



# LAC-FMB

Liga  
Acadêmica de  
Ciências

A sessão começará às 18h30

**Este documento foi gerado a partir de uma sessão da Liga Acadêmica de Ciências da Faculdade de Medicina da Bahia.**

**Todos os direitos estão reservados aos seus autores e à liga. Não edite o documento, não tire print. A utilização da sua imagem está proibida.**

LAC-FMB

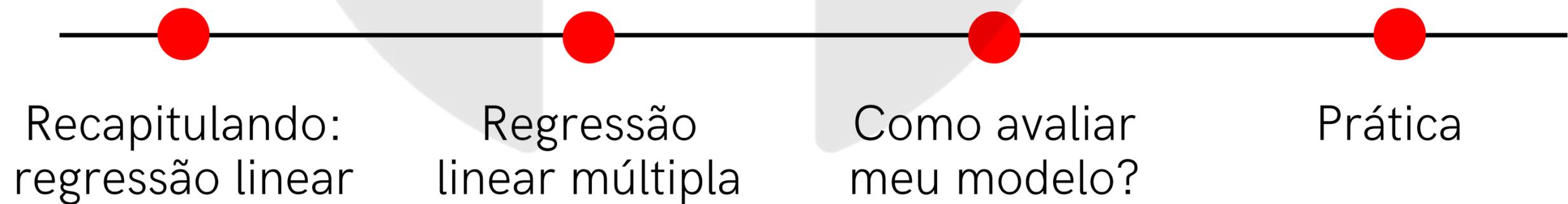
# Regressão linear múltipla

Módulo de Bioestatística

Marina Moniz  
Rebeca Borges



# Sumário



LAC-FMB

Liga Acadêmica de Ciências

# Recapitulando: regressão linear

#01

Módulo de Bioestatística

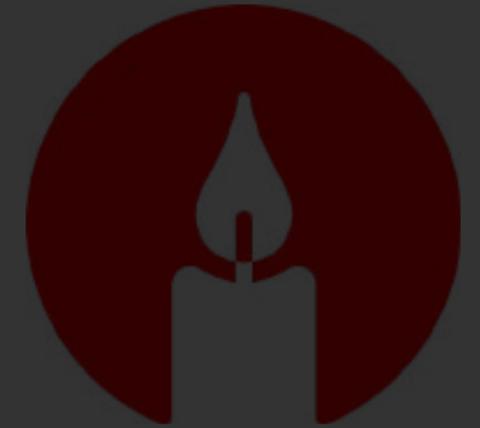


# Correlação X Regressão



LAG FMD

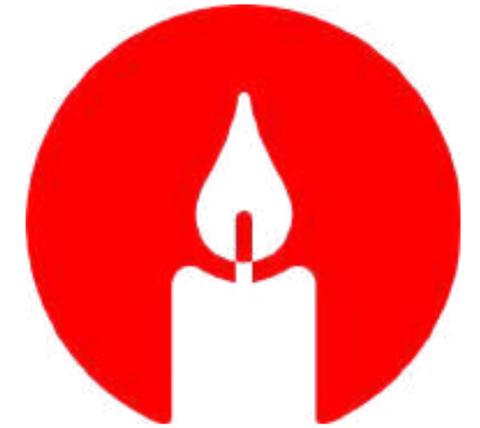
# Correlação X Regressão



Sávio

Qual a diferença entre correlação e regressão?

