บทที่ 10
สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ในการเลือกสถิติที่ใช้ในการวิเคราะห์ข้อมูล จำเป็นต้องทำความเข้าใจประเภทของสถิติและการนำไปใช้ในแต่ละประเภท ดังวัตถุประสงค์ของบทเรียนต่อไปนี้

1. สามารถอธิบายประเภทของสถิติที่จะนำมาใช้ในการประมวลผลข้อมูลได้
2. สามารถอธิบายคุณลักษณะของตัวอย่าง หรือ ประชากรจากค่าสถิติพรรณนาได้
3. สามารถอธิบายผลการทดสอบสมมุติฐานได้
4. สามารถเลือกใช้ค่าสถิติกับข้อมูลที่เก็บรวบรวมในงานวิจัยได้

ประเภทของสถิติ ตามความหมายของสถิติศาสตร์ มี 2 ประเภท คือ

1. สถิติพรรณนา (Descriptive statistics) ใช้ในการบรรยายลักษณะของกลุ่มตัวอย่างหรือกลุ่มประชากร ผลที่ได้จากการศึกษาจะไม่นำไปสรุปอ้างอิงถึงประชากร รูปแบบในการนำเสนอข้อมูล อาจเป็นรูปบทความ ตาราง ร้อยละ รูปกราฟประเภทต่าง ๆ ใช้วิธีการทางสถิติดังนี้

1.1 การแจกแจงความถี่ ได้แก่ จำนวน ร้อยละ

1.2 การวัดแนวโน้มเข้าสู่ส่วนกลาง เช่น ตัวกลางเลขคณิตหรือค่าเฉลี่ย ตัวกลางเรขาคณิต ฐานนิยม ควอไทล์ เดไซล์ เปอร์เซ็นต์ไทล์

1.3 การวัดการกระจายของข้อมูล เช่น พิสัย ส่วนเบี่ยงเบนมาตรฐาน ค่าแปรปรวน ค่าเบ้ ความโด่ง เป็นต้น

1.4 การหาความสัมพันธ์ระหว่างตัวแปร ค่าสถิติที่ใช้ สหสัมพันธ์ของเพียร์สัน (Pearson correlation) สหสัมพันธ์เชิงอันดับ (Spearman rank correlation)

2. สถิติอนุมานหรือสถิติอ้างอิง (Inferential statistics) คือ สถิติที่เกี่ยวกับการนำเสนอข้อมูลที่ได้จากกลุ่มตัวอย่าง (Sample) แล้วนำข้อเท็จจริงที่ได้ไปอธิบายลักษณะของประชากร (Population) ทั้งกลุ่ม การอนุมานทางสถิติมี 2 วิธี คือ

2.1 การอนุมานแบบพารามิเตอร์ (Parametric inference) เป็นการนำค่าที่ได้มาจากตัวอย่าง (Sample) ที่เรียกว่าค่าสถิติ (Statistics) ไปอธิบายคุณลักษณะ ประชากร (Population) ที่เรียกว่า ค่าพารามิเตอร์ (Parameter) โดยมีเงื่อนไขการเลือกใช้ค่าสถิติไปสรุปผลค่าพารามิเตอร์ ดังนี้

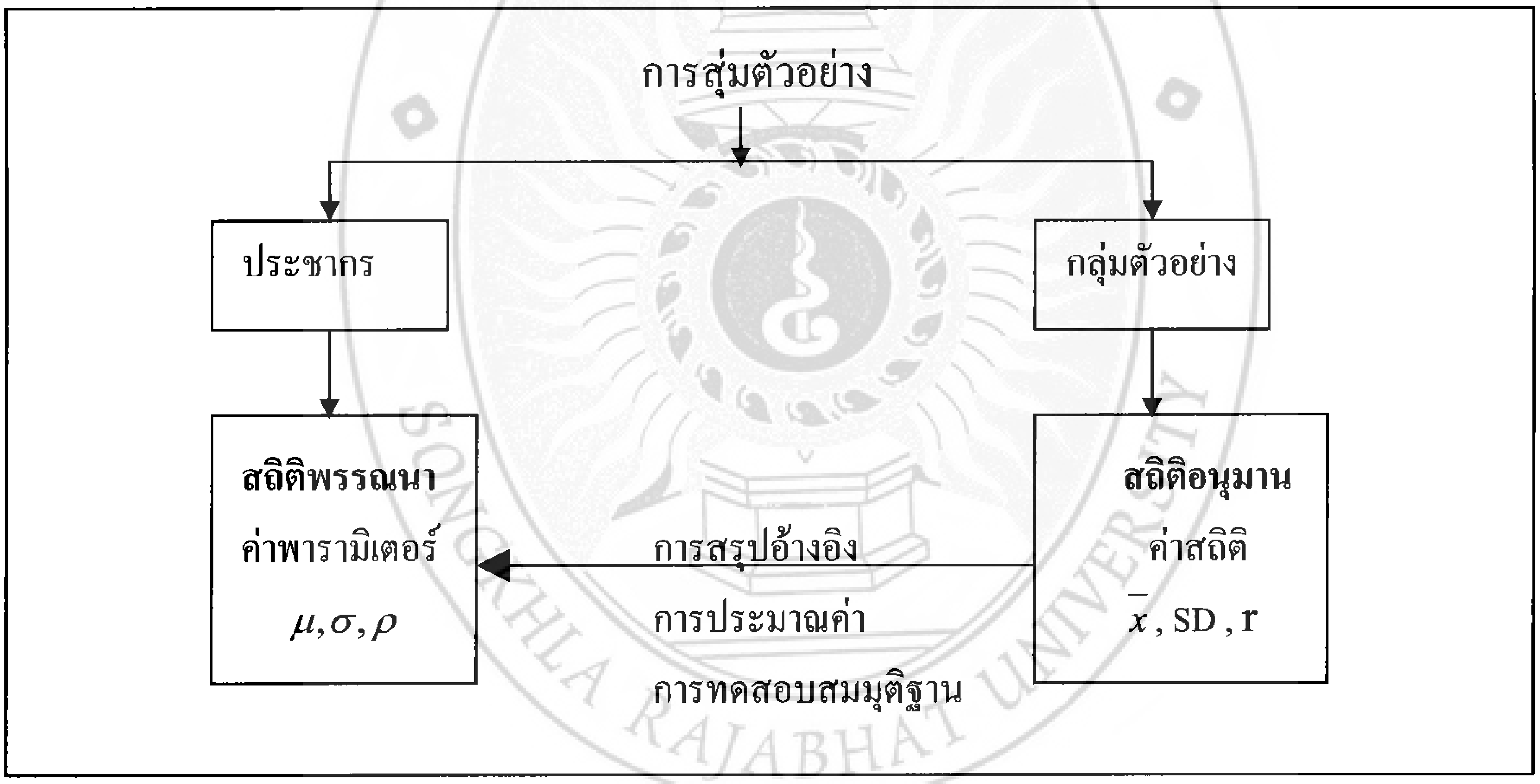
2.1.1 ค่าของประชากรควรมีการแจกแจงแบบปกติ (Normal Distribution)

2.1.2 การเลือกตัวอย่าง (Sample) เป็นไปอย่างอิสระและไม่มีความเอนเอียง (unbias)

2.1.3 ตัวแปรที่นำมาใช้เป็นตัวแปรระดับช่วง หรือ อัตราส่วน

2.2 การอนุมานแบบไม่มีพารามิเตอร์ (Non-Parametric inference) ใช้เมื่อเงื่อนไขหรือข้อมูลไม่สอดคล้องกับการอนุมานแบบพารามิเตอร์ เช่น ไม่ทราบค่าของข้อมูลจากประชากรที่สนใจว่ามีการแจกแจงแบบใด เป็นข้อมูลการวัดระดับกลุ่มหรือจัดอันดับ

โดยสรุป การเลือกสถิติที่ใช้ในการวิเคราะห์ข้อมูลให้พิจารณาว่า ข้อมูลมาจากประชากรหรือกลุ่มตัวอย่าง ถ้าข้อมูลได้มาจากประชากร สถิติที่จะนำมาใช้ในการวิเคราะห์ข้อมูล คือ สถิติเชิงพรรณนา (Descriptive statistics) เมื่อวิเคราะห์ข้อมูลจะได้ค่าพารามิเตอร์ แต่ถ้ารวบรวมจากกลุ่มตัวอย่างและต้องการที่จะสรุปอ้างอิงกลับไปหาประชากร สถิติที่นำมาใช้ในการวิเคราะห์ข้อมูลคือ สถิติอนุมาน (Inferential statistics) เมื่อวิเคราะห์ข้อมูลจะได้ค่าสถิติ ซึ่งแสดงความสัมพันธ์ระหว่างค่าสถิติและค่าพารามิเตอร์แสดงได้ดังภาพ



