Linear regression

NHH
TECH3

Statistical model:

$$Y_i = \mu + \epsilon_i, \qquad i = 1, \dots, n$$

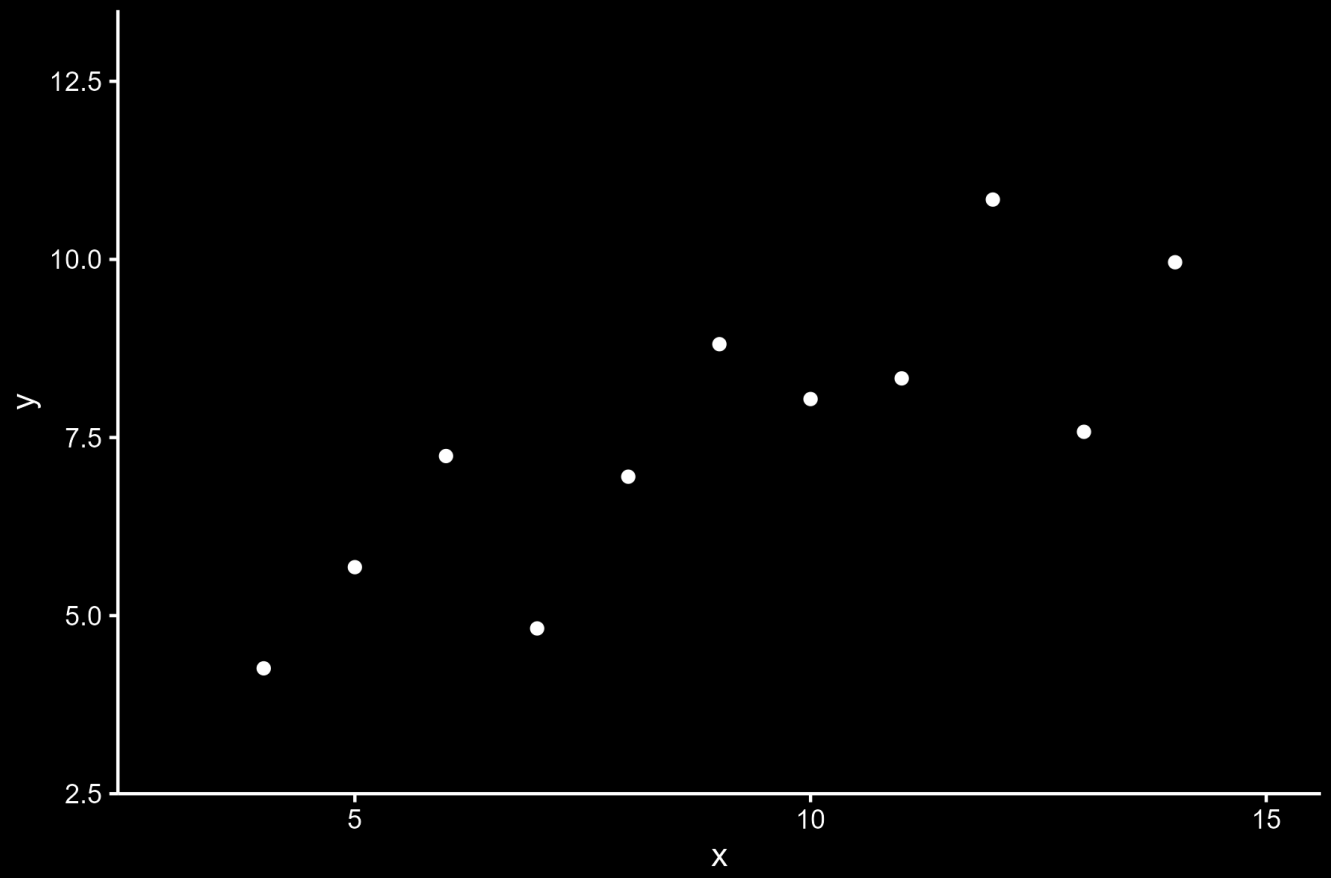
$$\text{Data} = \text{Model} + \text{Error}$$

where $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are independent and identically distributed.

$$\underbrace{f(\epsilon_1, \epsilon_2, \dots, \epsilon_n)}_{\text{Joint distribution}} = \underbrace{f(\epsilon_1) f(\epsilon_2) \cdots f(\epsilon_n)}_{\text{Marginal distributions}}$$

