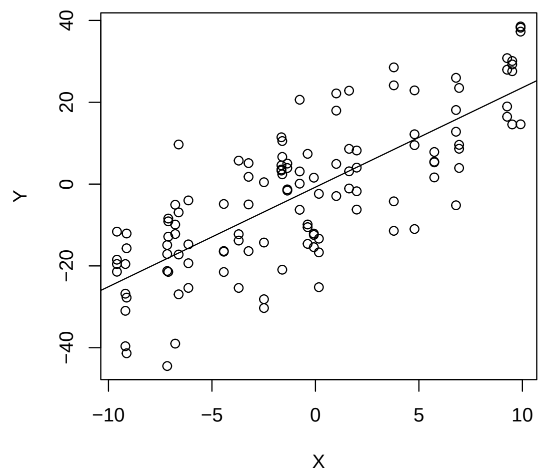
Lesson 8. Conditions for a Simple Linear Regression Model – Part 1

1 Conditions for a simple linear regression model

- Recall that the **simple linear regression model** is

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \text{where} \quad \varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)$$

  - The **errors** ε follow an identical normal distribution and are independent from one another



- When is a simple linear regression model reasonable?
  - Are we justified in using our model? How much can we trust predictions that come from the model?
- We check for the following conditions:

Condition	Explanation
Linearity	• The overall relationship between the variables has a linear pattern • The average values of the response Y for each value of X fall on a common straight line
Independence	• The errors are independent from one another • The distance of one point from the line has no influence on the distance of another point
Normality	• The errors follow a normal distribution
Equal variance	• The variability in the errors is the same for all values of X • In other words, the spread of the points around the line remains fairly constant
Randomness	• The data are obtained using a random process

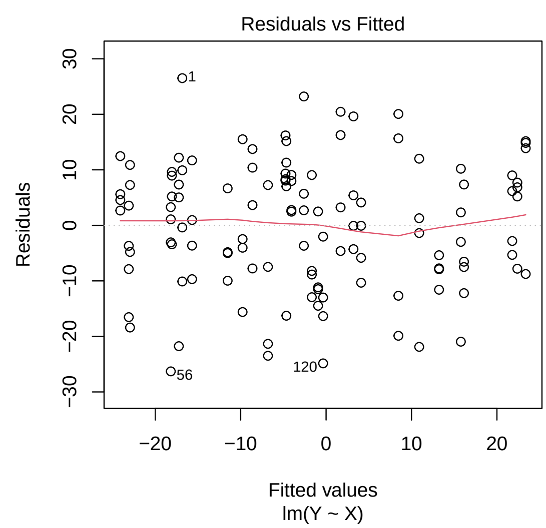
- Mnemonic: **LINER**
- Normality and randomness must be satisfied when we want to use the model for statistical inference (e.g., confidence intervals, hypothesis tests)

2 Assessing conditions for a simple linear regression model

- To easily assess the above conditions, we will use two diagnostic plots

2.1 Residuals vs. fitted values plot

- The **residuals vs. fitted values plot** reorients the axes so that the regression line is represented as a horizontal line through zero
  - $\left\{\begin{array}{c}\text{Positive}\\\text{Negative}\end{array}\right\}$ residuals are represented by points $\left\{\begin{array}{c}\text{above}\\\text{below}\end{array}\right\}$ the regression line
  - This plot lets us focus on any clear patterns in the estimated errors (i.e., residuals)

