บทที่ 5

หลักการประมาณค่าและการทดสอบสมมติฐาน

88520159

PROBABILITY AND STATISTICS FOR COMPUTING

ทบทวนสัญลักษณ์

พารามิเตอร์ (Parameter) สถิติ (Statistic)

	ประชากร	ตัวอย่าง
ขนาด	N	n
ค่าเฉลี่ย	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{n-1}$
ความแปรปรวน	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
ส่วนเบี่ยงเบนมาตรฐาน	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$
สัดส่วน	P	$\hat{p}$

พารามิเตอร์และสถิติ

พารามิเตอร์ (Parameter)

$$\boxed{\mu, \sigma^2, \sigma, P}$$

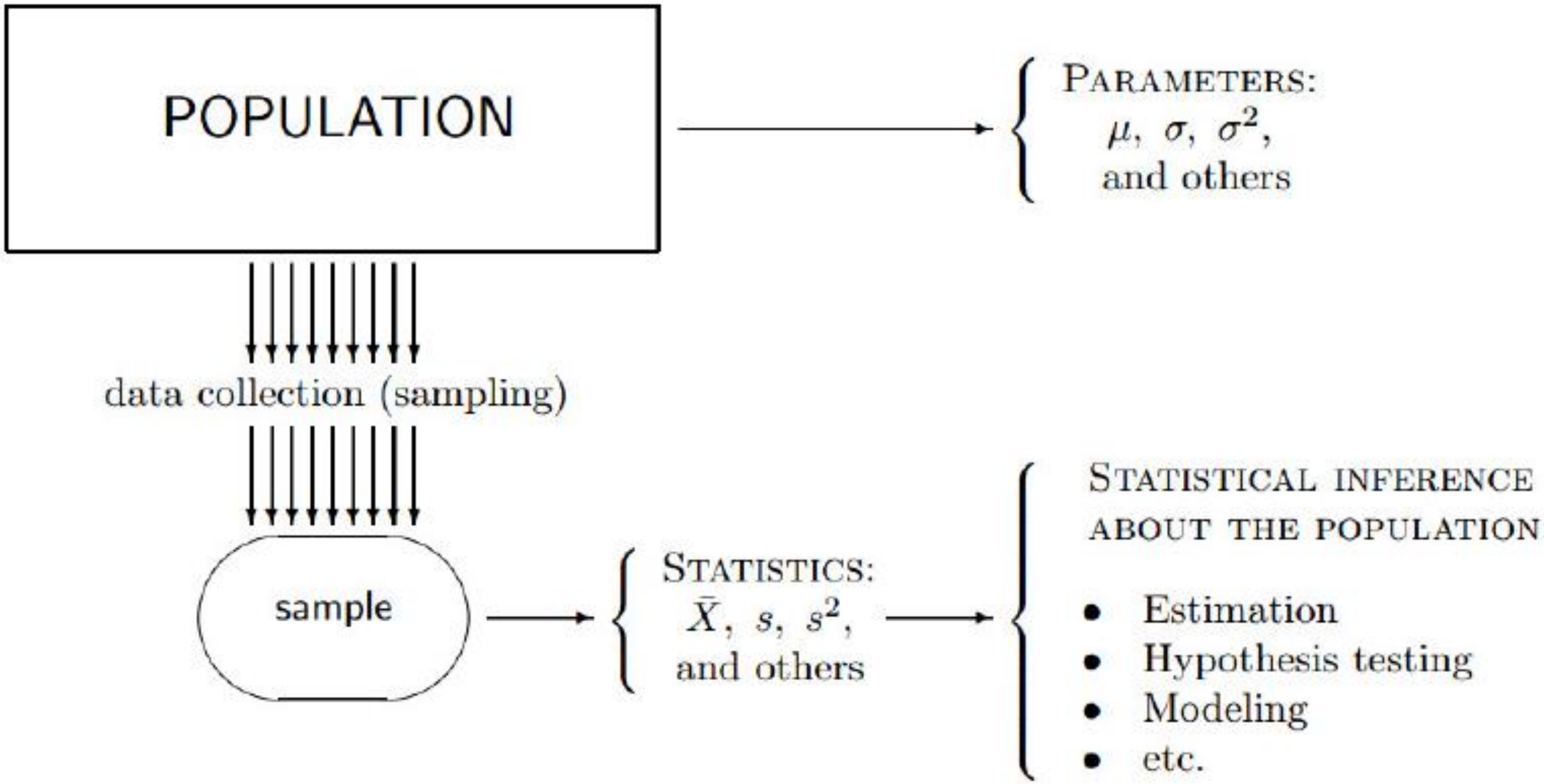
เป็นค่าที่คำนวณจากข้อมูลของประชากร เป็นค่าคงที่ที่แสดงลักษณะของประชากร

สถิติ (Statistic)

$$\boxed{\bar{x}, s^2, s, \hat{p}}$$

เป็นค่าที่คำนวณจากข้อมูลตัวอย่าง ไม่ใช่ค่าคงที่ แต่เป็นตัวแปรสุ่ม

พารามิเตอร์และสถิติ



สถิติเชิงอนุมาน

สถิติเชิงอนุมาน
(inferential Statistic)

การประมาณค่า
(Estimation)

การทดสอบสมมติฐาน
(Hypothesis Testing)

การศึกษาลักษณะบางประการของประชากร โดยอาศัยข้อมูลจากตัวอย่าง

1. การประมาณค่าพารามิเตอร์ (Parameter Estimation)

การประมาณค่า (Estimation) คือ การประมาณค่าพารามิเตอร์ของประชากร โดยใช้ข้อมูลที่ได้จากชุดตัวอย่าง

- เช่น การประมาณยอดขายเฉลี่ยรายเดือนของบริษัทในปีหน้าเพื่อการวางแผนการผลิต
- ต้องทำอย่างมีหลักเกณฑ์เพื่อให้ค่าประมาณที่ได้มีความคลาดเคลื่อนจากค่าพารามิเตอร์ที่แท้จริงน้อยที่สุด

- ทางสถิติการประมาณค่าพารามิเตอร์สามารถแบ่งได้เป็น 2 ชนิด คือ

  1.1 การประมาณค่าแบบจุด (point estimation)

  1.2 การประมาณค่าแบบช่วง (Interval estimation)

1.1 การประมาณค่าแบบจุด (Point Estimation)

การประมาณแบบจุด เป็นการประมาณค่าพารามิเตอร์ด้วยค่าประมาณเพียงค่าเดียว โดยใช้ข้อมูลจากตัวอย่าง

ตัวประมาณแบบจุดของ

ค่าเฉลี่ยของประชากร ($\boldsymbol{\mu}$) ⟵ ค่าเฉลี่ยของตัวอย่าง ($\overline{\boldsymbol{x}}$)

ค่าสัดส่วนของประชากร ($\boldsymbol{P}$) ⟵ ค่าสัดส่วนของตัวอย่าง ($\widehat{\boldsymbol{p}}$)

ค่าความแปรปรวนของประชากร ($\boldsymbol{\sigma}^{2}$) ⟵ ค่าความแปรปรวนของตัวอย่าง ($\boldsymbol{s}^{2}$)

ค่าส่วนเบี่ยงเบนมาตรฐานของประชากร ($\boldsymbol{\sigma}$) ⟵ ค่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง ($\boldsymbol{s}$)

1.1 การประมาณค่าแบบจุด (Point Estimation)

- ค่าประมาณแบบจุดนี้อาจจะมีค่าเท่าหรือไม่เท่ากับค่าพารามิเตอร์ก็ได้
- มีโอกาสที่จะคลาดเคลื่อนไปจากค่าพารามิเตอร์ที่แท้จริงได้มาก
- ถ้าชุดตัวอย่างเป็นตัวแทนที่ไม่ดีของประชากรแล้ว ค่าประมาณที่ได้จะคลาดเคลื่อนไปจากค่าพารามิเตอร์ที่แท้จริงมาก
- ดังนั้นการประมาณชนิดนี้จึงมีความเสี่ยงสูง

