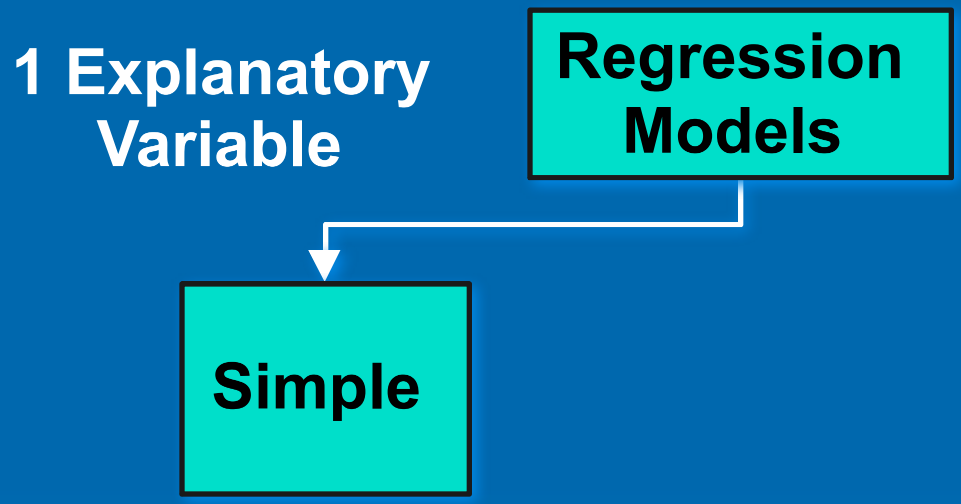
Regression

Types of Regression Models

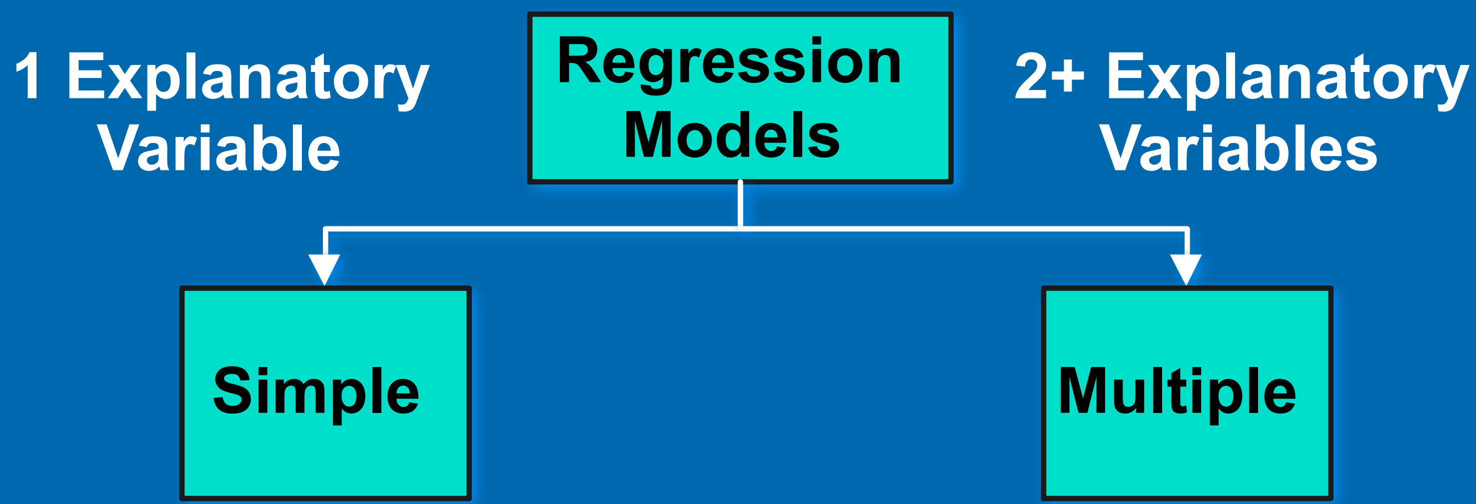
Types of Regression Models

Regression Models

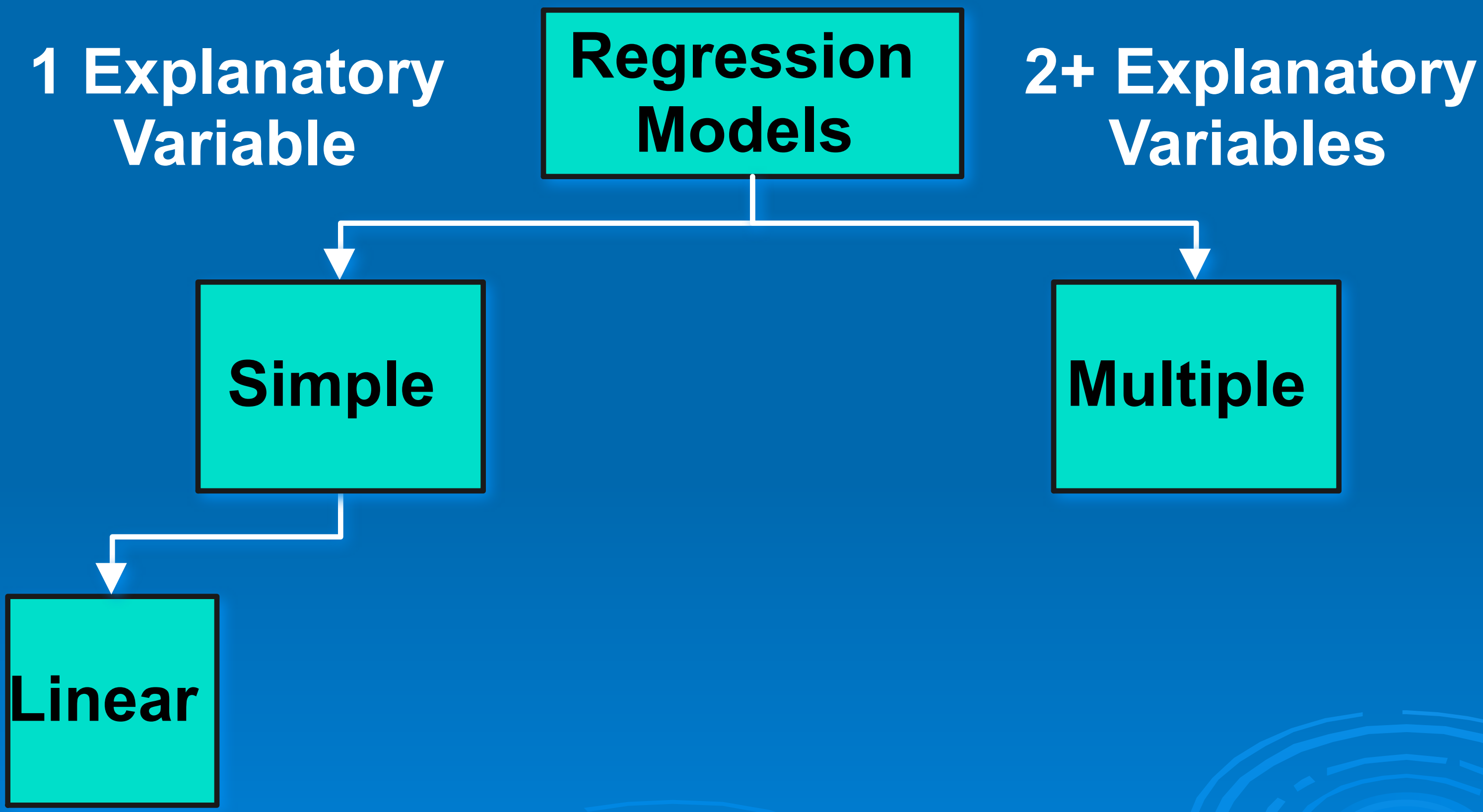
Types of Regression Models



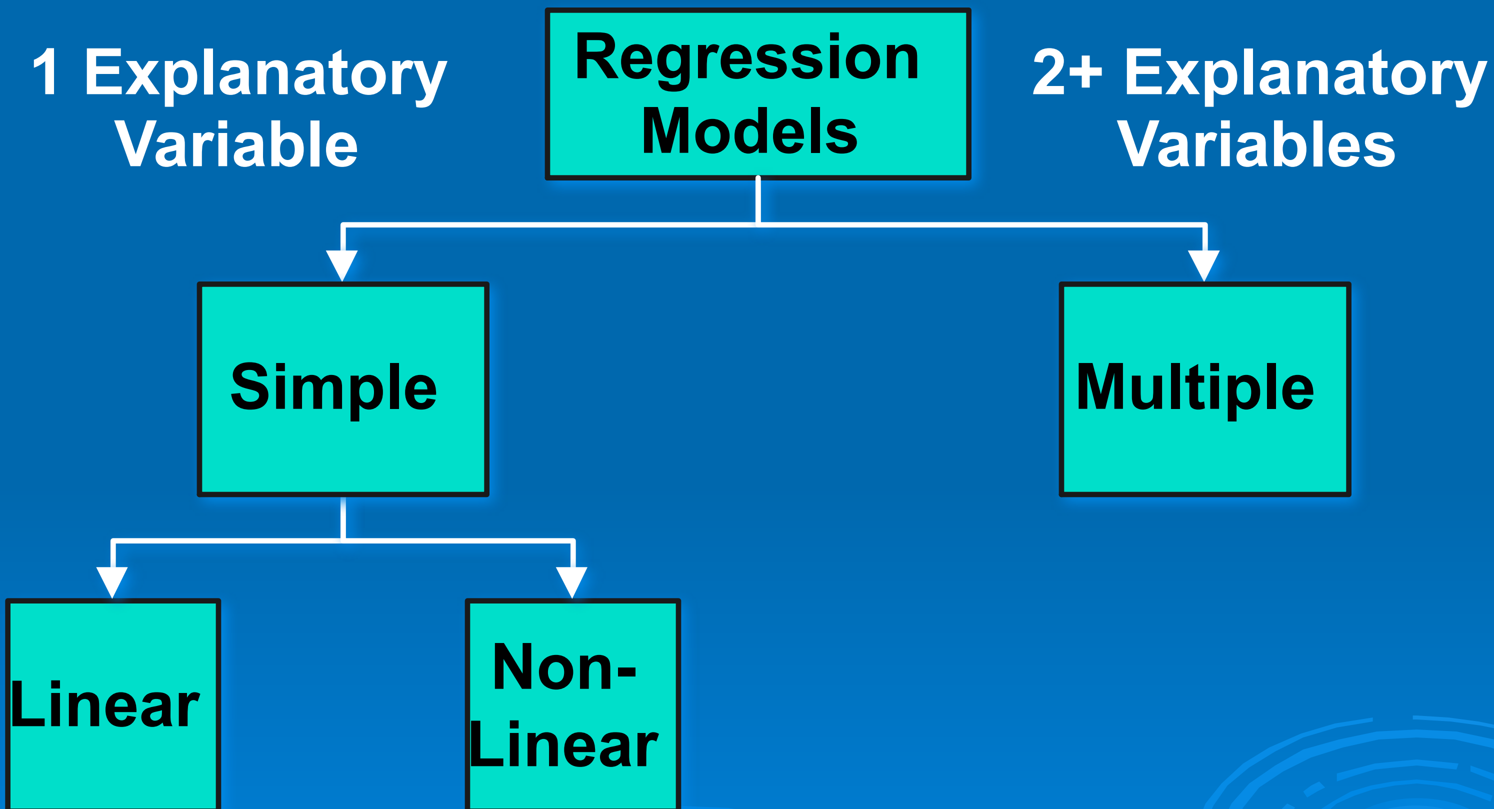
Types of Regression Models



