแนน X = 70

$$\hat{y} = 4.6264 + 0.8753\ X$$

$$\hat{y} = 4.6264 + 0.8753(70)$$

$$= 4.6264 + 61.261$$

$$= 65.8974$$

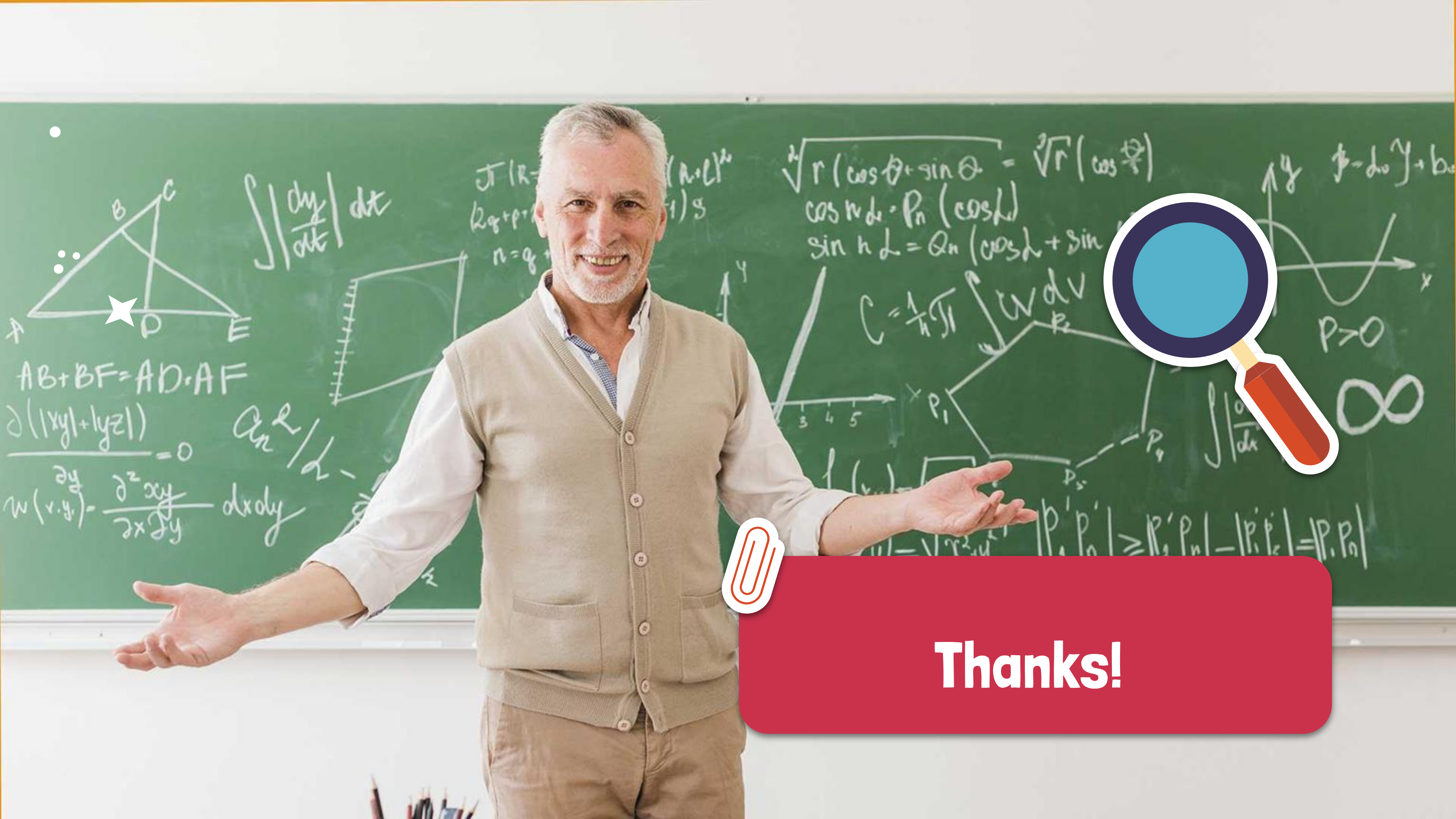
ดังนั้น นักศึกษาคนหนึ่งที่สอบวิชาสถิติได้ 70 คะแนน คาดว่าจะสอบวิชาภาษาอังกฤษได้ประมาณ 66 คะแนน

assignment

1. จากคะแนนสอบของนักศึกษาเอกคณิตศาสตร์ที่สอบวิชาแคลคูลัส 1 และแคลคูลัส 2 จำนวน 12 คน เป็นตารางด้านล่างอยากทราบว่าคะแนนสอบทั้ง 2 วิชามีคามสัมพันธ์กันเพียงใด

คนที่	แคลคูลัส1	แคลคูลัส2
1	51	74
2	68	70
3	72	88
4	97	93
5	55	67
6	73	73
7	95	99
8	74	73
9	20	33
10	91	91
11	74	80
12	80	86

คำตอบ = 0.935


Thanks!

บทที่ 7

ทฤษฎีการประมาณค่า

คณะวิทยาการจัดการ บริหารธุรกิจ
ผศ ดร มนันยา มีนคร

ทฤษฎีการประมาณค่า

เนื่องจากว่ามักจะหาค่าของประชากรที่เรียกว่า พารามิเตอร์ โยตรงไม่ได้ จึงตรงประมาณค่าพารามิเตอร์ ที่เราต้องการทราบจากค่าที่คำนวณได้จากกลุ่มตัวอย่างที่เรียกว่าค่าสถิติ การประมาณค่ามี 2 แบบ การประมาณค่าแบบจุด (Point Estimation) และการประมาณค่าแบบช่วง (Interval Estimation)

การประมาณค่า หมายถึง ฟังก์ชั่นของตัวแปรสุ่ม ซึ่งใช้ประมาณค่าพารามิเตอร์ที่สนใจจะศึกษา โดยใช้สัญลักษณ์ดังนี้

ค่าพารามิเตอร์	ตัวประมาณค่า
ค่าเฉลี่ยประชากร (μ)	ค่าเฉลี่ยกลุ่มตัวอย่าง ($\bar{X}$)
ค่าสัดส่วนประชากร (P)	ค่าสัดส่วนกลุ่มตัวอย่าง ($\hat{p}$)
ค่าความแปรปรวนประชากร (σ^2)	ค่าความแปรปรวนกลุ่มตัวอย่าง (S^2)
ค่าส่วนเบี่ยงเบนมาตรฐานประชากร (σ)	ค่าส่วนเบี่ยงเบนมาตรฐานกลุ่มตัวอย่าง (S)
ผลต่างระหว่างค่าเฉลี่ยประชากร 2 กลุ่ม ($\mu_1 - \mu_2$)	ผลต่างระหว่างค่าเฉลี่ยกลุ่มตัวอย่าง 2 กลุ่ม ($\bar{X}_1 - \bar{X}_2$)
ผลต่างระหว่างค่าสัดส่วนประชากร 2 กลุ่ม ($p_1 - p_2$)	ผลต่างระหว่างค่าสัดส่วนกลุ่มตัวอย่าง 2 กลุ่ม ($\hat{p}_1 - \hat{p}_2$)

การประมาณค่า

การประมาณค่าแบบจุด (Point Estimation) หรือเรียกว่าการประมาณค่าเดียว เป็นการคำนวณหาค่าสถิติตัวหนึ่งจากข้อมูลตัวอย่าง แล้วนำค่าที่ได้ไปเป็นค่าประมาณพารามิเตอร์ เช่น การใช้ค่าเฉลี่ยของตัวอย่าง ($\bar{x}$) เป็นตัวประมาณค่าเฉลี่ยของพารามิเตอร์ประชากร (μ) หรือการใช้ค่าสัดส่วนของตัวอย่าง $\hat{P}$ เป็นตัวประมาณค่าสัดส่วนของพารามิเตอร์ประชากร (P)

การประมาณค่าแบบจุด เป็นค่าประมาณค่า เพียงค่าเดียวที่มีโอกาสเท่ากับหรือไม่เท่ากับ ค่าพารามิเตอร์ก็ได้ มีโอกาสคลาดเคลื่อนจากค่าพารามิเตอร์ของประชากรสูง <u>จึงไม่นิยมนำมาใช้เพื่อประมาณค่ากันมากนัก</u>

การหาค่าประมาณแบบช่วง (**Interval Estimation)** เป็นการประมาณว่าค่าพารามิเตอร์ของประชากร จะอยู่ในช่วงใดช่วงหนึ่ง ที่มีค่าอยู่ระหว่างค่าสองค่า คือ

L = ขอบเขตต่ำ (**lower limit)** และ **U =** ขอบเขตสูง (**upper limit)**

สมมติว่าพารามิเตอร์ที่ต้องการประมาณในประชากรหนึ่ง คือ θ ค่าประมาณแบบเป็นช่วงของ θ นี้จะอยู่ในช่วง $L < \theta < U$ โดยใช้ข้อมูลจากกลุ่มตัวอย่าง เช่น

รายได้เฉลี่ยต่อเดือนของคนในเขตกรุงเทพมหานคร อยู่ในช่วง 3,500 ถึง 5,500 นั่นคือ $3{,}500 < \mu < 5{,}500$

สัดส่วนของนักเรียนที่จบชั้นมัธยมศึกษาปีที่ 3 ของประเทศไทยและเรียนต่อในสายอาชีพ อยู่ระหว่าง 0.39 ถึง 0.67 นั่นคือ $0.35 < p < 0.67$

การประมาณค่า

การประมาณค่าแบบช่วง

หมายถึง การประมาณค่าพารามิเตอร์ออกมาเป็นช่วงตัวเลข มีค่าต่ำสุด (a) และค่าสูงสุด (b) ที่คาดว่าจะคลุมค่าพารามิเตอร์ที่จะประมาณค่า อาจจะกว้างหรือแคบ

การประมาณค่าแคบ หมายถึง โอกาสที่ค่าพารามิเตอร์จะตกอยู่ในช่วงนั้นมีน้อย แต่ผลที่ได้รับจะมีความแม่นยำสูง

การประมาณค่ากว้าง หมายถึง โอกาสที่ค่าพารามิเตอร์จะตกอยู่ในช่วงนั้นมีมาก แต่ผลที่ได้รับจะมีความแม่นยำต่ำ หรืออาจใช้ไม่ได้เลยถ้าช่วงกว้างเกินไป

การประมาณค่า

การใช้ตารางค่า Z หาค่า α

ตารางที่ 5 UPPER CRITICAL VALUES OF THE t-DISTRIBUTION

$t_{\alpha;n}$

n	α = .10	α = .05	α = .025	α = .01	α = .005	n
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
∞	1.282	1.645	1.960	2.326	2.576	inf.

การประมาณค่าเฉลี่ยของประชากร

การประมาณค่าเฉลี่ยของประชากรกลุ่มเดียว

1 **การประมาณค่าแบบจุด เป็นการคำนวณหาค่าเฉลี่ย**

จากข้อมูลกลุ่มตัวอย่าง ไม่ว่าข้อมูลนั้นจะแจกแจงแบบใดก็ตาม โดยใช้สูตรการคำนวณ ดังนี้

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad \text{และ} \quad \bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

เมื่อ

μ แทน ค่าเฉลี่ยของประชากร

x แทน ค่าเฉลี่ยของตัวอย่าง

X_i แทน ค่าของลักษณะที่สนใจในหน่วย

N_i แทน ขนาดของประชากร

n แทน ขนาดของตัวอย่าง

ตัวอย่าง

คะแนนทดสอบวิชาสถิติและการวางแผนการทดลองของนักศึกษา จำนวน 12 คน ข้อมูลเป็นดังนี้ 16 19 15 14 13 11 13 8 12 17 16 14 จงหาค่าประมาณคะแนนการสอบโดยเฉลี่ยของนักศึกษากลุ่มนี้

<u>วิธีทำ</u> ให้ X_i แทน คะแนนการสอบแต่ละคน

$$= \frac{16 + 19 + 15 + 14 + 13 + 11 + 13 + 8 + 12 + 17 + 16 + 14}{12}$$

$$= \frac{168}{12} = 14 \#$$

ดังนั้น ค่าประมาณคะแนนสอบโดยเฉลี่ยของนักศึกษาเป็น 14 คะแนน

การประมาณค่าเฉลี่ยของประชากร

1. การหาค่าประมาณแบบช่วง การประมาณค่าเฉลี่ยของประชากรกลุ่มเดียว

แบบได้เป็น 3 กรณี

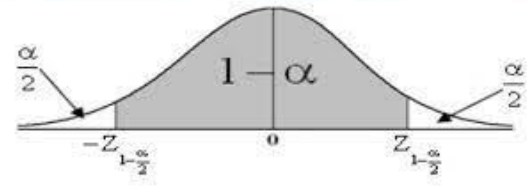
1.กรณีทราบความแปรปรวนของประชากร

ระดับความเชื่อมั่น (1- α)100% ที่นิยมใช้ในการประมาณค่าส่วนใหญ่มีค่าเป็น 90% 95% และ 99% จะได้ค่าประมาณแบบช่วงของค่าปกติมาตรฐาน Z อยู่ระหว่าง

สูตร

$$\bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

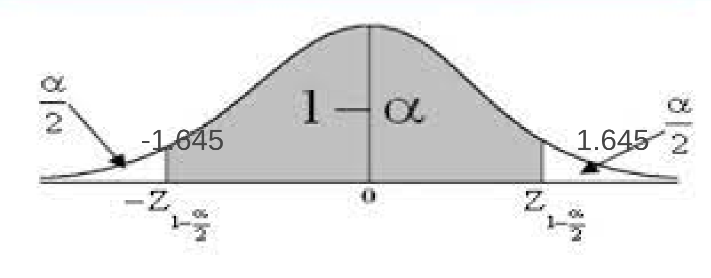
การสร้างช่วงความเชื่อมั่นในการประมาณค่าสัดส่วนของประชากร โดยกำหนดความระดับเชื่อมั่น $1-\alpha$ และพิจารณารูปต่อไปนี้



ตัวอย่าง จากคะแนนการเรียนวิชาสถิติและการวางแผนการทดลองของนักศึกษา ในรอบ 3 ปีที่ผ่านมา พบว่า คะแนนมีการแจกแจงปกติและมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 คะแนน ในการเรียนภาคเรียนนี้จึงสุ่มนักศึกษามา 20 คน เพื่อหาคะแนนเฉลี่ยและพบว่ามีคะแนนเฉลี่ยเท่ากับ 63.5 คะแนน จงประมาณคะแนนเฉลี่ยของนักศึกษา ที่ระดับความเชื่อมั่น 90% ($Z_{\alpha/2}$ = 1.645)

<u>วิธีทำ</u> จากโจทย์ พบว่าประชากรมีการแจกแจงปกติ และทราบส่วนเบี่ยงเบนมาตรฐานของประชากรและ

$\sigma = 10$, $n = 20$, $\overline{X} = 63.5$, ระดับความเชื่อมั่น 90% $Z_{\alpha/2}$ = 1.645



จากสูตรการประมาณค่า $\bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$

แทนค่า

$$= 63.5 - 1.645\left(\frac{10}{\sqrt{20}}\right) < \mu < 63.5 + 1.645\left(\frac{10}{\sqrt{20}}\right)$$

$$= 63.5 - 1.645\,(2.237) < \mu < 63.5 + 1.645\,(2.237)$$

$$= 63.5 - 3.680 < \mu < 63.5 + 3.680$$

$$= 59.82 < \mu < 67.18$$

นักศึกษามีคะแนนเฉลี่ยระหว่าง 59.82 - 67.18 ที่ระดับความเชื่อมั่น 90%

2. กรณีไม่ทราบความแปรปรวนของประชากร กลุ่มตัวอย่างขนาดใหญ่ (n ≥ 30)

สูตร

$$\bar{X} - Z_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + Z_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

☛ กรณีมีการแจกแจงแบบปกติ และไม่ทราบส่วนเบี่ยงเบนมาตราฐานของประชากร σ และขนาดของตัวอย่างมากกว่าหรือเท่ากับ **30** (n ≥ 30) ใช้ค่าสถิติ **Z**

