4. การอนุมานสัดส่วนของประชากร

กระบวนวิชา 208263: สถิติเบื้องต้น

โดย... ผศ. ดร.สุคนธ์ ประสิทธิ์วัฒนเสรี

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

วัตถุประสงค์

- นักศึกษาสามารถทำการอนุมานค่าสัดส่วนประชากร 1 กลุ่ม และ 2 กลุ่มได้อย่างถูกต้องเหมาะสม

2

เนื้อหา

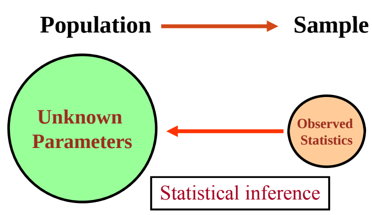
- บทนำ
- การอนุมานค่าสัดส่วน 1 กลุ่ม
  - การแจกแจงค่าสัดส่วนของตัวอย่าง
  - การประมาณค่า
  - การทดสอบสมมติฐาน
- การอนุมานค่าสัดส่วน 2 กลุ่ม
  - การแจกแจงค่าสัดส่วนของตัวอย่าง
  - การประมาณค่า
  - การทดสอบสมมติฐาน

3

Introduction

- **หัวใจหลักของงานวิจัย**

  ต้องการอ้างอิงผลการศึกษาที่ได้จากกลุ่มตัวอย่างที่ศึกษา ไปถึงกลุ่มประชากร



4

การอนุมานทางสถิติ

- **Estimation** – ทำการประมาณค่าพารามิเตอร์
  - ➢point estimate
  - ➢interval estimate
- **Hypothesis testing** – ทำการทดสอบเกี่ยวกับพารามิเตอร์ว่าข้อสมมติหรือข้อสงสัยเป็นจริงหรือไม่ โดยพิจารณาจากค่าสถิติที่ได้จากข้อมูลตัวอย่าง

5

การอนุมานเกี่ยวกับค่าสัดส่วน 1 กลุ่ม

- **Point and Interval Estimate**
- **Hypothesis Testing**

6

ค่าสัดส่วน (Proportion)

- จากการตรวจสอบสินค้าจำนวน 10 ชิ้น ว่าผ่านเกณฑ์มาตรฐานที่กำหนดหรือไม่? ได้ผลดังนี้

N

ชิ้นที่ 1	ผ่าน
ชิ้นที่ 2	ผ่าน
ชิ้นที่ 3	ไม่ผ่าน
ชิ้นที่ 4	ผ่าน
ชิ้นที่ 5	ผ่าน

ชิ้นที่ 6	ไม่ผ่าน
ชิ้นที่ 7	ไม่ผ่าน
ชิ้นที่ 8	ผ่าน
ชิ้นที่ 9	ผ่าน
ชิ้นที่ 10	ผ่าน

X

จำนวนสินค้าที่ได้มาตรฐาน = 7

สัดส่วนสินค้าได้มาตรฐาน = $^{7}/_{10} = 0.7$ P

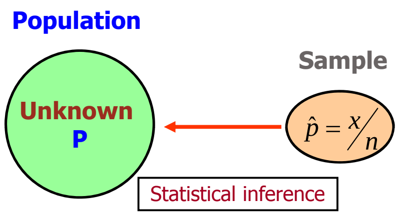
สัดส่วนสินค้าที่ไม่ได้มาตรฐาน = 1 – 0.7 = 0.3

Q

ค่าสัดส่วน (Proportion)

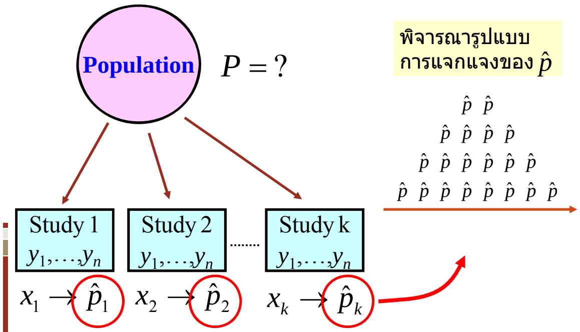
	ประชากร	ตัวอย่าง
จำนวนทั้งหมด	N	n
จำนวนครั้งที่เกิดเหตุการณ์ที่สนใจ	X	x
สัดส่วนที่เกิดเหตุการณ์ที่สนใจ	P = X/N	$\hat{p} = {}^{x}/_{n}$
สัดส่วนที่เกิดเหตุการณ์ไม่สนใจ	Q = 1-P	$\hat{q} = 1 - \hat{p}$

การอนุมานค่า P



โดยอาศัยลักษณะการแจกแจงความน่าจะเป็นของค่า $\hat{p}$

การแจกแจงของค่าสัดส่วนตัวอย่าง



การแจกแจงของค่าสัดส่วนตัวอย่าง

- เมื่อ n มีขนาดใหญ่ การแจกแจงของ $\hat{p} = \frac{x}{n}$ มีการแจกแจงแบบปกติ ที่มี

ค่าเฉลี่ย $\mu_{\hat{p}} = P$

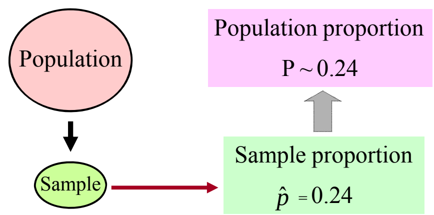
และความแปรปรวน $\sigma^2_{\hat{p}} = \frac{PQ}{n}$

$$Z = \frac{\hat{p} - P}{\sqrt{PQ/n}}$$

ดังนั้นจะอาศัยการแจกแจงปกติมาตรฐาน Z ไปช่วยในการประมาณค่าแบบช่วงและการทดสอบสมมติฐานเกี่ยวกับ P