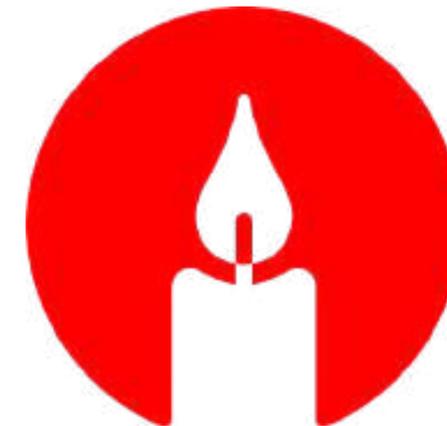
# Correlação X Regressão



Mede a força ou o grau de relacionamento entre as variáveis

Usa a força que foi estudada para promover um valor de desfecho

# Correlação X Regressão



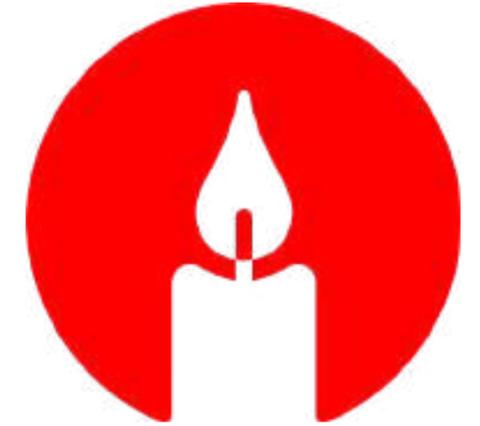
● Mede a **força** ou o grau de relacionamento entre as variáveis

● Dá um **número**

● Usa a **força** que foi estudada para promover um valor de desfecho

● Dá uma **equação matemática** que descreve essa relação

# Correlação X Regressão



● Mede a **força** ou o grau de relacionamento entre as variáveis

● Dá um **número**

● Usa a **força** que foi estudada para promover um valor de desfecho

● Dá uma **equação matemática** que descreve essa relação

● A partir dele podemos encontrar um **poder preditivo** das variáveis

# Correlação X Regressão



● Mede a **força** ou o grau de relacionamento entre as variáveis

● Dá um **número**

● Usa a **força** que foi estudada para promover um valor de desfecho

● Dá uma **equação matemática** que descreve essa relação

● A partir dele podemos encontrar um **poder preditivo** das variáveis

LAC-FMB

Liga Acadêmica de Ciências



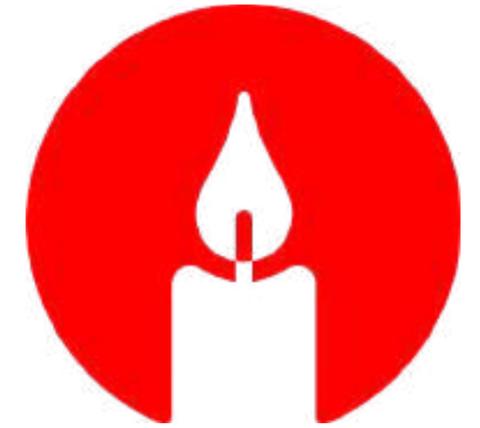
Evelyn

"Dá uma equação matemática que descreve essa relação entre variáveis"

**Que equação é essa?**

Que equação é essa?

