Prediction With Linear Regression

POLS 602

Dr. Mike Burnham
Texas A&M Political Science

Review

1. You have a bunch of data. The only thing you know about the data is that the mean of your sample is 50. A, observation from the sample is chosen at random and you have to guess its value. What is your best guess?

2. You have a bunch of data. The only thing you know about the data is that the mean of your sample is 50 and the median is 60. A, observation from the sample is chosen at random and you have to guess its value. What is your best guess?

3. In plain english, what is the standard deviation?

4. In plain english, what is Z in this formula? $Z = \frac{x - \mu}{\sigma}$

5. Why would we want to use units of standard deviations to communicate how far something is from the mean?

Review

$$Z = \frac{x - \mu}{\sigma}$$

Z = \frac{x - \mu}{\sigma}

$$Z_i^X = \frac{(X_i - \bar{X})}{sd(X)}$$

Z_{i}^{X} = \frac{(X_i - \bar{X})}{sd(X)}

Predictive Modeling

Why Predictive Modeling

- Uncover new causal mechanisms and generate new hypotheses
- Discover new measures and compare different operationalizations
- Improvement to existing explanatory models
- Assessing the distance between theory and practice
- Compare competing theories
- Quantifying the level of predictability of measurable phenomena by creating benchmarks

Source:

To Explain or to Predict? Galit Shmueli, 2010

What Can We Learn from Predictive Modeling? Cranmer and Desmarais, 2017

Notation and Vocabulary

- **X**: Predictors, features
- **Y**: Outcomes, observed outcomes
- **Model**: a mathematical expression that represents a relationship between data
- **Data Generating Process:** The mechanisms or processes in the real world that produce the data we observe

System Map

wmata.com
Customer Information Service: 202 637-7000
TTY Phone: 202 638-3780
Metro Transit Police: 202 962-2121

Legend

- Red Line • Glenmont / Shady Grove
- Orange Line • New Carrollton / Vienna
- Blue Line • Franconia-Springfield / Largo Town Center
- Green Line • Branch Ave / Greenbelt
- Yellow Line • Huntington / Fort Totten
- Silver Line • Wiehle-Reston East / Largo Town Center

Station Features
- Bus to Airport
- Parking
- Hospital
- Airport

Connecting Rail Systems: MARC, VRE

Transfer Station
Under Construction
Full-Time Service
Rush-Only Service: Monday-Friday 6:30am - 9:00am 3:30pm - 6:00pm
Station in Service

Shady Grove
Rockville
Twinbrook
White Flint
Grosvenor-Strathmore
Medical Center
Bethesda
Friendship Heights
Tenleytown-AU
Van Ness-UDC
Cleveland Park
Woodley Park Zoo/Adams Morgan
Dupont Circle
Farragut North
Metro Center
McPherson Sq
Farragut West
Foggy Bottom-GWU
Rosslyn
Arlington Cemetery
Pentagon
Pentagon City
Crystal City
Ronald Reagan Washington National Airport
Braddock Rd
King St-Old Town
Eisenhower Ave
Huntington
Van Dorn St
Franconia-Springfield
Glenmont
Wheaton
Forest Glen
Silver Spring
Takoma
Brookland-CUA
Rhode Island Ave
NoMa-Gallaudet U
Union Station
Capitol South
Eastern Market
Potomac Ave
Stadium-Armory
Benning Rd
Capitol Heights
Addison Rd
Morgan Blvd
Largo Town Center
Greenbelt
College Park-U of Md
West Hyattsville
Prince George's Plaza
Fort Totten
Georgia Ave-Petworth
Columbia Heights
U Street
Shaw-Howard U
Mt Vernon Sq
Gallery Place
Archives
Judiciary Square
L'Enfant Plaza
Waterfront
Navy Yard-Ballpark
Anacostia
Congress Heights
Southern Ave
Naylor Rd
Suitland
Branch Ave
New Carrollton
Landover
Cheverly
Deanwood
Minnesota Ave
Van Ness-UDC
Tenleytown-AU
Friendship Heights
Bethesda
Medical Center
Grosvenor-Strathmore
White Flint
Twinbrook
Rockville
Shady Grove
Route 772
Route 606
Dulles International Airport
Innovation Center
Herndon
Reston Town Center
Wiehle-Reston East
Spring Hill
Tysons Corner
Greensboro
McLean
West Falls Church
Virginia Sq-GMU
Clarendon
Court House
Ballston-MU
Arlington Cemetery
Pentagon
Pentagon City
Crystal City
Ronald Reagan Washington National Airport
Braddock Rd
King St-Old Town
Eisenhower Ave
Huntington
Van Dorn St
Franconia-Springfield
Glenmont
Wheaton
Forest Glen
Silver Spring
Takoma
Brookland-CUA
Rhode Island Ave
NoMa-Gallaudet U
Union Station
Capitol South
Eastern Market
Potomac Ave
Stadium-Armory
Benning Rd
Capitol Heights
Addison Rd
Morgan Blvd
Largo Town Center
Greenbelt
College Park-U of Md
West Hyattsville
Prince George's Plaza
Fort Totten
Georgia Ave-Petworth
Columbia Heights
U Street
Shaw-Howard U
Mt Vernon Sq
Gallery Place
Archives
Judiciary Square
L'Enfant Plaza
Waterfront
Navy Yard-Ballpark
Anacostia
Congress Heights
Southern Ave
Naylor Rd
Suitland
Branch Ave
New Carrollton
Landover
Cheverly
Deanwood
Minnesota Ave
Farragut North
Farragut West
Foggy Bottom-GWU
Rosslyn
Arlington Cemetery
Pentagon
Pentagon City
Crystal City
Ronald Reagan Washington National Airport
Braddock Rd
King St-Old Town
Eisenhower Ave
Huntington
Van Dorn St
Franconia-Springfield
Glenmont
Wheaton
Forest Glen
Silver Spring
Takoma
Brookland-CUA
Rhode Island Ave
NoMa-Gallaudet U
Union Station
Capitol South
Eastern Market
Potomac Ave
Stadium-Armory
Benning Rd
Capitol Heights
Addison Rd
Morgan Blvd
Largo Town Center
Greenbelt
College Park-U of Md
West Hyattsville
Prince George's Plaza
Fort Totten
Georgia Ave-Petworth
Columbia Heights
U Street
Shaw-Howard U
Mt Vernon Sq
Gallery Place
Archives
Judiciary Square
L'Enfant Plaza
Waterfront
Navy Yard-Ballpark
Anacostia
Congress Heights
Southern Ave
Naylor Rd
Suitland
Branch Ave
New Carrollton
Landover
Cheverly
Deanwood
Minnesota Ave
Farragut North
Farragut West
Foggy Bottom-GWU
Rosslyn
Arlington Cemetery
Pentagon
Pentagon City
Crystal City
Ronald Reagan Washington National Airport
Braddock Rd
King St-Old Town
Eisenhower Ave
Huntington
Van Dorn St
Franconia-Springfield

Metrorail Operating Times
Mon-Thu
5am-midnight
Fri
5am-3am
Sat
7am-3am
Sun
7am-midnight
Times are approximate

Metro is accessible.

Map is set to scale

No Smoking
No Eating or Drinking
No Animals (except service animals)
No Audio (without earphones)
No Littering or Spitting
No Dangerous or Flammable Items

System Map

wmata.com
Information: 202-GO-METRO | TTY: 202-962-2033
Metro Transit Police: 202-962-2121 | Text: MYMTPD (696873)

Terminal stations

- **Red Line** • Glenmont / Shady Grove
- **Orange Line** • New Carrollton / Vienna
- **Blue Line** • Franconia-Springfield / Downtown Largo
- **Green Line** • Branch Av / Greenbelt
- **Yellow Line** • Huntington / Greenbelt
- **Silver Line** • Ashburn / Downtown Largo & New Carrollton

Station Features

- Parking
- Hospital
- Airport

Connecting Rail Systems: Amtrak, VRE, MARC

Transfer Station
Station in Service
Served by every other train

Alternating Y trains to Greenbelt begin Dec. 2025

Metro is accessible.

