บทที่ 7
การอ้างอิงเกี่ยวกับสัดส่วน

7.1 ประชากรอ้างอิงคือประชากรแบบทวินาม ซึ่งมีเพียง 2 ลักษณะคือ Success และ failure

π คือสัดส่วนของลักษณะที่สนใจ (Success) ดังนั้น $1 - \pi$ คือ สัดส่วนของ Failure

ทฤษฎีที่ 7.1 $p = x/n =$ (จำนวน S)/n คือค่าประมาณแบบไม่เอียงเฉของ π โดยมี $E(p) = \pi$ และ $V(p) = \pi(1 - \pi)/n$ และเมื่อ n มีค่าโตพอสมควร จะใช้โค้งปกติประมาณการแจกแจงแบบทวินามด้วย $\mu_X = n\pi$, $V(X) = n\pi(1 - \pi)$

สำหรับการแปลงค่า X เป็น standard normal

$$Z = (X - n\pi)/\sqrt{n\pi(1 - \pi)} \sim N(0, 1)$$

เมื่อหารทุกตัวด้วย n จะได้

$$Z = (p - \pi)/\sqrt{\pi(1 - \pi)/n} \sim N(0, 1)$$

และ ช่วงเชื่อมั่นของ π คือ

$$p \pm Z_{\alpha/2}\sqrt{p(1 - p)/n} \text{ -------- (1) หรือ } p \pm Z_{\alpha/2}\sqrt{1/4n} \text{ ---- (2)}$$

ข้อสังเกต วิธีที่ (2) ให้ช่วงเชื่อมั่นกว้างกว่าวิธีที่ (1) เนื่องจากผลคูณของ p(1-p) จะสูงสุด เมื่อ p = .5, p(1-p) = .5(.5) = .25 = 1/4

ขนาดตัวอย่าง

ขนาดตัวอย่าง n เพื่อให้ค่าประมาณของ π คือ p ต่างจากค่าจริง คือ π ไม่เกิน d หน่วย ด้วยความมั่นใจ $(1- \alpha)100\%$ โดยหาจากสูตร 2 สูตร คือ

จากสูตรที่ (1) เมื่อทราบค่าประมาณของ π

$$d = Z_{\alpha/2}\sqrt{p(1-p)/n}$$

$$d^2 = Z^2_{\alpha/2}p(1 - p)/n \text{ ดังนั้น } n = Z^2_{\alpha/2}p(1 - p)/d^2$$

จากสูตรที่ (2) เมื่อไม่ทราบค่าประมาณของ π

$d^2 = Z^2_{\alpha/2}/4n$

$n = Z^2_{\alpha/2}/4d^2$

__ตัวอย่างที่ 1__ โรคจิตชนิดหนึ่งเป็นกับเพศชายมากกว่าเพศหญิง ซึ่งเกิดจากการถ่ายทอดโครโมโซม X (ที่ผิดปกติคือเปราะบางเป็นพิเศษ) จากมารดาสู่บุตรชาย จากการศึกษาชายที่เป็นโรคจิต 150 คน มี 4 คนที่มีโครโมโซม X ผิดปกติ ดังนั้น $p = X/n = 4/150 = .027 = 2.7\%$

__ตัวอย่างที่ 2__ แมลงวันพันธุ์หนึ่ง มีสีฟ้า-เทาและตัวใหญ่เป็น 3 เท่าของแมลงวันปกติ มันจะวางไข่ในแผลสดของสัตว์ ทำให้อักเสบร้ายแรง การควบคุมไม่ให้ขยายพันธุ์ ใช้วิธีฉายรังสีขนาด 2500 rad กับดักแด้เพื่อให้เป็นหมัน จะทำให้ไข่ที่ผสมกับตัวผู้ที่ได้รับรังสีจะเป็นหมันด้วย หลังจากฉายรังสีแล้ว มีไข่ 415 ฟองจาก 500 ฟองที่เป็นหมัน

นั่นคือ $p = 415/500 = .83$ และ 95% ช่วงเชื่อมั่นของ π คือ

วิธีที่ 1 $.83 \pm 1.96\sqrt{.83(.17)/500} = .83 \pm .03 = .80, .86$ หรือ 80%,86%

