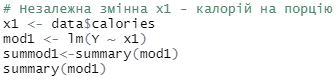
**Лабораторна робота №3**

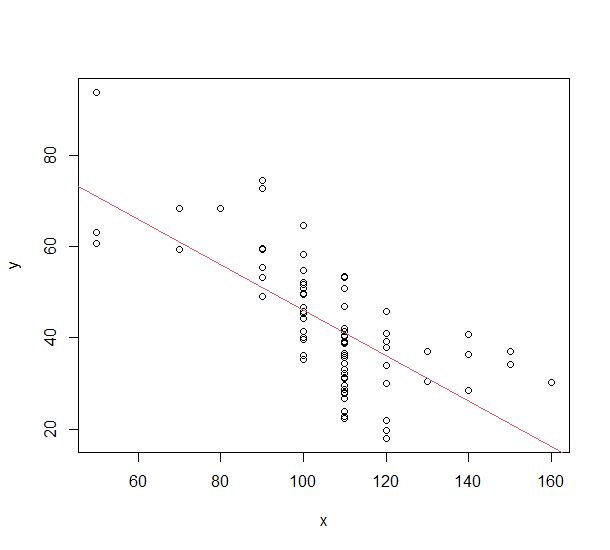
**А.** Для факторів 𝑥1, 𝑥2, 𝑥3 та 𝑦 побудувати в 𝑅 модель 𝑚𝑖: lm(𝑦~𝑥𝑖) та виконати наступні завдання

a. - l. та зробити висновки щодо припущень 1, 2, 3, 4 для кожної пари відповідно.

**Змінна Х1**

1. **Побудувати діаграму розсіювання plot(\*$𝑥𝑖, \*$y) та накласти регресійну лінію.**

****

****

1. **Перевірити значення \*$r.squared та зробити висновки**;

****

Залежна змінна пояснюється незалежною змінною x1 на

47.5%

1. **Перевірити sum(\*$residuals^2) та зробити висновки;**

****

Сума квадратів відхилень достатньо велика

1. **Обчислити var(\*$𝑥𝑖 );**

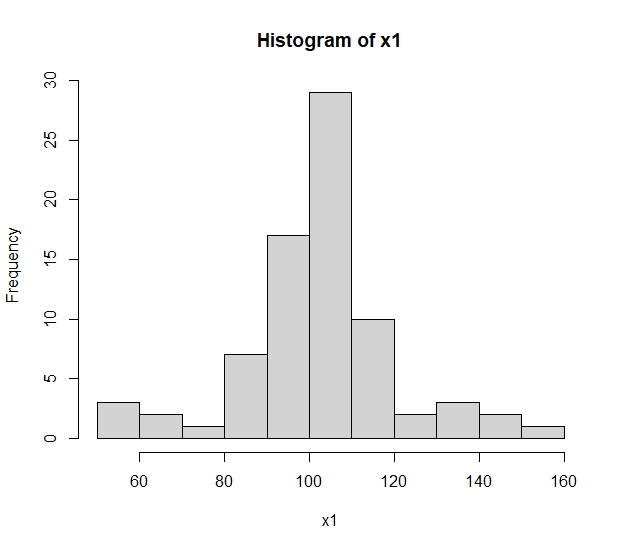
****

1. **Обчислити var(\*$y);**

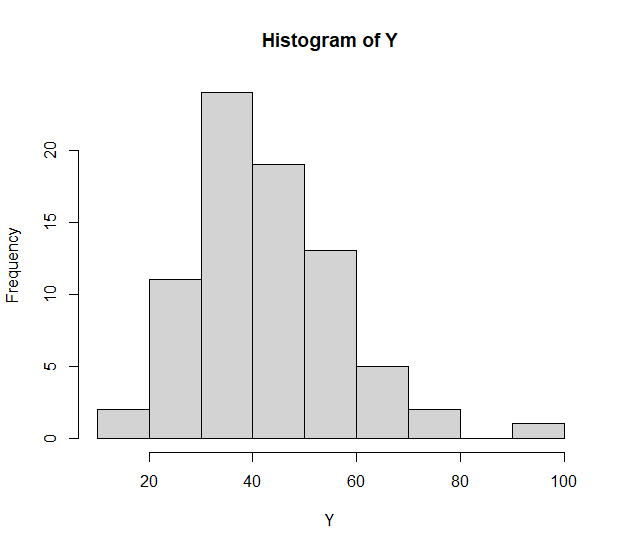
****

Варіація x більше, ніж варіація y, що добре

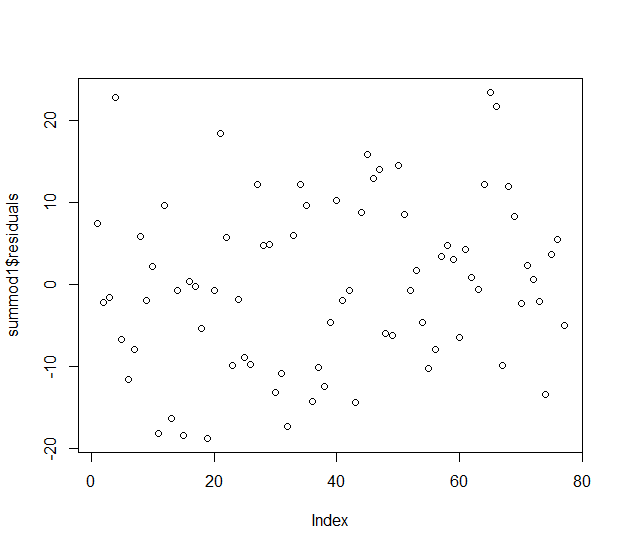
1. **Побудувати hist(\*$𝑥𝑖);**



1. **Побудувати hist(\*$y);**



1. **Побудувати plot(\*$residuals) зробити припущення чи відповідає 𝑁(0; 1);**



Схоже на нормальний розподіл, тобто на N(0, 1)