Map is not to scale

System Map

wmata.com
Information: 202-GO-METRO | TTY: 202-962-2033
Metro Transit Police: 202-962-2121 | Text: MYMTPD (696873)

Terminal stations

- **Red Line** • Glenmont / Shady Grove
- **Orange Line** • New Carrollton / Vienna
- **Blue Line** • Franconia-Springfield / Downtown Largo
- **Green Line** • Branch Av / Greenbelt
- **Yellow Line** • Huntington / Greenbelt
- **Silver Line** • Ashburn / Downtown Largo & New Carrollton

Station Features
- Parking
- Hospital
- Airport

Connecting Rail Systems
ADATRAN, VRE, MARC

Transfer Station
Station in Service
Served by every other train

Alternating Y trains to Greenbelt begin Dec. 2025

R: Glenmont, Wheaton, Forest Glen, Silver Spring, Takoma, Fort Totten, Brookland-CUA, Rhode Island Av, NoMa-Gallaudet U, Union Station, Judiciary Sq, Gallery Place, Metro Center, McPherson Sq, Farragut North, Dupont Circle, Woodley Park, Cleveland Park, Van Ness-UDC, Tenleytown-AU, Friendship Heights, Bethesda, Medical Center, Grosvenor-Strathmore, North Bethesda, Twinbrook, Rockville, Shady Grove

Y: Greenbelt, College Park-U of Md, Hyattsville Crossing, West Hyattsville, Fort Totten, Georgia Av-Petworth, Columbia Heights, U St, Shaw-Howard U, Mt Vernon Sq, Gallery Place, Judiciary Sq, L'Enfant Plaza, Waterfront, Navy Yard-Ballpark, Anacostia, Congress Heights, Southern Av, Naylor Rd, Suitland, Branch Av

G: Fort Totten, Gallery Place, L'Enfant Plaza, Federal Center SW, Capitol South, Eastern Market, Potomac Av, Stadium-Armory, Minnesota Av, Deanwood, Cheverly, Landover, New Carrollton

O: New Carrollton, Landover, Cheverly, Deanwood, Minnesota Av, Stadium-Armory, Potomac Av, Capitol South, Federal Center SW, L'Enfant Plaza, Smithsonian, Federal Triangle, Metro Center, McPherson Sq, Farragut West, Farragut North, Dupont Circle, Woodley Park, Cleveland Park, Van Ness-UDC, Tenleytown-AU, Friendship Heights, Bethesda, Medical Center, Grosvenor-Strathmore, North Bethesda, Twinbrook, Rockville, Shady Grove

S: New Carrollton, Landover, Cheverly, Deanwood, Minnesota Av, Stadium-Armory, Potomac Av, Capitol South, Eastern Market, Federal Center SW, L'Enfant Plaza, Waterfront, Navy Yard-Ballpark, Anacostia, Congress Heights, Southern Av, Naylor Rd, Suitland, Branch Av, Capitol Heights, Addison Rd, Morgan Blvd, Downtown Largo

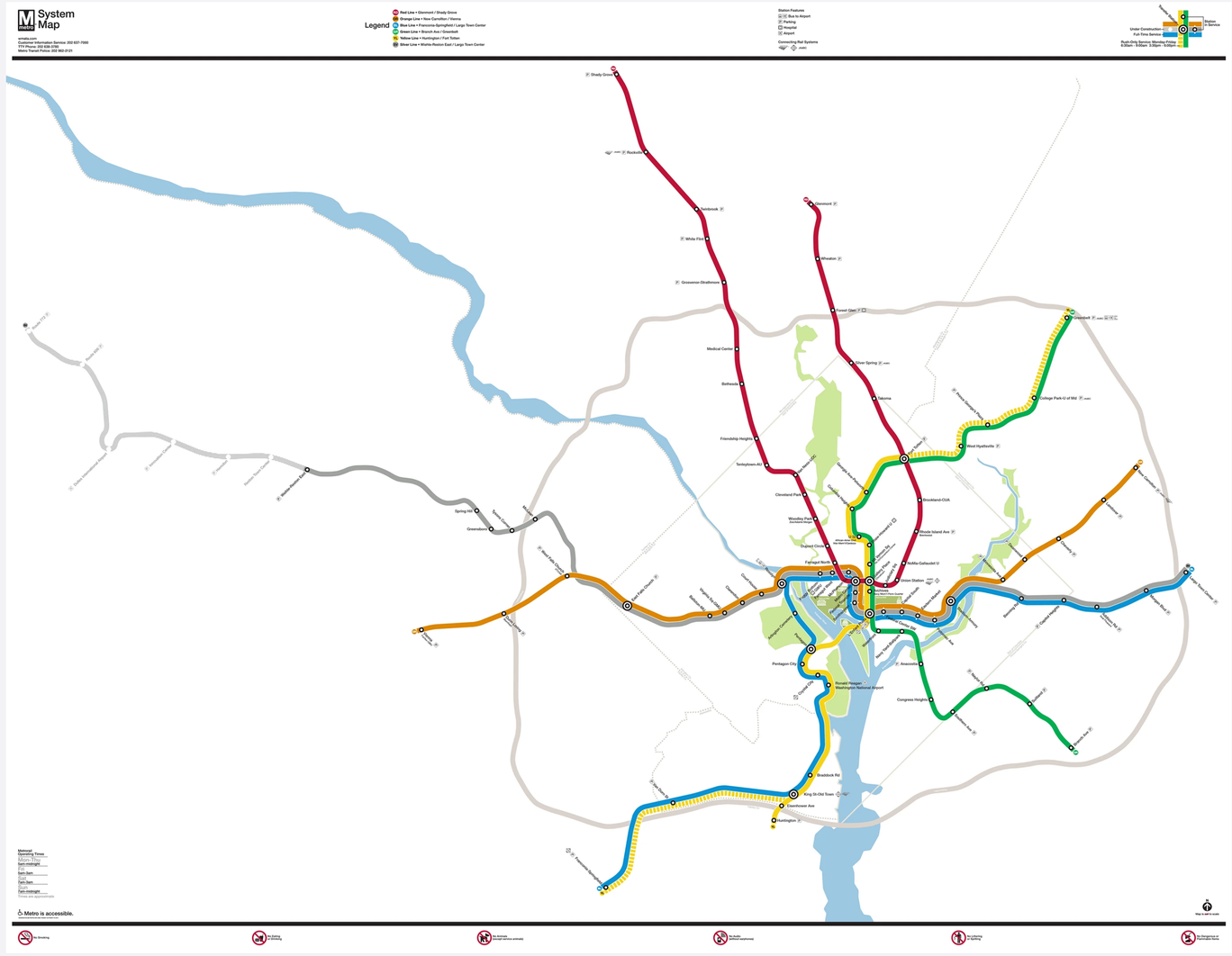
B: Downtown Largo, Morgan Blvd, Addison Rd, Capitol Heights, Benning Rd, Stadium-Armory, Potomac Av, Eastern Market, Capitol South, Federal Center SW, L'Enfant Plaza, Waterfront, Navy Yard-Ballpark, Anacostia, Congress Heights, Southern Av, Naylor Rd, Suitland, Branch Av, Pentagon City, Crystal City, Ronald Reagan Washington National Airport, Potomac Yard, Braddock Rd, King St-Old Town, Eisenhower Av, Huntington, Franconia-Springfield

