เทคนิคการวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติ

ขั้นต้นและขั้นกลาง

สำหรับงานวิจัย

โดย

รองศาสตราจารย์ ศิริชัย พงษ์วิชัย

ภาควิชา สถิติ คณะพาณิชย์ศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ความหมายของสถิติ

- **ตัวเลข (Numerical)**

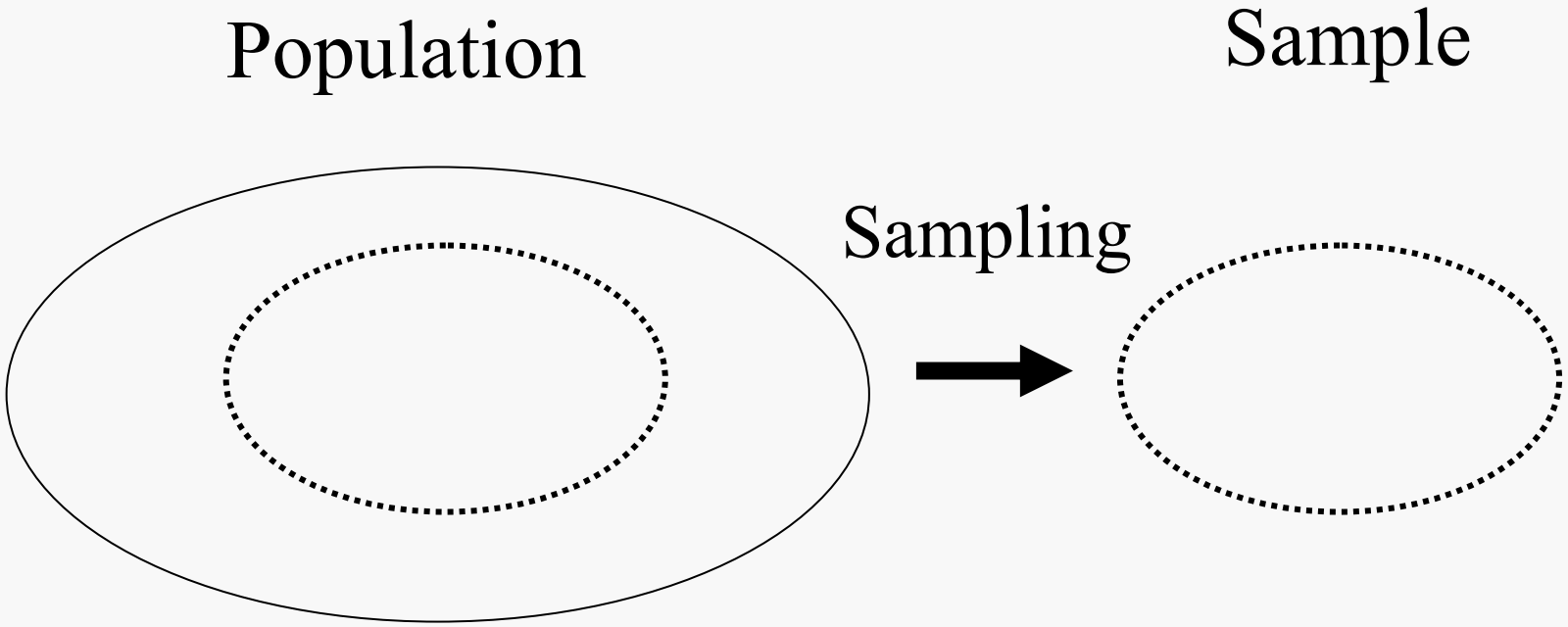
  ตัวเลขที่แสดงรายละเอียดต่าง ๆ เช่น *สถิติการจดทะเบียนรถยนต์ในกรุงเทพฯเดือน ก.ค. สถิติผู้ป่วยเป็นโรคเอดส์ในประเทศไทยปี 47*