- **ตัวอย่าง** การสุ่มตัวอย่างร้านตัดเสื้อสตรีมา 100 ร้าน มีรายได้เฉลี่ยเดือนละ 18,000 บาท ส่วยเบี่ยงเบนมาตรฐาน 3,000 บาท จงหาช่วงความเชื่อมั่น 95% ของรายได้เฉลี่ยของร้านตัดเสื้อสตรี
- **วิธีทำ**

$$\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

กำหนดให้ $\overline{X} = 18{,}000$ บาท , $S = 3{,}000$ บาท และ $n = 100$

$$\alpha = 1 - 0.95 = 0.05$$

$$\alpha/2 = 0.025$$

วิธีทำ

จากตาราง $Z_{\alpha/2} = Z_{0.025} = 1.96$

แทนค่า $18{,}000 - 1.96\dfrac{(3{,}000)}{\sqrt{100}} < \mu < 18{,}000 + 1.96\dfrac{(3{,}000)}{\sqrt{100}}$

$$18{,}000 - 588 < \mu < 18{,}000 + 588$$

$$17{,}412 < \mu < 18{,}588$$

ดังนั้นช่วงความเชื่อมั่น 95% รายได้เฉลี่ยของร้านตัดเสื้ออยู่ระหว่าง 17,412 บาท ถึง 18,588 บาท

3. กรณีไม่ทราบความแปรปรวนของประชากร กลุ่มตัวอย่างขนาดเล็ก (n ≤ 30)

สูตร

$$\bar{X} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

ซึ่งมีค่า df = n -1

☛ กรณีมีการแจกแจงแบบปกติหรือใกล้เคียงปกติ และไม่ทราบส่วนเบี่ยงเบนมาตรฐานของประชากร σ และขนาดของตัวอย่างน้อยกว่า 30 (n ≤ 30) ใช้ค่าสถิติ t

ตัวอย่าง จากการสอบถามแม่บ้าน 16 คน ที่ทำการเลือกมาจากแม่บ้านทั้งหมดที่อาศัยอยู่ในหมู่บ้านจัดสรรแห่งหนึ่งที่ดูหนังก่อนรายการข่าวทางโทรทัศน์ได้ค่าเฉลี่ยเป็น 25 วัน ส่วนเบี่ยงเบนมาตรฐาน 2 วัน จงหาช่วงความเชื่อมั่น 90% ของ μ **(ค่าเฉลี่ยของประชากร)** ถ้าประชากรมีการแจกแจงปกติ

วิธีทำ

$$\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

วิธีทำ

กำหนดให้ $\overline{X} = 25$ วัน , $S = 2$ วัน และ $n = 16$, $df = n - 1 = 15$

$$\alpha = 1 - 0.90 = 0.10$$

$$\alpha/2 = 0.05$$

จากตาราง $t_{\alpha/2} = t_{0.05} = 1.753$

แทนค่า $25 - 1.753\frac{(2)}{\sqrt{16}} < \mu < 25 + 1.753\frac{(2)}{\sqrt{16}}$

☐ วิธีทำ

$$25 - 0.8765 < \mu < 25 + 0.8765$$

$$24.1235 < \mu < 25.8765$$

ดังนั้นช่วงความเชื่อมั่น 90% ของ μ อยู่ระหว่าง 24.1235 ถึง 25.8765

2. การประมาณค่าความแตกต่างระหว่างค่าเฉลี่ยประชากร 2 กลุ่ม

คือการประมาณค่าแบบ μ_1 - μ_2

<u>แบบจุด</u>

$$\mu_1 - \mu_2 = \bar{X}_1 - \bar{X}_2$$

$$\mu_1 - \mu_2 = \left(\frac{\sum X}{n}\right)_1 - \left(\frac{\sum X}{n}\right)_2$$

ตัวอย่าง สุ่มตัวอย่างหลอดไฟ **2** ชนิด คือชนิด **A** และ ชนิด **B** อย่างละ **25** หลอด พบว่ามีอายุการใช้งานเฉลี่ยเท่ากับ **1,400** และ **1,200** ชั่วโมงตามลำดับ ถ้าอายุการใช้งานของหลอดไฟทั้งสองชนิด มีการแจกแจงปกติ โดยมีส่วนเบี่ยงเบนมาตราฐานของอายุการใช้งานเท่ากับ **200** และ **100** ชั่วโมงตามลำดับ จงคำนวณค่าผลต่างระหว่างอายุการใช้งานเฉลี่ยของหลอดไฟทั้ง **2** ชนิด แบบจุด

วิธีทำ

$$\mu_1 - \mu_2 = \bar{X}_1 - \bar{X}_2$$

$$= 1,400 - 1,200$$

$$= 200 \ \#$$

แบบช่วง แบ่งออกได้ 2 กรณี

- **การประมาณช่วงผลต่างระหว่างค่าเฉลี่ยของประชากรสองกลุ่ม**
- **สูตรในการคำนวณ**
  - 1. สูตรการคำนวณเนื่องจากเราทราบความแปรปรวนของประชากร จึงประมาณช่างของ $\mu_1 - \mu_2$

$$(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2}\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2}\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$$

- **ตัวอย่าง** โรงงานผลิตรถยนต์ต้องการศึกษาอายุการใช้งานเฉลี่ยของไฟหน้ารถที่ผลิตจากโรงงาน A และโรงงาน B จึงสุ่มหลอดไฟจากโรงงาน A มา 150 หลอด พบว่าอายุการใช้งานเฉลี่ย 1,400 ชม ส่วนเบี่ยงเบนมาตรฐานของประชากร 120 ชม และสุ่มหลอดไฟจากโรงงาน B มา 100 หลอด พบว่าอายุการใช้งานเฉลี่ย 1,200 ชม ส่วนเบี่ยงเบนมาตรฐานของประชากร 80 ชม จงประมาณช่วงของความแตกต่างอายุการใช้งานเฉลี่ยที่ผลิตจากโรงงาน A และ B โดยมีระดับความเชื่อมั่น 95%

- **วิธีทำ** ให้ $\mu_A - \mu_B$ เป็นอายุการใช้งานเฉลี่ยของหลอดไฟโรงงาน A และ B จากสูตร

$$(\bar{X}_A - \bar{X}_B) - Z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} < \mu_1 - \mu_2 < (\bar{X}_A - \bar{X}_B) + Z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

วิธีทำ

เนื่องจากเราทราบความแปรปรวนของประชากร จึงประมาณช่วงของ $\mu_1 - \mu_2$

กำหนดให้ $n_A = 150$, $n_B = 100$

$\bar{X}_A = 1{,}400$, $\bar{X}_B = 1{,}200$

$\sigma_A = 120$, $\sigma_B = 80$

$\alpha = 1 - 0.95 = 0.05$, $\alpha/2 = 0.025$

จากตาราง $Z_{\alpha/2} = Z_{0.025} = 1.96$

วิธีทำ จากสูตร

$$(\bar{X}_A - \bar{X}_B) - Z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} < \mu_1 - \mu_2 < (\bar{X}_A - \bar{X}_B) + Z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

แทนค่า $(1{,}400-1{,}200) - 1.96\sqrt{\frac{(120)^2}{150} + \frac{(80)^2}{100}} < \mu_A - \mu_B < (1{,}400-1{,}200) + 1.96\sqrt{\frac{(120)^2}{150} + \frac{(80)^2}{100}}$

$$200 - 1.96\,(12.649) < \mu_A - \mu_B < 200 + 1.96\,(12.649)$$

$$200 - 24.79 < \mu_A - \mu_B < 200 + 24.79$$

$$175.21 < \mu_A - \mu_B < 224.79$$

ดังนั้นความแตกต่างของอายุการใช้งานเฉลี่ยของหลอดไฟที่ผลิตจากโรงงาน A และ B จะอยู่ระหว่าง 175 ชม ถึง 225 ชม

สูตรในการคำนวณ

- 2. สูตรการคำนวณเนื่องจากขนาดของกลุ่มตัวอย่างเกิน 30 และไม่ทราบค่าความแปรปรวนของประชากรจึงประมาณค่าจากการแจกแจงแบบ Z โดยใช้ S_1^2 และ S_2^2 แทน σ_1^2 และ σ_2^2 ดังนั้นช่วงความเชื่อมั่นของ $\mu_1 - \mu_2$ คือ

$$(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

- **ตัวอย่าง** แบบสอบวิชาคณิตศาสตร์ชุดหนึ่งนำไปสอบนักเรียนชาย 49 คน และนักเรียนหญิง 64 คน พบว่า นักเรียนชายได้คะแนนเฉลี่ย 74 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 6 คะแนน นักเรียนหญิงได้คะแนนเฉลี่ย 65 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 8 คะแนน จงประมาณช่วงความเชื่อมั่น 99% ของผลต่างของคะแนนเฉลี่ยของคะแนนชาย – หญิง

- **วิธีทำ** ให้ μ_1 และ μ_2 เป็นคะแนนเฉลี่ยชาย – หญิงตามลำดับ

  เนื่องจากขนาดของกลุ่มตัวอย่างเกิน 30 และไม่ทราบค่าความแปรปรวนของประชากรจึงประมาณค่าจากการแจกแจงแบบ Z โดยใช้ S_1^2 และ S_2^2 แทน σ_1^2 และ σ_2^2 ดังนั้นช่วงความเชื่อมั่นของ $\mu_1 - \mu_2$ คือ

วิธีทำ

$$(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2}\sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2}\sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}}$$

กำหนดให้ $n_1 = 49$, $n_2 = 64$

$\bar{X}_1 = 74$, $\bar{X}_2 = 65$

$S_1 = 6$, $S_2 = 8$

$\alpha = 1 - 0.99 = 0.01$, $\alpha/2 = 0.005$

จากตาราง $Z_{\alpha/2} = Z_{0.005} = 2.575$

วิธีทำ จากสูตร

$$(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2}\sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2}\sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}}$$

แทนค่า $(74\text{-}65) - 2.575\sqrt{\frac{(6)^2}{49} + \frac{(8)^2}{64}} < \mu_A - \mu_B < (74\text{-}65) + 2.575\sqrt{\frac{(6)^2}{49} + \frac{(8)^2}{64}}$

$$9 - 2.575\ (1.317) < \mu_A - \mu_B < 9 + 2.575\ (1.317)$$

$$9 - 3.39 < \mu_A - \mu_B < 9 + 3.39$$

$$5.61 < \mu_A - \mu_B < 12.39$$

ดังนั้นผลต่างของคะแนนเฉลี่ยของนักเรียนชาย หญิง จะอยู่ระหว่าง 5.61 คะแนน ถึง 12.39 คะแนน

การประมาณค่าสัดส่วนของประชากร 1 กลุ่ม

1. การประมาณค่าสัดส่วนของประชากรกลุ่มเดียวแบบจุด จะแบ่งลักษณะของประชากรออกเป็น 2 กลุ่ม คือ ลักษณะประชากรที่สนใจและลักษณะประชากรที่ไม่สนใจ

ให้ p เป็นสัดส่วนของประชากรที่สนใจศึกษา

q เป็นสัดส่วนของประชากรที่ไม่สนใจศึกษา

โดยที่ $q = 1 - p$ หรือ $p + q = 1$

เนื่องจากไม่ทราบ p ของประชากร จึงใช้ $\hat{P}$ ซึ่งเป็นสัดส่วนของตัวอย่าง แทนการประมาณสัดส่วนของประชากร

$$\hat{P} = \frac{\text{จำนวนตัวอย่างที่มีลักษณะสนใจ}}{\text{จำนวนตัวอย่างทั้งหมด}}$$

หรือ $\hat{P} = \dfrac{X}{n}$

ตัวอย่าง จากการสอบถามแม่บ้าน 50 คน เกี่ยวกับการใช้น้ำยาล้างจานยี่ห้อ A และ B พบว่า แม่บ้านใช้น้ำยาล้างจานยี่ห้อ A จำนวน 38 คน จงประมาณสัดส่วนของแม่บ้านที่ใช้น้ำยาล้างจานยี่ห้อ B

วิธีทำ ให้ $\hat{P}$ เป็นตัวประมาณสัดส่วนของผู้ใช้น้ำยาล้างห้องน้ำยี่ห้อ B

$$\hat{P} = \frac{\text{จำนวนตัวอย่างที่มีลักษณะสนใจ}}{\text{จำนวนตัวอย่างทั้งหมด}}$$

$$= \frac{12}{50}$$

50-18 = 12

$$= 0.24 \ \#$$

นั่นคือ ค่าประมาณสัดส่วนแบบจุดของแม่บ้านที่ใช้น้ำยาล้างจานยี่ห้อ B เท่ากับ 0.24 หรือ 24 %

2. การประมาณค่าสัดส่วนของประชากรกลุ่มเดียวแบบช่วง

เมื่อ **n** มีขนาดใหญ่ **p** มีการแจกแจงปกติ จะได้ค่าประมาณสัดส่วนแบบช่วง ดังสูตรต่อไปนี้

$$\mu = \hat{p} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

และ $\widehat{q} = 1 - \hat{P}$

ตัวอย่าง สุ่มชาวนาในจังหวัดอุดรธานีมา 100 คน พบว่าเป็นหนี้ธนาคาร ธกส. จำนวน 60 คน จงประมาณช่วงความเชื่อมั่น 95% ($Z_{\alpha/2}$ = 1.96) ของสัดส่วนชาวนาในจังหวัดอุดรธานีที่เป็นหนี้ธนาคาร

ให้ $\widehat{P}$ เป็นวัดส่วนของชาวนาที่เป็นหนี้

$$\hat{P} = \frac{X}{n}$$

$$= \frac{60}{100}$$

$$= 0.60$$

และ $\widehat{P}$ $= 1 - 0.60$

$= 0.40$

$$\widehat{q} = 1 - \hat{P}$$

ดังนั้น การประมาณค่าสัดส่วนของประชากรคือ $\mu = \hat{p} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$= 0.60 \pm 1.96\left(\sqrt{\frac{(0.6)(0.40)}{100}}\right)$$

$$= 0.60 \pm 0.096$$

$= 0.504$ ถึง 0.696

ดังนั้น สัดส่วนชาวนาในจังหวัดอุดรธานีที่เป็นหนี้ ธกส. อยู่ในช่วง 0.504 - 0.696 ที่ระดับความความเชื่อมั่น 95%

การประมาณค่าสัดส่วนของประชากร 2 กลุ่ม

ให้ $\hat{p}_1$ แทน สัดส่วนของตัวอย่างที่สนใจกลุ่มที่ 1

$\hat{p}_2$ แทน สัดส่วนของตัวอย่างที่สนใจกลุ่มที่ 2

$$\hat{p}_1 = \frac{x_1}{n_1} \qquad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\hat{q}_1 = 1 - \hat{p}_1 \qquad \hat{q}_2 = 1 - \hat{p}_2$$

ดังนั้น การประมาณช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $P_1 - P_2$ คือ

$$\boxed{(\hat{p}_1 - \hat{p}_2) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < P_1 - P_2 < (\hat{p}_1 - \hat{p}_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

ตัวอย่าง จากการสุ่มตัวอย่างคนเชียงใหม่ผู้ชาย **400** คน และผู้หญิง **500** คน พบว่าชอบดูรายการข่าวการมือง จำนวน **348** คน และ **280** คน จงประมาณช่วงความเชื่อมั่น **99 %** ($Z_{\alpha/2}$ **= 2.576**) ผลต่างของสัดส่วนระหว่างผู้ชายกับผู้หญิงที่ชอบดูข่าวการเมือง

วิธีทำ ให้ $\widehat{P_1}$ แทน สัดส่วนของผู้ชายที่ชอบดูข่าวการเมือง

$\widehat{P_2}$ แทน สัดส่วนของผู้ผญิงที่ชอบดูข่าวการเมือง

จะได้

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{348}{400} = 0.87$$

$$\hat{q}_1 = 1 - \hat{p}_1 = 1 - 0.87 = 0.13$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{280}{500} = 0.56$$

$$\hat{q}_2 = 1 - \hat{p}_2 = 1 - 0.56 = 0.44$$

แทนค่าในสูตร

$$(\hat{p}_1 - \hat{p}_2) - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < P_1 - P_2 < (\hat{p}_1 - \hat{p}_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(0.87 - 0.56) - 2.576\sqrt{\frac{(0.87\times0.13)}{400} + \frac{(0.56\times0.44)}{500}} < P_1 - P_2 < (0.87 - 0.56) + 2.576\sqrt{\frac{(0.87\times0.13)}{400} + \frac{(0.56\times0.44)}{500}}$$

$$0.31 - 0.07 < P_1 - P_2 < 0.31 + 0.07$$

$$0.24 < P_1 - P_2 < 0.38$$

ดังนั้นสัดส่วนของผู้ชายและผู้หญิงที่ชอบดูรายการนี้ต่างกันประมาณ 0.24 ถึง 0.38 หรือ คิดเป็นร้อยละ 24 ถึง ร้อยละ 38 ที่ระดับความเชื่อมั่น 99%

THANKS!

