# Explore the differences between Simple, Multiple and Multivariate Linear Regression

# Difference between Simple and Multiple Linear Regression

**Simple Linear Regression** and **Multiple Linear Regression** are both *statistical methods* used to model the relationship between **dependent** and **independent** variables. The main difference lies in the *number of independent variables* involved.

## Simple Linear Regression

### Definition:

**Simple linear regression** models the relationship between **a single independent variable (predictor)** and **a dependent variable (response).**

### Equation: 𝛽 (beta) = intercept; ε (Greek epsilon) = residual part = error term

#### 𝑌 = 𝛽0 + 𝛽1𝑋 + ϵ

* **𝑌** is the dependent variable.
* **𝑋** is the independent variable.
* **𝛽0** is the **y**-intercept.
* **𝛽1** is the slope of the line.
* **𝜖** is the residual part or error term.

### Example:

Let's say you want to predict a person's **weight** based on their **height.** Here,

* **height** is the independent variable, and
* **weight** is the dependent variable.

### Data:

* **Height (in inches):** 60, 62, 64, 66, 68
* **Weight (in pounds):** 115, 120, 125, 130, 135

Using **simple linear regression**, you could find the equation that **best fits this data**.

#### Weight = 10 + 2 × Height

## Multiple Linear Regression

### Definition:

**Multiple linear regression** models the relationship **between two or more independent variables** and **a dependent variable.**

### Equation:

#### 𝑌 = 𝛽0 + 𝛽1𝑋1 + 𝛽2𝑋2 + . . . + 𝛽𝑘𝑋𝑘 + 𝜖

* **𝑌 is the dependent variable.**
* **𝑋1, 𝑋2, . . ., 𝑋𝑘 are the independent variables.**
* **𝛽0 is the y-intercept.**
* **𝛽1, 𝛽2, . . ., 𝛽𝑘 are the coefficients of the independent variables.**
* **𝜖 is the residual part or error term.**

### Example:

Let's say you want to predict a person's **weight** based on their **height** and **age.** Here,

* **height and age** are the **independent variables,** and
* **weight** is the **dependent variable.**

### Data:

* **Height (in inches):** 60, 62, 64, 66, 68
* **Age (in years):** 25, 30, 35, 40, 45
* **Weight (in pounds):** 115, 120, 125, 130, 135

Using **multiple linear regression**, you could find the equation that **best fits this data** considering both **height** and **age** as **predictors.**

## Practical Example

### Simple Linear Regression Example:

Imagine you are a data analyst at a fitness company. You have collected data on the **heights and weights** of a sample group of people and want to establish if there's a **linear relationship between height and weight.**

#### Data:

* **Height (X): 150, 160, 170, 180, 190 cm**
* **Weight (Y): 50, 60, 65, 70, 80 kg**

Performing **simple linear regression**, you might find an equation like:

#### Y = 𝛽0 + 𝛽1 \* X

#### Weight = 15 + 0.3 × Height

This equation suggests that for every additional **centimetre in height, the weight increases by 0.3 kg,** starting from an **intercept of 15 kg.**

### Multiple Linear Regression Example:

Now, suppose you also collected data on the **age** of the individuals and want to see how **both height and age predict weight.**

#### Data:

* **Height (X1): 150, 160, 170, 180, 190 cm**
* **Age (X2): 20, 25, 30, 35, 40 years**
* **Weight (Y): 50, 60, 65, 70, 80 kg**

Performing **multiple linear regression**, you might find an equation like:

#### Y = 𝛽0 + 𝛽1 \* X1 + 𝛽2 \* X2

#### Weight = 10 + 0.25 × Height + 0.5 × Age

This equation indicates that **weight increases by 0.25 kg** for every additional **centimetre in height and by**

**0.5 kg** for every additional **year of age,** starting from an **intercept of 10 kg.**

## Key Differences

### Number of Independent Variables:

**Simple Linear Regression:** One independent variable.  
**Multiple Linear Regression:** Two or more independent variables.

### Complexity:

**Simple Linear Regression:** Less complex, easier to interpret.  
**Multiple Linear Regression:** More complex, can capture relationships involving multiple factors.

