Lecture 4: Simple Linear Regression (3.1.1)

Introduction

- Linear regression is a simple approach for supervised learning.

Introduction

- Linear regression is a simple approach for supervised learning.
- It is useful for predicting a quantitative response.

Introduction

- Linear regression is a simple approach for supervised learning.
- It is useful for predicting a quantitative response.
- Despite being an old method, it is still widely used and serves as a good foundation for more complex learning methods.

Introduction

- Linear regression is a simple approach for supervised learning.
- It is useful for predicting a quantitative response.
- Despite being an old method, it is still widely used and serves as a good foundation for more complex learning methods.
- Understanding linear regression is crucial for studying more complex learning methods.

Introduction

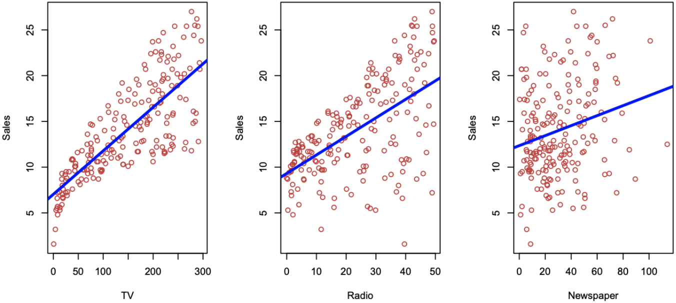
- Linear regression is a simple approach for supervised learning.
- It is useful for predicting a quantitative response.
- Despite being an old method, it is still widely used and serves as a good foundation for more complex learning methods.
- Understanding linear regression is crucial for studying more complex learning methods.
- Many advanced statistical learning approaches can be seen as generalizations or extensions of linear regression.

Introduction

- Linear regression is a simple approach for supervised learning.
- It is useful for predicting a quantitative response.
- Despite being an old method, it is still widely used and serves as a good foundation for more complex learning methods.
- Understanding linear regression is crucial for studying more complex learning methods.
- Many advanced statistical learning approaches can be seen as generalizations or extensions of linear regression.
- **True regression functions are never linear!**

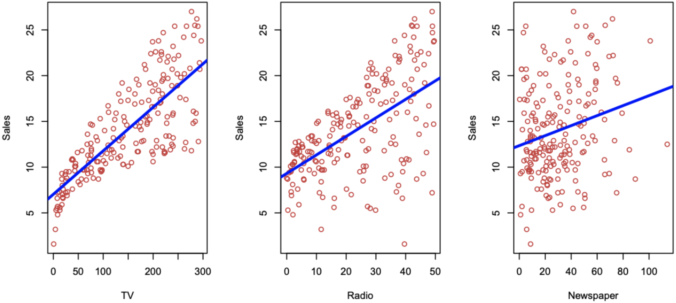
Example: Advertising dataset (Lecture 2)

- The Figure displays sales (in thousands of units) for a particular product as a function of advertising budgets (in thousands of dollars) for TV, radio, and newspaper media.



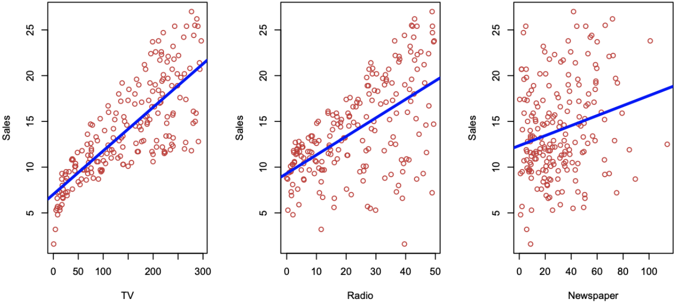
Example: Advertising dataset (Lecture 2)

- The Figure displays sales (in thousands of units) for a particular product as a function of advertising budgets (in thousands of dollars) for TV, radio, and newspaper media.
- Suppose that in our role as statistical consultants we are asked to suggest, on the basis of this data, a marketing plan for next year that will result in high product sales.



Example: Advertising dataset (Lecture 2)

- The Figure displays sales (in thousands of units) for a particular product as a function of advertising budgets (in thousands of dollars) for TV, radio, and newspaper media.
- Suppose that in our role as statistical consultants we are asked to suggest, on the basis of this data, a marketing plan for next year that will result in high product sales.
- What information would be useful in order to provide such a recommendation?.



Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?
  - *Assuming that there is a relationship between advertising and sales, we would like to know the strength of this relationship. Does knowledge of the advertising budget provide a lot of information about product sales?*

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?
  - *Assuming that there is a relationship between advertising and sales, we would like to know the strength of this relationship. Does knowledge of the advertising budget provide a lot of information about product sales?*
- Which media are associated with sales?

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?
  - *Assuming that there is a relationship between advertising and sales, we would like to know the strength of this relationship. Does knowledge of the advertising budget provide a lot of information about product sales?*
- Which media are associated with sales?
  - *Are all three media (TV, radio, and newspaper) associated with sales, or are just one or two of the media associated? To answer this question, we must find a way to separate out the individual contribution of each medium to sales when we have spent money on all three media.*

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?
  - *Assuming that there is a relationship between advertising and sales, we would like to know the strength of this relationship. Does knowledge of the advertising budget provide a lot of information about product sales?*
- Which media are associated with sales?
  - *Are all three media (TV, radio, and newspaper) associated with sales, or are just one or two of the media associated? To answer this question, we must find a way to separate out the individual contribution of each medium to sales when we have spent money on all three media.*
- How large is the association between each medium and sales?

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?
  - *Assuming that there is a relationship between advertising and sales, we would like to know the strength of this relationship. Does knowledge of the advertising budget provide a lot of information about product sales?*
- Which media are associated with sales?
  - *Are all three media (TV, radio, and newspaper) associated with sales, or are just one or two of the media associated? To answer this question, we must find a way to separate out the individual contribution of each medium to sales when we have spent money on all three media.*
- How large is the association between each medium and sales?
  - *For every dollar spent on advertising in a particular medium, by what amount will sales increase? How accurately can we predict this amount of increase?*

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?
  - *Assuming that there is a relationship between advertising and sales, we would like to know the strength of this relationship. Does knowledge of the advertising budget provide a lot of information about product sales?*
- Which media are associated with sales?
  - *Are all three media (TV, radio, and newspaper) associated with sales, or are just one or two of the media associated? To answer this question, we must find a way to separate out the individual contribution of each medium to sales when we have spent money on all three media.*
- How large is the association between each medium and sales?
  - *For every dollar spent on advertising in a particular medium, by what amount will sales increase? How accurately can we predict this amount of increase?*
- How accurately can we predict future sales?

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
  - *Our first goal should be to determine whether the data provide evidence of an association between advertising expenditure and sales. If the evidence is weak, then one might argue that no money should be spent on advertising!*
- How strong is the relationship between advertising budget and sales?
  - *Assuming that there is a relationship between advertising and sales, we would like to know the strength of this relationship. Does knowledge of the advertising budget provide a lot of information about product sales?*
- Which media are associated with sales?
  - *Are all three media (TV, radio, and newspaper) associated with sales, or are just one or two of the media associated? To answer this question, we must find a way to separate out the individual contribution of each medium to sales when we have spent money on all three media.*
- How large is the association between each medium and sales?
  - *For every dollar spent on advertising in a particular medium, by what amount will sales increase? How accurately can we predict this amount of increase?*
- How accurately can we predict future sales?
  - *For any given level of television, radio, or newspaper advertising, what is our prediction for sales, and what is the accuracy of this prediction?*

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
- How strong is the relationship between advertising budget and sales?
- Which media are associated with sales?
- How large is the association between each medium and sales?
- How accurately can we predict future sales?
- Is the relationship linear?

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
- How strong is the relationship between advertising budget and sales?
- Which media are associated with sales?
- How large is the association between each medium and sales?
- How accurately can we predict future sales?
- Is the relationship linear?
    - *If there is approximately a straight-line relationship between advertising expenditure in the various media and sales, then linear regression is an appropriate tool. If not, then it may still be possible to transform the predictor or the response so that linear regression can be used.*

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
- How strong is the relationship between advertising budget and sales?
- Which media are associated with sales?
- How large is the association between each medium and sales?
- How accurately can we predict future sales?
- Is the relationship linear?
  - *If there is approximately a straight-line relationship between advertising expenditure in the various media and sales, then linear regression is an appropriate tool. If not, then it may still be possible to transform the predictor or the response so that linear regression can be used.*
- Is there synergy among the advertising media?