$$Y = (b_0 + b_1 X_1) + \varepsilon_i$$



$X_n$

$Y$

$\epsilon_i$

LAC-FMB

Liga Acadêmica de Ciências



$X_n$  Variável preditora ou independente

$Y$

$\epsilon_i$

● Pode ser quantitativa

● Pode ser categórica



LAC-FMB

Liga Acadêmica de Ciências



$X_n$  Variável preditora ou independente

$Y$

$\epsilon_i$

● Pode ser quantitativa

● Pode ser categórica

└─ Variável  
   Dummy

LAC-FMB

Liga Acadêmica de Ciências

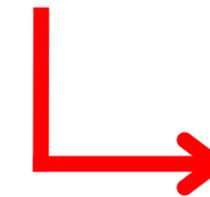


$X_n$

Variável preditora ou independente

● Pode ser quantitativa

● Pode ser categórica



Variável  
Dummy

Exemplo: níveis de atividade física

1- Intensa

2- Moderada

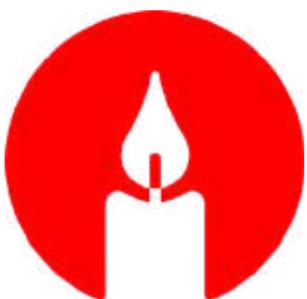
3- Leve

4- Nenhuma

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n)$$

LAC-FMIB

Liga Acadêmica de Ciências

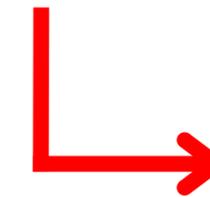


$X_n$

Variável preditora ou independente

● Pode ser quantitativa

● Pode ser categórica



Variável  
Dummy

Exemplo: níveis de atividade física

1- Intensa **Referência!**

2- Moderada  $X_1$

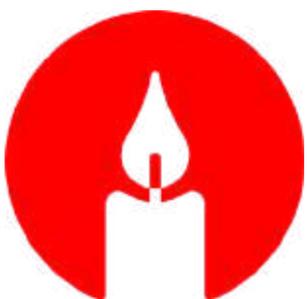
3- Leve  $X_2$

4- Nenhuma  $X_3$

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n)$$

LAC-FMIB

Liga Acadêmica de Ciências



