Math is fun

MR. HUYNH NAM

Định nghĩa 1.1

Xét tập hợp

$$D \subseteq \mathbb{R}^n; \quad D = \left\{ x = (x_1, x_2, \ldots, x_n) \,|\, x_i \in \mathbb{R}, i = \overline{1, n} \right\} \neq \emptyset$$

và tập hợp $E \subseteq \mathbb{R}$.
Ánh xạ

$$f : D \to E \text{ xác định bởi}$$
$$x \in D \longmapsto u = f(x) = f\left(x_1, x_2, \ldots, x_n\right) \in E$$

gọi là hàm số của n biến số độc lập $x_1, x_2, \ldots, x_n$ xác định trên tập hợp D.
Tập hợp D gọi là miền xác định của hàm số f, tập $f(D)$ gọi là miền giá trị của hàm số f.
Trong trường hợp $n = 2$ hay $n = 3$ ta thường ký hiệu $z = f(x, y)$ hay $u = f(x, y, z)$ tương ứng.

Ở đây ta chỉ xét đối với hàm số hai biến độc lập, số biến > 2 được suy ra tương tự. Ở đây

Ví dụ 1

Tìm miền xác định của hàm và tính giá trị của hàm tại điểm P:

- $z = f(x, y) = \dfrac{\sqrt{x + 2y - 3}}{x^2 - 1}, \quad P(-2; 10.5).$

  Giải:

$$D = \{(x, y) | x + 2y - 3 \geq 0, \ x \neq \pm 1\}; \quad f(P) = \frac{4}{3}$$

Giới hạn và liên tục

Giới hạn hàm nhiều biến

Định nghĩa 2.1

Ta nói dãy điểm $M_n(x_n, y_n)$ hội tụ đến điểm $M_0(x_0, y_0)$ trong $\mathbb{R}^2$ và ký hiệu là $M_n \to M_0$ nếu

$$\lim_{n\to\infty} x_n = x_0 \text{ và } \lim_{n\to\infty} y_n = y_0.$$

Định nghĩa 2.2

Hàm số $f(x,y)$ có giới hạn là L khi $(x,y) \to (x_0, y_0)$ nếu

$$\forall\varepsilon > 0, \exists\delta > 0 : 0 < \rho < \delta \Longrightarrow |f(x,y) - L| < \varepsilon,$$

trong đó $\rho = \sqrt{(x-x_0)^2 + (y-y_0)^2}$; hoặc

$$\forall M_n(x_n, y_n) \to M_0(x_0, y_0) \Longrightarrow \lim_{n\to\infty} f(x_n, y_n) = L.$$

Ký hiệu

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \lim_{\substack{x\to x_0\\ y\to y_0}} f(x,y) = L.$$

Ví dụ 1

Xác định $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ với $f(x,y) = \dfrac{2xy^2}{x^2+y^2}$.

Giải: Chú ý rằng $(0,0) \notin D$.

Với $0 < \varepsilon < 1$ và (x,y) sao cho $0 < \rho = \sqrt{(x-0)^2 + (y-0)^2} = \sqrt{x^2+y^2} < \dfrac{\varepsilon}{2}$ ta có

$$|f(x,y) - 0| = \frac{2|x|\,y^2}{x^2+y^2} < 2|x| < 2\rho < \varepsilon.$$

Vậy $\lim\limits_{(x,y)\to(0,0)} f(x,y) = 0$.

Ví dụ 2

Cho hàm $f(x,y) = \begin{cases} \dfrac{x^2-2y^2}{x^2+y^2}, & (x,y) \neq (0,0) \\ 1, & (x,y) = (0,0) \end{cases}$. Xác định $\lim\limits_{(x,y)\to(0,0)} f(x,y)$.

Giải: Chú ý rằng $(0,0) \in D$. Ta có

- $f(x,y) \to 1$ khi $(x,y) \to (0,0)$ dọc theo trục Ox;
- $f(x,y) \to -2$ khi $(x,y) \to (0,0)$ dọc theo trục Oy;
- $f(x,y) \to \dfrac{1-2k^2}{1+k^2}$ khi $(x,y) \to (0,0)$ dọc theo đường thẳng $y = kx$.

Do đó, $\nexists \lim\limits_{(x,y)\to(0,0)} f(x,y)$.

Định nghĩa 2.3

Hàm số $f(x,y)$ gọi là liên tục tại điểm $(a,b) \in D$ nếu

$$\lim_{\substack{x\to a\\ y\to b}} f(x,y) = f(a,b).$$

Hàm số liên tục tại mọi điểm trong miền $D \subset \mathbb{R}^2$ gọi là liên tục trên D.

Ví dụ 3

Cho $f(x,y) = \dfrac{\sqrt{x+2y+1}}{x-1}$. Chứng minh rằng f liên tục tại gốc toạ độ.

Giải: Ta có $f(0,0) = -1$ và

$$\lim_{\substack{x\to 0\\ y\to 0}} f(x,y) = \lim_{\substack{x\to 0\\ y\to 0}} \frac{\sqrt{x+2y+1}}{x-1} = -1 = f(0,0).$$

Đạo hàm và vi phân

Đạo hàm riêng

Định nghĩa 3.1

Cho hàm số $z = f(x, y)$ xác định và liên tục trên miền $\mathcal{D}$. Với $(x, y) \in \mathcal{D}$, giữ y không đổi, cho x số gia Δx ta có điểm $(x + \Delta x, y) \in \mathcal{D}$. Nếu tồn tại

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

thì giới hạn đó được gọi là đạo hàm riêng (cấp 1) của hàm số $f(x, y)$ tại điểm (x, y), ký hiệu

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial x} = f'_x = D_x f.$$

Tương tự ta cũng có đạo hàm riêng của hàm f theo biến số y:

$$\frac{\partial z}{\partial y} = \frac{\partial f}{\partial y} = f'_y = D_y f.$$

Ví dụ 1

Cho $f(x, y) = \sqrt{2x + 3y^2 + 1}$. Tìm $\dfrac{\partial f(x,y)}{\partial x}$, $\dfrac{\partial f(x,y)}{\partial y}$ và tính giá trị của chúng tại điểm $(2, 4)$.