วิธีที่ 2 $.83 \pm 1.96\sqrt{1/4(500)} = .83 \pm .04 = .79, .87$ หรือ 79%,87%

__ตัวอย่างที่ 3__ ยารักษาสิวชนิดใหม่สามมารถทำให้ผู้เป็นสิวขนาดหนัก 13 คนจาก 14 คน มีอาการดีขึ้นอย่างชัดแจ้ง นั่นคือ $p = 13/14 = .93$ และ 90% ช่วงเชื่อมั่นของ π คือ

วิธีที่ 1 $.93 \pm 2.57\sqrt{.93(.07)/14} = .93 \pm .18 = .75, 1.11 = 75\%, 111\%$

วิธีที่ 2 $.93 \pm 2.57\sqrt{1/4(14)} = .93 \pm .34 = .59, 1.27$ หรือ 59%, 127%

__ตัวอย่างที่ 4__ จากตัวอย่างที่ 3 ถ้าต้องการให้ค่าประมาณของ π แตกต่างจาก π ไม่เกิน $d = .02$ ด้วยความเชื่อมั่น 90% จะต้องใช้ขนาดตัวอย่างเท่าใด?

$p = .93, Z_{.05} = 1.645$

วิธีที่ 1 $n = (1.645)^2(.93)(.07)/(.02)^2 \approx 440$

วิธีที่ 2 ถ้าไม่ทราบว่า $p = .93$

$n = (1.645)^2/4(.02)^2 \approx 1691$

7.2 การทดสอบสมมติฐานเกี่ยวกับ π

(1) $H_0 : \pi \leq \pi_0$ (2) $H_0 : \pi \geq \pi_0$ (3) $H_0 : \pi = \pi_0$

$H_a : \pi > \pi_0$ $H_a : \pi < \pi_0$ $H_a : \pi \neq \pi_0$

(4) $Z = (p - \pi_0)/\sqrt{\pi_0(1 - \pi_0)/n} \sim N(0, 1)$

(5) ปฏิเสธ H_0 ได้ ถ้า Z ตกในเขตวิกฤต

__ตัวอย่างที่ 5__ ยารักษาโรคชนิดหนึ่ง โฆษณาว่ารักษาได้ผลเกิน 50% ถ้าผลการทดลองกับคนไข้ 100 คน มี 57 คนที่หายจากโรค จงสรุปผลโดยใช้ $\alpha = .10$

(1) $H_0 : \pi = .50$, $H_a : \pi > .50$ (2) $\alpha = .10$ (3) เขตวิกฤต เมื่อ $Z_c > Z_{.10} = 1.28$

(4) $Z = (.57 - .50)/ (.5)(.5)/100 = 1.4$

(5) เนื่องจาก $Z_c = 1.4 > 1.28$ จึงอยู่ในเขตวิกฤต จึงปฏิเสธH_0 และสรุปว่า ยาดังกล่าวมีผลในการรักษาสูงกว่า 50%

ข้อสังเกต เนื่องจากปฏิเสธ H_0 จึงมีโอกาสเกิด type I error นั่นคือมีโอกาส 10% ที่จะสรุปผิดพลาด คือ สรุปว่ายามีผลในการรักษาสูงกว่า 50% แต่แท้จริงแล้วมีผลการรักษาไม่เกิน 50%

7.3 การเปรียบเทียบระหว่างสัดส่วนของ 2 ประชากร (แบบทวินาม)

π_1 , π_2 คือ สัดส่วนของลักษณะที่สนใจจากประชากรทวินามที่ 1 และ 2 ตาม ลำดับ โดยมีค่าประมาณ คือ $p_1 = x_1/n_1$ และ $p_2 = x_2/n_2$ เมื่อต้องการเปรียบเทียบ 2 ประชากร คือ $\pi_1 - \pi_2$ ค่าประมาณแบบจุด คือ $p_1 - p_2$

ทฤษฎีบทที่ 7.2 สำหรับตัวอย่างขนาดใหญ่ ค่าประมาณ $(p_1 - p_2)$ จะมีการแจกแจงโดยประมาณแบบปกติด้วย $\mu(p_1 - p_2) = \pi_1 - \pi_2$ และ $V(p_1 - p_2) = \pi_1(1 - \pi_2)/n_1 + \pi_2(1 - \pi_2)/n_2$ และช่วงเชื่อมั่น 100% ของ $\pi_1 - \pi_2$ คือ

$$(p_1 - p_2) \pm Z_{\alpha/2}\sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}$$

