Regression

Md. Ismail Hossain Riday

Regression

- In the previous class, we covered the concept of "correlation," which helps us measure the relationship between two variables, denoted as "x" and "y”.

- A positive correlation implies that as "x" increases, "y" also increases, or vice versa.

- However, determining which variable is influencing the other is challenging.

Regression

- Relationship between two or more variables
- Cause and effect relationship
- Casual variables vs Affected variables
- Causal variables as independent variables; affected variables as dependent variables

Regression

- Relationship between two or more variables
- Cause and effect relationship
- Causal variables vs Affected variables
- Causal variables as independent variables; affected variables as dependent variables

Regression

$\boldsymbol{Dependent\ variable} = \boldsymbol{Y}$

$\boldsymbol{Independent\ variable} = \boldsymbol{X}$

- Regression refers to the cause-and-effect relationship between two or more variables, where affected variables are dependent variables and causal variables are independent variables.

- This relationship is mathematically expressed, highlighting the impact of independent variables on dependent variables.

Regression

- There is a positive relationship between income and expenditure, i.e. an increase in income increases expenditures.

- As increase in income causes an increase in expenditures

- “Income” as independent variable (X) and “Expenditures” as dependent variable (Y).

Correlation vs Regression

Correlation	Regression
1. Correlation is a statistical measure that determines the association between two variables.	1. Regression describe how to numerically relate an independent variable to the dependent variable.
2. There is no dependent variable and independent variable.	2. Must be one dependent variable and one independent variable.
3. To represent the linear relationship between variables.	3. To represent the cause and effect relationship between variables.

Types of Regression

- On the basis of variables

Types of Regression

- On the basis of variables (two types)

Types of Regression

- On the basis of variables (two types)

Simple regression

Types of Regression

- On the basis of variables (two types)

Simple regression

Multiple regression

Types of Regression

- On the basis of variables (two types)

Simple regression	• One dependent vs one independent
Multiple regression	

Types of Regression

- On the basis of variables (two types)

Simple regression

- One dependent vs one independent

Multiple regression

- One dependent vs multiple independent

Simple regression

- One dependent variable vs One independent variable
- Expressed this relationship as a mathematical form
- The mathematical form is,