- **ศาสตร์ (Sciences)**

  วิชาที่ศึกษาถึงวิธีการต่าง ๆ ดังนี้ ***การเก็บรวบรวมข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล การแปลความหมายของข้อมูล และ การสรุปผลข้อมูล***

สถิติกับงานวิจัย

- สถิติบรรยาย (**Descriptive Statistics**) 04
- สถิติอนุมาน (**Inference Statistics**) 05



ข้อมูลกับระดับการวัด

☞ *คุณภาพ (Qualitative Scale)*

- นามบัญญัติ (Nominal Scale)

  จำแนกข้อมูลเป็นกลุ่มเป็นพวกแต่ละกลุ่มมีความเท่าเทียมกัน

- เรียงอันดับ (Ordinal Scale)

  Nominal + สามารถบอกความแตกต่างระหว่างกลุ่มได้

☞ *ปริมาณ (Quantitative Scale)*

- ช่วง (Interval Scale)

  Ordinal + บอกความแตกต่างระหว่างกลุ่มเป็น *ตัวเลข*ได้ แต่ *ไม่มี* ศูนย์แท้ และ จุดเริ่มต้น *ไม่* เป็นธรรมชาติ

- อัตราส่วน (Ratio Scale)

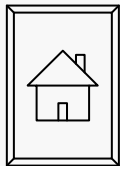
  Ordinal + บอกความแตกต่างระหว่างกลุ่มเป็นตัวเลขได้ มีศูนย์แท้ และ จุดเริ่มต้นเป็นธรรมชาติ

- **สถิติบรรยาย (Descriptive Statistics)**
  - **การนำเสนอข้อมูล (Presentation)**
  - **การแจกแจงความถี่ (Frequencies)**
  - **การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measure Central of Tendency)**
    - ☆ **การหาค่าเฉลี่ย(*Mean*)แบบต่างๆ A.M , G.M , H.M**
    - ☆ **การหาค่าฐานนิยม(*Mode*)**
    - ☆ **การหาค่าแสดงตำแหน่งของข้อมูล**

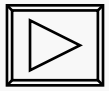
      Median , Quartiles , Deciles , Percentiles , N-Tiles
  - **การหาค่าการกระจายของข้อมูล (Dispersion)**

    Range , Quartiles Deviation , Mean Deviation ,

    Standard Deviation , Coefficient Variation



สถิติอนุมาน (Inference Statistics) 05

ค่า Statistics และ ค่า Parameter  06

- การอนุมานแบบพาราเมตริก (Parametric) ☺ ความน่าเชื่อถือสูง

  ข้อมูลอยู่ในระดับ Interval /Ratio

  ทราบการแจกแจงของข้อมูล หรือ ข้อมูลมีจำนวนมาก

  - ❑ *ใช้ประมาณค่า(**Estimation**) กำหนดขนาดตัวอย่าง* ▶ 08
  - ❑ *ใช้ทดสอบสมมติฐาน (**Hypothesis Testing**)* ▶ 11
  - ❑ *ใช้หาความสัมพันธ์ (**Correlation & Association**)* ▶ 20
  - ❑ *ใช้พยากรณ์ (**Forecasting & Prediction**)* ▶ 25
  - ❑ *ใช้วิเคราะห์แบบหลายตัวแปร (**Multivariate Analysis**)* ▶ 26

- การอนุมานแบบน็อนพาราเมตริก (Non-Parametric)

  ข้อมูลอยู่ในระดับ Nominal / Ordinal

  ไม่ทราบการแจกแจงของข้อมูล และ ข้อมูลมีจำนวนน้อย ☹ ความน่าเชื่อถือต่ำ

ค่า Statistics & ค่า Parameter

★ ค่า Statistics คือ

ค่าที่แสดงลักษณะของตัวอย่าง(Sample) โดยเป็นค่าที่คำนวณได้จากข้อมูลตัวอย่าง เช่น ค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ย ฯลฯ