1.2 การประมาณค่าแบบช่วง (Interval Estimation)

เมื่อต้องการประมาณค่าพารามิเตอร์ θ (พารามิเตอร์ใดๆ) มีความเป็นไปได้ที่

$\hat{\theta} = \theta$ ตัวสถิติจะมีค่าเท่ากับค่าพารามิเตอร์

และอาจเกิดความคลาดเคลื่อนจากการสุ่ม (sampling error)

$$\text{sampling error} = \theta - \hat{\theta}$$

<u>ตัวอย่าง</u>

- ซื้ออินเตอร์เน็ตความเร็ว 10MB ต่อวินาที เร็ว 10MB ต่อวินาที จริงหรือไม่
- การพยากรณ์อากาศ แม่นยำมากน้อยเพียงใด
- ความน่าจะเป็นที่ค่านี้จะเข้าใกล้ความจริงมีมากเพียงใด

1.2 การประมาณค่าแบบช่วง (Interval Estimation)

- เมื่อประมาณค่าพารามิเตอร์ θ แล้ว ค่าที่แท้จริงของ θ มีค่าเป็นเท่าใด
- นักสถิติสามารถใช้การประมาณช่วงความเชื่อมั่น ซึ่งจะครอบคลุมค่าพารามิเตอร์จากข้อมูลที่เก็บได้จากตัวอย่าง
- การประมาณค่าพารามิเตอร์แบบช่วง เป็นการใช้ช่วงประมาณที่เรียกว่า *ช่วงความเชื่อมั่น (confidence interval)*

$$L < \theta < U$$

โดย L คือค่าประมาณต่ำสุด และ U คือค่าประมาณสูงสุดของช่วงการประมาณ

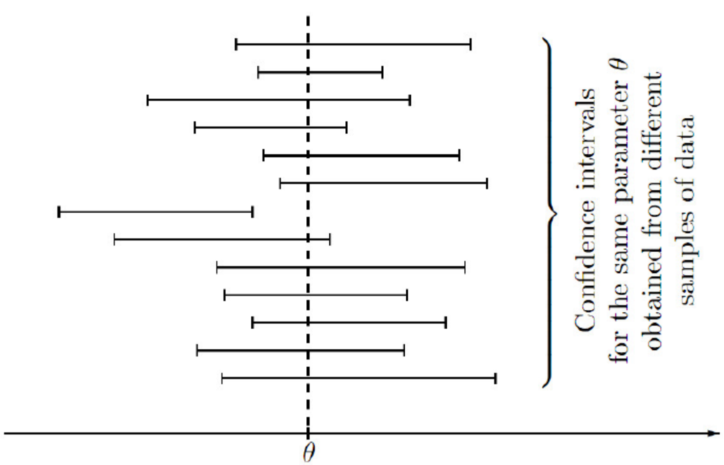
สัมประสิทธิ์ความเชื่อมั่น (Confidence coefficient)

สัมประสิทธิ์ความเชื่อมั่น (Confidence coefficient) หมายถึง ความน่าจะเป็นที่ช่วงประมาณจะครอบคลุมค่าพารามิเตอร์ของประชากร โดยมีค่าเท่ากับ $1-\alpha$ นั่นคือ $P(L \leq \theta \leq U) = 1-\alpha$

เช่น ค่าสัมประสิทธิ์ความเชื่อมั่น $(1-\alpha)$ เท่ากับ 0.90

- โอกาสที่จะครอบคลุมค่าพารามิเตอร์ที่เราสนใจ 90%
- โอกาสที่ไม่ครอบคลุม 10%

หมายความว่า สุ่มตัวอย่าง 100 ครั้ง มี 90 ครั้ง ที่ค่าประมาณแบบช่วงจะคลุม θ และอีก 10 ครั้ง ไม่คลุมค่า θ



และ เรียก $\boldsymbol{\alpha}$ ว่าระดับนัยสำคัญ

3. การทดสอบสมมติฐาน (Hypothesis test)

สมมติฐาน คือ ข้อสมมติที่เกิดจากความเชื่อ ซึ่งอาจจะจริงหรือไม่ก็ได้

สมมติฐานทางสถิติ คือ คำกล่าว หรือ ข้อความเกี่ยวกับค่าพารามิเตอร์

การทดสอบสมมติฐาน เป็นกระบวนการเพื่อใช้ทดสอบว่า พารามิเตอร์ของประชากรที่กล่าวอ้างเป็นจริงหรือไม่

<u>ตัวอย่าง</u>

- นักวิจัยทางการแพทย์ท่านหนึ่งกล่าวว่า อุณหภูมิเฉลี่ยของร่างกายของคนที่มีสุขภาพดีนั้นไม่เท่ากับ 98.6 F
- นักข่าวคนหนึ่งกล่าวว่าผู้ขับขี่รถยนต์ในกรุงเทพฯ ส่วนใหญ่ขับรถฝ่าไฟแดง
- เครื่องมือวัดความสูงในเครื่องบินชนิดใหม่ที่นำมาใช้นั้นดีกว่าชนิดเดิม เนื่องจากมีความแปรปรวนของความคลาดเคลื่อนต่ำกว่า 0.2

ส่วนประกอบของการทดสอบสมมติฐาน

1. สมมติฐานหลัก (Null Hypothesis, H_0) คือข้อความที่ว่า พารามิเตอร์ของประชากรมีค่าเท่ากับค่าที่กล่าวอ้าง (ใช้เครื่องหมาย =, $\leq$, $\geq$)

2. สมมติฐานรอง (Alternative Hypothesis, H_1) คือข้อความที่ว่า พารามิเตอร์มีค่าแตกต่างจากสมมติฐานหลัก (ใช้เครื่องหมาย $\neq$, $>$, $<$)

- จะทดสอบสมมติฐานหลักในความหมายที่สมมติว่าสมมติฐานหลักนี้เป็นจริงหรือถูกต้อง
- จะได้ข้อสรุปของประการ คือ ปฏิเสธ H_0 หรือไม่สามารถปฏิเสธ H_0 ได้ หรือยอมรับ H_0 นั่นเอง

ส่วนประกอบของการทดสอบสมมติฐาน

ตัวอย่าง ให้ตั้งสมมติฐานหลักและสมมติฐานรองสำหรับข้อความต่อไปนี้

1. ความสูงเฉลี่ยของนักบาสเกตบอลอาชีพนั้นไม่เกิน 7 ฟุต

H_0 : H_1 :

2. สัดส่วนของผู้ขับขี่รถยนต์ยอมรับว่าเคยขับรถฝ่าไฟแดงนั้นมากกว่า 0.5

H_0 : H_1 :

3. ค่าเบี่ยงเบนมาตรฐานของคะแนน IQ ของนักแสดงชายเท่ากับ 15

H_0 : H_1 :

ค่าสถิติทดสอบ (Test statistic)

ค่าสถิติทดสอบ คือ ค่าที่คำนวณมาจากข้อมูลตัวอย่าง และใช้ในการตัดสินใจเกี่ยวกับการปฏิเสธสมมติฐานหลัก
โดยค่าสถิติทดสอบนี้ได้จากการเปลี่ยนค่าสถิติของตัวอย่าง (เช่น $\hat{p}, \bar{x}$ หรือ s) ให้เป็นค่ามาตรฐาน (เช่น z, t หรือχ^2)

- ดังนั้นเราจึงสามารถใช้ค่าสถิติทดสอบเพื่อหาว่าจากข้อมูลตัวอย่างที่มีอยู่เรามีหลักฐานเพียงพอที่จะขัดแย้งกับสมมติฐานหลักหรือไม่
- โดยใช้ความรู้เกี่ยวกับการแจกแจงค่าจากตัวอย่างของค่าสถิติ ช่วยในการหาค่าสถิติทดสอบที่ใช้ในการทดสอบสมมติฐาน