<u>ตัวอย่างที่ 6</u> การทดลองใช้ยารักษาไตกับคนไข้ 72 คน โดยแบ่งเป็น 2 กลุ่ม กลุ่มแรกมี 34 คน ให้ใช้ยาดังกล่าว มี 1 คนที่มีอาการไตวาย อีก 38 คน ให้เป็นกลุ่มควบคุม คือได้รับยาหลอก มี 10 คนที่มีอาการไตวาย ดังนั้น

$$p_1 = 1/34 \approx .03,\ p_2 = 10/38 \approx .26,$$

$$(p_1 - p_2) = .03 - .26 = -.23$$

ช่วงเชื่อมั่น 95%ของ $(\pi_1 - \pi_2) = -.23 \pm 1.96\sqrt{(.03)(.97)/34 + (.26)(.74)/38}$

$$= -.23 \pm .15 = -.38, -.08$$

นั่นคือ อัตราผู้ตายแตกต่างกันระหว่าง -38% ถึง -8% พึงสังเกตว่าช่วงเชื่อมั่นไม่รวมค่า 0 แสดงว่าอัตราไตวายของ 2 กลุ่มมีความแตกต่างอย่างมีนัยสำคัญ คือกลุ่มที่ใช้ยาสามารถรักษาได้ผลสูงกว่ากลุ่มไม่ใช้ยา

7.4 การทดสอบสมมติฐานระหว่าง 2 สัดส่วน

(1) $H_0: \pi_1 - \pi_2 = d_0$

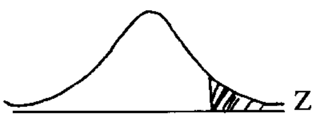
$H_a: \pi_1 - \pi_2 > d_0$

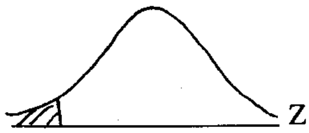
(2) $H_0: \pi_1 - \pi_2 = d_0$

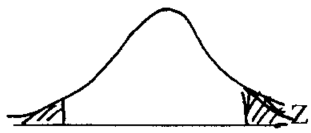
$H_a: \pi_1 - \pi_2 < d_0$

(3) $H_0: \pi_1 - \pi_2 = d_0$

$H_a: \pi_1 - \pi_2 \neq d_0$







(4) ตัวสถิติที่ใช้ทดสอบ

4.1 ถ้า $d_0 \neq 0$

$$Z = \frac{(p_1 - p_2) - d_0}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}}$$

4.2 ถ้า $d_0 = 0$

$$Z = \frac{(p_1 - p_2)}{\sqrt{\hat{\pi}(1-\hat{\pi})(1/n_1 + 1/n_2)}}$$

(5) ปฏิเสธ H_0 ได้ถ้าค่าสถิติตกในเขตวิกฤต , $\hat{\pi} = (x_1 + x_2)/(n_1 + n_2)$

<u>ตัวอย่างที่ 7</u> เชื่อกันว่าไวตามินซี ช่วยให้คนไข้โรคมะเร็งมีอาการดีขึ้นกว่ากลุ่มควบคุมถึง 4% ได้ทดลองโดยแบ่งคนไข้ 150 คนเป็น 2 กลุ่มๆ ละ 75 คน กลุ่มแรกให้กินไวตามินซีวันละ 10 กรัม พบว่าภายใน 4 สัปดาห์ มี 47 คนมีอาการดีขึ้น ส่วนกลุ่มที่ 2 ซึ่งไม่ได้กินยา มี 43 คนที่มีอาการดีขึ้น จงทดสอบโดยใช้ $\alpha = .10$