Giải:

$$\begin{aligned} f'_x &= \frac{\partial f(x,y)}{\partial x} = \frac{\partial}{\partial x}\sqrt{2x + 3y^2 + 1} = \frac{1}{\sqrt{2x + 3y^2 + 1}}; \quad f'_x(2,4) = \frac{1}{\sqrt{53}} \\ f'_y &= \frac{\partial f(x,y)}{\partial y} = \frac{\partial}{\partial y}\sqrt{2x + 3y^2 + 1} = \frac{3y}{\sqrt{2x + 3y^2 + 1}}; \quad f'_y(2,4) = \frac{12}{\sqrt{53}} \end{aligned}$$

Ví dụ 2

Cho $f(x, y) = \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases}$. Chứng minh $f'_x(0, 0) = 0$, $f'_y(0, 0) = 0$.

Từ định nghĩa ta thấy: tính đạo hàm riêng có quy tắc giống như tính đạo hàm của hàm số một biến số, nghĩa là khi tính theo x thì xem y là hằng số, còn khi tính theo y thì xem x là hằng số.

Cực trị hàm hai biến số

Cực trị địa phương

Định nghĩa 4.1

Hàm số $z = f(x, y)$ có miền xác định là $\mathcal{D} \subset \mathbb{R}^2$. Ta nói hàm số $f(x, y)$ đạt cực đại (cực tiểu) tại điểm $(x_0, y_0) \in \mathcal{D}$ nếu

$$\forall (x, y) \in \delta(x_0, y_0) := 0 < \rho = \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$$

thì

$$f(x, y) < f(x_0, y_0) \quad (f(x, y) > f(x_0, y_0)).$$

Định lý 4.1

(Điều kiện cần)
Nếu hàm số $z = f(x, y)$ khả vi tại điểm $M_0(x_0, y_0) \in \mathcal{D}$ mà tại đó hàm số đạt cực trị thì

$$f'_x(x_0, y_0) = 0; \quad f'_y(x_0, y_0) = 0. \qquad (*)$$

Các điểm thỏa mãn hệ (*) gọi là điểm dừng hoặc điểm tới hạn.

Ví dụ 1

Tìm cực trị của hàm số sau $z = \frac{1}{2}xy + (47 - x - y)\left(\frac{x}{3} + \frac{y}{4}\right)$.
Hàm số xác định $\forall (x, y) \in \mathbb{R}^2$. Tính

$$\frac{\partial z}{\partial x} = -\frac{1}{12}y - \frac{2}{3}x + \frac{47}{3}; \quad \frac{\partial z}{\partial y} = -\frac{1}{2}y - \frac{1}{12}x + \frac{47}{4}$$

Cho $\frac{\partial z}{\partial x} = 0, \quad \frac{\partial z}{\partial y} = 0 \Longrightarrow \begin{cases} 8x + y = 188 \\ x + 6y = 142 \end{cases}$, thu được điểm dừng $M_0(21, 20)$

Cực trị hàm hai biến số

Cực trị có điều kiện

Bài toán

Tìm cực trị của hàm số $z = f(x, y)$ với điều kiện $\varphi(x, y) = 0$.

Để giải bài toán này, ta sử dụng phương pháp nhân tử Lagrange:
Lập hàm

$$u = f(x, y) + \lambda\varphi(x, y),$$

trong đó λ là tham số chưa xác định.

Điều kiện cần của cực trị có điều kiện

$$\begin{cases} \dfrac{\partial u}{\partial x} = \dfrac{\partial f}{\partial x} + \lambda\dfrac{\partial \varphi}{\partial x} & = 0 \\ \dfrac{\partial u}{\partial y} = \dfrac{\partial f}{\partial y} + \lambda\dfrac{\partial \varphi}{\partial y} & = 0 \\ \qquad\quad \varphi(x, y) & = 0. \end{cases}$$

Giải hệ trên ta được điểm nghi ngờ $M_o(x_o, y_o)$ và λ_o

Ví dụ 2

Tìm cực trị của hàm $z = xy$ với điều kiện $2x + 3y - 5 = 0$.
Lập hàm nhân tử Lagrange $\Phi = xy + \lambda(2x + 3y - 5)$ và tính

$$\frac{\partial \Phi}{\partial x} = y + 2\lambda; \quad \frac{\partial \Phi}{\partial y} = x + 3\lambda.$$

Tìm điểm dừng từ hệ $\begin{cases} \dfrac{\partial \Phi}{\partial x} = y + 2\lambda = 0 \\ \dfrac{\partial \Phi}{\partial y} = x + 3\lambda = 0 \\ 2x + 3y - 5 = 0 \end{cases}$. Giải hệ trên ta được $\lambda = -\dfrac{5}{12}; \quad x = \dfrac{5}{4}; \quad y = \dfrac{5}{6}$.

Simple Linear Regression

MR. HUYNH NAM

Nội dung

1. Mô tả bài toán hồi quy
2. Mô hình hồi quy tuyến tính đơn giản
3. Đồ thị của hàm số hồi quy tuyến tính đơn giản
4. Biểu đồ Scatter
5. Hệ số tương quan
6. Nhận diện tương quan
7. Mô hình hóa bài toán dự báo
8. Ý tưởng giải mô hình hồi quy tuyến tính
9. Residual (phần dư)
10. Ordinary Least Squares (OLS) in Least Square Method
11. Tổng kết mô hình hồi quy tuyến tính trong fitting data
12. Thực hành với dữ liệu mẫu
13. Mô tả dữ liệu và bài toán dự báo
14. Đánh giá mô hình
15. Xây dựng ứng dụng dự báo

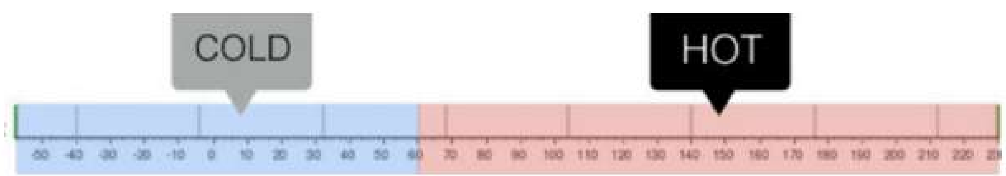
Regression

What will be the temperature tomorrow?