ภาพที่ 10.1 ความสัมพันธ์ระหว่างค่าสถิติและค่าพารามิเตอร์ (ดัดแปลงมาจาก มลิวัลย์ สมศักดิ์,2547)

จากภาพที่ 10.1 แสดงความสัมพันธ์ระหว่างค่าสถิติและค่าพารามิเตอร์ กล่าวคือ การวิจัยที่ศึกษาจากประชากรทั้งหมด เพื่อบรรยายสภาพที่เป็นอยู่ต้องใช้สถิติพรรณนา เมื่อวิเคราะห์ข้อมูลจะได้ค่าพารามิเตอร์ (Parameter) เช่น ค่าเฉลี่ยของประชากร () และส่วนเบี่ยงเบนมาตรฐานของประชากร () ซึ่งสามารถแปลความหมายและสรุปผลการวิจัยได้ แต่ในกรณีที่รวบรวมข้อมูลจากกลุ่มตัวอย่างที่สุ่มมาจากบางส่วนของประชากรต้องใช้สถิติอนุมาน ซึ่งค่าที่คำนวณได้เรียกว่า ค่าสถิติ (Statistic) เช่น ค่าเฉลี่ย () และส่วนเบี่ยงเบนมาตรฐาน (SD) แล้วจึงสรุปอ้างอิงค่าสถิติของกลุ่มตัวอย่างสู่ค่าพารามิเตอร์ของประชากร การสรุปอ้างอิงมี 2 วิธีคือ วิธีการประมาณค่า และการทดสอบสมมุติฐาน

สถิติพรรณนา (Descriptive statistics)

เป็นสถิติที่ใช้อธิบายคุณลักษณะหรือรายละเอียดของประชากรทางการวิจัย โดยเฉพาะที่นำมาใช้ในการวิจัยทางธุรกิจ มีทั้งการวัดเป็นร้อยละ การวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจาย โดยมีรายละเอียด ดังนี้

1. การวัดด้วยสถิติร้อยละ (Percentage)

ร้อยละเป็นสถิติที่นิยมใช้กันมาก ใช้ในการอธิบายประชากรเมื่อจำแนกตามเพศ อายุ การศึกษา อาชีพ เป็นต้น ซึ่งเป็นการเปรียบเทียบความถี่หรือจำนวนที่ต้องการ กับจำนวนทั้งหมดที่เทียบเป็น 100 โดยมีสูตรสำหรับคำนวณหาค่าร้อยละ ดังนี้

<u>สูตร</u> $P = \frac{f}{N} \times 100$

แทนค่า ค่า P = ร้อยละหรือ %

ค่า f = ความถี่ที่ต้องการเทียบค่าให้เป็นร้อยละ

ค่า N = จำนวนความถี่ทั้งหมด หรือ จำนวนประชากร

ตัวอย่างที่ 10.1

การเก็บข้อมูลจากสถานประกอบการ จังหวัดสงขลา จำนวน 450 แห่ง เป็นอุตสาหกรรมบริการ จำนวน 115 แห่ง โดยหาค่าร้อยละ ได้ดังนี้

$$P = \frac{115}{450} \times 100 = 25.56\ \%$$ หรือ ร้อยละ 25.56

2. การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of central tendency) เป็นการหาค่ากลางของข้อมูลชุดหนึ่งๆ เพื่อใช้เป็นตัวแทนของข้อมูลทั้งชุด ในการวิจัยทางธุรกิจมักใช้ในการวัดระดับความคิดเห็น ทัศนคติ เป็นต้น สถิติที่นิยมใช้ คือ ค่าเฉลี่ย (Mean) มัธยฐาน (Median) และฐานนิยม (Mode)

2.1 ค่าเฉลี่ย (Mean : $\overline{X}$)

<u>สูตร</u> $\overline{X} = \frac{\sum fX}{N}$

แทนค่า ค่า $\overline{X}$ = ค่าเฉลี่ย

ค่า $\sum fX$ = ผลรวมทั้งหมดของความถี่คูณค่าคะแนน

ค่า N = จำนวนประชากรในกลุ่ม

เหมาะสมที่ใช้กับข้อมูลระดับช่วง และ อัตราส่วน เป็นค่าที่ใกล้ความจริงมากที่สุด ถ้าข้อมูลแจกแจงเป็นโค้งปกติ เพราะนำทุกค่ามาเฉลี่ย แต่ไม่เหมาะกับข้อมูลที่มีค่าบางค่าสูงหรือต่ำกว่าข้อมูลอื่นๆ มาก

ตัวอย่างที่ 10.2 การหาค่าเฉลี่ยระดับความพึงพอใจ จงหาค่าเฉลี่ยข้อมูลชุดนี้

ระดับ	ค่าคะแนน(X)	ความถี่ (f)	ค่า fx
มากที่สุด	4	5	20
มาก	3	16	54
ปานกลาง	2	25	50
น้อย	1	20	20
ไม่พอใจ	0	4	0
		$\sum f = N = 70$	$\sum fx = 144$

$$\overline{X} = \frac{\sum fX}{N} = \frac{144}{70} = 2.06$$

ดังนั้น ระดับความพึงพอใจเฉลี่ย =2.06 ดังนั้นข้อมูลชุดนี้มีความพึงพอใจระดับปานกลาง

2.2 **ค่ามัธยฐาน (Median : Mdn)** เป็นค่าที่อยู่ตรงกลางโดยการเรียงลำดับค่าทั้งหมด จากน้อยไปมาก หรือ จากมากไปน้อย ถ้าข้อมูลเป็นเลขคี่ มัธยฐานจะเป็นค่าที่อยู่ตรงกลางพอดี แต่ถ้าข้อมูลเป็นเลขคู่ จะต้องนำเลขตรงกลาง 2 ตัวมาบวกกันแล้วหาร 2

เหมาะกับข้อมูลระดับอันดับ ช่วงและอัตราส่วน เป็นค่าที่ใกล้เคียงความจริงน้อยกว่าค่าเฉลี่ย แต่ก็ดีกว่าฐานนิยม เหมาะสำหรับข้อมูลที่มีค่าบางค่าสูงหรือต่ำกว่าข้อมูลอื่นๆ มาก

ตัวอย่างที่ 10.3 จงหาค่ามัธยฐาน ของน้ำหนักสินค้า จำนวน 6 กล่อง ดังนี้ 15 12 18 21 16 22

วิธีทำ จัดเรียงเลขใหม่ ได้ดังนี้ 12 15 16 18 21 22 ดังนั้นค่ากลางคือ 16 และ 18

ดังนั้น $\text{Mdn} = \frac{16+18}{2} = 17$

2.3 **ค่าฐานนิยม (Mode : Mo)** หมายถึง ค่าของคะแนนที่มีความถี่สูงสุดของข้อมูลชุดหนึ่งๆ ถ้าข้อมูลชุดใดมีค่าความถี่สูงสุดมากกว่า 1 ค่า ข้อมูลชุดนั้นก็มีฐานนิยมมากกว่า 1 ค่า หรือ ถ้าข้อมูลชุดใดมีค่าความถี่สูงสุดเท่าๆ กันทุกค่า ข้อมูลชุดนั้นไม่มีฐานนิยม

เหมาะกับข้อมูลระดับกลุ่ม อันดับ ช่วงและอัตราส่วน หาได้ง่ายที่สุด แต่เป็นค่าที่ใกล้เคียงความจริงน้อยที่สุด

ตัวอย่างที่ 10.4 จงหาค่าฐานนิยม ของราคา ส.ค.ส. ปี 25xx ดังนี้ 5 7 10 15 12 12 5 8 10 7 12 15