$X_n$  Variável preditora ou independente

$Y$

$\epsilon_i$

● Pode ser quantitativa

● Pode ser categórica

└─ Variável  
   Dummy



$X_n$  Variável preditora ou independente

$Y$  Variável de desfecho ou dependente

$\epsilon_i$

● Pode ser quantitativa

● Pode ser categórica

└─→ Variável  
Dummy

● Sempre quantitativa

●



$X_n$  Variável preditora ou independente

$Y$  Variável de desfecho ou dependente

$\epsilon_i$  Erros ou resíduos

● Pode ser quantitativa

● Pode ser categórica

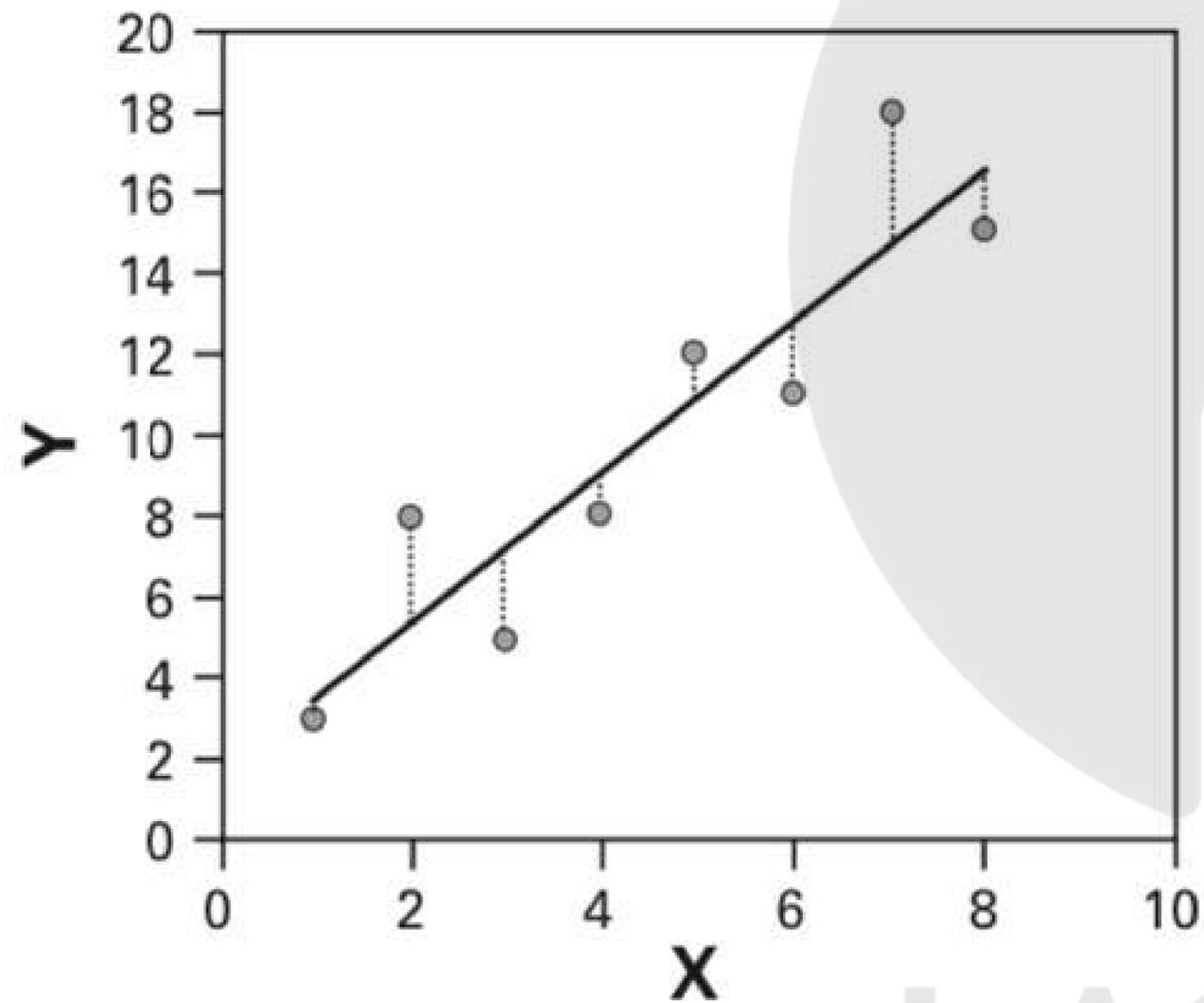
└─→ Variável  
Dummy

● Sempre quantitativa

● Mostra que o modelo  
