Assignment 1

การวิเคราะห์การถดถอยอย่างง่าย

1.1 จงสร้างแผนภาพการกระจาย (scatter plots) ระหว่างตัวแปร Price และ Floor

คำสั่ง R

```
plot(Homes$Floor, Homes$Price)
```

จากแผนภาพการกระจาย อธิบายความสัมพันธ์ระหว่างตัวแปร Price และ Floor ได้ดังนี้

ตัวแปร Price ไม่มีความสัมพันธ์เชิงเส้นกับ ตัวแปร Floor

1.2 จงหาสมการถดถอยของตัวอย่างสำหรับการทำนายค่าเฉลี่ยของราคาบ้าน (Price) ตามตัวแบบการถดถอยที่กำหนดให้ดังนี้ $Price = \beta_0 + \beta_1 Floor + \varepsilon_i$

คำสั่ง R

```
Homes_center <- data.frame(Price = Homes$Price,Floor_center = scale(Homes$Floor, scale = FALSE))
model_center <- lm(Price ~ Floor_center, data = Homes_center)
summary(model_center)
```

จากผลลัพธ์ของคำสั่ง R ข้างต้น

สมการถดถอยของตัวอย่าง คือ Y_hat = 285.796 + 57.216 * Floor_center

จงอธิบายความหมายของ $\hat{\beta}_0$ และ $\hat{\beta}_1$

$\hat{\beta}_0$ มีค่าเท่ากับ 285.796 หมายถึง บ้านที่มีขนาด (Floor) เท่ากับค่าเฉลี่ย (1.970395 พันตารางฟุต) จะมีราคาเท่ากับ 285.796 หน่วย

$\hat{\beta}_1$ มีค่าเท่ากับ 57.216 หมายถึง เมื่อค่า Floot_center เพิ่มขึ้น 1 หน่วย (พันตารางฟุต) ราคาบ้านจะเพิ่มขึ้น 57.216 หน่วย

1.3 จงเขียนอธิบายผลการทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์การถดถอยของสมการถดถอยในข้อ 1.2

จากผลลัพธ์ของ R ในข้อ 1.2 จงเติมตัวเลขลงในตาราง

ตัวแปร	ค่าประมาณสัมประสิทธิ์การถดถอย $\hat{\beta}_j$	ความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}_j$ ($SE(\hat{\beta}_j)$)	สถิติทดสอบที t-statistic	p-value
Intercept	285.796	6.824	41.878	<2e-16
Floor_center	57.216	32.341	1.769	0.081

จากตารางข้างต้น จงอธิบายการทดสอบสมมติฐานสำหรับ β_0 และ β_1

1 การทดสอบสมมติฐานสำหรับ β_0

สมมติฐานการทดสอบ $H_0: \beta_0 = 0$
$H_1: \beta_0 \neq 0$

สถิติทดสอบ คือ $t = \frac{\hat{\beta}_0 - \beta_0}{\hat{SE}(\hat{\beta}_0)}$

ค่า p-value เท่ากับ <2e-16

สรุปได้ว่า ปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.001 คือ บ้านที่มีพื้นที่ขนาดเท่ากับค่าเฉลี่ย จะไม่มีราคาเท่ากับ 0

2 การทดสอบสมมติฐานสำหรับ β_1

สมมติฐานการทดสอบ $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

สถิติทดสอบ คือ $t = \frac{\hat{\beta}_1 - \beta_1}{\hat{SE}(\hat{\beta}_1)}$

ค่า p-value เท่ากับ 0.081

สรุปได้ว่า ปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.1 คือ ค่าพื้นที่ของบ้านและราคาของบ้านมีความสัมพันธ์เชิงเส้นกัน

1.4 จงเขียนตารางวิเคราะห์ความแปรปรวนของการวิเคราะห์การถดถอยในข้อ 1.2

คำสั่ง R

```
variance ← anova(model_center); variance
sum(variance$`Sum Sq`)
```

สมมติฐานการทดสอบ $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

ตารางวิเคราะห์ความแปรปรวนของการวิเคราะห์การถดถอย

Source	df	SS	MS	F
Regression	1	11079	11078.6	3.1299
Residual	74	261929	3539.6	
Total	75	273007.3		

สรุปได้ว่า ปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.1 คือ ค่าพื้นที่ของบ้านและราคาของบ้านมีความสัมพันธ์เชิงเส้นกัน

1.5 จากผลลัพธ์ในข้อ 1.2 จงอธิบายค่าสัมประสิทธิ์การตัดสินใจและความคลาดเคลื่อนมาตรฐานของส่วนเหลือ

ค่าสัมประสิทธิ์การตัดสินใจ (R^2) มีค่าเท่ากับ 0.04058 หมายถึง เส้นถดถอยสามารถอธิบายความผันแปรของ Y ได้ 4%