$$Y_i = \alpha + \beta X_i + \epsilon_i \quad ; i = 1,2,3, \dots, n$$

Here,

$Y_i = Dependent\ variable$

$X_i = Independent\ variable$

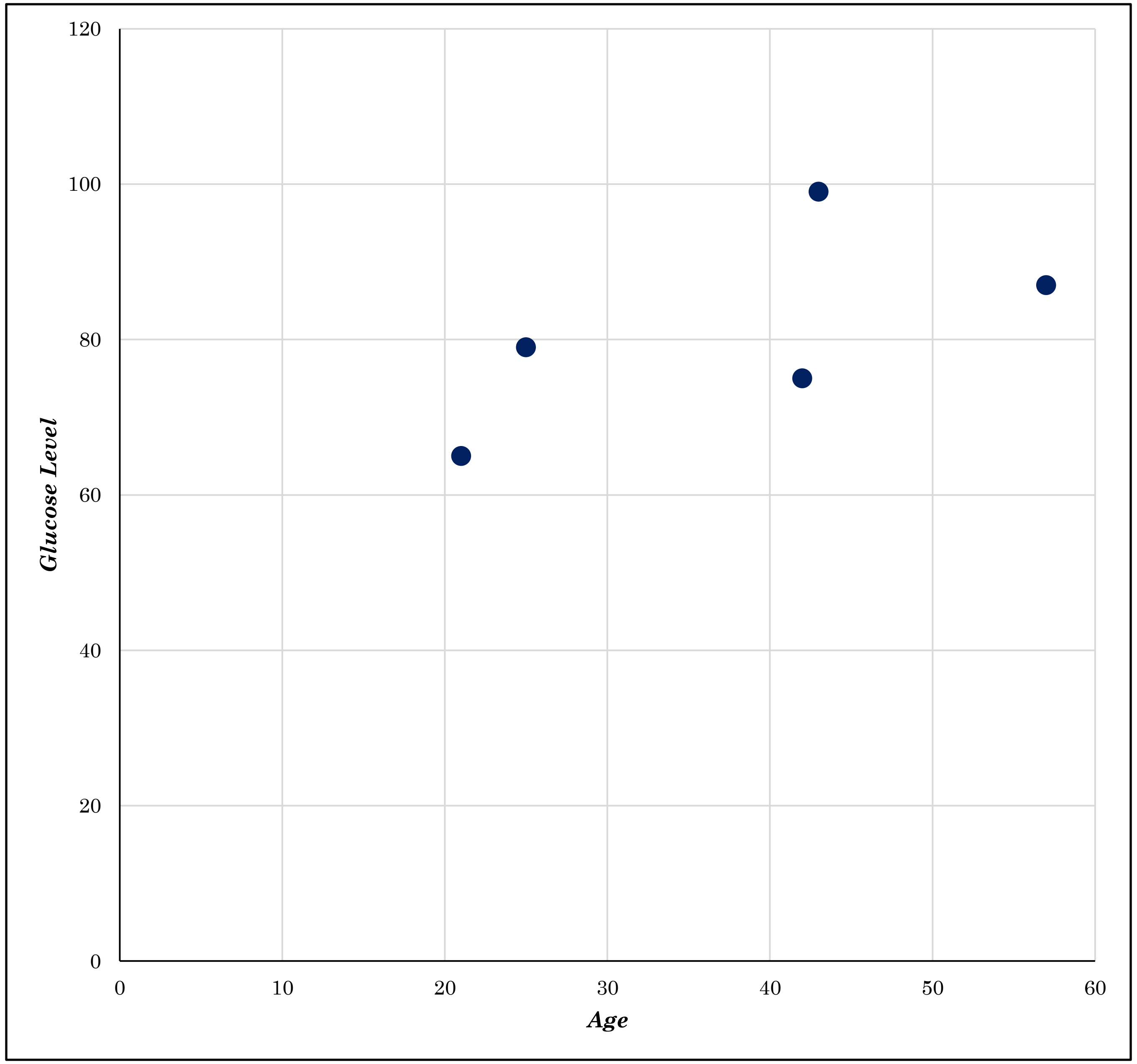
$\alpha = Intercept\ coefficient$

$\beta = Slope\ coefficient$

$\epsilon_i = Error\ term$

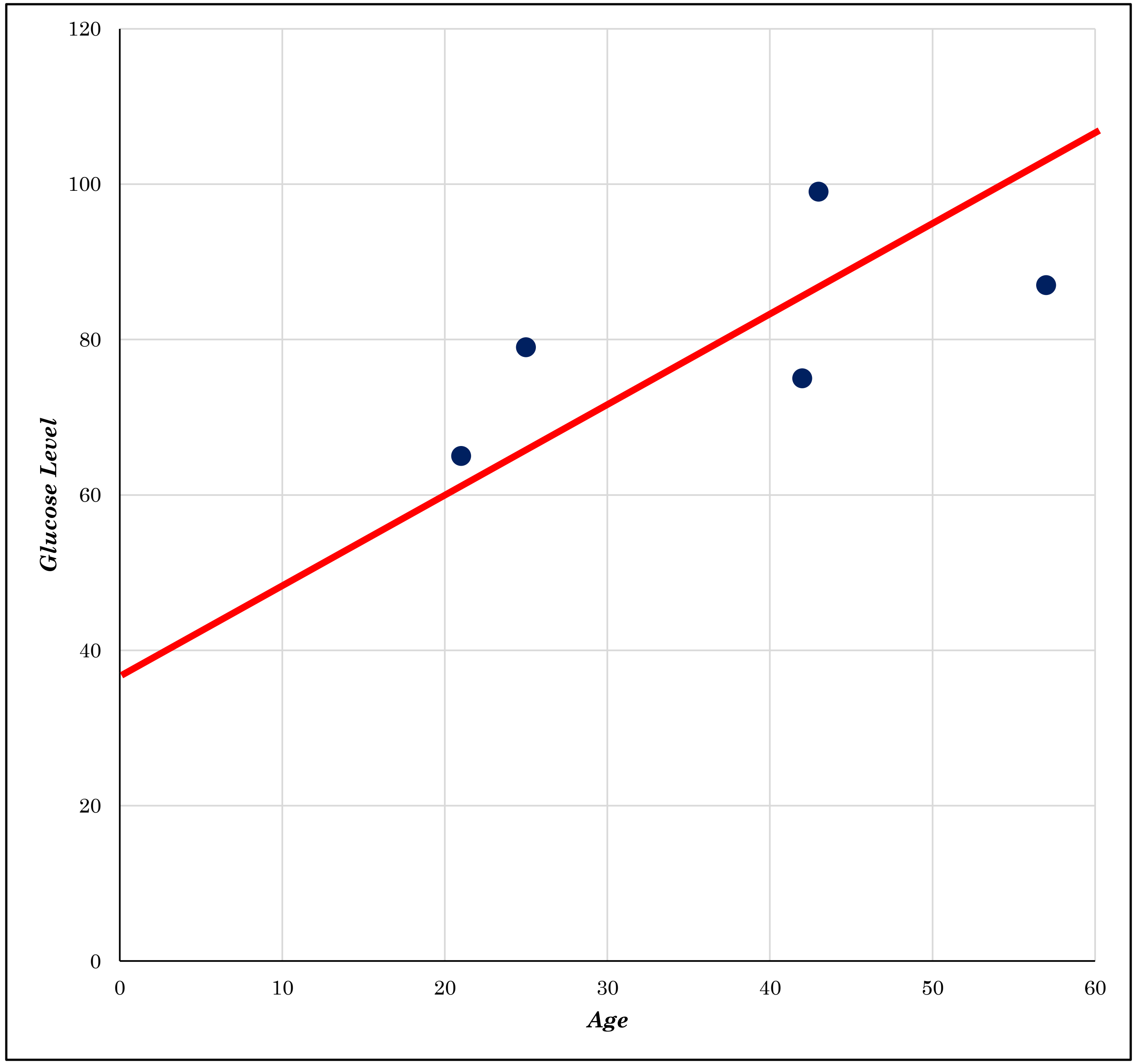
Multiple regression: $Y = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon$

Simple regression



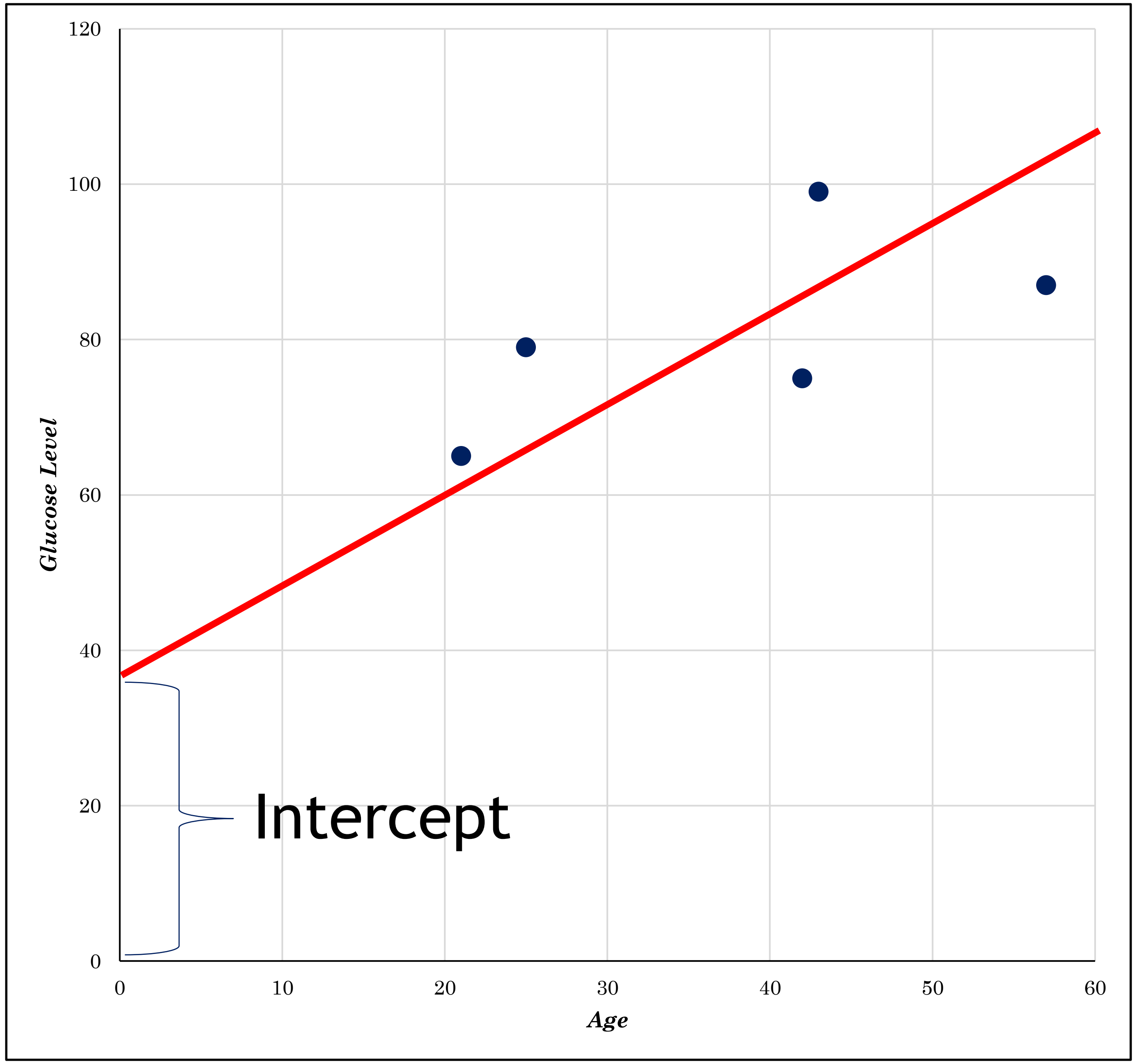
$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Simple regression