ค่าสถิติทดสอบ (Test statistic)

เขตวิกฤติหรือเขตปฏิเสธ H_0 คือ ช่วงของค่าสถิติทดสอบที่ทำให้เกิดการปฏิเสธสมมติฐานหลัก

ระดับนัยสำคัญ (α) คือความน่าจะเป็นที่ค่าสถิติทดสอบจะมีค่าอยู่ในเขตวิกฤติ เมื่อสมมติฐานหลักเป็นจริง ถ้าค่าสถิติทดสอบที่คำนวณได้มีค่าอยู่ในเขตวิกฤติเราจะตัดสินใจปฏิเสธสมมติฐานหลัก

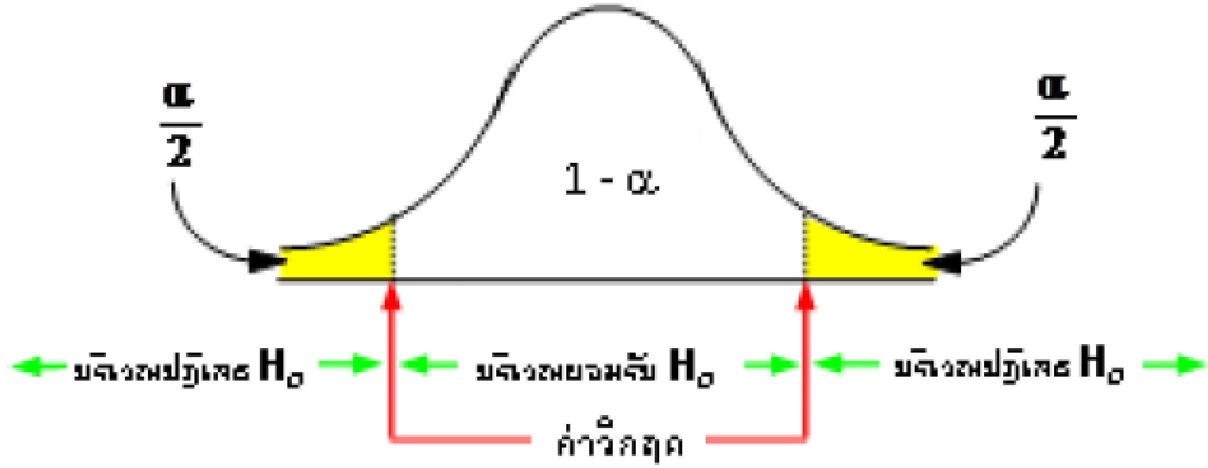
ค่าวิกฤติ คือ ค่าใด ๆ ที่ใช้ในการแบ่งเขตวิกฤติและเขตยอมรับ ค่าวิกฤตินี้ขึ้นอยู่กับลักษณะของสมมติฐานหลัก การแจกแจงค่าจากตัวอย่างของค่าสถิติที่ใช้และระดับนัยสำคัญ

การทดสอบสองทาง (Two-tailed test)

การทดสอบสองทาง เป็นการทดสอบสมมติฐาน ที่สมมติฐานรองเป็นเครื่องหมายไม่เท่ากับ (≠) โดยมีเขตวิกฤติอยู่ที่ปลายทั้งสองข้างของการแจกแจง

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

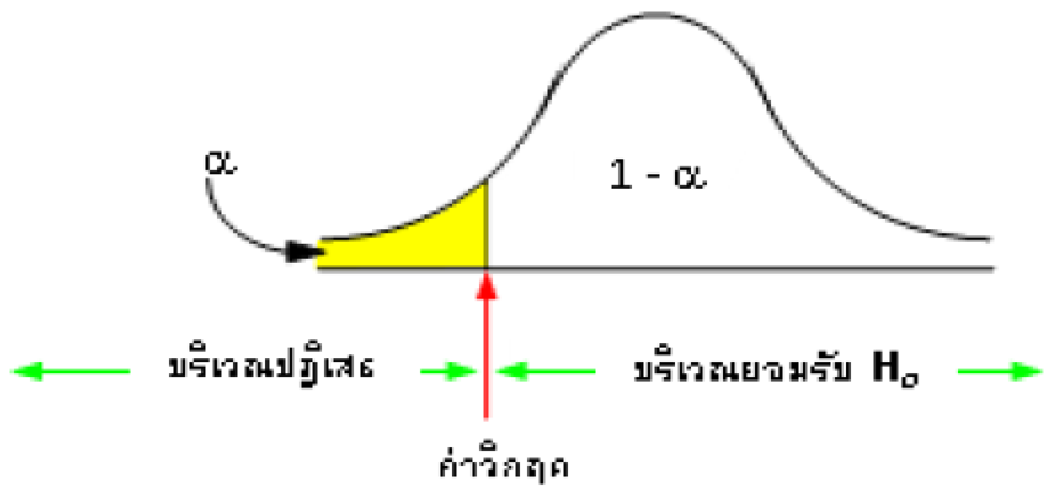


การทดสอบทางเดียวด้านซ้าย (Left-tailed test)

การทดสอบทางเดียวด้านซ้าย เป็นการทดสอบสมมติฐาน ที่สมมติฐานรองเป็นเครื่องหมายน้อยกว่า (<) โดยมีเขตวิกฤติอยู่ที่ปลายหางด้านซ้ายของการแจกแจง

$$H_0 : \theta \geq \theta_0$$

$$H_1 : \theta < \theta_0$$

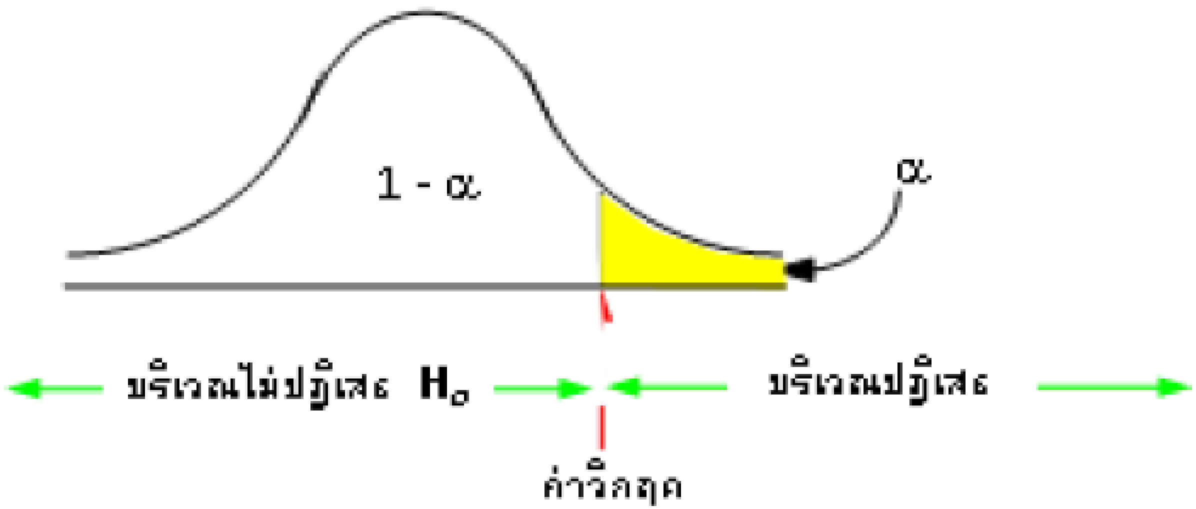


การทดสอบทางเดียวด้านขวา (Right-tailed test)