$p_1 = 47/75 = .63,\ p_2 = 43/75 = .57$

$H_0 : \pi_1 - \pi_2 = .04$, $H_a : \pi_1 - \pi_2 > .04$, CR : $Z_c > Z_{.10} = 1.28$

$$Z = \frac{(.63 - .57) - .04}{\sqrt{(.63)(.37)/75 + (.5))(.43)/75}} = .25$$

เนื่องจาก $Z_c = .25 \leq 1.28$ จึงยังไม่ปฏิเสธ H_0 นั่นคือหลักฐานจากการทดลองยังไม่พอที่จะสรุปว่าไวตามินซีช่วยในการรักษาโรงมะเร็งสูงกว่ากลุ่มควบคุม 4%

ข้อสังเกต ปกติ จะไม่ค่อยระบุค่า d_0 คือมักให้ $d_0 = 0$ นั่นคือไม่สนใจ**ขนาด**ของความแตกต่าง ดังนั้น ความหมายของสมมติฐาน คือ

(1) $H_0 : \pi_1 = \pi_2$	(2) $H_0 : \pi_1 = \pi_2$	(3) $H_0 : \pi_1 = \pi_2$
$H_a : \pi_1 > \pi_2$	$H_a : \pi_1 < \pi_2$	$H_a : \pi_1 \neq \pi_2$

<u>ตัวอย่างที่ 8</u> เชื่อกันว่า จำนวนแมลง black beetles ในท้องถิ่นต่างๆ จะมีสัดส่วนต่างกัน จากการเก็บตัวอย่างจาก Rhode Island 500 ตัว มีสีดำ 95 ตัว แต่จากตัวอย่าง 112 ตัว จาก New York มีสีดำเพียง 17 ตัว จงทดสอบแบบ 2 ด้าน, $\alpha = .05$

(1) $H_0 : \pi_1 = \pi_2$, $H_a : \pi_1 \neq \pi_2$ (2) $\alpha = .05$ (3) CR : $|Z| > Z_{.025} = 1.96$

(4) $p_1 = 95/500 = .19,\ p_2 = 17/112 = .15,\ \pi = (95 + 17)/(500 + 112) = 112/612 = .183$

$$Z = \frac{(.19 - .15)}{\sqrt{(.183)(.817)(1/500 + 1/112)}} = \frac{.04}{04} \approx 1$$

(5) $Z_c = 1.0 \leq 1.96$ จึงยังปฏิเสธ H_0 ไม่ได้ นั่นคือมีหลักฐานไม่เพียงพอที่จะสรุปว่าสัดส่วนแมลงสีดำระหว่างท้องถิ่น มีความแตกต่างกัน

7.5 การประมาณจำนวนสัตว์ป่าที่อาศัยในบริเวณทุ่งกว้าง

นิยมใช้วิธี จับ-ปล่อย-จับซ้ำ โดยมี N เป็นค่าพารามิเตอร์ที่ต้องการประมาณค่า ซึ่งคือจำนวนประชากรของสัตว์ชนิดหนึ่ง เช่น นก, ปลา, กระต่าย ฯลฯ เช่น ต้องการทราบจำนวนประชากรหมีในแถบภูเขาแห่งหนึ่งเพื่อวางโครงการอนุรักษ์ แต่ไม่สามารถใช้วิธีสำรวจประชากรตามปกติเนื่องจากสัตว์เคลื่อนย้ายที่ในบริเวณกว้างอยู่ตลอดเวลา จึงต้องใช้วิธีการทางสถิติ ประมาณค่า N ด้วยวิธีจับสัตว์มาติดเครื่องหมายแล้วปล่อยไประยะเวลาหนึ่ง (หลายๆ วัน หรือหลายสัปดาห์) จึงเก็บตัวอย่างอีกครั้งแล้วนับจำนวนสัตว์ที่มีเครื่องหมาย ซึ่งจะได้สัดส่วนจากการจับครั้งที่ 2 นี้เป็นค่าประมาณของประชากร โดยมีสัญญลักษณ์ ดังนี้

T = จำนวนสัตว์ที่จับได้ในตัวอย่างครั้งที่ 1

π = สัดส่วนสัตว์ที่ติดเครื่องหมายในประชากร = T/N

C = จำนวนสัตว์ที่จับได้ในตัวอย่างครั้งที่ 2 (capture)