não é perfeito



$$Y_i = (b_0 + b_1 X_1) + \varepsilon_i$$



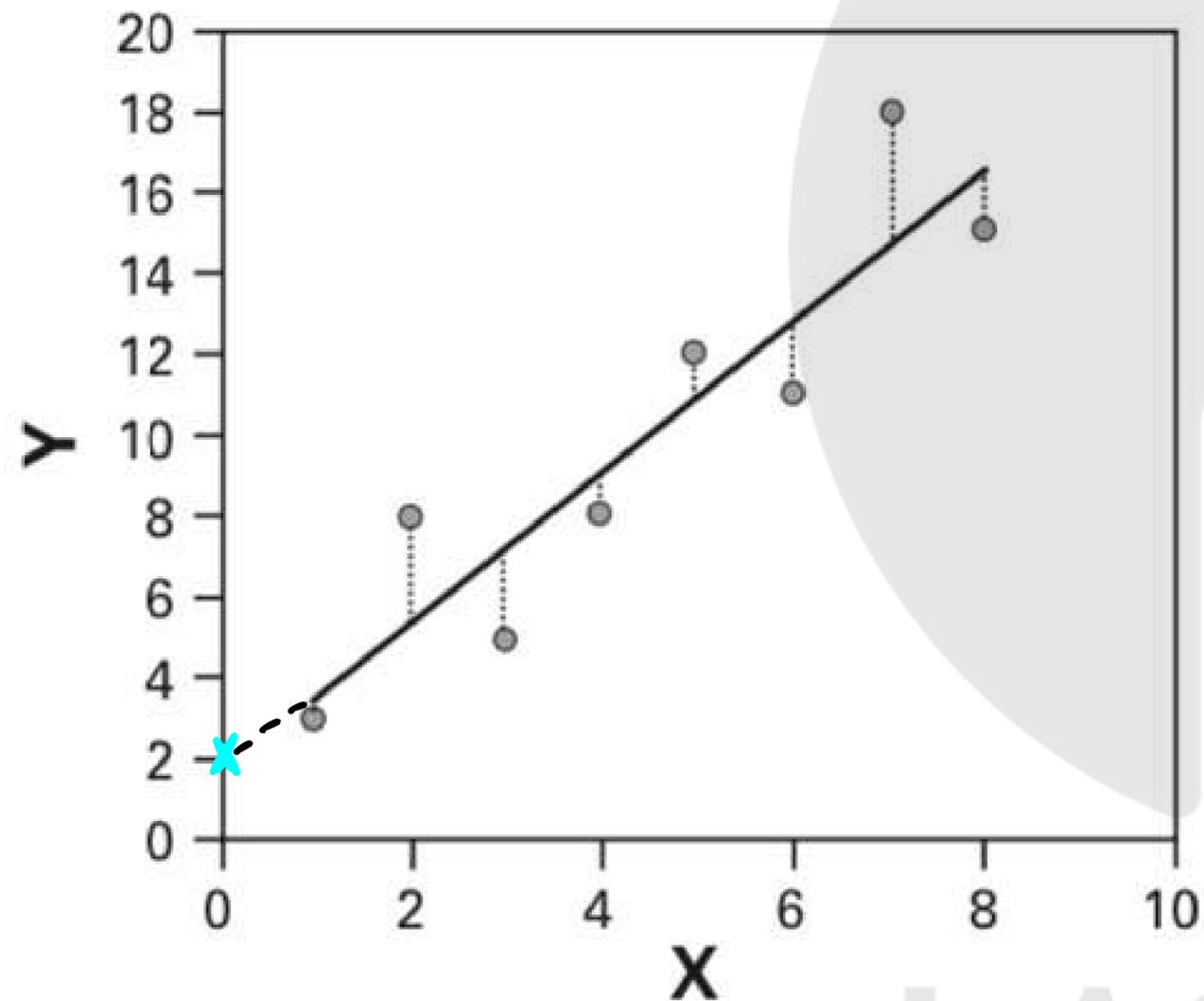
Demais coeficientes

$b_0 = ?$

$b_1 = ?$



$$Y_i = (b_0 + b_1 X_1) + \varepsilon_i$$



# Demais coeficientes

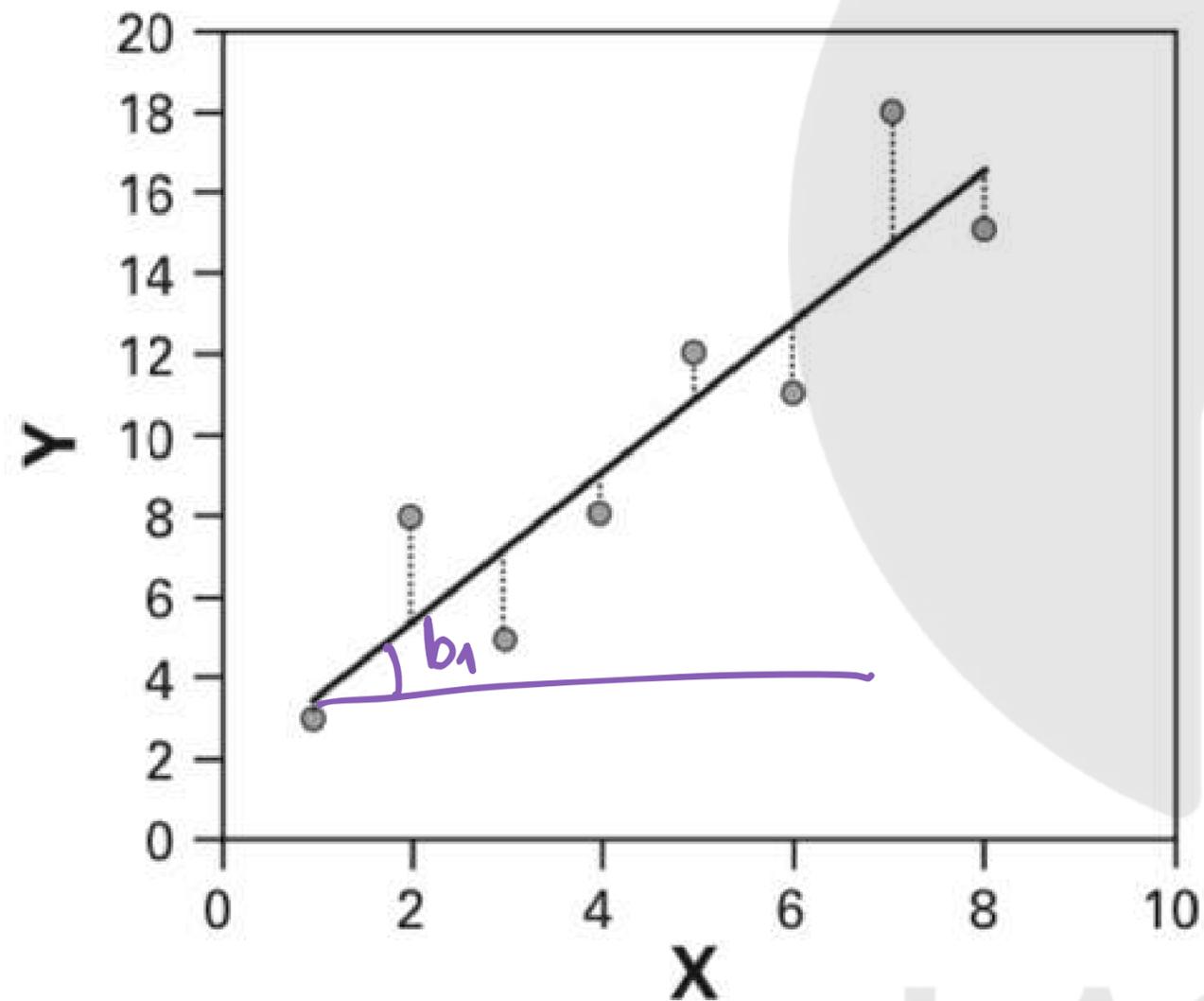
**$b_0$**

- Coeficiente linear
- Intercepto
- "b" em  $y = aX + b$

**$b_1 = ?$**



$$Y_i = (b_0 + b_1 X_1) + \varepsilon_i$$



# Demais coeficientes

**$b_0$**

- Coeficiente linear
- Intercepto
- "b" em  $y = aX + b$

**$b_1$**

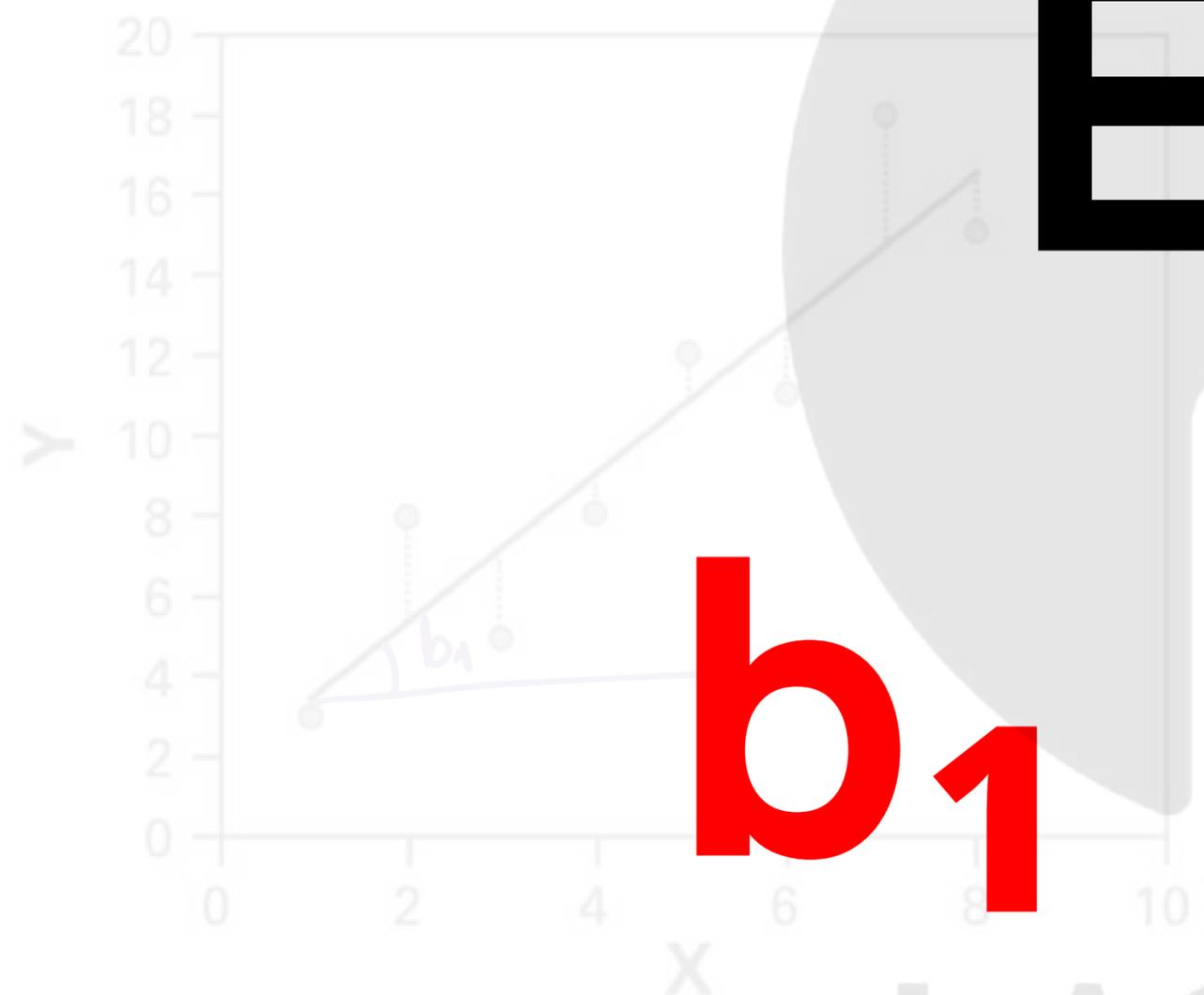
- Coeficiente angular
- Inclinação
- "a" em  $y = aX + b$
- Teremos mais de um