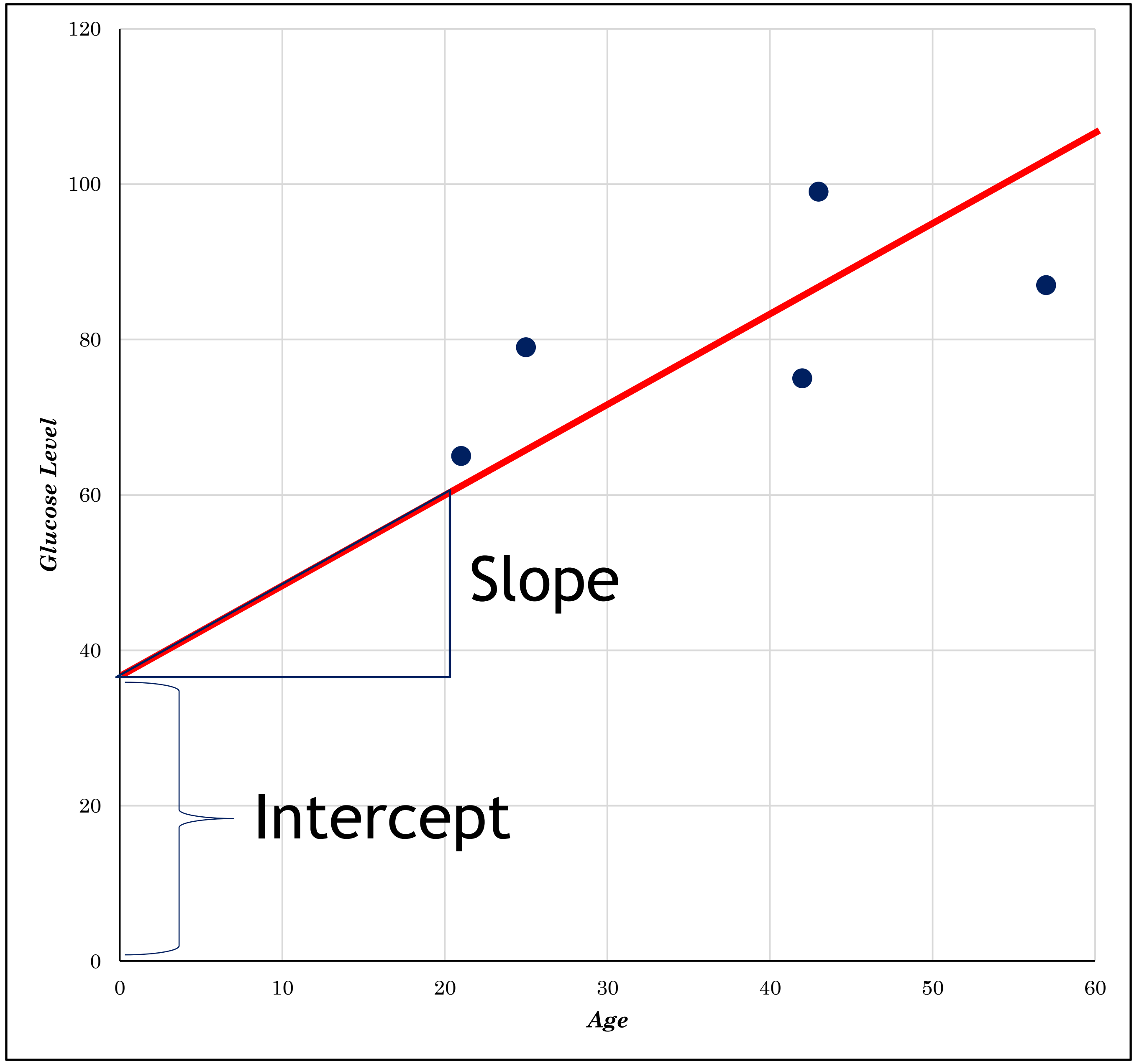
Simple regression

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



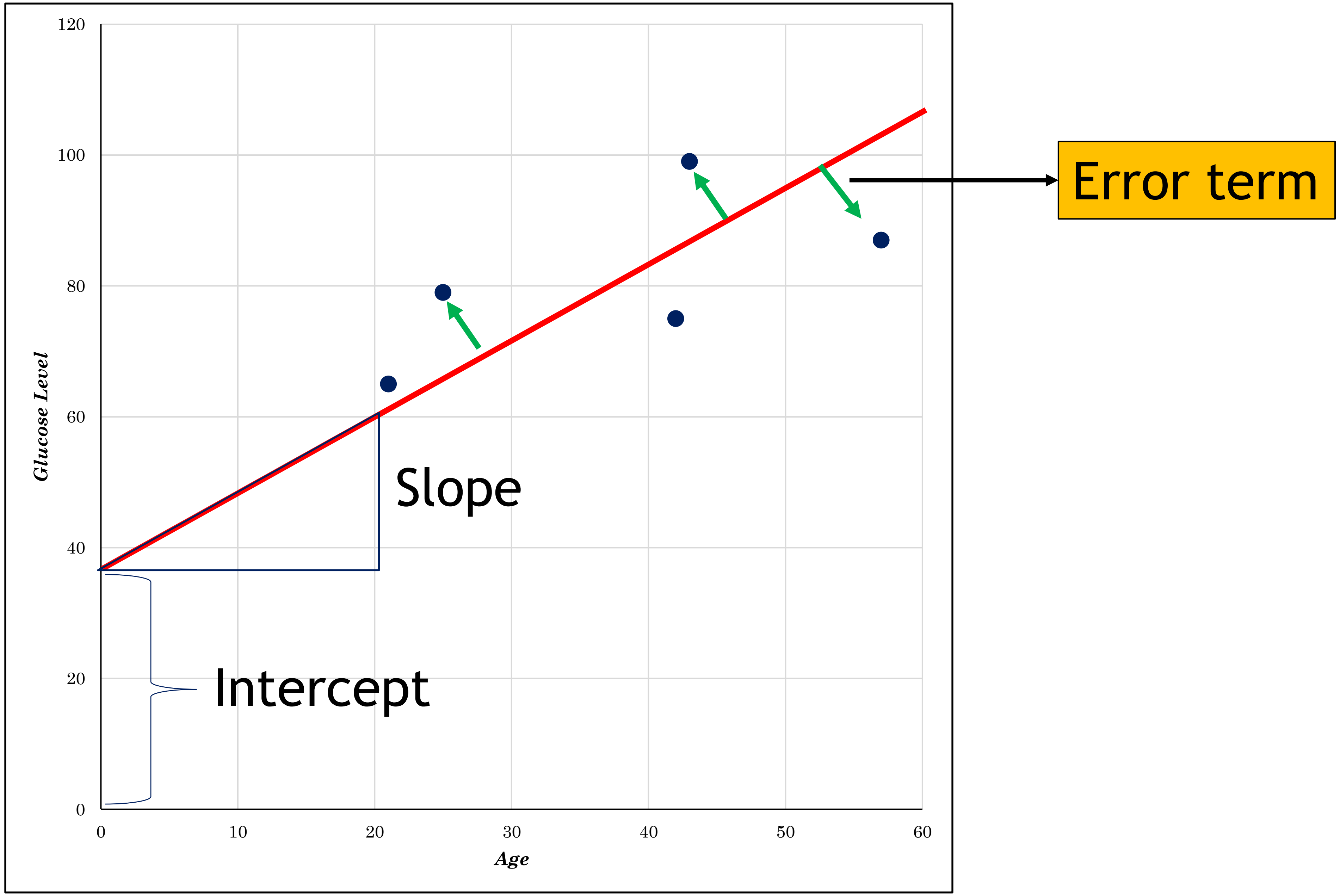
Simple regression