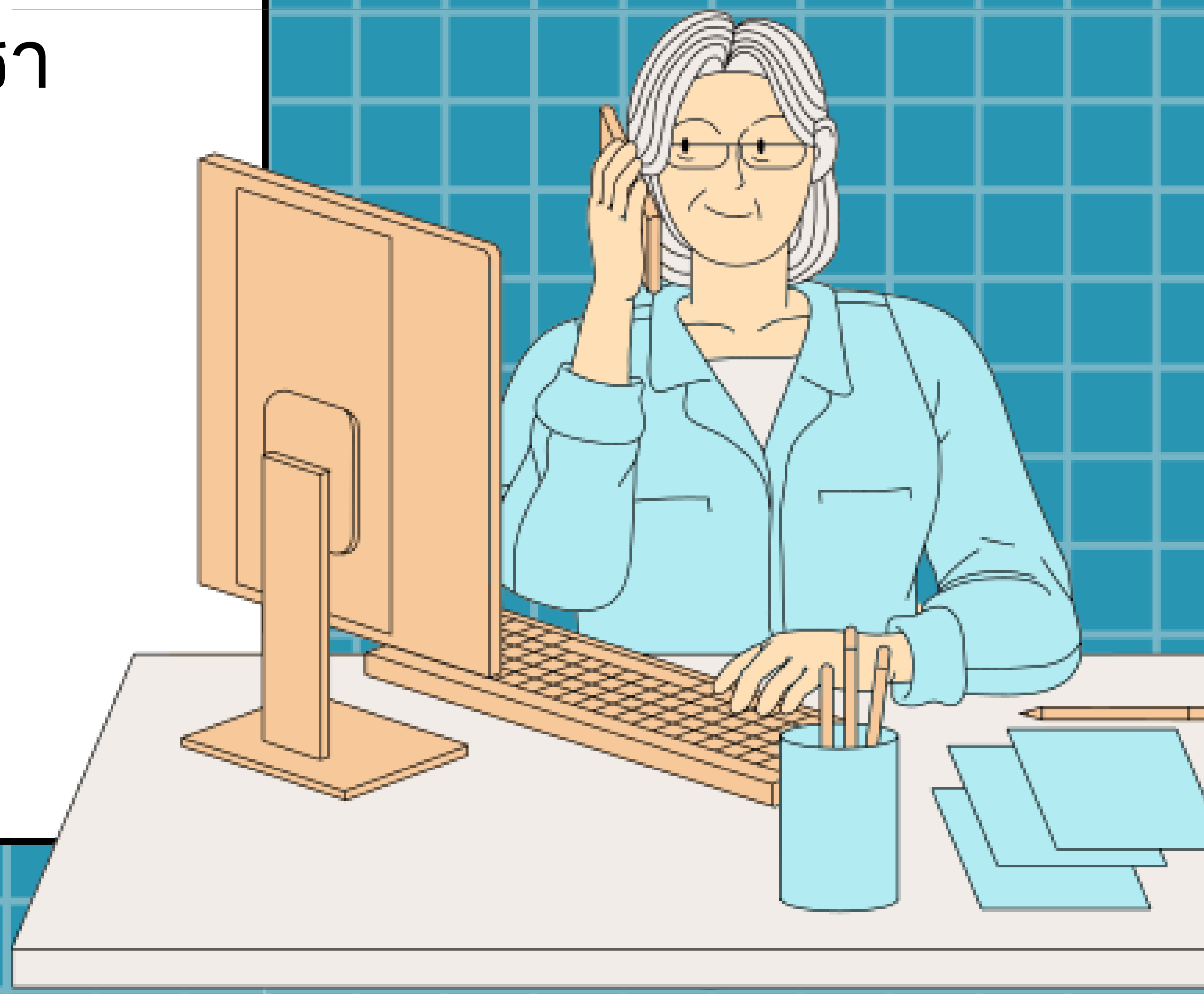
บทที่ 9

การทดสอบสมมติฐาน

นำเสนอโดย ผศ ดร มนันยา มีนคร

การทดสอบสมมติฐาน

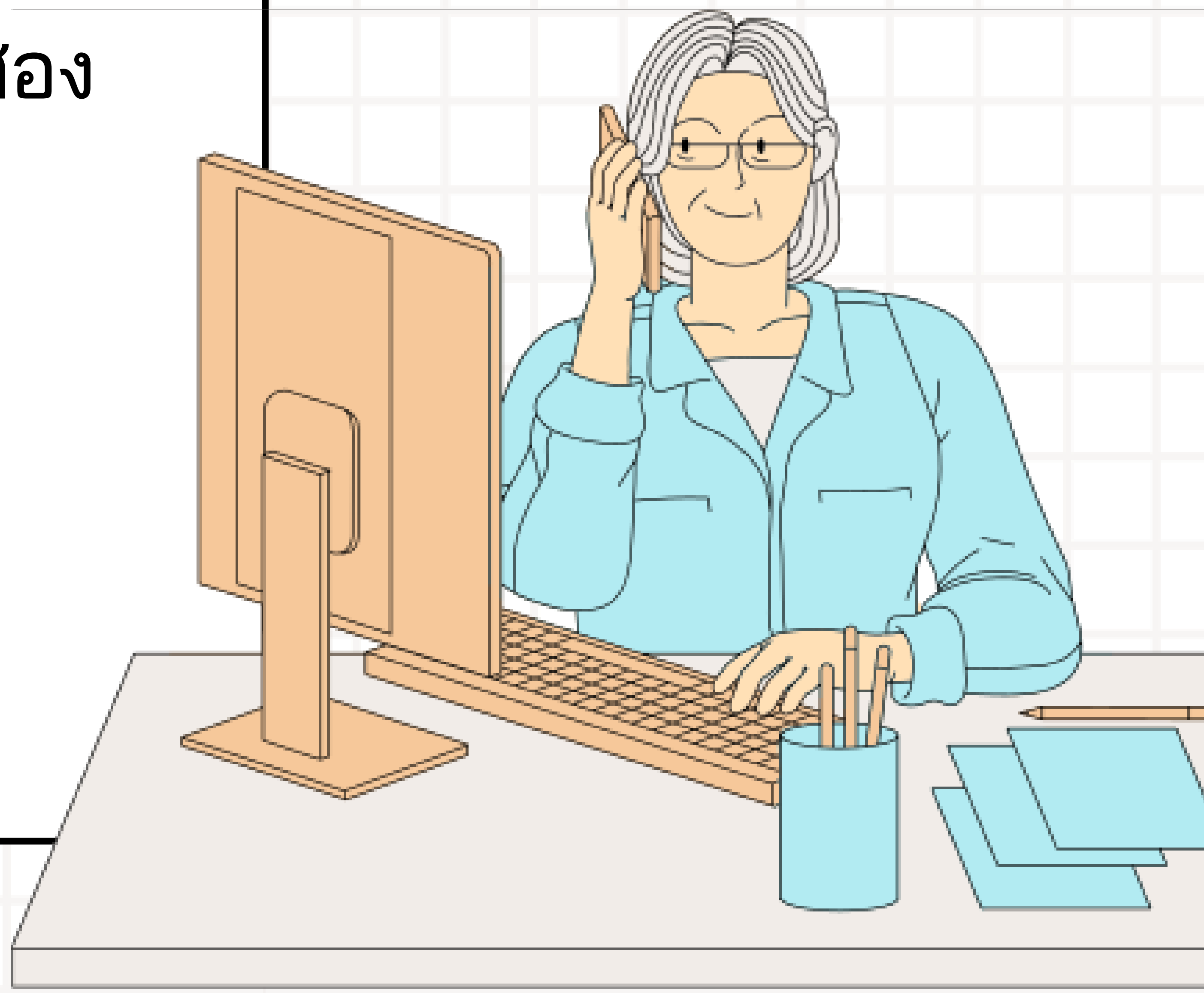
สมมุติฐาน คือ ข้อสมมุติเกี่ยวกับค่าพารามิเตอร์ของประชากร ที่กำหนดขึ้นเพื่อใช้ในการทดสอบว่าสิ่งที่เราพิจารณาอยู่นั้นจะเป็นไปตามที่คาดไว้หรือไม่



การทดสอบสมมติฐาน

สมมติฐานทางสถิติ (statistical hypothesis)

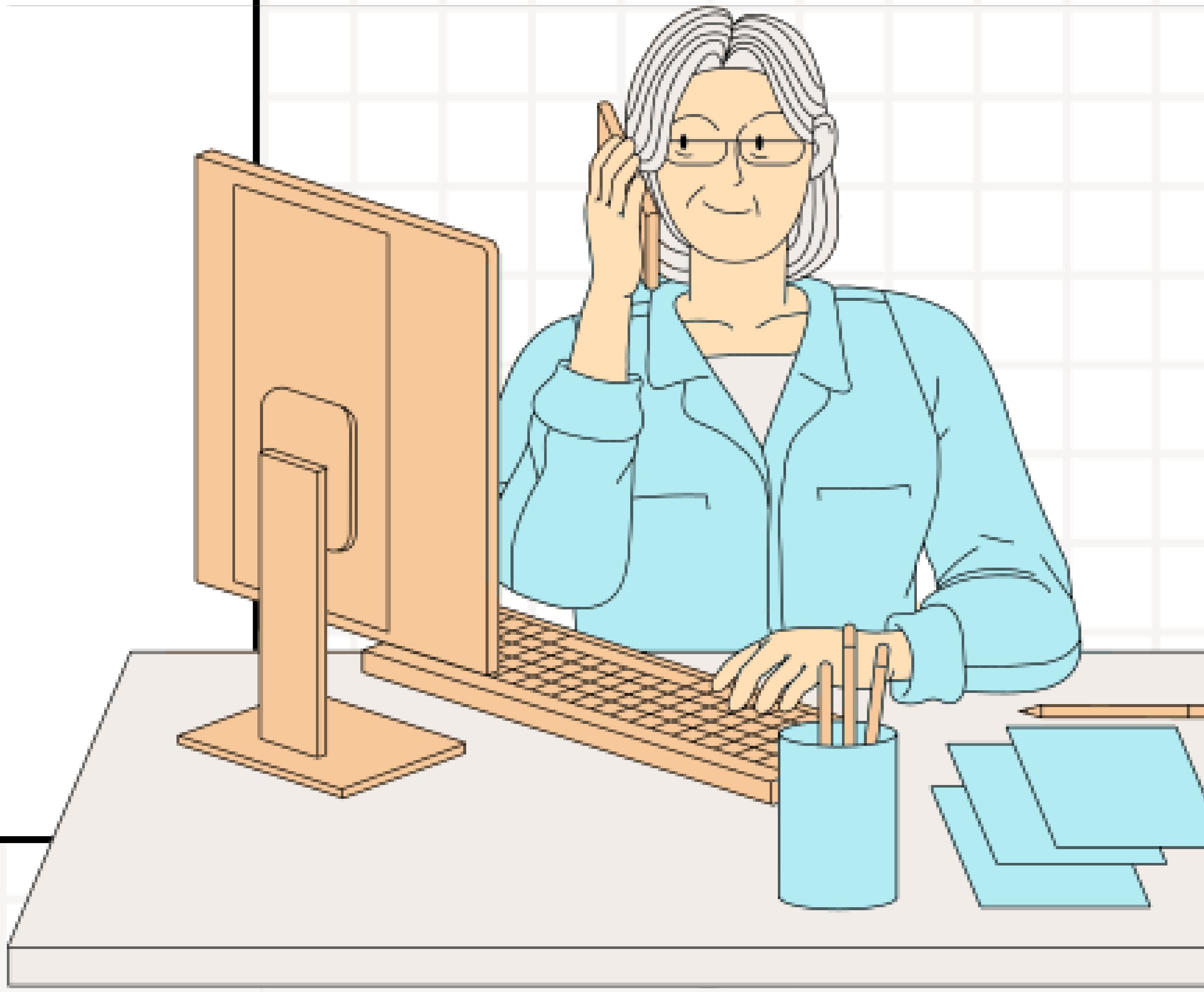
เป็นสมมุติฐานที่เขียนอธิบายคำตอบใน รูปโครงสร้างทางคณิตศาสตร์ที่คาดคะเนถึงความสัมพันธ์ระหว่างตัวแปรสองตัวหรือมากกว่าโดยเขียนในรูปลักษณะการแจกแจงหรือค่าพารามิเตอร์ของประชากรที่ต้องการทดสอบด้วยวิธีการทางสถิติ ซึ่งสมมติฐานทางสถิติประกอบด้วยส่วนประกอบที่สำคัญ 2 ส่วน ดังนี้



การทดสอบสมมติฐาน

โดยทั่วไปจะศึกษา 3 เรื่อง

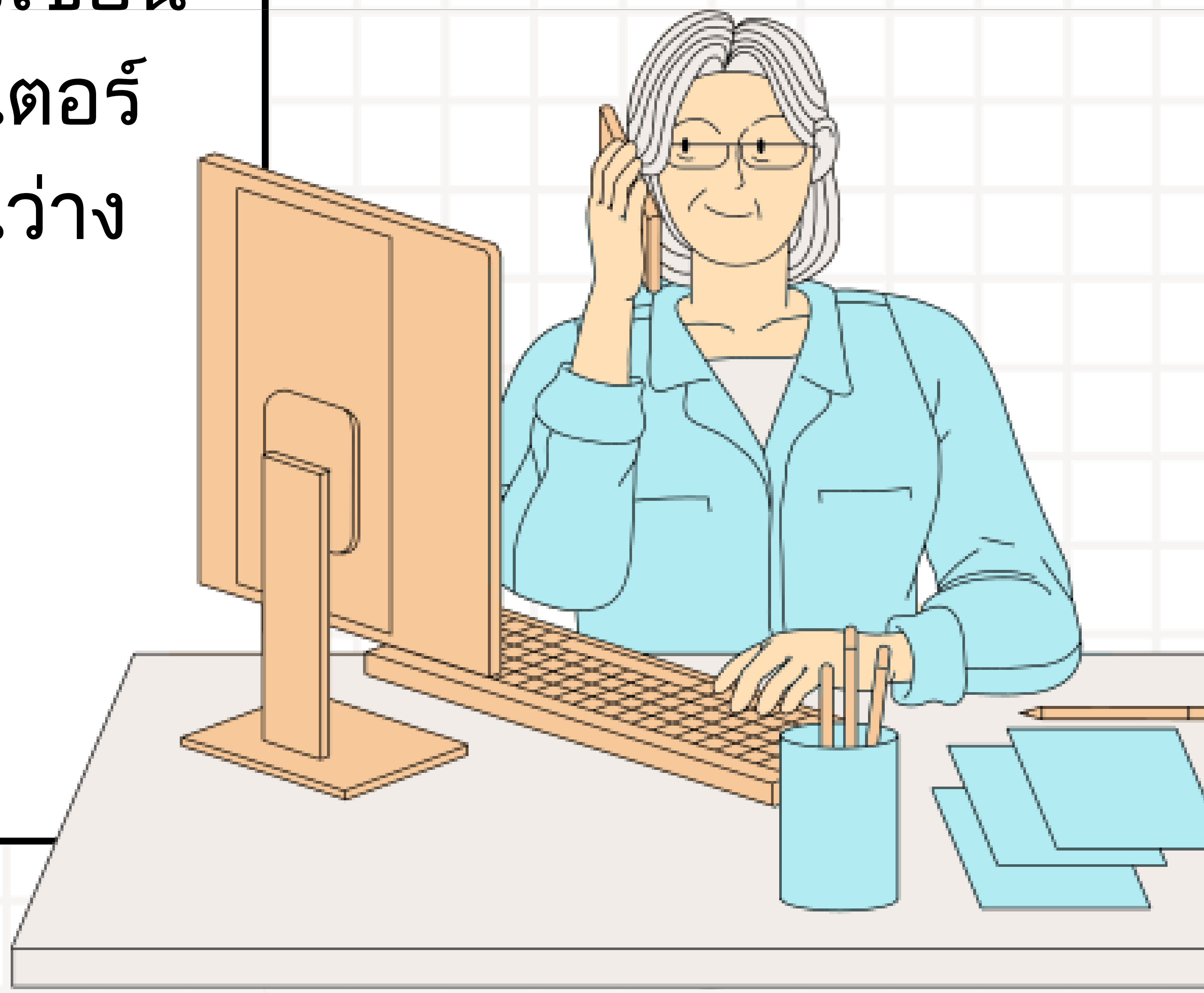
1. ค่าเฉลี่ยประชากร 1 กลุ่ม
2. ค่าเฉลี่ยประชากร 2 กลุ่ม
3. ค่าสัดส่วนประชากร 1 กลุ่ม



