เฉลยแบบฝึกหัดชุดที่ 13

Probability and Statistics

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

---

1. ละอองดาวเป็นเจ้าของธุรกิจร้านขายไอศกรีม เธอสงสัยว่าอุณหภูมิของอากาศจะมีความสัมพันธ์กับยอดขายไอศกรีม ละอองดาวบันทึกผลการสังเกตได้ตามตาราง 1

ตาราง 1: อุณหภูมิและยอดขายไอศกรีม

อุณหภูมิ (C)	**ยอดขาย (บาท)**
33	400
30	365
31	360
35	410
36	422

(a) จงหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอุณหภูมิและยอดขายไอศกรีม อุณหภูมิและยอดขายไอศกรีมมีความสัมพันธ์กันอย่างไร

วิธีทำ กำหนด x เป็นอุณหภูมิ และ y เป็นยอดขาย คำนวณค่าต่างๆ ดังนี้

$$\begin{aligned}
\overline{x} &= \frac{1}{5}(33+30+31+35+36) = 33 \\
\overline{y} &= \frac{1}{5}(400+365+360+410+422) = 391.40 \\
s_x &= \sqrt{\sum_{i=1}^{5}\frac{(x_i-33)^2}{4}} = 2.5495 \\
s_y &= \sqrt{\sum_{i=1}^{5}\frac{(y_i-391.40)^2}{4}} = 27.564
\end{aligned}$$

หาค่าสัมประสิทธิ์สหสัมพันธ์ ดังนี้

$$\begin{aligned}
r &= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{s_x}\right)\left(\frac{y_i-\overline{y}}{s_y}\right) \\
&= \frac{1}{4}\sum_{i=1}^{5}\left(\frac{x_i-33}{2.5495}\right)\left(\frac{y_i-391.40}{27.564}\right) \\
&= 0.9461
\end{aligned}$$

ดังนั้นได้ค่า $r = 0.9641$ เนื่องจากค่า r เป็นบวกจึงสรุปได้ว่าอุณหภูมิและยอดขายไอศกรีมมีความสัมพันธ์

เชิงบวก

(b) จงหาสมการถดถอยสำหรับข้อมูลในตาราง 1

วิธีทำ หาตามขั้นตอนดังนี้

1. หาค่าที่จำเป็นต่างๆ ดังนี้

$$\begin{aligned}
\overline{x} &= \frac{1}{5}(33+30+31+35+36) = 33 \\
\overline{y} &= \frac{1}{5}(400+365+360+410+422) = 391.40 \\
\sum_{i=1}^{5}(x_i-\overline{x})^2 &= 26 \\
\sum_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y}) &= 271
\end{aligned}$$

2. หาค่า $\widehat{\beta}_1$

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{5}(x_i-\overline{x})^2} \\
&= \frac{271}{26} = 10.423
\end{aligned}$$

3. หาค่า $\widehat{\beta}_0$

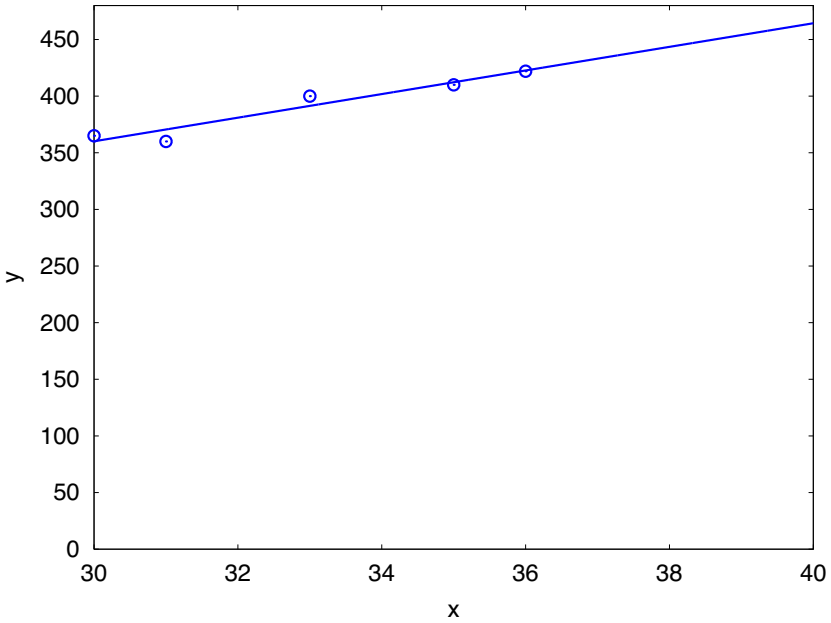
$$\begin{aligned}
\widehat{\beta}_0 &= \overline{y} - \widehat{\beta}_1\overline{x} \\
&= 391.4 - (10.423 \times 33) \\
&= 47.438
\end{aligned}$$

4. เขียนสมการถดถอย คือ

$$\begin{aligned}
y &= \widehat{\beta}_0 + \widehat{\beta}_1 x \\
y &= 47.438 + 10.423\,x
\end{aligned}$$