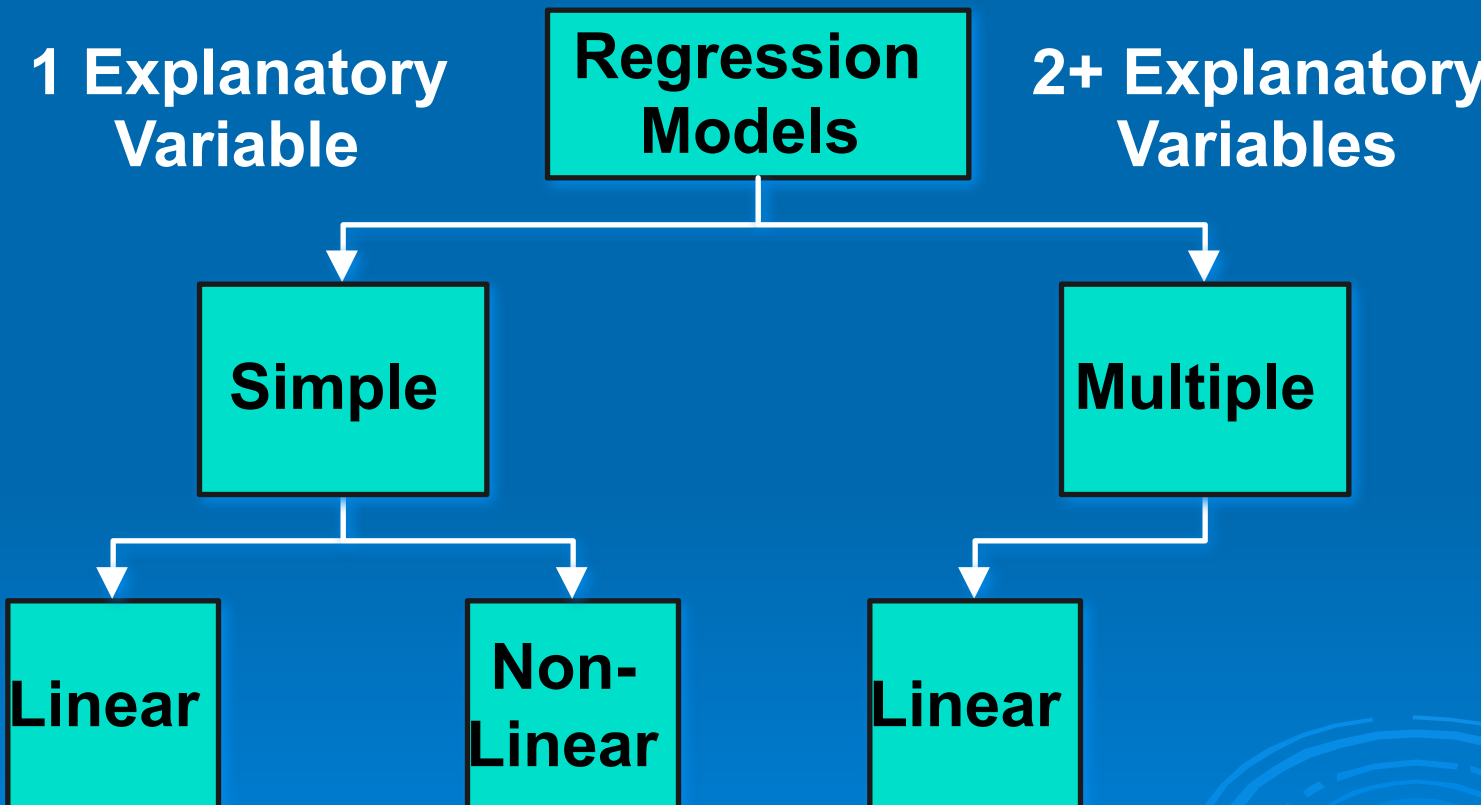
Types of Regression Models



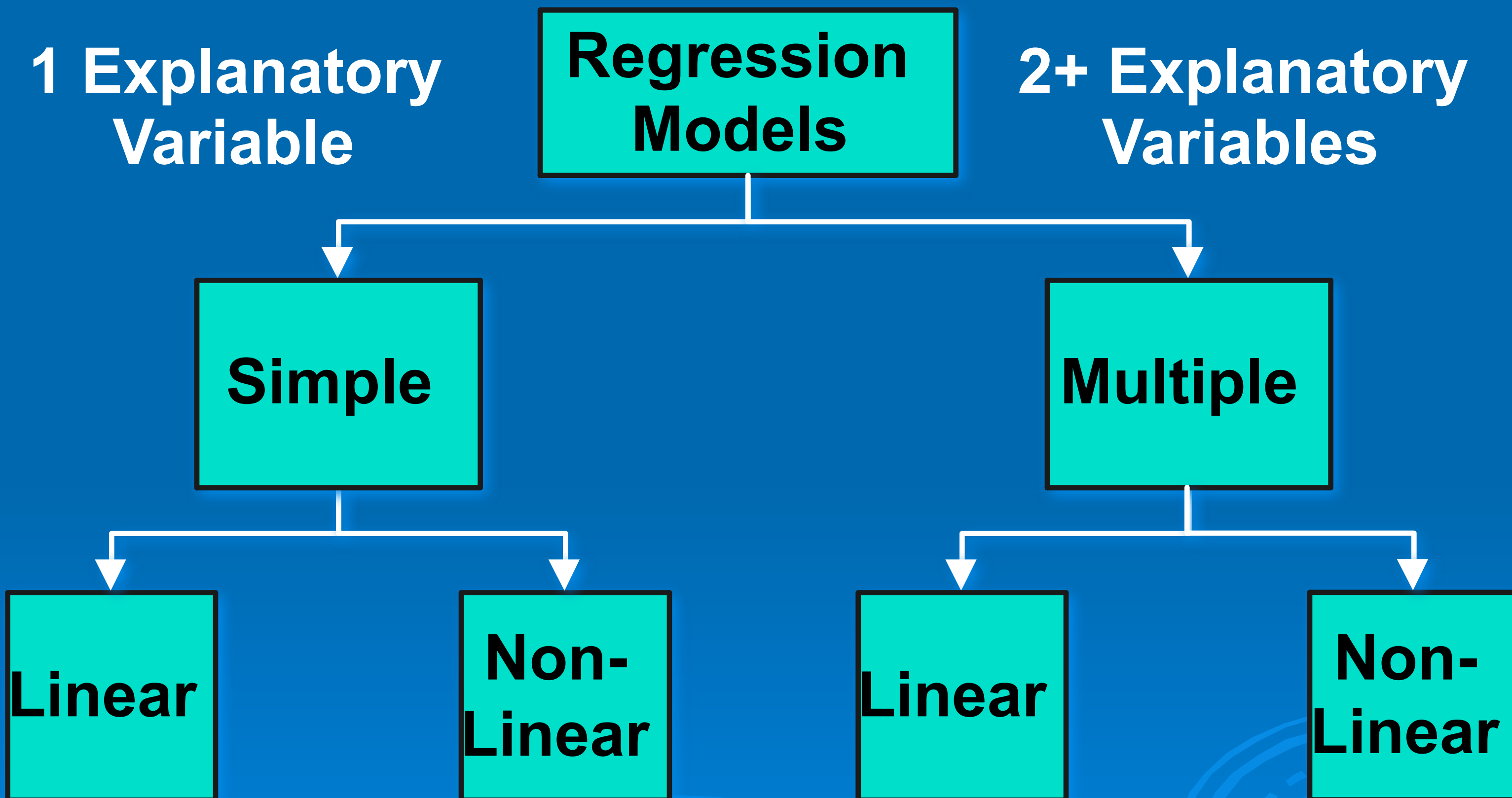
Types of Regression Models



Types of Regression Models

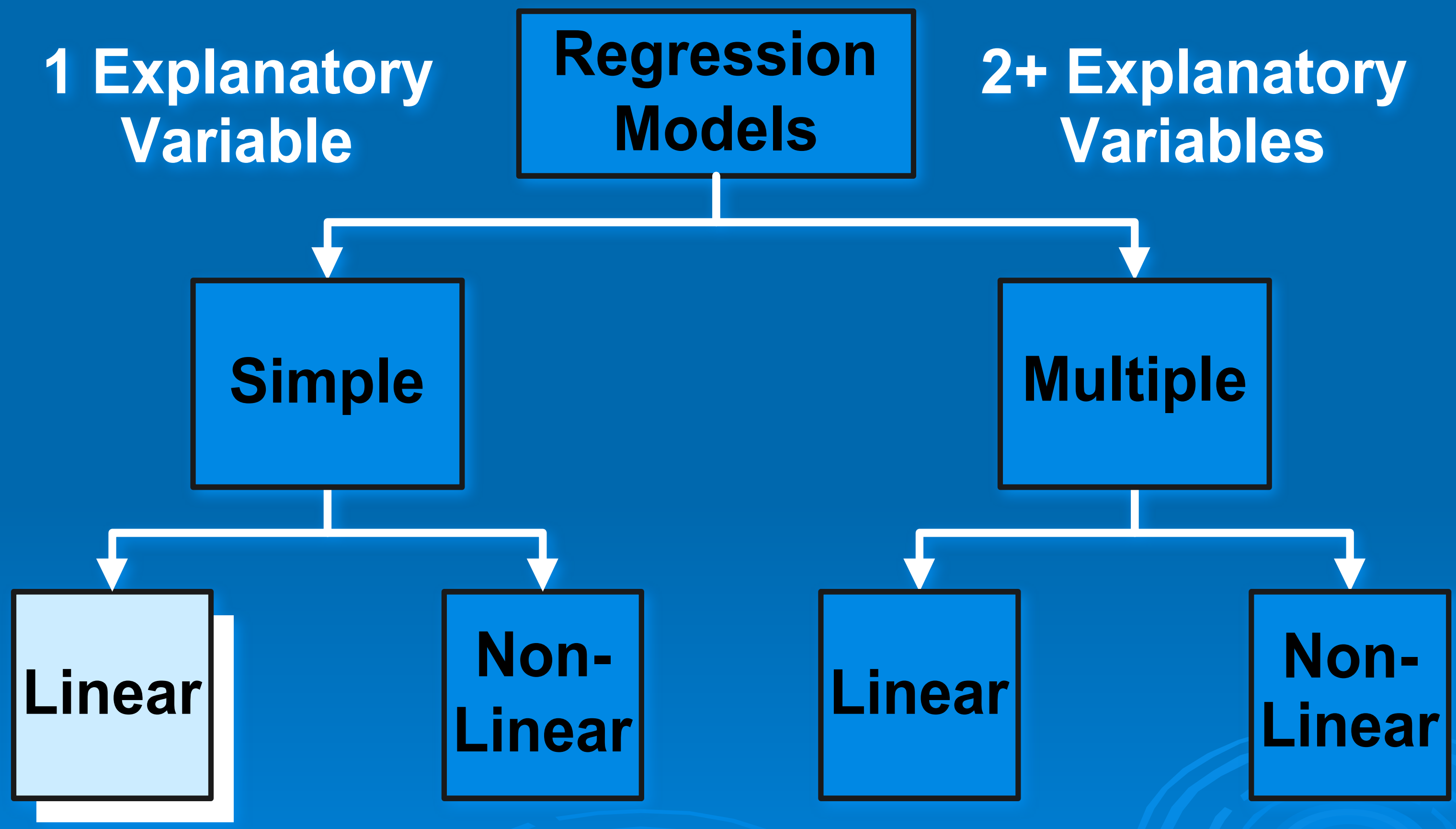


Types of Regression Models

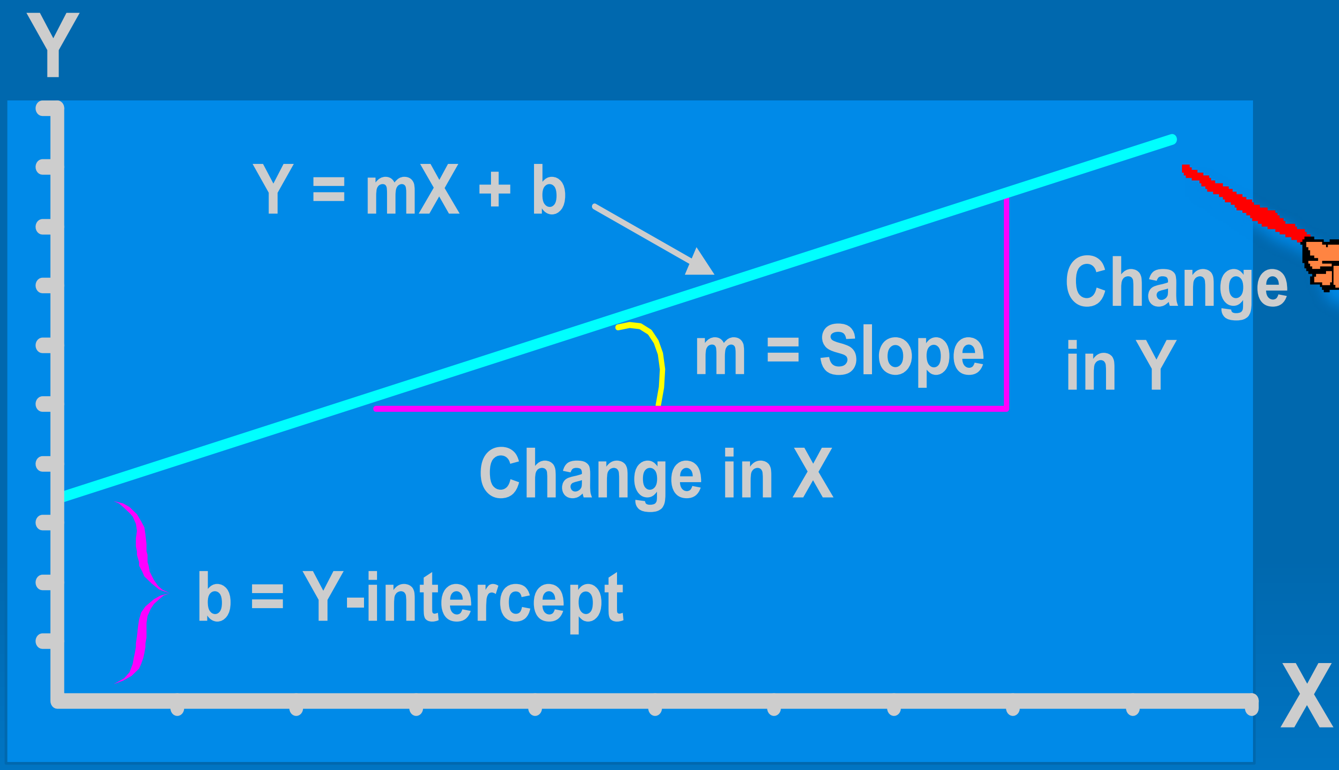


Linear Regression Model

Types of Regression Models