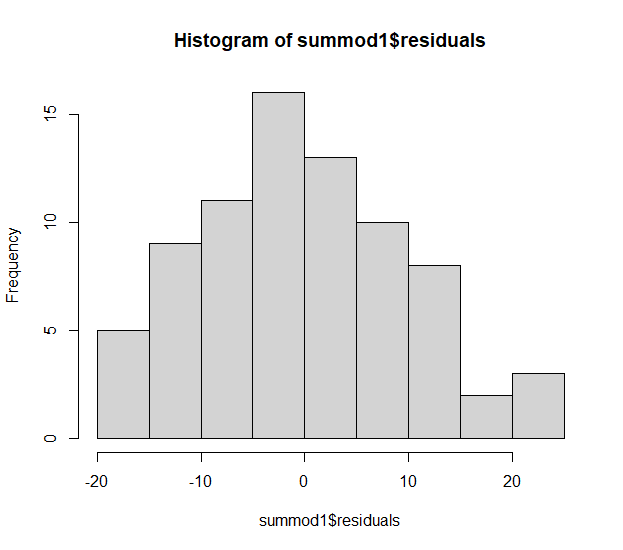
1. **Перевірити mean(\*$residuals);**

****

1. **Обчислити var(\*$residuals);**

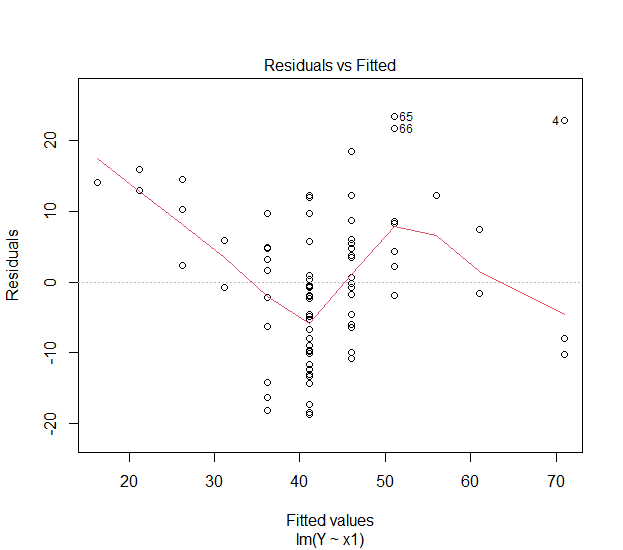
****

1. **Побудувати hist(\*$residuals) та перевірити чи відповідає 𝑁(0; 1);**

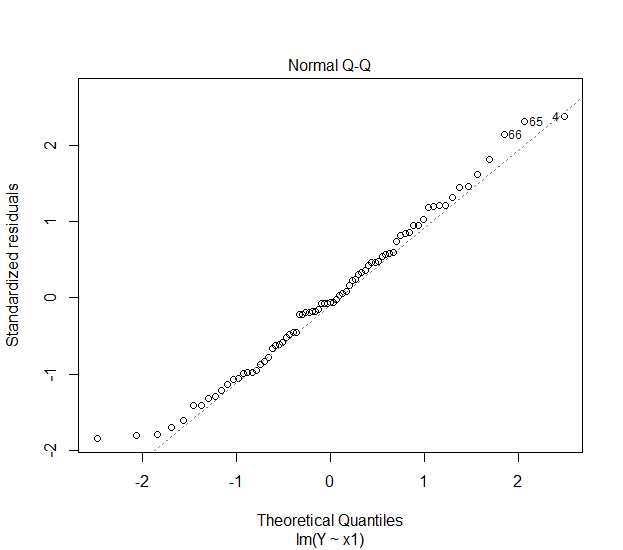


Схоже на нормальний розподіл, але хвости різні

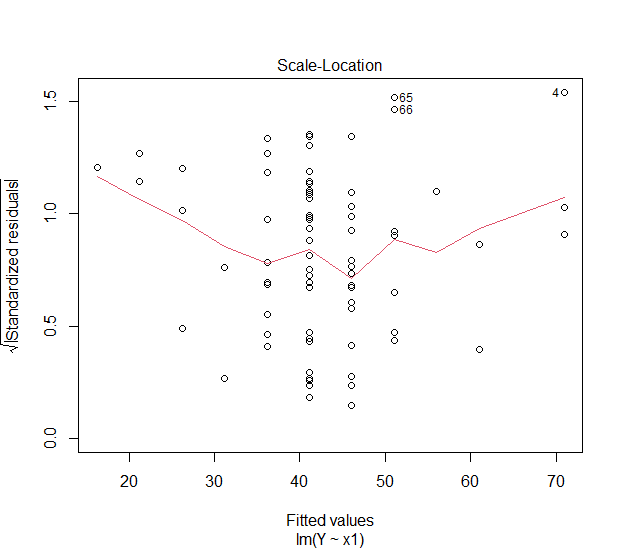
1. **Виконати перевірку 4-х припущень для МНК за допомогою plot(mod\_AGST, 1), plot(mod\_AGST, 2), plot(mod\_AGST, 3), plot(mod\_AGST$residuals, type = "o").**



