

1. [15 คะแนน] สมมติว่าสถานการณ์การระบาดของไวรัส covid-19 ร้ายแรงมากจนคนไทยทุกคนติดเชื้อ ตารางด้านล่างแสดงจำนวนประชากรของประเทศไทยแบ่งตามช่วงอายุ และอัตราการเสียชีวิตของโรคนี้สำหรับคนในช่วงอายุต่าง ๆ

(Hint: เพื่อความสะดวกให้ใช้ค่ากึ่งกลางของแต่ละช่วงแทนอายุของคนในช่วงนั้นได้ เช่นช่วงอายุ 20-30 ปีก็ให้ใช้ 25 ปีแทนอายุของคนในช่วงนั้น)

ช่วงอายุ	จำนวนประชากร (ล้านคน)	อัตราการเสียชีวิต (%)	จำนวนคนเสียชีวิต (ล้านคน)
0-10	7.55	0.0	
10-20	8.63	0.2	
20-30	9.62	0.2	
30-40	9.35	0.2	
40-50	11.07	0.4	
50-60	10.56	1.3	
60-70	7.30	3.6	
70-80	3.70	8.0	
80+	1.85	14.8	

- (a) [5 คะแนน] จงคำนวณหาอายุเฉลี่ยของประชากรไทยทั้งประเทศ
- (b) [5 คะแนน] จงคำนวณหาจำนวนผู้ที่คาดว่าจะเสียชีวิตในแต่ละช่วงอายุ และจำนวนผู้เสียชีวิตรวม
- (c) [5 คะแนน] จงคำนวณหาอายุเฉลี่ยของผู้เสียชีวิต
2. [10 คะแนน]
- (a) [5 คะแนน] โดยปกตินิยายที่พิมพ์โดยสำนักพิมพ์แม่น้ำมีอัตราการพิมพ์ผิด 1 จุดใน 50 หน้า จงหาความน่าจะเป็นที่นิยายเล่มหนึ่งที่หนา 300 หน้าจะมีจำนวนจุดที่พิมพ์ผิดเป็น 0, 1, 2 และ 3 จุดตามลำดับ
- (b) [5 คะแนน] ผู้บริหารสำนักพิมพ์แม่น้ำเพิ่งปรับเปลี่ยนระบบใหม่เพื่อแก้ไขการพิมพ์ผิด โดยนิยายเล่มล่าสุดมีความหนา 300 หน้าและมีจำนวนจุดที่พิมพ์ผิดเพียง 2 จุด ผู้บริหารจึงสรุปว่าระบบใหม่นี้มีประสิทธิภาพในการลดการพิมพ์ผิด คำอ้างนี้มีนัยสำคัญทางสถิติหรือไม่ถ้าเราตั้ง confidence level ไว้ที่ 5%
3. [10 คะแนน] ชายคนหนึ่งซื้อสลากกินแบ่งรัฐบาล (ล็อตเตอรี่) งวดละหนึ่งใบทุกงวดติดกันเป็นเวลา 15 ปี (เป็นจำนวนรวมกัน 360 งวด) ตลอดช่วงเวลานั้นเขาไม่เคยถูกรางวัลเลขท้ายสองตัวเลยแม้แต่ครั้งเดียว (รางวัลเลขท้ายสองตัวมีโอกาสถูกเป็น 0.01 ต่อการซื้อหนึ่งงวด) เขาจึงตีโพยตีพายว่ากองสลากกลับแก้งเขาด้วยการล็อกห่วยให้เขาไม่ถูกรางวัล (ในที่นี้ให้พิจารณาเฉพาะรางวัลเลขท้ายสองตัวเท่านั้น ไม่ต้องสนใจรางวัลอื่นๆ)
- (a) [5 คะแนน] คำกล่าวอ้างของเขามีนัยสำคัญหรือไม่ที่ confidence level 1%
- (b) [5 คะแนน] เขาต้องไม่ถูกรางวัลติดต่อกันกี่งวดขึ้นไป คำพูดของเขาจึงจะเริ่มมีนัยสำคัญที่ confidence level 1%

