Week 7 Quiz

CASA0006

Q1. What is the objective of linear regression?

a. To minimize the sum of absolute deviations between the predicted values and the actual values of the dependent variable.

b. To maximize the correlation coefficient between the dependent variable and the independent variable.

c. To minimize the sum of squared deviations between the predicted values and the actual values of the dependent variable.

d. To maximize the adjusted R-squared of the model.

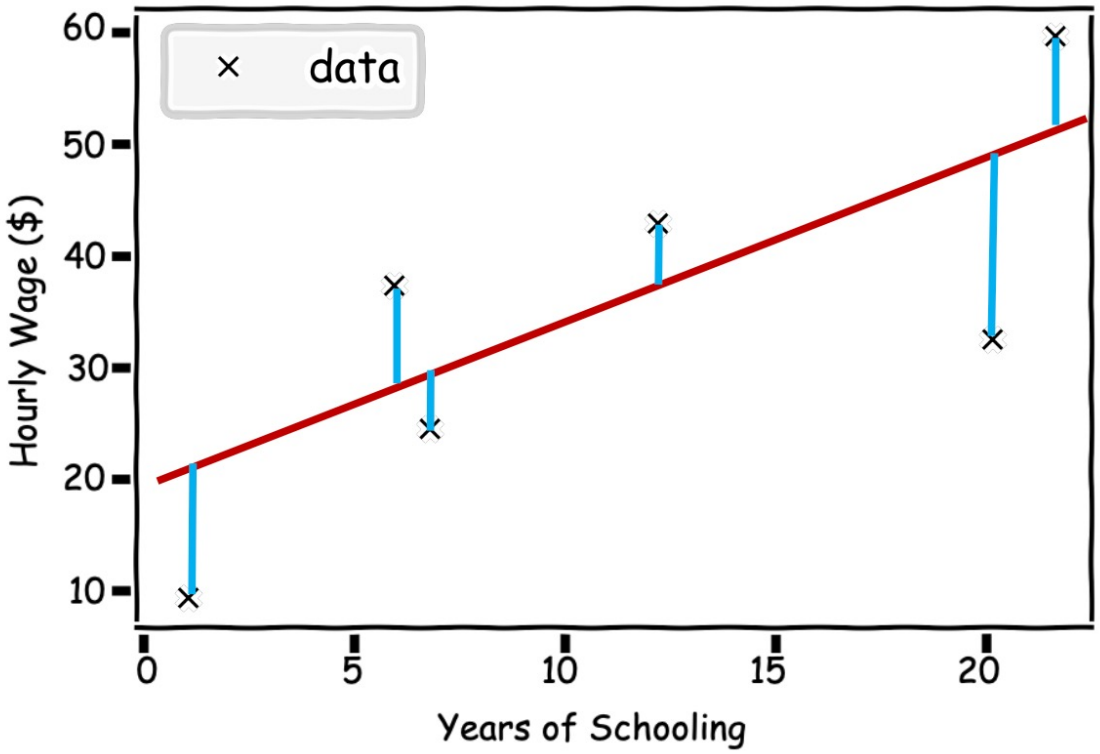
Q1. What is the objective of linear regression?

a. To minimize the sum of absolute deviations between the predicted values and the actual values of the dependent variable.

b. To maximize the correlation coefficient between the dependent variable and the independent variable.

c. To minimize the sum of squared deviations between the predicted values and the actual values of the dependent variable.

d. To maximize the adjusted R-squared of the model.



