

## SCPY208 การบ้านชุดที่ 4

ส่งเมื่อไหร่ก็ได้ก่อนวันสอบ final

1. ค่าของตัวแปร  $x$  และ  $y$  ที่ได้จากการวัดหลายๆ ครั้งมีค่าดังนี้

$$x : 5.7, 4.5, 4.8, 5.1, 4.9$$

$$y : 61.5, 60.1, 59.7, 60.3, 58.4$$

จงหา mean, standard deviation และ probable error ของปริมาณต่อไปนี้

(a)  $x$

(c)  $y - x$

(e)  $e^y$

(b)  $y$

(d)  $y/x$

(f)  $x/y^3$

2. ค่าของตัวแปร  $x$  และ  $y$  ที่ได้จากการวัดหลายๆ ครั้ง มีค่าดังนี้

$$x : 5.1, 4.9, 5.0, 5.2, 4.9, 5.0, 4.8, 5.1$$

$$y : 1.03, 1.05, 0.96, 1.00, 1.02, 0.95, 0.99, 1.01, 1.00, 0.99$$

จงหา mean, standard deviation และ probable error ของปริมาณต่อไปนี้

(a)  $x$

(c)  $x + y$

(e)  $x^3 \sin y$

(b)  $y$

(d)  $xy$

(f)  $\ln x$

3. ในปัญหานี้เราจะสาธิตผลของ Central Limit Theorem ในกรณีง่ายๆ โดยใช้คอมพิวเตอร์ในการสุ่มตัวเลขจาก uniform distribution

$$f(x) = \begin{cases} 1 & , \quad 0 < x < 1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

(เราสามารถแสดงได้ว่า expected value และ variance ของ distribution นี้มีค่าเป็น  $\mu = \frac{1}{2}$  และ  $\sigma^2 = \frac{1}{12}$ )

- (a) จงสุ่มตัวเลขจาก distribution นี้ขึ้นมา  $n = 20$  ตัว และหาค่า  $\bar{x}$  จากตัวเลขเหล่านี้ (การสุ่มตัวเลขจาก uniform distribution ควรจะทำได้โดยเรียกใช้ฟังก์ชันสำเร็จรูป ไม่ต้องเขียนโปรแกรมเอง)
- (b) ทำขั้นตอนในข้อ (a) ซ้ำ  $m = 100$  ครั้ง ซึ่งจะทำได้ค่า  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$
- (c) คำนวณค่า mean และ standard deviation ของ  $\bar{x}$  ทั้งหมด และเปรียบเทียบค่าที่ได้กับค่าที่ควรจะเป็น
- (d) สร้าง histogram ของค่า  $\bar{x}$  ทั้งหมด และแสดงว่า histogram นี้ทำตัวเหมือน Normal Distribution โดยการพล็อต Normal distribution ที่มี mean และ standard ที่เหมาะสมลงไปเทียบ

4. ค่าของปริมาณ  $x$  ถูกวัด  $n$  มีค่าเป็น  $x_1, x_2, \dots, x_n$  ตามลำดับ และค่าที่วัดเหล่านี้มี standard deviation เป็น  $\sigma_1, \sigma_2, \dots, \sigma_n$  ตามลำดับ ถ้าเราต้องการหา “ค่าเฉลี่ย”  $x_m$  เพื่อใช้เป็นตัวเลขตัวเดียวที่อธิบายข้อมูลชุดนี้ โดยที่  $x_m$  นี้มีคุณสมบัติที่ทำให้

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - x_m)^2}{\sigma_i^2}$$

มีค่าน้อยที่สุด จงหาค่า  $x_m$  และแสดงว่า  $x_m$  เป็นค่าเฉลี่ยถ่วงน้ำหนักในบางแบบของ  $x_1, x_2, \dots, x_n$

5. ข้อมูลชุดหนึ่งประกอบด้วยคู่อันดับ  $(x, y)$  ดังต่อไปนี้ โดยค่า  $\sigma_y$  คือค่าความคลาดเคลื่อนของค่า  $y$

$x$	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
$y$	10.5	2.3	3.9	5.6	9.8	13.1	17.3	21.2	27.7	34.1	32.1
$\sigma_y$	15.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	15.0