(drawn from the same distribution f)

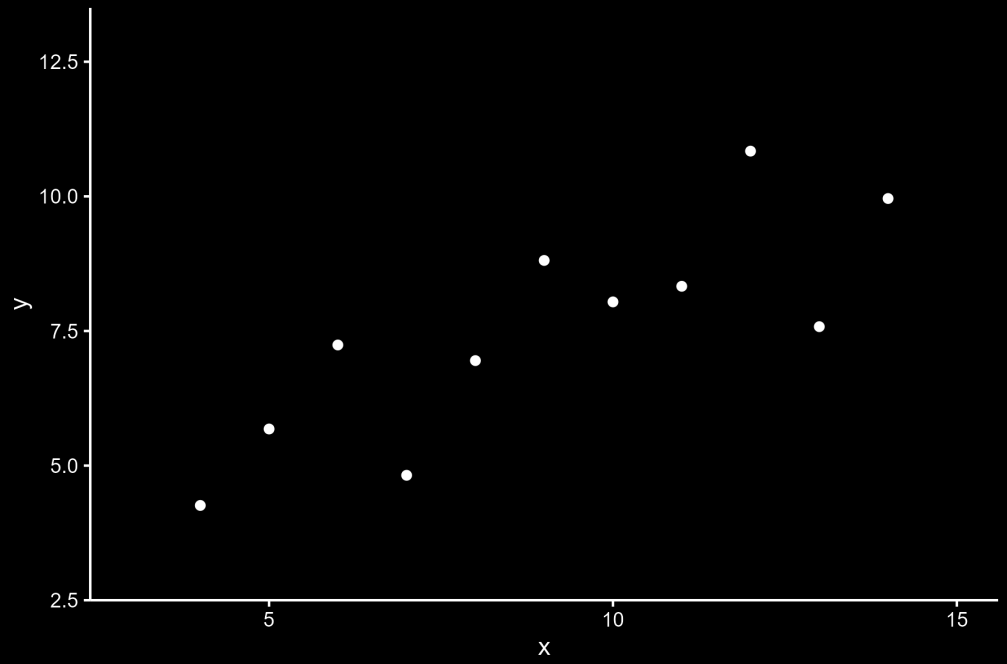
NHH
TECH3



NHH
TECH3

Data = Model + Error

Y: Dependent variable



X: Independent variable

$$Y_i = \beta_0 + X_i\ \beta_1 + \epsilon_i$$

$$\hat{y}_i = \beta_0 + x_i\ \beta_1$$

NHH
TECH3

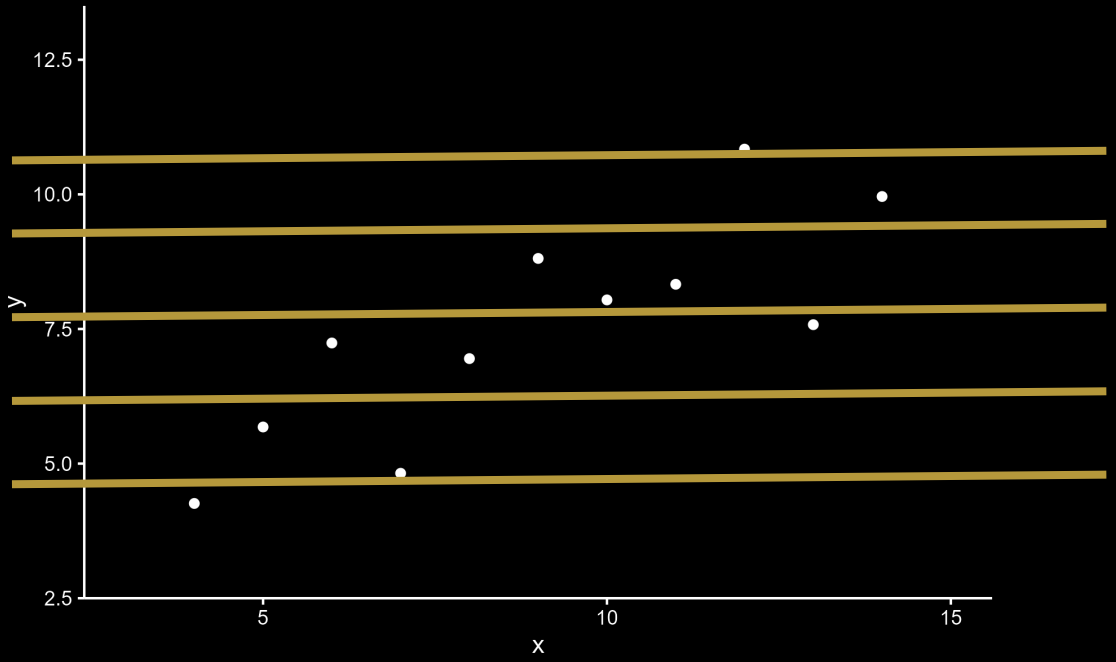
The three fundamental activities of statistics