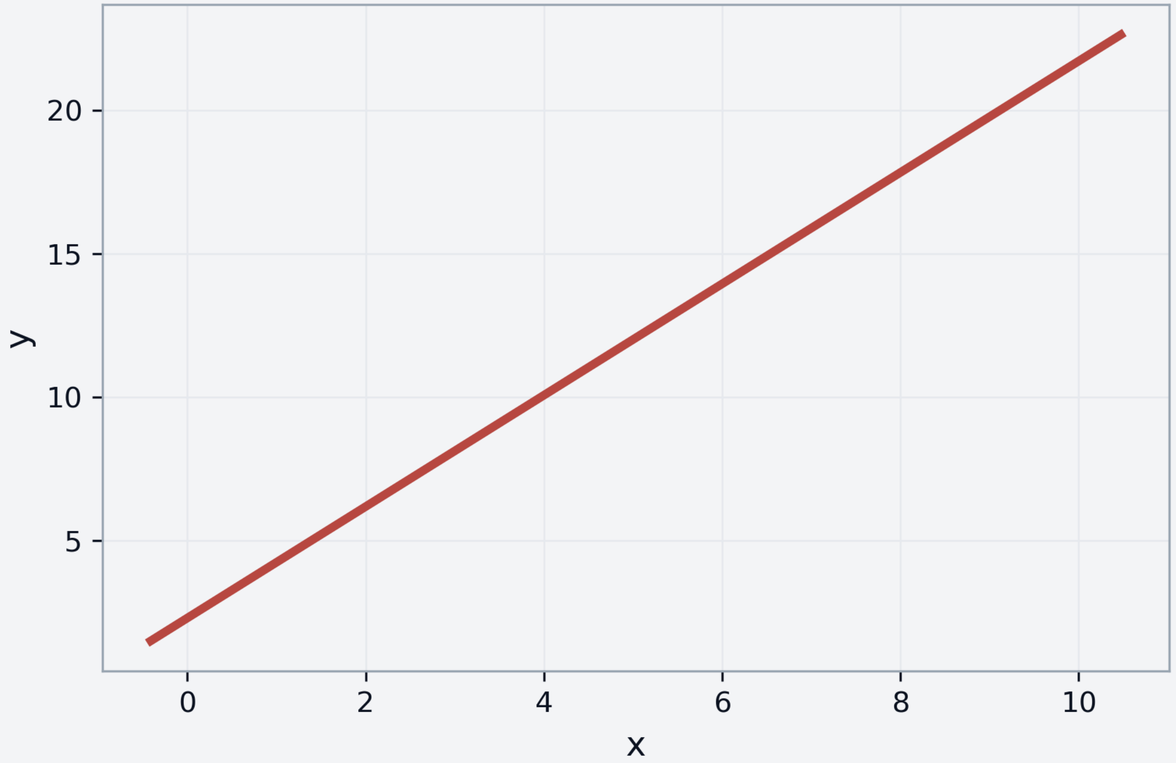
S: Ashburn, Loudoun Gateway, Washington Dulles International Airport, Innovation Center, Herndon, Reston Town Center, Wiehle-Reston East, Spring Hill, Greensboro, Tysons, McLean, Rosslyn, Arlington Cemetery, Pentagon, Pentagon City, Crystal City, Ronald Reagan Washington National Airport, Potomac Yard, Braddock Rd, King St-Old Town, Eisenhower Av, Huntington, Franconia-Springfield

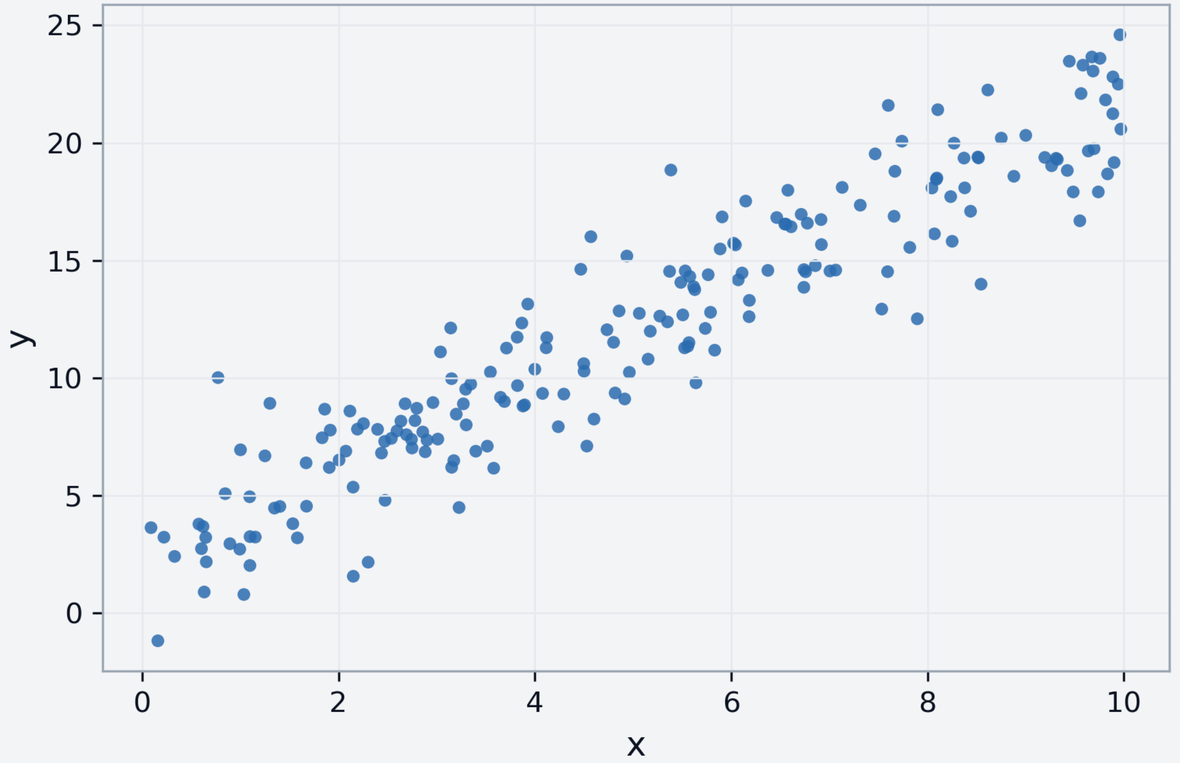
O: Vienna, Dunn Loring, West Falls Church, East Falls Church, Ballston-MU, Virginia Sq-GMU, Clarendon, Court House, Rosslyn, Farragut West, Farragut North, Dupont Circle, Woodley Park, Cleveland Park, Van Ness-UDC, Tenleytown-AU, Friendship Heights, Bethesda, Medical Center, Grosvenor-Strathmore, North Bethesda, Twinbrook, Rockville, Shady Grove

Metro is accessible.