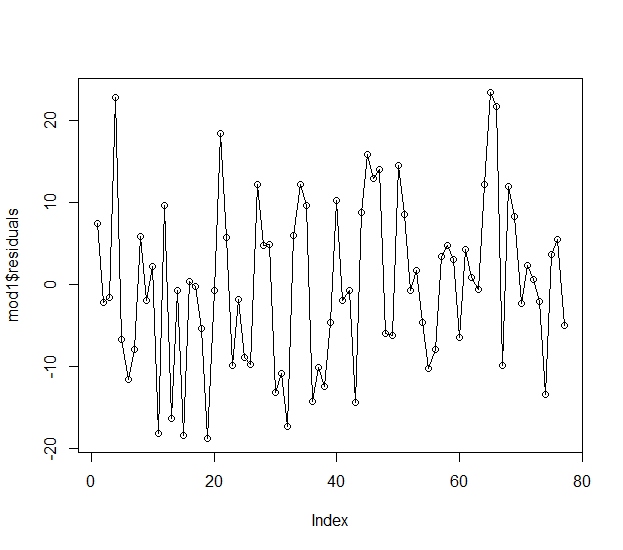
Оскільки червона лінія не близько до пунктирної, то лінійність не дотримується



Майже усі значення лежать на прямій, що означає що дані розподілені нормально

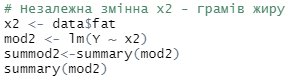


З графіку видно що припущення про гомоскедастичність не виконується, тобто ми маємо гетероскедастичність

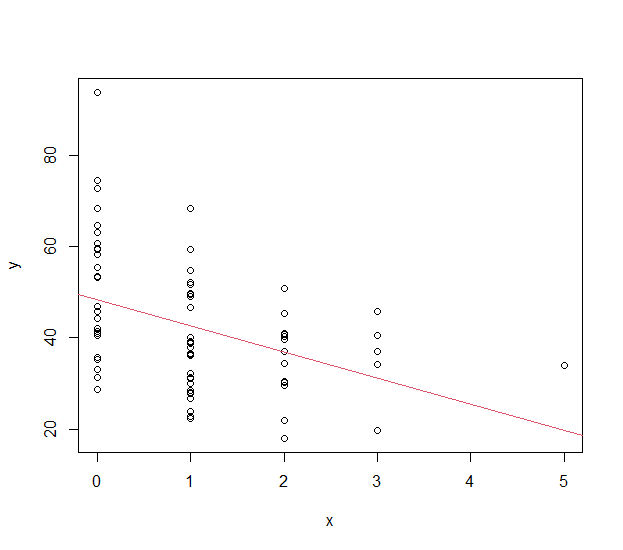


Графік майже повністю заповнює простір, це означає що припущення про незалежність виконується

**Змінна Х2**

****

1. **Побудувати діаграму розсіювання plot(\*$𝑥𝑖, \*$y) та накласти регресійну лінію;**