การทดสอบทางเดียวด้านขวา เป็นการทดสอบสมมติฐาน ที่สมมติฐานรองเป็นเครื่องหมายมากกว่า (>) โดยมีเขตวิกฤติอยู่ที่ปลายหางด้านขวาของการแจกแจง

$$H_0 : \theta \leq \theta_0$$

$$H_1 : \theta > \theta_0$$



การตัดสินใจและการสรุป

การทดสอบสมมติฐานจะทดสอบที่สมมติฐานหลักเสมอ ดังนั้นข้อสรุปเบื้องต้นในการทดสอบสมมติฐานจะเป็นไปได้สองแบบต่อไปนี้

1. ปฏิเสธสมมติฐานหลัก (Reject H_0)
2. ยอมรับสมมติฐานหลัก (Accept H_0)

การตัดสินใจที่จะปฏิเสธหรือยอมรับ H_0 มีเกณฑ์ดังนี้

ปฏิเสธ H_0 ถ้าค่าสถิติทดสอบมีค่าอยู่ภายในเขตวิกฤติ

ยอมรับ H_0 ถ้าค่าสถิติทดสอบไม่ได้อยู่ในเขตวิกฤติ

การตัดสินใจและการสรุป

ในการตัดสินใจและสรุปผล เมื่อทำการวิเคราะห์ข้อมูลด้วยโปรแกรม R จะใช้ค่า p-value ซึ่งเป็นค่าความน่าจะเป็นที่แปลงมาจากค่าสถิติทดสอบ โดยมีเกณฑ์ดังนี้

- ปฏิเสธ H_0 ถ้าค่า p-value มีค่าน้อยกว่า ระดับนัยสำคัญ
- ยอมรับ H_0 ถ้าค่า p-value มีค่ามากกว่าหรือเท่ากับ ระดับนัยสำคัญ

สรุปขั้นตอนการทดสอบสมมติฐาน

ขั้นตอนการทดสอบสมมติฐานมี 4 ขั้นตอนดังนี้

1. กำหนดมมมติฐานหลักและสมมติฐานรอง
2. กำหนดระดับนัยสำคัญ
3. กำหนดตัวสถิติทดสอบ และคำนวณหาค่า p-value
4. สรุปผล

การประมาณค่าและการทดสอบสมมติฐาน

การประมาณค่าและการทดสอบสมมติฐาน

1. สำหรับค่าเฉลี่ยประชากร 1 กลุ่ม
2. สำหรับค่าสัดส่วนของประชากร 1 กลุ่ม
3. สำหรับค่าความแปรปรวนประชากร 1 กลุ่ม
4. สำหรับค่าสัดส่วนของประชากร 2 กลุ่ม
5. สำหรับค่าความแปรปรวนประชากร 2 กลุ่ม
6. สำหรับค่าเฉลี่ยของประชากร 2 กลุ่มเป็นอิสระกัน
7. สำหรับค่าเฉลี่ยของประชากร 2 กลุ่มไม่เป็นอิสระกัน

บทที่ 6
การประมาณค่าและทดสอบสมมติฐานสำหรับค่าเฉลี่ยของประชากร 1 กลุ่ม

88520159

PROBABILITY AND STATISTICS FOR COMPUTING

1. การประมาณแบบจุดของค่าเฉลี่ยของประชากร 1 กลุ่ม

- ค่าประมาณแบบจุดของค่าเฉลี่ยของประชากร (μ)
  - แจกแจงปกติ ใช้ ค่าเฉลี่ยของตัวอย่าง ($\bar{x}$)
  - ไม่มีการแจกแจงปกติ ใช้ ค่ามัธยฐานของตัวอย่าง

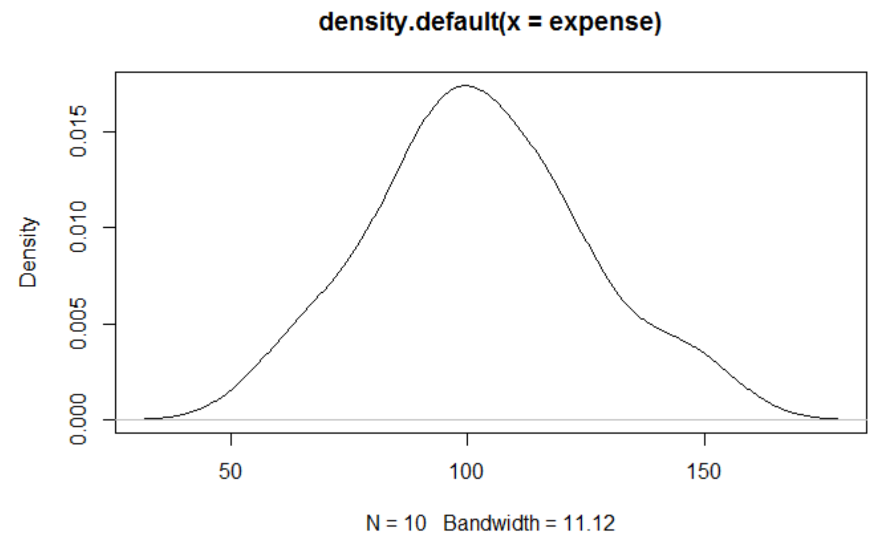
ตัวอย่าง การประมาณแบบจุดของค่าเฉลี่ยของประชากร

สุ่มตัวอย่างนิสิตจำนวน 10 คน เพื่อทำการประมาณค่าใช้จ่ายเฉลี่ยต่อวันของนิสิตคณะวิทยาการสารสนเทศ ได้ข้อมูลดังนี้

120 145 110 100 95 65 120 100 80 90

จากข้อมูลต้องการประมาณค่าเฉลี่ย

```
> expense =
c(120,145,110,100,95,65,120,100,80,90)
> plot(density(expense)
> mean(expense)
[1] 102.5
```



จากกราฟจะพบว่าข้อมูลค่าใช้จ่ายมีการแจกแจงสมมาตร ดังนั้น เราจะประมาณค่าใช้จ่ายเฉลี่ย ด้วยค่าเฉลี่ย จะได้ว่าค่าใช้จ่ายเฉลี่ยของนิสิตคณะวิทยาการสารสนเทศเป็น 102.5 บาทต่อวัน

2. การประมาณแบบช่วงค่าเฉลี่ยของประชากร 1 กลุ่ม

- ค่าประมาณแบบช่วงของค่าเฉลี่ยของประชากร (μ)
  - เมื่อทราบค่า σ^2 แจกแจงแบบใดก็ตาม ใช้ t test
  - เมื่อไม่ทราบค่า σ^2 และ $n \geq 30$ มีการแจกแจงปกติ ใช้ t test
  - เมื่อไม่ทราบค่า σ^2 และ $n < 30$

    หากข้อมูลมาจากการแจกแจงปกติ ใช้ t test

    หากข้อมูลไม่ได้มาจากการแจกแจงปกติ ใช้ sign test

การตรวจสอบการแจกแจงแบบปกติ

ในการทดสอบสมมติฐานเกี่ยวกับประชากร 1 กลุ่ม

- หากข้อมูลตัวอย่างมาจากประชากรที่มีการแจกแจงแบบปกติ หรือทราบค่า σ^2 ใช้สถิติที่อิงพารามิเตอร์ ได้แก่ t-test หรือ z-test
- หากข้อมูลตัวอย่างมาจากประชากรไม่มีการแจกแจงปกติ จะใช้สถิติไม่อิงพารามิเตอร์ได้แก่ sign test
- ใช้การทดสอบการแจกแจงปกติ Anderson Darling ด้วยโปรแกรม R

```
> install.packages("nortest")

> library(nortest)
```

การตรวจสอบการแจกแจงแบบปกติ

- ฟังก์ชันที่ใช้ในการทดสอบการแจกแจงปกติ Anderson Darling คือ ad.test()