★ ค่า Parameter คือ

ค่าที่แสดงลักษณะของประชากร(Population) โดยเป็นค่าที่คำนวณได้จากข้อมูลประชากร เช่น ค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ย ฯลฯ

ค่า Statistics และค่า Parameter

ค่า ที่ทำการคำนวณ	ค่า Statistics	ค่า Parameter
ค่าเฉลี่ย(Mean)	$\overline{X}$	μ
ค่าสัดส่วน(Proportion)	p	π
ค่าความคลาดเคลื่อน	s	σ
ค่าความสัมพันธ์	r	ρ
ค่าอื่นๆ	$\hat{\theta}$	θ

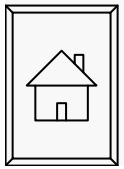
Estimation : การประมาณค่า

การประมาณค่าแบบจุด (Point Estimation)

- ประมาณค่าเฉลี่ย $\bar{x}$
- ประมาณค่าสัดส่วน p

การประมาณค่าแบบช่วง (Interval Estimation)

- ประมาณค่าเฉลี่ย $\bar{x} \pm t_{\alpha/2} * s / \sqrt{n}$
- ประมาณค่าสัดส่วน $p \pm z_{\alpha/2} * \sqrt{p*(1-p)/n}$



Sampling : การเลือกตัวอย่าง

แบบใช้ความน่าจะเป็น (Probability Techniques)

- แบบสุ่มอย่างง่าย(Simple Random Sampling)
- แบบแบ่งกลุ่ม(Stratified Sampling)
- แบบมีระบบ (Systematic Sampling)
- แบบกลุ่ม(Cluster Sampling)
- แบบอื่น ๆ

แบบไม่ใช้ความน่าจะเป็น (Non Probability Techniques)

- แบบตามสะดวก(Convenience Sampling)
- แบบใช้เกณฑ์ผู้วิจัย(Purposive Sampling)
- แบบกำหนดโควต้า(Quota Sampling)
- แบบสโนว์บอล(Snowball Sampling)

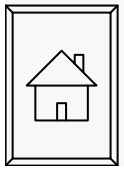
Definite Sampling : การกำหนดขนาดตัวอย่าง

- ใช้แนวทางของความน่าจะเป็น
- ใช้แนวทางของการทดสอบสมมติฐาน
- ใช้แนวทางของการประมาณค่า
  - จากการประมานค่าเฉลี่ย
  - จากการประมานค่าสัดส่วน

สิ่งที่ต้องกำหนดไว้ก่อนการกำหนดขนาดตัวอย่าง

(1) กำหนดช่วงความเชื่อมั่นของการประมาณค่า

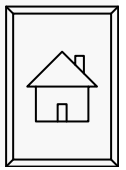
(2) กำหนดระดับความคลาดเคลื่อนของการประมาณค่า



Testing Statistical Hypothesis : การทดสอบสมมติฐานทางสถิติ 11

- กำหนดสมมติฐานทางสถิติ
- กำหนดเกณฑ์ที่จะใช้ในการทดสอบ
- เลือกวิธีการทดสอบทางสถิติที่เหมาะสม
- สร้างขอบเขตของการตัดสินใจ
- คำนวณค่าสถิติจากข้อมูลตัวอย่าง
- ตัดสินใจ ปฏิเสธ หรือ ยอมรับสมมติฐาน
- สรุปผล

- การทดสอบเกี่ยวกับความแตกต่าง
- การทดสอบเกี่ยวกับการหาความสัมพันธ์ 2 ตัวแปร



การทดสอบสมมติฐานสำหรับความแตกต่างของ ค่าเฉลี่ย & ค่าร้อยละ

1. การทดสอบเกี่ยวกับ *ค่าเฉลี่ย*

1.1 *ค่าเฉลี่ย 1 กลุ่มตัวอย่าง* ▶ 13

1.2 *ค่าเฉลี่ย 2 กลุ่มตัวอย่าง*

1.2.1 2 กลุ่มตัวอย่าง เป็นอิสระต่อกัน ▶ 14

1.2.2 2 กลุ่มตัวอย่าง มีความสัมพันธ์กัน ▶ 15

1.3 *ค่าเฉลี่ยตั้งแต่ 3 กลุ่มตัวอย่างขึ้นไป* ▶ 16

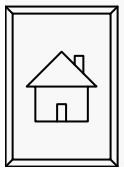
2. การทดสอบเกี่ยวกับ *ค่าร้อยละ* หรือ *สัดส่วน*