$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Find the line of best fit

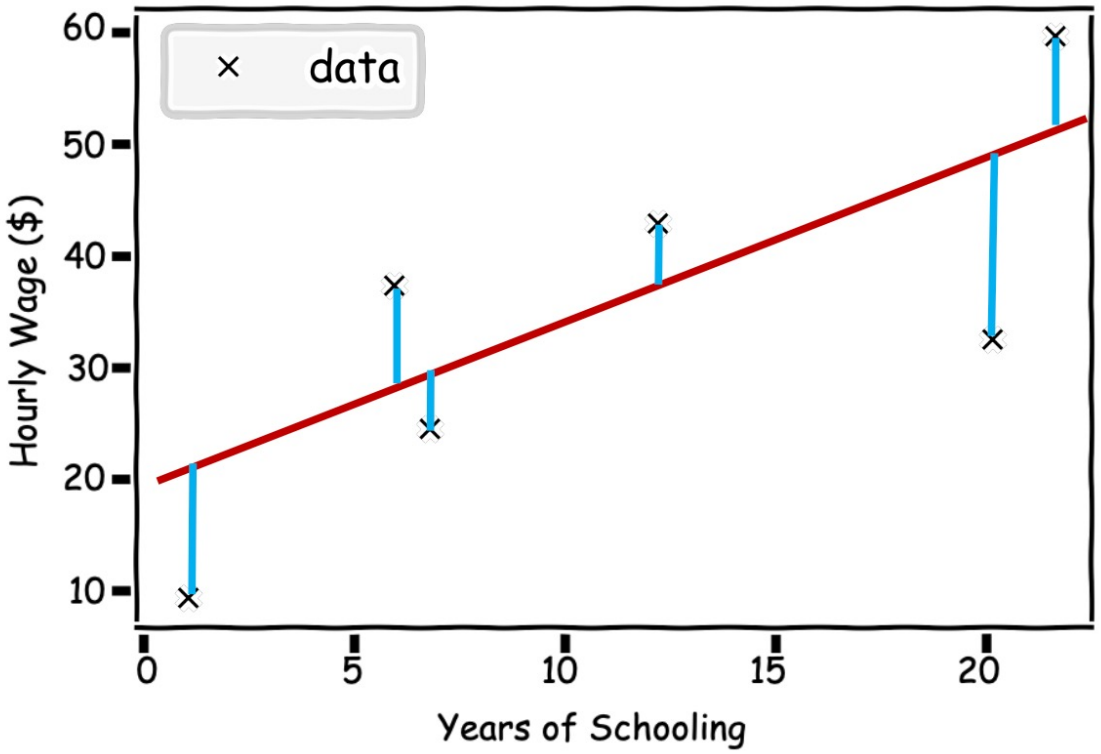
Q1. What is the objective of linear regression?

a. To minimize the sum of absolute deviations between the predicted values and the actual values of the dependent variable.

b. To maximize the correlation coefficient between the dependent variable and the independent variable.

c. To minimize the sum of squared deviations between the predicted values and the actual values of the dependent variable.

d. To maximize the adjusted R-squared of the model.



$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Find the line of best fit

Q2. Which of the following is the assumption of linear regression?

a. The variance of the residuals is constant across all values of the independent variable.

b. The relationship between the dependent variable and independent variable is nonlinear.

c. The independent variables are highly correlated with each other.

d. The residuals are normally distributed.

Q2. Which of the following is the assumption of linear regression?

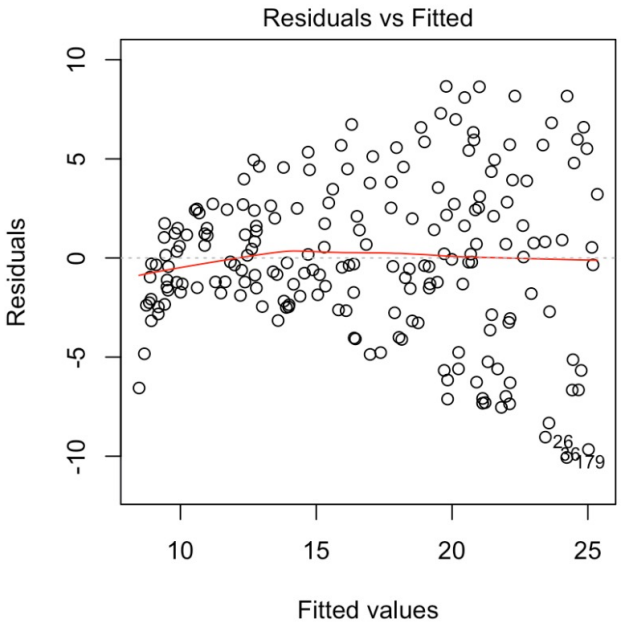
a. The variance of the residuals is constant across all values of the independent variable.

b. The relationship between the dependent variable and independent variable is nonlinear.