R = (recapture) จำนวนสัตว์ที่มีเครื่องหมายในการจับครั้งที่ 2

p = สัดส่วนสัตว์ที่มีเครื่องหมายในการจับครั้งที่ 2 = R/C

เพราะว่า $\hat{\pi} = p$ หรือ $p \approx \pi$

แทนค่า $R/C \approx T/N$

$N \approx CT/R$ หรือ $\hat{N} = CT/R$

เช่นครั้งที่ 1 จับหมีได้ 20 ตัว = T ติดเครื่องหมายแล้วปล่อยไปทั้ง 20 ตัว ต่อมาครั้งที่2 จับได้ 15 ตัว = C พบว่า มี 3 ตัวที่ติดเครื่องหมาย = R ดังนั้น $\hat{N} = CT/R = 15(20)/3 = 100$ ตัว

แบบฝึกหัดที่ 7

1. herpes simplex virus เป็นสาเหตุของโรคจากเพศสัมพันธ์ จากการทดลองให้หญิงที่เป็นโรคดังกล่าว 36 คนใช้ยาชนิดหนึ่ง ซึ่งมีส่วนผสมของน้ำตาล 2-deoxy-D-glucose พบว่าภายใน 4 วัน มี 32 คนมีอาการดีขึ้น จงหาค่าประมาณของ π ซึ่งแสดงผลการรักษา
2. กระทรวงสาธารณสุขต้องการประมาณค่า π ซึ่งคือสัดส่วนของผู้สูบบุหรี่ที่มีอายุไม่เกิน 16 ปีจากการสัมภาษณ์ 1000 คน พบว่ามี 200 คนที่สูบบุหรี่เป็นประจำ จงหา $\hat{\pi}$
3. รายงานแพทย์พบว่าความเครียดมีผลต่อโรคหัวใจและอาจถึงตาย จากสถิติผู้ตาย 15 ราย พบว่ามี 11 รายมีอาการกล้ามเนื้อหัวใจตาย จงหาค่าประมาณของ π ซึ่งคือสัดส่วนผู้ตายเนื่องจากกล้ามเนื้อหัวใจตาย และสร้างช่วงเชื่อมั่น 90% ของ π
4. จากข้อ 1. จงสร้างช่วงเชื่อมั่น 95% ของ π
5. จากข้อ 2. จงหา 96% ของช่วงเชื่อมั่นของ π โดยใช้วิธีที่ (2)
6. การตรวจนับเม็ดเลือดขาวใช้วิธีหยดเลือดบนแผ่นแก้ว เกลี่ยให้เสมอกัน ย้อมสีแล้วส่องกล้อง นับจำนวนเม็ดเลือดขาว ถ้าจากเซลล์เม็ดเลือดขาว 200 เซลล์จากบุคคลหนึ่ง พบเซลล์ neutrophils 125 เซลล์ จงหาสัดส่วนของเซลล์ neutrophils ในเม็ดเลือดขาว และสร้างช่วงเชื่อมั่น 90% โดยวิธีที่ (1)

   ในคนปกติจะมีเซลล์ neutrophils ในเม็ดเลือดขาว ประมาณ 60% ถึง 70% จากช่วงเชื่อมั่น 90% จะสนับสนุนว่าบุคคลผู้นี้ มีเซลล์ neutrophils ในระดับสมดุลย์หรือไม่
7. จากข้อ 3. จงหาขนาดตัวอย่างที่เหมาะสม เพื่อประมาณค่า π โดยให้มีความผิดพลาดไม่สูงกว่า .02 ด้วยความมั่นใจ 95%
8. ต้องการประมาณจำนวนนกเหยี่ยว falcon ในป่าแห่งหนึ่ง ครั้งแรกจับได้ 30 ตัว ติดเครื่องหมายแล้วปล่อยไป ครั้งที่ 2 จับได้ 20 ตัว และมี 2 ตัวที่ติดเครื่องหมาย จงประมาณค่า N ซึ่งคือจำนวนนกเหยี่ยวทั้งหมดที่อาศัยในป่านั้น
9. ในปี 1930 พบว่า หงส์ trumpeter ใกล้สูญพันธุ์ ฝ่ายอนุรักษ์พยายามใช้บริเวณอุทยาน Yellow Stone และทะเลสาป Montana เป็นที่หลบภัย ต่อมาในปี 1960 ได้ลองสำรวจจำนวน โดยจับครั้งแรกได้ 50 ตัว ติดเครื่องหมายแล้วปล่อยไป ครั้งที่ 2 จับได้ 30 ตัว และมี 5 ตัวที่มีเครื่องหมาย จงประมาณจำนวนประชากรของหงส์ดังกล่าวในบริเวณนั้น