การทดสอบสมมติฐาน

1) สมมติฐานว่าง (null hypothesis) หรือสมมติฐานหลัก เป็นสมมติฐานทางสถิติที่ตั้งเอาไว้เพื่อการทดสอบ ซึ่งเขียนไว้ในลักษณะที่ไม่แสดงความแตกต่างระหว่างค่าพารามิเตอร์ที่ต้องการทดสอบ นิยมใช้สัญลักษณ์ H_0 แทนสมมติฐานว่าง

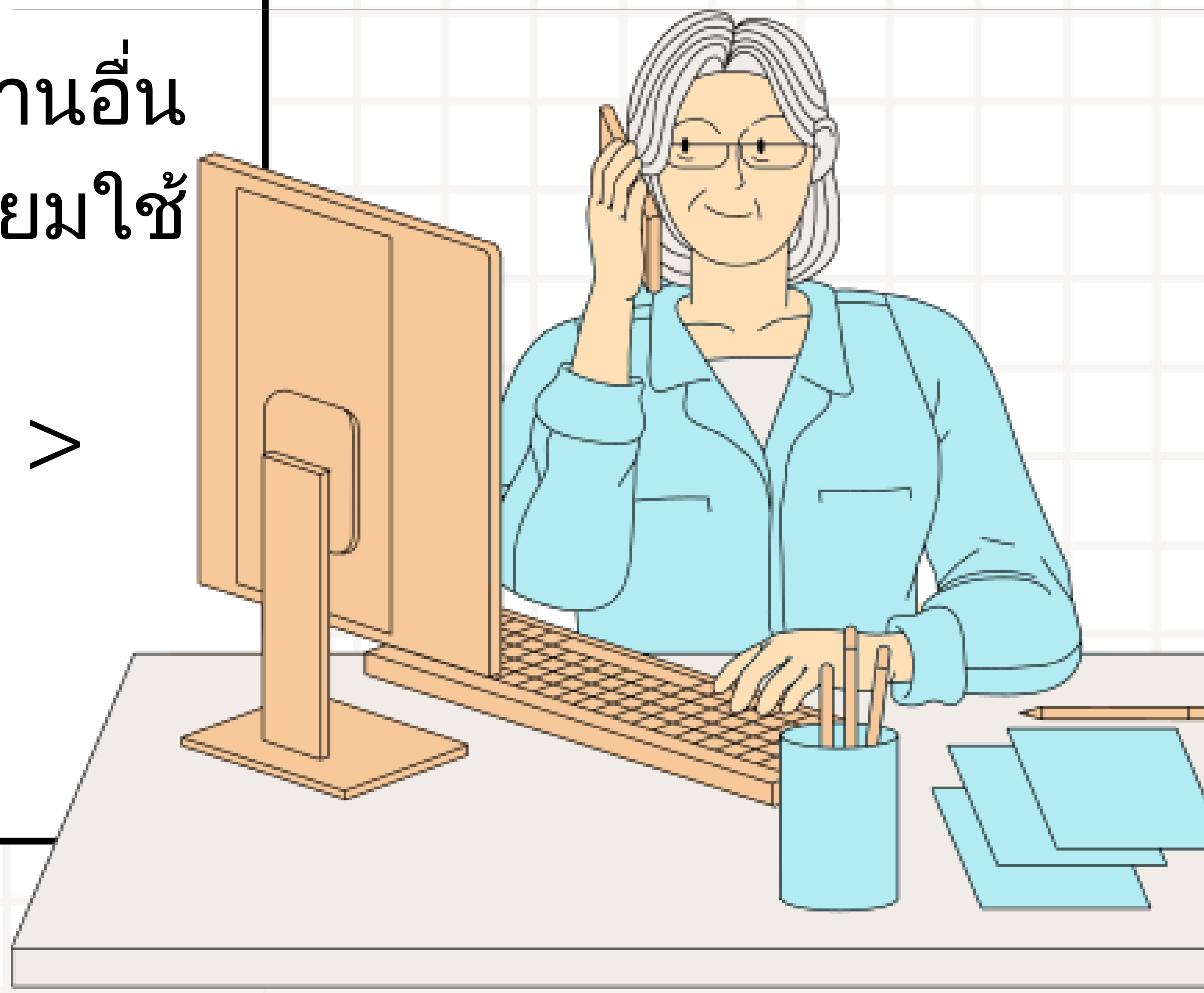
- สิ่งที่เขียนไว้มีเครื่องหมาย = , $\leq$ หรือ $\geq$



การทดสอบสมมติฐาน

2) สมมติฐานทางเลือกอื่น (alternative hypothesis) หรือสมมติฐานแย้ง เป็นสมมติฐานทางสถิติที่ตรงข้ามกับสมมติฐานว่างที่ต้องการทดสอบ โดยทั่วไปสมมติฐานอื่นมักจะเป็นสมมติฐานที่คาดว่าจะเป็นผลของการวิจัย นิยมใช้สัญลักษณ์ H_1 หรือ H_a แทนสมมติฐานอื่น

- จะมีเครื่องหมายตรงข้ามกับสมมติฐานหลัก $\neq, <$ หรือ $>$



การทดสอบสมมติฐาน

H_0 : ……… = …………

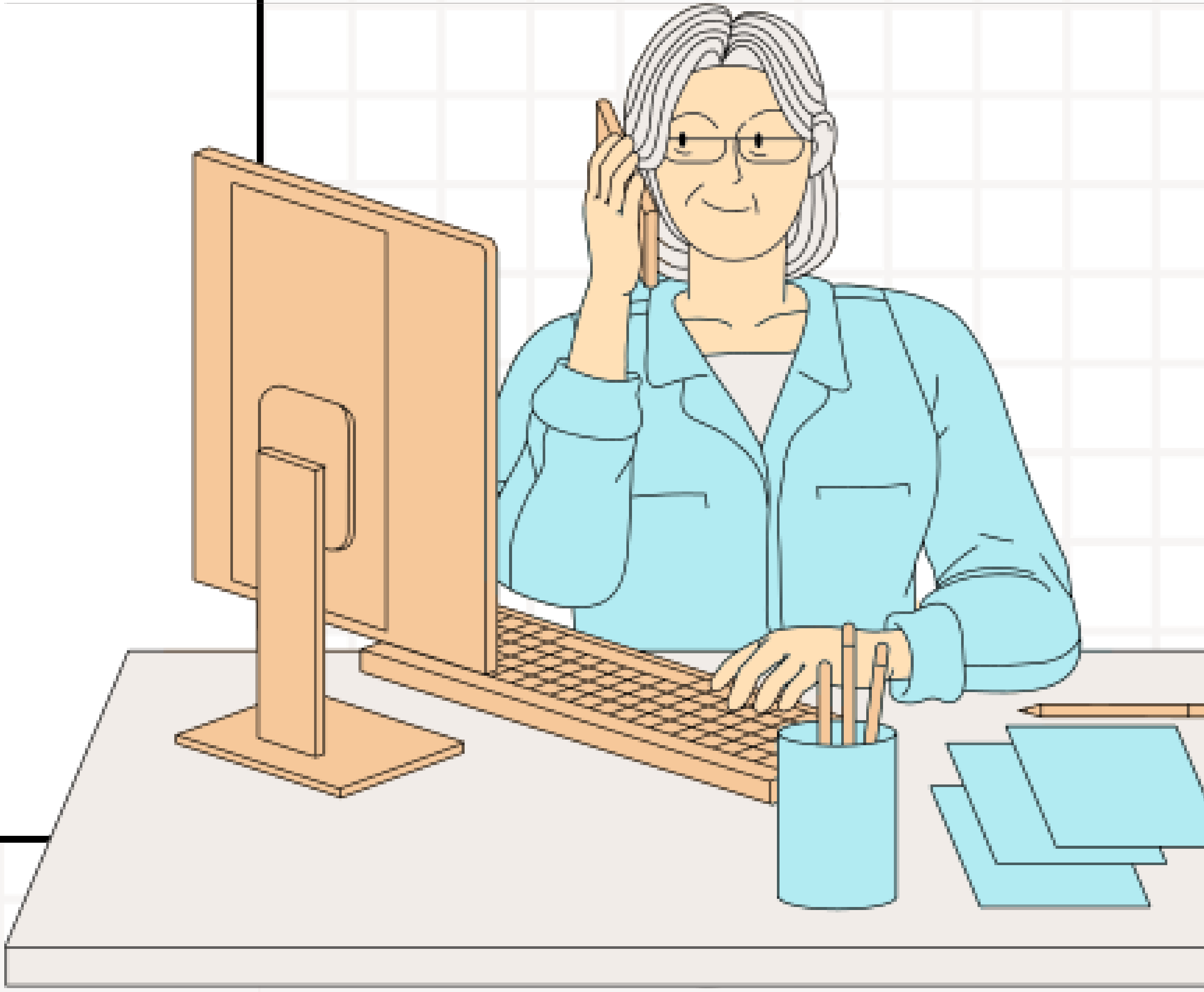
H_1 : ……… $\neq$ …………

H_0 : ……… $\leq$ …………

H_1 : ……… $>$ …………

H_0 : ……… $\geq$ …………

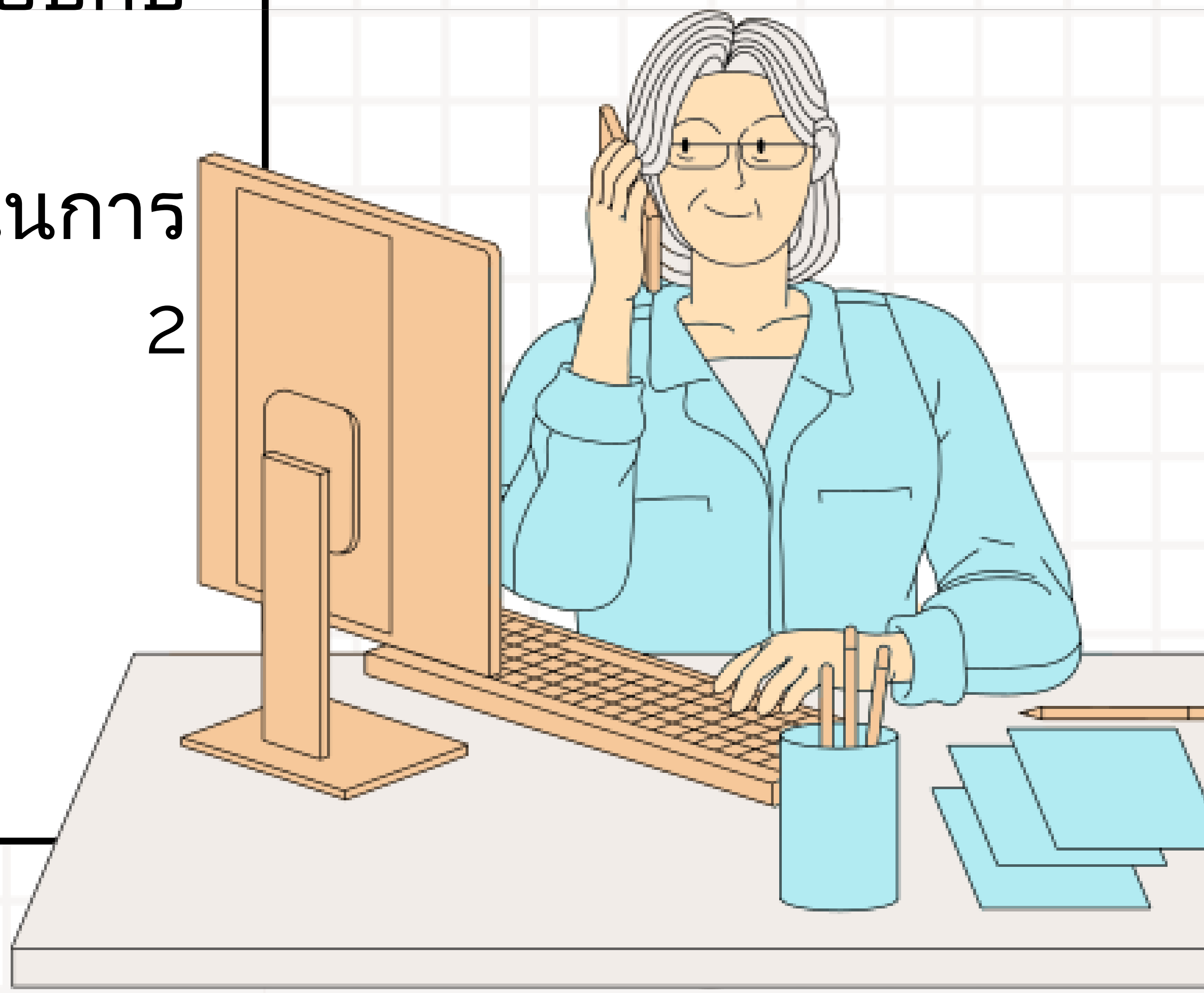
H_1 : ……… $<$ …………



การทดสอบสมมติฐาน

เป็นการตัดสินใจหรือประเมินความสำคัญของค่าสถิติจากกลุ่มตัวอย่างที่ได้มาโดยการสุ่มตัวอย่าง ด้วยการเปรียบเทียบกับสมมติฐานทางสถิติที่กำหนดไว้ล่วงหน้าโดยอาศัยเกณฑ์ (criteria) ที่ตั้งไว้เพื่อลดโอกาสของความคลาดเคลื่อนในการตัดสินใจ ผลที่ได้จากการทดสอบสมมติฐานทางสถิติมี 2 ลักษณะ ดังนี้

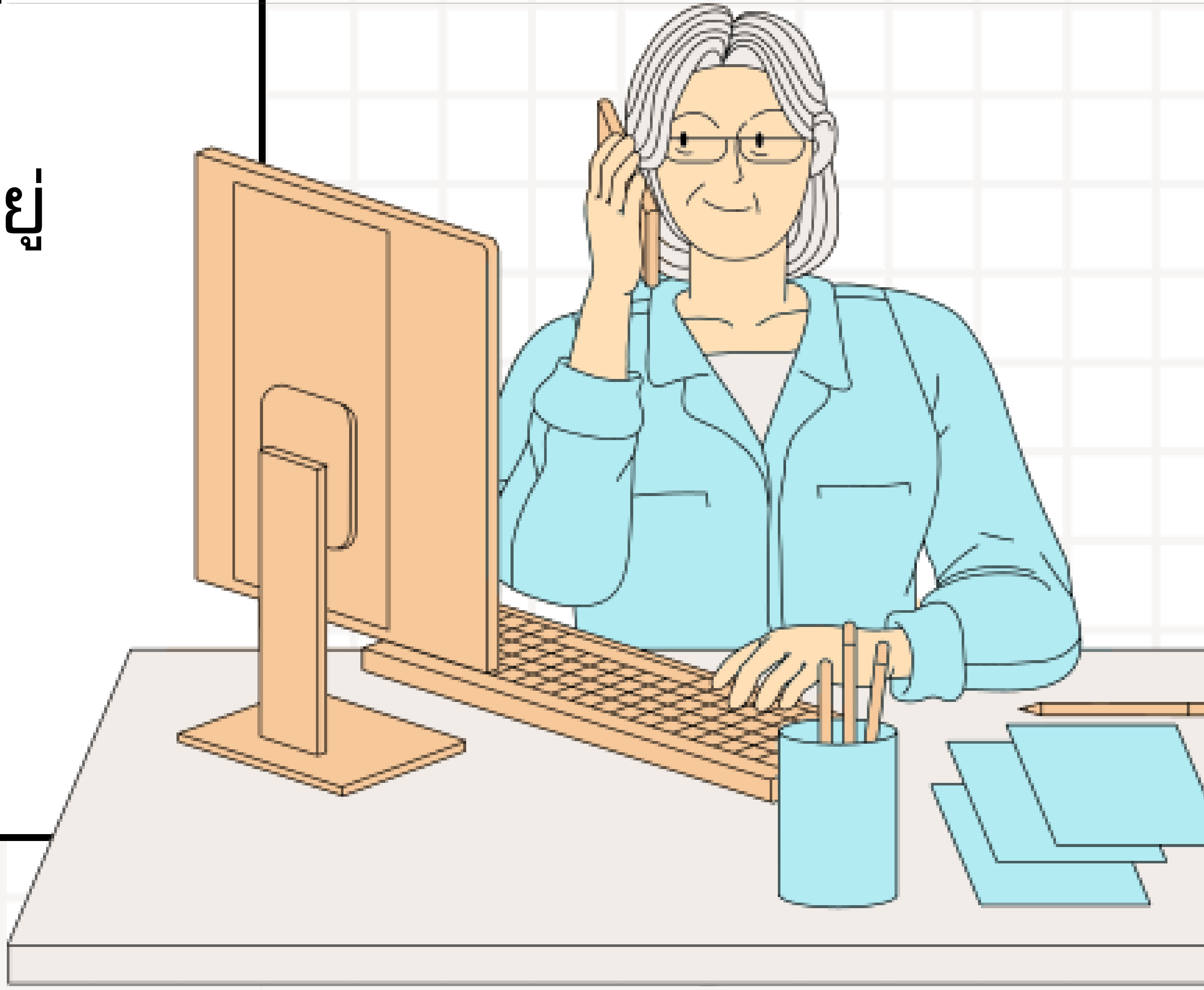
-



การทดสอบสมมติฐาน

การยอมรับสมมติฐาน (accept hypothesis or nosignificant) หมายถึง ความแตกต่างของค่าสถิติที่คำนวณได้จากกลุ่มตัวอย่างกับค่าพารามิเตอร์ ตามสมมติฐานว่างมีเพียงเล็กน้อย และความแตกต่างนั้นอยู่ภายในขอบเขตที่ยอมรับได้

•



การทดสอบสมมติฐาน

การปฏิเสธสมมติฐาน (reject hypothesis or significant) หมายถึง ความแตกต่างของค่าสถิติที่คำนวณได้จากกลุ่มตัวอย่างประชากรกับค่าพารามิเตอร์ ตามสมมติฐานว่างมีจำนวนมากและความแตกต่างนั้นมากจนเกินขอบเขตที่จะยอมรับได้

ประเภทของการทดสอบสมมติฐาน

1. การทดสอบสมมติฐานสองทาง (Two-Sided Tast)

- สมมติฐาน H_1 มีเครื่องหมาย $\neq$
- เขตการปฏิเสธการทดสอบจะมีสองด้าน

2. การทดสอบทางเดียว (One-Sided Tast)

การทดสอบทางขวา (A Right-Tailed Test)

- สมมติฐาน H_1 มีเครื่องหมาย $>$
- เขตปฏิเสธการทดสอบจะอยู่ทางด้านขวา

การทดสอบทางซ้าย (A Left-Tailed Test)

- สมมติฐาน H_1 มีเครื่องหมาย $<$
- เขตปฏิเสธการทดสอบจะอยู่ทางด้านซ้าย

H_0 : .... = ...
H_1 : .... $\neq$ ....