$$Y_i = (b_0 + b_1 X_1) + \epsilon_i$$

Demais coeficientes

E se



$b_1 = 0$ ?

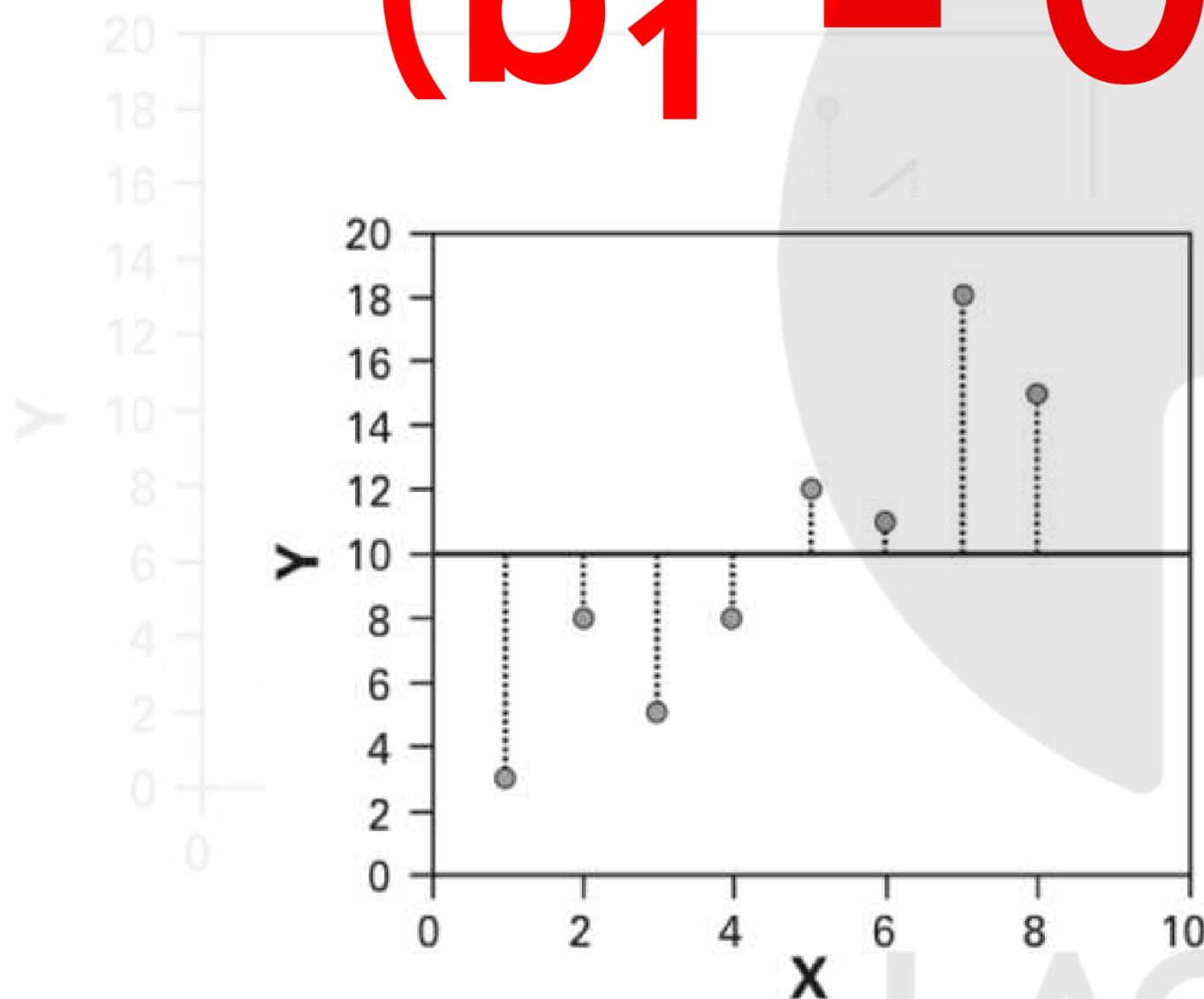
- Coeficiente linear
- Intercepto
- "b" em  $y = aX + b$