1. **Перевірити значення \*$r.squared та зробити висновки;**

****

Залежна змінна пояснюється незалежною змінною x1 на

16.7%

1. **Перевірити sum(\*$residuals^2) та зробити висновки;**

****

Сума квадратів відхилень достатньо велика

1. **Обчислити var(\*$𝑥𝑖 );**

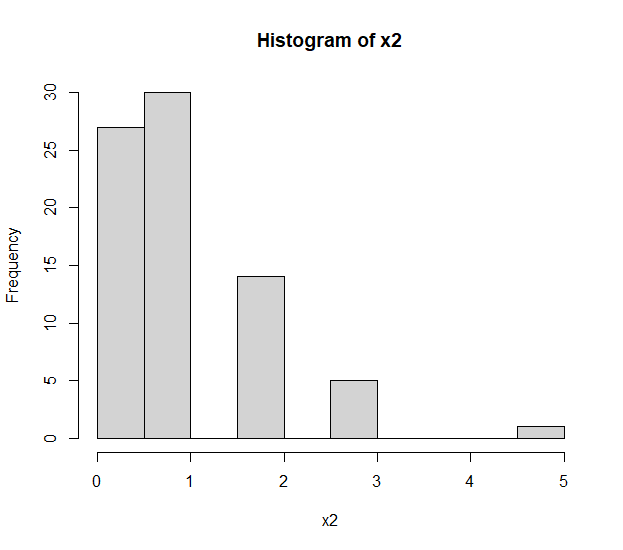
****

1. **Обчислити var(\*$y);**

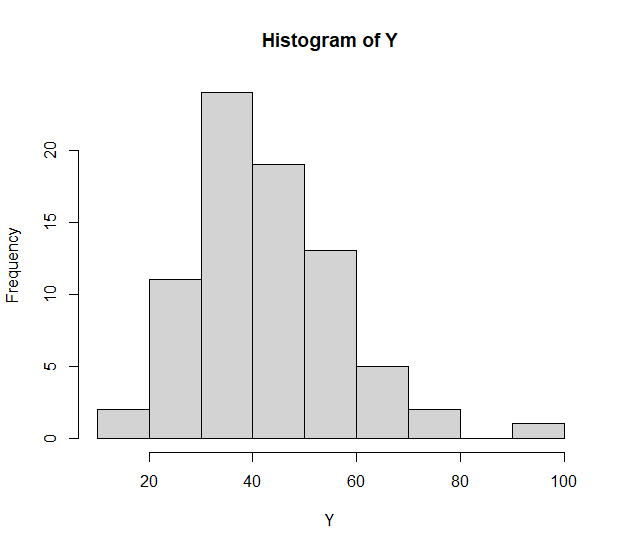
****

Варіація y більше, ніж варіація x, що погано

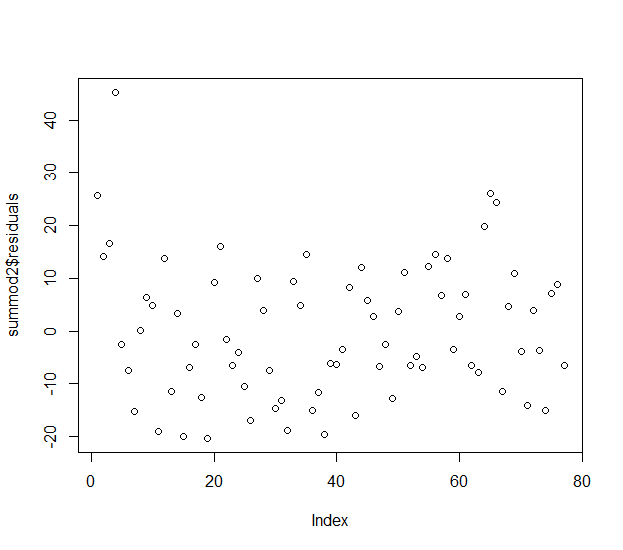
1. **Побудувати hist(\*$𝑥𝑖);**



1. **Побудувати hist(\*$y);**



1. **Побудувати plot(\*$residuals) зробити припущення чи відповідає 𝑁(0; 1);**



Схоже було б на нормальний розподіл, тобто на N(0, 1), але є викид

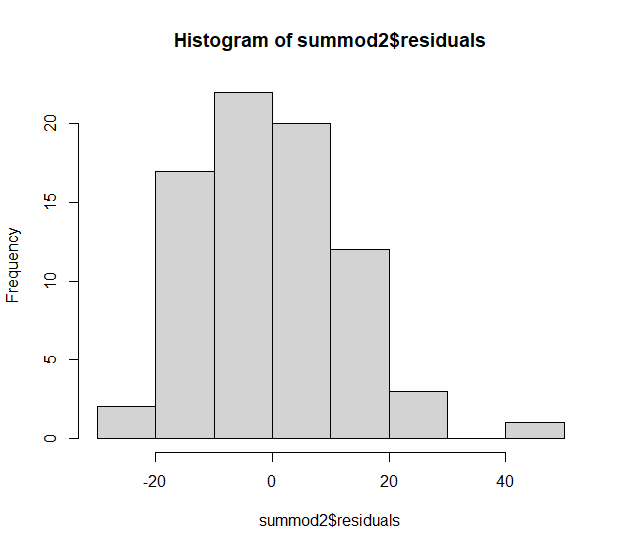
1. **Перевірити mean(\*$residuals);**

****

1. **Обчислити var(\*$residuals);**

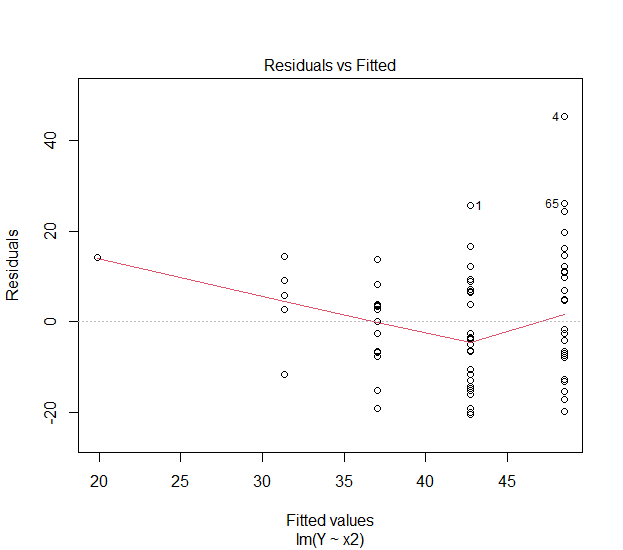