Point Estimation

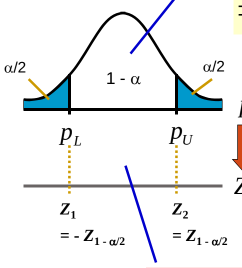
- ประมาณ**ค่าสัดส่วนประชากร**แบบจุดด้วย <u>**ค่าสัดส่วนของตัวอย่าง**</u>



Interval Estimation

(1-α)100% ช่วงความเชื่อมั่น ((1-α)100 % CI) ของค่าสัดส่วน

$$\Rightarrow p_L < P < p_U$$



$\hat{p}$

$$Z = \frac{\hat{p} - P}{\sqrt{PQ/n}}$$

$$P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1 - \alpha$$

Interval Estimation

(1-α)100 % CI ของค่าสัดส่วน

$$P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1 - \alpha$$

แทนค่า $Z = \dfrac{\hat{p} - P}{\sqrt{PQ/n}}$

$$P\left(-Z_{1-\alpha/2} < \frac{\hat{p} - P}{\sqrt{PQ/n}} < Z_{1-\alpha/2}\right) = 1 - \alpha$$

$$P\left(\boxed{\hat{p} - Z_{1-\alpha/2}\sqrt{PQ/n} < P < \hat{p} + Z_{1-\alpha/2}\sqrt{PQ/n}}\right) = 1 - \alpha$$

Interval Estimation

(1 - α)100% CI ของค่าสัดส่วน

$$\hat{p} - Z_{1-\alpha/2}\sqrt{PQ/n} < P < \hat{p} + Z_{1-\alpha/2}\sqrt{PQ/n}$$

ไม่ทราบค่า

จำเป็นต้องประมาณค่าด้วย $\hat{p}$

$$\hat{p} - Z_{1-\frac{\alpha}{2}}\sqrt{\hat{p}\hat{q}/n} < P < \hat{p} + Z_{1-\frac{\alpha}{2}}\sqrt{\hat{p}\hat{q}/n}$$

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

Hypothesis Testing

- **สมมติฐานทางสถิติ**

เมื่อ c=ค่าคงที่ใด ๆ [0,1]

$H_0 : P = c$
$H_1 : P \neq c$

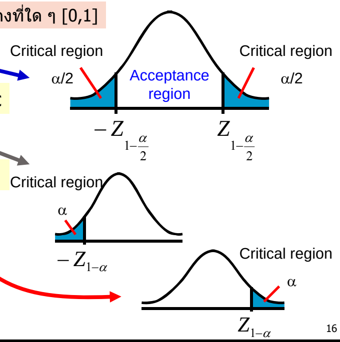
$H_0 : P = c$ or $P \geq c$
$H_1 : P < c$

$H_0 : P = c$ or $P \leq c$
$H_1 : P > c$

- **สถิติทดสอบ**

$$Z = \frac{\hat{p} - P}{\sqrt{PQ/n}}$$

- **เขตวิกฤต** ที่ α



ทดลองทำ

- ต้องการทดสอบว่าสัดส่วนของเด็กที่มีน้ำหนักแรกเกิดอยู่ในเกณฑ์ปกติเท่ากับ 90% หรือไม่ จงตั้งสมมติฐานทางสถิติเพื่อทำการทดสอบ
  ก) H_0: P = 90% vs H_1: P ≠ 90%
  ข) H_0: P = 90% vs H_1: P > 90%
  ค) H_0: P = 0.9 vs H_1: P ≠ 0.9
  ง) H_0: P = 0.9 vs H_1: P > 0.9

ทดลองทำ

- ตามเกณฑ์มาตรฐานกำหนดว่า สัดส่วนของสินค้าที่บกพร่องซึ่งผลิตจากเครื่อง A ต้องมีค่าไม่เกินร้อยละ 10 ต้องการทดสอบว่าเครื่อง A เป็นไปตามเกณฑ์ที่กำหนดไว้หรือไม่
  จงตั้งสมมติฐานทางสถิติเพื่อทำการทดสอบ
  ก) H_0: P = .10 vs H_1: P ≠ .10
  ข) H_0: P = .10 vs H_1: P > .10
  ค) H_0: P = .10 vs H_1: P < .10
  ง) H_0: P = .10 vs H_1: P ≤ .10

ลองพิจารณา สัญลักษณ์ของข้อความ

- ค่าสัดส่วนมีค่า**ไม่เกิน** .05
  ก. P < .05 ข. $P \le .05$ ค. $P \ge .05$
- ค่าสัดส่วนมีค่า**ไม่น้อยกว่า** .05
  ก. P < .05 ข. P > .05 ค. $P \ge .05$
- ค่าสัดส่วนมีค่า**อย่างมาก** .40
  ก. P < .40 ข. $P \le .40$ ค. $P \ge .40$
- ค่าสัดส่วนมีค่า**อย่างน้อย** .40
  ก. P < .40 ข. $P \le .40$ ค. $P \ge .40$
- ค่าสัดส่วนมีค่า**เกินกว่า** .60
  ก. P > .60 ข. $P \le .60$ ค. $P \ge .60$
- ค่าสัดส่วนมีค่า**ต่ำกว่า** .60
  ก. P < .60 ข. $P \le .60$ ค. $P \ge .60$