ขั้นตอนในการทดสอบสมมติฐานมีดังนี้

1. ตั้งสมมติฐานในการทดสอบ
   - H_0 : ข้อมูลมาจากประชากรที่มีการแจกแจงปกติ
   - H_1 : ข้อมูลไม่ได้มาจากประชากรที่มีการแจกแจงปกติ
2. กำหนดระดับนัยสำคัญ (α) เช่น 0.05, 0.01, 0.1
3. คำนวณค่า p-value
4. สรุปผล โดยพิจารณาค่า p-value ดังนี้
   - ถ้า p-value $\geq \alpha$ ยอมรับ H_0 แสดงว่าข้อมูลมีการแจกแจงปกติ
   - ถ้า p-value $< \alpha$ ปฏิเสธ H_0 แสดงว่าข้อมูลไม่มีการแจกแจงปกติ

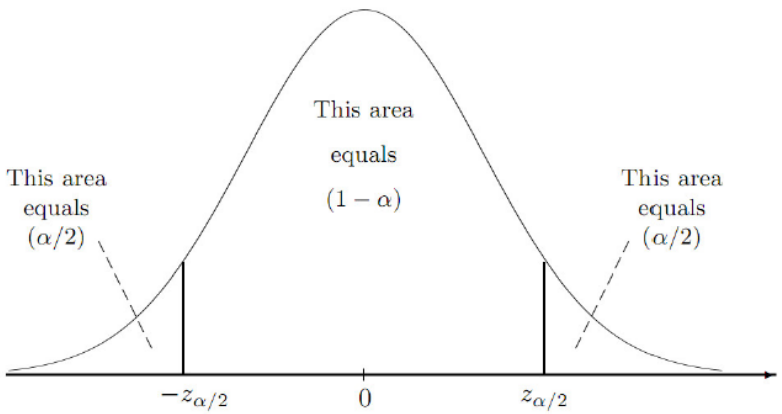
ช่วงความเชื่อมั่น $(1-\alpha)$ 100% ของ μ เมื่อทราบค่า σ^2

- เมื่อทราบค่าความแปรปรวนของประชากร (σ^2)
- หรือตัวอย่างมีขนาดใหญ่ $(n \geq 30)$
- ตัวอย่างถูกสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติหรือการแจกแจงแบบใดก็ตาม

ช่วงความเชื่อมั่น $(1-\alpha)$ 100% ของ μ เมื่อทราบค่า σ^2

$$\left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

- ใช้สถิติ t test ในการประมาณช่วงความเชื่อมั่นของค่าเฉลี่ย



ช่วงความเชื่อมั่น $(1-\alpha)$ 100% ของ μ เมื่อไม่ทราบค่า σ^2 และ $n \geq 30$

- ไม่ทราบค่า σ^2
- และ ตัวอย่างมีขนาดใหญ่ ($n \geq 30$)
- ใช้ค่าความแปรปรวนของตัวอย่าง (s^2) เป็นค่าประมาณของ σ^2
- ยิ่งตัวอย่างมีขนาดใหญ่ขึ้น ($n \rightarrow \infty$) การแจกแจงค่าจาก ตัวอย่าง ของ $\bar{x}$ จะเข้าใกล้การแจกแจงแบบปกติมากขึ้น

ช่วงความเชื่อมั่น $(1-\alpha)$ 100% ของ μ เมื่อไม่ทราบค่า σ^2 และ $n \geq 30$

$$\left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

- ใช้สถิติ t test ในการประมาณช่วงความเชื่อมั่นของค่าเฉลี่ย

ช่วงความเชื่อมั่น $(1-\alpha)$ 100% ของ μ เมื่อไม่ทราบค่า σ^2 และ $n < 30$

- ไม่ทราบค่า σ^2
- และตัวอย่างสุ่มขนาดเล็ก (n < 30)
- ตัวอย่างถูกสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติหรือใกล้เคียงปกติ
- จะมีการแจกแจงแบบที (t-distribution) ที่มีองศาเสรีเท่ากับ n – 1

ช่วงความเชื่อมั่น $(1-\alpha)$ 100% ของ μ เมื่อไม่ทราบค่า σ^2 และ $n < 30$

โดยประชากรมีการแจกแจงแบบปกติ

$$\left(\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right)$$

- ใช้สถิติ t test ในการประมาณช่วงความเชื่อมั่นของค่าเฉลี่ย

ค่าประมาณแบบช่วงของค่าเฉลี่ยประชากร ($n <$ 30, ไม่ปกติ)

- ข้อมูลไม่เป็นไปตามเงื่อนไข ทั้ง 3 สูตรข้างต้น
- ข้อมูลไม่ได้มาจากการแจกแจงปกติ
- ตัวอย่างมีขนาดเล็ก (n < 30)
- การประมาณค่ากลางของข้อมูลด้วยค่าเฉลี่ยจะไม่เหมาะสม
- ใช้ค่ามัธยฐานในการนำเสนอค่าประมาณแบบจุด
- ใช้สถิติ sign test ในการประมาณช่วงความเชื่อมั่นของค่ามัธยฐาน

```
> install.packages("signmedian.test")

> library(signmedian.test)
```

คำสั่งในการทดสอบ t test ด้วยโปรแกรม R

- ค่าประมาณแบบช่วงของค่าเฉลี่ยของประชากร (μ)

```
t.test(x, conf.level = .....)
```

- ทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร 1 กลุ่ม

```
t.test(x, mu = ....., alternative = .....)
```

x คือ ข้อมูลตัวเลขเป็น vector

conf.level คือ สัมประสิทธิ์ความเชื่อมั่น เช่น 0.95

mu คือ ค่าเฉลี่ย

alternative คือ สัญลักษณ์ของสมมติฐานรอง "two.sided" หรือ "less" หรือ "greater"

คำสั่งในการทดสอบ sign test ด้วยโปรแกรม R

- **ค่าประมาณแบบช่วงของค่ามัธยฐานประชากร (μ)**

```
signmedian.test(x, conf.level = .....)
```

- **ทดสอบสมมติฐานเกี่ยวกับค่ามัธยฐานประชากร 1 กลุ่ม**

```
signmedian.test(x, mu = ...., alternative = .....)
```

x คือ ข้อมูลตัวเลขเป็น vector

conf.level คือ สัมประสิทธิ์ความเชื่อมั่น เช่น 0.95

mu คือ ค่าเฉลี่ย

alternative คือ สัญลักษณ์ของสมมติฐานรอง "two.sided" หรือ "less" หรือ "greater"

ตัวอย่างที่ 1

นักวิจัยคนหนึ่งต้องการประมาณระยะเวลาเฉลี่ยในการสูบบุหรี่ของคนไข้อายุ 40-45 ปีที่ป่วยเป็นโรคถุงลมโป่งพอง (หน่วย: ปี) โดยสุ่มตัวอย่างผู้ป่วยมา 10 คนแล้วสอบถามระยะเวลาที่เขาสูบบุหรี่ได้ข้อมูลดังนี้

22 21 19 25 24 26 23 21 23 22

ให้นิสิตช่วยนักวิจัยคนนี้ประมาณระยะเวลาการสูบบุหรี่เฉลี่ยของคนไข้โรคถุงลมโป่งพองที่มีอายุ 40-45 ปี ที่ระดับความเชื่อมั่น 95%

วิเคราะห์โจทย์

- ต้องการประมาณแบบช่วงของค่าเฉลี่ย
- ไม่ทราบความแปรปรวนของประชากร σ^2
- จำนวนข้อมูลมีขนาดเล็กเป็น 10 ค่า

ต้องตรวจสอบการแจกแจงของข้อมูลว่ามีการแจกแจงปกติหรือไม่

ตัวอย่างที่ 1

ทดสอบสมมติฐานเพื่อตรวจสอบการแจกแจงปกติ

1. ตั้งสมมติฐานในการทดสอบ

H_0 : ข้อมูลระยะเวลาในการสูบบุหรี่มาจากประชากรที่มีการแจกแจงปกติ

H_1 : ข้อมูลระยะเวลาในการสูบบุหรี่ไม่ได้มาจากประชากรที่มีการแจกแจงปกติ

2. กำหนดระดับนัยสำคัญ (α) = 0.05

3.คำนวณค่า p-value

p-value=0.9067 ซึ่งมีค่ามากกว่าระดับนัยสำคัญ

4. สรุปผล

ยอมรับ H_0 แสดงว่าที่ระดับนัยสำคัญ 0.05

ข้อมูลระยะเวลาในการสูบบุหรี่มาจากประชากรที่มีการแจกแจงปกติ

```
> smoke=c(22,21,19,25,24,26,23,21,23,22)

> ad.test(smoke)

Anderson-Darling normality test

data: smoke

A = 0.16909, p-value = 0.9067
```