- (a) จงหาฟังก์ชันเชิงเส้น  $y_m = ax + b$  ที่อธิบายข้อมูลได้ดีที่สุด
- (b) จงหาฟังก์ชันพหุนามกำลังสอง  $y_m = ax^2 + bx + c$  ที่อธิบายข้อมูลได้ดีที่สุด
- (c) จะเกิดอะไรขึ้นถ้าเราไม่ได้ถ่วงน้ำหนักจุดข้อมูลด้วยค่าความคลาดเคลื่อนอย่างเหมาะสม
6. ข้อมูลชุดหนึ่งประกอบด้วยคู่อันดับ  $(x, y)$  ดังต่อไปนี้ โดยให้สมมติว่าค่าความคลาดเคลื่อนของค่า  $y$  ( $\sigma_y$ ) สำหรับทุกจุดมีค่าเท่ากันหมด

$x$	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
$y$	1.5	1.3	2.1	3.0	5.2	6.7	11.5	16.2	24.6	36.5

- (a) จงหาสมการในรูป  $y_m = a e^{bx}$  ที่อธิบายข้อมูลได้ดีที่สุด โดยสัมประสิทธิ์  $a$  และ  $b$  สามารถหาได้จากการเปลี่ยนรูปสมการให้เป็นเชิงเส้น และทำ linear regression โดยถ่วงน้ำหนักอย่างเหมาะสม
- (b) จงหาสมการในรูป  $y_m = ax^b$  ที่อธิบายข้อมูลได้ดีที่สุด โดยสัมประสิทธิ์  $a$  และ  $b$  สามารถหาได้โดยการเปลี่ยนรูปสมการให้เป็นเชิงเส้นและทำ linear regression คล้ายๆ กับข้อ (a)

ย้าว่าอย่าลืมถ่วงน้ำหนักอย่างเหมาะสมในขั้นตอนการทำ linear regression เพราะสิ่งที่เราต้องการ minimize คือค่า  $\sum_i (y - y_m)^2$  ไม่ใช่  $\sum_i (\ln y - \ln y_m)^2$

7. นาย A กับนาย B เล่นเกมโยนเหรียญพนันหัวก้อยกัน โดยถ้าโยนแล้วเหรียญออกหัวนาย A จะชนะ แต่ถ้าเหรียญออกก้อยนาย B จะชนะ
- (a) ในการโยนเหรียญ 4 ครั้งแรก เหรียญออกหัวและนาย A เป็นฝ่ายชนะทั้ง 4 ครั้ง ทำให้นาย B เริ่มสงสัยว่า อาจจะมีการโกงกัน นาย B สามารถยืนยันข้อสงสัยนี้ได้ที่ระดับ confidence level 5% หรือไม่
- (b) ต้องโยนเหรียญและออกหัวติดกันไปเรื่อยๆ เป็นจำนวนกี่ครั้ง นาย B จึงจะสามารถยืนยันข้อสงสัยว่ามีการโกงกันได้ที่ระดับ confidence level 5%, 1% หรือ 0.1%

- 
8. บริษัทผลิตหลอดไฟแห่งหนึ่งโฆษณาว่าหลอดไฟที่บริษัทผลิตมีค่าเฉลี่ยของอายุการใช้งานที่ 2400 ชั่วโมง โดยมี standard deviation ของอายุการใช้งาน 300 ชั่วโมง วิศวกรตรวจสอบคุณภาพหลอดไฟของบริษัทนี้ว่าเป็นไปตามค่าโฆษณาหรือไม่ จึงสุ่มตัวอย่างมา 100 หลอด และพบว่าอายุการใช้งานเฉลี่ยของหลอดไฟ 100 หลอดนี้มีค่าเป็น 2300 ชั่วโมง วิศวกรตรวจสอบควรจะตัดสินใจว่าบริษัทนี้โฆษณาเกินจริงหรือไม่ ถ้าตั้งระดับ confidence level ไว้ที่ 5%
9. มีข้อมูลในทางการแพทย์ว่าความเข้มข้นของสารโพพรอมบรินในคนปกติมีระดับเฉลี่ย 20 mg/ml โดยมีค่า standard deviation 4 mg/ml บริษัทยาต้องการทดลองยาตัวหนึ่งที่เชื่อว่าจะเปลี่ยนระดับของสารโพพรอมบรินไปจากปกติได้ (โดยการเปลี่ยนนี้อาจจะทำให้มีปริมาณสารนี้มากขึ้นหรือลดลงก็ได้) บริษัทนี้เก็บข้อมูลจากคน 40 คนที่ได้รับยาที่กำลังทดลอง และพบว่าคน 40 คนนี้มีค่าเฉลี่ยของสารโพพรอมบรินเป็น 18.5 mg/ml บริษัทนี้ควรจะสรุปว่าสารนี้เปลี่ยนแปลงระดับโพพรอมบรินอย่างมีนัยสำคัญหรือไม่ถ้าตั้งระดับ confidence level ไว้ที่ 5%