19

ตัวอย่างที่ 1

- เพื่อทดสอบเชื้อไวรัสชนิดหนึ่งที่ฉีดเข้าไปในกระต่ายว่า ทำให้กระต่ายที่ได้รับเชื้อนี้มีโอกาสเป็นเนื้องอกที่หน้าอกในอัตราร้อยละเท่าใด ผู้ทำการวิจัยจึงได้สุ่มตัวอย่างกระต่ายที่ได้รับเชื้อนี้มา 128 ตัว แล้วทำการตรวจสอบ ปรากฏว่า มีอยู่ 72 ตัว ที่เป็นเนื้องอกที่หน้าอก

จงประมาณค่าสัดส่วนของกระต่ายทั้งหมดที่ได้รับเชื้อไวรัสนี้แล้วจะเป็นเนื้องอกที่หน้าอก โดยใช้ช่วงความเชื่อมั่น 95%

20

- เพื่อทดสอบเชื้อไวรัสชนิดหนึ่งที่ฉีดเข้าไปในกระต่ายว่า ทำให้กระต่ายที่ได้รับเชื้อนี้มีโอกาสเป็นเนื้องอกที่หน้าอกในอัตราร้อยละเท่าใด ผู้ทำการวิจัยจึงได้สุ่มตัวอย่างกระต่ายที่ได้รับเชื้อนี้มา 128 ตัว แล้วทำการตรวจสอบ ปรากฏว่า มีอยู่ 72 ตัว ที่เป็นเนื้องอกที่หน้าอก

n

x

จงประมาณค่าสัดส่วนของกระต่ายทั้งหมดที่ได้รับเชื้อไวรัสนี้แล้วจะเป็นเนื้องอกที่หน้าอก โดยใช้ช่วงความเชื่อมั่น 95%

P = สัดส่วนของกระต่ายที่ได้รับเชื้อไวรัสแล้วเป็นเนื้องอกที่หน้าอก

95% CI ของ P คือ $\boxed{P = \hat{p} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}}$

$\hat{p}$ = สัดส่วนของตัวอย่างกระต่ายที่ได้รับเชื้อไวรัสแล้วเป็นเนื้องอก

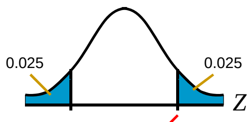
$= \frac{x}{n} = \frac{72}{128} = 0.563 \Rightarrow \hat{q} = 1 - \hat{p} = 0.437$

21

$\boxed{\text{95\% CI}}$ ของ P คือ $P = \hat{p} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$\hat{p} = 0.563$
$\hat{q} = 0.437$
$n = 128$

$\alpha = 0.05$



$Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$

95% CI ของ P คือ

$0.563 \pm (1.96 \cdot \sqrt{\frac{(0.563)(0.437)}{128}})$

0.563 ± 0.086

$\Rightarrow [0.477 - 0.649]$

- ที่ระดับความเชื่อมั่น 95% สัดส่วนของกระต่ายที่ได้รับเชื้อไวรัสจะเป็นเนื้องอกที่หน้าอกมีค่าอยู่ในช่วง 0.477 ถึง 0.649

22

ตัวอย่างที่ 2

- ตามลักษณะพันธุกรรมกล่าวว่า พืชที่ได้มาจากการผสมพันธุ์ระหว่างเมล็ดพันธุ์ ก และ ข จะมีลักษณะเตี้ย 80% ได้มีผู้ทดลองผสมพันธุ์พืช 2 ชนิดนี้ คือ ก และ ข พบว่า พืชที่ได้จากการผสมพันธุ์นี้ 200 ต้น ปรากฏว่า มี 64 ต้น ที่ไม่ใช่ต้นเตี้ย

จงทดสอบว่า การทดลองนี้จะสอดคล้องตามทฤษฎีดังกล่าวข้างต้นนี้หรือไม่ โดยใช้ระดับนัยสำคัญ 0.10

ต้องการให้ทำการทดสอบสมมติฐานทางสถิติ

23

- ตามลักษณะพันธุกรรมกล่าวว่า พืชที่ได้มาจากการผสมพันธุ์ระหว่างเมล็ดพันธุ์ ก และ ข จะมีลักษณะเตี้ย 80% ได้มีผู้ทดลองผสมพันธุ์พืช 2 ชนิดนี้ คือ ก และ ข พบว่า พืชที่ได้จากการผสมพันธุ์นี้ 200 ต้น ปรากฏว่า มี 64 ต้น ที่ไม่ใช่ต้นเตี้ย
  จงทดสอบว่า การทดลองนี้จะสอดคล้องตามทฤษฎีดังกล่าวข้างต้นนี้หรือไม่ โดยใช้ระดับนัยสำคัญ 0.10

n

x = 200-64 = 136

คำถาม ผลการผสมพันธุ์ระหว่าง ก กับ ข จะมีต้นเตี้ย = 80% ?

ให้ P คือสัดส่วนของต้นเตี้ยที่เกิดจากการผสมระหว่าง ก กับ ข

ดังนั้นต้องการทดสอบ P = 80% ? (P = 0.80)

สมมติฐานทางสถิติ $H_0 : P = 0.80$

$H_1 : P \neq 0.80$

สถิติทดสอบ ภายใต้ $H_0 : P = 0.80 \Rightarrow Q = 0.20$

$Z = \frac{\hat{p} - P}{\sqrt{PQ/n}}$ คำนวณ $\hat{p} = \frac{x}{n} = \frac{136}{200} = 0.68$

24

สมมติฐานทางสถิติ $H_0 : P = 0.80$

$H_1 : P \neq 0.80$

$\hat{p} = 0.68$
$n = 200$
$\alpha = 0.10$

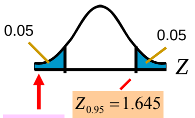
สถิติทดสอบ $Z = \dfrac{\hat{p} - P}{\sqrt{PQ/n}}$