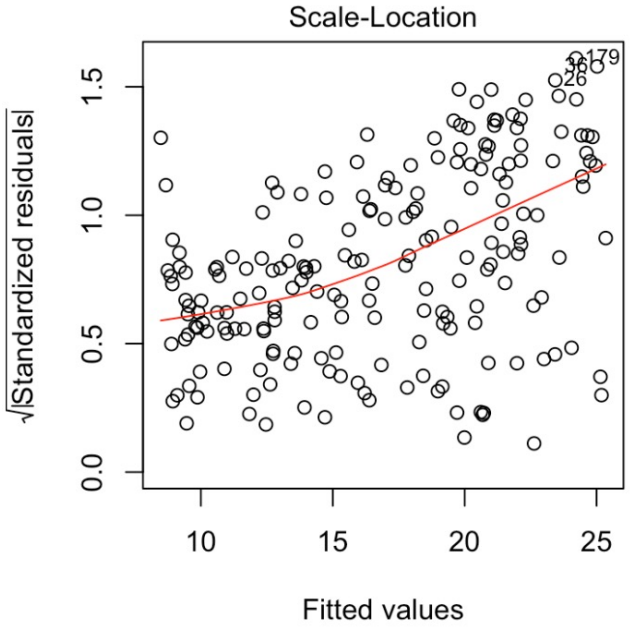
c. The independent variables are highly correlated with each other.

d. The residuals are normally distributed.

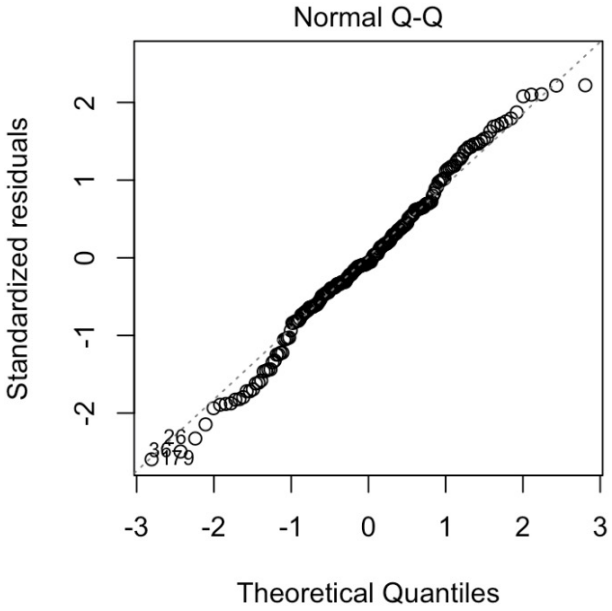
Linearity of the data



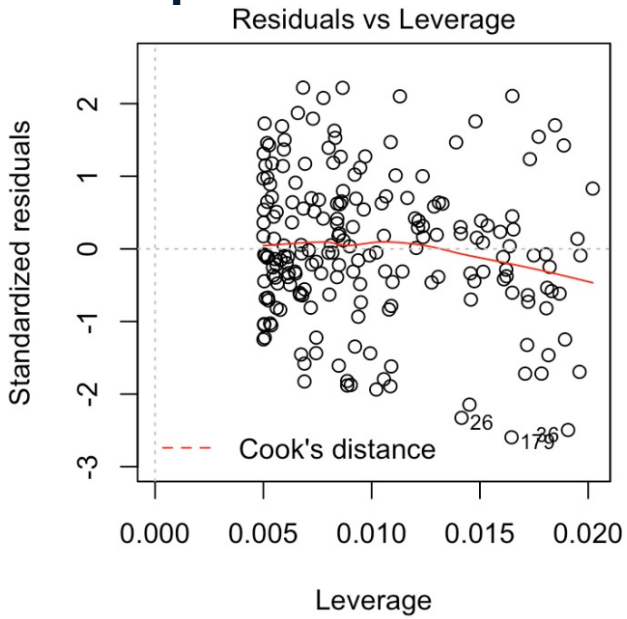
Homogeneity of variance



Normality of residuals



Independence of residuals



Q2. Which of the following is the assumption of linear regression?