Fahrenheit

Classification

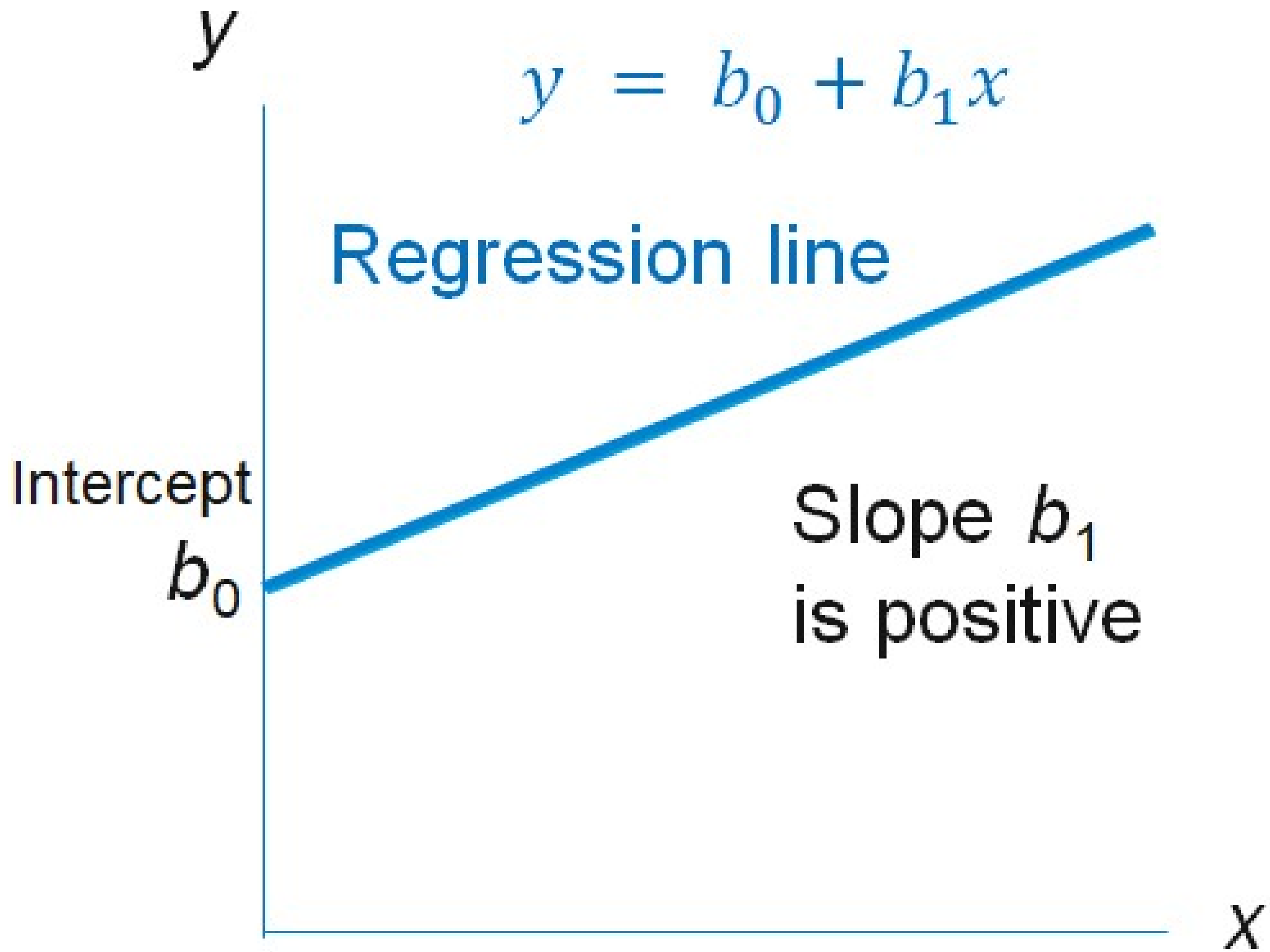
Will it be hot or cold tomorrow?