Linear Equations



© 1984-1994 T/Maker Co.

Linear Regression Model

- 1. Relationship Between Variables Is a Linear Function

Population Y-Intercept

Population Slope

Random Error

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

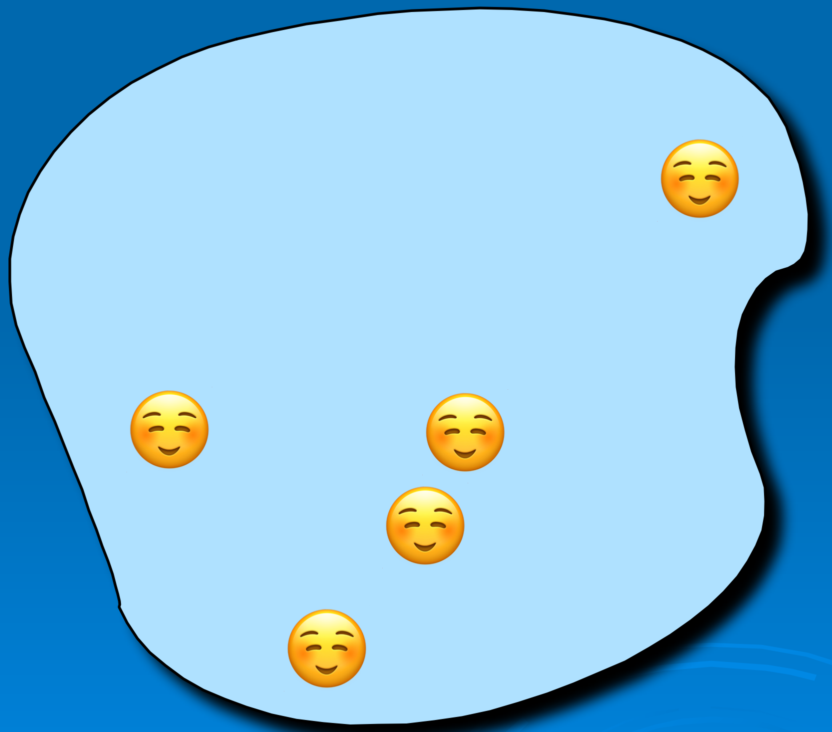
Dependent (Response) Variable (e.g., CD+ c.)