Map is not to scale

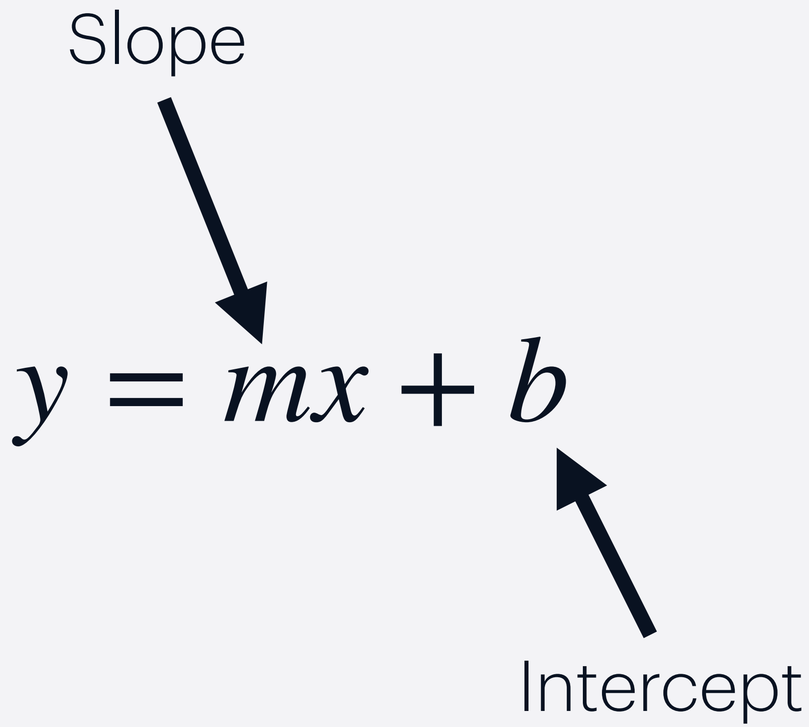


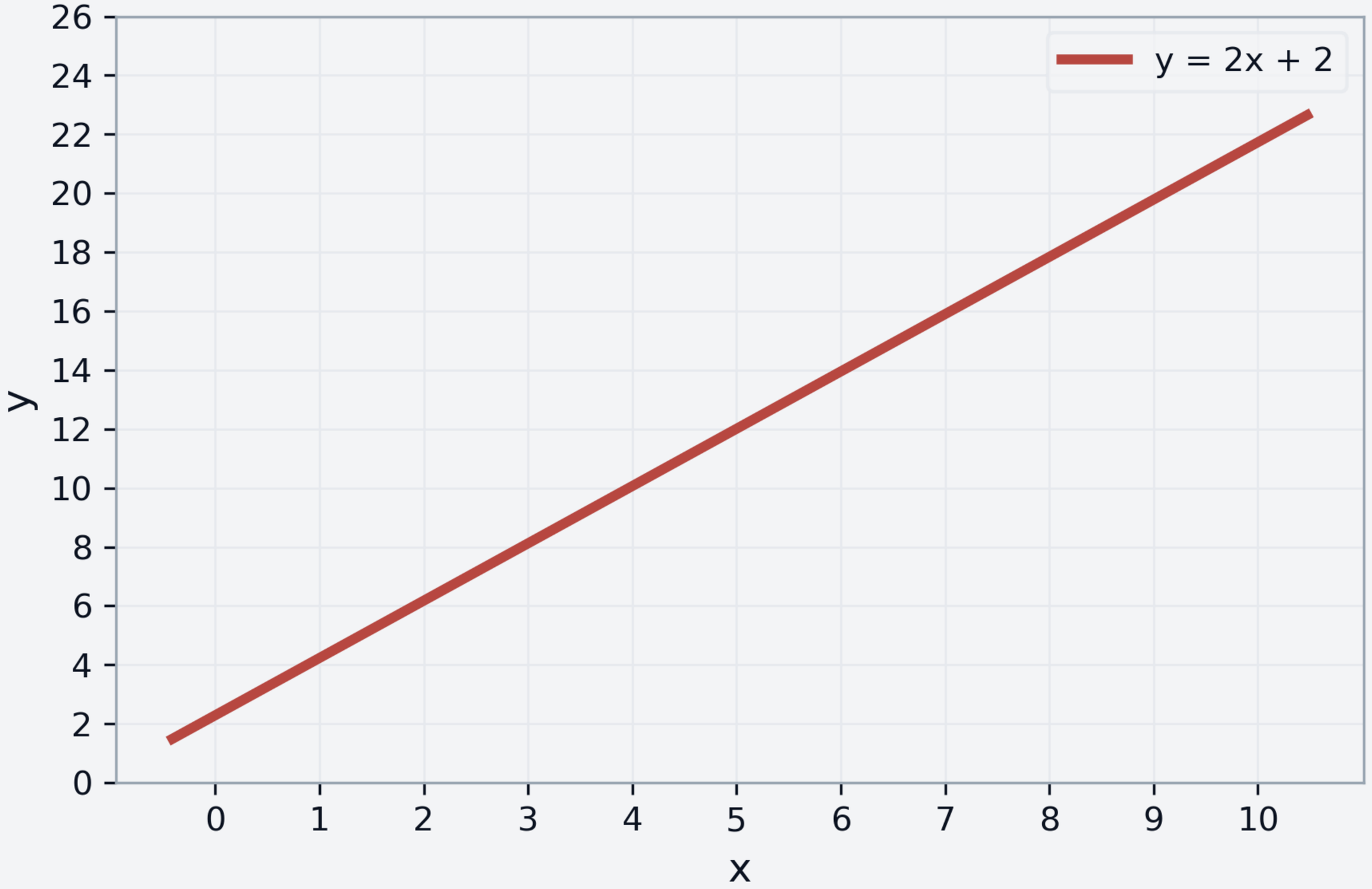

20
15
10
5
y
0
2
4
6
8
10
x

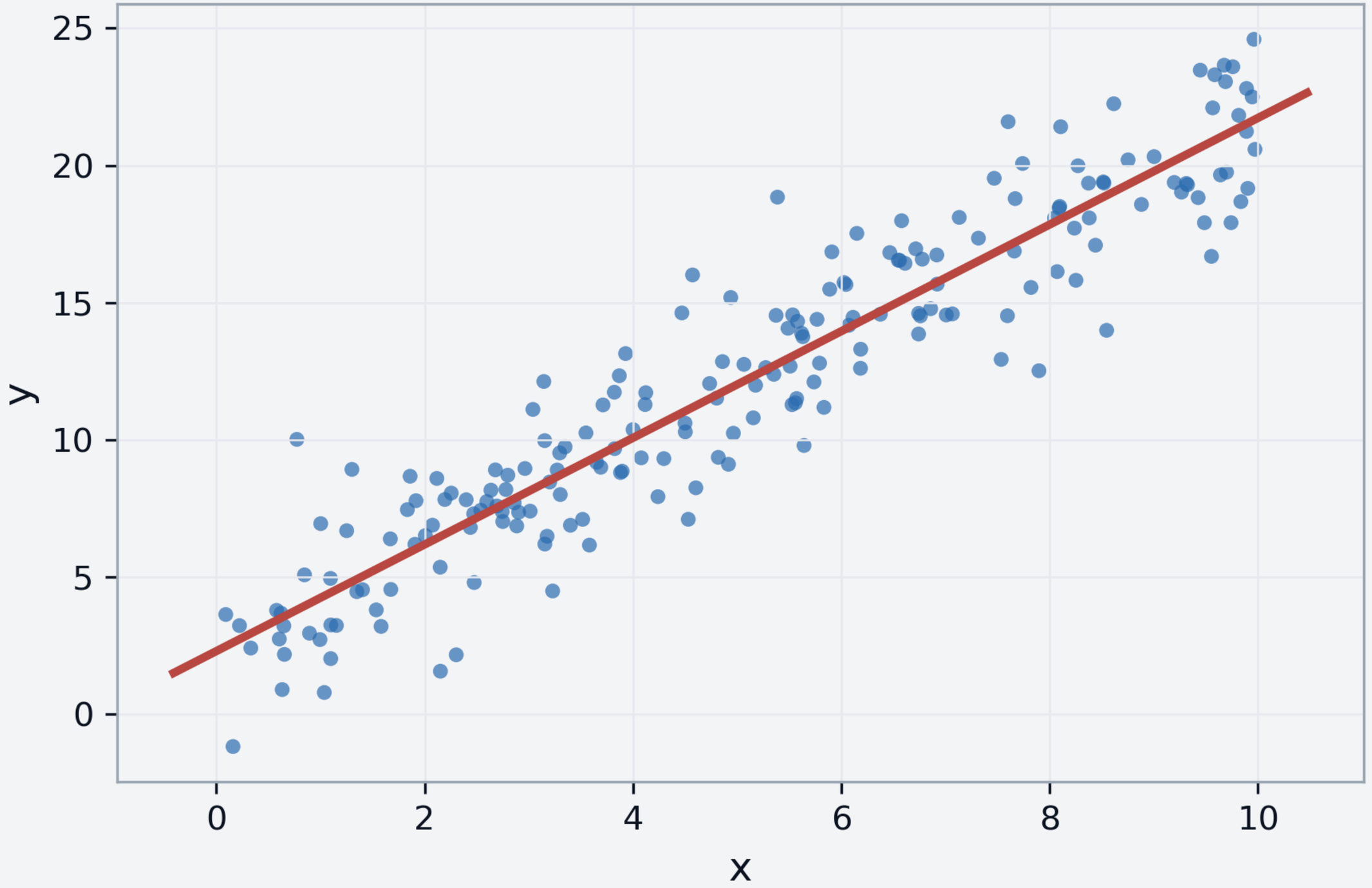

25
20
15
10
5
0
y
0
2
4
6
8
10
x

Linear Regression

$$y = mx + b$$


Slope
$y = mx + b$
Intercept


y = 2x + 2
y
x
0
2
4
6
8
10
12
14
16
18
20
22
24
26
0
1
2
3
4
5
6