คำนวณค่าสถิติทดสอบ

$$z_c = \frac{0.68 - 0.8}{\sqrt{0.8(0.2)/200}} = -4.242$$

เขตวิกฤต ที่ α = .10

Z < -1.645 หรือ Z > 1.645



สรุปผล

ปฏิเสธ $H_0 : P = 0.80$

ที่ระดับนัยสำคัญ 0.10 สัดส่วนของต้นเตี้ยแตกต่างจาก 0.8

25

ข้อที่ 1 (แบบฝึกหัดท้ายบท)

- ในถุงใบหนึ่งมีลูกหินสีแดงและสีดำรวมกัน 10 ลูก สุ่มหยิบลูกหินครั้งละ 1 ลูก แบบใส่กลับคืน จำนวน 100 ครั้ง ปรากฏว่าเป็นลูกหินสีแดง 62 ครั้ง จงประมาณว่าในถุงดังกล่าวจะมีลูกหินสีแดงกี่ลูกด้วยความเชื่อมั่น 95%

26

ข้อที่ 2 (แบบฝึกหัดท้ายบท)

- จากการสอบถามนักธุรกิจในเขตกรุงเทพมหานคร กลุ่มหนึ่ง จำนวน 200 คน ปรากฏว่า 40% ใช้โทรศัพท์มือถือยี่ห้อ A ที่ระดับความเชื่อมั่น 96% จะมีผู้ใช้โทรศัพท์มือถือยี่ห้อ A ร้อยละเท่าใด

27

ข้อที่ 3 (แบบฝึกหัดท้ายบท)

- ทันตแพทย์ผู้หนึ่งเชื่อว่า อย่างมากที่สุด 10% ของเด็กที่มีอายุระหว่าง 3 ถึง 4 ขวบ เป็นโรคฟันผุ เพื่อทดสอบความเชื่อของเขา จึงสุ่มตัวอย่างเด็กอายุระหว่าง 3 ถึง 4 ขวบ มาตรวจฟันจำนวน 100 คน ปรากฏว่า มี 14 คน เป็นโรคฟันผุ จากข้อมูลที่เก็บรวบรวมมาได้นี้ จงทดสอบความเชื่อของทันตแพทย์ผู้นี้ ที่ระดับนัยสำคัญ 1%

28

ข้อที่ 4 (แบบฝึกหัดท้ายบท)

- บริษัทผู้ผลิตเมล็ดพันธุ์ข้าวโพดแห่งหนึ่งจะคัดเมล็ดพันธุ์ข้าวโพดไว้ทำพันธุ์ก็ต่อเมื่ออัตราการงอกของเมล็ดพันธุ์นั้น ๆ ต้องไม่ต่ำกว่าร้อยละ 80 จากการตรวจสอบของบริษัทได้ทำการทดลองเพาะเมล็ดพันธุ์ข้าวโพด ดงดก-1 จำนวน 100 เมล็ด ปรากฏว่า บริษัทยอมรับที่จะใช้พันธุ์ดงดก-1 เป็นเมล็ดพันธุ์ อยากทราบว่าการทดลองดังกล่าวมีเมล็ดพันธุ์ข้าวโพดงอกอย่างน้อยเท่าไร กำหนดระดับนัยสำคัญ 0.05

29

ค่าสัดส่วน (Proportion) 2 กลุ่ม

- จากการตรวจสอบคุณภาพผลิตภัณฑ์ที่ผลิตจากเครื่องจักร 2 เครื่อง ว่าได้มาตรฐานหรือไม่ ? ได้ผลดังนี้

เครื่อง A	
ชิ้นที่ 1	ผ่าน
ชิ้นที่ 2	ผ่าน
ชิ้นที่ 3	ไม่ผ่าน
ชิ้นที่ 4	ผ่าน
ชิ้นที่ 5	ผ่าน

เครื่อง B	
ชิ้นที่ 1	ไม่ผ่าน
ชิ้นที่ 2	ไม่ผ่าน
ชิ้นที่ 3	ผ่าน
ชิ้นที่ 4	ผ่าน
ชิ้นที่ 5	ผ่าน

สัดส่วนผลิตภัณฑ์ที่ได้มาตรฐานจากเครื่อง A = 4/5 = 0.80

สัดส่วนผลิตภัณฑ์ที่ได้มาตรฐานจากเครื่อง B = 3/5 = 0.60

30

ค่าสัดส่วน (Proportion) 2 กลุ่ม

	ประชากร 1	ตัวอย่าง 1	ประชากร 2	ตัวอย่าง 2
จำนวนทั้งหมด	N_1	n_1	N_2	n_2
จำนวนครั้งที่เกิดเหตุการณ์ที่สนใจ	X_1	x_1	X_2	x_2
สัดส่วนที่เกิดเหตุการณ์ที่สนใจ	$P_1 = X_1/N_1$	$\hat{p}_1 = x_1/n_1$	$P_2 = X_2/N_2$	$\hat{p}_2 = x_2/n_2$
สัดส่วนที่เกิดเหตุการณ์ไม่สนใจ	$Q_1 = 1-P_1$	$\hat{q}_1 = 1-\hat{p}_1$	$Q_2 = 1-P_2$	$\hat{q}_2 = 1-\hat{p}_2$