- Describe: How strong is the relationship between Y and X?
- Decide: Is there a statistically significant relationship between Y and X?
- Predict: Given a particular value of X, what value of Y do we expect?

NHH
TECH3

Data = Model + Error

Y: Dependent variable



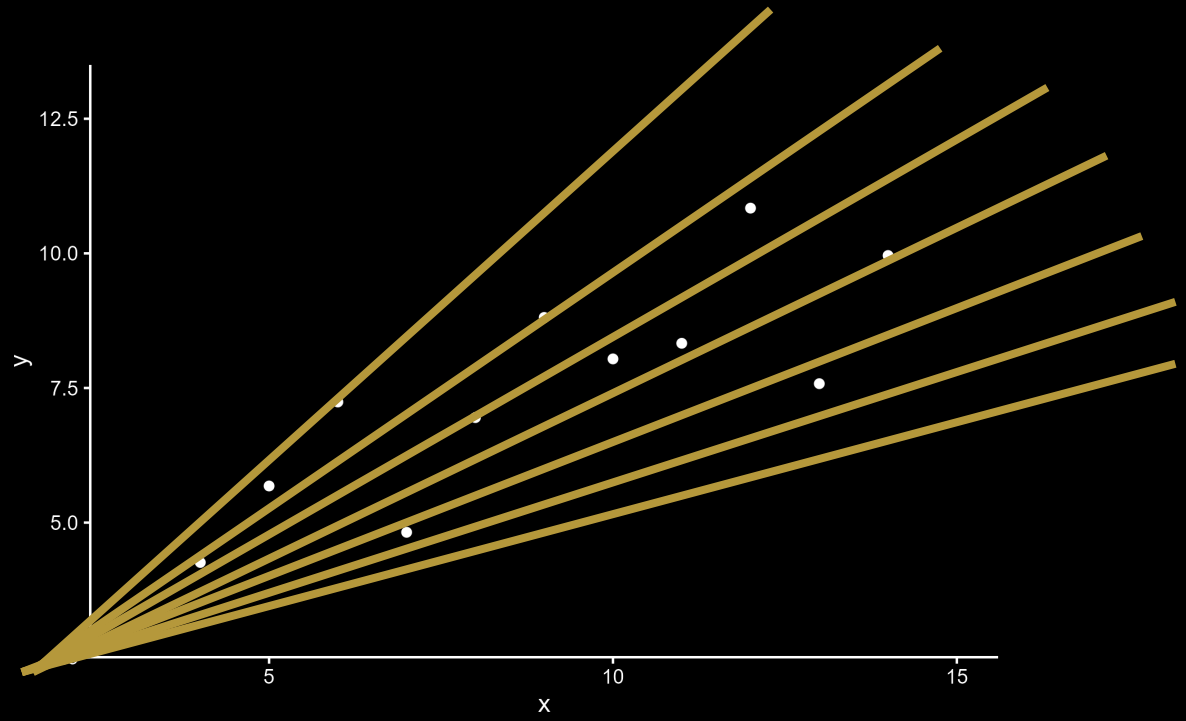
X: Independent variable

$$\hat{y}_i = \beta_0 + x_i \beta_1^{=0}$$

NHH
TECH3

Data = Model + Error