H_0 : .... $\leq$ ....
H_1 : .... $>$ ....

H_0 : .... $\geq$ ....
H_1 : .... $<$ ....

การทดสอบสมมติฐานทางสถิติ

จะเกิดสมมติฐานได้ 3 คู่

$$
\begin{array}{ll|c|c}
H_0 : & = & \geq & \leq \\
H_1 : & \neq & < & >
\end{array}
$$

สองทาง ทางเดียวทางซ้าย ทางเดียวทางขวา

สมมติฐานทางสถิติ

วิธีการตั้งสมมติฐาน

ตัวอย่าง 1 นักวิชาการเชื่อว่าเด็กไทยเฉลี่ยเล่นอินเตอร์เน็ต 4.5 ชั่วโมงต่อวัน

$H_0 : \mu = 4.5$

$H_1 : \mu \neq 4.5$

การทดสอบแบบสองทาง

สมมติฐานทางสถิติ

วิธีการตั้งสมมติฐาน

ตัวอย่าง 2 สัตว์แพทย์คาดว่าสุนัขจะใช้เวลาฟื้นตัวหลังผ่าตัดเฉลี่ยอย่างน้อย   5ชม

$H_0 : \mu \geq 5$

$H_1 : \mu < 5$

การทดสอบแบบทางเดียว
แบบทางซ้าย

สมมติฐานทางสถิติ

วิธีการตั้งสมมติฐาน

ตัวอย่าง 3 แม่ค้าคิดว่าราคาไข่เฉลี่ยเพิ่มขึ้นไม่เกิน 2.50 บาทต่อฟอง

$H_0 : \mu \leq 2.50$

$H_1 : \mu > 2.50$

การทดสอบแบบทางเดียว
แบบทางขวา

สมมติฐานทางสถิติ

$\mu < 4$

เครื่องหมาย น้อยกว่า ต้องใส่ H1

วิธีการตั้งสมมติฐาน

ตัวอย่าง 4 โดยปกติยาแก้ปวดจะมีฤทธิ์เฉลี่ยน้อยกว่า 4 ชม

$H_0 ; \mu \geq 4$

$H_1 : \mu < 4$

การทดสอบแบบทางเดียว
แบบทางซ้าย

สมมติฐานทางสถิติ

$\mu > 6500$

วิธีการตั้งสมมติฐาน

ตัวอย่าง 5 รายได้เฉลี่ยต่อเดือนของพนักงานทำความสะอาดมากกว่า 6500บาทต่อเดือน

$$H_0 : \mu \leq 6500$$

$$H_1 : \mu > 6500$$

การทดสอบแบบทางเดียว
แบบทางขวา

สมมติฐานทางสถิติ

วิธีการตั้งสมมติฐาน

$\mu_1 - \mu_2 = 50$

ตัวอย่าง 6 นักเรียนชายและนักเรียนหญิงใช้เวลาทบทวนหนังสือก่อนสอบต่างกัน 50 นาที

$H_0 : \mu_1 - \mu_2 = 50$

$H_1 : \mu_1 - \mu_2 \neq 50$

การทดสอบแบบสองทาง

สมมติฐานทางสถิติ

วิธีการตั้งสมมติฐาน

$p =$ สัดส่วนประชากร

ตัวอย่าง 7 ประชาชนในเขตสวนอ้อย 45% เป็นคนกรุงเทพโดยกำเนิด

$H_0 : p = 0.45$

$H_1 : p \neq 0.45$

การทดสอบแบบสองทาง

ความผิดพลาดในการทดสอบสมมติฐาน

โดยส่วนใหญ่จะแบ่งเป็น **2** ประเภท

1. ความผิดพลาดประเภทที่ **1 (Type I Error)**

- จะไม่ยอมรับ H_0 เมื่อ H_0 เป็นจริง
- ความน่าจะเป็นจะมีค่าเท่ากับ α หรือ ระดับนัยสำคัญ

ความผิดพลาดในการทดสอบสมมติฐาน

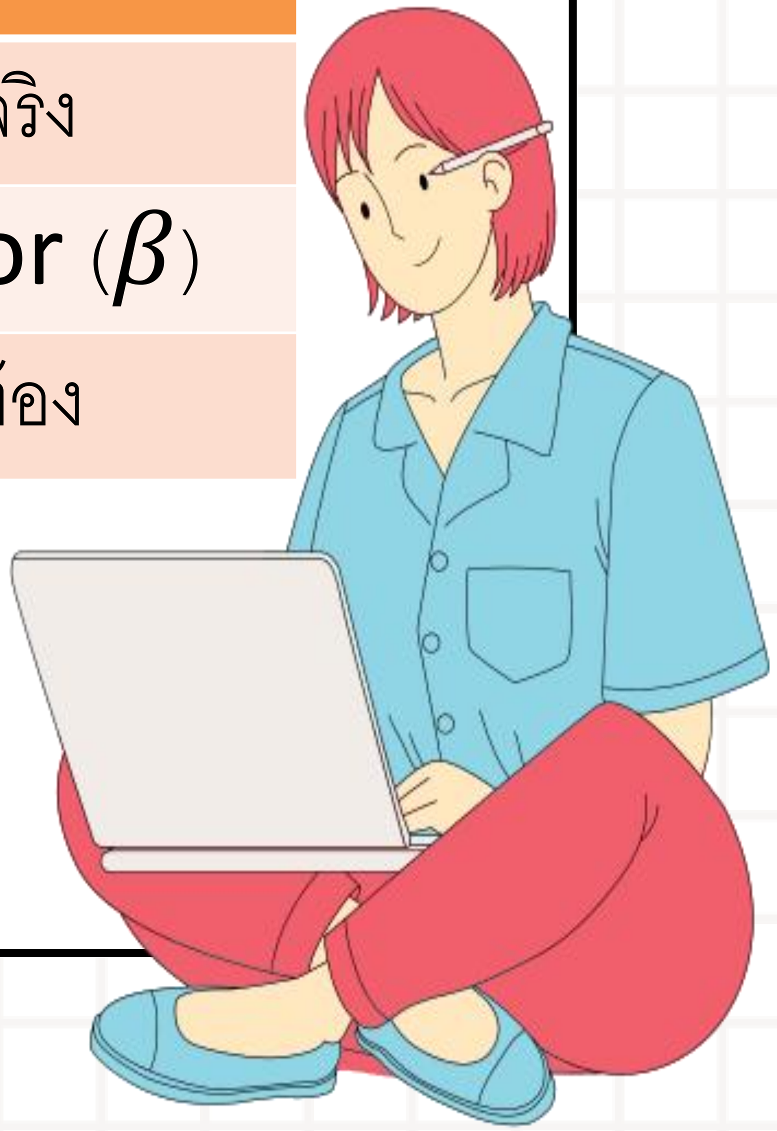
โดยส่วนใหญ่จะแบ่งเป็น **2** ประเภท

2. ความผิดพลาดประเภทที่ **2 (Type II Error)**

- จะยอมรับ H_0 เมื่อ H_0 ไม่จริง
- ความน่าจะเป็นจะมีค่าเท่ากับ β

ความผิดพลาดในการทดสอบสมมติฐาน

ผลการทดสอบ	**ความเป็นจริง**	**ความเป็นจริง**
	H_0 เป็นจริง	H_0 ไม่เป็นจริง
ยอดรับ H_0	ทดสอบถูกต้อง	Type II Error (β)
ปฏิเสธ H_0	Type I Error (α)	ทดสอบถูกต้อง



ขั้นตอนการทดสอบสมมติฐาน

1. ตั้งสมมติฐานหลักและสมมติฐานรอง
2. กำหนดตัวสถิติทดสอบเป็น **Z** หรือ **t** (ดูจากสูตรในแต่ละกรณีของเรื่อง)
3. หาค่าวิกฤติ
   - ค่าวิกฤตคือค่า $Z_{\propto}$, $Z_{\propto/2}$, $t_{\propto}$ หรือ $t_{\propto/2}$
   - ซึ่งค่าวิกฤติจะขึ้นอยู่กับ สถิติทดสอบในข้อ **2** ,ระดับนัยสำคัญ (α)
   - ประเภทของข้อมูลสมมติฐาน

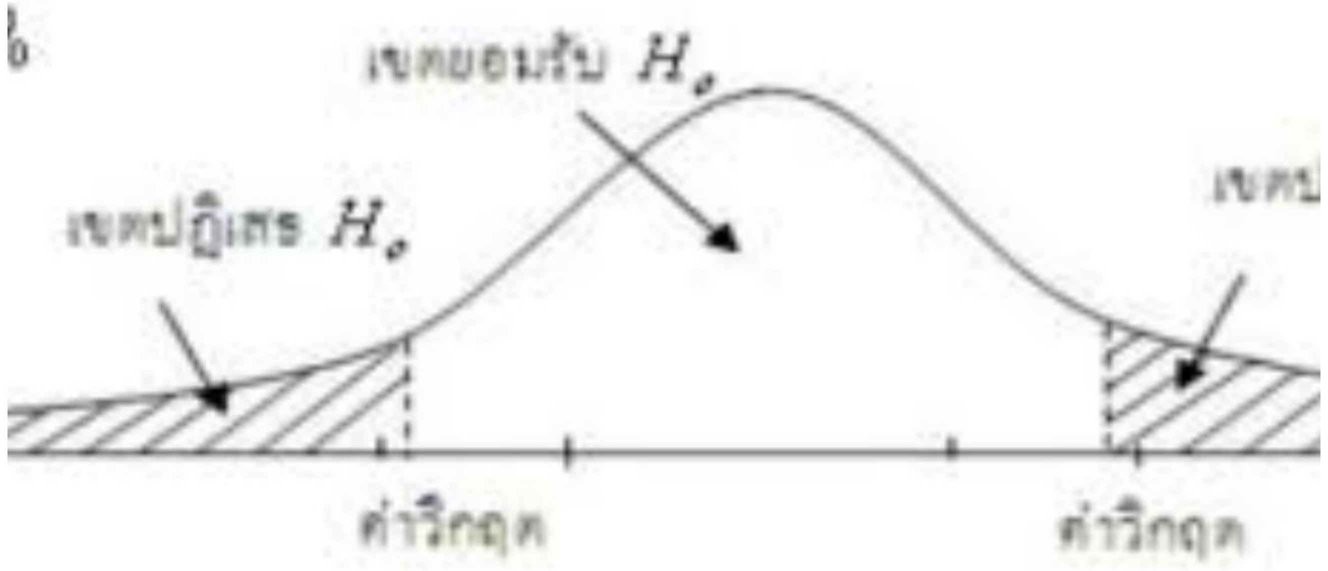
ขั้นตอนการทดสอบสมมติฐาน

		ค่าวิกฤติ	
		การทดสอบสองทาง	การทดสอบทางเดียว
สถิติทดสอบ	**Z**	$Z_{\propto/2}$	$Z_{\propto}$
ข้อ 2	**t**	$t_{\propto/2}$	$t_{\propto}$