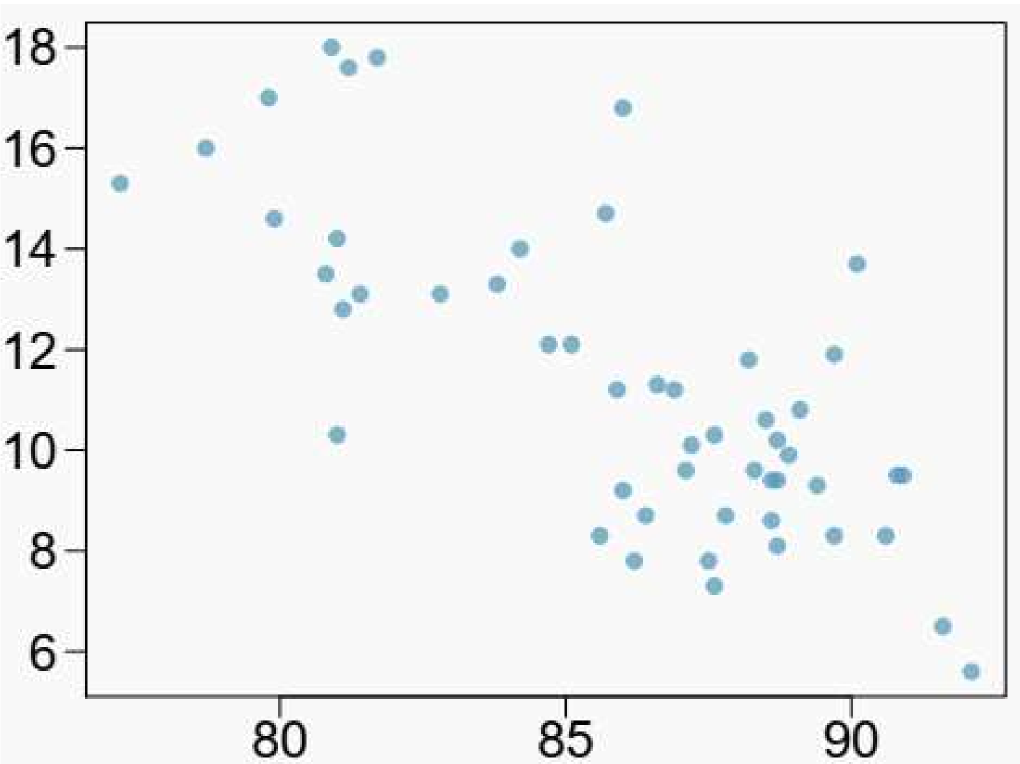
Fahrenheit

Simple Linear Regression

$$y = b_0 + b_1 * x$$



y
$y = b_0 + b_1 x$
Regression line
Intercept
b_0
Slope b_1
is positive
x



Response variable?

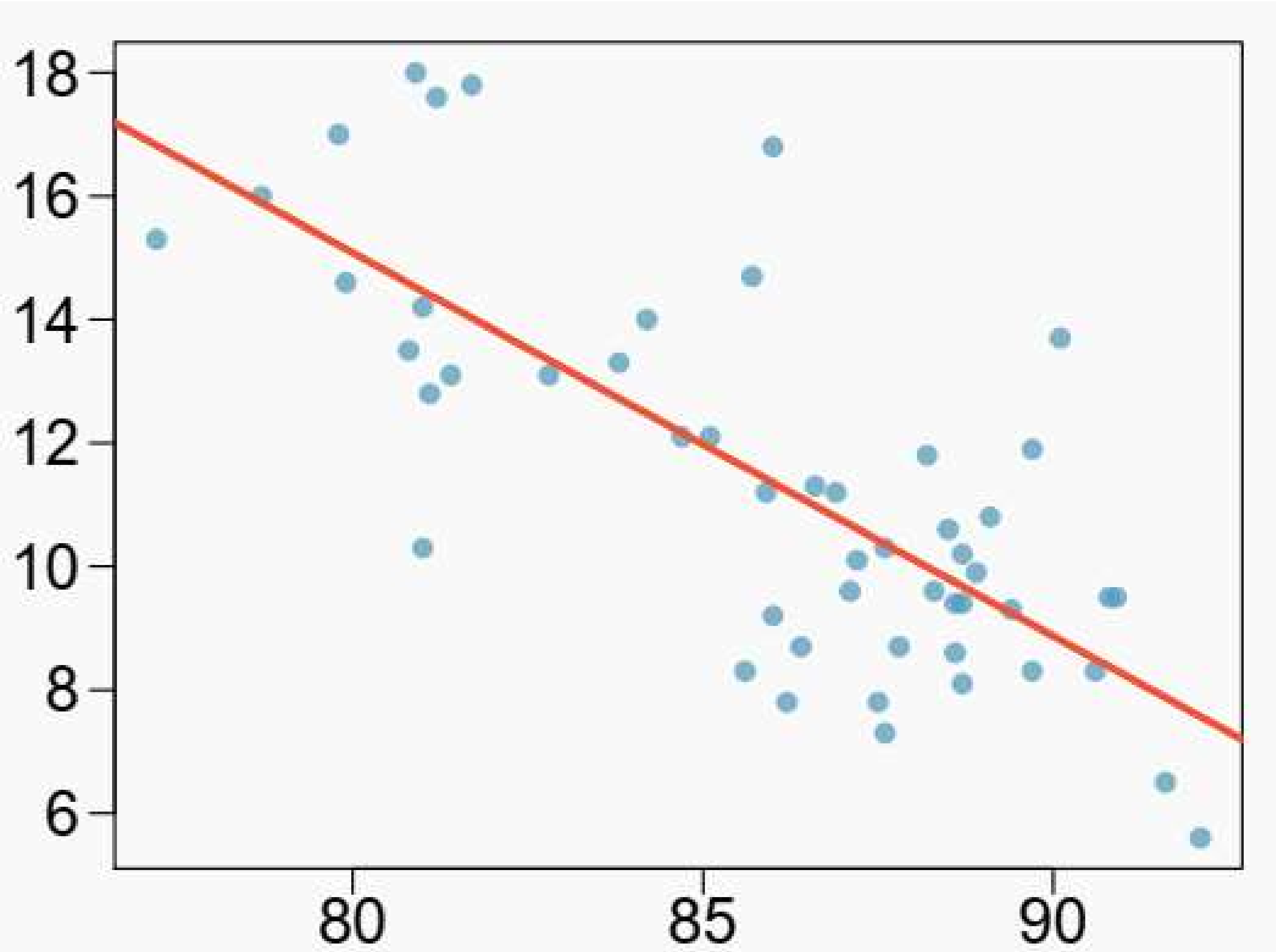
Explanatory variable?

Relationship?

linear, negative, moderately strong

Quantifying the relationship

- *Correlation* describes the strength of the *linear* association between two variables.
- It takes values between -1 (perfect negative) and +1 (perfect positive).
- A value of 0 indicates no linear association.