ดังนั้น Mo = 12

3. การวัดการกระจาย (Measures of variability) การวัดการกระจายเป็นการบอกให้ทราบว่าข้อมูลที่ได้จากการวัดมีค่าใกล้เคียงหรือแตกต่างกันหรือกระจายจากกันมากน้อยเพียงใด เพราะการวัดแนวโน้มเข้าสู่ส่วนกลางเพียงอย่างเดียวจะได้เพียงค่าที่เป็นตัวแทน ไม่เพียงพอที่จะอธิบายลักษณะของข้อมูลได้ จึงต้องวัดการกระจายควบคู่ไปด้วย เพราะแม้ค่าที่ได้จากการวัดเข้าส่วนกลางเท่ากันแต่การกระจายของข้อมูลอาจจะไม่เหมือนกันได้ ดังตัวอย่าง

ตัวอย่างที่ 10.5 ข้อมูลชุดที่ 1 2 5 6 9 13 มี $\overline{X} = 7$

ข้อมูลชุดที่ 2 5 6 7 8 9 มี $\overline{X} = 7$

จะเห็นได้ว่าข้อมูล 2 ชุดมีค่าเฉลี่ยเท่ากัน แต่ลักษณะการกระจายข้อมูลแตกต่างกัน คือ ข้อมูลชุดที่ 1 มีการกระจายของข้อมูลสูงกว่า ข้อมูลชุดที่ 2 ดังนั้นการวัดการกระจายจึงช่วยให้ทราบลักษณะของข้อมูลได้ชัดเจนยิ่งขึ้น การวัดการกระจายที่นิยม 4 วิธี คือ พิสัย (Range) ส่วนเบี่ยงเบนควอไทล์ (Quartile deviation) ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) และความแปรปรวน (Variance)

3.1 พิสัย (Range) หมายถึง ผลต่างของข้อมูลที่มีค่าสูงสุดกับข้อมูลที่มีค่าต่ำสุด เป็นการวัดการกระจายของข้อมูลอย่างคร่าว ๆ และทำได้อย่างรวดเร็วที่สุด อาจแปลความอย่างหยาบ ๆ ได้ว่า ถ้าค่าพิสัยมาก ข้อมูลก็มีการกระจายมากไปด้วย

<u>สูตร</u> พิสัย = คะแนนสูงสุด-คะแนนต่ำสุด

ตัวอย่างที่ 10.6 ข้อมูล 2 7 8 6 4 9 6 ดังนั้นพิสัย = 9-2 = 7

3.2 ส่วนเบี่ยงเบนควอไทล์ (Quartile deviation) หมายถึง ครึ่งหนึ่งของความแตกต่างของควอไทล์ที่ 3 กับควอไทล์ที่ 1 เหมาะสำหรับใช้สำหรับกรณีที่ใช้ค่ามัธยฐานเป็นสถิติที่ใช้วัดแนวโน้มเข้าสู่ส่วนกลาง

<u>สูตร</u> $$\text{QD} = \frac{Q3 - Q1}{2}$$

แทนค่า ค่า QD = ส่วนเบี่ยงเบนควอไทล์

ค่า Q1 = ค่าที่แบ่งข้อมูลออกเป็น $\frac{1}{4}$ หรือ 25%

ค่า Q3 = ค่าที่แบ่งข้อมูลออกเป็น $\frac{3}{4}$ หรือ 75%

ตัวอย่างที่ 10.7 นักธุรกิจ 8 คน ทำกำไรในรอบปีได้ดังนี้ (หน่วย: ล้านบาท)

$$\begin{array}{cccccccc} 22 & 26 & 29 & 33 & 34 & 38 & 40 & 43 \\ & \downarrow & & & \downarrow & & \downarrow & \\ & Q1 & & & \text{Mdn} = \frac{33+34}{2} = 33.5 & & Q3 & \end{array}$$

$$QD = \frac{Q3-Q1}{2} = \frac{40-26}{2} = 7$$

3.3 ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) หมายถึงรากที่สองของค่าเบี่ยงเบนของข้อมูลแต่ละตัวจากค่าเฉลี่ยยกกำลังสอง (ความแปรปรวน) เป็นการวัดการกระจายที่บอกว่าข้อมูลกระจายจากค่าเฉลี่ยมากน้อยเพียงใด ถ้าค่าเบี่ยงเบนมาตรฐานน้อยแสดงว่าข้อมูลของแต่ละคนในกลุ่มมีความใกล้เคียงกัน ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่างใช้สัญลักษณ์ SD และใช้สัญลักษณ์ σ ในข้อมูลประชากร

3.3.1 กรณีเก็บข้อมูลจากกลุ่มตัวอย่างและใช้ค่าเฉลี่ยในการคำนวณ

<u>สูตร</u>

$$SD = \sqrt{\frac{\sum (X-\overline{X})^2}{n-1}}$$

แทนค่า
ค่า SD = ส่วนเบี่ยงเบนมาตรฐาน
ค่า X = คะแนนของแต่ละคน
ค่า $\overline{X}$ = ค่าเฉลี่ยของกลุ่มตัวอย่าง
ค่า n = จำนวนตัวอย่าง

ตัวอย่างที่ 10.8 จงหาส่วนเบี่ยงเบนมาตรฐานของกำไรของนักธุรกิจ 6 คน ดังนี้ (หน่วย : ล้านบาท)

5 8 4 10 7 8

วิธีทำ ค่า $\overline{X} = \dfrac{5+8+4+10+7+8}{6} = 7$

นักธุรกิจคนที่	X	$(X-\overline{X})$	$(X-\overline{X})^2$
1	5	-2	4
2	8	1	1
3	4	-3	9
4	10	3	9
5	7	0	0
6	8	1	1
$n = 6$			$\sum(X-\overline{X})^2 = 24$

ดังนั้น $SD = \sqrt{\dfrac{\sum(X-\overline{X})^2}{n-1}}$

$= \sqrt{\dfrac{24}{6-1}}$

$= 2.19$

3.3.2 กรณีเก็บข้อมูลจากกลุ่มตัวอย่างและใช้คะแนนดิบในการคำนวณ

<u>สูตร</u> $SD = \sqrt{\dfrac{n\sum X^2-(\sum X)^2}{n(n-1)}}$

แทนค่า ค่า SD = ส่วนเบี่ยงเบนมาตรฐาน

ค่า $\sum X$ = ผลรวมของคะแนนแต่ละคน

ค่า $\sum X^2$ = ผลรวมของคะแนนแต่ละคนยกกำลังสอง

ค่า n = จำนวนตัวอย่าง

ตัวอย่างที่ 10.9 จงหาส่วนเบี่ยงเบนมาตรฐานของกำไรของนักธุรกิจ 6 คน ดังนี้ (หน่วย:ล้านบาท)
5 8 4 10 7 8

นักธุรกิจคนที่	X	X^2
1	5	25
2	8	64
3	4	16
4	10	100
5	7	49
6	8	64
$n = 6$	$\sum X = 42$	$\sum X^2 = 318$

$$SD = \sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n-1)}}$$

$$= \sqrt{\frac{(6)(318) - (42)^2}{6(6-1)}}$$

$$= 2.19$$

3.3.3 กรณีเก็บข้อมูลจากประชากรและใช้คะแนนดิบในการคำนวณ

<u>สูตร</u> $\sigma = \sqrt{\frac{\sum X^2 - (\sum X)^2}{N^2}}$

แทนค่า ค่า σ = ส่วนเบี่ยงเบนมาตรฐานของประชากร

ค่า $\sum X$ = ผลรวมของคะแนนแต่ละคน

ค่า $\sum X^2$ = ผลรวมของคะแนนแต่ละคนยกกำลังสอง

ค่า N = จำนวนประชากร

3.4 ความแปรปรวน (Variance) หมายถึง ค่ากำลังสองของส่วนเบี่ยงเบนมาตรฐาน เป็นการวัดการกระจายของคะแนนโดยมองในลักษณะของพื้นที่ ใช้สัญลักษณ์ S^2 หรือ SD^2 เมื่อเป็นความแปรปรวนของกลุ่มตัวอย่าง และใช้สัญลักษณ์ σ^2 เมื่อเป็นความแปรปรวนของประชากร