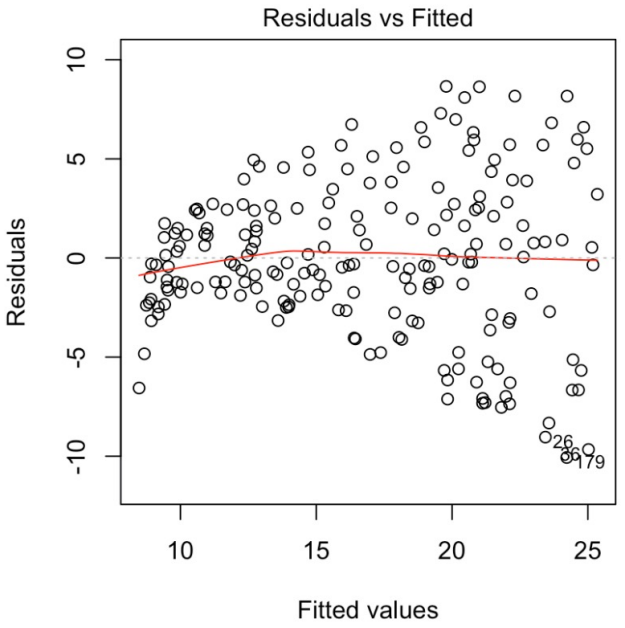
a. The variance of the residuals is constant across all values of the independent variable.

b. The relationship between the dependent variable and independent variable is nonlinear.

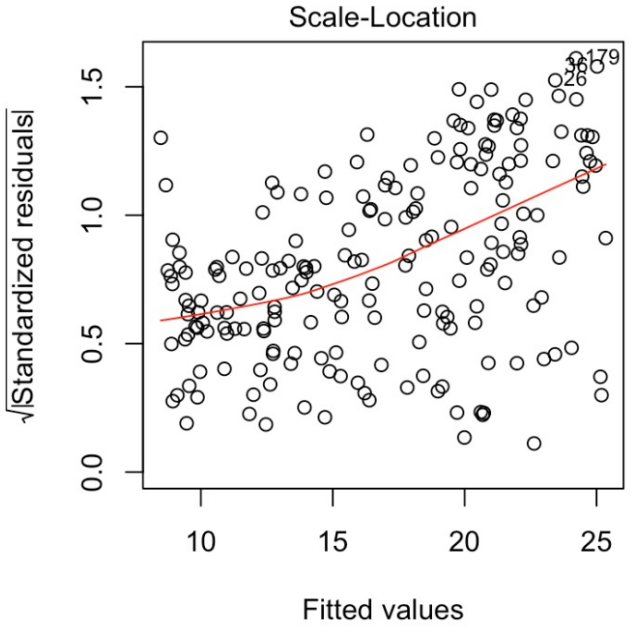
c. The independent variables are highly correlated with each other.

d. The residuals are normally distributed.