Y: Dependent variable

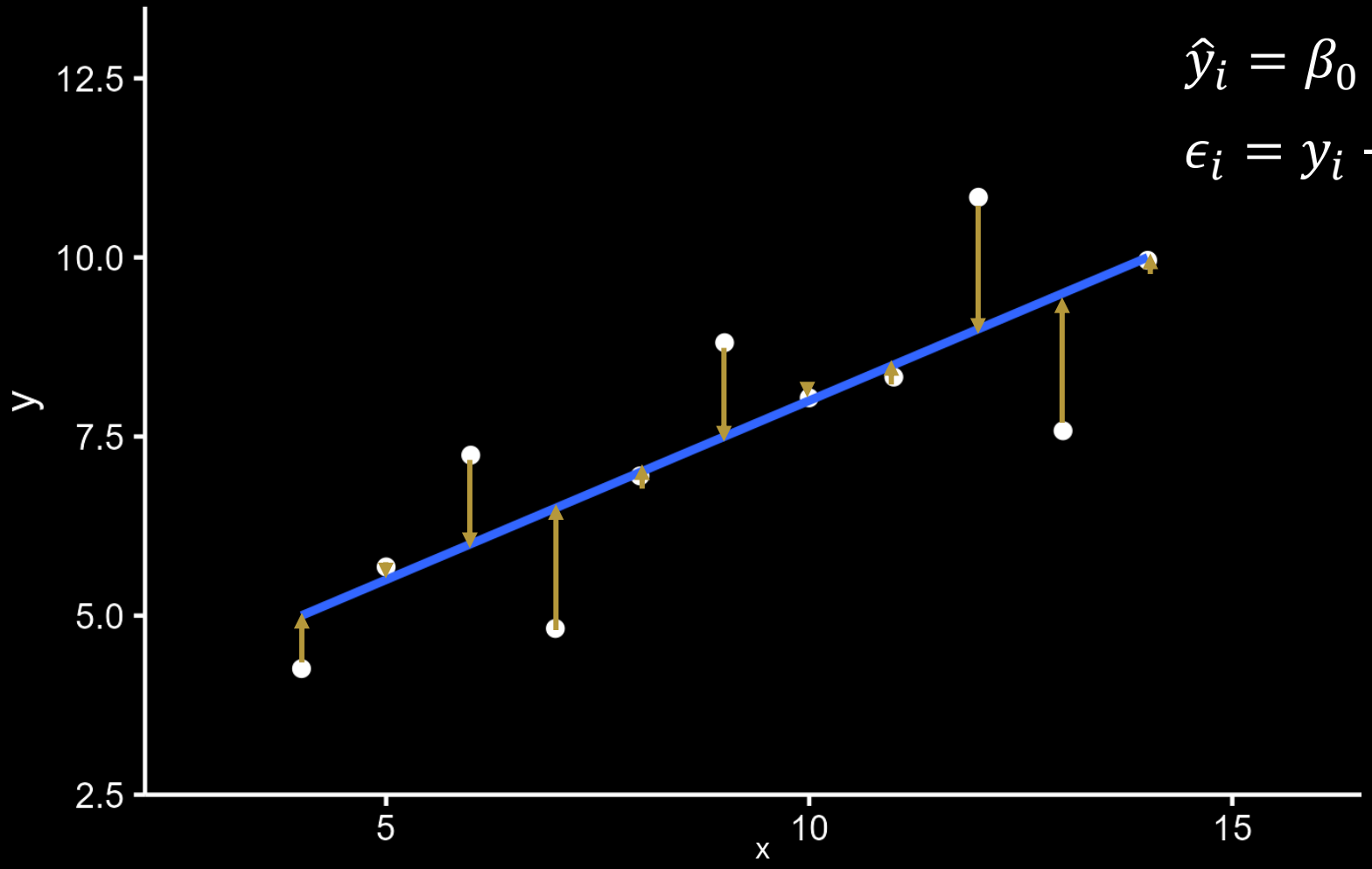


X: Independent variable

$$\hat{y}_i = \beta_0 + x_i \beta_1$$

=fixed

NHH
TECH3

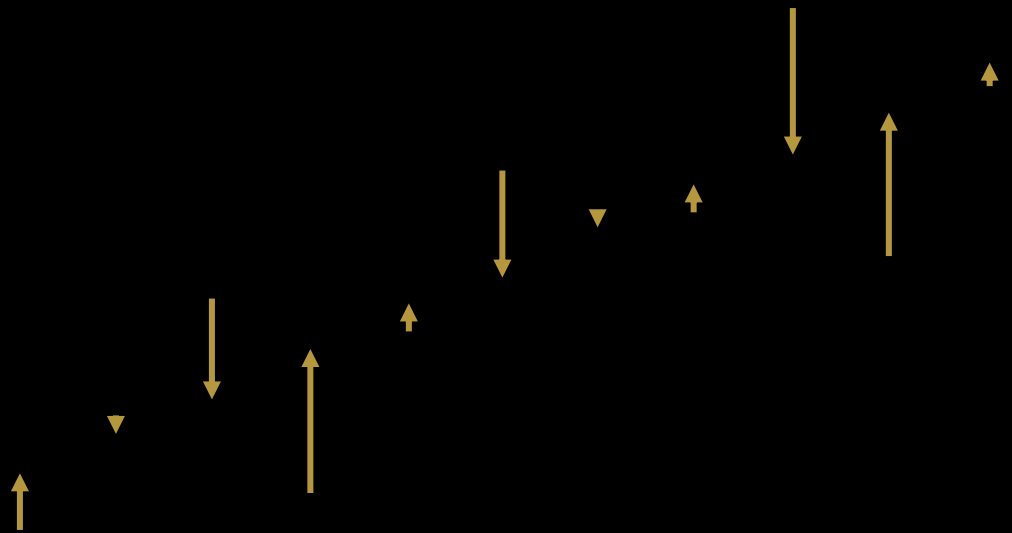