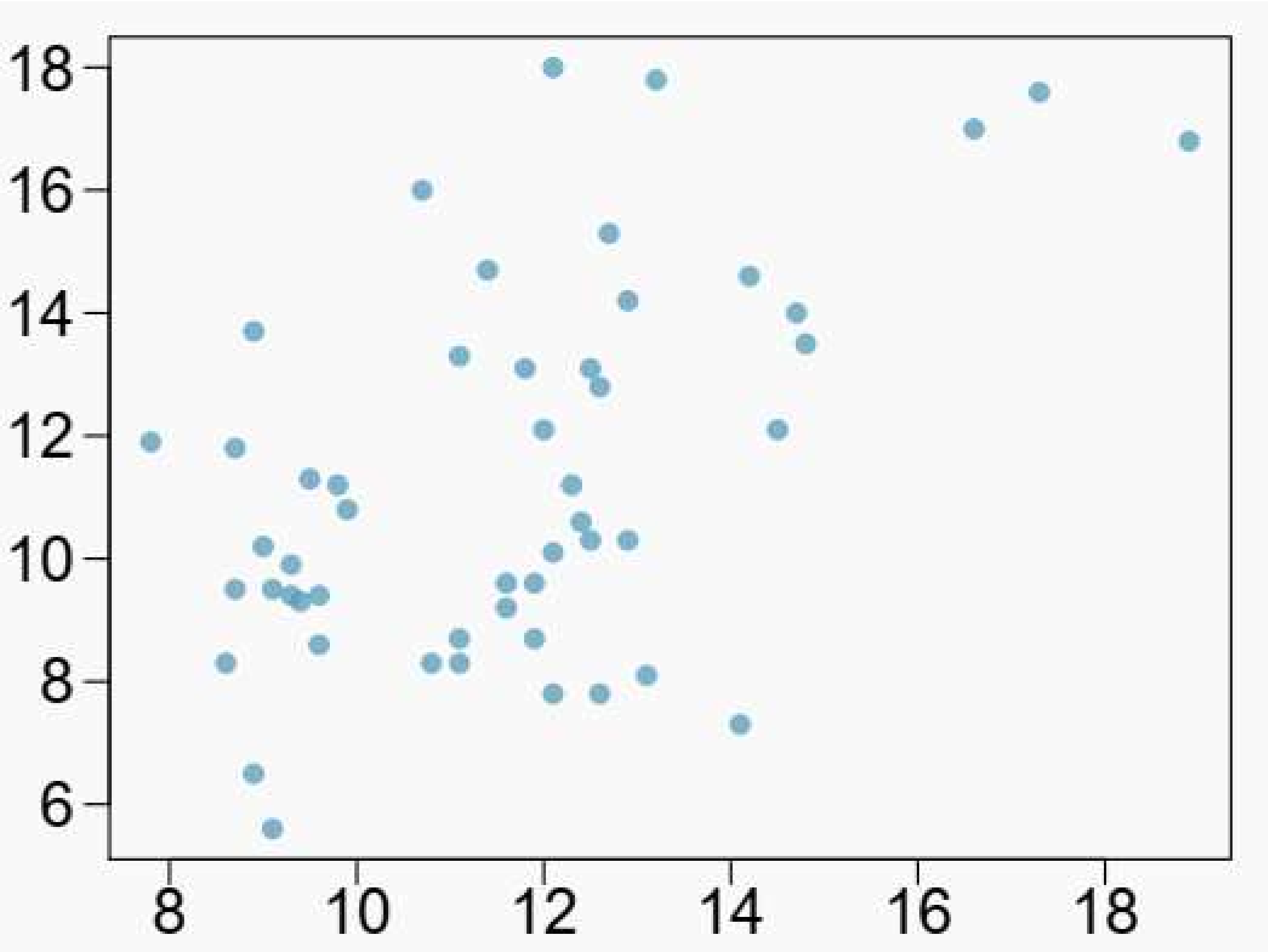
(a) 0.6

(b) -0.75

(c) -0.1

(d) 0.02

(e) -1.5



(a) 0.1

(b) -0.6

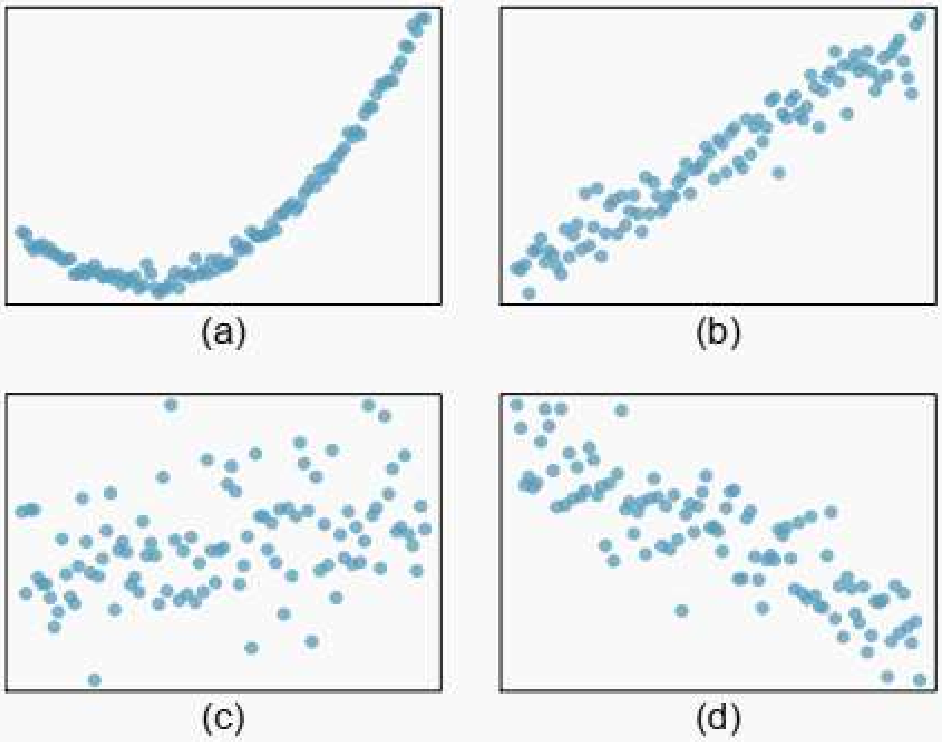
(c) -0.4

(d) 0.9

(e) 0.5

Assessing the correlation

Which of the following is has the strongest correlation, i.e. correlation coefficient closest to +1 or -1?