2.1 *ค่าสัดส่วน 1 กลุ่มตัวอย่าง* ▶ 17

2.2 *ค่าสัดส่วน ตั้งแต่ 2 กลุ่มตัวอย่างขึ้นไป*

2.2.1 กลุ่มตัวอย่าง เป็นอิสระต่อกัน ▶ 18

2.2.2 กลุ่มตัวอย่าง มีความสัมพันธ์กัน ▶ 19



1.1 การทดสอบเกี่ยวกับ *ค่าเฉลี่ย* 1 กลุ่มตัวอย่าง

ต้องการทดสอบความเชื่อที่ว่า *คนขับรถแท็กซี่ใน เขตกรุงเทพมหานคร น่าจะมีรายได้เฉลี่ยตกวันละ 1000 บาท*

H_0 : รายได้คนขับรถแท็กซี่ประมาณวันละ 1000 บาท หรือ $\mu = 150$

H_1 : รายได้คนขับรถแท็กซี่ไม่เท่ากับ 1000 บาท หรือ $\mu \neq 150$

วิธีการทดสอบสามารถจำแนกได้ 2 กรณีดังนี้

- *กรณีทราบการกระจายของข้อมูล (ทราบค่า σ)*

ตัวทดสอบคือ $z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$

- *กรณี ไม่ ทราบการกระจายของข้อมูล (ไม่ทราบค่า σ)*

ตัวทดสอบคือ $t = \dfrac{\bar{X} - \mu}{s / \sqrt{n}}$

1.2.1 การทดสอบเกี่ยวกับ *ค่าเฉลี่ย 2 กลุ่มตัวอย่าง อิสระกัน* 14

ต้องการทดสอบ *ค่าใช้จ่ายของนักศึกษา 2 คณะ*

H_0 : ค่าใช้จ่ายของนิสิต 2 คณะ *ไม่* แตกต่างกัน

H_1 : ค่าใช้จ่ายของนิสิต 2 คณะ แตกต่างกัน

วิธีการทดสอบสามารถจำแนกได้ 2 กรณีดังนี้

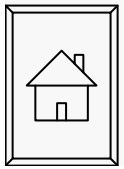
- *กรณีทราบการกระจายของข้อมูลทั้ง 2 กลุ่ม (ทราบค่า σ)*

  ตัวทดสอบคือ Z

- *กรณี ไม่ ทราบการกระจายของข้อมูล (ไม่ทราบค่า σ)*

  ตัวทดสอบคือ t ซึ่งจะต้องเลือกใช้ต่างกันดังนี้คือ

  - เมื่อทราบว่ามีการกระจายที่แตกต่างกัน(แต่ไม่ทราบค่าจริง)
  - เมื่อทราบว่ามีการกระจาย ไม่ แตกต่างกัน(แต่ไม่ทราบค่าจริง)



แต่ถ้าไม่ทราบ การกระจายแตกต่างกันหรือไม่ ต้องทดสอบก่อนด้วย **F test**

1.2.2 การทดสอบเกี่ยวกับ *ค่าเฉลี่ย 2 กลุ่มที่ สัมพันธ์กัน*

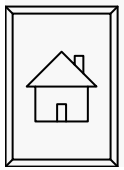
ต้องการทดสอบ ***การอบรมมีผลทำให้คะแนนแตกต่างกันหรือไม่***