### Application:

**Simple Linear Regression:** Best used when there's a single factor influencing the outcome.  
**Multiple Linear Regression:** Used when multiple factors are believed to influence the outcome.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Number of Independent Variables** | **Complexity** | **Application** |
| **Simple Linear Regression** | One independent variable. | Less complex, easier to interpret. | Best used when there's a single factor influencing the outcome. |
| **Multiple Linear Regression** | Two or more independent variables | More complex, can capture relationships involving multiple factors | Used when multiple factors are believed to influence the outcome. |

# Difference between Simple and Multivariate Linear Regression with example

**Simple Linear Regression** and **Multivariate Linear Regression** (often referred to as multiple linear regression when discussing multiple **predictors**) are both statistical methods used to model relationships between variables, but they differ in the **number** and **type** of variables involved.

## Simple Linear Regression:

### Definition:

**Simple linear regression** models the relationship between **a single independent variable (predictor)** and **a single dependent variable (response).**

### Equation:

#### 𝑌 = 𝛽0 + 𝛽1𝑋 + 𝜖

* **𝑌** is the dependent variable.
* **𝑋** is the independent variable.
* **𝛽0** is the **y**-intercept.
* **𝛽1** is the slope of the line.
* **𝜖** is the residual part or error term.

### Example:

Predicting a person's **weight** based on their **height.**

### Data:

* **Height (in inches):** 60, 62, 64, 66, 68
* **Weight (in pounds):** 115, 120, 125, 130, 135

Using **simple linear regression**, you could find the equation that **best fits this data**.

#### Weight = 10 + 2 × Height

## Multivariate Linear Regression

### Definition:

**Multivariate linear regression** models the relationship between **multiple dependent variables** and **one or more independent variables.** This distinguishes it from **multiple linear regression,** which typically involves ***multiple independent variables predicting a single dependent variable.***

### Equation:

**𝑌1 = 𝛽01 + 𝛽11𝑋1 + 𝛽21𝑋2 + . . . + 𝛽𝑘1𝑋𝑘 + 𝜖1**

**𝑌2 = 𝛽02 + 𝛽12𝑋1 + 𝛽22𝑋2 + . . . + 𝛽𝑘2𝑋𝑘 + 𝜖2**

**.  
.  
.**

**𝑌𝑚 = 𝛽0𝑚 + 𝛽1𝑚𝑋1 + 𝛽2𝑚𝑋2 + . . . + 𝛽𝑘𝑚𝑋𝑘 + 𝜖𝑚**

* **𝑌1, 𝑌2, . . ., 𝑌𝑚** are the dependent variables.
* **𝑋1, 𝑋2, . . ., 𝑋𝑘** are the independent variables.
* **𝛽01, 𝛽02, . . ., 𝛽0𝑚** are the **y**-intercepts for *each dependent variable.*
* **𝛽11, 𝛽21, . . ., 𝛽𝑘1** are the coefficients for the independent variables in the first equation, and so on.
* **𝜖1, 𝜖2, . . . 𝜖𝑚** are the residual parts or the error terms.

### Example:

Predicting a person's **weight** and **body fat percentage** based on their **height and age.**

### Data:

* **Height (in inches):** 60, 62, 64, 66, 68
* **Age (in years):** 25, 30, 35, 40, 45
* **Weight (in pounds):** 115, 120, 125, 130, 135
* **Body fat percentage:** 20, 18, 16, 14, 12

Using **multivariate linear regression,** the equations might be:

#### Weight = 10 + 0.25 × Height + 0.5 × Age Body Fat Percentage = 30 − 0.1 × Height − 0.2 × Age

Predicting a person's **weight** and **body fat percentage** based on their **height and age.**

These equations indicate that **weight** and **body fat percentage** are both influenced by **height and age.**

## Key Differences

### Number of Dependent Variables:

**Simple Linear Regression:** One dependent variable.  
**Multivariate Linear Regression:** Multiple dependent variables.