ขั้นตอนการทดสอบสมมติฐาน

4. กฎการตัดสินใจ

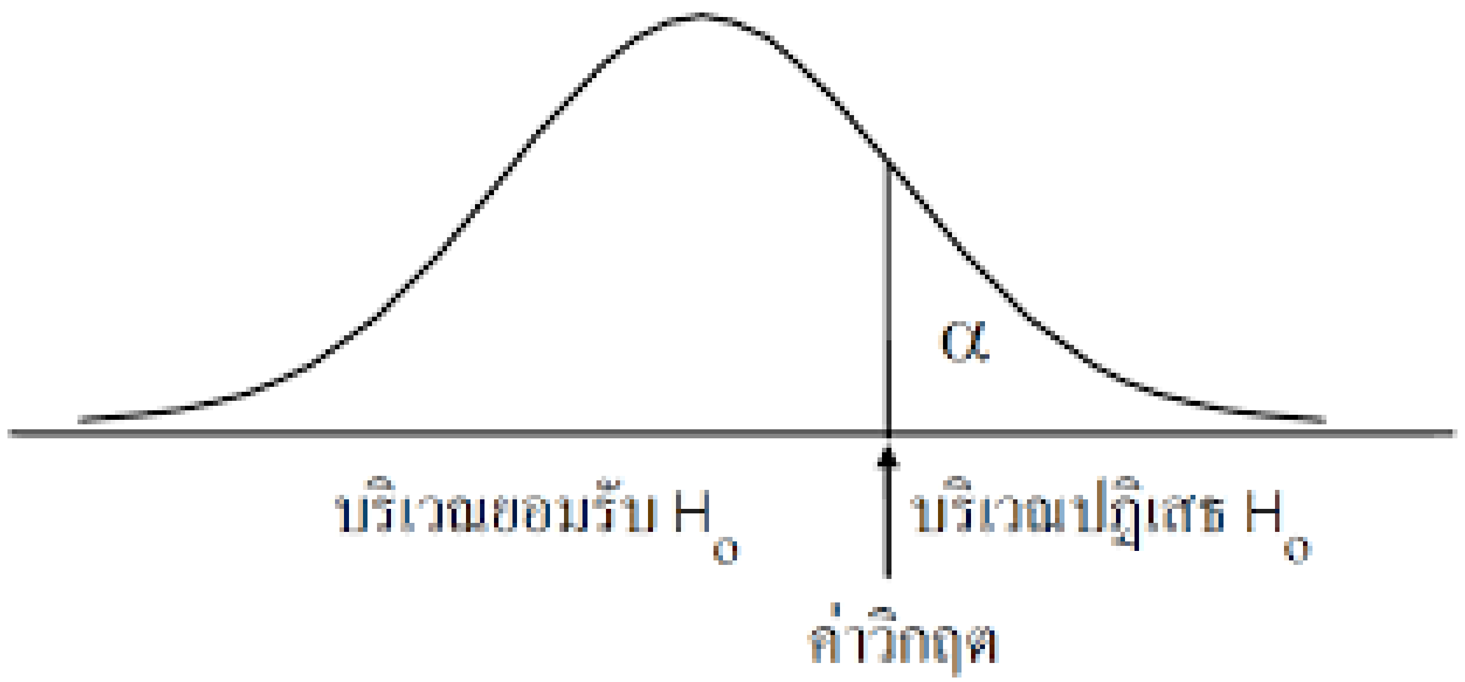
กฎการตัดสินใจสำหรับการทดสอบสองทาง



ขั้นตอนการทดสอบสมมติฐาน

4. กฎการตัดสินใจ

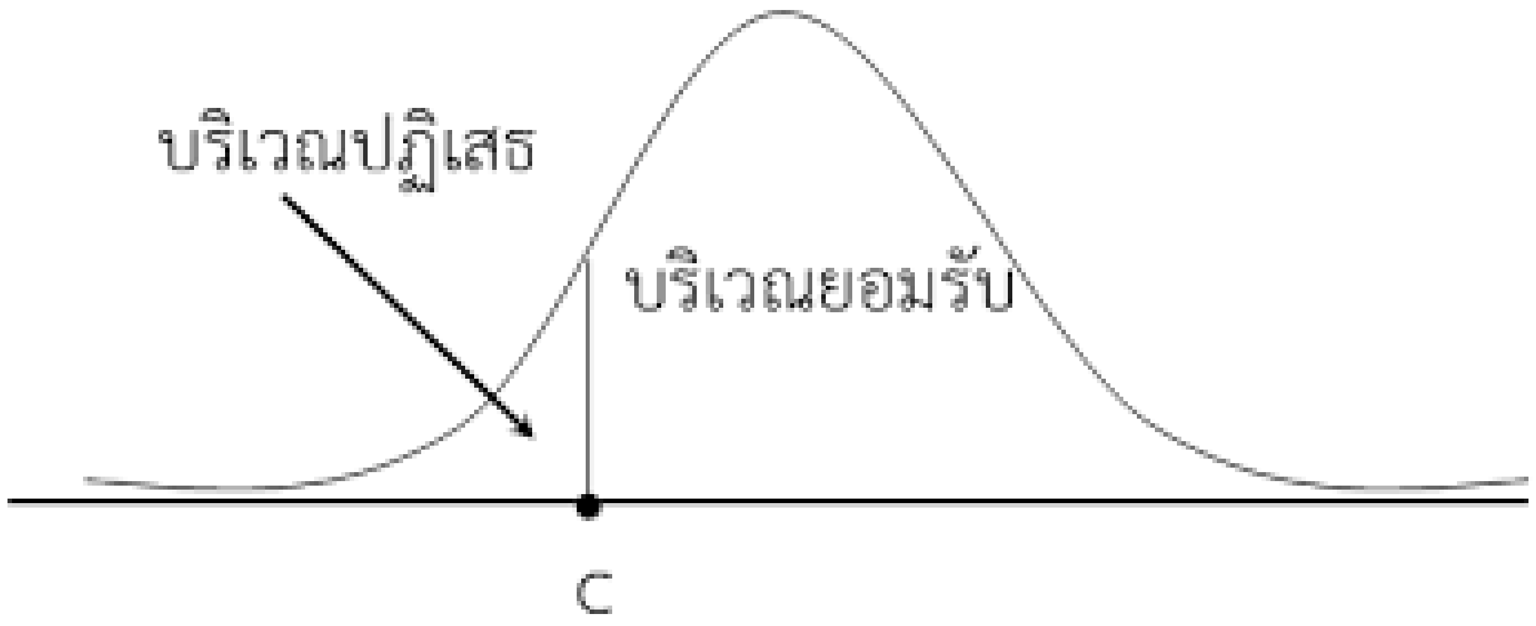
กฎการตัดสินใจสำหรับการทดสอบทางเดียวแบบทางขวา



ขั้นตอนการทดสอบสมมติฐาน

4. กฎการตัดสินใจ

กฎการตัดสินใจสำหรับการทดสอบทางเดียวแบบทางซ้าย



ขั้นตอนการทดสอบสมมติฐาน

5. คำนวณค่าสถิติทดสอบในสมการ **(ขั้น2)**

6. สรุปผล (ปฏิเสธ หรือ ยอมรับ)

การทดสอบสมมติฐาน

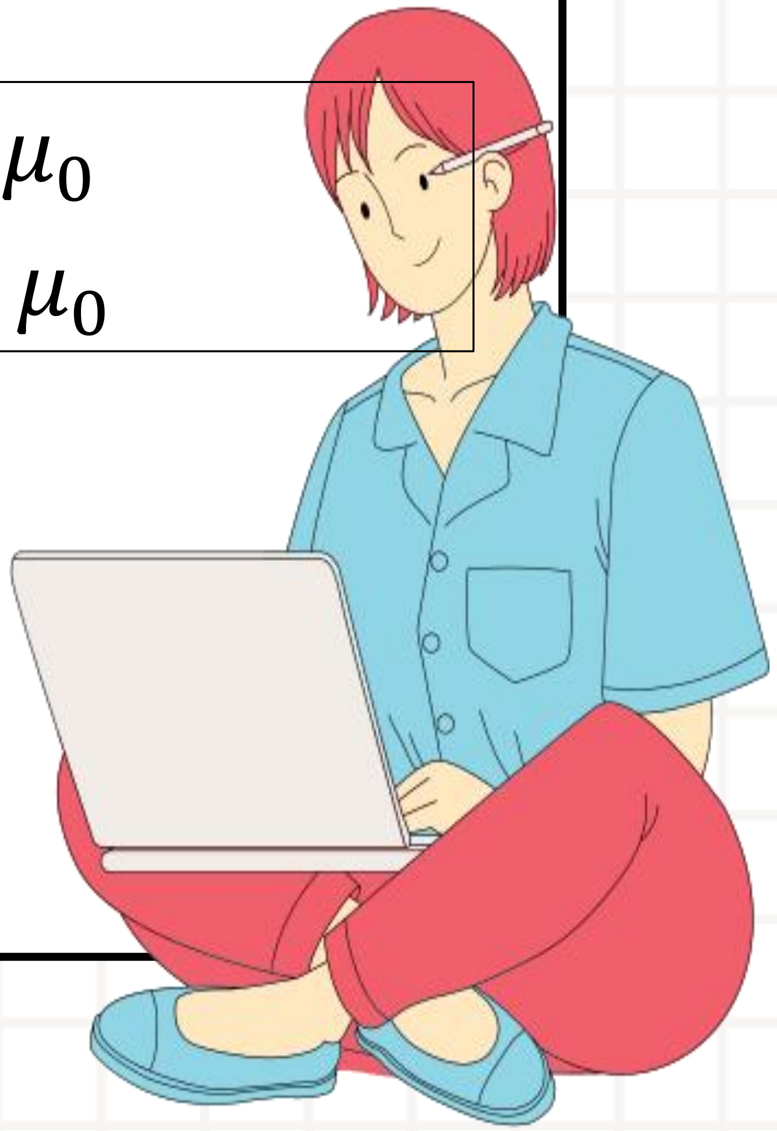
การทดสอบสมมติฐานของค่าเฉลี่ยประชากรหนึ่งกลุ่ม

สมมติฐานที่ตั้ง

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

$$H_0 : \mu. \geq \mu_0$$
$$H_1 : \mu. < \mu_0$$



ขั้นตอนการทดสอบสมมติฐาน

สถิติทดสอบ

① กรณีทราบความแปรปรวนของประชากร(σ^2)

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

โจทย์ให้คำว่าส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) มาก่อนหน้าคำว่า 'สุ่ม'

② กรณีไม่ทราบความแปรปรวนของประชากร(σ^2), $n \geq 30$

$$Z = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

โจทย์ให้คำว่าส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) มาหลังคำว่า 'สุ่ม' และ $n \geq 30$

③ กรณีไม่ทราบความแปรปรวนของประชากร(σ^2), $n < 30$

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}}, \quad df = n - 1$$

โจทย์ให้คำว่าส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) มาหลังคำว่า 'สุ่ม' และ $n < 30$

การทดสอบสมมติฐาน

การทดสอบสมมติฐานของค่าเฉลี่ยประชากรหนึ่งกลุ่ม

ตัวอย่าง ร้านเบเกอรี่เชื่อว่าช็อคโกแลตที่ซื้อมาจากร้านค้าได้ค่ามาตราฐานในการนำไปผสมกับขนม จะมีน้ำหนักเฉลี่ยเท่ากับ **8.5** กรัมต่อแท่ง และค่าเบี่ยงเบนมาตราฐานเท่ากับ **0.72** กรัมต่อแท่ง จึงทำการสุ่มตัวอย่างช็อคโกแลตจำนวน **16** แท่ง พบว่ามีน้ำหนักเฉลี่ย **8.96** กรัมต่อแท่ง จงทดสอบสมมติฐานนี้ ที่ระดับความเชื่อมั่น **94%** ($Z_{\propto/2}$ = **1.88**)

การทดสอบสมมติฐาน

วิธีทำ

n = 16 , $\bar{X}$ **= 8.96 ,** $\theta = $ **0.72**

1. ตั้งสมมติฐาน

$H_0 : \mu = 8.5$

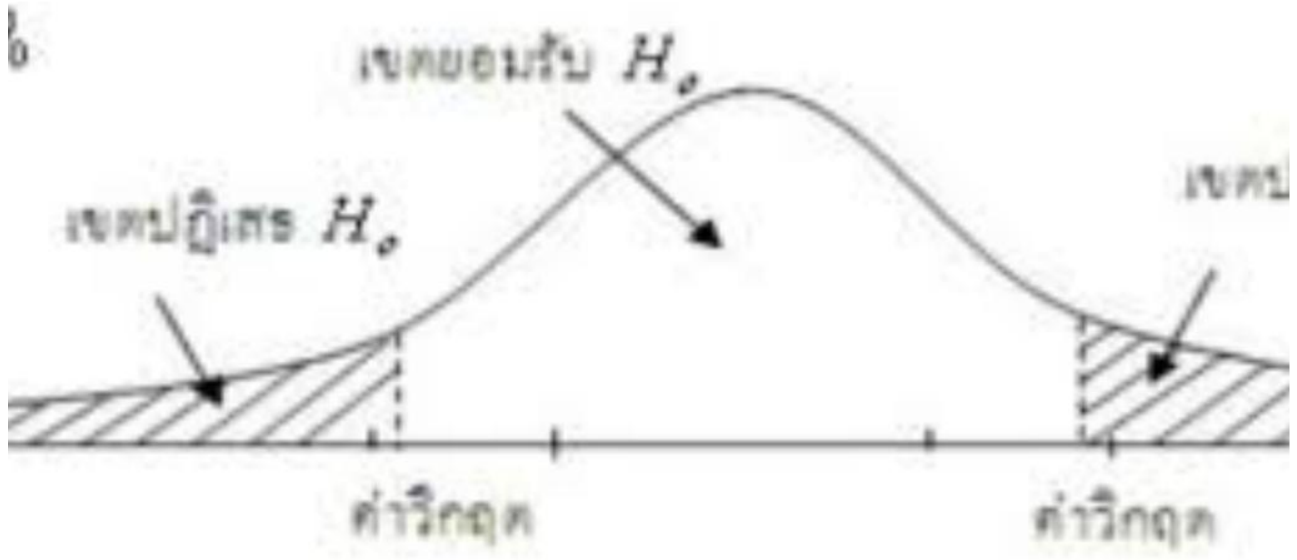
$H_1 : \mu \neq 8.5$

2. กำหนดสถิติทดสอบ **Z**

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

การทดสอบสมมติฐาน

3. หาค่าวิกฤต

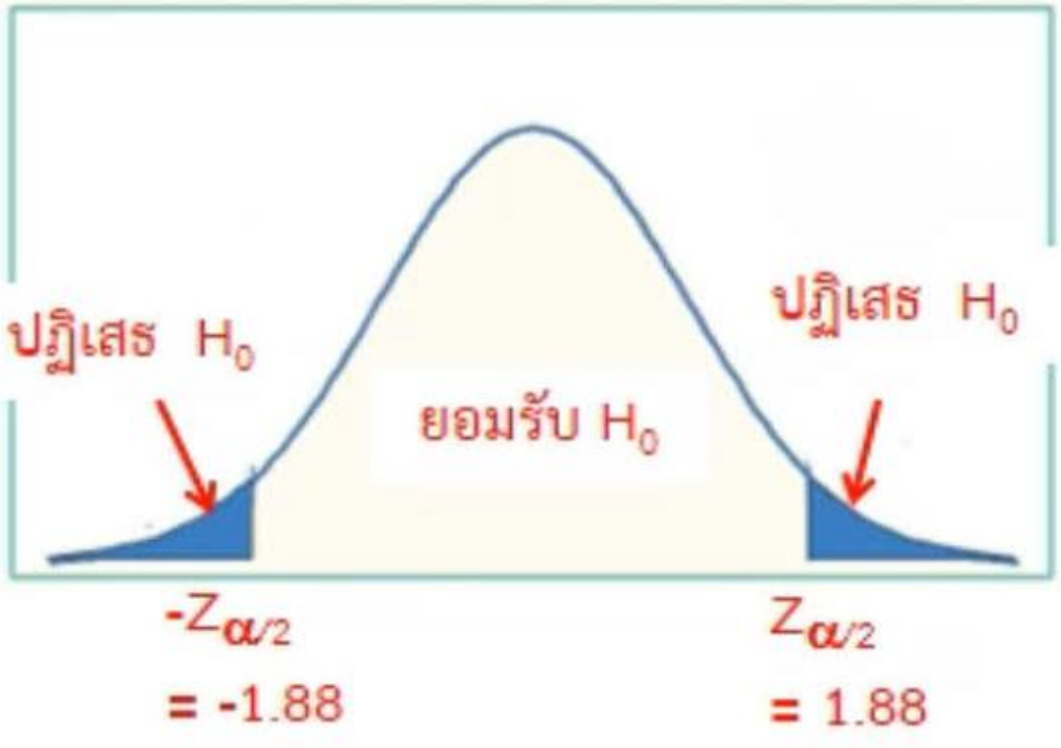


$-Z_{\propto/2}$ = -1.88 $Z_{\propto/2}$ = 1.88

การทดสอบสมมติฐาน

4. กำหนดกฎของหารตัดสินใจ

ปฏิเสธ H_0 เมื่อ $z < -1.88$ หรือ $Z > 1.88$



5. คำนวณค่าสถิติทดสอบ

$$Z = \frac{8.96 - 8.5}{0.72/\sqrt{16}}$$

$= 2.56$

6. สรุปผล

ปฏิเสธ H_0 น้ำหนักเฉลี่ยของช็อคโกแลตไม่เท่ากับ 8.5 กรัมต่อแท่ง แปลว่า ยอมรับ H_1

การทดสอบสมมติฐาน

ตัวอย่าง ผู้จัดการรีสอร์ท A กำหนดระดับความพึงพอใจเฉลี่ยของผู้ใช้บริการไว้ว่าจะต้องไม่ต่ำกว่า **4.20** คะแนน จึงต้องสำรวจความพึงพอใจของผู้ใช้บริการอย่างสม่ำเสมอ เมื่อสุ่มสำรวจความพึงพอใจของผู้ใช้บริการจำนวน **49** คน พบว่ากลุ่มตัวอย่างมีความพึงพอใจเฉลี่ยเท่ากับ **4.02** คะแนน ค่าเบี่ยงเบนมาตราฐาน **1.24** จงทดสอบว่าความคิดของผู้จัดการรีสอร์ท A เป็นจริงหรือไม่ ที่ระดับนัยสำคัญทางสถิติ **0.10** (Z_{α} = **1.28**)

② กรณีไม่ทราบความแปรปรวนของประชากร(σ^2), $n \geq 30$

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

โจทย์ให้คำว่าส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) มา<u>หลัง</u>คำว่า 'สุ่ม' และ $n \geq 30$

การทดสอบสมมติฐาน

วิธีทำ

n = 49 , $\bar{x}$ = 4.02 , S= 1.24

1. ตั้งสมมติฐาน (ทางเดียว)