3.4.1 กรณีเก็บข้อมูลจากกลุ่มตัวอย่างและใช้คะแนนเฉลี่ยในการคำนวณ

<u>สูตร</u> $SD^2 = \frac{\sum(X-\overline{X})^2}{n-1}$

แทนค่า ค่า SD^2 = ความแปรปรวน

ค่า X = คะแนนของแต่ละคน

ค่า $\overline{X}$ = ค่าเฉลี่ยของกลุ่มตัวอย่าง

ค่า n = จำนวนตัวอย่าง

3.4.2 กรณีเก็บข้อมูลจากกลุ่มตัวอย่างและใช้คะแนนดิบในการคำนวณ

<u>สูตร</u> $SD^2 = \frac{n\sum X^2-(\sum X)^2}{n(n-1)}$

แทนค่า ค่า SD^2 = ความแปรปรวน

ค่า $\sum X$ = ผลรวมของคะแนนแต่ละคน

ค่า $\sum X^2$ = ผลรวมของคะแนนแต่ละคนยกกำลังสอง

ค่า n = จำนวนตัวอย่าง

3.4.3 กรณีเก็บข้อมูลจากประชากรและใช้คะแนนเฉลี่ยในการคำนวณ

<u>สูตร</u> $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$

แทนค่า ค่า σ^2 = ความแปรปรวนของประชากร

ค่า X = คะแนนของแต่ละคน

ค่า μ = ค่าเฉลี่ยของประชากร

ค่า N = จำนวนประชากร

3.4.4 กรณีเก็บข้อมูลจากประชากรและใช้คะแนนดิบในการคำนวณ

<u>สูตร</u> $\sigma^2 = \frac{N\sum X^2-(\sum X)^2}{N^2}$

แทนค่า ค่า σ^2 = ความแปรปรวนของประชากร

ค่า $\sum X$ = ผลรวมของคะแนนแต่ละคน

ค่า $\sum X^2$ = ผลรวมของคะแนนแต่ละคนยกกำลังสอง

ค่า N = จำนวนประชากร

4. การวัดความสัมพันธ์ระหว่างตัวแปร เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัวหรือมากกว่าที่มาจากแหล่งเดียวกันว่ามีความเกี่ยวพันหรือมีความแปรปรวนร่วมกันมากน้อยเพียงใด ถ้ามีความแปรปรวนร่วมกันหรือซ้อนทับกันทั้งหมดแปลว่าสองสิ่งนั้นสัมพันธ์กันสูงมาก ถ้าความแปรปรวนซ้อนทับกันบางส่วนแปลว่าสองสิ่งนั้นสัมพันธ์กันน้อย ในกรณีความแปรปรวนไม่ซ้อนทับกันเลยแปลว่าสองสิ่งนั้นไม่สัมพันธ์กันเลย (ล้วน สายยศ และอังคณา สายยศ,2538) ลักษณะของความสัมพันธ์ดังกล่าวมีดัชนีที่ชี้ให้เห็นความสัมพันธ์เรียกว่า สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) สัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ระหว่าง - 1.00 ถึง + 1.00 ในการแปลความหมายสัมประสิทธิ์สหสัมพันธ์ต้องพิจารณาทั้งปริมาณและทิศทางของความสัมพันธ์ ถ้าพิจารณาปริมาณของความสัมพันธ์ให้พิจารณาเฉพาะตัวเลข ส่วนเครื่องหมาย บวก หรือ ลบ แสดงทิศทางของสหสัมพันธ์

การแปลความหมายปริมาณของสัมประสิทธิ์สหสัมพันธ์ว่ามากน้อยเพียงใดอาจจะพิจารณาจากเกณฑ์ดังนี้ (มลิวัลย์ สมศักดิ์,2547)

.91	-	1.00	หมายถึง	มีความสัมพันธ์สูงมาก
.71	-	.90	หมายถึง	มีความสัมพันธ์สูง
.31	-	.70	หมายถึง	มีความสัมพันธ์ปานกลาง
.00	-	.30	หมายถึง	มีความสัมพันธ์ต่ำ

การแปลความหมายทิศทางของสัมประสิทธิ์สหสัมพันธ์ พิจารณาจากเครื่องหมาย คือ ถ้าสัมประสิทธิ์สหสัมพันธ์เป็นบวก หมายความว่า ตัวแปร 2 ตัว มีความสัมพันธ์ในลักษณะตามกัน คือ ถ้าใครได้คะแนนสูงในตัวแปรหนึ่งแล้วจะได้คะแนนสูงในอีกตัวแปรหนึ่งด้วย และถ้าใครได้คะแนนต่ำในตัวแปรหนึ่งแล้วจะได้คะแนนต่ำในอีกตัวแปรหนึ่งด้วย เช่น ถ้าโฆษณาสินค้าใดมีความถี่ในการออกอากาศมาก ก็จะมียอดขายของสินค้าสูงด้วย แต่ถ้าโฆษณาใดมีความถี่ในการออกอากาศน้อยก็จะมียอดขายของสินค้าต่ำด้วย และถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบ หมายความว่าตัวแปร 2 ตัว มีความสัมพันธ์ในลักษณะตรงข้ามกัน คือ ถ้าใครได้คะแนนสูงในตัวแปรหนึ่งแล้วจะได้คะแนนต่ำในอีกตัวแปรหนึ่ง และถ้าใครได้คะแนนต่ำในตัวแปรหนึ่งแล้วจะได้คะแนนสูงในอีกตัวแปรหนึ่ง เช่น ถ้าพนักงานฝ่ายผลิตคนใดมีผลผลิต(Yield) สูง จะมีสัมพันธภาพกับเพื่อนต่ำ แต่ถ้าพนักงานฝ่ายผลิตคนใดมีผลผลิตต่ำ จะมีสัมพันธภาพกับเพื่อนสูง เป็นต้น

การแปลความหมายสัมประสิทธิ์สหสัมพันธ์ จะแปลความหมายเกี่ยวกับปริมาณและทิศทางของความสัมพันธ์ระหว่างตัวแปรเท่านั้น ซึ่งจะแปลความหมายได้แต่เพียงว่าตัวแปร 2 ตัวมีความสัมพันธ์กันหรือไม่ มีความสัมพันธ์ระดับใดและสัมพันธ์กันในทิศทางใด จะไม่แปลความหมายว่าตัวแปรหนึ่งเป็นสาเหตุของอีกตัวแปรหนึ่ง

วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ มีหลายวิธีขึ้นอยู่กับข้อมูลว่าอยู่ในมาตรการวัดระดับใด ในที่นี้จะกล่าวถึงการหาค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple correlation coefficient) คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรเพียงสองตัว ซึ่งนิยมใช้กันมาก คือ สหสัมพันธ์แบบเพียร์สัน และสหสัมพันธ์เชิงอันดับ

4.1 สหสัมพันธ์แบบเพียร์สัน (Pearson product moment correlation) ใช้หาความสัมพันธ์ระหว่างตัวแปรสองตัวที่มีความสัมพันธ์เป็นเส้นตรง (Linear relationship) ซึ่งมีข้อมูลอยู่ในระดับช่วง หรือ อัตราส่วน สัญลักษณ์ที่ใช้คือ r_{xy}

สูตร
$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

แทนค่า ค่า r_{xy} = ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

ค่า $\sum X, \sum Y$ = ผลรวมของคะแนน X และ Y ตามลำดับ

ค่า $\sum XY$ = ผลรวมของผลคูณระหว่างคะแนน X และคะแนน Y

ค่า $\sum X^2, \sum Y^2$ = ผลรวมของคะแนนแต่ละคนยกกำลังสอง

ค่า n = จำนวนข้อมูล

ตัวอย่างที่ 10.10 จงหาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนวิชาสถิติเพื่อการวิจัย กับวิชาวิธีวิทยาการวิจัยของนักศึกษาปริญญาโท 10 คนดังนี้

คนที่	วิชาสถิติ	วิชาวิธีวิทยาการวิจัย			
	X	Y	X^2	Y^2	XY
1	15	20	225	400	300
2	16	16	256	256	256
3	15	17	225	289	255
4	17	18	285	324	306
5	14	15	196	225	210
6	18	21	324	441	378
7	13	14	169	196	182
8	18	20	324	400	360
9	17	25	289	625	425
10	18	22	324	484	396
	$\sum X$ =161	$\sum Y$ =188	$\sum X^2$ =2621	$\sum Y^2$ =3640	$\sum XY$ =3068

<u>สูตร</u> $$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{10(3068) - (161)(188)}{\sqrt{[10(2621) - (161)^2][10(3640) - (188)^2}} = 0.75$$

สรุป คะแนนวิชาสถิติเพื่อการวิจัย กับ วิชาวิธีวิทยาการวิจัยของนักศึกษาปริญญาโท มีความสัมพันธ์กันสูงในทางบวก