$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

$$\epsilon_i = y_i - \hat{y}_i$$

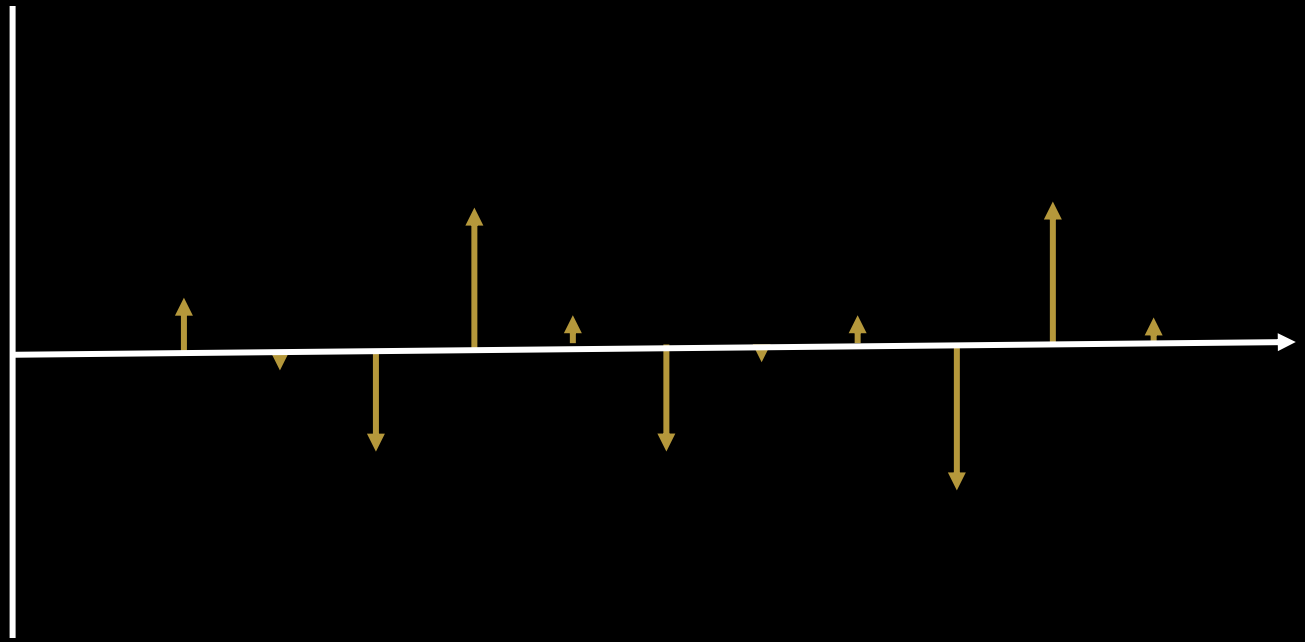
NHH
TECH3

$$\epsilon_i = y_i - \hat{y}_i$$

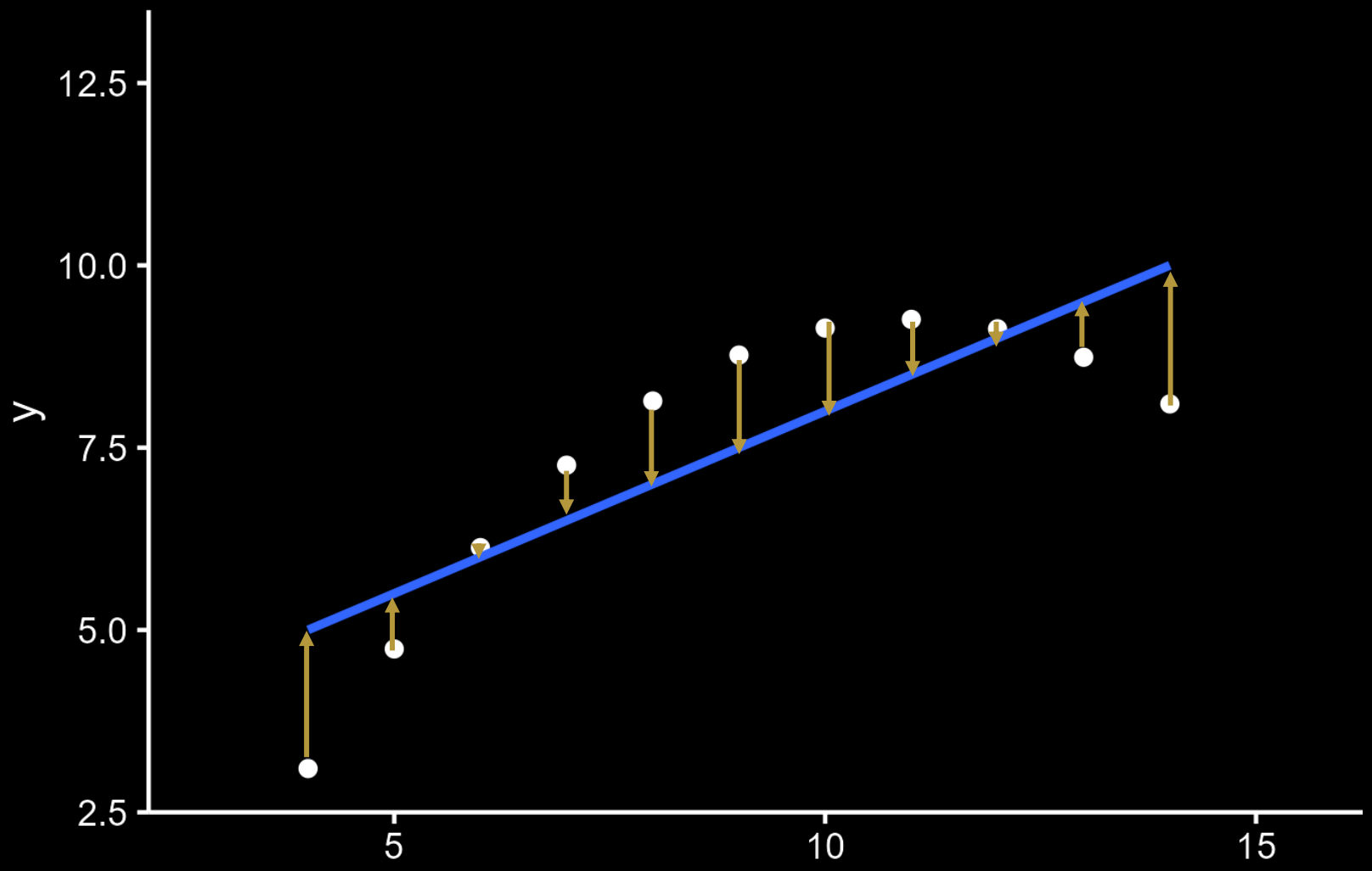


NHH