7
8
9
10


25
20
15
10
5
0
y
0
2
4
6
8
10
x

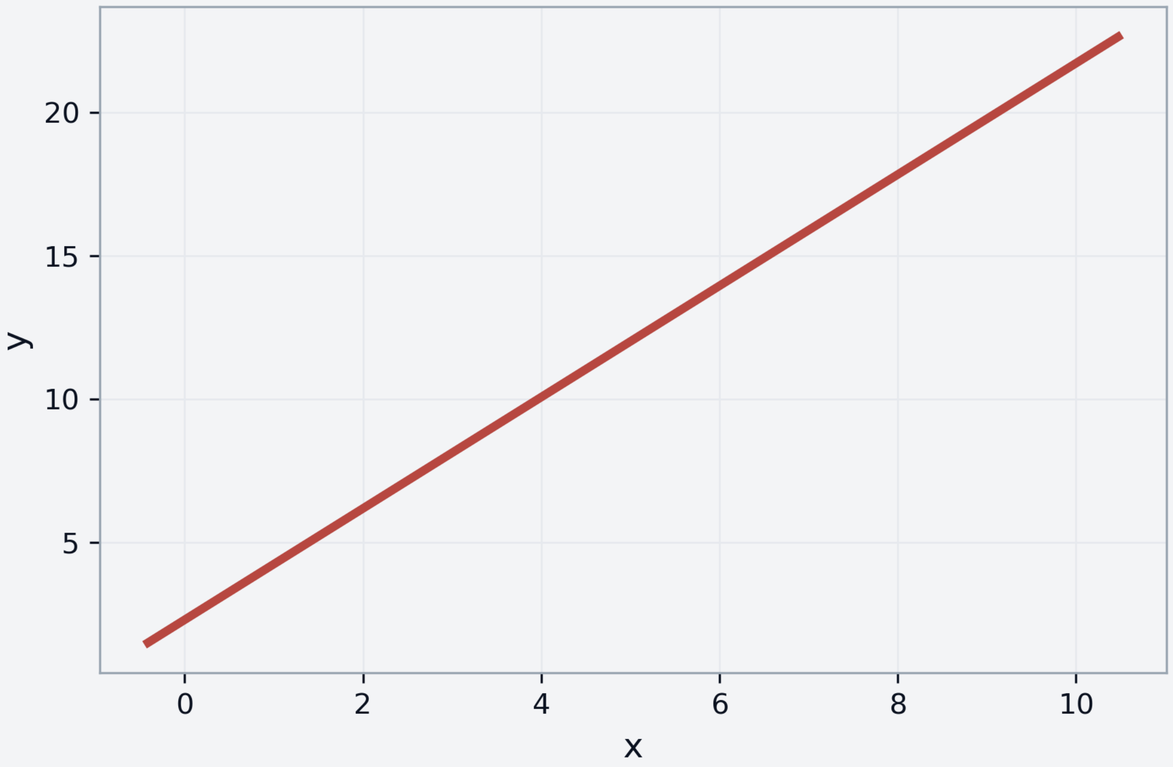
Estimated Values

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

- $\hat{Y}$: Predicted values, predicted outcomes, fitted values. The values of Y predicted by a model. \hat{Y}
- $\hat{\alpha}$: The intercept
- $\hat{\beta}$: The slope coefficient