### Application:

**Simple Linear Regression:** Used when analyzing the effect of a **single predictor on a single outcome.**  
**Multivariate Linear Regression:** Used when analyzing the effect of **multiple predictors on multiple outcomes.**

### Complexity:

**Simple Linear Regression:** Simpler model, easier to interpret and visualize.  
**Multivariate Linear Regression:** More complex, can capture relationships involving **multiple outcomes** and **predictors**.

|  |  |  |
| --- | --- | --- |
|  | **Simple Linear Regression** | **Multivariate Linear Regression** |
| **Number of Dependent Variables** | One dependent variable. | Multiple dependent variables. |
| **Application** | Used when analyzing the effect of a **single predictor on a single outcome.** | Used when analyzing the effect of **multiple predictors on multiple outcomes.** |
| **Complexity** | Simpler model, easier to interpret and visualize. | More complex, can capture relationships involving **multiple outcomes** and **predictors**. |

## Practical Example

### Simple Linear Regression Example:

**Predicting the price of a house based on its size.**

#### Data:

* **Size (in square feet):** 1500, 1600, 1700, 1800, 1900
* **Price (in $1000):** 300, 320, 340, 360, 380

Performing **simple linear regression,** you might find an equation like:

#### Price = 100 + 0.1 × Size

### Multivariate Linear Regression Example:

Predicting both the **price** and the **rental value** of a house based on its **size and number of bedrooms.**

#### Data:

* **Size (in square feet):** 1500, 1600, 1700, 1800, 1900
* **Bedrooms:** 3, 3, 4, 4, 5
* **Price (in $1000):** 300, 320, 340, 360, 380
* Rental Value (in $1000): 2, 2.2, 2.4, 2.6, 2.8

Performing **multivariate linear regression,** you might find equations like:

#### Price = 50 + 0.2 × Size + 10 × Bedrooms

#### Rental Value = 1 + 0.001 × Size + 0.5 × Bedrooms

These equations show how both **size** and **number of bedrooms** influence both the **price** and the **rental value** of the house.

# Difference Between Multiple and Multivariate Linear Regression

**Multiple Linear Regression** and **Multivariate Linear Regression** are statistical methods used to **model the relationships between variables,** *but they differ in the****number and type of dependent variables****involved.*

## Multiple Linear Regression

### Definition:

**Multiple linear regression** models the relationship between **one dependent variable** and **two or more independent variables.**

### Equation:

#### 𝑌 = 𝛽0 + 𝛽1𝑋1 + 𝛽2𝑋2 + . . . + 𝛽𝑘𝑋𝑘 + 𝜖

* **𝑌 is the dependent variable.**
* **𝑋1, 𝑋2, . . ., 𝑋𝑘 are the independent variables.**
* **𝛽0 is the y-intercept.**
* **𝛽1, 𝛽2, . . ., 𝛽𝑘 are the coefficients of the independent variables.**
* **𝜖 is the residual part or error term.**

### Example:

Predicting a person's **weight** based on their **height** and **age**.

### Data:

* **Height (in inches):** 60, 62, 64, 66, 68
* **Age (in years):** 25, 30, 35, 40, 45
* **Weight (in pounds):** 115, 120, 125, 130, 135

Using **multiple linear regression**, you could find the equation that **best fits this data** considering both **height** and **age** as **predictors.**

#### Y = 𝛽0 + 𝛽1 \* X1 + 𝛽2 \* X2

#### Weight = 10 + 0.25 × Height + 0.5 × Age

This equation indicates that **weight increases by 0.25 pounds** for every additional **inch of height and by**

**0.5 pounds** for every additional **year of age,** starting from an **intercept of 10 pounds.**

## Multivariate Linear Regression

### Definition:

**Multivariate linear regression** models the relationship between **multiple dependent variables** and **one or more independent variables.** This distinguishes it from **multiple linear regression,** which typically involves ***multiple independent variables predicting a single dependent variable.***