(a) (b) (c) (d)

Simple Linear Regression

X	Y
2	8
3	8
7	4
5	1
9	4

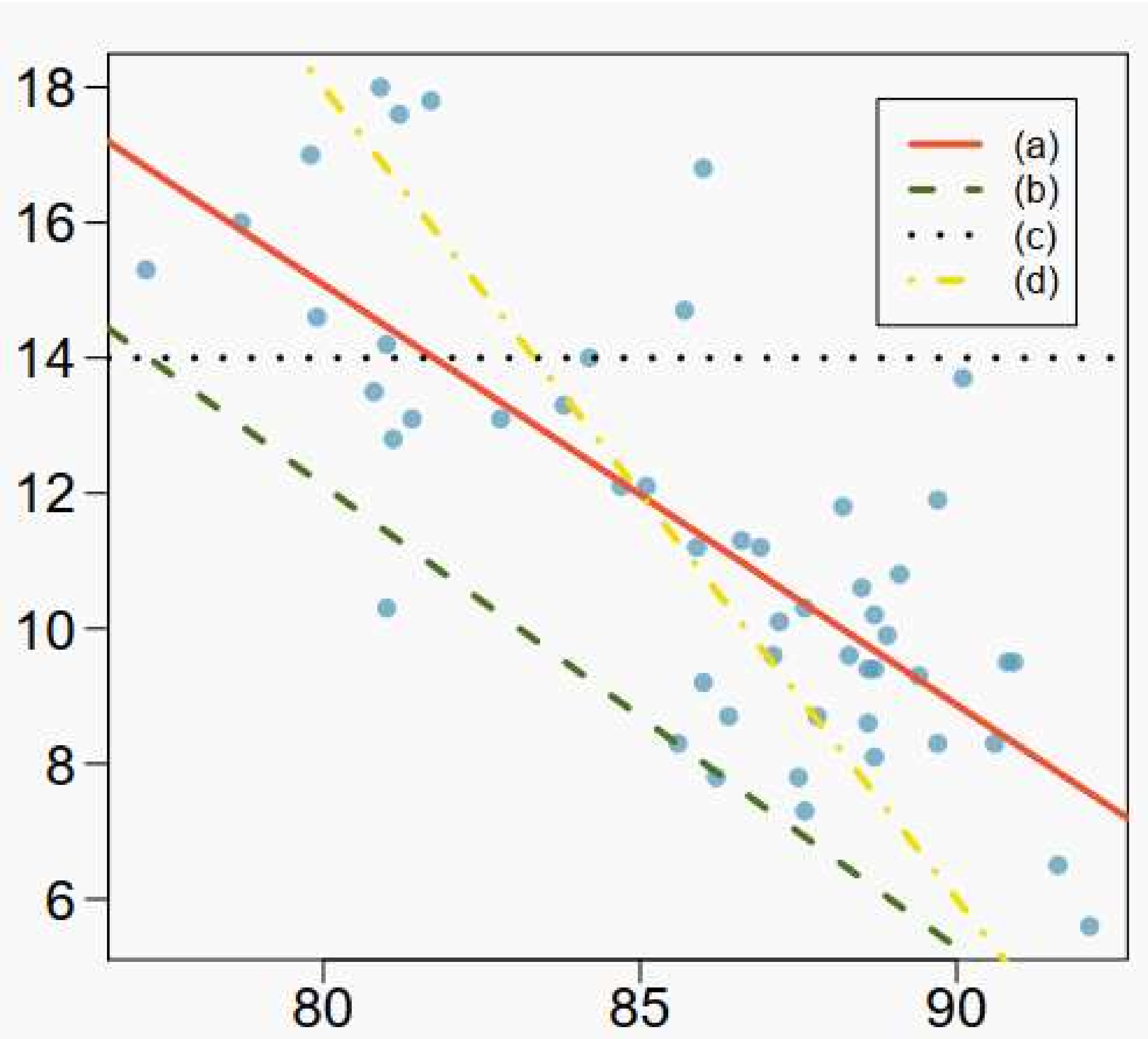
$$\hat{y} = b_0 + b_1 X$$

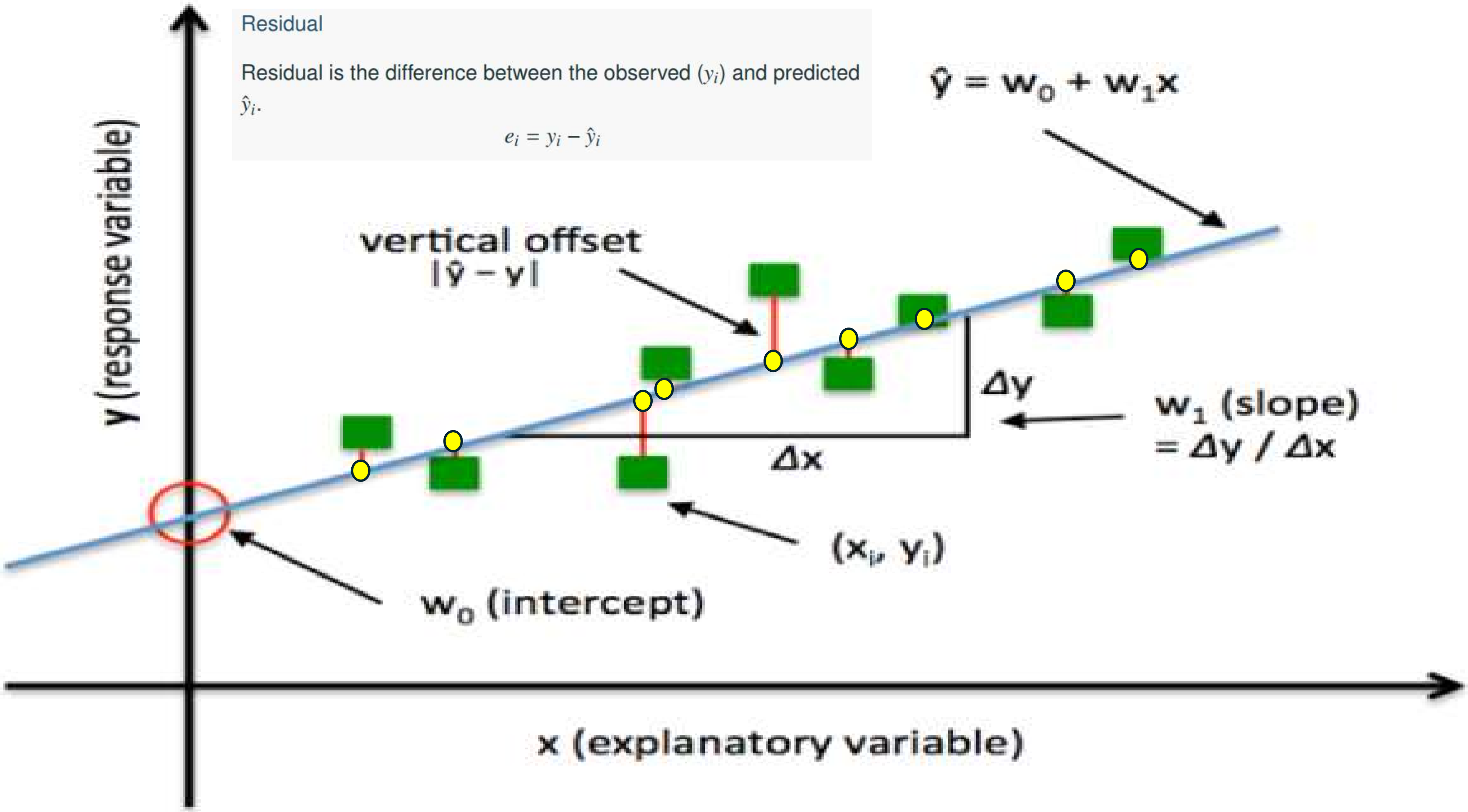
Dependent variable

Independent variable

y-intercept (constant)

Slope coefficient


18
16
14
12
10
8
6
80
85
90
(a)
(b)
(c)
(d)


Residual
Residual is the difference between the observed (y_i) and predicted $\hat{y}_i$.
$$e_i = y_i - \hat{y}_i$$
$\hat{y} = w_0 + w_1 x$
vertical offset
$|\hat{y} - y|$
Δy
Δx
w_1 (slope)
$= \Delta y / \Delta x$
(x_i, y_i)
w_0 (intercept)
y (response variable)