TECH3

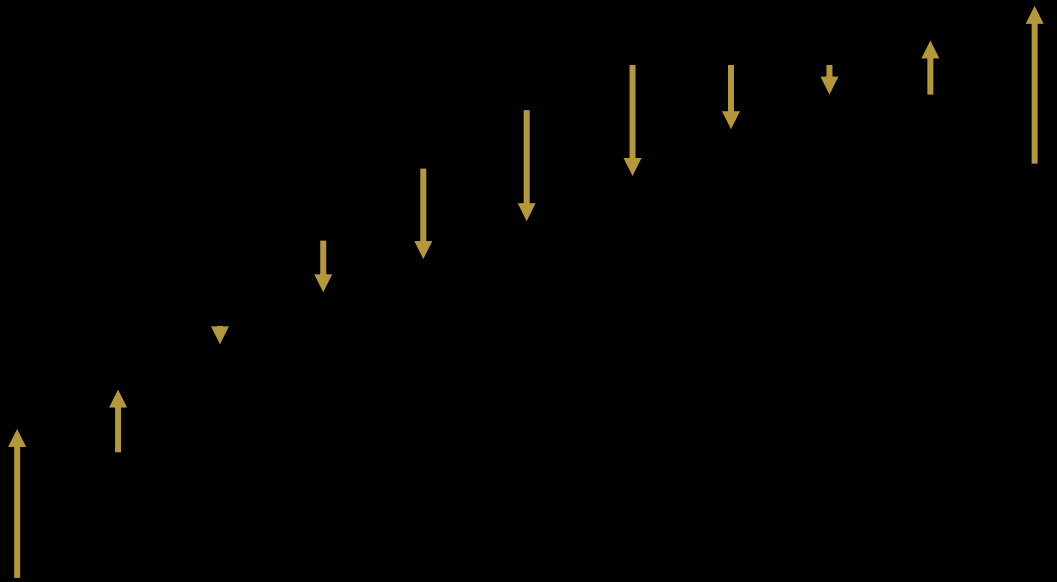
$$\epsilon_i = y_i - \hat{y}_i$$



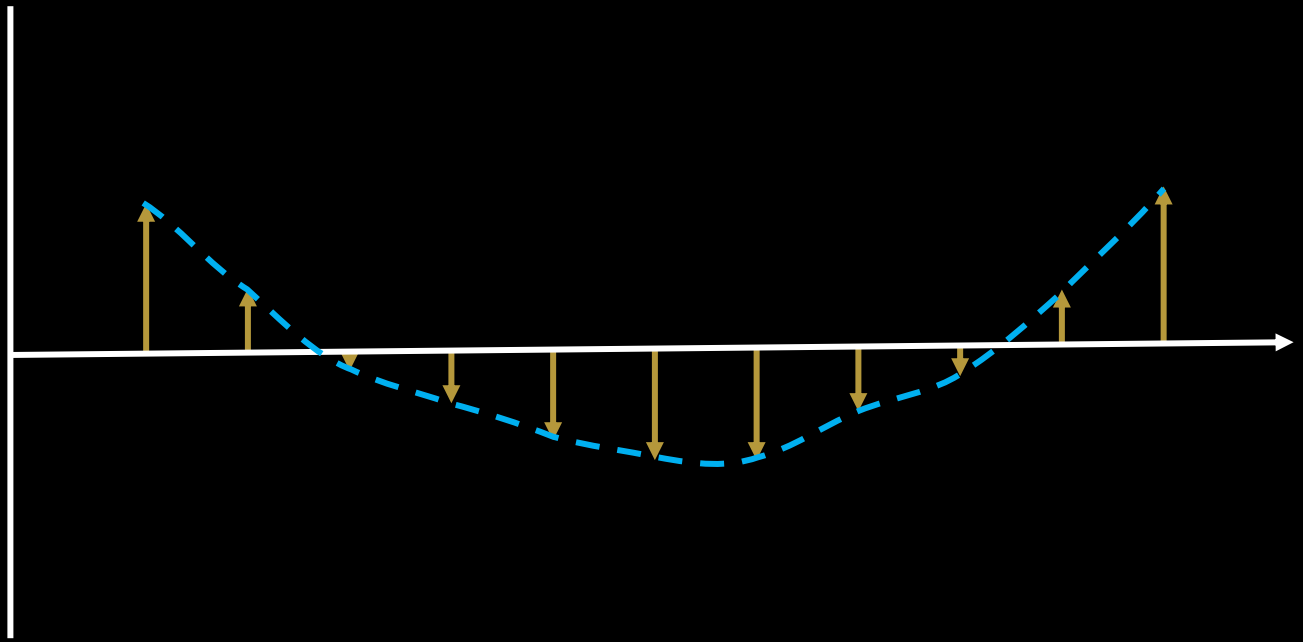
NHH
TECH3



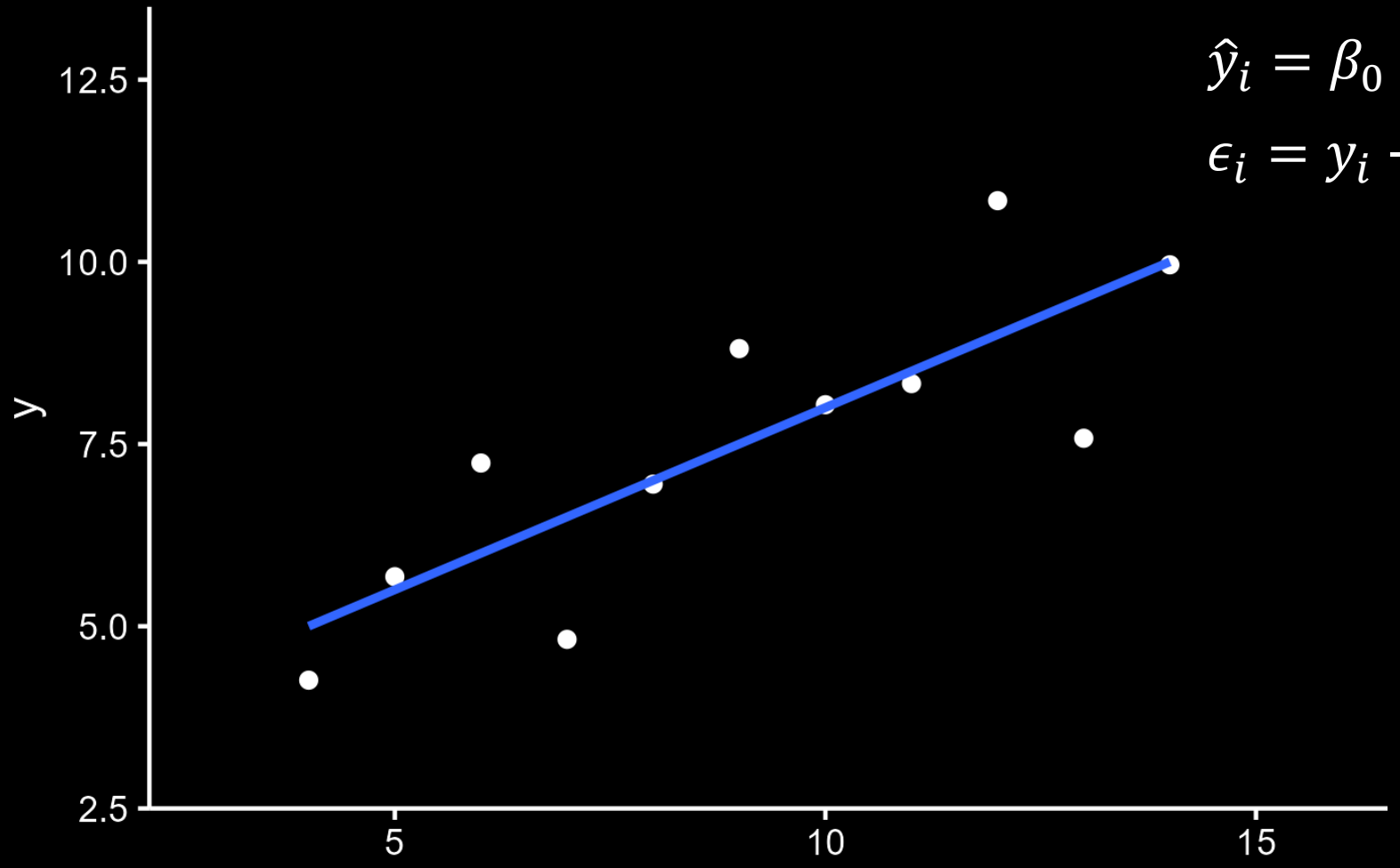
NHH
TECH3



NHH
TECH3



NHH
TECH3



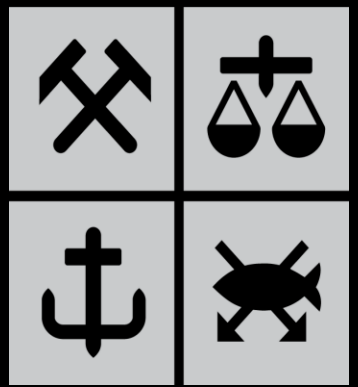
$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

$$\epsilon_i = y_i - \hat{y}_i$$

NHH
TECH3

NHH

TECH3



Sondre Hølleland
Geir Drage Berentsen