H_0 : คะแนนเฉลี่ย ก่อน และ หลัง การอบรม *ไม่* แตกต่างกัน

H_1 : คะแนนเฉลี่ย ก่อน และ หลัง การอบรม แตกต่างกัน

วิธีการทดสอบ

ตัวทดสอบคือ t แบบจับคู่ (Paired t–test)



1.3 การทดสอบเกี่ยวกับ *ค่าเฉลี่ย* ตั้งแต่ 3 กลุ่มขึ้นไป

ต้องการทดสอบ *ราคาสินค้า 4 ยี่ห้อ*

H_0 : ราคาเฉลี่ยของสินค้าทั้ง 4 ยี่ห้อไม่แตกต่างกัน

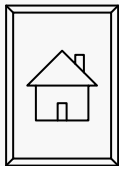
H_1 : ราคาเฉลี่ยของสินค้าทั้ง 4 ยี่ห้อแตกต่างกัน

วิธีการทดสอบ ใช้ตัวทดสอบคือ F

โดยการวิเคราะห์ความแปรรวน(Analysis Of Variance : ANOVA)

ข้อกำหนดของการวิเคราะห์ความแปรปรวน

- *ข้อมูลของแต่ละกลุ่มจะต้องมาจากประชากรที่มีการกระจาย(σ ไม่แตกต่างกัน*
- *ข้อมูลแต่ละกลุ่มต้องมาจากประชากรที่มีการแจกแจงแบบปกติ*
- *ตัวอย่างที่เลือกมาแต่ละกลุ่มจะต้องเป็นอิสระต่อกัน*



2.1 การทดสอบเกี่ยวกับ *ค่าสัดส่วน* 1 กลุ่มตัวอย่าง

- *กรณีข้อมูลมีค่าเป็นไปได้เพียง 2 ค่า เช่น เพศ ชาย & หญิง*

เช่น ต้องการทดสอบว่า *ในการผลิตสินค้าจะได้มาตรฐานไม่ต่ำกว่า 90%*

H_0 : จำนวนสินค้าที่ได้มาตรฐานมีไม่ต่ำกว่า 90 %

H_1 : จำนวนสินค้าที่ได้มาตรฐานมีต่ำกว่า 90 %

วิธีการทดสอบใช้ **Binomial** แต่ถ้า **n** มีมาก ใช้ **Z** หรือ **Poisson**

- *กรณีข้อมูลเป็นไปได้ตั้งแต่ 2 ค่าขึ้นไป เช่น อาชีพ , สถานะภาพ*

เช่น ต้องการทดสอบว่า *ยอดขาย TV 3 ยี่ห้อเป็นอัตราส่วน* 3 : 5 : 2

H_0 : ยอดขาย TV 3 ยี่ห้อ เป็น อัตราส่วน 3 : 5 : 2

H_1 : ยอดขาย TV 3 ยี่ห้อ ไม่เป็น อัตราส่วน 3 : 5 : 2

วิธีการทดสอบใช้ χ^2

2.2.1 การทดสอบ *ค่าสัดส่วน* ตั้งแต่ 2 กลุ่มตัวอย่าง *อิสระกัน*

- ***กรณีข้อมูลมีค่าเป็นไปได้เพียง 2 ค่าทั้ง 2 กลุ่ม เช่น เพศ & ความชอบ***

  เช่น ต้องการทดสอบว่า *เพศ ชาย & หญิง มีความ ชอบ & ไม่ชอบ ไม่ต่างกัน*

  H_0 : สัดส่วนของเพศ ชาย & หญิง มี ความชอบ ไม่ แตกต่างกัน

  H_1 : สัดส่วนของเพศ ชาย & หญิง มี ความชอบ แตกต่างกัน

  วิธีการทดสอบใช้ Z

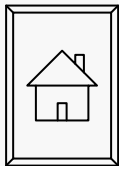
- ***กรณีข้อมูลเป็นไปได้ตั้งแต่ 2 ค่าขึ้นไป เช่น อาชีพ , สถานะภาพ***