Independent (Explanatory) Variable (e.g., Years s. serocon.)

Population & Sample Regression Models

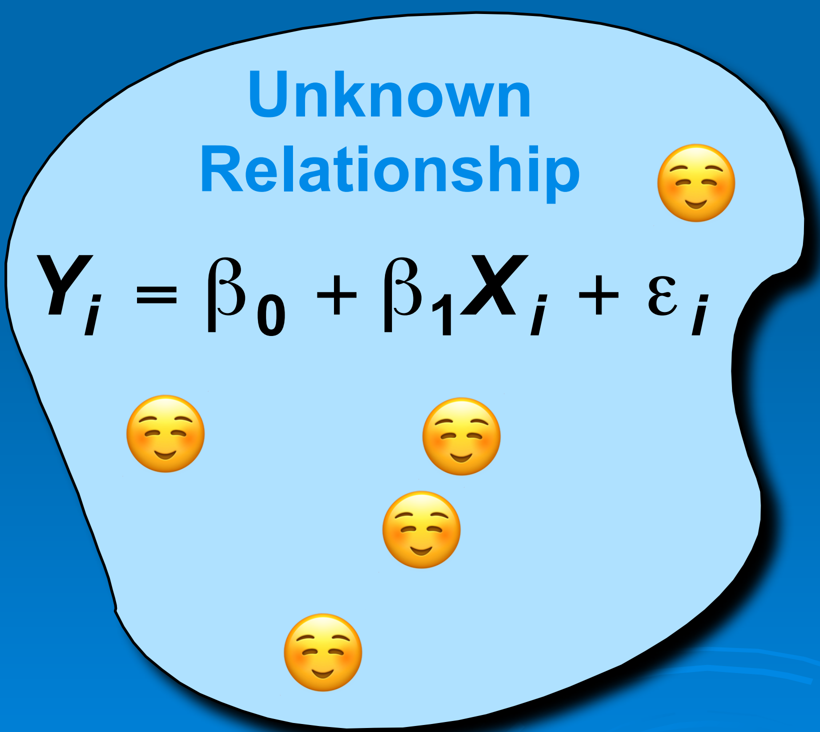
Population & Sample Regression Models

Population



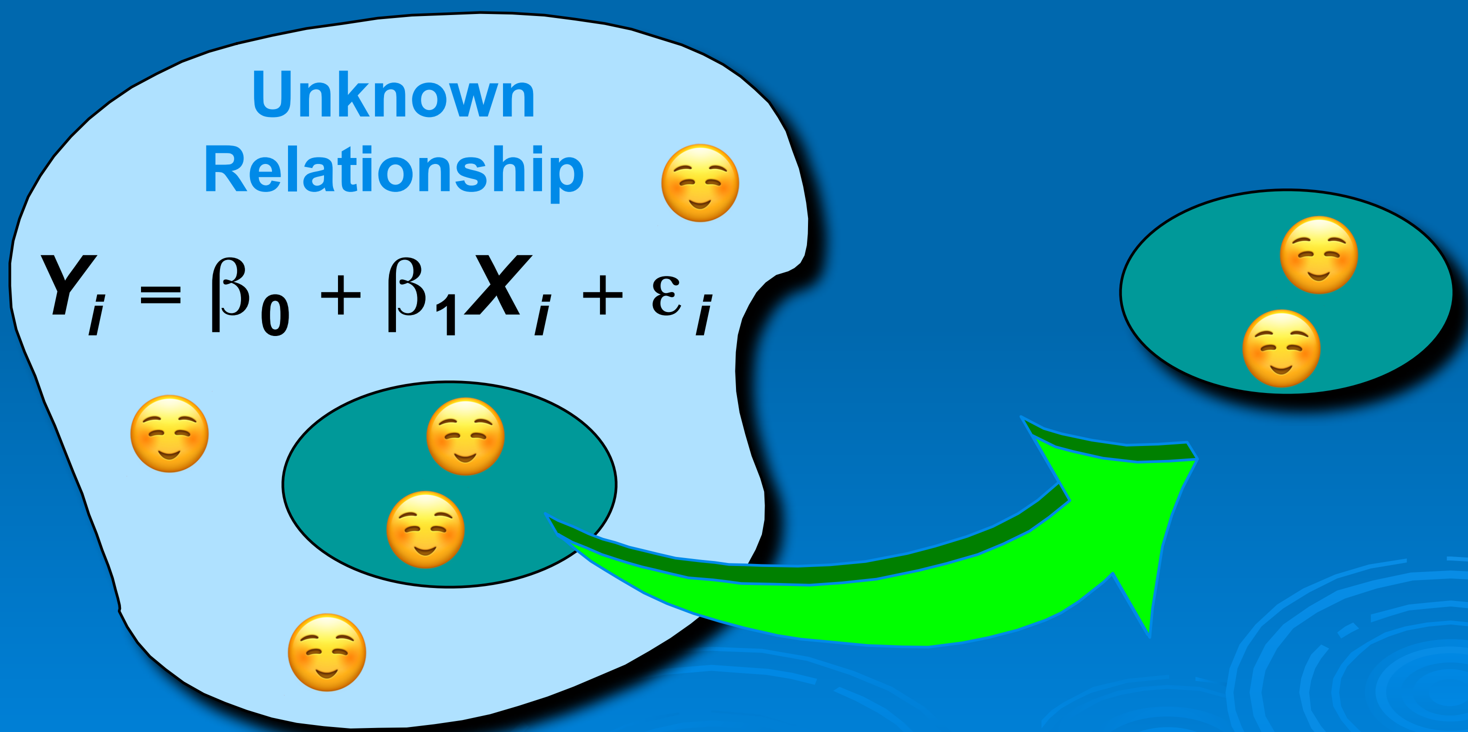
Population & Sample Regression Models

Population

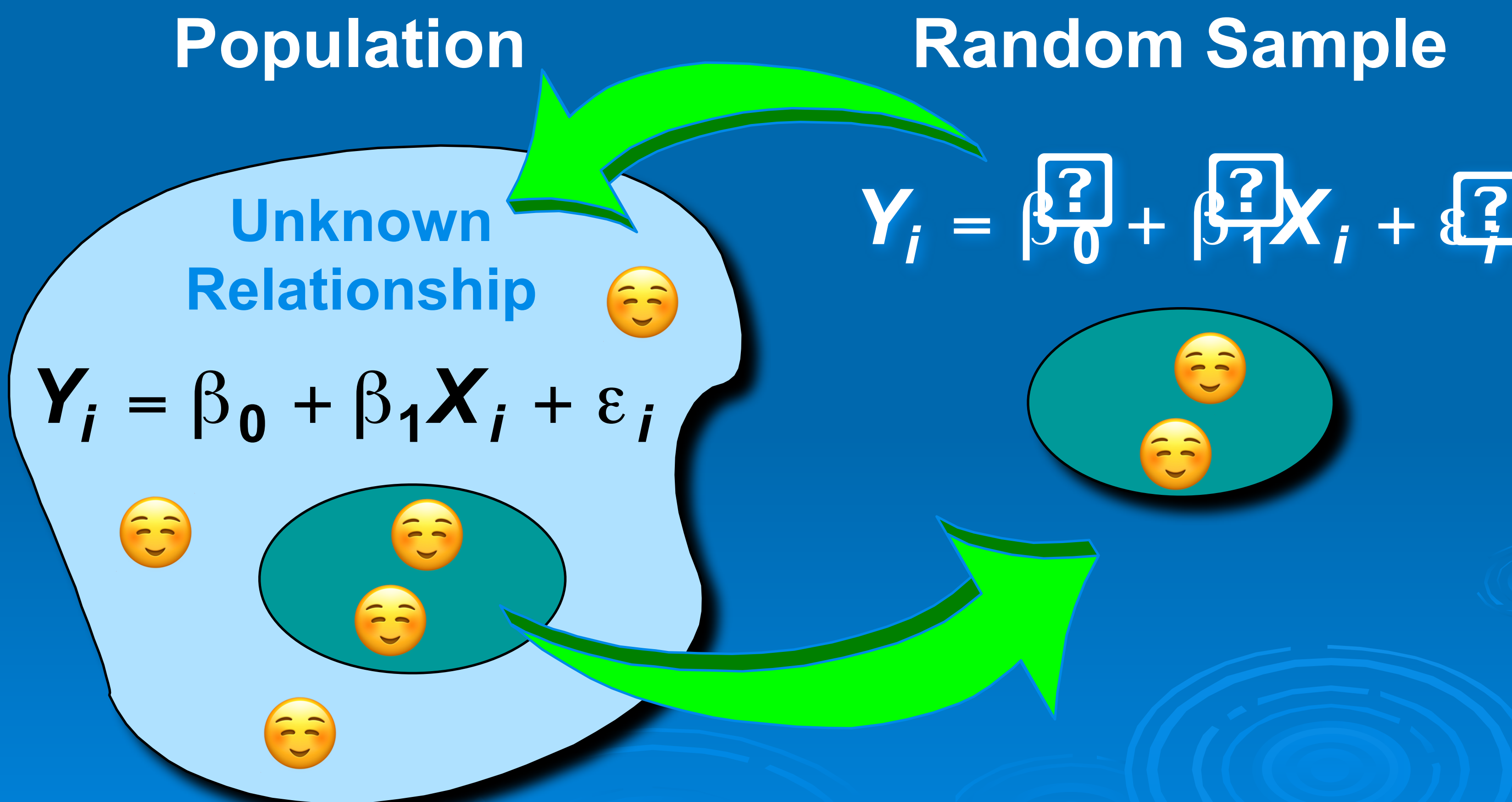