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
- How strong is the relationship between advertising budget and sales?
- Which media are associated with sales?
- How large is the association between each medium and sales?
- How accurately can we predict future sales?
- Is the relationship linear?
  - *If there is approximately a straight-line relationship between advertising expenditure in the various media and sales, then linear regression is an appropriate tool. If not, then it may still be possible to transform the predictor or the response so that linear regression can be used.*
- Is there synergy among the advertising media?
  - *Perhaps spending $50,000 on television advertising and $50,000 on radio advertising is associated with higher sales than allocating $100,000 to either television or radio individually. In marketing, this is known as a synergy effect, while in statistics it is called an interaction effect.*

Example: Advertising dataset (Lecture 2)

- Is there a relationship between advertising budget and sales?
- How strong is the relationship between advertising budget and sales?
- Which media are associated with sales?
- How large is the association between each medium and sales?
- How accurately can we predict future sales?
- Is the relationship linear?
- Is there synergy among the advertising media?

Simple Linear Regression: Model

- We assume a model
$$Y = \beta_0 + \beta_1 X + \epsilon,$$
β_0 and β_1 are **unknown** constants that represent the intercept and slope terms in the linear model.

Simple Linear Regression: Model

- We assume a model
$$Y = \beta_0 + \beta_1 X + \epsilon,$$
β_0 and β_1 are **unknown** constants that represent the intercept and slope terms in the linear model.
- β_0 and β_1 are the coefficients or parameters of the model, and ϵ is the error term.

Simple Linear Regression: Model

- We assume a model
$$Y = \beta_0 + \beta_1 X + \epsilon,$$
β_0 and β_1 are **unknown** constants that represent the intercept and slope terms in the linear model.
- β_0 and β_1 are the coefficients or parameters of the model, and ϵ is the error term.
- Once we have used our training data to produce estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, we can predict future values of Y based on a particular value of X. The prediction is computed as:
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Simple Linear Regression: Model

- We assume a model

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

  β_0 and β_1 are **unknown** constants that represent the intercept and slope terms in the linear model.
- β_0 and β_1 are the coefficients or parameters of the model, and ϵ is the error term.
- Once we have used our training data to produce estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, we can predict future values of Y based on a particular value of X. The prediction is computed as:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

- The hat symbol denotes the estimated value for an unknown parameter or coefficient, or the predicted value of the response.

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.
- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for Y based on the i-th value of X.

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.
- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for Y based on the i-th value of X.
  - Then $e_i = \hat{y}_i - y_i$ represents the i-th residual.

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.
- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for Y based on the i-th value of X.
  - Then $e_i = \hat{y}_i - y_i$ represents the i-th residual.
- We define the residual sum of squares (RSS):

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.
- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for Y based on the i-th value of X.
  - Then $e_i = \hat{y}_i - y_i$ represents the i-th residual.
- We define the residual sum of squares (RSS):
  - $$RSS = e_1^2 + \ldots + e_n^2$$

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.
- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for Y based on the i-th value of X.
  - Then $e_i = \hat{y}_i - y_i$ represents the i-th residual.
- We define the residual sum of squares (RSS):
  - $$RSS = e_1^2 + \ldots + e_n^2$$
- The least squares approach chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the RSS:

Estimating coefficients

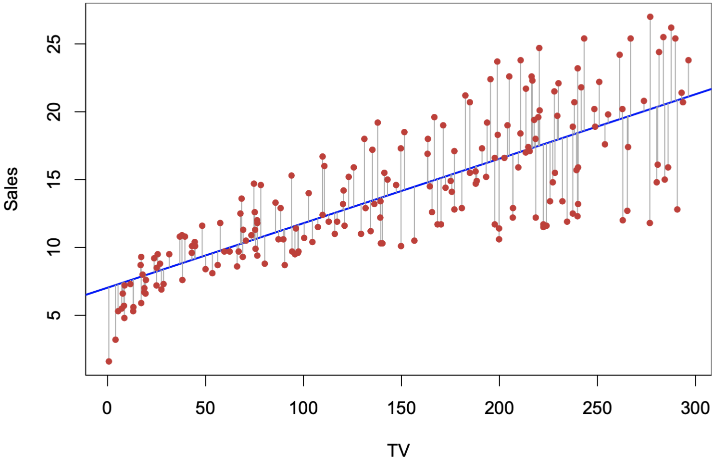
- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.
- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for Y based on the i-th value of X.
  - Then $e_i = \hat{y}_i - y_i$ represents the i-th residual.
- We define the residual sum of squares (RSS):
  - $$RSS = e_1^2 + \ldots + e_n^2$$
- The least squares approach chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the RSS:
  - $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