4.2 สหสัมพันธ์เชิงอันดับ (Spearman rank correlation coefficient) ใช้หาความสัมพันธ์ระหว่างตัวแปรสองตัวที่มีข้อมูลอยู่ในการวัดระดับจัดอันดับ สัญลักษณ์ที่ใช้คือ r_s

<u>สูตร</u> $$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

แทนค่า ค่า r_s = ค่าสัมประสิทธิ์สหสัมพันธ์อันดับที่ของสเปียร์แมน

ค่า D = ผลต่างของอันดับที่ของข้อมูลแต่ละคู่

ค่า n = จำนวนข้อมูล

ตัวอย่างที่ 10.11 จงหาสัมประสิทธิ์สหสัมพันธ์ระหว่างอันดับที่ของไอคิวกับความสามารถในการปรับตัวเข้ากับสังคมของพนักงานฝ่ายบัญชี 10 คนดังนี้

คนที่	อันดับที่ไอคิว	อันดับที่การปรับตัวเข้ากับสังคม	D	D^2
1	8	6	2	4
2	1	3	-2	4
3	6	9	-3	9
4	9	2	7	49
5	10	8	2	4
6	4	1	3	9
7	5	10	-5	25
8	3	5	-2	4
9	7	7	0	0
10	2	4	-2	4
				$\sum D^2 = 112$

<u>สูตร</u> $$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6(112)}{10(100-1)} = 0.32$$

สรุป อันดับที่ของไอคิวกับความสามารถในการปรับตัวเข้ากับสังคม มีความสัมพันธ์กันปานกลางในทางบวก

สถิติอนุมาน (Inferential statistics)

สถิติอนุมานหรือสถิติอ้างอิง เป็นสถิติที่วิเคราะห์ข้อมูลที่รวบรวมจากกลุ่มตัวอย่างแล้วนำผลมาสรุปอ้างอิงกลับไปยังประชากรโดยมีขั้นตอนดังนี้ คือ เลือกกลุ่มตัวอย่างที่ศึกษาให้เป็นตัวแทนที่ดีของประชากร ศึกษาคุณลักษณะของกลุ่มตัวอย่างโดยคำนวณค่าสถิติต่าง ๆ เช่น $\overline{X}$,SD และ r แล้วค่าสถิติที่ได้สรุปอ้างอิงไปถึงคุณลักษณะของประชากรด้วยวิธีการทางสถิติ ซึ่งมี 2 วิธี ได้แก่ การประมาณค่า และ การทดสอบสมมุติฐาน การวิจัยส่วนใหญ่จะใช้การทดสอบสมมุติฐานในที่นี้จะกล่าวเฉพาะการทดสอบสมมุติฐานเท่านั้น

1. การทดสอบสมมุติฐาน (Hypothesis testing) เป็นกระบวนการทดสอบสมมุติฐานทางสถิติ เพื่อตัดสินใจว่าสมมุติฐานการวิจัยที่ตั้งไว้เกี่ยวกับประชากรที่ศึกษานั้นถูกต้องเป็นจริงหรือไม่ เริ่มจากการแปลงสมมุติฐานการวิจัยให้เป็นสมมุติฐานทางสถิติ แล้วใช้เทคนิคทางสถิติแบบใดแบบหนึ่งแล้วแต่กรณีมาทดสอบโดยอาศัยข้อมูลจากกลุ่มตัวอย่าง ซึ่งมีขั้นตอนการทดสอบดังนี้

1.1 ตั้งสมมุติฐานทางสถิติ โดยแปลงมาจากสมมุติฐานการวิจัย สมมุติฐานทางสถิติจะต้องตั้งทั้งสมมุติฐานหลัก (H_0) และสมมุติฐานรอง (H_1) โดยเฉพาะสมมุติฐานรองต้องตั้งให้สอดคล้องกับสมมุติฐานการวิจัยเสมอ

1.2 กำหนดระดับนัยสำคัญ (Level of significance) ในการทดสอบสมมุติฐานจะต้องตัดสินใจว่าจะยอมรับหรือปฏิเสธสมมุติฐานหลัก (H_0) ที่ตั้งขึ้น ถ้ายอมรับสมมุติฐานหลักที่เป็นจริงหรือปฏิเสธสมมุติฐานหลักที่ไม่เป็นจริงแล้วผลการตัดสินใจดังกล่าวก็ถูกต้อง แต่ในทางปฏิบัติโอกาสที่จะตัดสินใจถูกต้องทั้งหมดเป็นไปได้ยาก เนื่องจากมีโอกาสที่จะเกิดความคลาดเคลื่อน 2 ประเภท คือ

1) ความคลาดเคลื่อนประเภทที่ 1 (Type I error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมุติฐานหลัก (H_0) ที่เป็นจริง ความน่าจะเป็นในการเกิดความคลาดเคลื่อนประเภทที่ 1 เรียกว่าระดับนัยสำคัญ แทนด้วยสัญลักษณ์ α (อ่านว่า แอลฟา)

2) ความคลาดเคลื่อนประเภทที่ 2 (Type II error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการยอมรับสมมุติฐานหลัก (H_0) ที่ไม่เป็นจริง ความน่าจะเป็นในการเกิดความคลาดเคลื่อน

ประเภทที่ 2 เรียกว่าระดับนัยสำคัญ แทนด้วยสัญลักษณ์ β (อ่านว่า เบต้า)

ความสัมพันธ์ระหว่างความคลาดเคลื่อนประเภทที่ 1 และประเภทที่ 2 และความน่าจะเป็นในการเกิดความคลาดเคลื่อนแต่ละประเภทแสดงได้ดังภาพ 10.2

		สถานการณ์ที่ถูกต้องของ H_o	
		H_o เป็นจริง	H_o ไม่จริง
การตัดสินใจ	ยอมรับ H_o	ตัดสินใจถูกต้อง $(1 - \alpha)$	ความคลาดเคลื่อนประเภทที่ 2 (β)
	ปฏิเสธ H_o	ความคลาดเคลื่อนประเภทที่ 1 (α)	ตัดสินใจถูกต้อง $(1 - \beta)$

ภาพที่ 10.2 ความสัมพันธ์ระหว่างความคลาดเคลื่อนประเภทที่ 1 และประเภทที่ 2 และความน่าจะเป็นในการเกิดความคลาดเคลื่อนแต่ละประเภท (ดัดแปลงมาจาก มลิวัลย์ สมศักดิ์,2547)