### Equation:

**𝑌1 = 𝛽01 + 𝛽11𝑋1 + 𝛽21𝑋2 + . . . + 𝛽𝑘1𝑋𝑘 + 𝜖1**

**𝑌2 = 𝛽02 + 𝛽12𝑋1 + 𝛽22𝑋2 + . . . + 𝛽𝑘2𝑋𝑘 + 𝜖2**

**.  
.  
.**

**𝑌𝑚 = 𝛽0𝑚 + 𝛽1𝑚𝑋1 + 𝛽2𝑚𝑋2 + . . . + 𝛽𝑘𝑚𝑋𝑘 + 𝜖𝑚**

Here;

* **𝑌1, 𝑌2, . . ., 𝑌𝑚** are the dependent variables.
* **𝑋1, 𝑋2, . . ., 𝑋𝑘** are the independent variables.
* **𝛽01, 𝛽02, . . ., 𝛽0𝑚** are the y-intercepts for each dependent variable.
* **𝛽11, 𝛽21, . . .,𝛽𝑘1** are the coefficients for the independent variables in the first equation, and so on.
* **𝜖1, 𝜖2, . . . 𝜖𝑚** are the error terms.

### Example:

Predicting a person's **weight** and **body fat percentage** based on their **height and age.**

### Data:

* **Height (in inches):** 60, 62, 64, 66, 68
* **Age (in years):** 25, 30, 35, 40, 45
* **Weight (in pounds):** 115, 120, 125, 130, 135
* **Body fat percentage:** 20, 18, 16, 14, 12

Using **multivariate linear regression,** the equations might be:

#### Weight = 10 + 0.25 × Height + 0.5 × Age Body Fat Percentage = 30 − 0.1 × Height − 0.2 × Age

Predicting a person's **weight** and **body fat percentage** based on their **height and age.**

These equations indicate that **weight** and **body fat percentage** are both influenced by **height and age.**

## Key Differences

### Number of Dependent Variables:

**Multiple Linear Regression:** One dependent variable.  
**Multivariate Linear Regression:** **Multiple dependent** variables.

### Application:

**Multiple Linear Regression:** Used when analyzing the effect of **multiple predictors** on a **single outcome**.  
**Multivariate Linear Regression:** Used when analyzing the effect of **multiple predictors** on **multiple outcomes.**

### Complexity:

**Multiple Linear Regression:** Simpler model with **one outcome variable.**  
**Multivariate Linear Regression:** More complex model with **multiple outcome variables.**

|  |  |  |
| --- | --- | --- |
|  | **Multiple Linear Regression** | **Multivariate Linear Regression** |
| **Number of Dependent Variables** | One dependent variable. | Multiple dependent variables. |
| **Application** | Used when analyzing the effect of **multiple predictor on a single outcome.** | Used when analyzing the effect of **multiple predictors on multiple outcomes.** |
| **Complexity** | Simpler model, with **one outcome variable.** | More complex with, **multiple outcomes variables.** |

## Practical Example

### Multiple Linear Regression Example:

Predicting the **price** of a house based on its **size** and **number of bedrooms.**

#### Data:

* **Size (in square feet):** 1500, 1600, 1700, 1800, 1900
* **Bedrooms:** 3, 3, 4, 4, 5
* **Price (in $1000):** 300, 320, 340, 360, 380

Performing **multiple linear regression,** you might find an equation like:

#### Price = 50 + 0.2 × Size + 10 × Bedrooms

This equation shows how both **size** and **number of bedrooms** influence the **price.**

### Multivariate Linear Regression Example:

Predicting both the **price** and the **rental value** of a house based on its **size** and **number of bedrooms.**

#### Data:

* **Size (in square feet):** 1500, 1600, 1700, 1800, 1900
* **Bedrooms:** 3, 3, 4, 4, 5
* **Price (in $1000):** 300, 320, 340, 360, 380
* **Rental Value (in $1000):** 2, 2.2, 2.4, 2.6, 2.8

Performing **multivariate linear regression,** you might find equations like:

#### Price = 50 + 0.2 × Size + 10 × Bedrooms

#### Rental Value = 1 + 0.001 × Size + 0.5 × Bedrooms

These equations indicate how both **size** and **number of bedrooms** influence both the **price** and the **rental value** of the house.