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



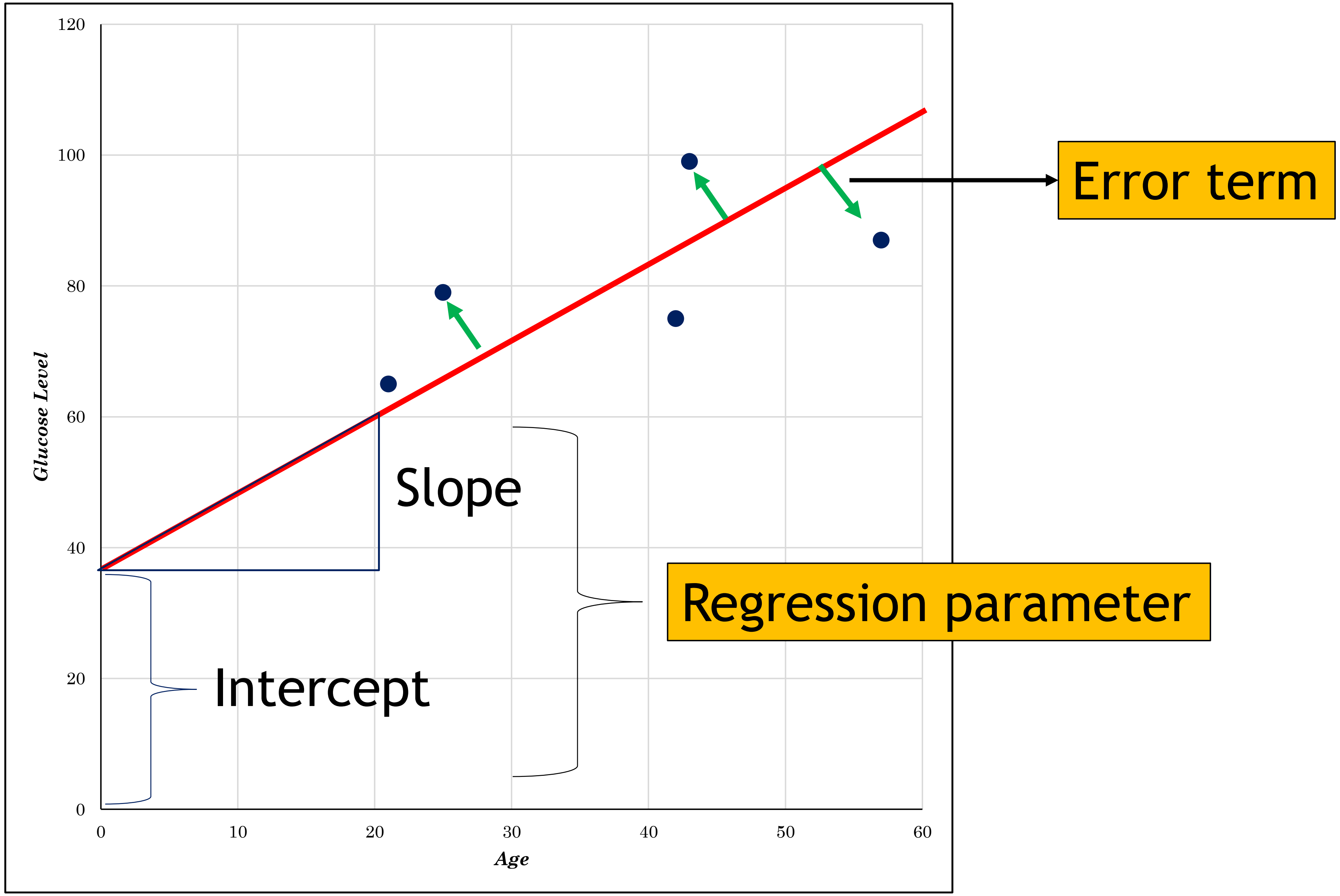
Simple regression

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



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