10. พวกต่อต้านการสร้างเขื่อน จะอ้างว่าประชาชนส่วนใหญ่ไม่เห็นด้วย ถ้าสัมภาษณ์ 500 คน มี 270 คนที่ต่อต้าน จะสรุปว่าส่วนใหญ่ต่อต้านได้หรือไม่ $\alpha = .05$ ข้อสรุปที่ได้อาจเกิดความผิดพลาดแบบใด จงอธิบาย

11. ผู้ผลิตเครื่องดักแมลงอ้างว่าสามารถกำจัดแมลงในรัศมี 30 ฟุต ได้เกิน 90% ถ้าทดลองปล่อยแมลง 900 ตัวใกล้ๆ เครื่อง อยากทราบว่า ถ้า H_0 เป็นจริง จะคาดหมายว่าจำนวนแมลงสูงสุดที่ดักจับได้ ควรเป็นเท่าใด $\alpha = .10$ ถ้าดักจับได้ 825 ตัว จะสรุปว่าอย่างไร ข้อสรุปอาจเกิดความผิดพลาดแบบใด

12. 90% ของผู้เป็นมะเร็งในปอดจะเสียชีวิตภายใน 3 ปี มีการค้นพบยารักษาโรคนี้ และเชื่อว่าสามารถลดอัตราผู้เสียชีวิต ถ้ามีคนไข้ใช้ยาชนิดใหม่นี้ 150 คน มี 128 คนที่เสียชีวิตภายใน 3 ปี จะสรุปผลว่าอย่างไร ถ้าใช้ $\alpha = .10$ ถ้าใช้ $\alpha = .05$ จะสรุปว่าอย่างไร จะสรุปว่ายาดังกล่าวช่วยลดอัตราผู้เสียชีวิตได้หรือไม่?

13. ได้มีการศึกษา การใช้ยา Anturane (ซึ่งเป็นยาใช้รักษาโรคเก๊าท์) ป้องกัน heart attack ครั้งที่ 2 โดยทดลองกับคนไข้ 733 คน ภายใน 8 เดือน พบว่า มี 13 คนที่ตายเพราะ heart attack ครั้งที่ 2 ส่วนกลุ่มควบคุม (กินยาหลอก) หรือ placebo 742 คน มี 29 คน ที่ตายเพราะ heart attack ครั้งที่ 2 จงหาผลต่างของอัตราการตายของ 2 กลุ่มนี้

14. ยาปฏิชีวนะ doxycycline สามารถป้องกันโรคท้องร่วงในระหว่างเดินทาง ได้มีการทดลองกับอาสาสมัครหมวดสันติภาพที่กำลังเดินทางไปประเทศเคนยา จำนวน 38 คน โดยแบ่งเป็น 2 กลุ่มๆ ละ 19 คน กลุ่มหนึ่งให้กินยา อีกกลุ่มให้กินเม็ดแป้ง พบว่าพวกกินยาไม่มีผู้ใดมีอาการท้องร่วง แต่พวกกินเม็ดแป้ง มี 11 คนที่ไม่มีอาการท้องร่วง จงประมาณค่าผลต่างของอัตราการป้องกันระหว่างผู้กินยาและไม่กินยา

15. จากข้อ 13. จงสร้างช่วงเชื่อมั่น 95% ของ $\pi_1 - \pi_2$ โดยวิธีที่ (1)

16. จากข้อ 13. จงสร้างช่วงเชื่อมั่น 90% ของ $\pi_1 - \pi_2$ โดยวิธีที่ (1) และ เปรียบเทียบกับข้อ 13. ว่าต่างกันอย่างไร มีหลักฐานเพียงพอจะสรุปได้ว่า heart attack ครั้งที่ 2 ในกลุ่มผู้กินยาต่ำกว่าผู้ไม่กินยาหรือไม่ จงอธิบาย

17. จากข้อ 14. จงสร้างช่วงเชื่อมั่น 90% ของ $\pi_1 - \pi_2$ (โดยวิธีที่ 1) มีหลักฐานเพียงพอที่จะสนับสนุนทฤษฎีที่ว่า ยา doxycycline ช่วยป้องกันท้องร่วงระหว่างเดินทางหรือไม่ จงอธิบาย!