Estimating coefficients

- We have n observation $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Our goal is to estimate β_0 and β_1 such that the linear model fits the data well. The most common approach is to minimize the least squares criterion. Alternative approaches will be considered later.
- Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the prediction for Y based on the i-th value of X.
  - Then $e_i = \hat{y}_i - y_i$ represents the i-th residual.
- We define the residual sum of squares (RSS):
  - $$RSS = e_1^2 + \ldots + e_n^2$$
- The least squares approach chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the RSS:
  - $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$
  - $\bar{y}$ and $\bar{x}$ are the sample means.

Estimating coefficients

For the Advertising data, the fit is found by minimizing the residual sum of squares. Each grey line segment represents a residual. In this case a linear fit captures the essence of the relationship, although it overestimates the trend in the left of the plot.



Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:
  - $Y = \beta_0 + \beta_1 X + \epsilon$.

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
    - f is an unknown function and ϵ is a mean-zero random error term.
    - If f is approximated by a linear function, we can write this relationship:
    - $Y = \beta_0 + \beta_1 X + \epsilon$.
    - β_0 is the intercept term: the expected value of Y when $X = 0$.

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:
  - $Y = \beta_0 + \beta_1 X + \epsilon$.
  - β_0 is the intercept term: the expected value of Y when $X = 0$.
  - β_1 is the slope: the average increase in Y ass. with a unit increase in X.

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:
  - $Y = \beta_0 + \beta_1 X + \epsilon$.
  - β_0 is the intercept term: the expected value of Y when $X = 0$.
  - β_1 is the slope: the average increase in Y ass. with a unit increase in X.
  - The error term is a catch-all for what we miss with this simple model: the true relationship is probably not linear, there may be other variables that cause variation in Y , and there may be measurement error.

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:
  - $Y = \beta_0 + \beta_1 X + \epsilon$.
  - β_0 is the intercept term: the expected value of Y when $X = 0$.
  - β_1 is the slope: the average increase in Y ass. with a unit increase in X.
  - The error term is a catch-all for what we miss with this simple model: the true relationship is probably not linear, there may be other variables that cause variation in Y , and there may be measurement error.
  - The true relationship is probably not linear,

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:
  - $Y = \beta_0 + \beta_1 X + \epsilon$.
  - β_0 is the intercept term: the expected value of Y when $X = 0$.
  - β_1 is the slope: the average increase in Y ass. with a unit increase in X.
  - The error term is a catch-all for what we miss with this simple model: the true relationship is probably not linear, there may be other variables that cause variation in Y , and there may be measurement error.
  - The true relationship is probably not linear,
  - There may be other variables that cause variation in Y,

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:
  - $Y = \beta_0 + \beta_1 X + \epsilon$.
  - β_0 is the intercept term: the expected value of Y when $X = 0$.
  - β_1 is the slope: the average increase in Y ass. with a unit increase in X.
  - The error term is a catch-all for what we miss with this simple model: the true relationship is probably not linear, there may be other variables that cause variation in Y , and there may be measurement error.
  - The true relationship is probably not linear,
  - There may be other variables that cause variation in Y,
  - There may be measurement errors.

Summary

- we assume that the true relationship between X and Y takes the form $Y = f(X) + \epsilon$.
  - f is an unknown function and ϵ is a mean-zero random error term.
  - If f is approximated by a linear function, we can write this relationship:
  - $Y = \beta_0 + \beta_1 X + \epsilon$.
  - β_0 is the intercept term: the expected value of Y when $X = 0$.
  - β_1 is the slope: the average increase in Y ass. with a unit increase in X.
  - The error term is a catch-all for what we miss with this simple model: the true relationship is probably not linear, there may be other variables that cause variation in Y , and there may be measurement error.
  - The true relationship is probably not linear,
  - There may be other variables that cause variation in Y,
  - There may be measurement errors.
  - We typically assume that the error term is independent of X.

Question

If the true relationship between X and Y is actually linear, can we have $\hat{\beta}_0 = \beta_0$ and $\hat{\beta}_1 = \beta_1$?