Population & Sample Regression Models

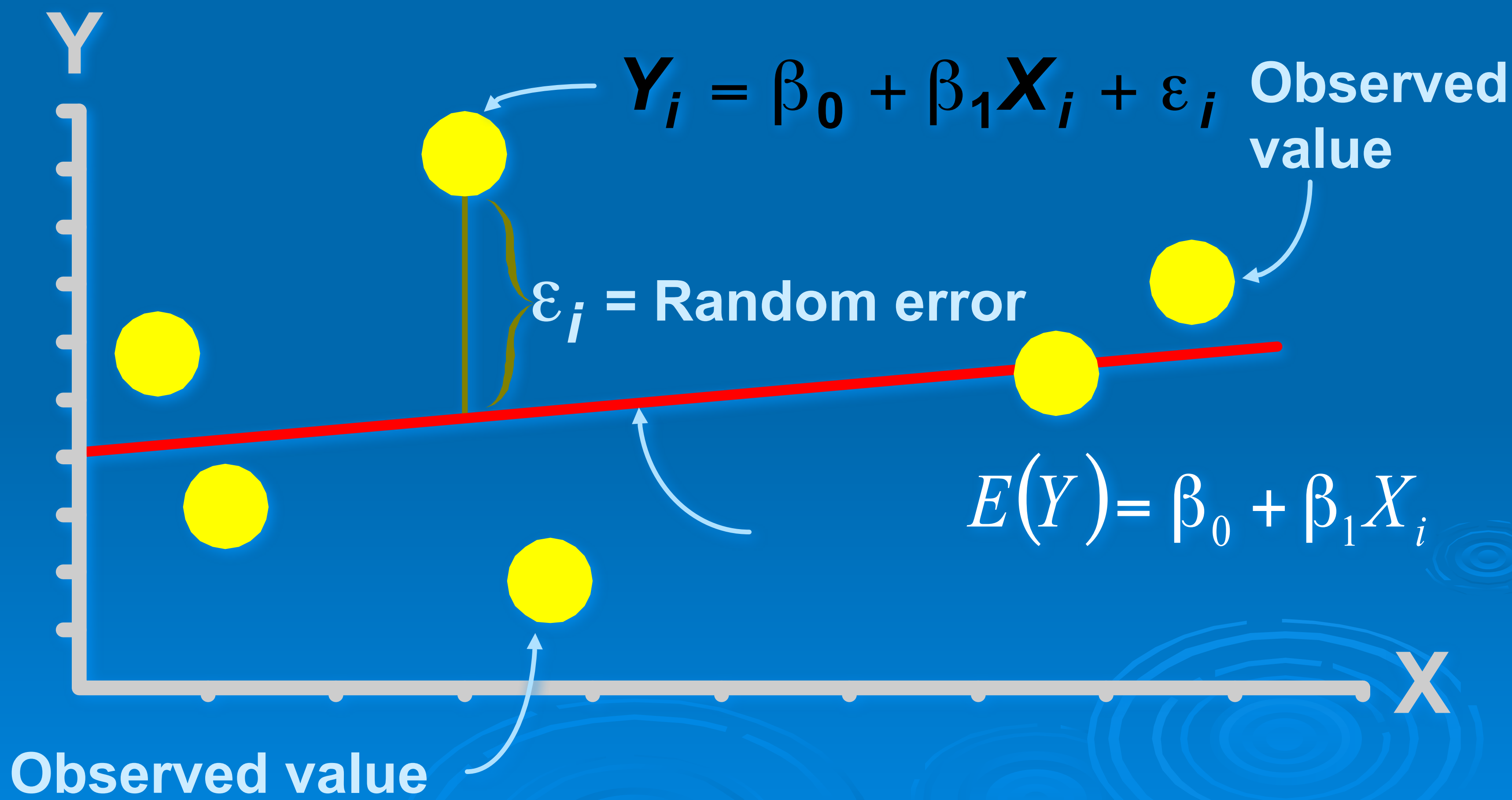
Population **Random Sample**



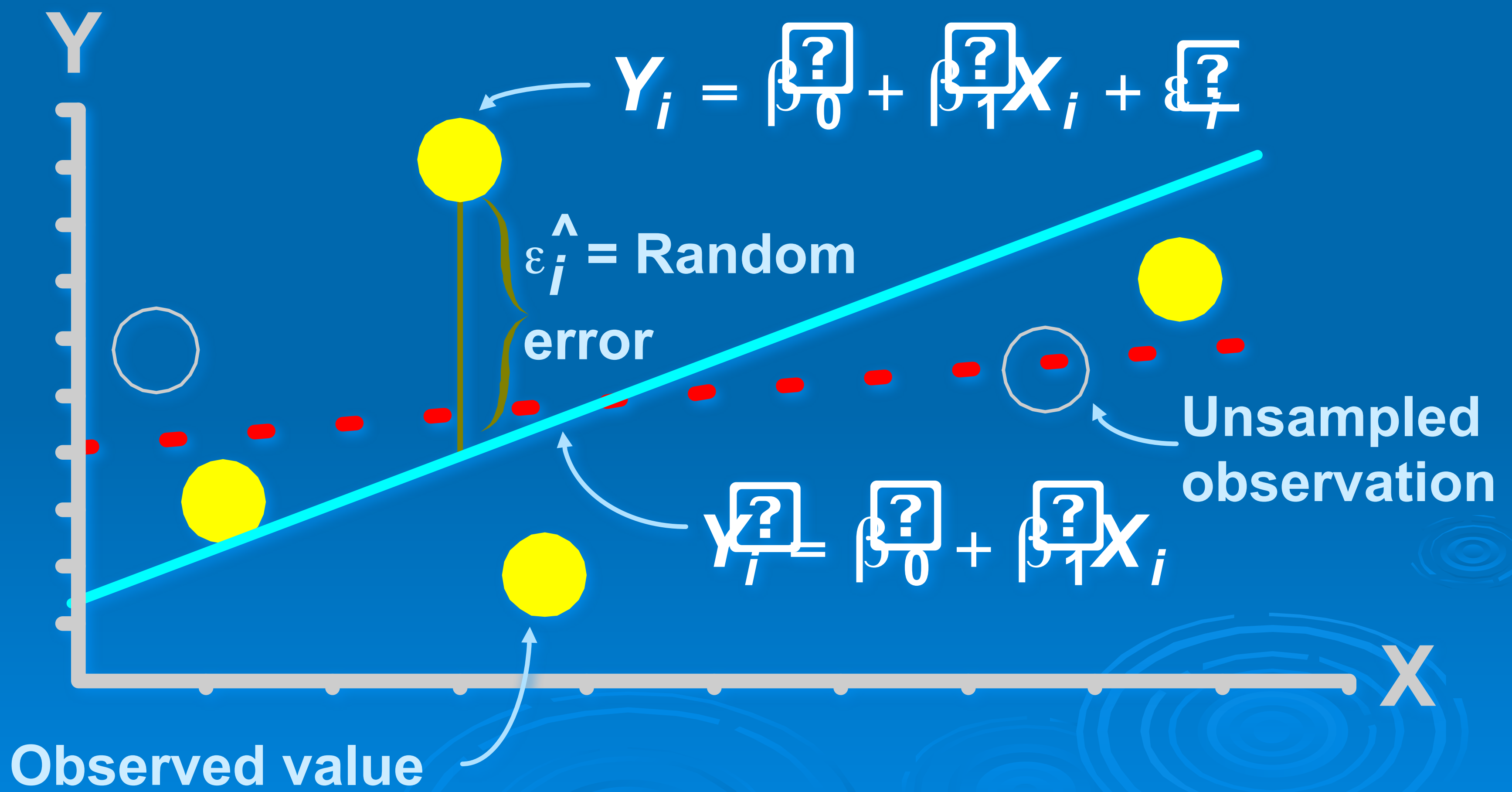
Population & Sample Regression Models



Population Linear Regression Model



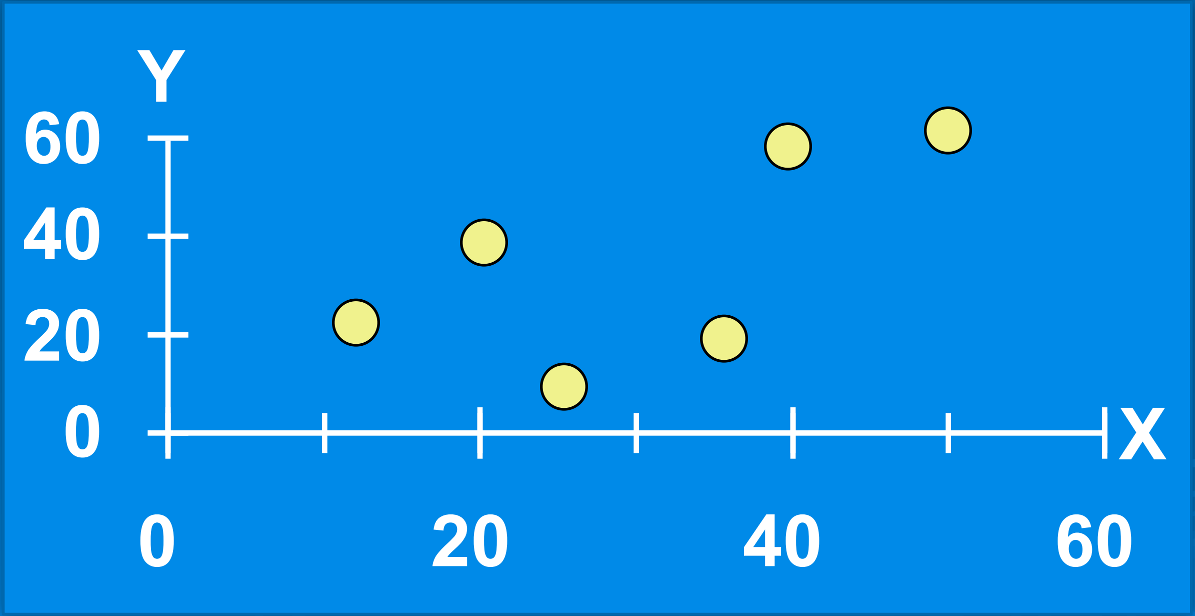
Sample Linear Regression Model



Estimating Parameters: Least Squares Method