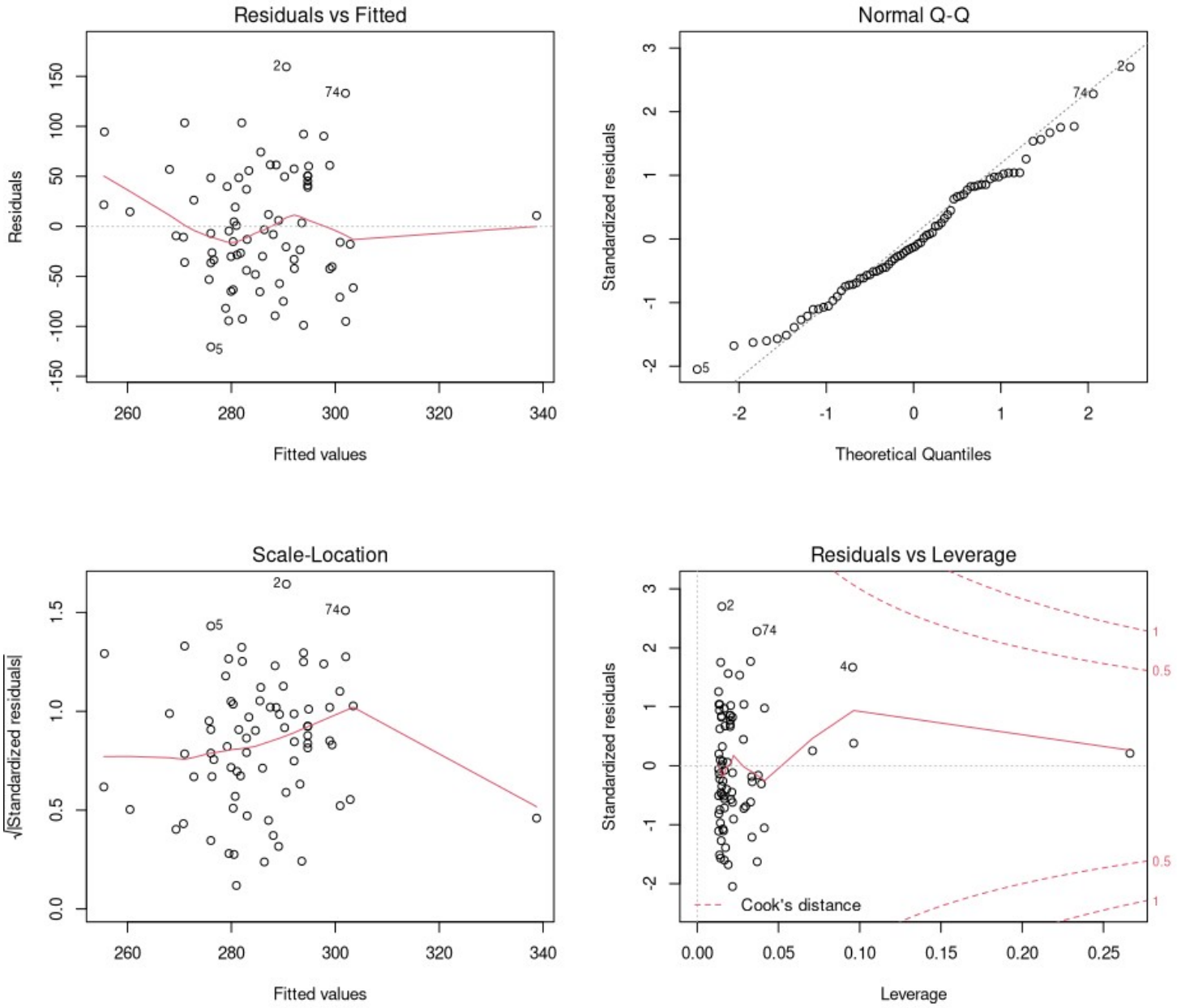
ความคลาดเคลื่อนมาตรฐานของส่วนเหลือ (Residual standard error) มีค่าเท่ากับ 59.49

1.6 จงตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยด้วยแผนภาพของส่วนเหลือ (residual plots)

คำสั่ง R

```
par(mfrow=c(2, 2))
plot(model_center)
```

รูปกราฟที่ได้ คือ



จงอธิบายลักษณะของแผนภาพว่า มีความสอดคล้องกับข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยหรือไม่อย่างไร

Residual vs fitted plot

ไม่สอดคล้อง เพราะ ค่าเฉลี่ยของความคลาดเคลื่อนไม่เท่ากับ 0 (เส้นสีแดงไม่ราบเรียบตามเส้นประ)

Normal Q-Q plot

ไม่สอดคล้อง เพราะ ค่าความคลาดเคลื่อนไม่มีการกระจายแบบการแจกแจงปกติ (จุดจำนวนมากไม่อยู่บนเส้นประ)

Scale-location plot

ไม่สอดคล้อง เพราะ ค่าความคลาดเคลื่อนไม่คงที่ (เส้นสีแดงไม่ราบเรียบ)

Residual vs leverage plot

สอดคล้อง เพราะ ไม่มีค่านอกกลุ่ม (ไม่มีค่าคลาดเคลื่อนใดอยู่เลยเส้นประสีแดง)