

Problem 6: 6210422036 ธนัท เอี่ยมปรีดี

Sheet: dataP6 (data file: LM7204Final2020.xlsx)

HouseNo	Income	LotSize	MowerOwner
1	60	18.4	Owner
2	85.5	16.8	Owner
3	64.8	21.6	Owner
4	61.5	20.8	Owner
...			
ในข้อสอบ แสดงไม่ครบทั้งหมด. ให้ใช้ข้อมูลที่ครบถ้วนจาก Sheet dataP6			
...			
23	51	14	Non-Owner
24	63	14.8	Non-Owner

บริษัทขายเครื่องตัดหญ้า (mower) สนใจพยากรณ์ความต้องการเครื่องตัดหญ้า. ตารางข้างต้นเป็นข้อมูลของ 24 ครั้วเรือน รายได้ (Income หน่วยเป็น 1000 USD/year), ขนาดสนามหญ้า (หน่วยเป็น 1000 square feet), และครั้วเรือนนี้เป็นเจ้าของเครื่องตัดหญ้า (Owner) หรือไม่เครื่องตัดหญ้า (Non-Owner). สมมติว่าแต่ละครั้วเรือนที่มีเครื่องตัดหญ้ามียี่ห้อเดียวกัน.

6.1 จงทำ logistic regression เพื่อพยากรณ์โอกาสเป็นเจ้าของเครื่องตัดหญ้า (mower owner) โดยใช้ independent variables ทั้งหมด. เขียนสมการ logistic regression ที่ได้ พร้อมทั้งแปลผล coefficients

```
logreg <- glm(MowerOwner ~ Income + LotSize, data=d, family="binomial")
summary(logreg)
```

```
Call:
glm(formula = MowerOwner ~ Income + LotSize, family = "binomial",
    data = d)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.74044  -0.29685   0.00439   0.44750   1.86821

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -25.9382    11.4871  -2.258  0.0239 *
Income       0.1109     0.0543   2.042  0.0412 *
LotSize      0.9638     0.4628   2.083  0.0373 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 33.271  on 23  degrees of freedom
Residual deviance: 15.323  on 21  degrees of freedom
AIC: 21.323

Number of Fisher Scoring iterations: 6
```

MowerOwner: Non-Owner = 0, MowerOwner = 1

$$P(\text{MowerOwner} = 1) = \frac{1}{1 + e^{-(-25.9382 + 0.1109 * \text{Income} + 0.9638 * \text{LotSize})}}$$

Coefficient Interpretation

- **Income:** ถ้า income เพิ่มขึ้น 1 หน่วย (USD 1,000/year) -> odds ของการเป็นเจ้าของเครื่องตัดหญ้า **จะเพิ่มขึ้น 11.09%**
- **LotSize:** ถ้า lot size เพิ่มขึ้น 1 หน่วย (1,000 sq. feet) -> odds ของการเป็นเจ้าของเครื่องตัดหญ้า **จะเพิ่มขึ้น 96.38%**

6.2 สมมติว่าหมู่บ้านจัดสรรโครงการหนึ่ง มีบ้านขนาดใหญ่จำนวน 10 หลัง มีพื้นที่หลังละ 22.2 (หน่วย 1000 square feet) และโดยเฉลี่ยเจ้าของบ้านมีรายได้ 68.7 (1000USD/ปี). มีบ้านขนาดเล็กจำนวน 25 หลังมีพื้นที่หลังละ 18.0 (หน่วย 1000 square feet) และโดยเฉลี่ยเจ้าของบ้านขนาดใหญ่มีรายได้ 49.5 (1000 USD/ปี).

6.2.1 จงหาความน่าจะเป็นที่ บ้านขนาดใหญ่จะมีเครื่องตัดหญ้า

6.2.2 จงหาความน่าจะเป็นที่ บ้านขนาดเล็กจะมีเครื่องตัดหญ้า

```
newdata <- data.frame(Income=c(68.7,49.5),  
                      LotSize=c(22.2,18))  
predict(logreg, newdata = newdata, type="response")
```

```
> predict(logreg, newdata = newdata, type="response")  
      1      2  
0.95579265 0.04299572
```

6.2.1 ความน่าจะเป็นที่ บ้าน**ขนาดใหญ่**จะมีเครื่องตัดหญ้า (LotSize=22.2, Income=68.7) = **0.95579265**

6.2.2 ความน่าจะเป็นที่ บ้าน**ขนาดเล็ก**จะมีเครื่องตัดหญ้า (LotSize=18, Income=49.5) = **0.04299572**

ให้ X เป็น random variable แทนจำนวนเครื่องตัดหญ้ารวมของบ้านขนาดใหญ่

ให้ Y เป็น random variable แทนจำนวนเครื่องตัดหญ้ารวมของบ้านขนาดเล็ก

สมมติให้เหตุการณ์ที่บ้านแต่ละหลังมีเครื่องตัดหญ้าเป็นอิสระ (independent) กัน

6.3.1 จงระบุการแจกแจง (distribution) ของ X และ Y (เช่น normal, Poisson, binomial, exponential เป็นต้น) พร้อมทั้ง parameters ของ distribution นั้น

Binomial Distribution?

$$P(X: n, p) = P(X: 10, 0.96)$$

$$P(Y: n, p) = P(Y: 25, 0.04)$$

6.3.2 ให้จำนวนเครื่องตัดหญ้ารวมทั้งหลังเล็กและใหญ่แทนด้วย $Z = X + Y$

จงคำนวณค่าเฉลี่ย (expectation) ของจำนวนเครื่องตัดหญ้ารวมทั้งหลังเล็กและใหญ่ พร้อมทั้ง variance และ standard deviation.

Distribution	Parameters	Mean	Variance	Standard deviation
N/A	N/A			

$$X : \text{Mean} = n \cdot p = 10 \cdot 0.96 = 9.6$$

$$Y : \text{Mean} = n \cdot p = 25 \cdot 0.04 = 1$$

$$\text{Expected of } Z = 10 \cdot 0.96 + 25 \cdot 0.04 = 10.6?$$

(มั่วมาได้แค่นี้ครับ สิ้นหมดแล้ว 😊)