ตัวอย่างที่ 1

- ข้อมูลระยะเวลาในการสูบบุหรี่มาจากประชากรที่มีการแจกแจงปกติ
- ดังนั้น เลือกใช้สถิติ t test เพื่อประมาณแบบช่วงของค่าเฉลี่ย

```
> t.test(smoke, conf.level=0.95)

One Sample t-test

data: smoke

t = 34.599, df = 9, p-value = 6.95e-11

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

21.12237 24.07763

sample estimates:

mean of x

22.6
```

สรุป ที่ระดับความเชื่อมั่น 95% ระยะเวลาการสูบบุหรี่เฉลี่ยของคนไข้โรคถุงลมโป่งพองที่มีอายุ 40-45 ปีมีค่าอยู่ในช่วง 21.12 ถึง 24.08 ปี

ตัวอย่างที่ 2

ในการสำรวจต้นโพธิ์ที่มีอายุ 4 ปีมา 50 ต้น ได้ค่าความกว้างของลำต้น (หน่วย : นิ้ว) ดังนี้

8.7, 12.3, 10.3, 11.8, 6.3, 10.5, 18.6, 12.0, 10.2, 15.7, 12.8, 6.3, 8.1, 11.7, 13.6, 11.9,13.9, 10.0, 15.8, 13.6, 16.0, 10.6, 1.0, 9.3, 15.4, 9.4, 7.0, 11.0, 16.2, 15.8, 13.1, 13.7, 9.3, 13.3, 8.6, 14.4, 13.3, 11.7, 12.3, 6.3, 12.4, 14.8, 6.4, 3.2, 12.9, 14.9, 12.1, 13.5, 16.5, 6.7

ตัวอย่างที่ 2

อยากทราบว่าค่าประมาณแบบจุดของค่าเฉลี่ยความกว้างของลำต้นโพธิ์ที่มีอายุ 4 ปีเท่ากับเท่าใด

```
>width=c(8.7,12.3,10.3,11.8,6.3,10.5,18.6,12.0,10.2,15.7,12.8,6.3,8.1,11.7,13.6,11.9,13.9,10.0,15.8,13.6,16.0,10.6,1.0,9.3,15.4,9.4,7.0,11.0,16.2,15.8,13.1,13.7,9.3,13.3,8.6,14.4,13.3,11.7,12.3,6.3,12.4,14.8,6.4,3.2,12.9,14.9,12.1,13.5,16.5,6.7)
> plot(density(width))
> mean(width)
[1] 11.504
```

จากกราฟพบว่ามีการกระจายค่อนข้างสมมาตร จึงประมาณค่าด้วยค่าเฉลี่ย

สรุป ความกว้างเฉลี่ยของลำต้นโพธิ์ที่มีอายุ 4 ปี มีค่าประมาณ 11.5 นิ้ว

ตัวอย่างที่ 2

จงประมาณช่วงความเชื่อมั่นที่ 99% ว่าต้นโพธิ์ที่มีอายุ 4 ปี น่าจะมีลำต้นกว้างเท่าไร

วิเคราะห์โจทย์

- ต้องการประมาณแบบช่วงของค่าเฉลี่ย
- ไม่ทราบค่า σ^2
- ข้อมูลมีขนาดใหญ่เป็น 50 ค่า → n>30 ข้อมูลมาจากการแจกแจงปกติ
- ดังนั้น เลือกใช้สถิติ t test เพื่อประมาณแบบช่วงของค่าเฉลี่ย

ตัวอย่างที่ 2

```
> t.test(width,conf.level = 0.99 )

One Sample t-test

data: width

t = 22.57, df = 49, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

99 percent confidence interval:

10.13805 12.86995

sample estimates:

mean of x

11.504
```

สรุป ที่ช่วงความเชื่อมั่นที่ 99% ต้นโพธิ์ที่มีอายุ 4 ปี จะมีลำต้นกว้างอยู่ในช่วง 10.14 ถึง 12.87 นิ้ว

ตัวอย่างที่ 3

จากการสอบถามนิสิตจำนวน 10 คนที่สุ่มมาจากนิสิตคณะวิทยาศาสตร์เกี่ยวกับการเวลาที่ใช้อินเทอร์เน็ตใน 1 วัน ได้ข้อมูลดังนี้

3 11 4 7 10 11 9 10 4 11

ให้หาค่าประมาณของจำนวนชั่วโมงที่นิสิตคณะวิทยาศาสตร์ใช้ในการเล่นอินเทอร์เน็ตเฉลี่ยต่อวัน ที่สัมประสิทธิ์ความเชื่อมั่น 0.90

วิเคราะห์โจทย์

- ต้องการประมาณแบบช่วงของค่าเฉลี่ย
- ไม่ทราบความแปรปรวนของประชากร σ^2
- จำนวนข้อมูลมีขนาดเล็กเป็น 10 ค่า

ต้องตรวจสอบการแจกแจงของข้อมูลว่ามีการแจกแจงปกติหรือไม่

ตัวอย่างที่ 3

ทดสอบสมมติฐานเพื่อตรวจสอบการแจกแจงปกติ

1. ตั้งสมมติฐานในการทดสอบ
   - H_0 : เวลาที่ใช้ในการเล่นอินเตอร์เน็ตมาจากประชากรที่มีการแจกแจงปกติ
   - H_1 : เวลาที่ใช้ในการเล่นอินเตอร์เน็ตไม่ได้มาจากประชากรที่มีการแจกแจงปกติ

2. กำหนดระดับนัยสำคัญ (α) = 0.1

3. คำนวณค่า p-value

p-value=0.03403 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ

```
> time=c(3,11,4,7,10,11,9,10,4,11)

> ad.test(time)

Anderson-Darling normality test

data: time

A = 0.748, p-value = 0.03403
```

4. สรุปผล

ปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.1 แสดงว่าข้อมูลเวลาที่ใช้ในการเล่นอินเตอร์เน็ตไม่ได้มาจากประชากรที่มีการแจกแจงปกติ

ตัวอย่างที่ 3

- จะได้ว่าข้อมูลไม่ได้มาจากการแจกแจงปกติ
- จึงใช้สถิติ sign test เพื่อประมาณแบบช่วงของมัธยฐาน

```
> signmedian.test(time,conf.level = 0.9)

Exact sign test

data: time

#(x!=0) = 10, mu = 0, p-value = 0.001953

alternative hypothesis: the median of x is not equal to mu

89.0625 percent confidence interval:

4 11

sample estimates:

point estimator

9.5
```

สรุป ที่สัมประสิทธิ์ความเชื่อมั่น 0.90 จำนวนชั่วโมงที่นิสิตคณะวิทยาศาสตร์ใช้ในการเล่นอินเตอร์เน็ตต่อวันมีค่าอยู่ในช่วง 4 ถึง 11 ชั่วโมงต่อวัน

3. การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร 1 กลุ่ม

เมื่อข้อมูลมาจากการแจกแจงปกติ หรือตัวอย่างมีขนาดใหญ่ $\bar{x}$ เป็นค่าสถิติและเป็นค่าประมาณของ μ สำหรับสมมติฐานที่ใช้ในการทดสอบจะมี 3 รูปแบบดังนี้

สมมติฐานหลัก	สมมติฐานรอง
$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$
$H_0 : \mu \leq \mu_0$	$H_1 : \mu > \mu_0$
$H_0 : \mu \geq \mu_0$	$H_1 : \mu < \mu_0$