x (explanatory variable)

A measure for the best line

- We want a line that has small residuals:
  1. Option 1: Minimize the sum of magnitudes (absolute values) of residuals

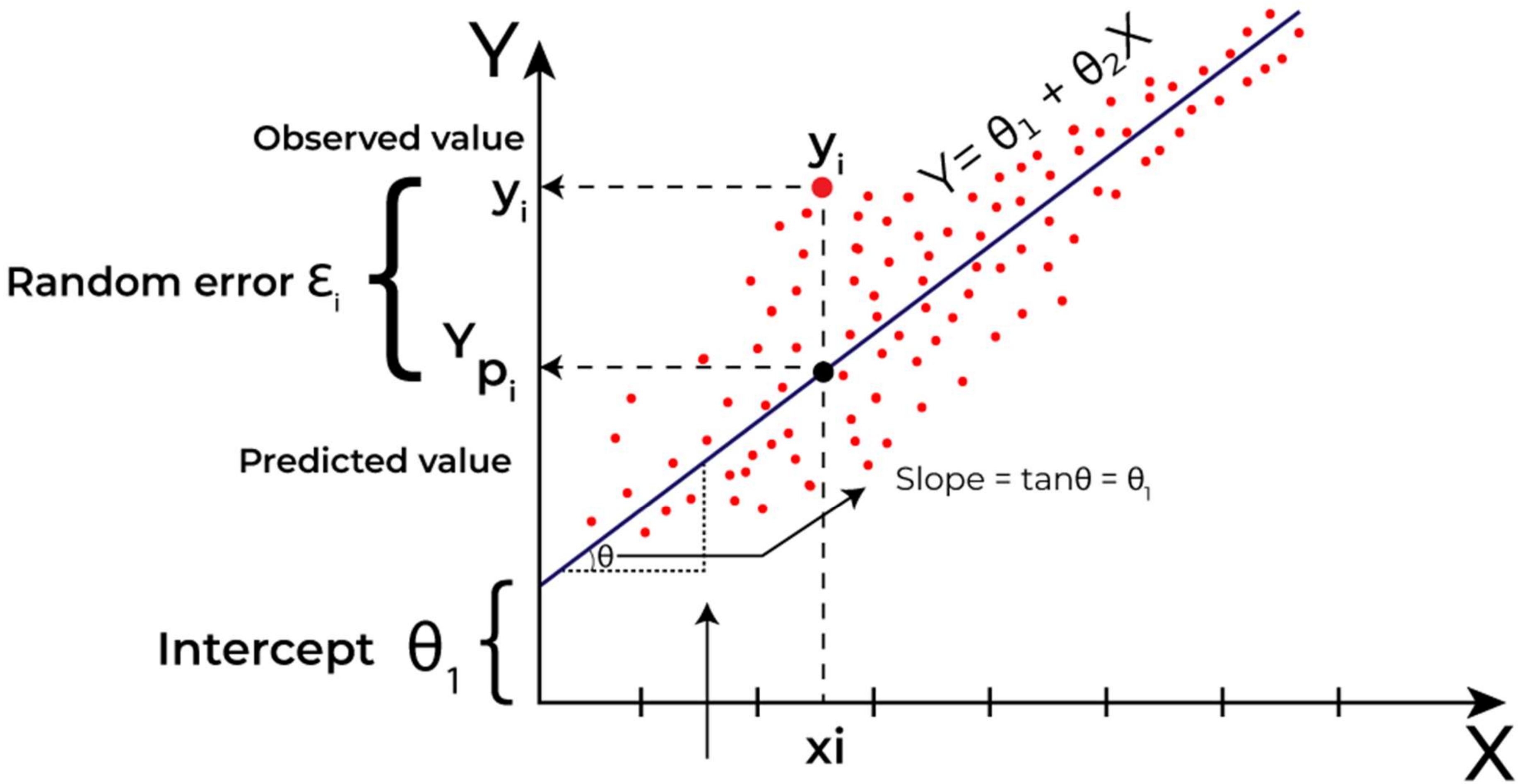
     $$|e_1| + |e_2| + \cdots + |e_n|$$

  2. Option 2: Minimize the sum of squared residuals – *least squares*

     $$e_1^2 + e_2^2 + \cdots + e_n^2$$

- Why least squares?
  1. Most commonly used
  2. Easier to compute by hand and using software
  3. In many applications, a residual twice as large as another is usually more than twice as bad