  ต้องการทดสอบว่า *จำนวนผู้ชมรายการแต่ละภาคมีสัดส่วนไม่แตกต่างกัน*

  H_0 : สัดส่วนของผู้ชมรายการแต่ละภาค ไม่ แตกต่างกัน

  H_1 : สัดส่วนของผู้ชมรายการแต่ละภาคแตกต่างกัน

  วิธีการทดสอบใช้ χ^2



2.2.2 การทดสอบ *ค่าสัดส่วน* ตั้งแต่2กลุ่มตัวอย่างที่ *สัมพันธ์กัน*

ต้องการทดสอบว่า

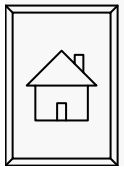
สัดส่วนของผู้ที่พอใจสินค้าก่อน/หลังการปรับปรุงแตกต่างกันหรือไม่

H_0 : สัดส่วนผู้ที่พอใจ ก่อน และ หลัง การปรับปรุง *ไม่* แตกต่างกัน

H_1 : สัดส่วนผู้ที่พอใจ ก่อน และ หลัง การปรับปรุง แตกต่างกัน

วิธีการทดสอบ

ตัวทดสอบคือ χ^2



การหาความสัมพันธ์สำหรับ 2 ตัวแปร (Associations) 20

1. กรณีทั้ง 2 ตัวแปร ไม่สามารถคำนวณได้

1.1 แบบ Nominal 21

เพศ & อาชีพ : ระดับการศึกษา & สถานะภาพ

1.2 แบบ Ordinal 22

ระดับความพึงพอใจมารยาทคนขับ & ระดับความพึงพอใจการแต่งกาย

2. กรณีทั้ง 2 ตัวแปร สามารถคำนวณได้ทั้งคู่ 23

รายได้ & รายจ่าย : งบโฆษณา & ยอดขาย

3. กรณีสามารถคำนวณได้เพียง 1 ตัวแปร 24

รายได้ & เพศ : ยอดขาย & สถานที่จำหน่าย

1.1 กรณีทั้ง 2 ตัวแปร ไม่สามารถคำนวณได้ *แบบ Nominal*

หาความสัมพันธ์ระหว่าง รสชาติลูกอมที่ชอบ & วัย

H_0 : รสชาติลูกอมที่ชอบ *ไม่ขึ้นกับ* วัย

H_1 : รสชาติลูกอมที่ชอบ *ขึ้นกับ* วัย

	เปรี้ยว	หวาน	เค็ม	รวม
เด็ก	30	40	10	80
ผู้ใหญ่	30	60	30	120
รวม	60	100	40	200

ถ้า รสชาติลูกอมที่ชอบ *ไม่ขึ้นกับ* วัย แล้ว หมายความว่า ?????

การชอบรสชาติต่าง ๆจะต้องคงที่คือ เปรี้ยว หวาน เค็ม ควรจะเป็นอัตราส่วน

60 : 100 :40 หรือ 3 : 5 : 2 นั่นคือไม่ว่า ในวัยไหน ๆ คน 10 คนจะชอบ เปรี้ยว 3 คน หวาน 5 คน เค็ม 2 คน ดังนั้นจะได้ค่าความถี่ที่ควรจะเป็นคือ

1.2 กรณีทั้ง 2 ตัวแปร ไม่สามารถคำนวณได้ *แบบ Ordinal*

หาความสัมพันธ์ระหว่าง ระดับความพึงพอใจ คนขับ & กระเป๋า

ค่าสถิติที่เกี่ยวข้องกับการหาความสัมพันธ์แบบนี้มี 2 ค่าคือ

(1) ค่า Spearman Correlation (r_s) สำหรับแสดงระดับความสัมพันธ์

ซึ่งค่า r_s จะมีค่าอยู่ระหว่าง –1 ถึง 1

ค่า + แสดงความสัมพันธ์มีทิศทาง *ไปทางเดียวกัน*

ค่า – แสดงความสัมพันธ์มีทิศทาง *ไปในทางตรงกันข้าม*

(2) ค่า t–test สำหรับตรวจสอบว่ามีความสัมพันธ์กันจริงหรือไม่ โดยกำหนดสมมติฐานทางสถิติดังนี้