ใช้ t test

เมื่อข้อมูลไม่ได้มาจากการแจกแจงปกติ ค่ามัธยฐานจะเป็นค่าสถิติและเป็นค่าประมาณของค่ากลางของประชากร สำหรับสมมติฐานที่ใช้ในการทดสอบจะมี 3 รูปแบบคือ

สมมติฐานหลัก	สมมติฐานรอง
$H_0 : M = M_0$	$H_1 : M \neq M_0$
$H_0 : M \leq M_0$	$H_1 : M > M_0$
$H_0 : M \geq M_0$	$H_1 : M < M_0$

ใช้ sign test

ตัวอย่างที่ 4

วารุณีเป็นผู้ประกอบการธุรกิจนมพาสเจอไรซ์บรรจุถุงรายย่อยหรือ SME เพื่อส่งขายตามโรงเรียนโดยรับซื้อนมดิบจากเกษตรกร จำนวน 24 รายเป็นประจำทุกวัน เพื่อนำมาแปรรูป หลังจากที่ทราบประกาศของมาตรฐานน้ำนมดิบของกรมปศุสัตว์ทราบว่าต้องมีไขมันนม (fat) มากกว่า 3.25% เขาจึงได้วิเคราะห์หาปริมาณนมดิบจากเกษตรกรทั้ง 24 รายพบว่ามีปริมาณไขมันดังนี้

3.08	3.19	3.44	3.23	3.32	5.53	4.33	3.28
3.92	3.77	3.44	3.80	5.64	4.60	5.02	4.11
5.62	5.16	4.09	3.70	3.48	3.24	3.79	4.09

จงทดสอบว่าไขมันนมเป็นไปตามประกาศของมาตรฐานนมดิบของกรมปศุสัตว์หรือไม่ที่ระดับนัยสำคัญ 0.05

ตัวอย่างที่ 4

วิเคราะห์โจทย์

- ต้องการทดสอบว่าไขมันนมเป็นไปตามประกาศของมาตรฐานนมดิบของกรมปศุสัตว์หรือไม่
- ไขมันนมตามมาตรฐานนมดิบ คือ ไขมันนมมากกว่า 3.25%
- สมมติฐานทางสถิติคือ $\mu > 3.25$
- ไม่ทราบความแปรปรวนของประชากร σ^2
- จำนวนข้อมูลมีขนาดเล็กเป็น 24 ค่า → ต้องตรวจสอบการแจกแจงของข้อมูลว่ามีการแจกแจงปกติหรือไม่

```
>milk=c(3.08,3.19,3.44,3.23,3.32,5.53,4.33,3.28,3.92,3.77,3.44,3.80,5.64,4.60,5.02,4.11,5.62,5.16,4.09,3.70,3.48,3.24,3.79,4.09)
```

ตัวอย่างที่ 4

ทดสอบสมมติฐานเพื่อตรวจสอบการแจกแจงปกติ

1. ตั้งสมมติฐานในการทดสอบ

   H_0: ข้อมูลไขมันนมมาจากประชากรที่มีการแจกแจงปกติ

   H_1: ข้อมูลไขมันนมไม่ได้มาจากประชากรที่มีการแจกแจงปกติ

2. กำหนดระดับนัยสำคัญ (α) = 0.05

3. คำนวณค่า p-value

```
> ad.test(milk)
Anderson-Darling normality test
data: milk
A = 1.0308, p-value = 0.008422
```

p-value=0.008422

ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ

4. สรุปผล ปฏิเสธ H_0 แสดงว่าข้อมูลไขมันนมไม่ได้มาจากประชากรที่มีการแจกแจงปกติ

ตัวอย่างที่ 4

- ข้อมูลไม่ได้มาจากการแจกแจงปกติ ดังนั้นจึงใช้สถิติ sign test ในการทดสอบสมมติฐาน

<u>การทดสอบสมมติฐาน</u>

1. ตั้งสมมติฐานในการทดสอบ

$H_0 : M \leq 3.25$

$H_1 : M > 3.25$

0.05

2. กำหนดระดับนัยสำคัญ (α) =

ตัวอย่างที่ 4

3.คำนวณค่า p-value

```
> signmedian.test(milk,mu=3.25,conf.level=0.95,alternative = "greater")
Exact sign test
data: milk
#(x>3.25) = 20, mu = 3.25, p-value = 0.0007719
alternative hypothesis: the median of x is greater than mu
93.60853 percent confidence interval:
3.44 4.11
sample estimates:
point estimator
3.795
```

ค่า p-value=0.0007719 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ

4. สรุปผล ปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 แสดงว่าไขมันนมเป็นไปตามประกาศของมาตรฐานนมดิบของกรมปศุสัตว์ นั่นคือ มีไขมันนมมากกว่า 3.25%

ตัวอย่างที่ 5

บริษัทขายรถยนต์ใช้แล้ว (used car) โฆษณาขายรถยนต์ใช้แล้วยี่ห้อหนึ่งที่มีอายุการใช้งาน 5 ปีว่าระยะทางเฉลี่ยต่อน้ำมัน 1 ลิตร เท่ากับ 10 กิโลเมตร นักสถิติต้องการตรวจสอบว่า คำโฆษณาดังกล่าวน่าเชื่อถือหรือไม่ จึงสุ่มตัวอย่างรถยนต์ดังกล่าว จำนวน 36 คัน ทดลองขับรถภายใต้เงื่อนไขต่าง ๆ ที่เหมือนกัน และบันทึกระยะทาง (กิโลเมตร) ต่อการใช้น้ำมัน 1 ลิตร ได้ดังนี้

12, 9, 11, 7, 7, 7, 8, 9, 10, 12, 13, 12, 7, 6, 8, 9, 10,

11, 7, 6, 8, 9, 8, 9, 11, 8, 7, 7, 6, 7, 8, 10, 14, 12, 11, 9

คำโฆษณาของบริษัทดังกล่าว น่าเชื่อถือหรือไม่ กำหนดระดับนัยสำคัญ 0.05

ตัวอย่างที่ 5

วิเคราะห์โจทย์

- ต้องการทดสอบว่าคำโฆษณาของบริษัทดังกล่าว น่าเชื่อถือหรือไม่
- รถยนต์ใช้แล้วยี่ห้อหนึ่งที่มีอายุการใช้งาน 5 ปีว่าระยะทางเฉลี่ยต่อน้ำมัน 1 ลิตร เท่ากับ 10 กิโลเมตร
- สมมติฐานทางสถิติคือ $\mu = 10$
- ไม่ทราบความแปรปรวนของประชากร σ^2
- จำนวนข้อมูลมีขนาดใหญ่เป็น 36 ค่า → n>30 ข้อมูลมาจากการแจกแจงปกติ

```
>car=c(12,9,11,7,7,7,8,9,10,12,13,12,7,6,8,9,10,11,7,6,8,9,8,9,11,8,7,7,6,7,
8,10,14,12,11,9)
```

ตัวอย่างที่ 5

- ข้อมูลมาจากการแจกแจงปกติ ดังนั้นจึงใช้สถิติ t test ในการทดสอบสมมติฐาน

การทดสอบสมมติฐาน

1. ตั้งสมมติฐานในการทดสอบ

$H_0 : \mu = 10$

$H_1 : \mu \neq 10$

2. กำหนดระดับนัยสำคัญ (α) = 0.05

ตัวอย่างที่ 5

3.คำนวณค่า p-value

```
> t.test(car,mu=10)
One Sample t-test
data: car
t = -2.72, df = 35, p-value = 0.01009
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
8.302158 9.753398
sample estimates:
mean of x
9.027778
```

