

Quiz 7 - Multiple regression – python code

ต้องการวิเคราะห์ว่า มี factor ไตบ้างที่ส่งผลต่อราคาขายของสัตว์แปลก ๆ ชนิดหนึ่ง
มีการเก็บข้อมูล น้ำหนัก (กิโลกรัม) และอายุ (เดือน) ของสัตว์ ว่าส่งผลต่อราคาขายหรือไม่
ข้อมูลเก็บไว้ในไฟล์ data/quiz7.csv

ประกอบด้วยข้อมูล 3 คอลัมน์คือ price, weight และ age โดย

price คือราคาขายสัตว์ตัวนั้น (บาท)

weight คือน้ำหนักของสัตว์ (กิโลกรัม)

age คืออายุของสัตว์ (เดือน)



ต้องการวิเคราะห์ทั้งรูปแบบ multiple regression และ simple linear regression

เพื่อหาสมการที่เหมาะสมที่สุด

ตอบตามรูปแบบที่กำหนดเท่านั้น ถ้าไม่ตรงจะไม่ให้คะแนน ตัวเลขให้ตอบเป็นทศนิยม 2 ตำแหน่ง

วิเคราะห์ด้วย multiple linear regression

1. สมการ multiple regression คือ ตอบในรูปแบบเช่น $Y = 0.25\text{Weight} + 0.77\text{Age} + 324.25$
2. ค่า $R^2 =$
3. ค่า MSE =
4. ใช้ multiple regression model วิเคราะห์ว่าน้ำหนัก และอายุ ส่งผลต่อราคาขายสัตว์อย่างไร

วิเคราะห์ด้วย simple regression โดยคิดเฉพาะ weight

5. สมการ linear regression คือ
6. ค่า $R^2 =$
7. ค่า MSE =

วิเคราะห์ด้วย simple regression โดยคิดเฉพาะ age

8. สมการ linear regression คือ
9. ค่า $R^2 =$
10. ค่า MSE =

สัตว์ที่มีน้ำหนัก 120 กิโลกรัม อายุ 40 เดือน จะขายได้ราคาเท่าไร

11. คำนวณโดยใช้ multiple linear regression

12. คำนวณโดยใช้ simple regression (Weight)

13. คำนวณโดยใช้ simple regression (Age)

14. ข้อมูลนี้ควรวิเคราะห์โดยใช้โมเดล regression ไດ เพราะเหตุใด อธิบายเหตุผลให้เหมาะสม

ข้อ 15. – 20. จาก code ที่กำหนดให้ (หน้าถัดไป) เขียนคำสั่งเพื่อให้โปรแกรมสมบูรณ์

ต้องการแสดงกราฟของข้อมูล และ simple linear regression ที่วิเคราะห์โดย Weight หรือ Age

จุดสี pink แสดงข้อมูล weight – price

จุดสี yellow แสดงข้อมูล age – price

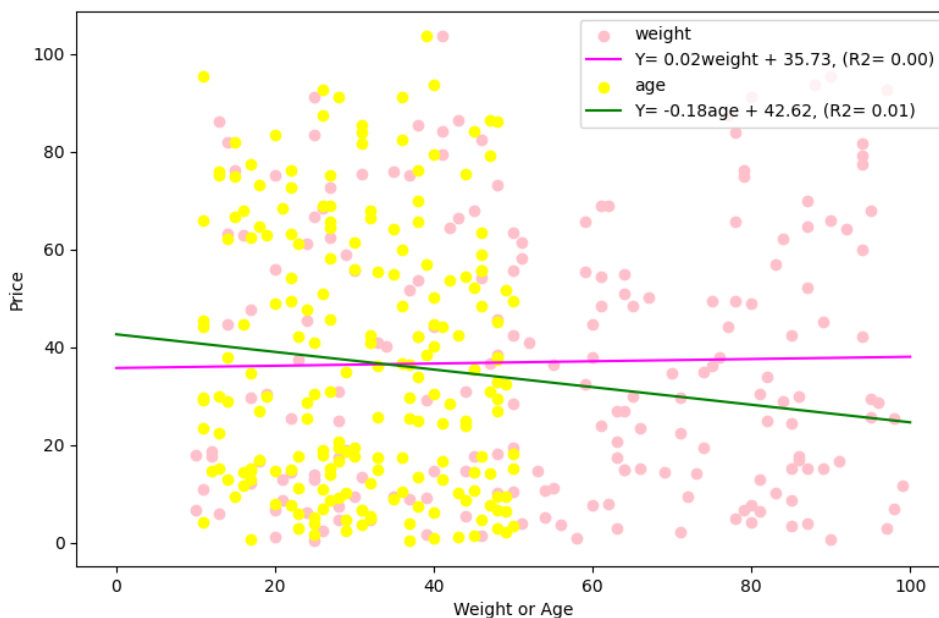
เส้นสี magenta แสดง linear regression (weight)

เส้นสี green แสดง linear regression (age)

แต่ละข้อได้เพียง 1 คำสั่งเท่านั้น ดังนั้นต้องคิดให้ดี ๆ ว่าจะต้องมีคำสั่งใดบ้าง

*การแสดงตัวแปรชนิด String ใช้ %s เช่น ทดสอบคำสั่ง print ('I am %s' % ('Apple'))

ตัวอย่างการแสดงกราฟ



```

// import ...

df = pd.read_csv("data/quiz7.csv")
y = df['price']
xd = ['weight', 'age']
dc = ['pink', 'yellow']      // สีของจุด data
lc = ['magenta', 'green']   // สีของเส้นสมการ

for i in range(len(xd)):
    _ 15. กำหนดค่า x

    _ 16. _____
    _ 17. _____
    b = model.coef_
    a = model.intercept_
    y_pred = model.predict(x.values)
    _ 18. _____

    x1 = np.linspace(0,100,2).reshape(-1,1)
    y1 = model.predict(x1)
    _ 19. แสดงจุดข้อมูล
    _ 20. แสดงเส้นสมการ

plt.xlabel('Weight or Age')
plt.ylabel('Price')
plt.legend()
plt.show()

```