H_0 ; $\mu \geq$ **4.20**

H_1 : $\mu <$ **4.20**

2. กำหนดสถิติทดสอบเป็น **Z**

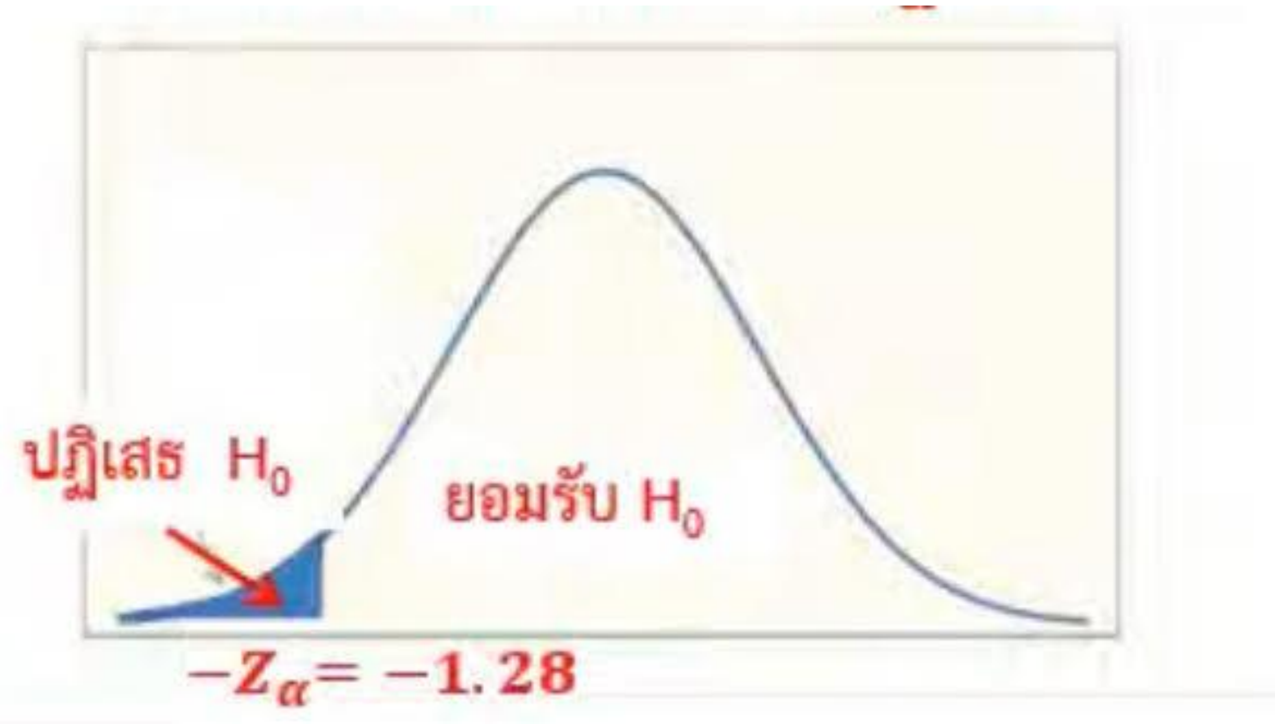
$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

การทดสอบสมมติฐาน

3. หาค่าวิกฤต

เป็นการทดสอบทางเดียว แบบทางซ้ายใช้ $Z_{\propto}$

โจทย์กำหนด $\propto$ = **0.10** จะได้ $Z_{\propto}$ = **1.28**



การทดสอบสมมติฐาน

4. กำหนดกฎของหารตัดสินใจ

จะปฏิเสธ H_0 เมื่อ $z < -1.28$

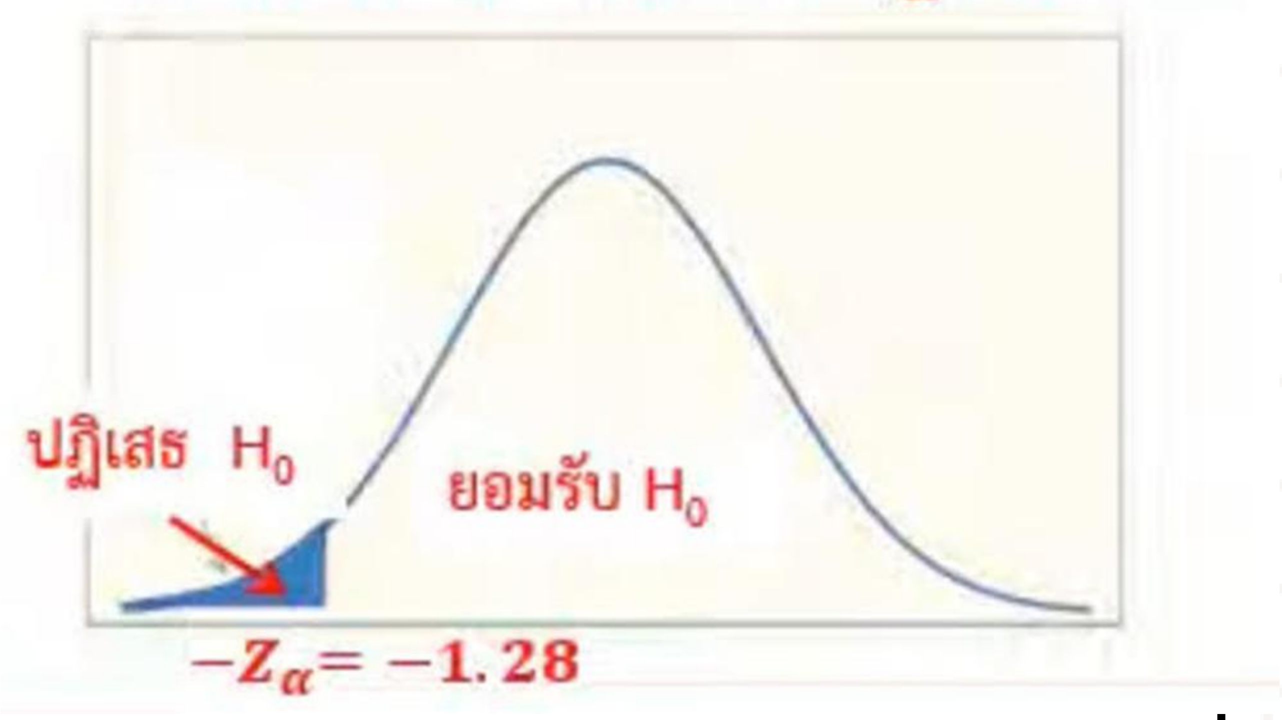
5. คำนวณค่าสถิติทดสอบ

$$Z = \frac{4.02 - 4.20}{1.24 / \sqrt{49}}$$

$$= -1.02$$

6. สรุปผล

ยอมรับ H_0 ความพึงพอใจของผู้ใช้บริการไม่ต่ำกว่า 4.20 ยอมรับ H_1



การทดสอบสมมติฐาน

ตัวอย่าง นักวิจัยเชื่อว่าจำนวนชั่วโมงในการนอนตอนกลางคืนมากกว่า **8** ชมจะสั่งผลต่อการเรียนให้มีประสิทธิภาพที่ดีขึ้น จึงทำการสุ่มตัวอย่างนึกศึกษา **25** คน จากมหาวิทยาลัย เอ พบว่านักศึกษาใช้เวลานอนโดยเฉลี่ย **6.5** ชม มีส่วนเบี่ยงเบนมาตราฐาน **2.5** ชม ให้ทดสอบว่าความเชื่อของนักวิจัยถูกต้องหรือไม่ที่ระดับนัยสำคัญทางสถิติ **0.01** ($t_{\propto} = 2.492$)

③ กรณีไม่ทราบความแปรปรวนของประชากร(σ^2), n < 30

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad ,df = n - 1$$

โจทย์ให้คำว่าส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) มาหลังคำว่า 'สุ่ม' และ n < 30

การทดสอบสมมติฐาน

วิธีทำ

n = **25** , $\bar{x}$ = **6.5** , **s= 2.5**

1. ตั้งสมมติฐาน (ทางเดียวขวา)

H_0 ; $\mu \leq$ **8**

H_1 : $\mu > 8$

2. กำหนดสถิติทดสอบเป็น **t**

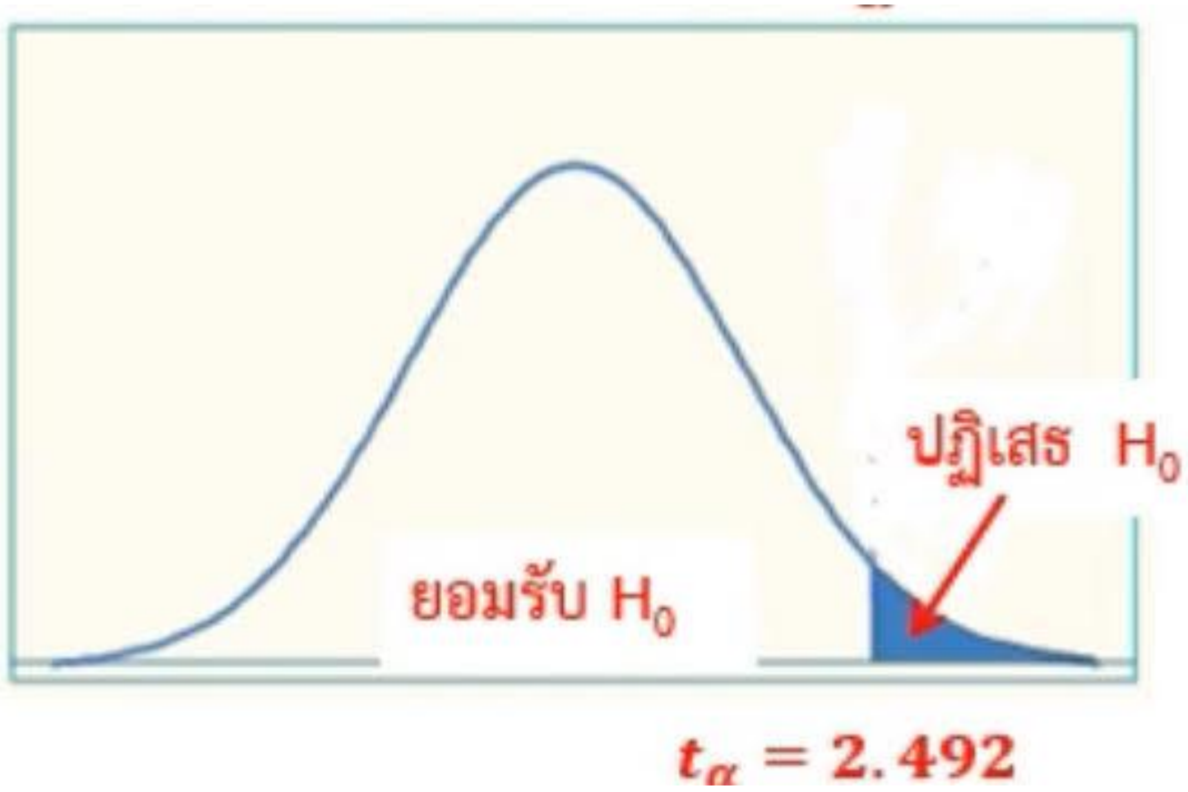
$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

การทดสอบสมมติฐาน

3. หาค่าวิกฤต

เป็นการทดสอบทางเดียว แบบทางขวาใช้ $t_{\propto}$

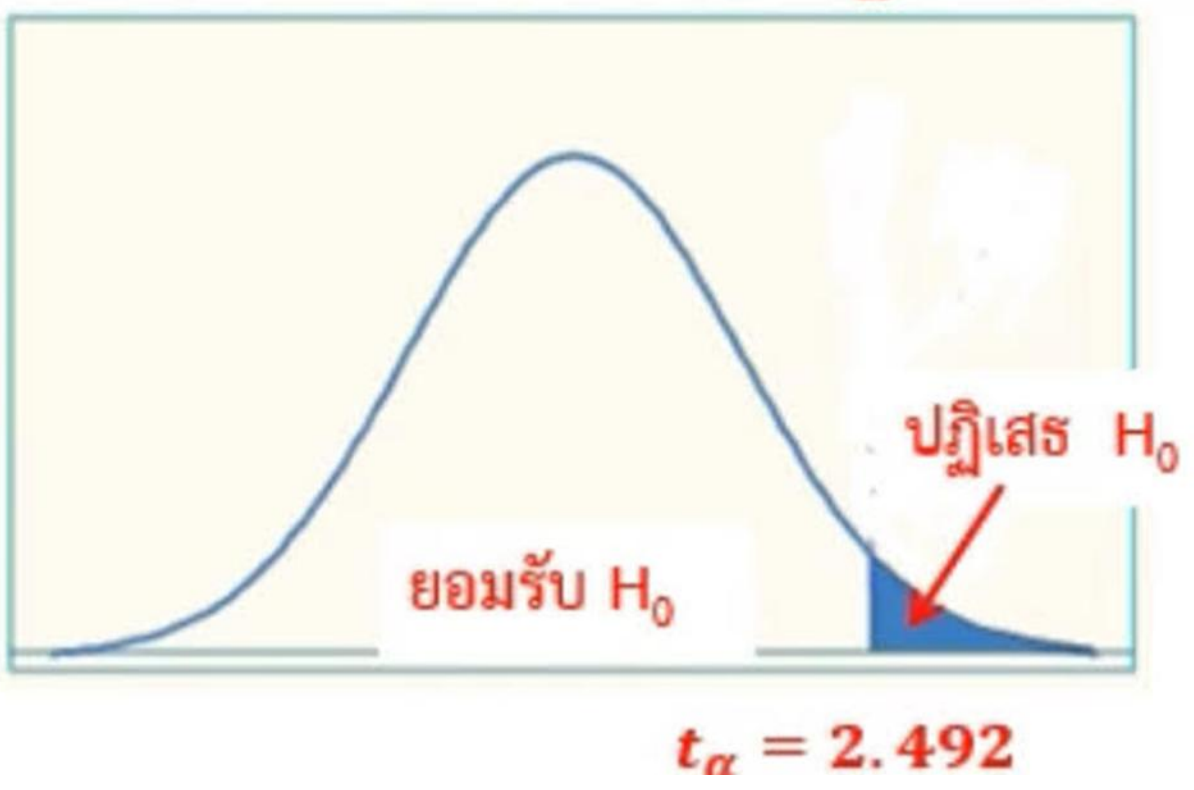
โจทย์กำหนด $\propto$ = **0.01** จะได้ $t_{\propto}$ = **2.492**



การทดสอบสมมติฐาน

4. กำหนดกฎของหารตัดสินใจ

จะปฏิเสธ H_0 เมื่อ **t** $>$ **2.492**



5. คำนวณค่าสถิติทดสอบ

$$t = \frac{6.5 - 8}{2.5/\sqrt{25}}$$

$$= -3.00$$

6. สรุปผล

ยอมรับ H_0 จำนวนชั่วโมงการนอนน้อยกว่าหรือเท่ากับ **8** ชั่วโมงจะส่งผลให้การเรียนมีประสิทธิภาพดีขึ้น

การทดสอบสมมติฐาน

การทดสอบสมมติฐานของผลต่างค่าเฉลี่ยของประชากรสองกลุ่ม

สมมติฐานที่ตั้ง

$H_0 : \mu_1 - \mu_2 = \mu_0$	$H_0 : \mu_1 - \mu_2 \le \mu_0$	$H_0 : \mu_1 - \mu_2 \ge \mu_0$
$H_1 : \mu_1 - \mu_2 \ne \mu_0$	$H_1 : \mu_1 - \mu_2 > \mu_0$	$H_1 : \mu_1 - \mu_2 < \mu_0$

สถิติทดสอบ

1. กรณีทราบค่าความแปรปรวนของประชากรทั้งสองกลุ่ม (σ_1^2, σ_2^2)

$$\boxed{Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

โจทย์ให้ค่าว่าส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) มาก่อนหน้าคำว่า 'สุ่ม'

2. กรณีไม่ทราบค่าความแปรปรวนของประชากร และขนาดตัวอย่างใหญ่ (n≥30)

$$\boxed{Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}}$$

โจทย์ให้ค่าว่าส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) มาหลังคำว่า 'สุ่ม' และ n≥30

การทดสอบสมมติฐาน

การทดสอบสมมติฐานของผลต่างค่าเฉลี่ยของประชากรสองกลุ่ม

ตัวอย่าง นักวิจัยเชื่อว่าน้ำหนักของเด็กแรกเกิดที่มีมารดาสูบบุหรี่เป็นประจำจะน้อยกว่าน้ำหนักของเด็กที่มารดาไม่สูบบุหรี่ จากเอกสารงานวิจัยที่ทำไว้แล้วพบว่าความแปรปรวนของน้ำหนักเด็กแรกเกิดที่มารดาสูบบุหรี่เป็นประจำกับมารดาที่ไม่สูบคือ **8** และ **4** ตามลำดับ นักวิจัยสุ่มตัวอย่างจากเด็กทั้งสองกลุ่มๆละ **75** คน พบว่า ค่าเฉลี่ยของน้ำหนักของเด็กกลุ่มที่มารดาสูบบุหรี่เป็นประจำดับมารดาที่ไม่สูบคือ **2.5** และ **3.3** กิโลกรัม จงทดสอบว่าความเชื่อของนักวิจัยถูกต้องหรือไม่ โยใช้ระดับนัยสำคัญที่ **0.02** $Z_{\propto} = 2.05$

การทดสอบสมมติฐาน

การทดสอบสมมติฐานของผลต่างค่าเฉลี่ยของประชากรสองกลุ่ม

วิธีทำ

สูบประจำ **(1)**	ไม่สูบ **(2)**
$n_1 = 75$	n_2 **= 75**
$\bar{X}_1$ **= 2.5**	$\bar{X}_2$ **= 3.3**
${\sigma^2}_1$**= 8**	${\sigma^2}_2$**= 4**