18. คนไข้อายุ 40-49 ปี ซึ่งเป็นมะเร็งที่เต้านม จำนวน 31 ราย ใช้เครื่องตรวจมะเร็งเต้านมที่เรียกว่า mammographies มี 6 รายที่เครื่องตรวจพบ ส่วนกลุ่มอายุ 49 ปีขึ้นไปซึ่งเป็นมะเร็งเต้านม จำนวน 101 คน เครื่องตรวจพบ 38 คน จงใช้ $\alpha = .05$ จะมีหลักฐานเพียงพอที่จะสรุปได้หรือไม่ว่าเครื่องดังกล่าวสามารถตรวจพบในกลุ่มผู้สูงอายุได้ดีกว่าหรือไม่

เฉลยแบบฝึกหัดที่ 7

1. $\hat{\pi} = p = 32/36 = .8889 = 88.89\%$

2. $\hat{\pi} = p = 200/1000 = .20 = 20\%$

3. $\hat{\pi} = p = 11/15 = .73$ ช่วงเชื่อมั่น 90% $= .73 \pm 1.645\sqrt{(.73)(.27)/15}$

   $= .73 \pm .19 = .54, .92$ หรือ 54% ถึง 92%

4. 95% ช่วงเชื่อมั่นของ $\pi = .89 \pm 1.96\sqrt{(.89)(.11)/36} = .89 \pm .10 = .79, .99$

5. 90% ช่วงเชื่อมั่นของ $\pi = .20 \pm 2.055\sqrt{1/4(1000)}$

   $= .20 \pm .03 = .17, .23$

6. $\hat{\pi} = p = 125/200 = .625$

   90% ช่วงเชื่อมั่นของ $\pi = .625 \pm 1.645\sqrt{(.625)(.375)/200}$

   $= .625 \pm .056 = .57, .68$

   ถ้าใช้ระดับปกติคือ 60% จะเห็นว่าอยู่ในช่วงเชื่อมั่น จึงถือว่าอยู่ในระดับสมดุลย์

7. $d = .02, \alpha = .05, Z_{.025} = 1.96$

   $n = Z^2_{.025}/4d^2 = (1.96)^2/4(.02)^2 = 2{,}401$

8. $T = 30, C = 20, R = 2, \hat{N} = CT/R = 20(30)/2 = 300$ ตัว

9. $T = 50, C = 30, R = 5, \hat{N} = CT/R = 30(50)/5 = 300$ ตัว

10. $H_0 : \pi = .50, H_a : \pi > .50, CR : Z > Z_{.05} = 1.645$

    $p = 270/500 = .54, Z = (.54 - .50)/\sqrt{(.5)(.5)/500} = 1.789$

    $1.794 > 1.645$ จึงปฏิเสธ H_0 สรุปได้ว่าประชาชนส่วนใหญ่ต่อต้านการสร้างเขื่อน

    ข้อสรุปนี้อาจเกิด type I error คือ แท้จริงผู้ต่อต้านไม่เกิน 50%

11. $H_0 : \pi = .90, H_a : \pi > .90, CR : Z > Z_{.10} = 1.282$

    $Z = (X - n\pi_0)/\sqrt{n\pi_0(1 - \pi_0)} = (X - 900(.90))/\sqrt{900(.9)(.1)}$

    $= (X - 810)/9$

    จากเขตวิกฤต แสดงว่าจะยอมรับ H_0 เมื่อ $(X - 810)/9 \leq 1.282$ หรือ

    $X \leq 1.282(9) + 810 \approx 822$ นั่นคือถ้า H_0 เป็นจริง ควรดักจับแมลงได้อย่างมากที่สุดประมาณ

822 ตัว จาก 900 ตัว ดังนั้น ถ้าจับได้ 825 ตัวจะตกในเขตวิกฤต นั่นคือ มีประสิทธิภาพตามอ้าง อนึ่งอาจตรวจสอบได้โดยวิธีทดสอบสมมติฐาน, $p = 825/900 = .9167$,