Estimated Values

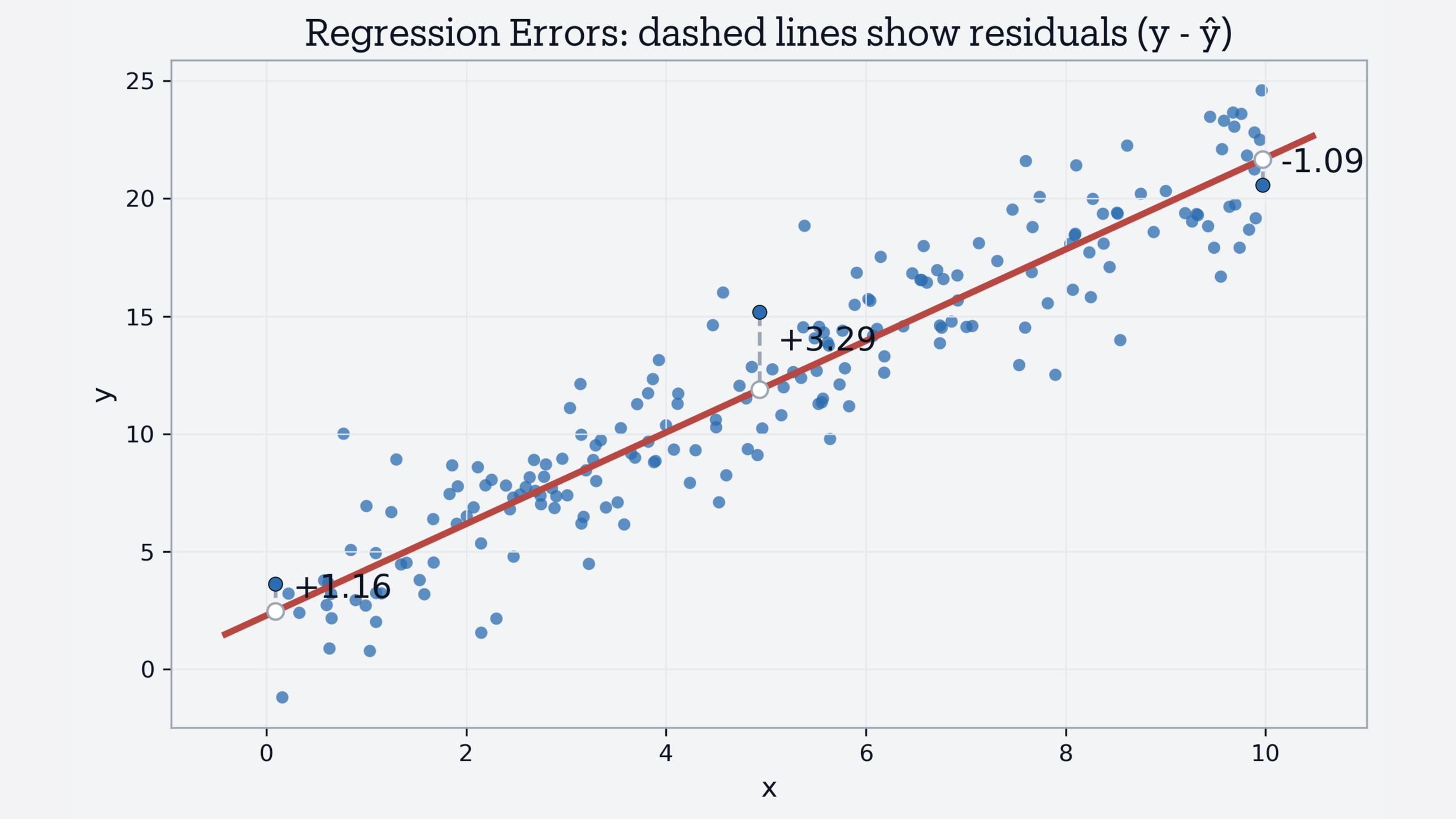


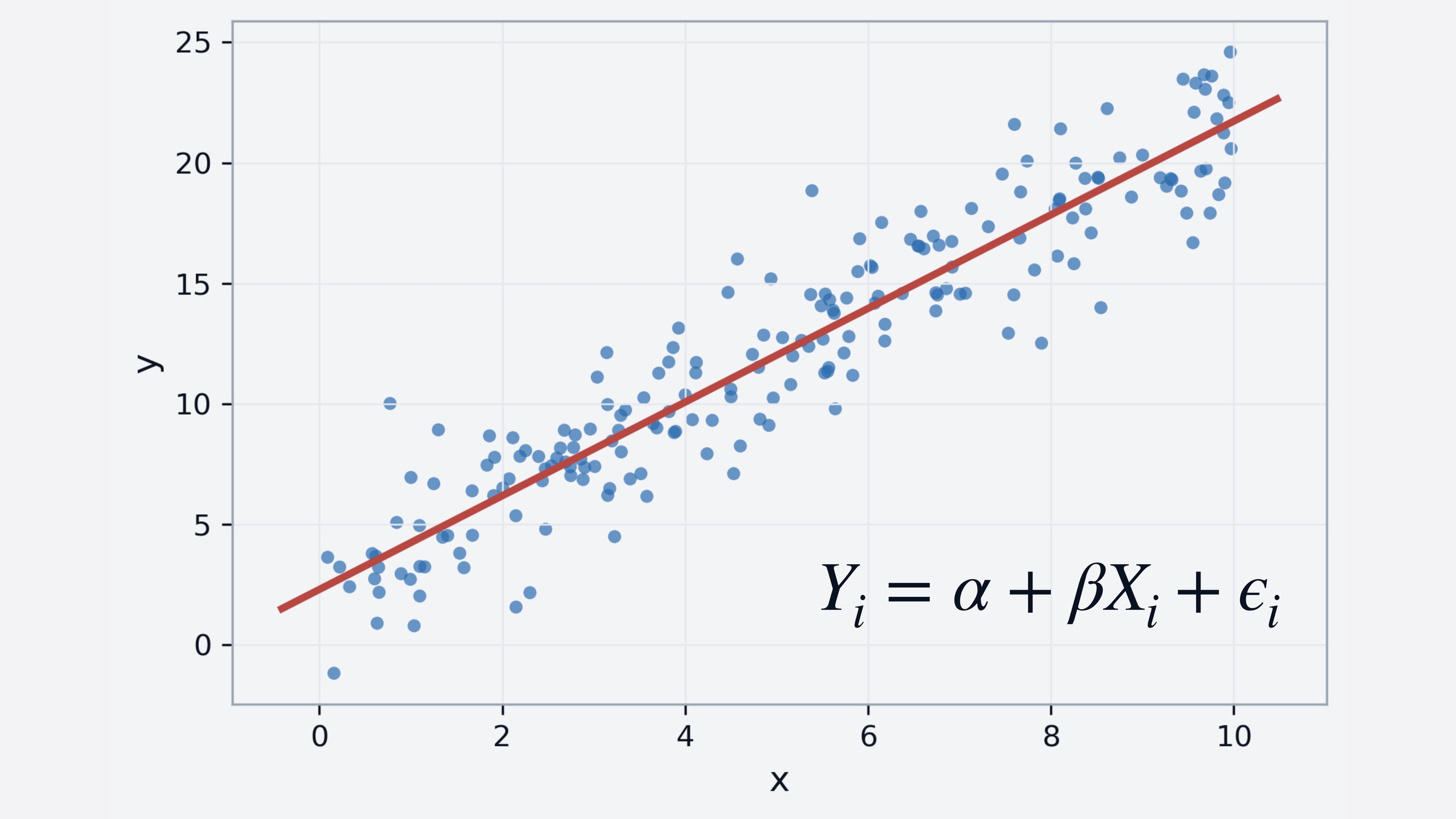
$$\hat{Y}_i = 2 + 2X_i$$

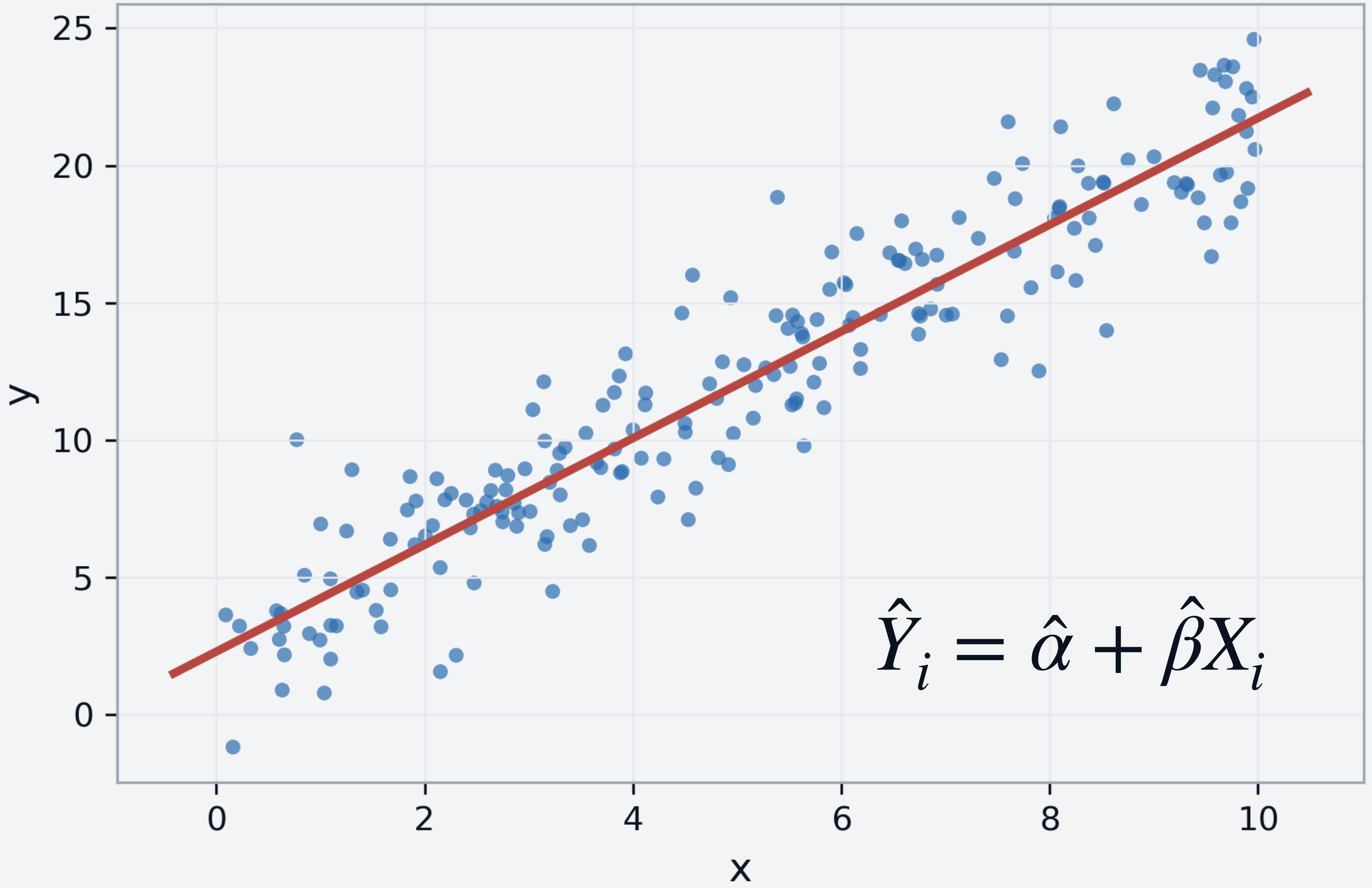
The Linear Regression Model

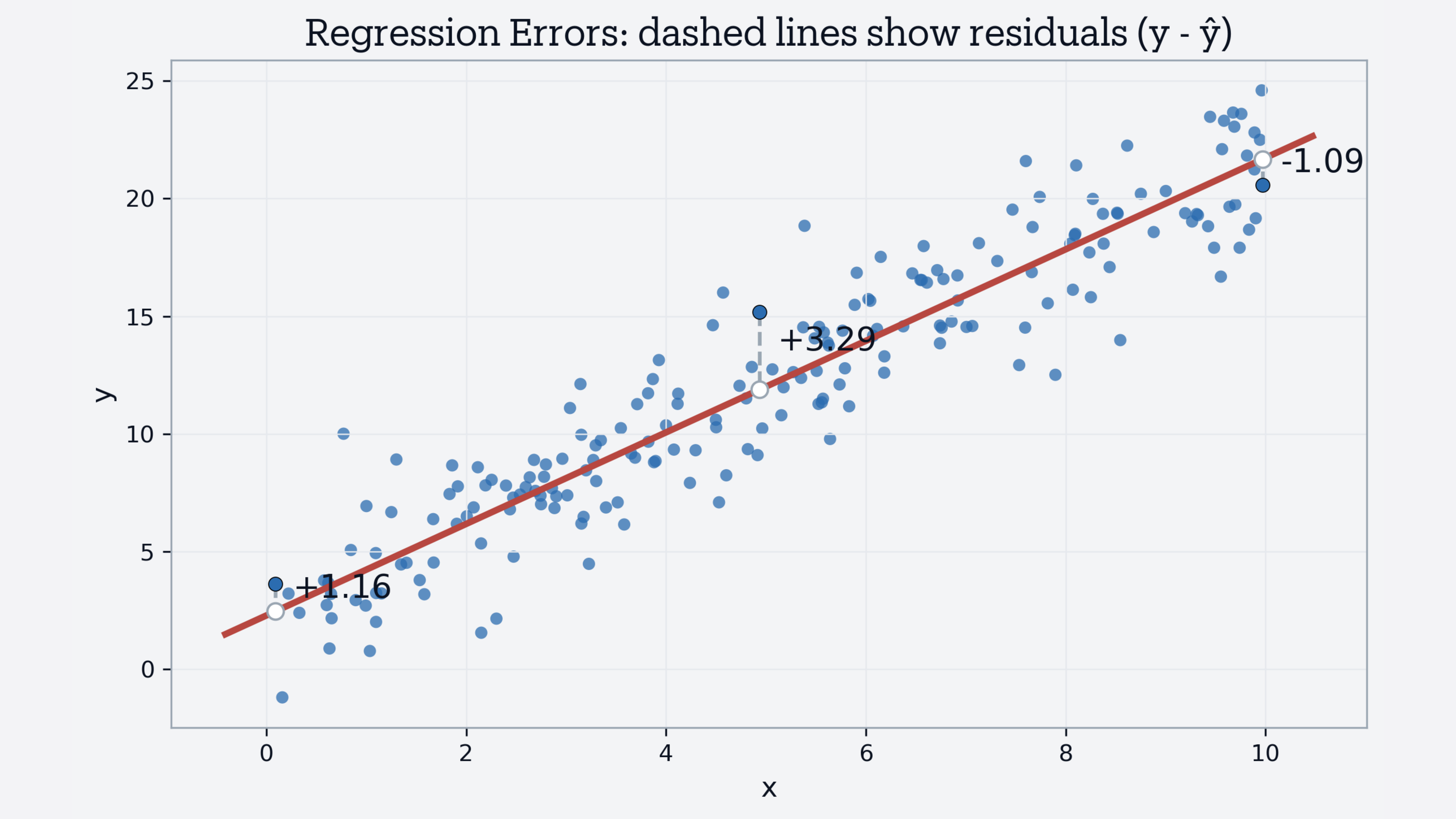