****

1. **Побудувати hist(\*$residuals) та перевірити чи відповідає 𝑁(0; 1);**

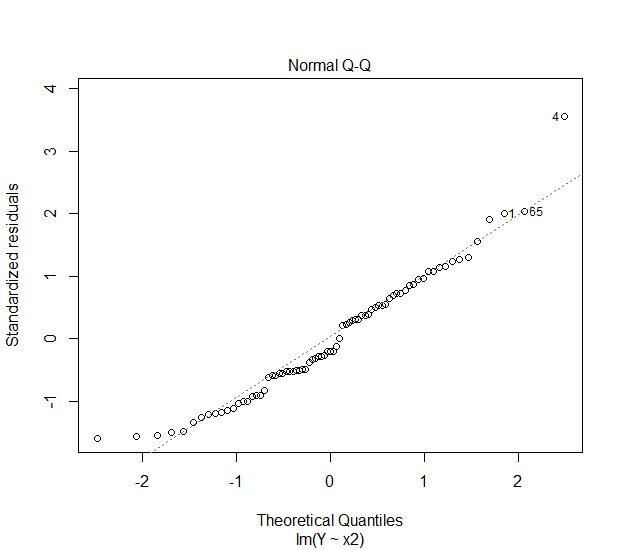


Схоже на нормальний розподіл, але на правому хвості є викид

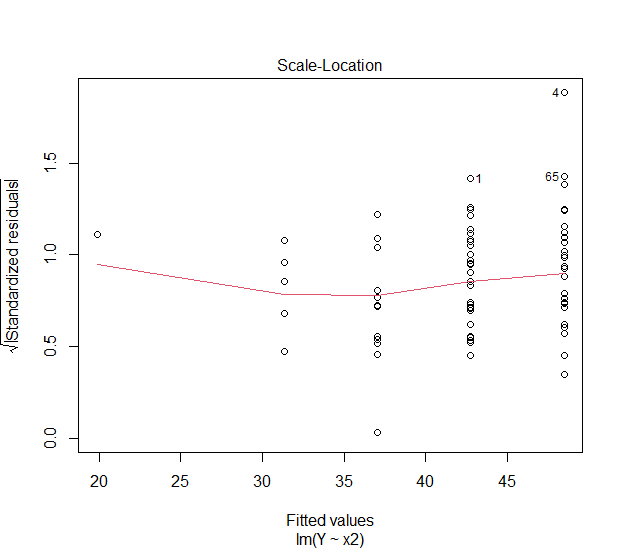
1. **Виконати перевірку 4-х припущень для МНК за допомогою plot(mod\_AGST, 1), plot(mod\_AGST, 2), plot(mod\_AGST, 3), plot(mod\_AGST$residuals, type = "o");**



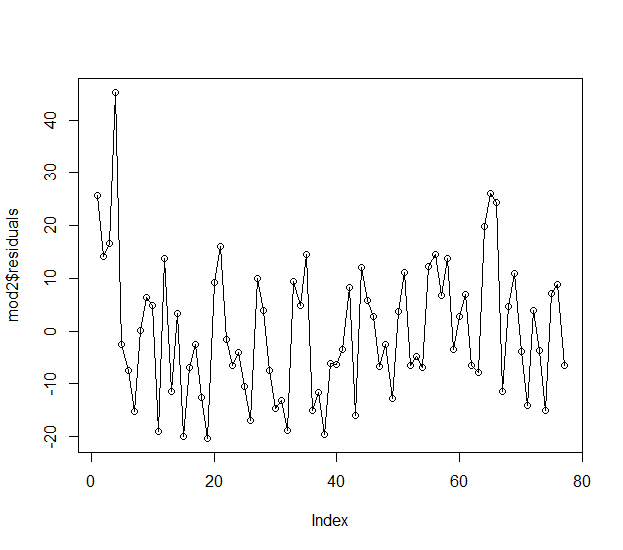
Оскільки червона лінія не близько до пунктирної, то лінійність не дотримується



Майже усі значення лежать на прямій, окрім одного викиду, що означає що данні розподілені нормально

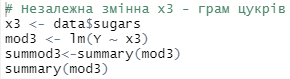


З графіку схоже, що припущення про гомоскедастичність виконується

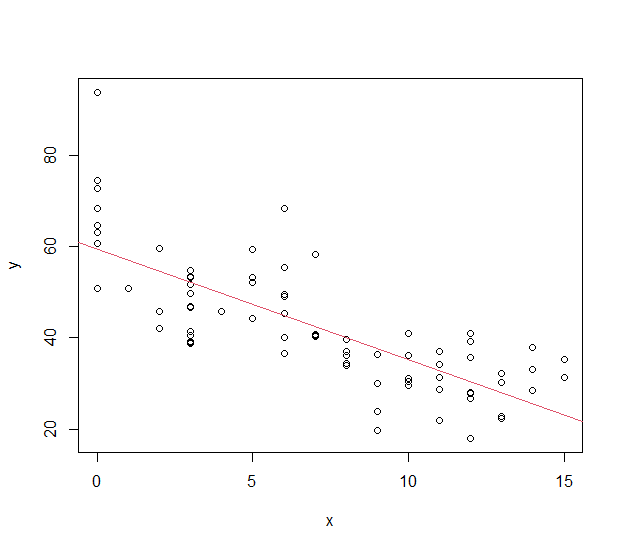


Спочатку є викид, якщо його не було то виконувалось б припущення про незалежність похибок