Simple regression

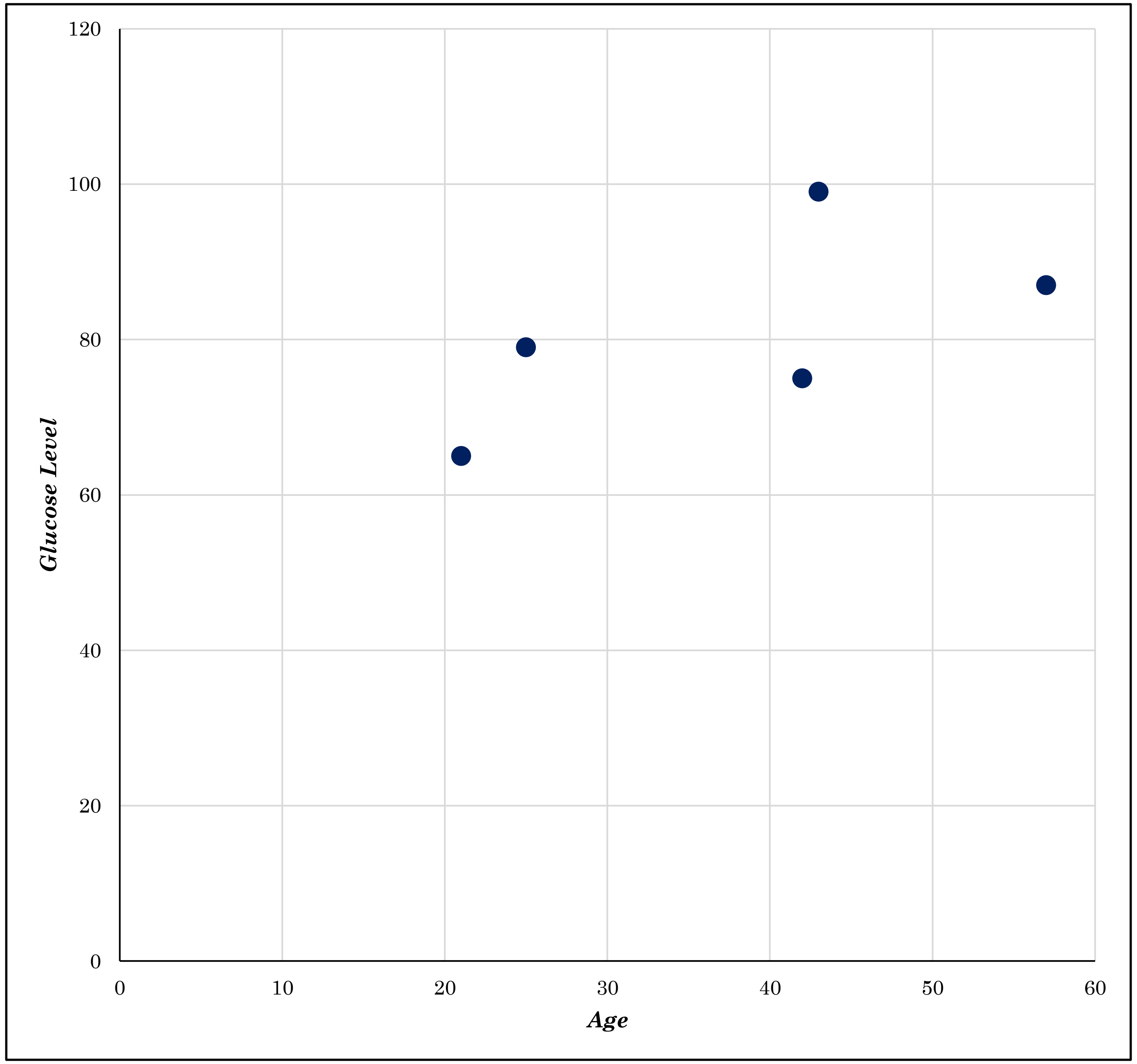


Estimating parameters

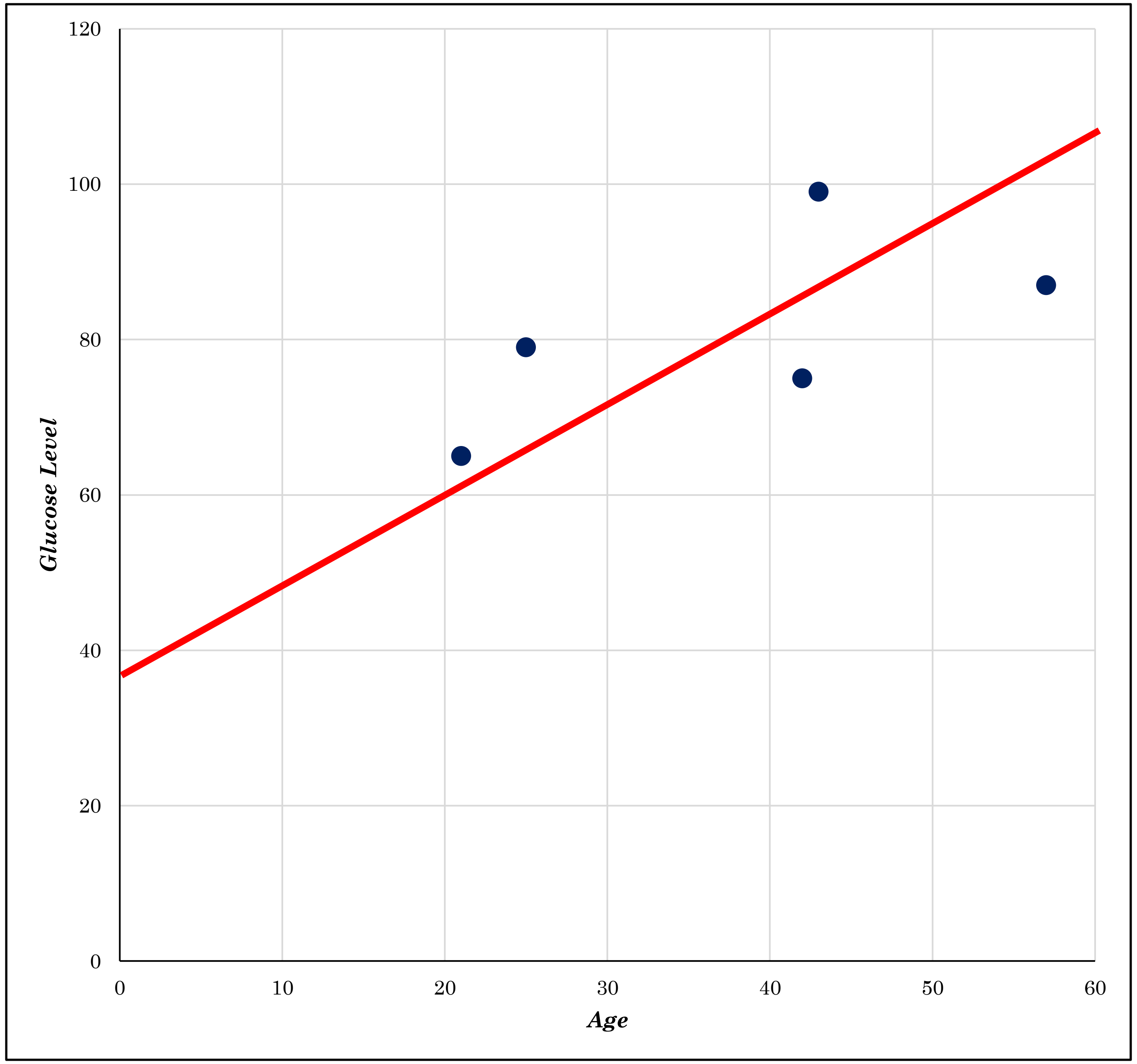
- Estimating regression parameters:

1. Least Square Method
2. Graphical Method

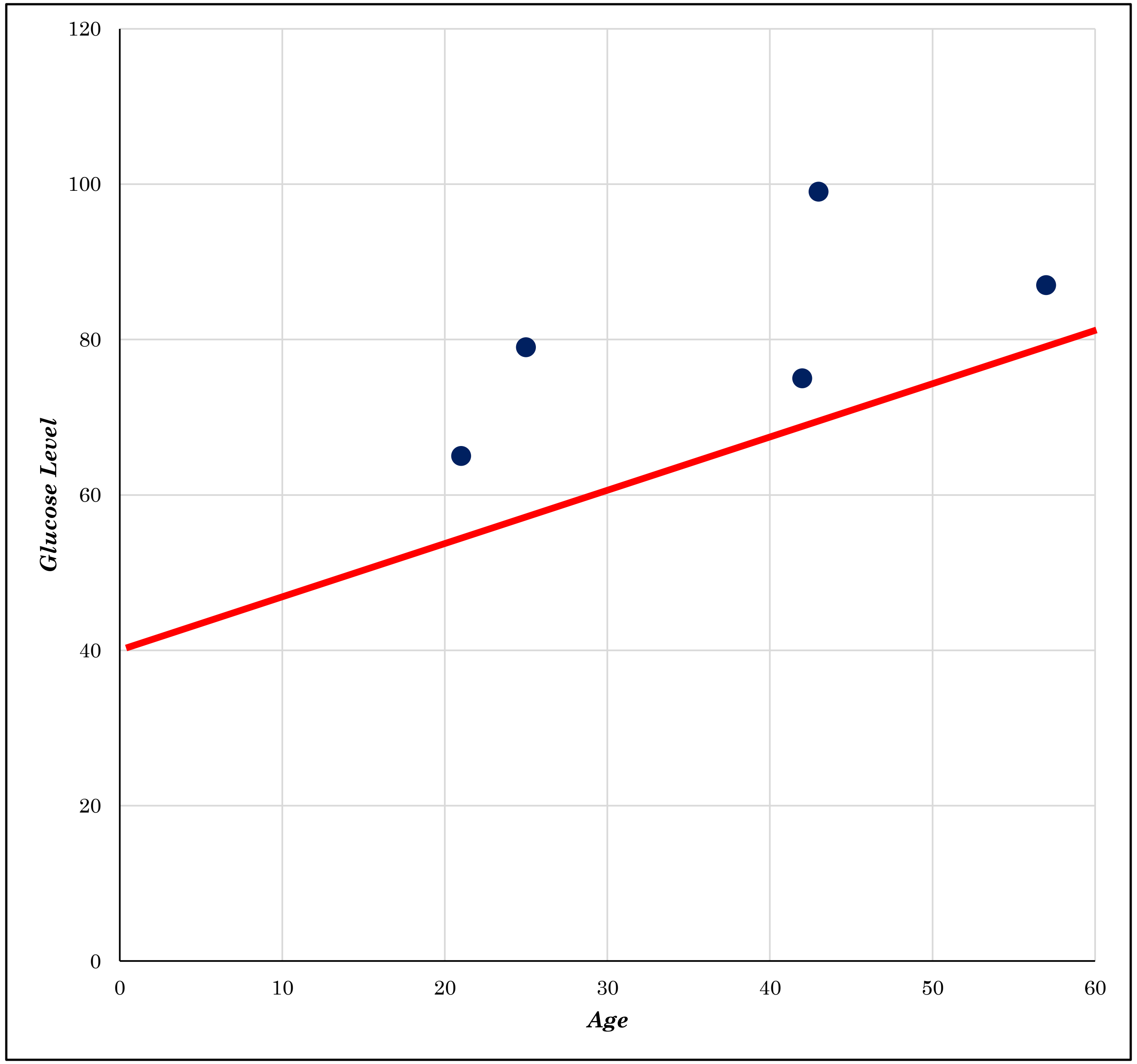
Estimating parameters



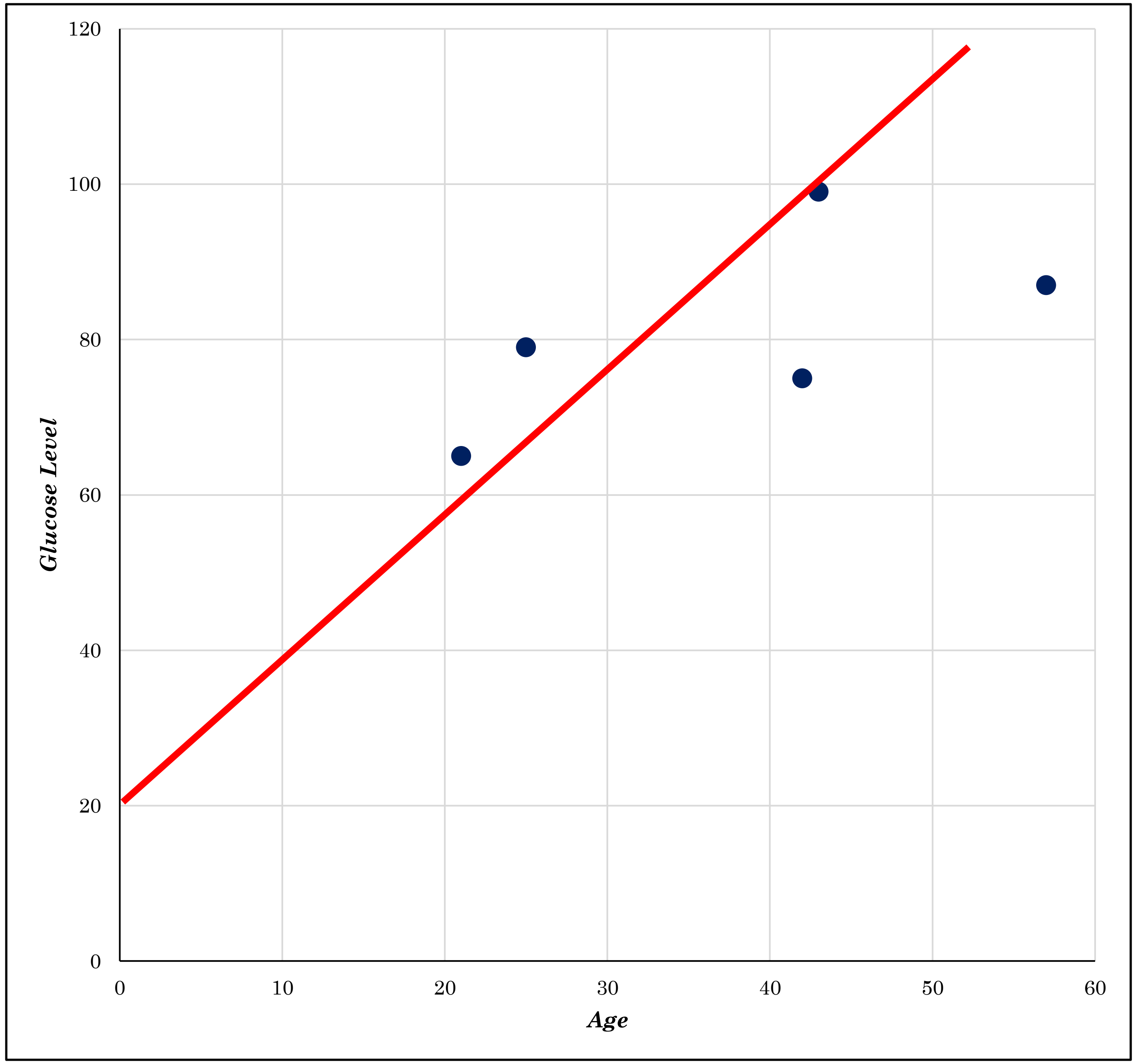
Estimating parameters



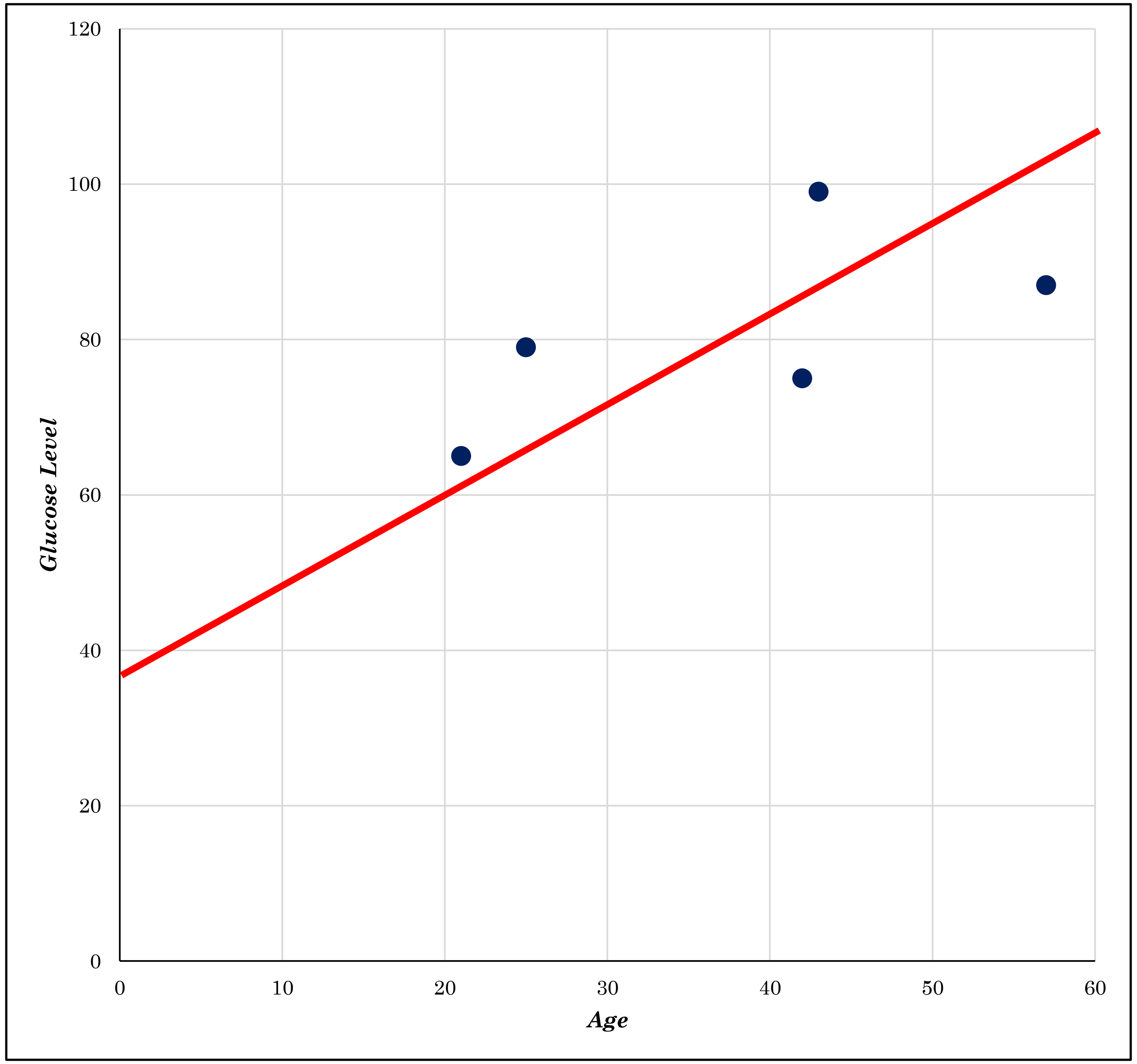
Estimating parameters



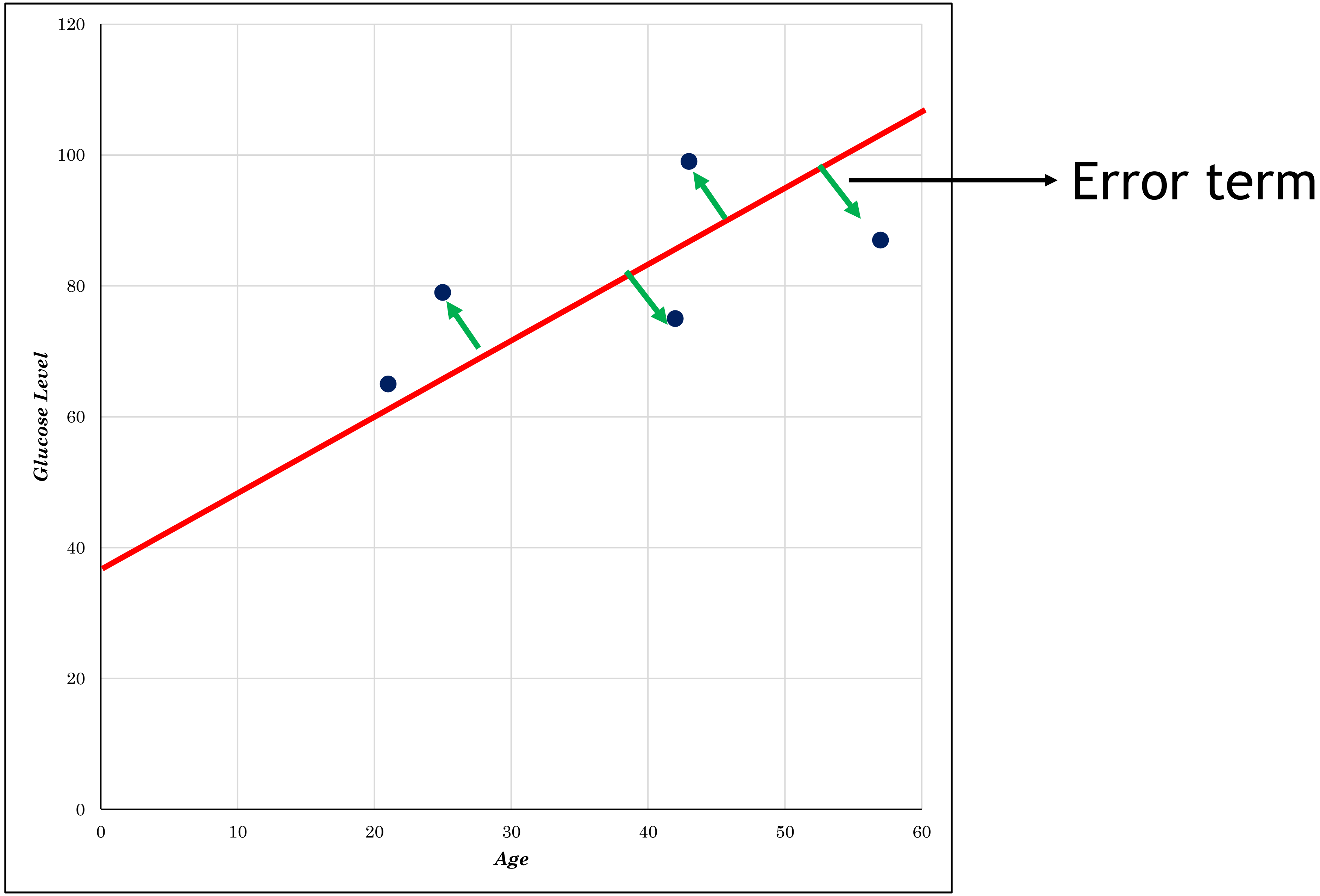
Estimating parameters



Estimating parameters



Estimating parameters



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

Estimating parameters

- Let the estimator of α and β is $\hat{\alpha}$ and $\hat{\beta}$.

- $\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$

- $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

Estimating parameters

- Let the estimator of α and β is $\hat{\alpha}$ and $\hat{\beta}$.
- $\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$
- $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$$\widehat{Y_i} = \hat{\alpha} + \hat{\beta}X_i$$

Estimating parameters

- Let the estimator of α and β is $\hat{\alpha}$ and $\hat{\beta}$.
- $\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$
- $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$$\widehat{Y_i} = \hat{\alpha} + \hat{\beta} X_i$$

Fitted model

Example

Advertising cost ($ million)	Sales revenue ($ million)
2	7
1	3
3	8
4	10

a) Draw the scatter diagram.

b) Determine the fitted regression model of sales revenue on advertising cost. $\widehat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$

c) Interpret the regression parameter.

d) Estimate sales revenue when advertising cost is $9.