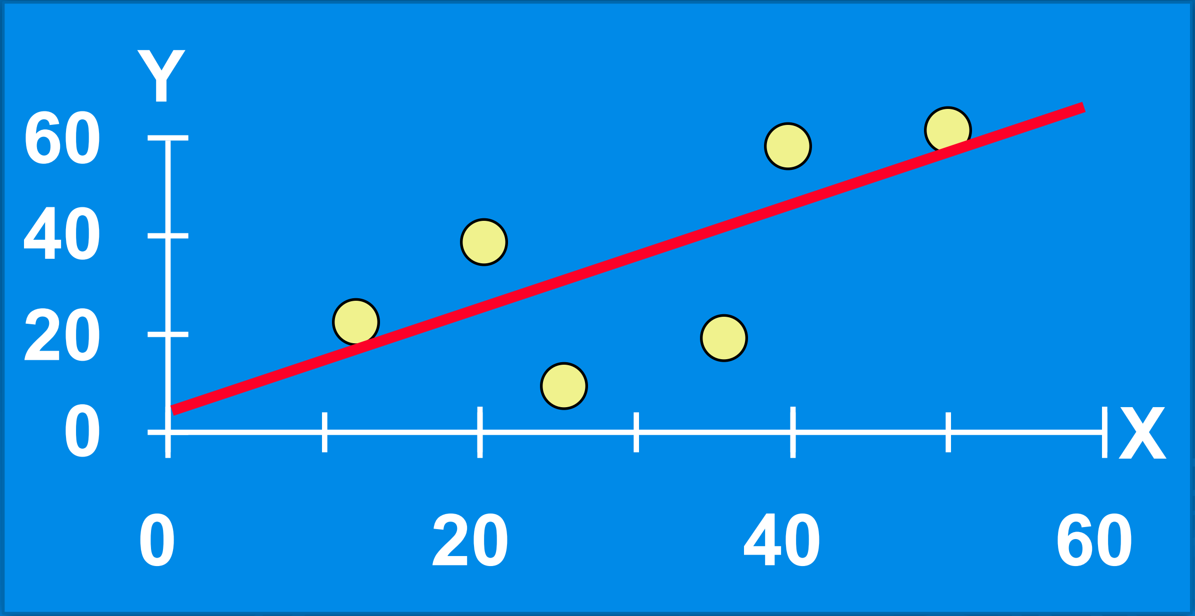
Scatter plot

- 1. Plot of All (X_i, Y_i) Pairs
- 2. Suggests How Well Model Will Fit



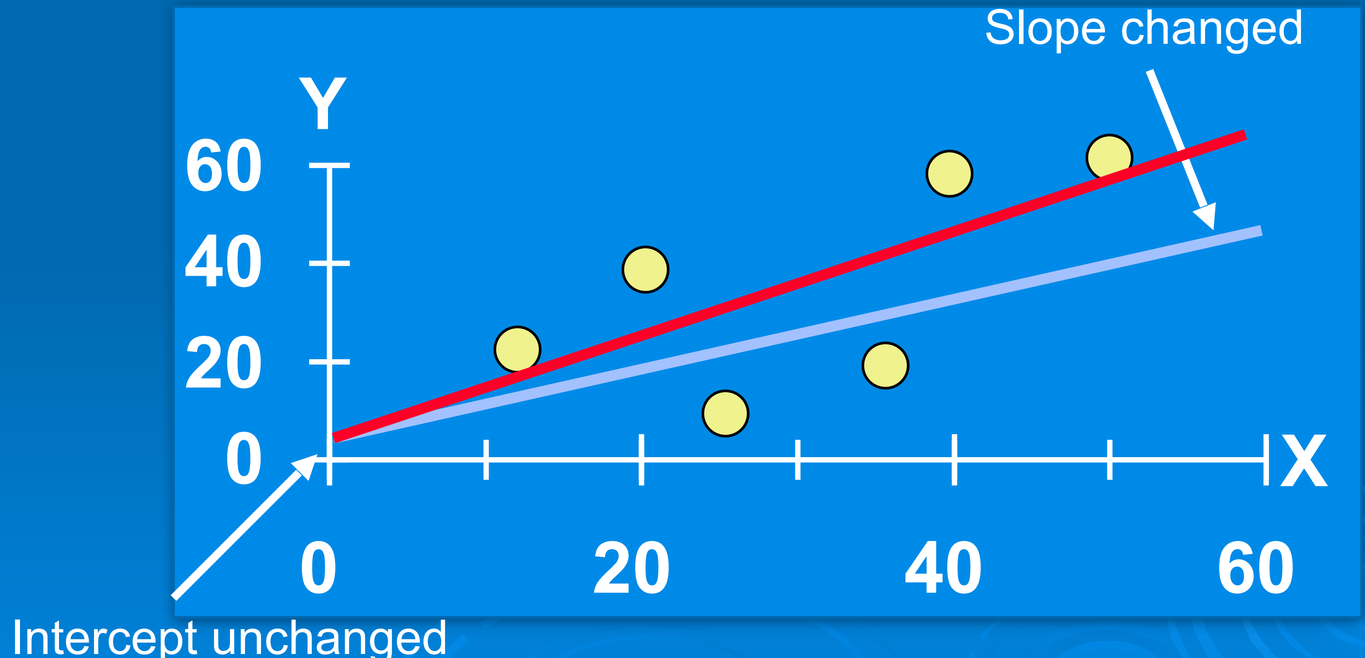
Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



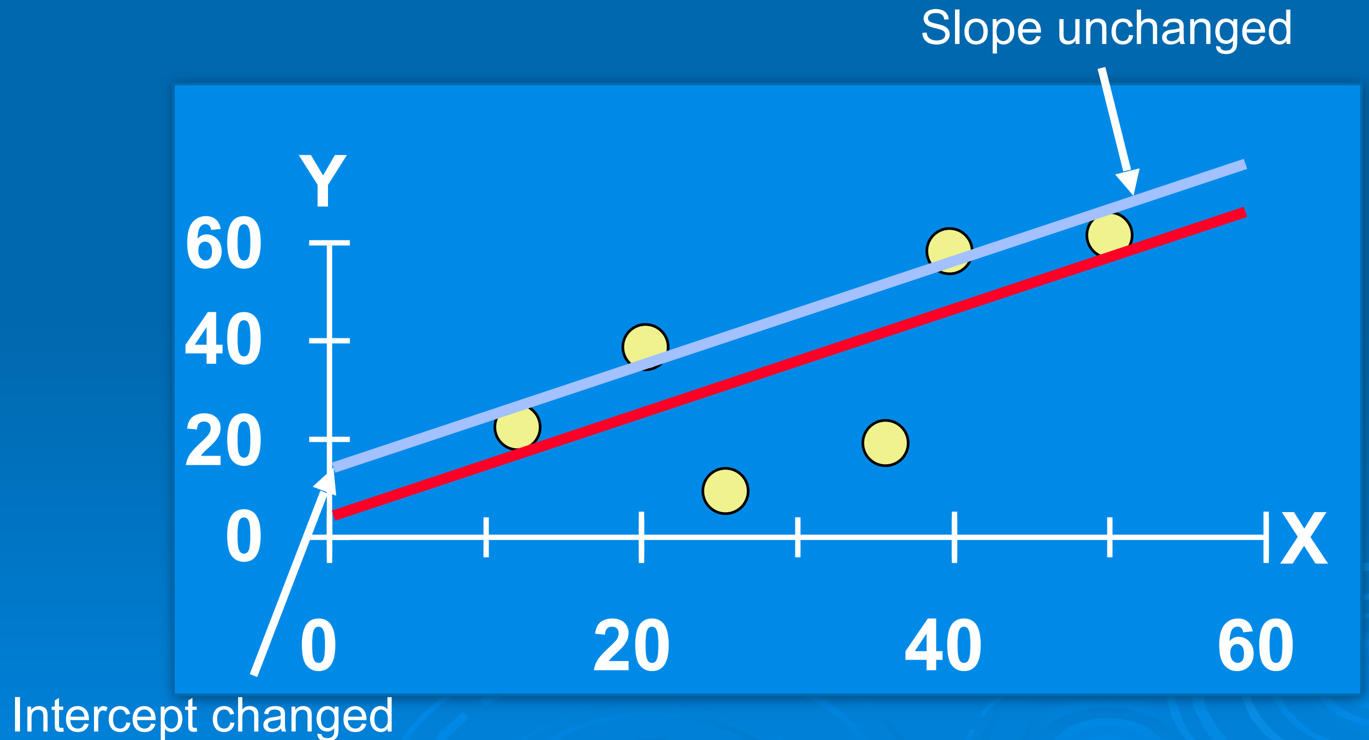
Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