**Змінна Х3**

****

1. **Побудувати діаграму розсіювання plot(\*$𝑥𝑖, \*$y) та накласти регресійну лінію;**



1. **Перевірити значення \*$r.squared та зробити висновки;**

****

Залежна змінна пояснюється незалежною змінною x1 на

58%

1. **Перевірити sum(\*$residuals^2) та зробити висновки;**

****

Сума квадратів відхилень достатньо велика

1. **Обчислити var(\*$𝑥𝑖 );**

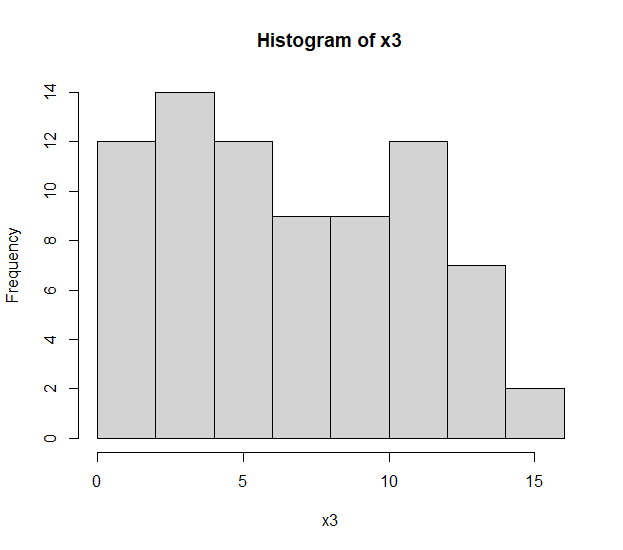
****

1. **Обчислити var(\*$y);**

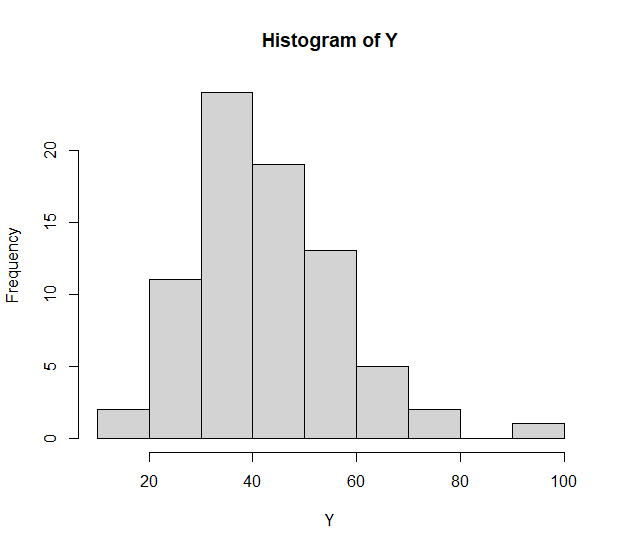
****

Варіація y більше, ніж варіація x, що погано

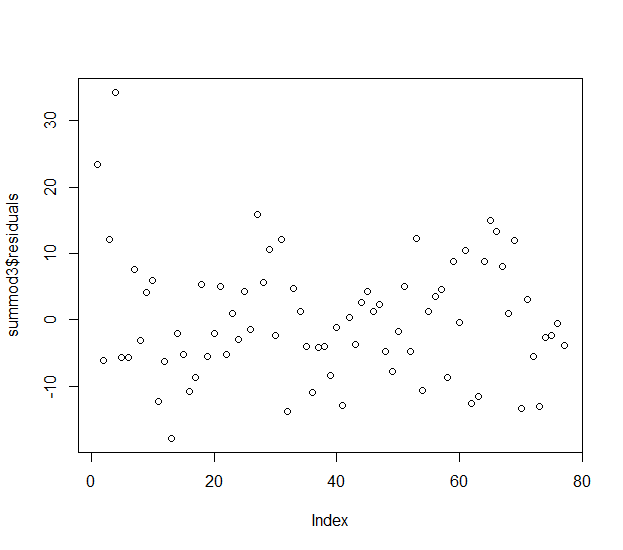
1. **Побудувати hist(\*$𝑥𝑖);**



1. **Побудувати hist(\*$y);**



1. **Побудувати plot(\*$residuals) зробити припущення чи відповідає 𝑁(0; 1);**