31

การอนุมานเกี่ยวกับค่าสัดส่วน 2 กลุ่ม

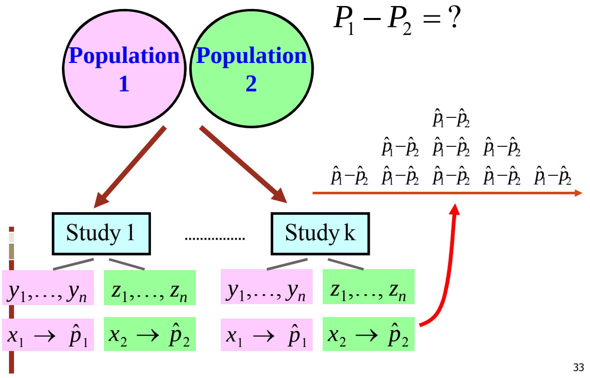
- **Point and Interval Estimate**
- **Hypothesis Testing**

โดยทั่วไปมักสนใจการอนุมานค่าผลต่างระหว่างสัดส่วน 2 กลุ่ม นั่นคือ อนุมานค่า $P_1 - P_2$

โดยอาศัยลักษณะการแจกแจงความน่าจะเป็นของค่า $\hat{p}_1 - \hat{p}_2$

32

Sampling Distribution of $\hat{p}_1 - \hat{p}_2$



33

การแจกแจงของผลต่างค่าสัดส่วนตัวอย่าง 2 กลุ่ม

- ค่าเฉลี่ยและความแปรปรวนของ $\hat{p}_1 - \hat{p}_2$

$$\mu_{\hat{p}_1-\hat{p}_2} = P_1 - P_2, \quad \sigma^2_{\hat{p}_1-\hat{p}_2} = \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}$$

- เมื่อขนาดตัวอย่าง (n_1, n_2) มีขนาดใหญ่พบว่า

$$\hat{p}_1 - \hat{p}_2 \sim N(\mu_{\hat{p}_1-\hat{p}_2}, \sigma^2_{\hat{p}_1-\hat{p}_2})$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}} \sim N(0,1)$$

ดังนั้นจะใช้ Z ช่วยในการประมาณค่าแบบช่วงและการทดสอบสมมติฐานเกี่ยวกับ $P_1 - P_2$

34

Point Estimation

- ประมาณค่าแบบจุดของ**<u>ผลต่างระหว่างค่าสัดส่วนประชากร 2 กลุ่ม</u>** ด้วยค่า**<u>ผลต่างระหว่างค่าสัดส่วนตัวอย่าง 2 กลุ่ม</u>**

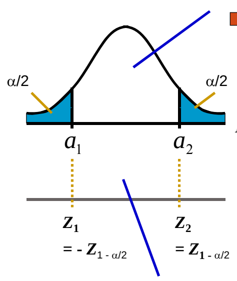
Sample proportion $\hat{p}_1 - \hat{p}_2 = 0.24$ → Population proportion $P_1 - P_2 \sim 0.24$

35

Interval Estimation

(1 - α)100 % CI ของค่าผลต่างสัดส่วน 2 กลุ่ม

$\boldsymbol{a_1 < P_1 - P_2 < a_2}$



$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}$$

$\mathbf{P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1 - \alpha}$

36

Interval Estimation

$(1-\alpha)100$ % CI ของ $P_1 - P_2$

$P(-Z_{1-\alpha/2} < Z < Z_{1-\alpha/2}) = 1-\alpha$

แทนค่า $Z = \dfrac{(\hat{p}_1-\hat{p}_2)-(P_1-P_2)}{\sqrt{\dfrac{P_1Q_1}{n_1}+\dfrac{P_2Q_2}{n_2}}}$

$$P\left(-Z_{1-\alpha/2} < \frac{(\hat{p}_1-\hat{p}_2)-(P_1-P_2)}{\sqrt{(P_1Q_1/n_1)+(P_2Q_2/n_2)}} < Z_{1-\alpha/2}\right) = 1-\alpha$$

$$P\left(\boxed{(\hat{p}_1-\hat{p}_2)-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}} < P_1-P_2 < (\hat{p}_1-\hat{p}_2)+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}}\right) = 1-\alpha$$

$(1-\alpha)100$ % CI ของ $P_1 - P_2$

ยังไม่ทราบค่า

$$(\hat{p}_1-\hat{p}_2)-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}} < P_1-P_2 < (\hat{p}_1-\hat{p}_2)+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$$

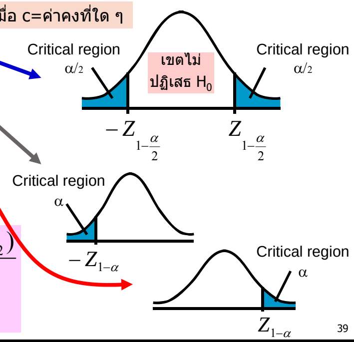
จำเป็นต้องประมาณค่าด้วย $\hat{p}_1, \hat{p}_2$

$$(\hat{p}_1-\hat{p}_2)-Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} < P_1-P_2 < (\hat{p}_1-\hat{p}_2)+Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$P_1-P_2 = (\hat{p}_1-\hat{p}_2) \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Hypothesis Testing

- **สมมติฐานทางสถิติ**
  - $H_0: P_1 - P_2 = c$ เมื่อ c=ค่าคงที่ใด ๆ
    $H_1: P_1 - P_2 \neq c$
  - $H_0: P_1 - P_2 = c$
    $H_1: P_1 - P_2 < c$
  - $H_0: P_1 - P_2 = c$
    $H_1: P_1 - P_2 > c$
- **เขตวิกฤต** ที่ α



- **สถิติทดสอบ**

$$Z = \frac{(\hat{p}_1-\hat{p}_2)-(P_1-P_2)}{\sqrt{\dfrac{P_1Q_1}{n_1}+\dfrac{P_2Q_2}{n_2}}}$$