ค่า p-value=0.01009 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ

4. สรุปผล ปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 แสดงว่ารถยนต์ใช้แล้วยี่ห้อหนึ่งที่มีอายุการใช้งาน 5 ปี ว่าระยะทางเฉลี่ยต่อน้ำมัน 1 ลิตร เท่ากับ 10 กิโลเมตร ไม่เป็นความจริง หรือคำโฆษณาของบริษัทดังกล่าวไม่น่าเชื่อถือ

ตัวอย่างที่ 6

ผู้จัดการคิดว่าการลดน้ำหนักของลูกค้าจะต้องลดลงอย่างน้อย 45 ปอนด์ ภายใน 6 เดือน สุ่มตัวอย่างลูกค้าที่ใช้บริการการลดน้ำหนัก 27 คน ได้ข้อมูลดังนี้

31 31 52 32 34 37 52 40 29 25 42 31 38 44 29 56 36 39 14 25 33 27 25 26 36 29 49

ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่าความเชื่อของผู้จัดการเกินความเป็นจริงหรือไม่

ตัวอย่างที่ 6

วิเคราะห์โจทย์

- ต้องการทดสอบว่าความเชื่อของผู้จัดการเกินความเป็นจริงหรือไม่
- ผู้จัดการคิดว่าการลดน้ำหนักของลูกค้าจะต้องลดลงอย่างน้อย 45 ปอนด์
- สมมติฐานทางสถิติคือ $\mu \geq 45$
- ไม่ทราบความแปรปรวนของประชากร σ^2
- จำนวนข้อมูลมีขนาดเล็กเป็น 27 ค่า → ต้องตรวจสอบการแจกแจงของข้อมูลว่ามีการแจกแจงปกติหรือไม่

```
>weight=c(31,31,52,32,34,37,52,40,29,25,42,31,38,44,29,56,36,39,14,25,33,27,25,26,36,29,49)
```

ตัวอย่างที่ 6

ทดสอบสมมติฐานเพื่อตรวจสอบการแจกแจงปกติ

1. ตั้งสมมติฐานในการทดสอบ

   H_0: ข้อมูลน้ำหนักมาจากประชากรที่มีการแจกแจงปรกติ

   H_1:ข้อมูลน้ำหนักไม่ได้มาจากประชากรที่มีการแจกแจงปรกติ

2. กำหนดระดับนัยสำคัญ (α) = 0.05

3.คำนวณค่า p-value

```
> ad.test(weight)
Anderson-Darling normality test
data: weight
A = 0.48549, p-value = 0.2081
```

p-value= 0.2081

ซึ่งมีค่ามากกว่าระดับนัยสำคัญ

4. สรุปผล ยอมรับ H_0 ที่ระดับนัยสำคัญ 0.01 แสดงว่าข้อมูลน้ำหนักมาจากประชากรที่มีการแจกแจงปกติ

ตัวอย่างที่ 6

- ข้อมูลมาจากการแจกแจงปกติ ดังนั้นจึงใช้สถิติ t test ในการทดสอบสมมติฐาน

<u>การทดสอบสมมติฐาน</u>

1. ตั้งสมมติฐานในการทดสอบ

$H_0 : \mu \geq$ 45

$H_1 : \mu <$ 45

0.05

2. กำหนดระดับนัยสำคัญ (α) =

ตัวอย่างที่ 6

3.คำนวณค่า p-value

```
> t.test(weight,mu=45,alternative = "less")
One Sample t-test
data: weight
t = -5.405, df = 26, p-value = 5.789e-06
alternative hypothesis: true mean is less than 45
95 percent confidence interval:
-Inf 38.07959
sample estimates:
mean of x
34.88889
```

p-value= 5.789e-06 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ

4. สรุปผล ปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 แสดงว่าน้ำหนักของลูกค้าลดลงอย่างน้อย 45 ปอนด์ ไม่เป็นความจริง หรือ ความเชื่อของผู้จัดการเกินความเป็นจริง

สรุป ค่าประมาณแบบจุดของค่าเฉลี่ยของประชากร (μ)

ดูลักษณะการแจกแจง

plot(density(ข้อมูล))

- หากแจกแจงปกติ ใช้ ค่าเฉลี่ยของตัวอย่าง ($\bar{x}$)

  mean(ข้อมูล)

- หากไม่มีการแจกแจงปกติ ใช้ ค่ามัธยฐานของตัวอย่าง

  median(ข้อมูล)

สรุป ค่าประมาณแบบช่วงของค่าเฉลี่ยของประชากร (μ)

- เมื่อทราบค่า σ^2 แจกแจงแบบใดก็ตาม

  t.test(ข้อมูล, conf.level = ?)

- เมื่อไม่ทราบค่า σ^2 และ $n \geq 30$ มีการแจกแจงปกติ

  t.test(ข้อมูล, conf.level = ?)

- เมื่อไม่ทราบค่า σ^2 และ $n < 30$

  ทดสอบการแจกแจงด้วย ad.test(ข้อมูล)

  มีการแจกแจงปกติ ใช้ t.test(ข้อมูล, conf.level = ?)

  ไม่มีการแจกแจงปกติ ใช้ signmedian.test(ข้อมูล, conf.level = ?)

สรุป การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร 1 กลุ่ม

1. ตั้งสมมติฐานในการทดสอบ (H_0 และ H_1)

2. กำหนดระดับนัยสำคัญ (α)

3. คำนวณค่า p-value

- เมื่อทราบค่า σ^2 แจกแจงแบบใดก็ตาม

  t.test(ข้อมูล, ค่าเฉลี่ย, alternative = "two.sided" หรือ "less" หรือ "greater"))

- เมื่อไม่ทราบค่า σ^2 และ $n \geq 30$ มีการแจกแจงปกติ

  t.test(ข้อมูล, ค่าเฉลี่ย, alternative = "two.sided" หรือ "less" หรือ "greater"))

- เมื่อไม่ทราบค่า σ^2 และ $n < 30$

  ทดสอบการแจกแจงด้วย ad.test(ข้อมูล)

  มีการแจกแจงปกติ ใช้ t.test(ข้อมูล, ค่าเฉลี่ย, alternative = "two.sided" หรือ "less" หรือ "greater"))

  ไม่มีการแจกแจงปกติ ใช้ signmedian.test(ข้อมูล, ค่าเฉลี่ย, alternative = "two.sided" หรือ "less" หรือ "greater")

4. สรุป

สรุป การตัดสินใจและการสรุป

ในการตัดสินใจและสรุปผล เมื่อทำการวิเคราะห์ข้อมูลด้วยโปรแกรม R จะใช้ค่า p-value ซึ่งเป็นค่าความน่าจะเป็นที่แปลงมาจากค่าสถิติทดสอบ โดยมีเกณฑ์ดังนี้

- ปฏิเสธ H_0 ถ้าค่า p-value มีค่าน้อยกว่า ระดับนัยสำคัญ
- ยอมรับ H_0 ถ้าค่า p-value มีค่ามากกว่าหรือเท่ากับ ระดับนัยสำคัญ

สรุป การตรวจสอบการแจกแจงแบบปกติ

- ฟังก์ชันที่ใช้ในการทดสอบการแจกแจงปกติ Anderson Darling คือ ad.test()

ขั้นตอนในการทดสอบสมมติฐานมีดังนี้

1. ตั้งสมมติฐานในการทดสอบ
   - H_0 : ข้อมูลมาจากประชากรที่มีการแจกแจงปกติ
   - H_1 : ข้อมูลไม่ได้มาจากประชากรที่มีการแจกแจงปกติ
2. กำหนดระดับนัยสำคัญ (α) เช่น 0.05, 0.01, 0.1
3. คำนวณค่า p-value
4. สรุปผล โดยพิจารณาค่า p-value ดังนี้
   - ถ้า p-value $\geq \alpha$ ยอมรับ H_0 แสดงว่าข้อมูลมีการแจกแจงปกติ
   - ถ้า p-value $< \alpha$ ปฏิเสธ H_0 แสดงว่าข้อมูลไม่มีการแจกแจงปกติ