(c) จงวาดแผนภูมิกระจาย (Scatter Plot) แสดงความสัมพันธ์ของค่าอุณหภูมิและยอดขายไอศกรีม รวมทั้งวาดเส้นตรงที่ได้จากสมการถดถอย

วิธีทำ เมื่อวาดเส้นถดถอยและจุดข้อมูล (x, y) จะได้ดังรูป 1



รูป 1: จุดข้อมูลแสดงความสัมพันธ์ระหว่างอุณหภูมิ (x) และยอดขายไอศกรีม (y) และเส้นถดถอย

ตาราง 2: ราคาน้ำมันและปริมาณรถ

ราคาน้ำมัน (บาท)	**ปริมาณรถ (คัน)**
23.5	250
22.0	270
21.0	262
25.0	231
26.0	240
20.0	284

2. สกาวเดือนสังเกตราคาน้ำมันกับปริมาณรถบนถนนที่ผ่านหน้าบ้านของเธอ ได้ผลตามตาราง 2

(a) จงหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างราคาน้ำมันและปริมาณรถ ราคาน้ำมันและปริมาณรถมีความสัมพันธ์กันอย่างไร

วิธีทำ กำหนด x เป็นราคาน้ำมัน และ y เป็นปริมาณรถ คำนวณค่าต่างๆ ดังนี้

$$\begin{aligned}
\overline{x} &= \frac{1}{6}(23.5+22+21+25+26+20) = 22.917 \\
\overline{y} &= \frac{1}{6}(250+270+262+231+240+284) = 256.17 \\
s_x &= \sqrt{\sum_{i=1}^{6}\frac{(x_i-22.917)^2}{5}} = 2.3327 \\
s_y &= \sqrt{\sum_{i=1}^{6}\frac{(y_i-256.17)^2}{5}} = 19.661
\end{aligned}$$

หาค่าสัมประสิทธิ์สหสัมพันธ์ ดังนี้

$$\begin{aligned}
r &= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{s_x}\right)\left(\frac{y_i-\overline{y}}{s_y}\right) \\
&= \frac{1}{5}\sum_{i=1}^{6}\left(\frac{x_i-22.917}{2.3327}\right)\left(\frac{y_i-256.17}{19.661}\right) \\
&= -0.9197
\end{aligned}$$

ดังนั้นได้ค่า $r=-0.9197$ เนื่องจากค่า r เป็นลบจึงสรุปได้ว่า ราคาน้ำมันและปริมาณรถมีความสัมพันธ์เชิงลบ

(b) จงหาสมการถดถอยสำหรับข้อมูลในตาราง 2

วิธีทำ หาตามขั้นตอนดังนี้

1. หาค่าที่จำเป็นต่างๆ ดังนี้

$$\begin{aligned}
\overline{x} &= \frac{1}{6}(23.5+22+21+25+26+20) = 22.917 \\
\overline{y} &= \frac{1}{6}(250+270+262+231+240+284) = 256.17 \\
\sum_{i=1}^{6}(x_i-\overline{x})^2 &= 27.208 \\
\sum_{i=1}^{6}(x_i-\overline{x})(y_i-\overline{y}) &= -210.92
\end{aligned}$$

2. หาค่า $\widehat{\beta}_1$

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{6}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{6}(x_i-\overline{x})^2} \\ &= \frac{-210.92}{27.208} = -7.7519\end{aligned}$$

3. หาค่า $\widehat{\beta}_0$

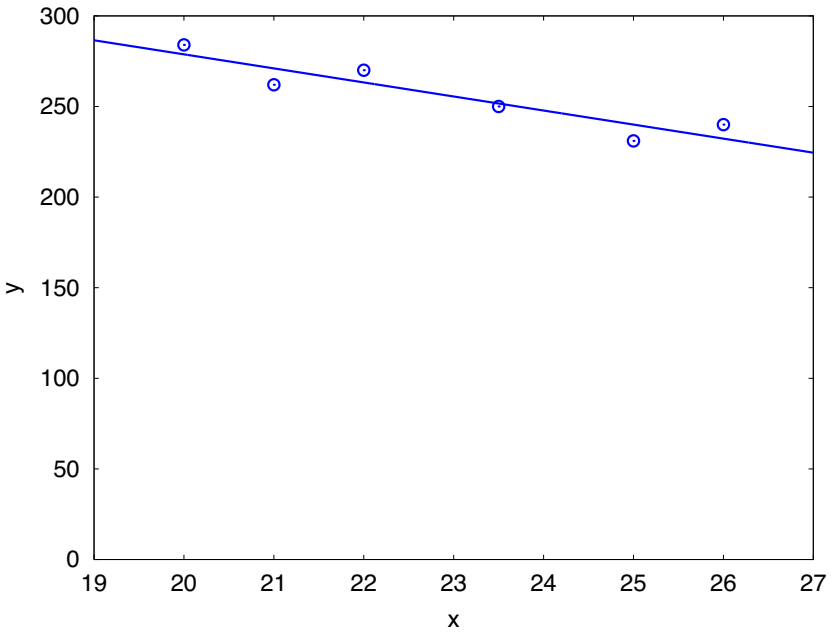
$$\begin{aligned}\widehat{\beta}_0 &= \overline{y}-\widehat{\beta}_1\overline{x} \\ &= 256.17-(-7.7519\times 22.917) \\ &= 433.81\end{aligned}$$