Example

$$\hat{\beta} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

Advertising cost (x_i)	Sales revenue (y_i)	x_i^2	$x_i \times y_i$
2	7	4	14
1	3	1	3
3	8	9	24
4	10	16	40
$\sum x_i = 10$	$\sum y_i = 28$	$\sum x_i^2 = 30$	$\sum x_i y_i = 81$

Example

$$Fitted\ model:\ \widehat{Y}_i = 1.5 + 2.2X_i$$

$$If\ X = 9:\ \widehat{Y}_i = 1.5 + 2.2 \times 9 = 21.3$$

$$\hat{\beta} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{(4 \times 81) - (10 \times 28)}{(4 \times 30) - (10)^2} = 2.2$$

$\widehat{\boldsymbol{\beta}} = \mathbf{2.2}$ **means that an increase of \$1 million in advertising cost, the average sales revenue will increase \$2.2 million**

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 7 - (2.2 \times 2.5) = 1.5$$

$\widehat{\boldsymbol{\alpha}} = \mathbf{1.5}$ **means that, if there is no advertising cost then average sales revenue would be \$1.5 million**

Self practice

Investment	5	10	15	20	25
Profit	3	4	8	12	18

Estimate linear regression line of $\underbrace{\text{profit}}_{Y}$ on $\underbrace{\text{investment}}_{X}$

Goodness of Fit

$$If\ X = 9:\ \widehat{Y}_i = 1.5 + 2.2 \times 9 = 21.3$$

- How precise such predictions are?

- Is this regression equation is useful for prediction?

- To answer these questions, we need “Coefficient of Determination”- Denoted by $\boldsymbol{R^2}$

Coefficient of determination

- The coefficient of determination tells the percent of the variation in the dependent variable that is explained (determined) by the model and the explanatory variable. It is denoted by R^2.