4. [15 คะแนน] เมื่อไม่นานมานี้มีข่าวที่แถลงผ่านเฟซบุ๊กของ ศ. นพ. ยง ภู่วรวรรณ ว่ามีการศึกษาการเกิดภูมิคุ้มกันหลังฉีดวัคซีนโควิดของ AstraZeneca ว่าจากกลุ่มตัวอย่าง 61 คนที่ฉีด มี 59 คนที่เกิดภูมิคุ้มกัน จึงสรุปว่าวัคซีนยี่ห้อนี้มีประสิทธิภาพ 96.7% ในข้อนี้เราจะศึกษาว่าค่าประสิทธิภาพที่ว่ามีมีความน่าเชื่อถือแค่ไหน โดยการคำนวณ 95% confidence interval ของมัน

(ในข้อ (c) ต้องมีการแก้หาคำตอบของสมการในเชิงตัวเลข อนุญาตให้ใช้วิธีใดๆ ก็ได้ในการหาคำตอบเชิงตัวเลขนี้ รวมถึงการใช้โปรแกรมสำเร็จรูปเช่น Mathematica หรือ WolframAlpha ด้วย)

ให้สมมติว่าที่จริงแล้ววัคซีนมีประสิทธิภาพ p โดยที่ p เป็นความน่าจะเป็นที่คนหนึ่งคนที่จะเกิดภูมิคุ้มกัน และจำนวนคนที่เกิดภูมิคุ้มกันหลังจากฉีดวัคซีนให้กลุ่มตัวอย่าง 61 คนนี้สามารถอธิบายได้ด้วย Binomial distribution ที่มี $n = 61$ และ success probability p

- (a) [5 คะแนน] จงเขียนความน่าจะเป็นที่จำนวน success x มีค่าดังต่อไปนี้ โดยให้ติดคำตอบในรูปของ p

i. $x = 59$

ii. $x = 60$

iii. $x = 61$

- (b) [5 คะแนน] จงเขียนความน่าจะเป็นที่จำนวน success x มีค่าในช่วงต่อไปนี้ โดยให้ติดคำตอบในรูปของ p

i. $x \geq 59$ (เราจะเรียกค่านี้ว่า $P(x \geq 59 | p)$)

ii. $x \leq 59$ (เราจะเรียกค่านี้ว่า $P(x \leq 59 | p)$)

- (c) [5 คะแนน]

i. จงหาค่า p_{\min} ซึ่งเป็นค่า p ที่ทำให้ $P(x \geq 59 | p) = 0.025$ (หมายความว่า ถ้าที่จริงแล้ว $p < p_{\min}$ จะทำให้ความน่าจะเป็นที่วัดได้ 59 จาก 61 คนหรือดีกว่า มีค่าน้อยกว่า 2.5%)

ii. จงหาค่า p_{\max} ซึ่งเป็นค่า p ที่ทำให้ $P(x \leq 59 | p) = 0.025$ (หมายความว่า ถ้าที่จริงแล้ว $p > p_{\max}$ จะทำให้ความน่าจะเป็นที่วัดได้ 59 จาก 61 คนหรือแย่กว่า มีค่าน้อยกว่า 2.5%)

ช่วง $[p_{\min}, p_{\max}]$ ที่หามาคือ 95% confidence interval ของค่า p เพราะว่าถ้าค่า p อยู่นอกช่วงนี้ จะสามารถถูก reject ได้ด้วย hypothesis testing แบบ two-tailed ที่ 95% confidence level

5. [15 คะแนน] สี่เหลี่ยมผืนผ้ารูปหนึ่งกว้าง x ยาว y โดยความยาว x และ y ได้มาจากการวัดซ้ำหลายๆ ครั้งดังข้อมูลต่อไปนี้

$$x : 5.3, 4.8, 5.2, 5.1, 4.6$$

$$y : 30.9, 31.5, 28.8, 32.3, 28.2, 30.3, 29.4, 28.6$$

เราจะรายงานค่าเฉลี่ยของปริมาณต่อไปนี้ รวมถึงค่า standard deviation ของค่าเฉลี่ยดังกล่าวได้อย่างไร

- | | |
|----------------------------------|----------------------|
| (a) [3 คะแนน] x | (ความยาวด้านกว้าง) |
| (b) [2 คะแนน] y | (ความยาวด้านยาว) |
| (c) [2 คะแนน] $2(x + y)$ | (ความยาวรอบรูป) |
| (d) [2 คะแนน] xy | (พื้นที่รูป) |
| (e) [3 คะแนน] y/x | (aspect ratio) |
| (f) [3 คะแนน] $\sqrt{x^2 + y^2}$ | (ความยาวเส้นทแยงมุม) |