$$Y_i = \alpha + \beta X_i + \epsilon_i$$

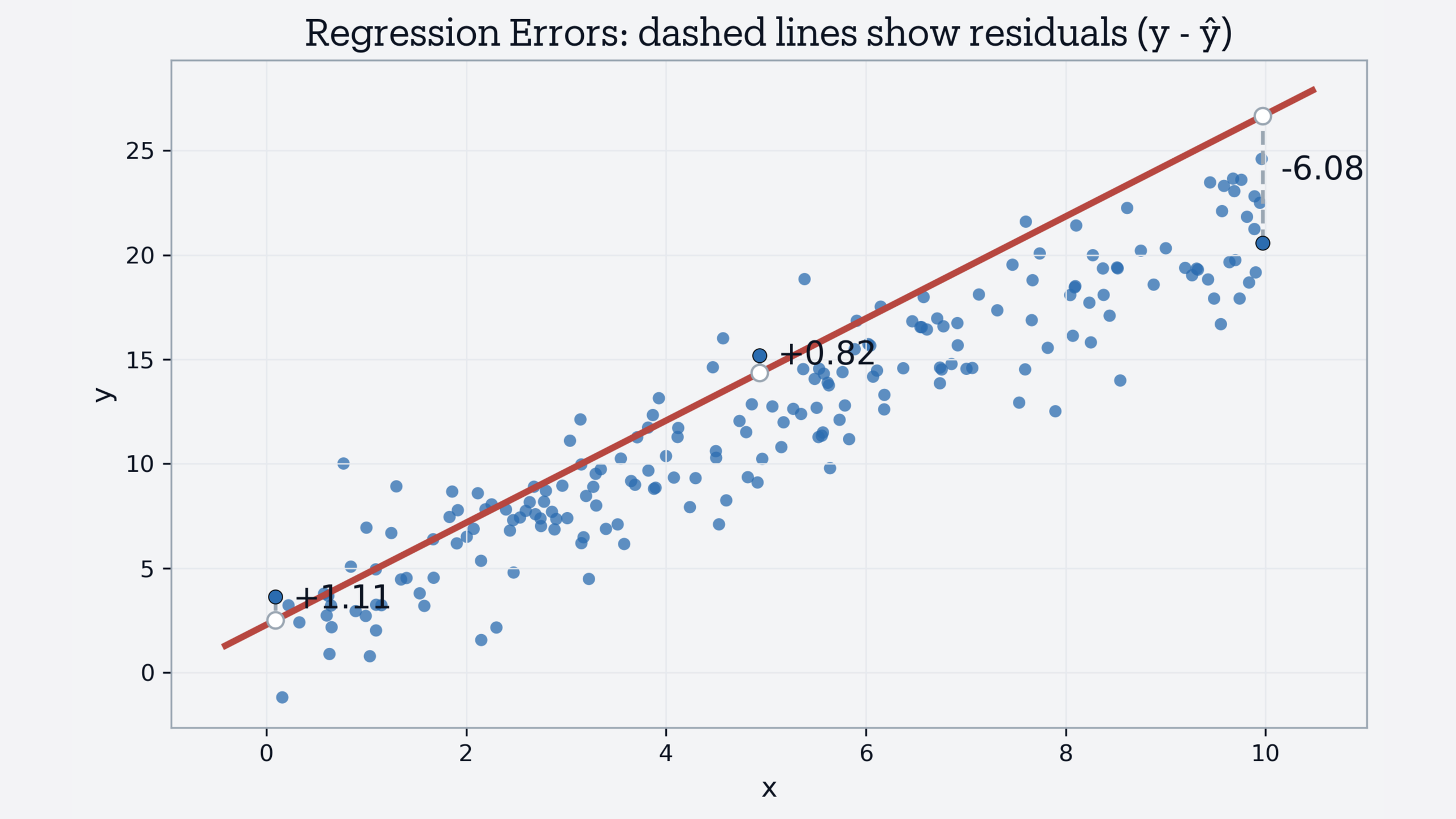
- Notice no hats!
- The new term ϵ_i is the prediction error, the error term, or the residuals
- $\epsilon_i = Y_i - \hat{Y}_i$