Range of R^2: [0 *to* 1]

$$R^2 = 0: Equation\ is\ not\ useful\ for\ predictions$$

$$R^2 = 1: Equation\ is\ useful\ for\ predictions$$

Coefficient of determination

$$R^2 = 1 - \frac{SSE}{SST}$$

SSE = Error Sum of Square

$$SSE = \sum (Y_i - \hat{Y})^2$$

SST = Total Sum of Square

$$SST = \sum (Y_i - \bar{Y})^2$$

Coefficient of determination

$$\text{Fitted model: } \widehat{Y}_i = 1.5 + 2.2X_i$$

Advertising cost (x_i)	Sales revenue (y_i)	$\hat{y} = 1.5 + 2.2x_i$	$(y_i - \hat{y}_i)^2$	$(y_i - \bar{y})^2$
2	7	5.9	1.21	0
1	3	3.7	0.49	16
3	8	8.1	0.01	1
4	10	10.3	0.09	9
$\sum x_i = 10$	$\sum y_i = 28$		$SSE = 1.8$	$SST = 26$

$$SSE = \sum (Y_i - \hat{Y})^2$$

$$SST = \sum (Y_i - \bar{Y})^2$$

Coefficient of determination

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R^2 = 1 - \frac{1.8}{26}$$

$$R^2 = 0.93$$

Interpretation of $\boldsymbol{R^2}$: For example, $R^2 = 0.93$ or 93%. It indicates that, almost 93% of the variability of the dependent variables explained by the independent variables.

Error Calculation

$$Fitted\ model\colon \widehat{Y}_i = 1.5 + 2.2X_i$$

Advertising cost (x_i)	Sales revenue (y_i)	$\hat{y} = 1.5 + 2.2x_i$	$\epsilon_i = (y_i - \hat{y}_i)$
2	7	5.9	1.1
1	3	3.7	-0.7
3	8	8.1	-0.1
4	10	10.3	-0.3

Properties of Regression Coefficients

1. Regression coefficient has unit.

2. Regression coefficients are not symmetric function, $i.e., \beta_{yx} \neq \beta_{xy}$

Thank You