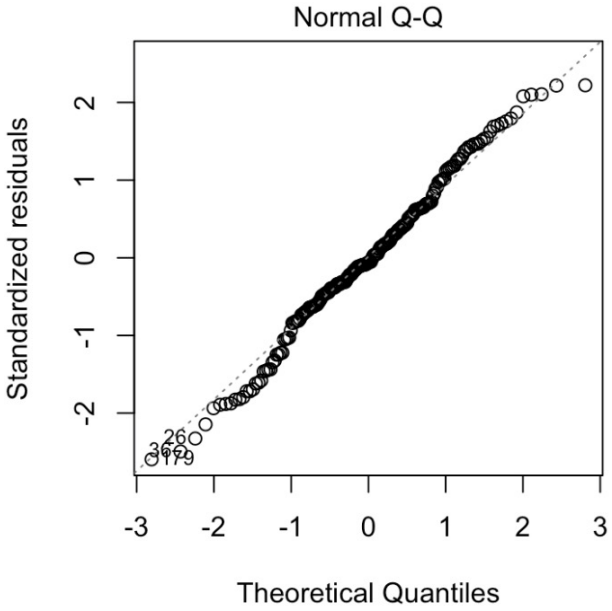
Linearity of the data



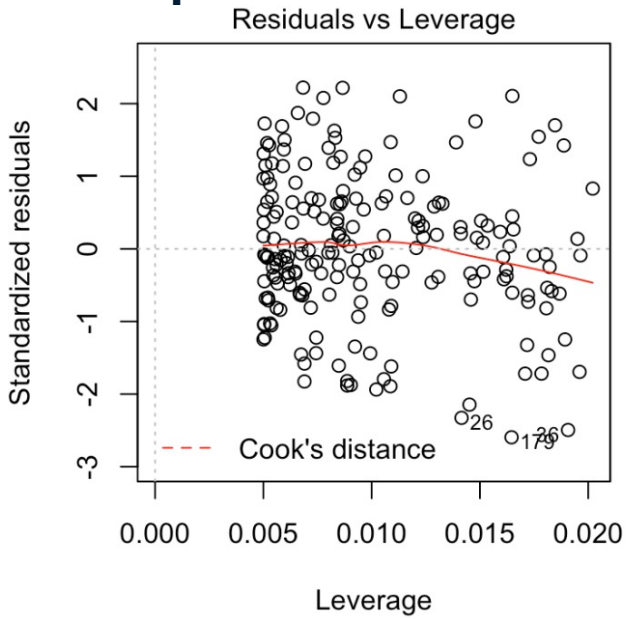
Homogeneity of variance



Normality of residuals



Independence of residuals



Q3. What does the coefficient of determination (R-squared) measure in linear regression?

a. The mean squared error of the model

b. The correlation between the dependent variable and the independent variable

c. The proportion of the variation in the dependent variable that is explained by the independent variable

d. The standard deviation of the residuals

Q3. What does the coefficient of determination (R-squared) measure in linear regression?

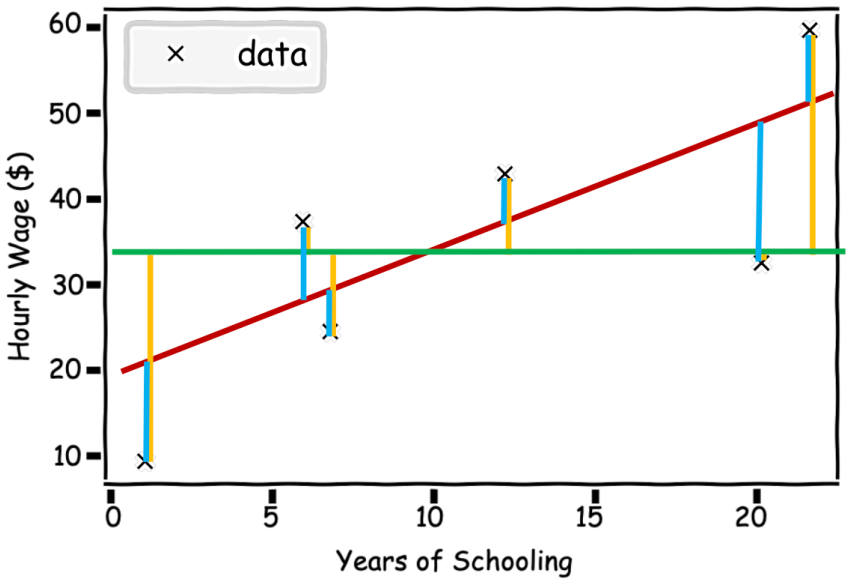
a. The mean squared error of the model

b. The correlation between the dependent variable and the independent variable

c. The proportion of the variation in the dependent variable that is explained by the independent variable

d. The standard deviation of the residuals

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$



$$R^2 = 1 - \frac{RSS}{TSS}$$

$$R^2 = 1 - \frac{(+ + + + + +)^2}{(+ - + + - +)^2}$$

Q3. What does the coefficient of determination (R-squared) measure in linear regression?