H_0 : ระดับความพึงพอใจ คนขับ และ กระเป๋า ไม่มีความสัมพันธ์เชิงเส้นตรงต่อกัน

H_1 : ระดับความพึงพอใจ คนขับ และ กระเป๋า มีความสัมพันธ์เชิงเส้นตรงต่อกัน

2. กรณีทั้ง 2 ตัวแปร สามารถคำนวณได้ทั้งคู่

หาความสัมพันธ์ระหว่าง จำนวนลูกค้าที่เข้าห้าง & ยอดขาย

ค่าสถิติที่เกี่ยวข้องกับการหาความสัมพันธ์แบบนี้มี 2 ค่าคือ

(1) ค่า **Pearson Correlation (r)** สำหรับแสดงระดับความสัมพันธ์

ซึ่งค่า r จะมีค่าอยู่ระหว่าง –1 ถึง 1

ค่า + แสดงความสัมพันธ์มีทิศทาง *ไปทางเดียวกัน*

ค่า – แสดงความสัมพันธ์มีทิศทาง *ไปในทางตรงกันข้าม*

(2) ค่า **t–test** สำหรับตรวจสอบว่ามีความสัมพันธ์กันจริงหรือไม่ โดยกำหนดสมมติฐานทางสถิติดังนี้

H_0 : จำนวนลูกค้าที่เข้าห้าง *ไม่มีความสัมพันธ์แบบเส้นตรงกับ ยอดขาย*

H_1 : จำนวนลูกค้าที่เข้าห้าง *มีความสัมพันธ์แบบเส้นตรงกับ ยอดขาย*

3. กรณีสามารถคำนวณได้เพียง 1 ตัวแปร

หาความสัมพันธ์ระหว่าง ค่าใช้จ่าย & เพศ หรือ กับ อาชีพ

ค่าสถิติที่เกี่ยวข้องกับการหาความสัมพันธ์แบบนี้มี 2 ค่าคือ

(1) ค่า **Eta** สำหรับแสดงระดับความสัมพันธ์

(2) ค่า สถิติ ที่ใช้ในการทดสอบว่าตัวแปรทั้ง 2 มีความสัมพันธ์กันจริงหรือไม่ โดยตรงนั้นไม่มี แต่สามารถที่จะใช้วิธีของการทดสอบค่าเฉลี่ยแทน โดยใช้ตัวแปรที่คำนวณได้มาทดสอบค่าเฉลี่น จำแนกตามกลุ่มของตัวแปรที่คำนวณได้ซึ่งจะใช้ตัวสถิติที่แตกต่างกัน ขึ้นอยู่กับจำนวนกลุ่ม

กรณีตัวแปรที่คำนวณไม่ได้จำแนกได้ 2 กลุ่ม เช่น ค่าใช้จ่าย & เพศ จะใช้ทดสอบด้วย **t–test**

กรณีตัวแปรที่คำนวณไม่ได้จำแนกได้ >2 กลุ่ม เช่น ค่าใช้จ่าย & อาชีพ จะใช้ทดสอบด้วย **F–test**

1. ใช้ข้อมูลในอดีตที่ต่อเนื่องกัน Time Series Analysis

2. ใช้ข้อมูลที่มีความสัมพันธ์กัน Regression Analysis

2.1 Linear Regression

- Simple Linear Regression
- Multiple Linear Regression

2.2 Non-Linear Regression

2.3 Probit/Logistic Regression

การวิเคราะห์ข้อมูลแบบหลายตัวแปร(Multivariate)

1. **Regression Analysis :** Hierarchical : Probit/Logistic
2. **Path Analysis**
3. **Factor Analysis**
4. **Discriminant Analysis**
5. **Cluster Analysis**
6. **Canonical analysis**