6. [10 คะแนน] นายหมีขาวทำแล็บฟิสิกส์วัดค่า x มา 5 ครั้ง หลังจากนั้นก็นำมาหาค่า mean และค่า standard deviation ของ mean ได้ค่าเป็น $\bar{x} = 5.0$ และ $\sigma_{mx} = \frac{1}{\sqrt{2}}$ ตามลำดับ

แต่อนิจจา หลังจากคำนวณ \bar{x} กับ σ_{mx} เสร็จหมาของนายหมีขาวก็แย่งสมุดบันทึกผลการทดลองวิ่งหนีไป กว่านายหมีขาวจะตามหามาทันและแย่งสมุดกลับมาได้ก็พบว่ากระดาษได้ขาดไปส่วนหนึ่งแล้ว ข้อมูลที่วัดมาจึงหายไปสองค่า โดยสามค่าที่เหลืออยู่คือ $x = 5, 3, 6$ จงหาว่าข้อมูลสองค่าที่หายไปเป็นเท่าไร

7. [15 คะแนน] พิจารณาข้อมูลคู่อันดับจำนวน n จุด ซึ่งเขียนอยู่ในรูป (x_i, y_i) โดยที่ $i = 1, 2, \dots, n$ และให้สมมติว่าความคลาดเคลื่อนของค่า y ของทุกจุดข้อมูลมีค่าเท่ากัน

- (a) [5 คะแนน] จงแสดงว่าเราสามารถหาสมการเส้นตรง $y = ax + b$ ที่เป็น least square fit ของชุดข้อมูลนี้ได้จากการแก้สมการเมทริกซ์ต่อไปนี้เพื่อหาสัมประสิทธิ์ a และ b

$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}$$

และจงแก้สมการนี้เพื่อหา a และ b

- (b) [3 คะแนน] จงแสดงว่าจุด (\bar{x}, \bar{y}) อยู่บนสมการเส้นตรงนี้ (โดยที่ \bar{x} และ \bar{y} คือค่าเฉลี่ยของ x และ y ของชุดข้อมูล)
- (c) [7 คะแนน] เป็นไปได้หรือไม่ที่เมทริกซ์ด้านบนจะมี determinant เป็นศูนย์ และไม่สามารถคำตอบของสมการได้ หรือพูดในอีกแง่หนึ่งก็คือ มีชุดข้อมูลใดหรือไม่ที่ไม่สามารถหาเส้นตรงที่พิตได้ดีที่สุดด้วยวิธีนี้ได้

8. [10 คะแนน] พิจารณาข้อมูลในตารางต่อไปนี้ (โดยให้สมมติว่าความคลาดเคลื่อนของค่า y ของทุกจุดมีค่าเท่ากัน)

x	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	-1	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$
y	0	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	-1	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$

- (a) [1 คะแนน] จงพล็อตจุดข้อมูลเหล่านี้ลงบนระนาบ $x - y$
- (b) [5 คะแนน] จงแสดงว่าเมื่อใช้ least square fit เพื่อหาเส้นตรงในรูป $y = ax + b$ ที่ดีที่สุดสำหรับข้อมูลชุดนี้ เราจะได้ว่าสมการ $y = 0$ พิตได้ดีที่สุด ($a = 0$ และ $b = 0$)
- (c) [4 คะแนน] เมื่อพิจารณาจากกราฟ ข้อมูลชุดนี้มีความสมมาตรการหมุนรอบจุด $(0, 0)$ ดังนั้นถ้าเส้นตรง $y = 0$ สามารถพิตข้อมูลได้ดี เส้นตรง $y = \pm \frac{1}{\sqrt{3}}x$ และเส้นตรง $y = \pm \sqrt{3}x$ ก็ควรจะพิตข้อมูลได้ดีเช่นเดียวกัน แต่ least square fit กลับให้คำตอบเดียวออกมา จงอธิบายความขัดแย้งที่เกิดขึ้นนี้

หน้านี้เป็นกระดาษเปล่า สามารถใช้ได้ถ้าต้องการ ขอให้เขียนเลขข้อที่กำลังทำให้ชัดเจน