การทดสอบสมมติฐาน

วิธีทำ

1. ตั้งสมมติฐาน

   H_0 ; μ_1 - μ_2 $\geq$ 0

   H_1 : μ_1 - μ_2 $<$ 0

2. กำหนดสถิติทดสอบเป็น z

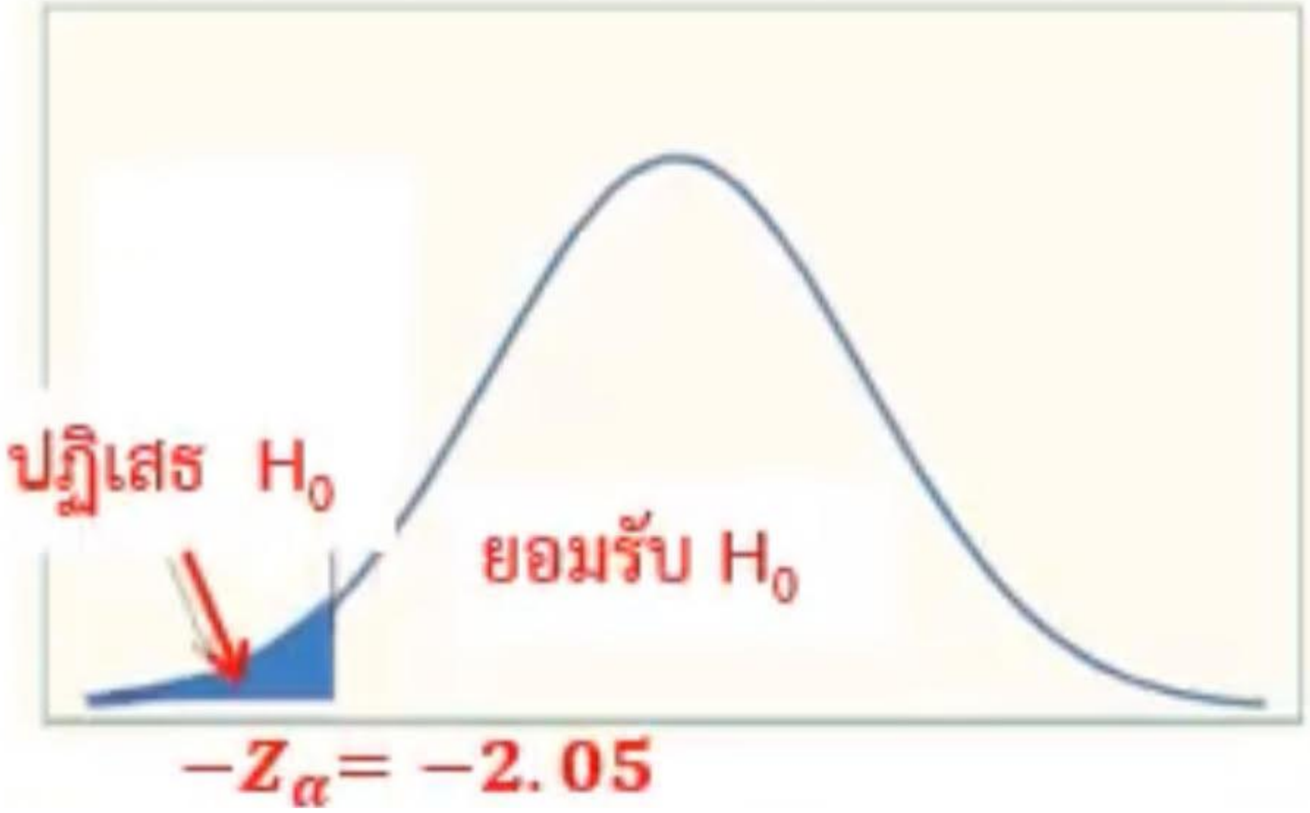
$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

การทดสอบสมมติฐาน

3. หาค่าวิกฤต

เป็นการทดสอบทางเดียว แบบทางขวาใช้ $Z_{\propto}$

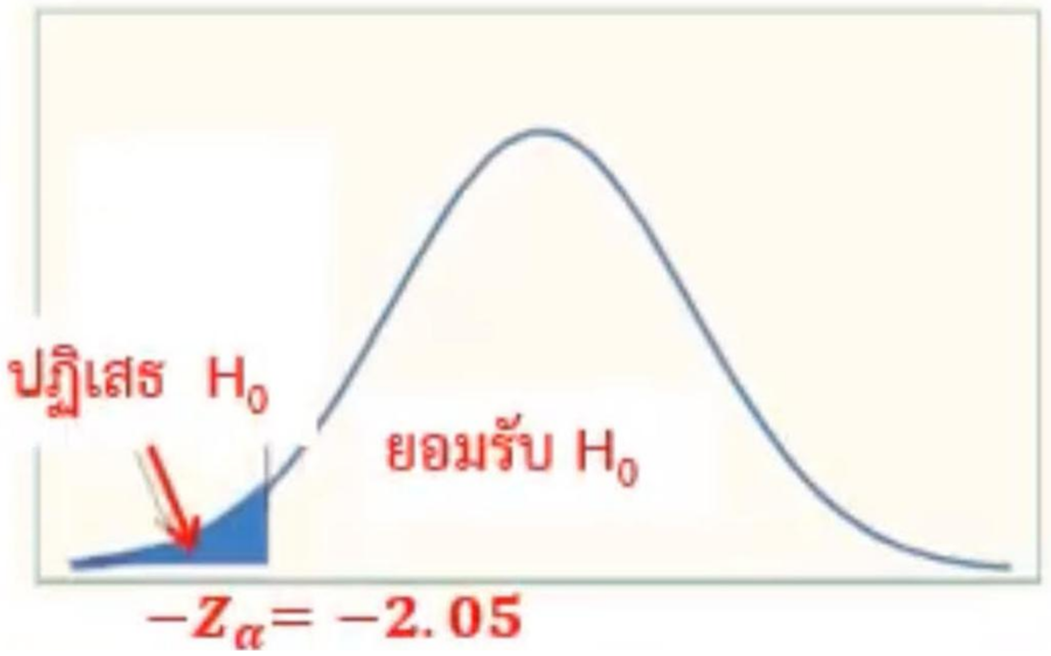
โจทย์กำหนด $\propto$ = **0.02** จะได้ $Z_{\propto}$ = **2.05**



การทดสอบสมมติฐาน

4. กำหนดกฎของหารตัดสินใจ

จะปฏิเสธ H_0 เมื่อ $Z < -2.05$



5. คำนวณค่าสถิติทดสอบ

$$Z = \frac{(2.5 - 3.3) - 0}{\sqrt{\frac{8}{75} + \frac{4}{75}}}$$

$$= -2.00$$

6. สรุปผล

ยอมรับ H_0 น้ำหนักของเด็กทารกแรกเกิดที่มีมารดาสูบบุหรี่เป็นประจำจะมากกว่าเท่ากับน้ำหนักของเด็กที่มารดาไม่สูบ

การทดสอบสมมติฐาน

การทดสอบสมมติฐานของค่าสัดส่วนของประชากร

สมมติฐานที่ตั้ง

$H_0 : p = p_0$	$H_0 : p \leq p_0$	$H_0 : p \geq p_0$
$H_1 : p \neq p_0$	$H_1 : p > p_0$	$H_1 : p < p_0$

สถิติทดสอบ

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

เมื่อ $q = 1 - p$

การทดสอบสมมติฐาน

การทดสอบสมมติฐานของค่าสัดส่วนของประชากร

ตัวอย่าง จากการสำรองโพลของ ABC เชื่อว่า **72%** ของคนที่เล่นอินเตอร์เน็ตจะถูกขโมยข้อมูลทางอินเตอร์เน็ต จึงทำการสุ่มตัวอย่างนักศึกษาที่เล่นอนเตอร์เน็ตจำนวน **300** คนจากมหาวิทยาลัยแห่งหนึ่งพบว่า **228** คน ถูกขโมยข้อมูลทางอินเตอร์เน็ต จงทดสอบความเชื่อนี้ได้หรือไม่ ที่ระดับรัยสำคัญ **0.10** $Z_{\propto/2} = 1.645$

การทดสอบสมมติฐาน

วิธีทำ **n = 300 , x = 228 , p = 0.72 (72/100) ,** $\hat{p} = \frac{228}{300} = 0.76$

1. ตั้งสมมติฐาน (สองทาง)

H_0 ; p = 0.72

H_1 : p $\neq$ 0.72

2. กำหนดสถิติทดสอบเป็น **z**

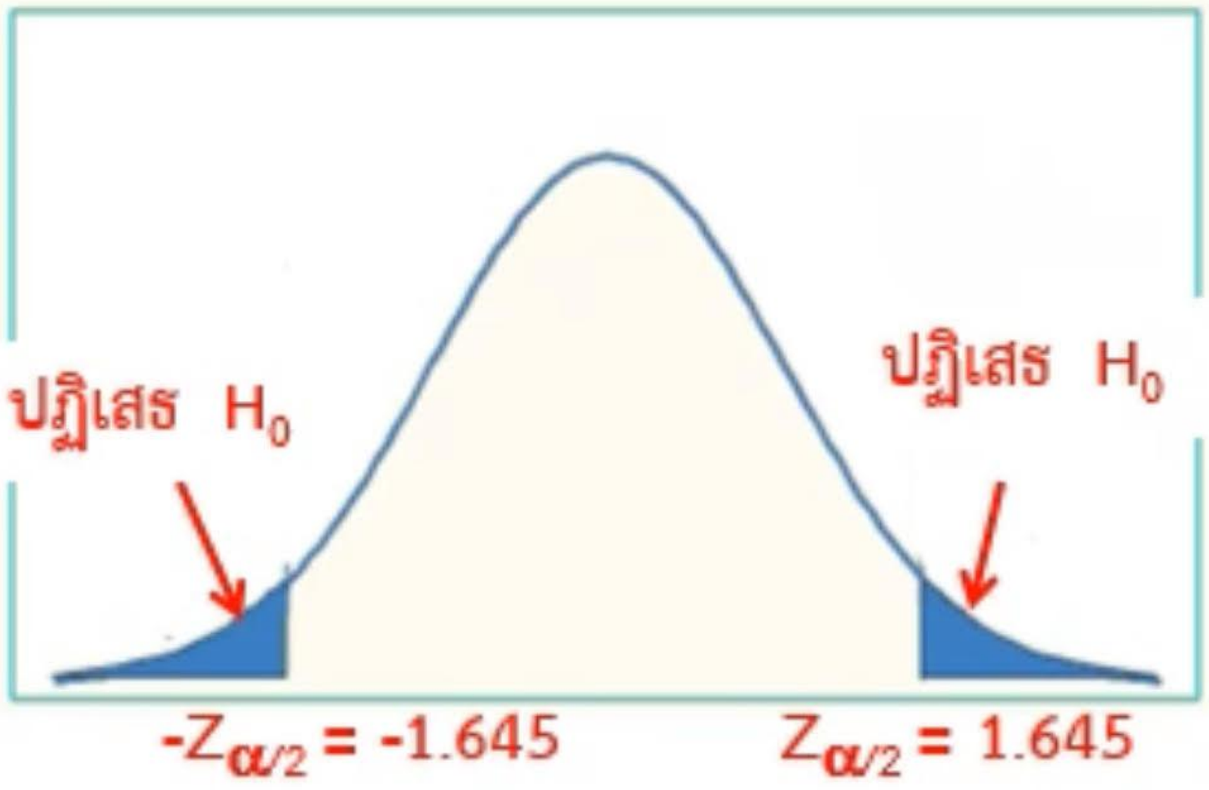
$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

การทดสอบสมมติฐาน

3. หาค่าวิกฤต

เป็นการทดสอบสองทาง ใช้ $Z_{\propto/2}$

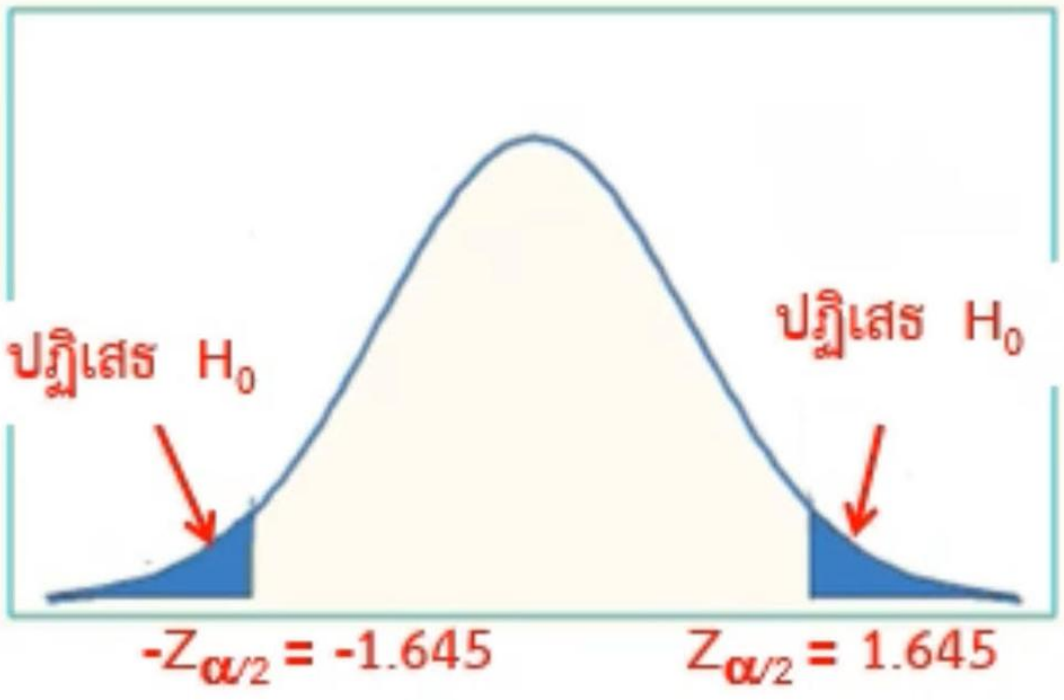
โจทย์กำหนด $\propto$ = **0.10 ,** $\propto/2$ = **0.05** จะได้ $Z_{\propto/2}$ = **1.645**



การทดสอบสมมติฐาน

4. กำหนดกฎของหารตัดสินใจ

จะปฏิเสธ H_0 เมื่อ **Z < -1.645** หรือ **Z > 1.645**



5. คำนวณค่าสถิติทดสอบ

P = 0.72 เลขที่ตั้งในสมมติฐาน

q = 1-0.72

= 0.28

$$Z = \frac{0.76 - 0.72}{\sqrt{\frac{0.72x0.28}{300}}}$$

= **1.54**

6. สรุปผล

ยอมรับ H_0 มีคนเล่นอินเตอร์เน็ตและถูกขโมยข้อมูลทางอินเตอร์เน็ต **72%** จริง

แบบฝึกหัด

1. น้ำตาลชนิดบรรจุขวดตราหนึ่ง พิมพ์ข้อความบนฉลากว่า “ น้ำหนักสุทธิ 200 กรัม ” ผู้บริโภครายหนึ่งสงสัยว่าข้อความดังกล่าวเกินความเป็นจริง จึงทำการสุ่มน้ำตาลตราดังกล่าวมา จำนวน 36 ขวด พบว่าน้ำหนักสุทธิมีค่าเฉลี่ย 199 กรัม และมีความแปรปรวนเท่ากับ 25 กรัม ที่ระดับนัยสำคัญ 0.05 ข้อสงสัยของผู้บริโภครายนี้เป็นความจริงหรือไม่ $Z_{\frac{\alpha}{2}} = 1.645$

แบบฝึกหัด

2. กาแฟชนิดบรรจุซองยี่ห้อหนึ่ง พิมพ์ข้อความบนฉลากว่า “ น้ำหนักสุทธิ 200 กรัม ” ผู้บริโภครายหนึ่งสงสัยว่าข้อความดังกล่าวเกินความเป็นจริง จึงทำการสุ่มกาแฟตราดังกล่าวมา จำนวน 20 ซอง พบว่า น้ำหนักสุทธิมีค่าเฉลี่ย 197 กรัม และมีความแปรปรวนเท่ากับ 25 กรัม ที่ระดับนัยสำคัญ 0.05 ข้อสงสัยของผู้บริโภครายนี้เป็นความจริงหรือไม่ $t_{\frac{\alpha}{2}} = 1.729$

Thank you.