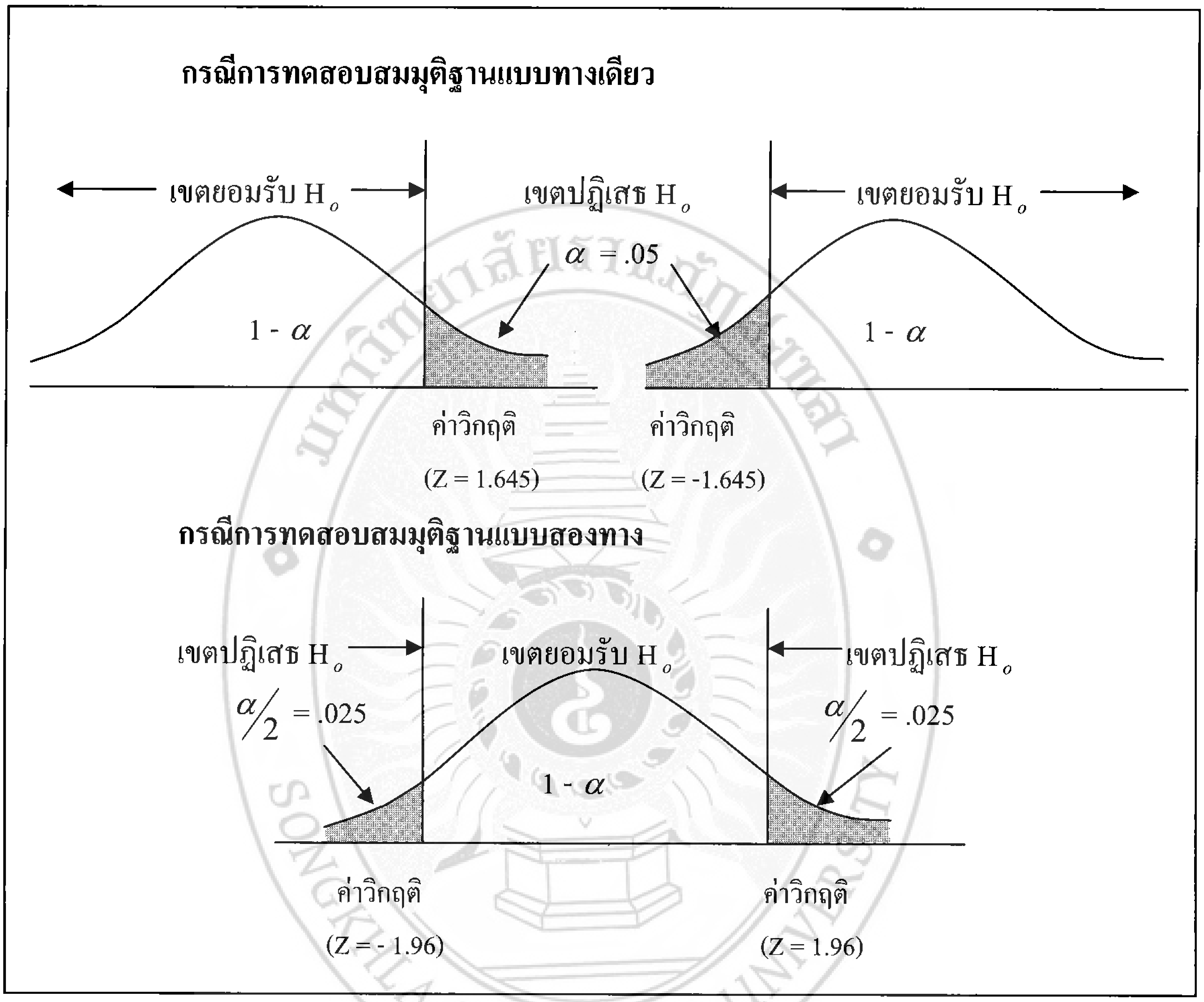
จากภาพ การตัดสินใจในการทดสอบสมมุติฐานมีโอกาสที่จะตัดสินใจคลาดเคลื่อนได้ 2 ประเภท ประเภทแรก เมื่อปฏิเสธ H_0 ที่เป็นจริงทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 (Type I error) มีความน่าจะเป็น = α กับอีกประเภทหนึ่งคือยอมรับ H_0 ที่ไม่จริงทำให้เกิดความคลาดเคลื่อนประเภทที่ 2 (Type II error) มีความน่าจะเป็น = β ในการวิจัยทางธุรกิจจะยอมรับให้เกิดความคลาดเคลื่อนประเภทที่ 1 หรือกำหนดนัยสำคัญ (α) ไว้ที่ระดับ .05 และ .01

1.3 เลือกเทคนิคทางสถิติ ที่จะใช้ในการทดสอบสมมุติฐาน เช่น Z-test, t-test หรือ F-test

1.4 กำหนดเขตวิกฤติ (Critical region) เขตวิกฤติเป็นพื้นที่หรือบริเวณที่ใช้ในการปฏิเสธสมมุติฐานหลัก (H_0) การกำหนดเขตวิกฤติพิจารณาจาก 3 ขั้น ขั้นตอนแรก คือ พิจารณาว่าสมมุติฐานรอง (H_1) ที่ผู้วิจัยตั้งไว้เป็นสมมุติฐานแบบมีทิศทางหรือไม่มีทิศทาง ขั้นตอนที่สอง พิจารณาว่ากำหนดนัยสำคัญ (α) ไว้ที่ระดับใด และขั้นตอนที่สาม พิจารณาว่าใช้สถิติแบบใด เมื่อพิจารณาทั้ง 3 ประการนี้แล้วผู้วิจัยก็จะไปอ่านค่าวิกฤติ (Critical value) จากตารางแจกแจงของค่าสถิตินั้นๆ ซึ่งค่าวิกฤติ หมายถึง ค่าที่เป็นจุดแบ่งระหว่างเขตยอมรับสมมุติฐาน H_0 และเขตปฏิเสธสมมุติฐาน H_0 เช่น ถ้าผู้วิจัยทดสอบสมมุติฐานเพื่อเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยที่กำหนดระดับนัยสำคัญหรือ α = .05 และใช้ Z-test ในการทดสอบ อ่านค่าวิกฤติจากตารางการแจกแจงของค่า Z กรณีทดสอบทางเดียวจะได้ค่าวิกฤติเท่ากับ -1.645 หรือ 1.645 และกรณีทดสอบสองทางต้องหาค่าวิกฤติโดยใช้ $\alpha/2$ ซึ่งจะได้ค่าวิกฤติเท่ากับ 1.96 ดังภาพที่ 10.3

1.5 คำนวณค่าสถิติ โดยการนำข้อมูลที่ได้จากกลุ่มตัวอย่างมาคำนวณหาค่าสถิติ

1.6 สรุปผล การสรุปผลมีหลักในการตัดสินใจ 2 กรณีคือ ถ้าค่าสถิติที่คำนวณได้อยู่ในเขตปฏิเสธสมมุติฐาน H_0 สรุปผลได้ว่าปฏิเสธสมมุติฐาน H_0 และยอมรับสมมุติฐาน H_1 แต่ถ้าค่าสถิติที่คำนวณได้อยู่ในเขตยอมรับสมมุติฐาน H_0 สรุปผลได้ว่ายอมรับสมมุติฐาน H_0



ภาพที่ 10.3 เขตวิกฤติในการใช้ Z-test ทดสอบสมมุติฐานที่นัยสำคัญทางสถิติที่ระดับ .05

2. การทดสอบสมมุติฐานเกี่ยวกับความแตกต่างระหว่างค่าเฉลี่ย

2.1 การทดสอบค่าเฉลี่ยในกลุ่มตัวอย่างเดียว (One sample) เป็นการทดสอบเพื่อเปรียบเทียบว่าค่าเฉลี่ยของกลุ่มตัวอย่างแตกต่างจากค่าเฉลี่ยของประชากรหรือไม่ ทดสอบโดยใช้สูตร t-test แบบ One sample test ดังนี้

<u>สูตร</u> $$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

$df = n - 1$

แทนค่า ค่า μ = ค่าเฉลี่ยของประชากร

ค่า $\overline{X}$ = ค่าเฉลี่ยของกลุ่มตัวอย่าง

ค่า S = ส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง

ค่า n = ขนาดของกลุ่มตัวอย่าง

ข้อตกลงเบื้องต้น

1. กลุ่มตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงเป็นโค้งปกติ
2. ไม่ทราบค่าความแปรปรวนของประชากร
3. ค่าของตัวแปรตามแต่ละหน่วยเป็นอิสระต่อกัน
4. กลุ่มตัวอย่างมีขนาดเล็ก ($n < 100$)

ตัวอย่างที่ 10.12 เมื่อปีที่แล้วพบว่านักเรียนได้รับเงินไปโรงเรียนมีค่าเฉลี่ยเท่ากับ50บาทต่อวัน ในปีนี้ผู้วิจัยได้สุ่มตัวอย่างมา20คน เมื่อสอบถามแล้วได้ค่าเฉลี่ยเท่ากับ54บาทและ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4 อยากทราบว่าในปีนี้ค่าเฉลี่ยของเงินที่นักเรียนได้รับต่อวันสูงกว่าปีที่แล้วหรือไม่

วิธีทำ ขั้นที่ 1 ตั้งสมมุติฐาน $H_o : \mu_1 = 50$

$H_1 : \mu_1 > 50$

ขั้นที่ 2 กำหนดระดับนัยสำคัญทางสถิติที่ .01

ขั้นที่ 3 กำหนดสถิติทดสอบโดยใช้ t-test แบบ One sample test

ขั้นที่ 4 หาค่าวิกฤติของ t โดยเปิดตารางแจกแจงของ t ที่นัยสำคัญทางสถิติที่ระดับ .01 และ df = n-1 = 19 ได้ค่า t ตาราง = 2.539

ขั้นที่ 5 คำนวณค่าสถิติ

<u>สูตร</u> $$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{54 - 50}{4/\sqrt{20}} = 4.49$$

ขั้นที่ 6 สรุปผล ค่า t คำนวณ (4.49) > t ตาราง (2.539) จึงปฏิเสธ H_0 และ ยอมรับ H_1 นั่นคือ นักเรียนได้รับเงินต่อวันสูงกว่าปีที่แล้วอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