$Z = (.9167 - .90)/\sqrt{(.9)(.1)/900} = .0167/.01 = 1.67$ และ $1.67 > 1.282$ จึงปฏิเสธ H_0 ข้อสรุปนี้อาจเกิด type I error คือแท้จริงดักจับได้ไม่เกิน 90%

12. $H_0: \pi = .90, H_a: \pi < .90, \alpha = .10, CR: Z < -1.282$ ($\alpha = .10$)

และ $CR: Z < -1.645$ ($\alpha = .05$), $p = 128/150 = .8533$

$Z = (.8533 - .90)/\sqrt{(.9)(.6)/150} = -1.92$

-1.92 อยู่ในเขตวิกฤต ทั้ง $\alpha = .05$ และ $\alpha = .01$ จึงปฏิเสธ H_0 สรุปว่าวิธีใหม่ช่วยลดอัตราการตาย

13. $p_1 = 13/733 = .0177$, $p_2 = 29/742 = .0391$, $\widehat{\pi_1 - \pi_2} = p_1 - p_2 = .0177 - .0391 = -.0214$ หรือ 2.14% หมายถึงอัตราการตายของผู้กินยาต่ำกว่าผู้ไม่กินยา 2.14%

14. $p_1 = 17/19 = .8947$, $p_2 = 11/19 = .5789$, $p_1 - p_2 = .8947 - .5789 = .3158 = 31.58\%$

15. ช่วงเชื่อมั่น 95% ของ $\pi_1 - \pi_2$

$= -.0214 \pm 1.96\sqrt{(.0177)(.9823)/733 + (.0391)(.9609)/742}$

$= -.0214 \pm (1.96)(.0086) = -.0214 \pm .0168 = -.0382, -.0046$

16. ช่วงเชื่อมั่น 90% ของ $(\pi_1 - \pi_2) = -.0214 \pm 1.645(.0086) = -.0214 + .0141 = -.0355, -.0073$

จะเห็นว่า ช่วงเชื่อมั่น 95% จะยาวกว่า 90% เพราะมีความมั่นใจสูงกว่า จึงต้องใช้ความระมัดระวังมากกว่า เพื่อให้ช่วงเชื่อมั่นครอบคลุมพารามิเตอร์ สูงกว่า 90% จากช่วงเชื่อมันทั้ง 2 อัน จะเห็นว่าไม่รวม 0 เพราะเครื่องหมายขีดจำกัดบนและล่าง ไม่เปลี่ยนแปลง แสดงว่าไม่รวมค่ากล่าว $H_0: \pi_1 - \pi_2 = 0$ แต่รวม $H_a: \pi_1 - \pi_2 \neq 0$ จึงสรุปได้ว่า อัตรา heart attack ครั้งที่ 2 มีความแตกต่างกัน คือกลุ่มที่กินยาจะป้องกันได้ดีกว่า

17. ช่วงเชื่อมั่น 90% ของ $(\pi_1 - \pi_2)$

$= .3158 \pm 1.645\sqrt{(.89)(.11)/19 + (.58)(.42)/19}$

$= .3158 \pm 1.645(.134) = .3158 \pm .2205 = .0953, .5363$

เนื่องจากขีดจำกัดบน-ล่าง มีเครื่องหมายบวกเหมือนกัน จึงไม่รวม $\pi_1 - \pi_2 = 0$

แต่รวม $\pi_1 - \pi_2 \neq 0$ แสดงว่ามีความแตกต่างระหว่างผลการป้องกัน คือกลุ่มที่กินยาจะป้องกันได้มากกว่า

18. $H_0: \pi_1 = \pi_2$, $H_a: \pi_1 > \pi_2$, $\alpha = .05$, CR : $Z > 1.645$

$p_1 = 38/101 = .3762$, $p_2 = 6/31 = .1935$, $\hat{\pi} = 44/132 = .33$

$Z = (.3762 - .1935)/\sqrt{(.33)(.67)(1/101 + 1/31)} = .1827/.9065$

$= 1.89$

$Z_c = 1.89$ อยู่ในเขตวิกฤต จึงปฏิเสธ H_0 และสรุปว่าเครื่องดังกล่าวตรวจพบมะเร็งในกลุ่มผู้สูงอายุได้ดีกว่า