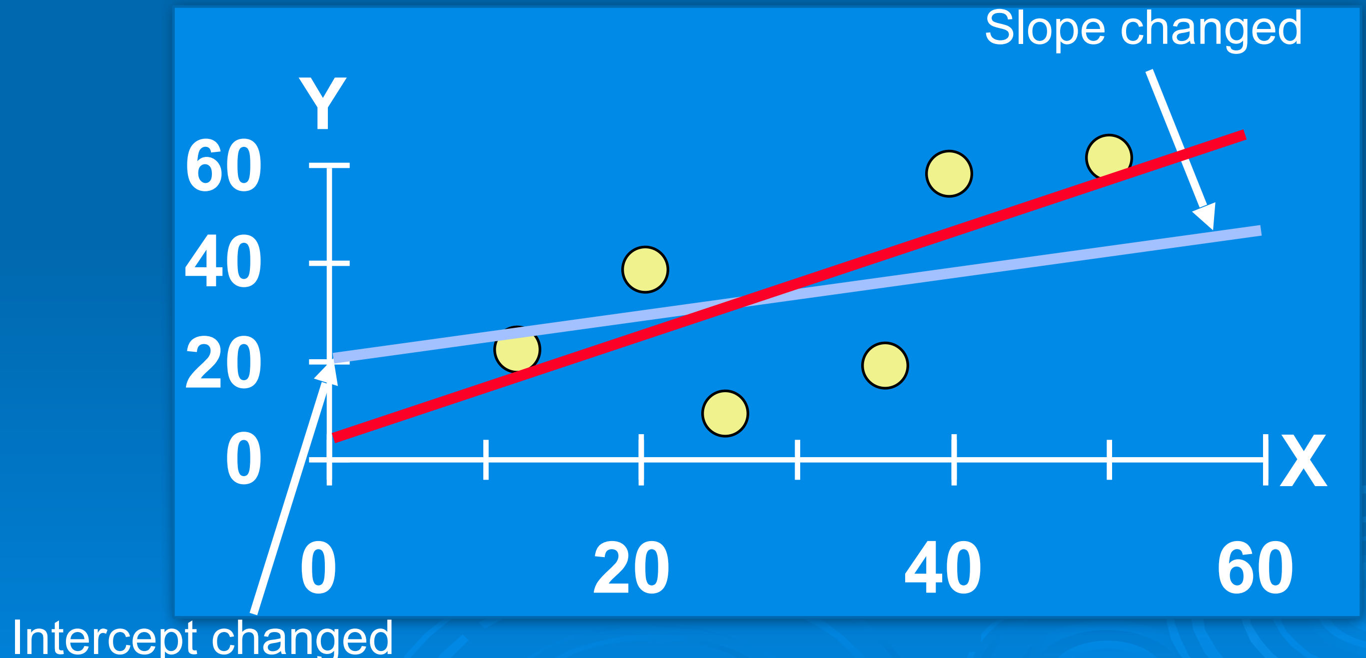
Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



Thinking Challenge

How would you draw a line through the points? How do you determine which line 'fits best'?



Least Squares

- 1. ‘Best Fit’ Means Difference Between Actual Y Values & Predicted Y Values Are a Minimum. *But* Positive Differences Off-Set Negative ones

Least Squares

- 1. 'Best Fit' Means Difference Between Actual Y Values & Predicted Y Values is a Minimum. *But* Positive Differences Off-Set Negative ones. So square errors!

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

Least Squares

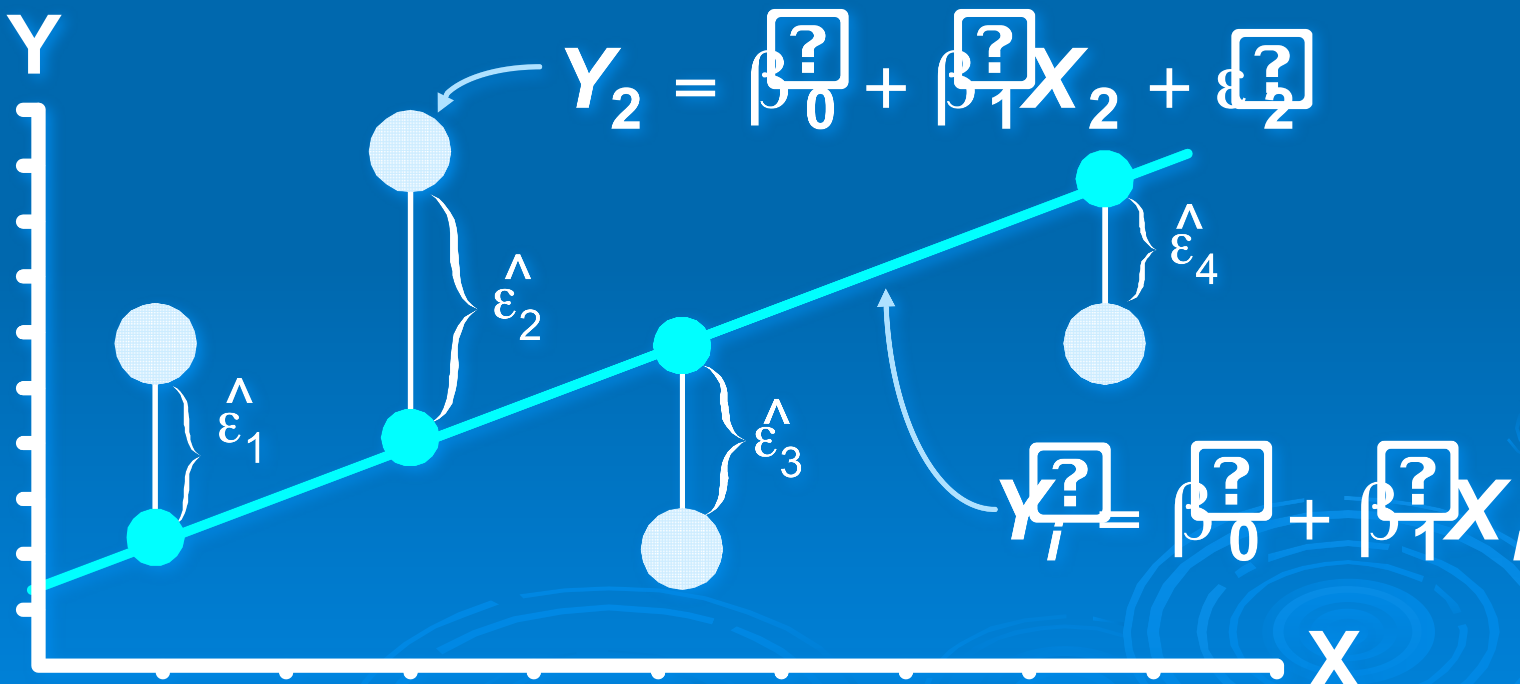
- 1. 'Best Fit' Means Difference Between Actual Y Values & Predicted Y Values Are a Minimum. *But* Positive Differences Off-Set Negative. So square errors!

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

- 2. LS Minimizes the Sum of the Squared Differences (errors) (SSE)

Least Squares Graphically

LS minimizes $\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$



Coefficient Equations

Prediction equation

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Sample slope

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Sample Y - intercept

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Interpretation of Coefficients

Interpretation of Coefficients

- 1. Slope ($\hat{\beta}_1$)
  - Estimated *Y* Changes by $\hat{\beta}_1$ for Each 1 Unit Increase in *X*
    - If $\hat{\beta}_1 = 2$, then *Y* Is Expected to Increase by 2 for Each 1 Unit Increase in *X*