Review



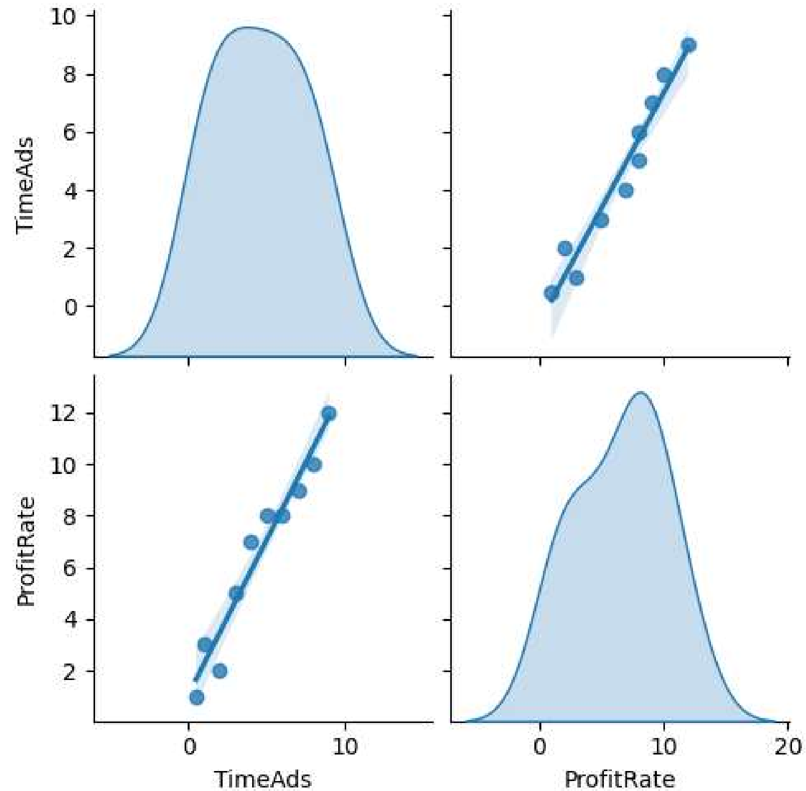
Y
Observed value
y_i
y_i
Random error ε_i
Y_{p_i}
Predicted value
$Y = \theta_1 + \theta_2 X$
Slope = tanθ = θ_1
θ
Intercept θ_1
xi
X

Thực hành với Data Sample

MR. HUYNH NAM

	TimeAds	ProfitRate
0	0.5	1
1	2.0	2
2	4.0	7
3	5.0	8
4	3.0	5
5	6.0	8
6	7.0	9
7	9.0	12
8	8.0	10
9	1.0	3

```
sns.pairplot(df[['TimeAds','ProfitRate']],
             diag_kind='kde', kind='reg')
plt.show()
```



	TimeAds	ProfitRate
0	0.5	1
1	2.0	2
2	4.0	7
3	5.0	8
4	3.0	5
5	6.0	8
6	7.0	9
7	9.0	12
8	8.0	10
9	1.0	3

```
X_train       y_train

array([[8.],  array([[10.],
       [3.],         [ 5.],
       [5.],         [ 8.],
       [2.],         [ 2.],
       [6.],         [ 8.],
       [1.]])        [ 3.]])
```

```
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.4,
                                                    random_state=16)
```

```
X_test          y_test

array([[7. ],   array([[ 9.],
       [4. ],          [ 7.],
       [0.5],          [ 1.],
       [9. ]])         [12.]])
```

```python
# Xây dựng mô hình hồi quy tuyến tính đơn biến: ProfitRate = f(TimeAds)
# ProfitRate = A0 + A1*TimeAds + epsilon
from sklearn import linear_model

# Khai báo mô hình SLR
model = linear_model.LinearRegression()

# Huấn luyện mô hình
model.fit(X_train,y_train)
```

```
#To retrieve the intercept:
print(model.intercept_)

#For retrieving the slope (coefficient):
print(model.coef_)
```

```
y_test

array([[ 9.],
       [ 7.],
       [ 1.],
       [12.]])
```

```
X_test

array([[7. ],
       [4. ],
       [0.5],
       [9. ]])
```

```python
y_test_pred = model.predict(X_test)
y_test_pred
```

```python
import numpy as np
from sklearn import metrics

# Tính giá trị dự báo ProfitRate dựa trên dữ liệu TimeAds của tập test
y_test_pred = model.predict(X_test)

print('Score or R-Squared:', model.score(X_test, y_test))

print('Mean Absolute Error:',
      metrics.mean_absolute_error(y_test, y_test_pred))
print('Root Mean Squared Error:',
      np.sqrt(metrics.mean_squared_error(y_test, y_test_pred)))
print('Mean Absolute Percentage Error:',
      metrics.mean_absolute_percentage_error(y_test, y_test_pred))
```

```python
# Lưu trữ mô hình đã huấn luyện xuống ổ đĩa
'''
Lưu trữ mô hình hồi quy xuống thiết bị lưu trữ với tên
model_SLR.sav
'''
import pickle
pickle.dump(model, open('model_SLR.sav', 'wb'))
```

```python
import pickle
#Load model từ storage
loaded_model = pickle.load(open('model_SLR.sav', 'rb'))

vTimeAds = float(input('Nhập thời lượng quảng cáo: '))

predicted_ProfitRate = loaded_model.predict([[vTimeAds]])

print(f'Dự báo điểm ProfitRate là: {predicted_ProfitRate}')
```