ตัวสถิติทดสอบ

จาก $Z = \dfrac{(\hat{p}_1-\hat{p}_2)-(P_1-P_2)}{\sqrt{\dfrac{P_1Q_1}{n_1}+\dfrac{P_2Q_2}{n_2}}}$

- ภายใต้ $\mathbf{H_0 : P_1 - P_2 = 0}$ นั่นคือ $P_1 = P_2 = P$

เพราะฉะนั้น

$$Z = \frac{(\hat{p}_1-\hat{p}_2)}{\sqrt{PQ\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \Rightarrow Z = \frac{(\hat{p}_1-\hat{p}_2)}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$$

ไม่ทราบค่า P และ Q จึงประมาณค่าด้วย

$P \approx \hat{p} = \frac{x_1+x_2}{n_1+n_2}, Q \approx \hat{q} = 1-\hat{p}$

ตัวสถิติทดสอบ

จาก $Z = \dfrac{(\hat{p}_1-\hat{p}_2)-(P_1-P_2)}{\sqrt{\dfrac{P_1Q_1}{n_1}+\dfrac{P_2Q_2}{n_2}}}$

- ภายใต้ $\mathbf{H_0 : P_1 - P_2 = c}$ เพราะฉะนั้น

$$Z = \frac{(\hat{p}_1-\hat{p}_2)-c}{\sqrt{\dfrac{P_1Q_1}{n_1}+\dfrac{P_2Q_2}{n_2}}} \Rightarrow Z = \frac{(\hat{p}_1-\hat{p}_2)-c}{\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1}+\dfrac{\hat{p}_2\hat{q}_2}{n_2}}}$$

เนื่องจาก P_1, P_2 ไม่ทราบค่า
จึงประมาณค่าด้วย $\hat{p}_1, \hat{p}_2$

ทดลองทำ

- สงสัยว่า สัดส่วนของสินค้าเสียที่ผลิตจากโรงงาน A แตกต่างจากโรงงาน B หรือไม่

  จงตั้งสมมติฐานทางสถิติเพื่อทำการทดสอบ

ก) H_0: $P_1 = P_2$ vs H_1: $P_1 \neq P_2$
ข) H_0: $\hat{p}_1 = \hat{p}_2$ vs H_1: $\hat{p}_1 \neq \hat{p}_2$
ค) H_0: $P_1 - P_2 = 0$ vs H_1: $P_1 - P_2 \neq 0$
ง) H_0: $\hat{p}_1 - \hat{p}_2 = 0$ vs H_1: $\hat{p}_1 - \hat{p}_2 \neq 0$

ตัวสถิติทดสอบ

1) $Z = \dfrac{(\hat{p}_1-\hat{p}_2)}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ 2) $Z = \dfrac{(\hat{p}_1-\hat{p}_2)-c}{\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1}+\dfrac{\hat{p}_2\hat{q}_2}{n_2}}}$

ทดลองทำ

- สงสัยว่า สัดส่วนของสินค้าเสียที่ผลิตจากโรงงาน A มากกว่าโรงงาน B เท่ากับร้อยละ 10 หรือไม่

จงตั้งสมมติฐานทางสถิติเพื่อทำการทดสอบ

ก) $H_0: P_1 - P_2 = .10$ vs $H_1: P_1 - P_2 > .10$

ข) $H_0: P_1 - P_2 = .10$ vs $H_1: P_1 - P_2 \neq .10$

ค) $H_0: P_1 \leq P_2$ vs $H_1: P_1 > P_2$

ง) $H_0: P_1 - P_2 > .10$ vs $H_1: P_1 - P_2 \leq .10$

ตัวสถิติทดสอบ

1) $Z = \dfrac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$ 2) $Z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$

ทดลองทำ

- สงสัยว่า สัดส่วนของสินค้าเสียที่ผลิตจากโรงงาน A สูงกว่าโรงงาน B เกินกว่าร้อยละ 10 หรือไม่

จงตั้งสมมติฐานทางสถิติเพื่อทำการทดสอบ

ก) $H_0: P_1 - P_2 = .10$ vs $H_1: P_1 - P_2 > .10$

ข) $H_0: P_1 - P_2 = .10$ vs $H_1: P_1 - P_2 \neq .10$

ค) $H_0: P_1 \leq P_2$ vs $H_1: P_1 > P_2$

ง) $H_0: P_1 - P_2 > .10$ vs $H_1: P_1 - P_2 \leq .10$

ตัวสถิติทดสอบ

1) $Z = \dfrac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$ 2) $Z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$

ตัวอย่างที่ 3

- จากการสุ่มตัวอย่างผู้ลงคะแนนเสียงเลือกตั้งสมาชิกสภาผู้แทนราษฎรของจังหวัดเชียงใหม่ครั้งหนึ่ง จากอำเภอเมือง 300 คน และอำเภอสารภี 200 คน พบว่า มี 56% และ 48% ตามลำดับ ที่เลือกผู้สมัครรับเลือกตั้งเบอร์ 1

ที่ระดับความเชื่อมั่น 95% จงประมาณค่าความแตกต่างระหว่างสัดส่วนที่ผู้ลงคะแนนเลือกผู้สมัครรับเลือกตั้งเบอร์ 1 ของทั้งสองอำเภอ