4. เขียนสมการถดถอย คือ

$$\begin{aligned}y &= \widehat{\beta}_0+\widehat{\beta}_1 x \\ y &= 433.81+-7.7519\,x\end{aligned}$$

(c) จงวาดแผนภูมิกระจาย (Scatter Plot) แสดงความสัมพันธ์ของราคาน้ำมันและปริมาณรถ รวมทั้งวาดเส้นตรงที่ได้จากสมการถดถอย

วิธีทำ เมื่อวาดเส้นดถอยและจุดข้อมูล (x, y) จะได้ดังรูป 2



รูป 2: จุดข้อมูลแสดงความสัมพันธ์ระหว่างราคาน้ำมัน (x) และปริมาณรถ (y) และเส้นถดถอย

3. รูป 3 แสดงอัตราการบริโภคช็อคโกแล็ตบนแกน x และจำนวนผู้ได้รับรางวัลโนเบลต่อประชากร 10 ล้านคนของแต่ละประเทศบนแกน y

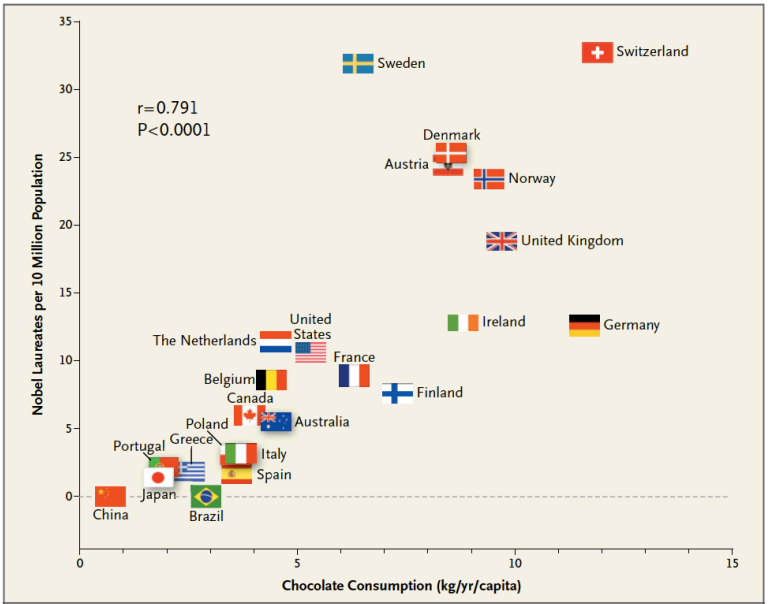


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

รูป 3: อัตราการบริโภคช็อคโกแล็ตและจำนวนผู้ได้รับรางวัลโนเบลต่อประชากร 10 ล้านคนของแต่ละประเทศ (ที่มาของภาพ: New England Journal of Medicine)

(a) ประเทศใดมีอัตราการบริโภคช็อคโกแล็ตสูงที่สุด และประเทศใดมีจำนวนผู้ได้รับรางวัลโนเบลต่อประชากร 10 ล้านคนสูงที่สุด

วิธีทำ ประเทศสวิสเซอร์แลนด์มีอัตราการบริโภคช็อคโกแล็ตสูงที่สุด และประเทศสวิสเซอร์แลนด์มีจำนวนผู้ได้รับรางวัลโนเบลต่อประชากร 10 ล้านคนสูงที่สุด

(b) ปริมาณการบริโภคช็อคโกแล็ตและจำนวนผู้ได้รับรางวัลโนเบล มีความสัมพันธ์กันในลักษณะใด

วิธีทำ มีความสัมพันธ์กันในลักษณะเชิงบวก กล่าวคือ มีแนวโน้มว่าประเทศที่บริโภคช็อคโกแล็ตมากจะมีผู้ได้รับรางวัลโนเบลมากตามไปด้วย

(c) เราจะสามารถสรุปได้หรือไม่ว่า การบริโภคช็อคโกแล็ตส่งผลให้มีจำนวนผู้ได้รับรางวัลโนเบลมากขึ้น กล่าวคือ ยิ่งบริโภคมากขึ้นก็ยิ่งมีจำนวนผู้ได้รับรางวัลมากขึ้น

วิธีทำ ไม่สามารถสรุปเช่นนั้นได้ เพราะสหสัมพันธ์ระหว่างตัวแปรสองตัว ไม่ได้เป็นความสัมพันธ์ในเชิงสาเหตุและผลที่ตามมา (Correlation is not causation) เราไม่อาจสรุปจากข้อมูลนี้ได้ว่าการบริโภคช็อคโกแล็ตส่งผลให้มีจำนวนผู้ได้รับรางวัลโนเบลมากขึ้น