2.2 การทดสอบความแตกต่างของค่าเฉลี่ย 2 กลุ่ม เพื่อเปรียบเทียบว่าค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มแตกต่างกันหรือไม่ สถิติที่ใช้ทดสอบคือ t-test ซึ่งมีหลายกรณีดังนี้

2.2.1 กรณีกลุ่มตัวอย่างเป็นอิสระกัน (Independent sample) กลุ่มตัวอย่างเป็นอิสระกัน หมายถึง กลุ่มตัวอย่าง 2 กลุ่มไม่เกี่ยวข้องกัน เช่น นักศึกษา ชาย-หญิง ใช้สูตร t-test แบบ Independent sample ดังนี้

1) กรณีทราบถึงความแปรปรวนของประชากร ใช้สูตร

<u>สูตร</u>

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$df = n_1 + n_2 - 2$$

2) กรณีไม่ทราบค่าความแปรปรวนของประชากรแต่ assume ว่าความแปรปรวนเท่ากัน ($\sigma_1^2 = \sigma_2^2$) ใช้ความแปรปรวนร่วม (Pooled variance) สูตร

<u>สูตร</u>

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]}{n_1 + n_2 - 2}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$$df = n_1 + n_2 - 2$$

ข้อตกลงเบื้องต้น

1. กลุ่มตัวอย่างแต่ละกลุ่มเป็นอิสระต่อกันและได้มาโดยสุ่ม
2. กลุ่มตัวอย่างแต่ละกลุ่มได้มาจากประชากรที่มีการแจกแจงเป็นโค้งปกติ
3. ความแปรปรวนของประชากรแต่ละกลุ่มเท่ากัน
4. กลุ่มตัวอย่างมีขนาดเล็ก ($n < 100$)

3) กรณีไม่ทราบค่าความแปรปรวนของประชากรและ assume ว่าความแปรปรวนไม่เท่ากัน ($\sigma_1^2 \neq \sigma_2^2$) ใช้สูตรดังนี้

<u>สูตร</u> $$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$df = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}}$$

ข้อตกลงเบื้องต้น

1. กลุ่มตัวอย่างแต่ละกลุ่มเป็นอิสระต่อกันและได้มาโดยสุ่ม
2. กลุ่มตัวอย่างแต่ละกลุ่มได้มาจากประชากรที่มีการแจกแจงเป็นโค้งปกติ
3. ความแปรปรวนของประชากรแต่ละกลุ่มไม่เท่ากัน
4. กลุ่มตัวอย่างมีขนาดเล็ก (n < 100)

หมายเหตุ ในการทดสอบความแตกต่างของค่าเฉลี่ยของประชากรสองกลุ่ม ควรจะตรวจสอบลักษณะข้อมูลเสียก่อนว่าเป็นไปตามข้อตกลงเบื้องต้นเกี่ยวกับความแปรปรวนของประชากรหรือไม่ เพื่อตัดสินใจว่าจะใช้สูตรตามข้อ 2) หรือ 3) โดยการทดสอบสมมุติฐานเกี่ยวกับความเท่ากันของความแปรปรวนของประชากร ใช้สูตร

<u>สูตร</u> $$F = \frac{S_1^2}{S_2^2}$$

$df_1 = n_1 - 1$ และ $df_2 = n_2 - 1$

แทนค่า
- ค่า S_1^2 = ค่าความแปรปรวนตัวที่มีค่ามากกว่า
- ค่า S_2^2 = ค่าความแปรปรวนตัวที่มีค่าน้อยกว่า
- ค่า n_1 = ค่าจำนวนตัวอย่างกลุ่มที่มีความแปรปรวนมากกว่า
- ค่า n_2 = ค่าจำนวนตัวอย่างกลุ่มที่มีความแปรปรวนน้อยกว่า

หาค่าวิกฤติโดยเปิดตารางการแจกแจงของ F ที่ $df_1 = n_1 - 1$ และ $df_2 = n_2 - 1$ ถ้าค่า F ที่คำนวณได้น้อยกว่าค่า F จากตารางแสดงว่าความแปรปรวนของประชากรทั้ง 2 กลุ่มเท่ากัน ($\sigma_1^2 = \sigma_2^2$) เลือกใช้สูตร 2) ถ้าค่า F ที่คำนวณได้มากกว่าค่า F จากตารางแสดงว่าความแปรปรวนของประชากรทั้ง 2 กลุ่มไม่เท่ากัน ($\sigma_1^2 \neq \sigma_2^2$) เลือกใช้สูตร 3)

ตัวอย่างที่ 10.13 จงทดสอบว่าพนักงานชายเก่งกว่าพนักงานหญิงหรือไม่จากการทดสอบการขายหนังสือ ซึ่งได้ค่าสถิติพื้นฐานดังนี้

พนักงานชาย	พนักงานหญิง
n = 38	n = 45
$\overline{X}$ = 28.34	$\overline{X}$ = 26.03
SD = 3.87	SD = 4.08

วิธีการทดสอบความแปรปรวน

ก่อนตัดสินใจว่าจะเลือกใช้สูตร 2) หรือ 3) ต้องทดสอบความแปรปรวนก่อนเพื่อให้ทราบว่าความแปรปรวนเท่ากันหรือแตกต่างกัน โดยวิธีการดังนี้

1. ตั้งสมมุติฐาน $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

2. กำหนดนัยสำคัญทางสถิติที่ระดับ .05

3. กำหนดสถิติทดสอบคือ F-test

4. หาค่าวิกฤติของ F จากตารางการแจกแจงของ F ที่ระดับนัยสำคัญทางสถิติ .05 และ df_1 = 40 (ใกล้เคียง 37) และ df_2 =40 (ใกล้เคียง 44) ได้ค่า F ตาราง = 1.69

5. คำนวณค่าสถิติใช้สูตร

$$F = \frac{S_1^2}{S_2^2} = \frac{4.08^2}{3.87^2} = 1.11$$

6. สรุปค่า F คำนวณ (1.11) < F ตาราง (1.69) จึงยอมรับ H_0 สรุปได้ว่าความแปรปรวนของประชากรทั้งสองกลุ่มไม่แตกต่างกัน จึงใช้ t-test แบบ Independent sample 2)

วิธีทำ ขั้นที่ 1 ตั้งสมมุติฐาน $H_o : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

กำหนดให้ μ_1 แทนค่าเฉลี่ยของยอดขายหนังสือของพนักงานชาย

μ_2 แทนค่าเฉลี่ยของยอดขายหนังสือของพนักงานหญิง

ขั้นที่ 2 กำหนดระดับนัยสำคัญทางสถิติที่ .05

ขั้นที่ 3 กำหนดสถิติทดสอบโดยใช้ t-test แบบ Independent sample

ขั้นที่ 4 หาค่าวิกฤติของ t โดยเปิดตารางแจกแจงของ t ที่นัยสำคัญทางสถิติที่ระดับ .05 และ df = 38+45-2 = 81 ได้ค่า t ตาราง = 1.671

ขั้นที่ 5 คำนวณค่าสถิติ

<u>สูตร</u> $$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{[(n_1-1)S_1^2 + (n_2-1)S_2^2]}{n_1+n_2-2}\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$$

$$= \frac{28.34 - 26.03}{\sqrt{\frac{[(38-1)(3.87)^2 + (45-1)(4.08)^2]}{38+45-2}\left[\frac{1}{38}+\frac{1}{45}\right]}} = 2.60$$

ขั้นที่ 6 สรุปผล ค่า t คำนวณ (2.60) > t ตาราง (1.671) จึงปฏิเสธ H_0 และ ยอมรับ H_1 นั่นคือ พนักงานชายขายหนังสือได้มากกว่าพนักงานหญิงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2.2.2 กรณีกลุ่มตัวอย่างไม่เป็นอิสระกัน (Dependent sample) หมายถึง กลุ่มตัวอย่าง 2 กลุ่มมาจากประชากรที่มีความสัมพันธ์กัน เช่น กลุ่มที่มีลักษณะเหมือนกันเป็นคู่ เช่น ฝาแฝด หรือเป็นกลุ่มเดียวแต่มีการสอบสองครั้ง เช่น การสอบก่อนเรียนและหลังเรียน ใช้สูตร t-test แบบ Dependent sample ดังนี้