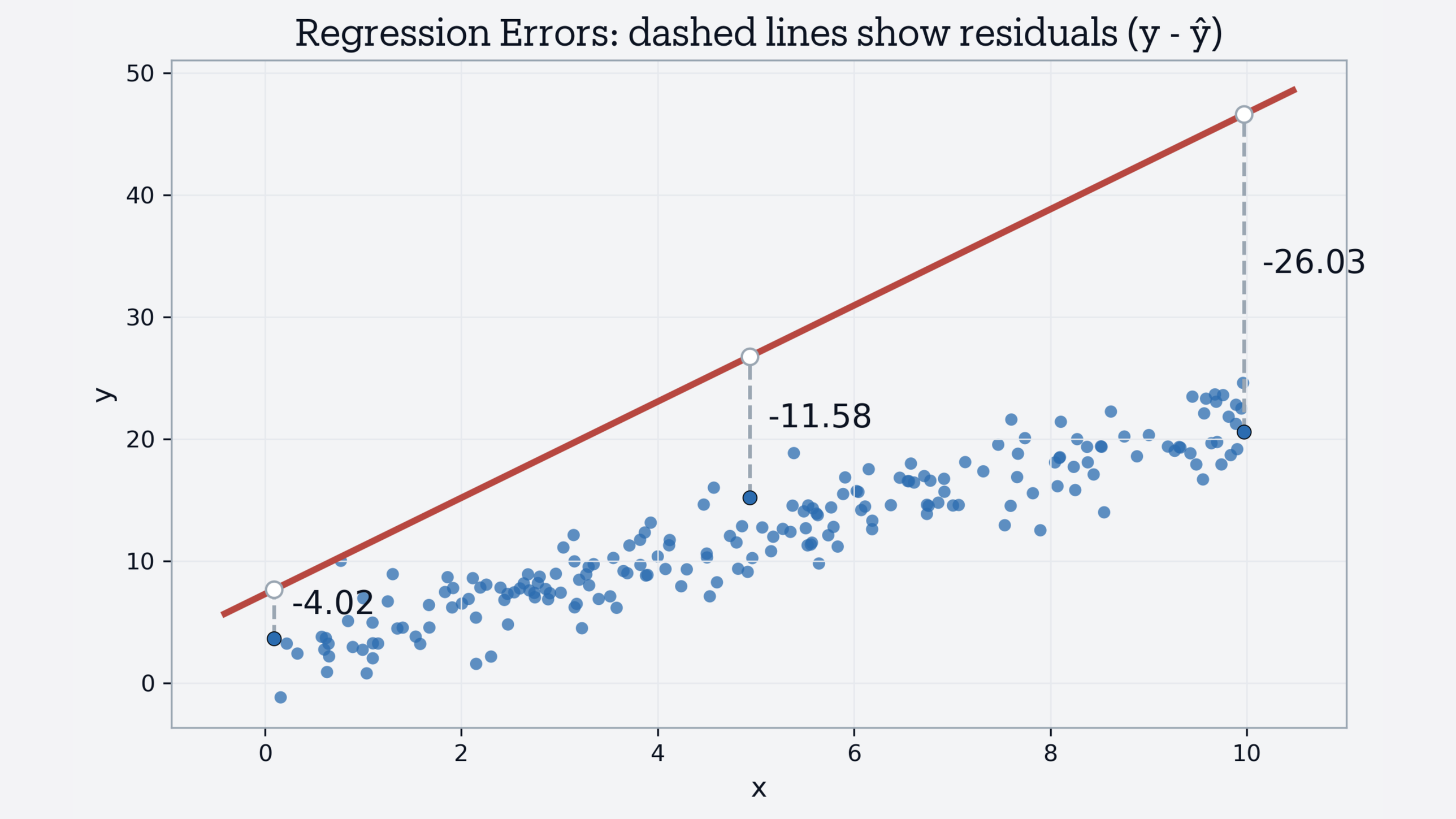

Regression Errors: dashed lines show residuals (y - ŷ)
25
20
15
10
5
0
y
0
2
4
6
8
10
x
+1.16
+3.29
-1.09

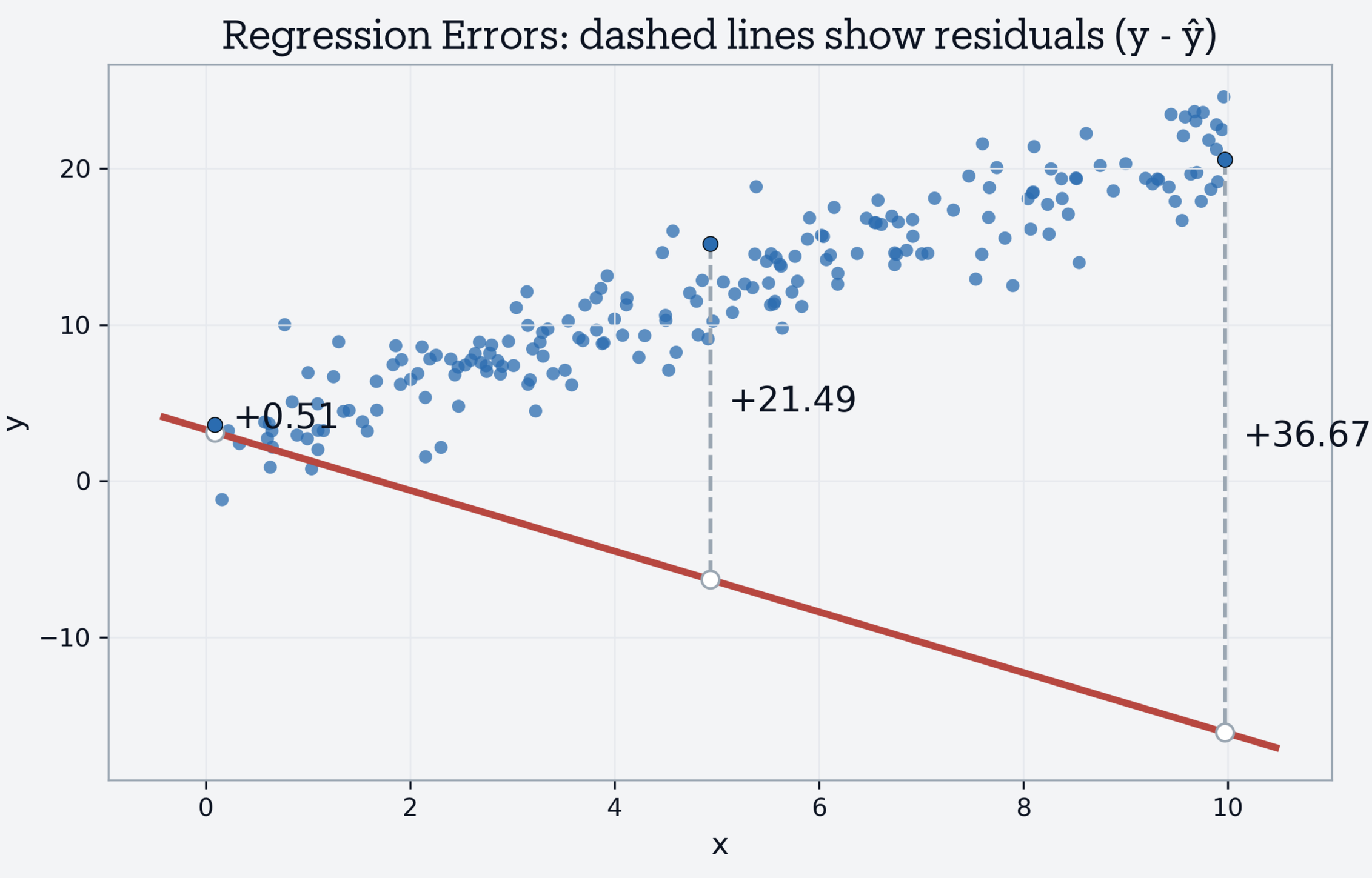

25
20
15
10
5
0
y
0
2
4
6
8
10
x
$Y_i = \alpha + \beta X_i + \epsilon_i$


25
20
15
10
5
0
y
0
2
4
6
8
10
x
$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$


Regression Errors: dashed lines show residuals (y - ŷ)
25
20
15
10
5
0
y
0
2
4
6
8
10
x
+1.16
+3.29
-1.09


Regression Errors: dashed lines show residuals (y - ŷ)
+1.11
+0.82
-6.08
y
x
0
5
10
15
20
25
0
2
4
6
8
10


Regression Errors: dashed lines show residuals (y - ŷ)
50
40
30
20
10
0
y
0
2
4
6
8
10
x
-4.02
-11.58
-26.03


Regression Errors: dashed lines show residuals (y - ŷ)
+0.51
+21.49
+36.67
y
x
20
10
0
−10
0
2
4
6
8
10