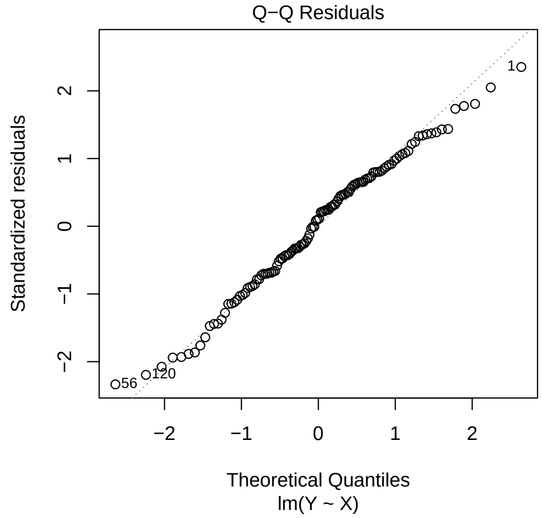


- The linearity condition is satisfied if

- The equal variance condition is satisfied if

2.2 Normal Q-Q plot of residuals

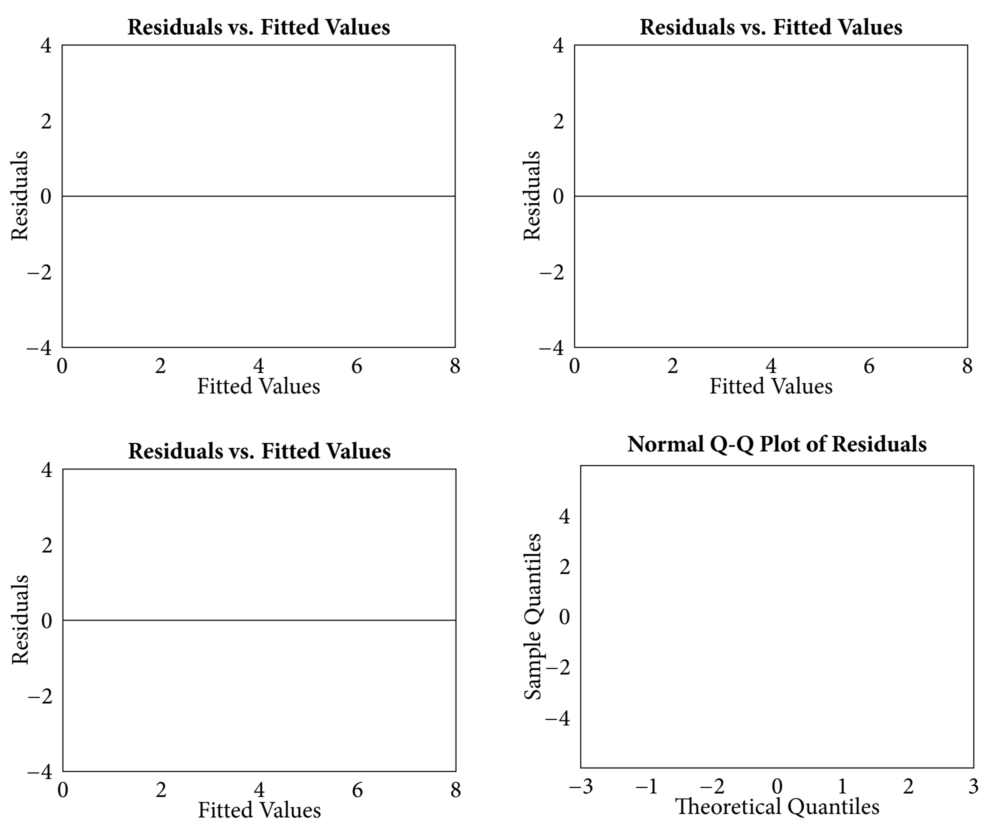
- An ideal Normal Q-Q plot of the residuals looks like



  - The larger the sample size, the more lenient we can be about normality

Example 1. On the blank plots below, sketch the following:

a. A residuals vs. fitted values plot where linearity is met, but equal variance is violated.
b. A residuals vs. fitted values plot where equal variance is met, but linearity is violated.
c. A residuals vs. fitted values plot where both equal variance and linearity are violated.
d. A Normal Q-Q plot that shows dramatic violation of the normality condition.



2.3 Putting it all together...

Condition	Where to check	What we want
Linearity	Residuals vs. fitted values plot	Points randomly and evenly distributed above and below residual = 0 line, moving from left to right
Independence	Description of data collection	No indication that the errors influence each other
Normality	Normal Q-Q plot of residuals	Points in approximately straight line
Equal variance	Residuals vs. fitted values plot	Points span constant vertical width, moving from left to right
Randomness	Description of data collection	Data obtained using a random process, such as random sampling from a population or randomization in an experiment

- No model is perfect, and linear regression is fairly robust to slight violations.
- We will only be concerned with blatant violations.