<u>สูตร</u> $$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D)^2}{n-1}}}$$

df = n-1

แทนค่า ค่า D = ผลต่างของคะแนนแต่ละคู่

ค่า n = จำนวนคู่ของกลุ่มตัวอย่าง

ข้อตกลงเบื้องต้น

1. กลุ่มตัวอย่างแต่ละกลุ่มเกี่ยวข้องกัน
2. การแจกแจงของความแตกต่างของคะแนนแต่ละคู่แจกแจงเป็นโค้งปกติ
3. คู่ของสิ่งที่จะศึกษาได้มาด้วยวิธีการสุ่ม

ตัวอย่างที่ 10.14 ในการอบรมเชิงปฏิบัติการการใช้คอมพิวเตอร์แก่พนักงานฝ่ายธุรการ ได้สุ่มตัวอย่างพนักงานที่เข้ารับการอบรม 10 คน ทดสอบความรู้และทักษะการใช้คอมพิวเตอร์ ก่อนและหลังอบรม ปรากฏว่าได้คะแนน ดังนี้

คนที่	1	2	3	4	5	6	7	8	9	10
ก่อนอบรม	4	6	8	8	4	7	3	4	6	9
หลังอบรม	9	11	14	16	10	10	7	14	8	9

จงทดสอบว่าหลังการอบรมครั้งนี้พนักงานมีความรู้ และทักษะในการใช้คอมพิวเตอร์เพิ่มขึ้นกว่าก่อนการอบรมหรือไม่

วิธีทำ ขั้นที่ 1 ตั้งสมมุติฐาน $H_o : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

กำหนดให้ μ_1 แทนค่าเฉลี่ยของคะแนนความรู้ และทักษะการใช้คอมพิวเตอร์หลังอบรม

μ_2 แทนค่าเฉลี่ยของคะแนนความรู้ และทักษะการใช้คอมพิวเตอร์ก่อนอบรม

ขั้นที่ 2 กำหนดระดับนัยสำคัญทางสถิติที่ .01

ขั้นที่ 3 กำหนดสถิติทดสอบโดยใช้ t-test แบบ Dependent sample

ขั้นที่ 4 หาค่าวิกฤติของ t โดยเปิดตารางแจกแจงของ t ที่นัยสำคัญทางสถิติที่ระดับ .01 และ df = 10-1 = 9 ได้ค่า t ตาราง = 2.821

ขั้นที่ 5 คำนวณค่าสถิติ

คนที่	หลังอบรม	ก่อนอบรม	D	D^2
1	9	4	5	25
2	11	6	5	25
3	14	8	6	36
4	16	8	8	64
5	10	4	6	36
6	10	7	3	9
7	7	3	4	12
8	14	4	10	100
9	8	6	2	4
10	9	9	0	0
			$\sum D = 49$	$\sum D^2 = 315$

<u>สูตร</u>

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

$$= \frac{49}{\sqrt{\frac{10(315)-(49)^2}{10-1}}} = 5.37$$

ขั้นที่ 6 สรุปผล ค่า t คำนวณ (5.37) > t ตาราง (2.821) จึงปฏิเสธ H_0 และ ยอมรับ H_1 นั่นคือ พนักงานธุรการมีความรู้และทักษะการใช้คอมพิวเตอร์หลังอบรมสูงกว่าก่อนอบรมอย่าง มีนัยสำคัญทางสถิติที่ระดับ .01

3. การทดสอบสมมุติฐานเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร

จากตัวอย่างที่ 10.10 สรุปว่าคะแนนวิชาสถิติเพื่อการวิจัยกับวิชาวิธีวิทยาการวิจัยของนักศึกษาปริญญาโทมีความสัมพันธ์กันสูงในทางบวก คือมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .75 ถ้ากลุ่มตัวอย่าง 10 คน สุ่มมาจากประชากรต้องทดสอบว่าความสัมพันธ์ดังกล่าวมีนัยสำคัญทางสถิติที่ระดับ .05 หรือไม่ การทดสอบนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์ทดสอบโดยใช้ t-test

<u>สูตร</u>

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

$$df = n - 2$$

วิธีทำ ขั้นที่ 1 ตั้งสมมุติฐาน $H_o : \rho = 0$

$H_1 : \rho \neq 0$

กำหนดให้ ρ แทนความสัมพันธ์ระหว่างคะแนนวิชาสถิติเพื่อการวิจัยกับวิชาวิธีวิทยาการวิจัยของนักศึกษาปริญญาโท

ขั้นที่ 2 กำหนดระดับนัยสำคัญทางสถิติที่ .05

ขั้นที่ 3 กำหนดสถิติทดสอบโดยใช้ t-test

ขั้นที่ 4 หาค่าวิกฤติของ t โดยเปิดตารางแจกแจงของ t ที่นัยสำคัญทางสถิติที่ระดับ .05 และ df = n-2 = 10-2 = 8 ได้ค่า t ตาราง = 2.306

ขั้นที่ 5 คำนวณค่าสถิติ

<u>สูตร</u>

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

$$= \frac{.75\sqrt{10-2}}{\sqrt{1-(.75)^2}} = 2.78$$

ขั้นที่ 6 สรุปผล ค่า t คำนวณ (2.78) > t ตาราง (2.306) จึงปฏิเสธ H_0 และ ยอมรับ H_1 นั่นคือ คะแนนวิชาสถิติเพื่อการวิจัยกับวิชาวิธีวิทยาการวิจัยของนักศึกษาปริญญาโทมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

บทสรุป

ประเภทของสถิติ มี 2 ประเภท คือ 1) สถิติพรรณนา ใช้บรรยายลักษณะของกลุ่มประชากร มีวิธีการดังนี้ การแจกแจงความถี่ [ค่าร้อยละ] การวัดแนวโน้มเข้าสู่ส่วนกลาง [ค่าเฉลี่ย มัธยฐาน และฐานนิยม] การวัดการกระจายของข้อมูล [พิสัย ส่วนเบี่ยงเบนควอไทล์ ส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวน] และการหาความสัมพันธ์ระหว่างตัวแปร [สหสัมพันธ์ของเพียร์สัน และ สหสัมพันธ์เชิงอันดับ] 2) สถิติอนุมานหรือสถิติอ้างอิง คือ สถิติที่เกี่ยวกับการนำเสนอข้อมูลที่ได้จากกลุ่มตัวอย่าง แล้วนำข้อเท็จจริงที่ได้ไปอธิบายลักษณะของประชากรทั้งกลุ่ม การอนุมานทางสถิติมี 2 วิธี คือ การอนุมานแบบพารามิเตอร์เป็นการนำค่าที่ได้มาจากตัวอย่าง ที่เรียกว่าค่าสถิติ ไปอธิบายคุณลักษณะประชากรที่เรียกว่า ค่าพารามิเตอร์ และ การอนุมานแบบไม่มีพารามิเตอร์ ใช้เมื่อเงื่อนไขหรือข้อมูลไม่สอดคล้องกับการอนุมานแบบพารามิเตอร์ เช่น ไม่ทราบค่าของข้อมูลจากประชากรที่สนใจว่ามีการแจกแจงแบบใด เป็นข้อมูลการวัดระดับกลุ่มหรือระดับจัดอันดับ การสรุปอ้างอิงไปถึงคุณลักษณะของประชากรด้วยวิธีการทางสถิติ มี 2 วิธี ได้แก่ การประมาณค่า และ การทดสอบสมมุติฐาน การวิจัยส่วนใหญ่จะใช้การทดสอบสมมุติฐาน ในที่นี้กล่าวถึง การทดสอบสมมุติฐานเกี่ยวกับความแตกต่างระหว่างค่าเฉลี่ย และการทดสอบสมมุติฐานเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร

คำถามท้ายบท

1. สถิติหมายถึงอะไร แบ่งออกเป็นกี่ประเภท
2. การอธิบายคุณลักษณะของข้อมูล ใช้การวัดวิธีใดบ้าง มีความหมายอย่างไร
3. การทดสอบสมมุติฐาน ใช้ในกรณีใด มีขั้นตอนอย่างไร
4. ระดับนัยสำคัญ คืออะไร
5. จากข้อมูลการใช้จ่ายซื้อขนมของเด็ก 10 คน ดังนี้ (หน่วย:บาท) 9 12 35 50 42 8 64 41 60 19 ควรหาแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจายด้วยวิธีใดจึงเหมาะสมและคำตอบเป็นเท่าใด