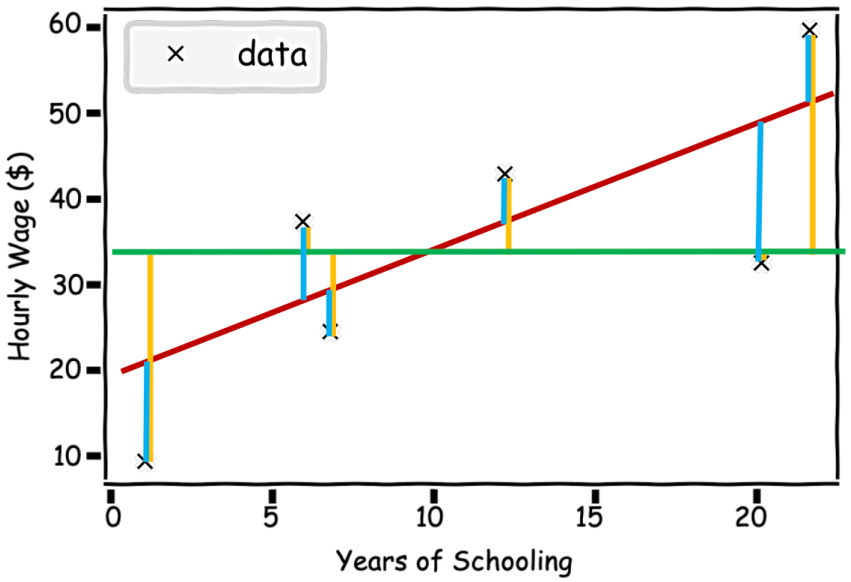
a. The mean squared error of the model

b. The correlation between the dependent variable and the independent variable

c. The proportion of the variation in the dependent variable that is explained by the independent variable

d. The standard deviation of the residuals

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2$$



$$R^2 = 1 - \frac{RSS}{TSS}$$

$$R^2 = 1 - \frac{(+ + + + + +)^2}{(+ - + + - +)^2}$$

Q4. What is the purpose of calculating the p-value of the coefficients in linear regression?

a. To test the hypothesis that the residuals are normally distributed

b. To test the null hypothesis that the coefficient is equal to zero

c. To test the hypothesis that the model fits the data well

d. To test the alternative hypothesis that the coefficient is different from zero

Q4. What is the purpose of calculating the p-value of the coefficients in linear regression?

a. To test the hypothesis that the residuals are normally distributed

b. To test the null hypothesis that the coefficient is equal to zero

c. To test the hypothesis that the model fits the data well

d. To test the alternative hypothesis that the coefficient is different from zero

Null hypothesis (H0):
the coefficients are equal to zero
(i.e., no relationship between x and y)

Alternative Hypothesis (Ha):
the coefficients are not equal to zero
(i.e., there is some relationship between x and y)

Q4. What is the purpose of calculating the p-value of the coefficients in linear regression?

a. To test the hypothesis that the residuals are normally distributed

b. To test the null hypothesis that the coefficient is equal to zero

c. To test the hypothesis that the model fits the data well

d. To test the alternative hypothesis that the coefficient is different from zero

Null hypothesis (H0):
the coefficients are equal to zero
(i.e., no relationship between x and y)

Alternative Hypothesis (Ha):
the coefficients are not equal to zero
(i.e., there is some relationship between x and y)

The p-value, or probability value, is a measure used in statistical hypothesis testing to determine the significance of the results of a hypothesis test.

Q4. What is the purpose of calculating the p-value of the coefficients in linear regression?

a. To test the hypothesis that the residuals are normally distributed

b. To test the null hypothesis that the coefficient is equal to zero

c. To test the hypothesis that the model fits the data well

d. To test the alternative hypothesis that the coefficient is different from zero

For a given beta coefficient (b), the t-statistic is computed as

$$t = (b - 0)/SE(b)$$

where SE(b) is the standard error of the coefficient b.

The t-statistic measures the number of standard deviations that b is away from 0.

A low P-value (< 0.05) means that the coefficient is likely not to equal zero and we reject H0

A high P-value (> 0.05) means that we cannot conclude that the x variable affects the y variable, so we fail to reject H0

Q5. How can we check for homoscedasticity?

a. By plotting the residuals against the independent variable

b. By plotting the predicted values against the actual values of the dependent variable

c. By plotting the independent variable against the dependent variable

d. By plotting the residuals against the dependent variable

Q5. How can we check for homoscedasticity?