Схоже було б на нормальний розподіл, тобто на N(0, 1), але є декілька викидів

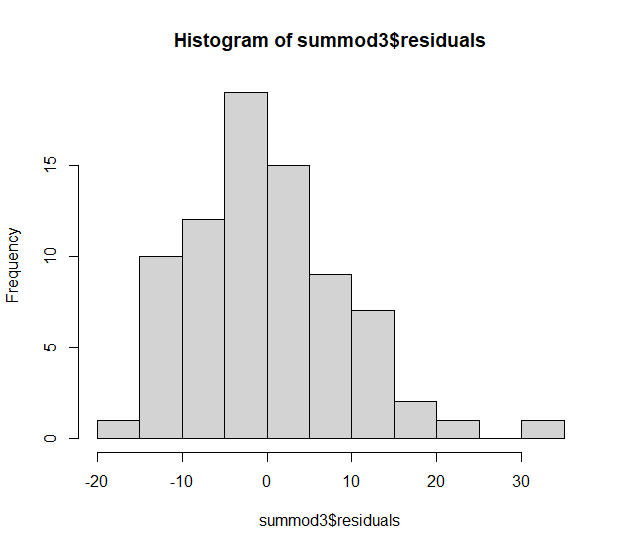
1. **Перевірити mean(\*$residuals);**

****

1. **Обчислити var(\*$residuals);**

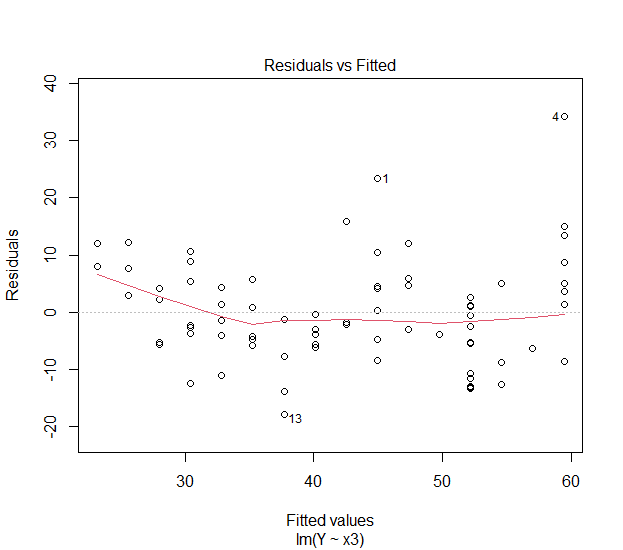
****

1. **Побудувати hist(\*$residuals) та перевірити чи відповідає 𝑁(0; 1);**

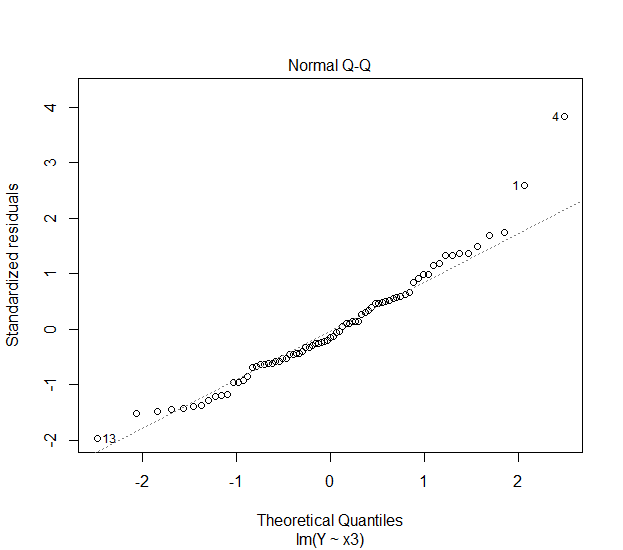


Схоже на нормальний розподіл, але на правому хвості є декілька викидів

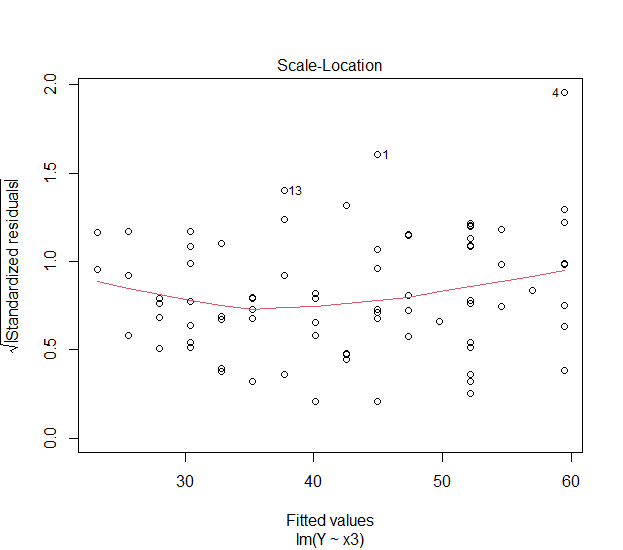
1. **Виконати перевірку 4-х припущень для МНК за допомогою plot(mod\_AGST, 1), plot(mod\_AGST, 2), plot(mod\_AGST, 3), plot(mod\_AGST$residuals, type = "o").**



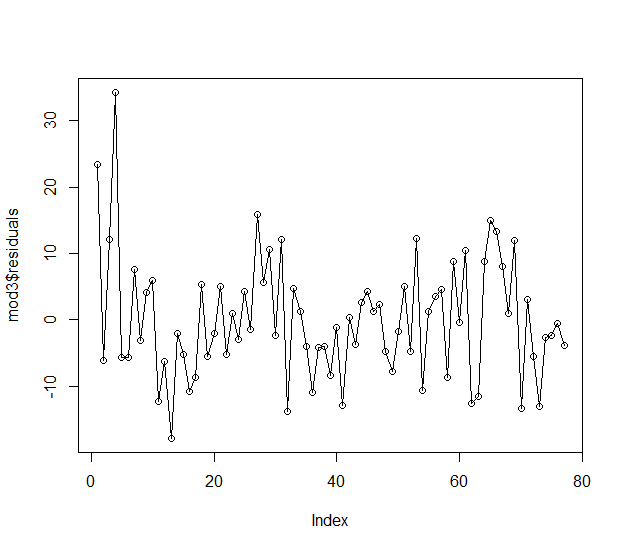
Майже вся червона лінія лежать на пунктирній лінії, тобто можна зробити висновок що лінійність дотримується