ต้องการให้ทำการประมาณค่าแบบช่วงของ $P_1 - P_2$

- จากการสุ่มตัวอย่างผู้ลงคะแนนเสียงเลือกตั้งสมาชิกสภาผู้แทนราษฎรของจังหวัดเชียงใหม่ครั้งหนึ่ง จากอำเภอเมือง 300 คน และอำเภอสารภี 200 คน พบว่า มี 56% และ 48% ตามลำดับ ที่เลือกผู้สมัครรับเลือกตั้งเบอร์ 1 ที่ระดับความเชื่อมั่น 95% จงประมาณค่าความแตกต่างระหว่างสัดส่วนที่ผู้ลงคะแนนเลือกผู้สมัครรับเลือกตั้งเบอร์ 1 ของทั้งสองอำเภอ

$\hat{p}_1$ $\hat{p}_2$

P_1 = สัดส่วนของผู้ลงคะแนนเลือกเบอร์ 1 ในอำเภอเมือง

P_2 = สัดส่วนของผู้ลงคะแนนเลือกเบอร์ 1 ในอำเภอสารภี

95% CI ของ $P_1 - P_2$ คือ

$$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$\hat{p}_1$ = สัดส่วนของตัวอย่างที่เลือกเบอร์ 1 ในอำเภอเมือง = 0.56 ➡ $\hat{q}_1 = 0.44$

$\hat{p}_2$ = สัดส่วนของตัวอย่างที่เลือกเบอร์ 1 ในอำเภอสารภี = 0.48 ➡ $\hat{q}_2 = 0.52$

95% CI ของ $P_1 - P_2$ คือ

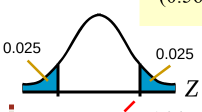
$\hat{p}_1 = 0.56, \hat{q}_1 = 0.44, n_1 = 300$

$\hat{p}_2 = 0.48, \hat{q}_2 = 0.52, n_2 = 200$

$$P_1 - P_2 = (\hat{p}_1 - \hat{p}_2) \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$\alpha = 0.05$

$$= (0.56 - 0.48) \pm (1.96.\sqrt{\frac{(0.56)(0.44)}{300} + \frac{(0.48)(0.52)}{200}})$$



➡ 0.08 ± 0.089 ➡ [-0.009 – 0.169]

$Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$

- ที่ระดับความเชื่อมั่น 95% ความแตกต่างของสัดส่วนผู้ที่เลือกเบอร์ 1 ของ 2 อำเภอมีค่าอยู่ในช่วง -0.009 ถึง 0.169

ตัวอย่างที่ 4

- ผู้ผลิตยาบริษัทแห่งหนึ่งโฆษณาว่า ได้คิดค้นยาชนิดใหม่ ซึ่งมีสรรพคุณในการบรรเทาอาการปวดได้ดีกว่ายาชนิดเก่า โดยทดลองให้ยาชนิดใหม่แก่ผู้ป่วยกลุ่มที่หนึ่ง 80 คน และให้ยาชนิดเก่าแก่ผู้ป่วยกลุ่มที่สอง 80 คนเท่ากัน ปรากฏว่าหลังรับประทานยาผู้ป่วยกลุ่มที่หนึ่งจำนวน 64 คน มีอาการดีขึ้น และผู้ป่วยกลุ่มที่สองจำนวน 56 คน มีอาการดีขึ้น

ท่านคิดว่ามีเหตุผลเพียงพอหรือไม่ที่จะสรุปว่า ยาชนิดใหม่ให้ผลในการรักษาดีกว่ายาชนิดเก่า โดยใช้ระดับนัยสำคัญ 5%

ต้องการให้ทำการทดสอบสมมติฐาน

- ผู้ผลิตยาบริษัทแห่งหนึ่งโฆษณาว่า ได้คิดค้นยาชนิดใหม่ ซึ่งมีสรรพคุณในการบรรเทาอาการปวดได้ดีกว่ายาชนิดเก่า โดยทดลองให้ยาชนิดใหม่แก่ผู้ป่วยกลุ่มที่หนึ่ง 80 คน และให้ยาชนิดเก่าแก่ผู้ป่วยกลุ่มที่สอง 80 คนเท่ากัน ปรากฏว่า หลังรับประทานยาผู้ป่วยกลุ่มที่หนึ่งจำนวน 64 คน มีอาการดีขึ้น x_1 และผู้ป่วยกลุ่มที่สองจำนวน 56 คน มีอาการดีขึ้น x_2

  ท่านคิดว่ามีเหตุผลเพียงพอหรือไม่ที่จะสรุปว่า ยาชนิดใหม่ให้ผลในการรักษาดีกว่ายาชนิดเก่า โดยใช้ระดับนัยสำคัญ 5%

คำถาม ยาชนิดใหม่ให้ผลในการรักษาดีกว่ายาเก่า ?

$\Rightarrow$ สัดส่วนของคนที่มีอาการดีขึ้นเมื่อทานยาใหม่ มากกว่ายาเก่า

ดังนั้นต้องการทดสอบ $P_1 > P_2$?

สมมติฐานทางสถิติ $\boxed{H_0 : P_1 = P_2}$

$H_1 : P_1 > P_2$

สถิติทดสอบ $$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{64}{80} = 0.8$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{56}{80} = 0.7$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{64+56}{80+80} = 0.75$$

49

สมมติฐานทางสถิติ $H_0 : P_1 = P_2$

$H_1 : P_1 > P_2$

$\hat{p}_1 = 0.8, \hat{q}_1 = 0.2, n_1 = 80$

$\hat{p}_2 = 0.7, \hat{q}_2 = 0.3, n_2 = 80$

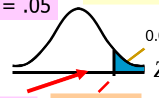
$\hat{p} = 0.75, \hat{q} = 0.25$

สถิติทดสอบ

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$$

เขตวิกฤต

$\alpha = .05$ 　 $Z > 1.645$



คำนวณค่าสถิติทดสอบ

$$Z_c = \frac{0.8 - 0.7}{\sqrt{0.75(0.25)(\frac{1}{80} + \frac{1}{80})}} = 1.46$$