Interpretation of Coefficients

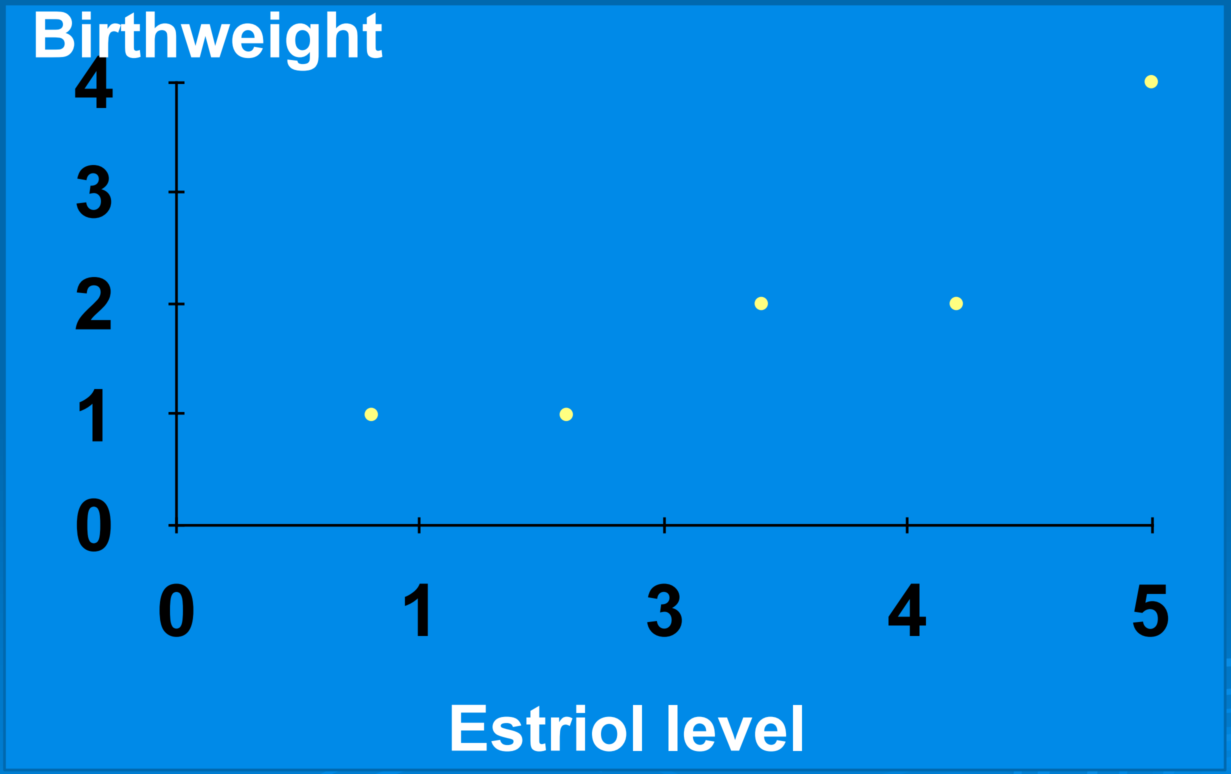
- 1. Slope ($\hat{\beta}_1$)
  - Estimated *Y* Changes by $\hat{\beta}_1$ for Each 1 Unit Increase in *X*
    - If $\hat{\beta}_1 = 2$, then *Y* Is Expected to Increase by 2 for Each 1 Unit Increase in *X*
- 2. Y-Intercept ($\hat{\beta}_0$)
  - Average Value of *Y* When *X* = 0
    - If $\hat{\beta}_0 = 4$, then Average *Y* Is Expected to Be 4 When *X* Is 0

Parameter Estimation Example

- Obstetrics: What is the **relationship** between Mother's Estriol level & Birthweight using the following data?

Estriol (mg/24h)	Birthweight (g/1000)
1	1
2	1
3	2
4	2
5	4

Scatterplot
Birthweight vs. Estriol level



Parameter Estimation Solution Table

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

Parameter Estimation Solution

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{37 - \frac{(15)(10)}{5}}{55 - \frac{(15)^2}{5}} = 0.70$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 2 - (0.70)(3) = -0.10$$

Coefficient Interpretation Solution

Coefficient Interpretation Solution

- 1. Slope ($\hat{\beta}_1$)
  - Birthweight (*Y*) Is Expected to Increase by .7 Units for Each 1 unit Increase in Estriol (*X*)

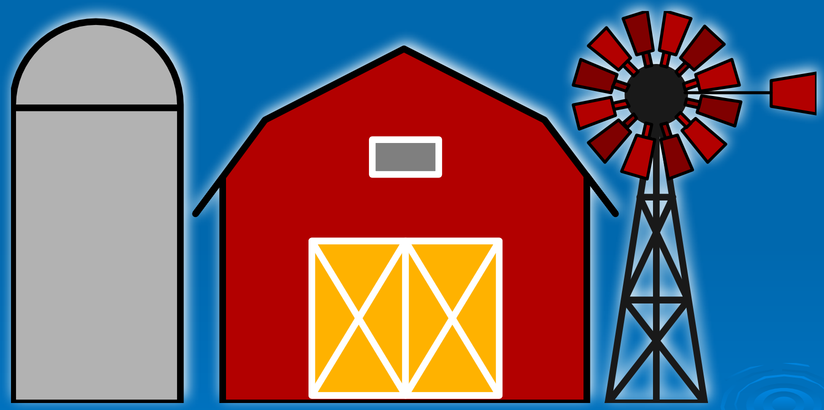
Coefficient Interpretation Solution

- 1. Slope ($\hat{\beta}_1$)
  - Birthweight (*Y*) Is Expected to Increase by .7 Units for Each 1 unit Increase in Estriol (*X*)
- 2. Intercept ($\hat{\beta}_0$)
  - Average Birthweight (*Y*) Is -.10 Units When Estriol level (*X*) Is 0
    - Difficult to explain
    - The birthweight should always be positive

Parameter Estimation Thinking Challenge

- You're a Vet epidemiologist for the county cooperative. You gather the following data:
- Food (lb.) Milk yield (lb.)

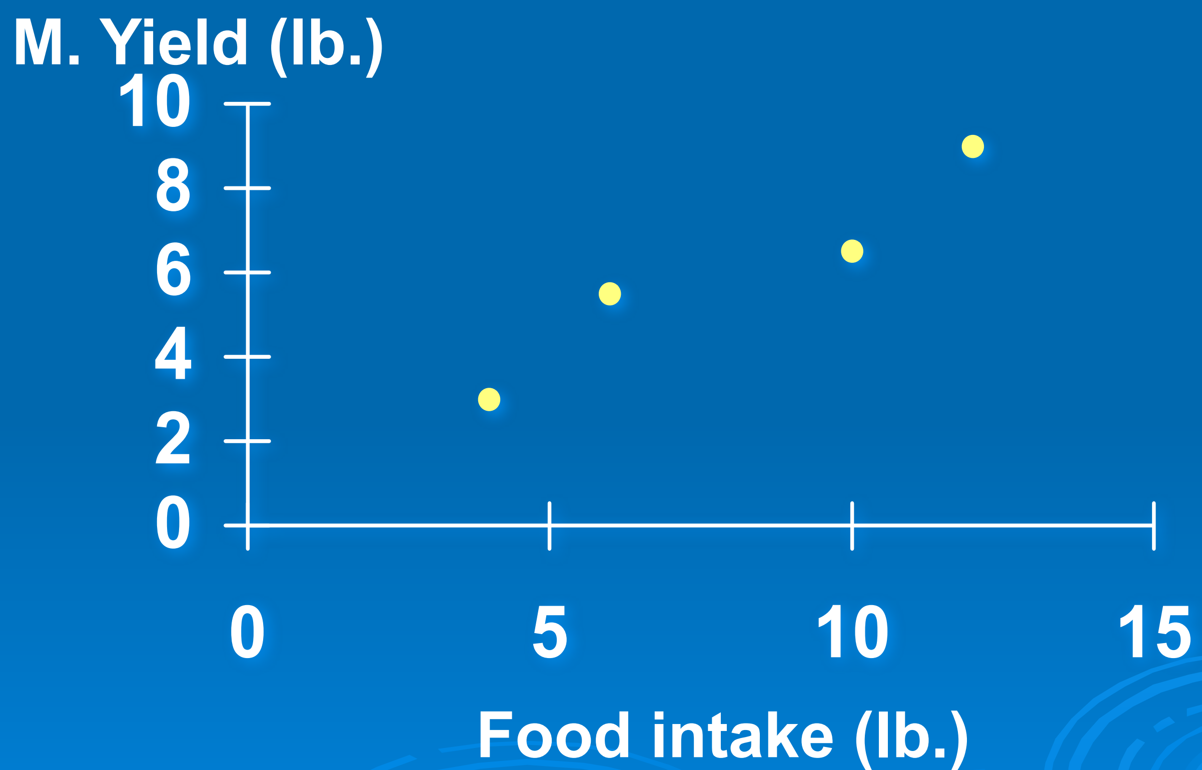
Food (lb.)	Milk yield (lb.)
4	3.0
6	5.5
10	6.5
12	9.0



© 1984-1994 T/Maker Co.

- What is the **relationship** between cows' food intake and milk yield?

Scattergram
Milk Yield vs. Food intake*



Parameter Estimation Solution Table*

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
4	3.0	16	9.00	12
6	5.5	36	30.25	33
10	6.5	100	42.25	65
12	9.0	144	81.00	108
32	24.0	296	162.50	218

Parameter Estimation Solution*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{218 - \frac{(32)(24)}{4}}{296 - \frac{(32)^2}{4}} = 0.65$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 6 - (0.65)(8) = 0.80$$

Coefficient Interpretation Solution*

Coefficient Interpretation Solution*

- 1. Slope ($\hat{\beta}_1$)
  - Milk Yield (*Y*) Is Expected to Increase by .65 lb. for Each 1 lb. Increase in Food intake (*X*)

Coefficient Interpretation Solution*

- 1. Slope ($\hat{\beta}_1$)
  - Milk Yield (Y) Is Expected to Increase by .65 lb. for Each 1 lb. Increase in Food intake (X)
- 2. Y-Intercept ($\hat{\beta}_0$)
  - Average Milk yield (Y) Is Expected to Be 0.8 lb. When Food intake (X) Is 0

Coefficient of determination

– To measure the strength of the linear relationship we use the coefficient of determination.

$$R^2 = \frac{\left[\sum (x_i - \bar{x})(y_i - \bar{y})\right]^2}{s_x^2 s_y^2}$$

$$or \quad R^2 = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2}$$

Note that the coefficient of determination is r^2

Coefficient of determination

- R^2 measures the proportion of the variation in y that is explained by the variation in x.

$$R^2 = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2} = \frac{\sum (y_i - \bar{y})^2 - SSE}{\sum (y_i - \bar{y})^2} = \frac{SSR}{\sum (y_i - \bar{y})^2}$$

- R^2 takes on any value between zero and one.
  $R^2 = 1$: Perfect match between the line and the data points.
  $R^2 = 0$: There are no linear relationship between x and y.