Майже усі значення лежать на прямій, окрім декількох викидів в кінці, що означає що данні розподілені нормально



З графіку схоже, що припущення про гомоскедастичність виконується

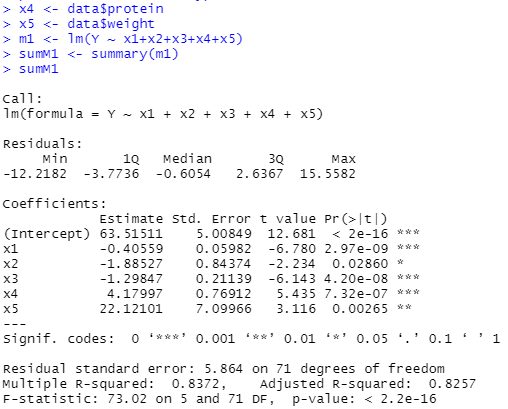


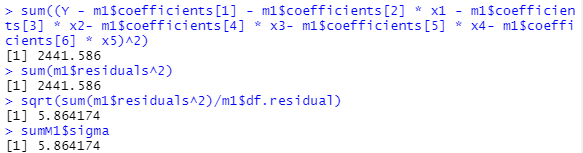
Спочатку є декілька викидів, якщо їх не було то виконувалось б припущення про незалежність похибок

# **B.** Побудувати лінійну модель (m1) за не менше ніж 5-ма параметрами;

𝑌̂ = 𝑏0 + 𝑏1𝑋1 + 𝑏2𝑋2 + 𝑏3𝑋3 + 𝑏4𝑋4 + 𝑏5𝑋5;

# Визначити з summary() чому дорівнює 𝑅𝑆𝐸 та порахувати вручну, а також перевірити чи вони співпадають.

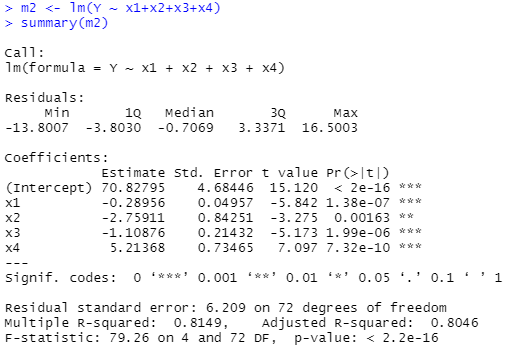
****

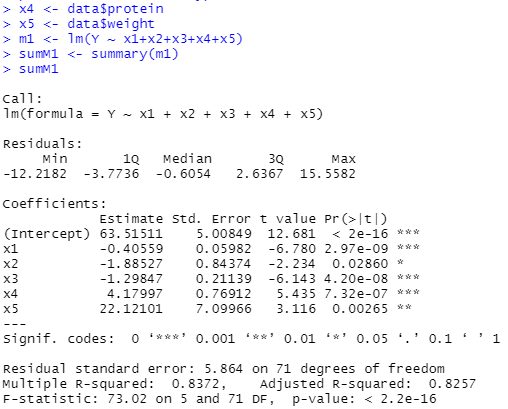
****

RSE = 5.864174

Значення яке ми отримали з summary збіглось з значенням яке ми порахували вручну

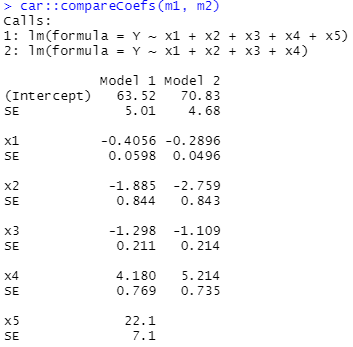
**C.** Створити модель (m2) в якої на 1-н параметр менше; Визначити ступені вільності для (m1) та (m2).

****

****

Для m1 ступені вільності дорівнюють 71, а для m2 - 72

**D.** Порівняти моделі (m1) та (m2) за допомогою функцій summary() та car::compareCoefs(m1, m2) на предмет: 𝑅2, 𝑅𝑆𝐸, 𝑆𝐸(𝛽𝑖 ). Зробити висновок, яка модель краща

****

Summary:

R^2: m1 — 0.8372, m2 — 0.8149. R^2 в моделі m1 вищий, тобто кращий

Adjusted R^2: m1 — 0.8257, m2 — 0.8046. Adjusted R^2 в моделі m1 вищий, тобто кращий

RSE: m1 — 5.864, m2 — 6.209. RSE в моделі m1 менший, ніж в m2, тобто кращий

Для моделі m2 усі SE трішки менше, окрім х3   
  
Можна зробити висновок, що модель m1 є краща, бо вона є кращою майже у всіх параметрах які ми перевіряли