สรุปผล

ไม่ปฏิเสธ $H_0 : P_1 = P_2$

ที่ระดับนัยสำคัญ 0.05 ไม่สามารถกล่าวได้ว่า ผลการรักษาด้วยยาใหม่ดีกว่ายาเก่า

50

ตัวอย่างที่ 5

- โรงแรมแห่งหนึ่งได้ทำการสำรวจตัวอย่างลูกค้าสำรองห้องพักในช่วงฤดูร้อนจำนวน 120 คน พบว่า มี 18 คน ที่บอกยกเลิกการสำรองห้องพัก และลูกค้าสำรองห้องพักในช่วงฤดูหนาว จำนวน 200 คน พบว่า มี 21 คน ที่ยกเลิกการสำรองห้องพัก

  จงทดสอบว่าสัดส่วนลูกค้าสำรองห้องพักจะบอกยกเลิกการสำรองห้องในช่วงฤดูร้อนสูงกว่าในช่วงฤดูหนาวร้อยละ 1 หรือไม่ ที่ระดับนัยสำคัญ 10%

ต้องการให้ทำการทดสอบสมมติฐาน

51

- โรงแรมแห่งหนึ่งได้ทำการสำรวจตัวอย่างลูกค้าสำรองห้องพักในช่วงฤดูร้อน x_1 จำนวน 120 คน พบว่า มี 18 คน ที่บอกยกเลิกการสำรองห้องพัก และลูกค้าสำรองห้องพักในช่วงฤดูหนาว จำนวน 200 คน พบว่า มี 21 คน ที่ยกเลิกการ x_2 สำรองห้องพัก

  จงทดสอบว่าสัดส่วนลูกค้าสำรองห้องพักจะบอกยกเลิกการสำรองห้องในช่วงฤดูร้อนสูงกว่าในช่วงฤดูหนาวร้อยละ 1 หรือไม่ ที่ระดับนัยสำคัญ 10%

คำถาม สัดส่วนลูกค้าที่ยกเลิกการสำรองห้องในฤดูร้อนสูงกว่าฤดูหนาว = 1% ?

ให้ P_1 = สัดส่วนการยกเลิกในฤดูร้อน และ P_2 = สัดส่วนในฤดูหนาว

ดังนั้นต้องการทดสอบ **$P_1 - P_2 = 1\%$** ?

สมมติฐานทางสถิติ $\boxed{H_0 : P_1 - P_2 = 0.01}$

$H_1 : P_1 - P_2 \neq 0.01$

สถิติทดสอบ $$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{18}{120} = 0.15$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{21}{200} = 0.105$$

52

สมมติฐานทางสถิติ $H_0 : P_1 - P_2 = 0.01$

$H_1 : P_1 - P_2 \neq 0.01$

$\hat{p}_1 = 0.15, \hat{q}_1 = 0.85, n_1 = 120$

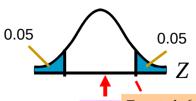
$\hat{p}_2 = 0.105, \hat{q}_2 = 0.895, n_2 = 200$

สถิติทดสอบ

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - c}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

เขตวิกฤต

$\alpha = .1$ 　 $Z < -1.645$ หรือ $Z > 1.645$



คำนวณค่าสถิติทดสอบ

$$Z_c = \frac{(0.15 - 0.105) - 0.01}{\sqrt{\frac{0.15(0.85)}{120} + \frac{0.105(0.895)}{200}}} = 0.9$$

สรุปผล ไม่ปฏิเสธ H_0

ที่ระดับนัยสำคัญ 0.10 ไม่อาจกล่าวว่า สัดส่วนของการยกเลิกห้องในฤดูร้อนสูงกว่าฤดูหนาวแตกต่างจาก 1%

53

ทดลองทำ (แบบฝึกหัดท้ายบท)

- ในการทดลองผสมพันธุ์สัตว์ชนิดหนึ่งระหว่างพ่อพันธุ์ที่มีลักษณะสูงและแม่พันธุ์ที่มีลักษณะเตี้ย ได้ลูกพันธุ์ผสมจำนวน 100 ตัว ซึ่งมีลักษณะสูงจำนวน 35 ตัว และลักษณะเตี้ยจำนวน 65 ตัว จากผลการทดลองดังกล่าว จะสรุปที่ระดับนัยสำคัญ 5% ได้หรือไม่ว่า การผสมพันธุ์สัตว์ในลักษณะข้างต้น จะให้ลูกพันธุ์ผสมลักษณะสูงน้อยกว่าลักษณะเตี้ย

54

ทดลองทำ (แบบฝึกหัดท้ายบท)

- โรงงานผลิตกลอนประตูแห่งหนึ่งต้องการศึกษาเกี่ยวกับประสิทธิภาพของเครื่องจักร 2 เครื่องที่ผลิตกลอนประตู จึงสุ่มตัวอย่างกลอนประตูที่ผลิตโดยเครื่องจักร ก และ ข เครื่องละ 200 อัน พบว่ากลอนประตูที่ใช้ไม่ได้จากเครื่องจักร ก และ ข มี 16 และ 8 อัน ตามลำดับ จากข้อมูลเหล่านี้ จะสามารถยืนยันได้ว่า เครื่องจักร ก มีสัดส่วนกลอนประตูที่ใช้ไม่ได้สูงกว่าเครื่องจักร ข เกินกว่า 5% อย่างมีนัยสำคัญที่ระดับ 0.05

55