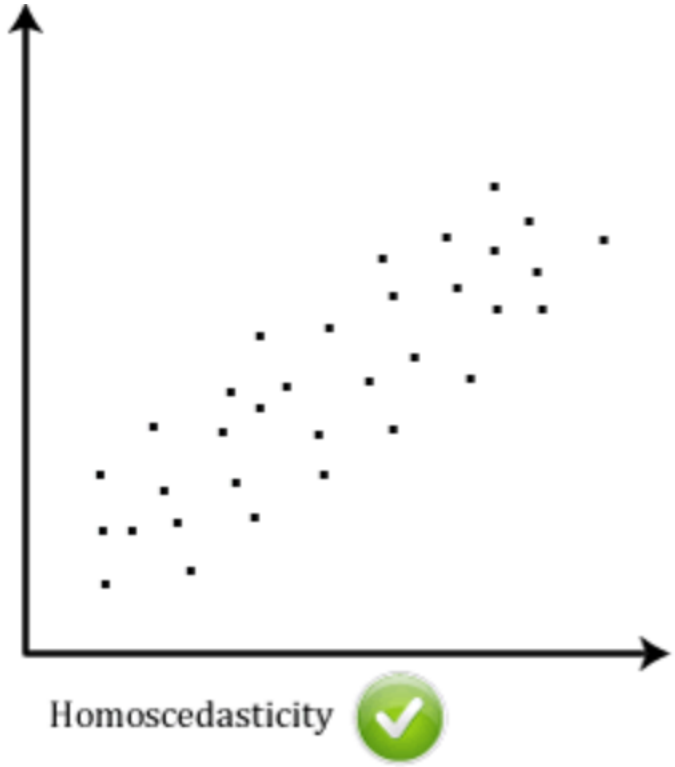
Homoscedasticity means that the variance of the error term, conditional on X, is constant.

a. By plotting the residuals against the independent variable

b. By plotting the predicted values against the actual values of the dependent variable

c. By plotting the independent variable against the dependent variable

d. By plotting the residuals against the dependent variable



$$Var[\varepsilon_i \mid X = x] = \sigma^2$$

Q5. How can we check for homoscedasticity?

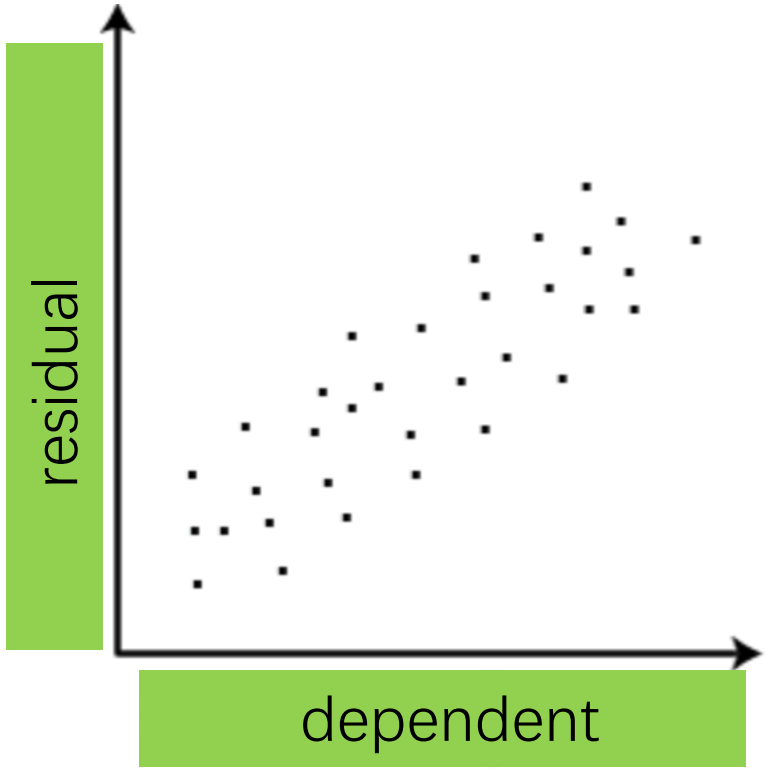
Homoscedasticity means that the variance of the error term, conditional on X, is constant.

a. By plotting the residuals against the independent variable

b. By plotting the predicted values against the actual values of the dependent variable

c. By plotting the independent variable against the dependent variable

d. By plotting the residuals against the dependent variable



$$\mathrm{Var}[\varepsilon_i \mid X = x] = \sigma^2$$