- Coeficiente angular
- Inclinação
- "a" em  $y = aX + b$
- Teremos mais de um



$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n)$$

# $(b_1 = 0)$ - média



- Hipótese nula  $b_1 = 0$
- Hipótese alternativa  $b_1 \neq 0$

Demais coeficientes:

- Intercepto
- "b" em  $y = aX + b$

- Coeficiente angular
- Inclinação
- "a" em  $y = aX + b$
- Teremos mais de um



LAC-FMB

Liga Acadêmica de Ciências

## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>
1	0.805	0.648

### Model Coefficients - Peso

Predictor	Estimate	SE	t	p
Intercept *	48.943	6.059	8.078	< .001
Doce (X/mês)	0.688	0.215	3.196	0.002
Sexo:				
M – F	6.657	2.963	2.247	0.027
Exercício:				
leve – intenso	10.694	4.466	2.394	0.019
moderado – intenso	-1.353	4.112	-0.329	0.743
nenhum – intenso	24.521	4.243	5.779	< .001

\* Represents reference level

# Demais coeficientes

p: ?

R: ?

R<sup>2</sup>: ?



## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>
1	0.805	0.648

### Model Coefficients - Peso

Predictor	Estimate	SE	t	p
Intercept *	48.943	6.059	8.078	< .001
Doce (X/mês)	0.688	0.215	3.196	0.002
Sexo:				
M – F	6.657	2.963	2.247	0.027
Exercício:				
leve – intenso	10.694	4.466	2.394	0.019
moderado – intenso	-1.353	4.112	-0.329	0.743
nenhum – intenso	24.521	4.243	5.779	< .001

\* Represents reference level

# Demais coeficientes

p:

- Teste F
- Modelo é significativo?
- $p < 0,05$

R: ?

R<sup>2</sup>: ?



## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>
1	0.805	0.648

### Model Coefficients - Peso

Predictor	Estimate	SE	t	p
Intercept *	48.943	6.059	8.078	< .001
Doce (X/mês)	0.688	0.215	3.196	0.002
Sexo:				
M – F	6.657	2.963	2.247	0.027
Exercício:				
leve – intenso	10.694	4.466	2.394	0.019
moderado – intenso	-1.353	4.112	-0.329	0.743
nenhum – intenso	24.521	4.243	5.779	< .001

\* Represents reference level

# Demais coeficientes

p:

- Teste F
- Modelo é significativo?
- $p < 0,05$

R:

- Coeficiente de correlação
- Análise de correlação de Pearson

R<sup>2</sup>: ?



## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>
1	0.805	0.648

### Model Coefficients - Peso

Predictor	Estimate	SE	t	p
Intercept *	48.943	6.059	8.078	.01
Doce (X/mês)	0.688	0.215	3.196	.02
Sexo:				
M - F	6.657	2.963	2.247	.07
Exercício:				
leve - intenso	10.694	4.466	2.394	.05
moderado - intenso	-1.353	4.112	-0.329	.74
nenhum - intenso	24.521	4.243	5.779	< .001

\* Represents reference level

# Demais coeficientes

- Teste F
- Modelo é significativo?
- $p < 0,05$

Coficiente de correlação  
Análise de correlação de  
Pearson



# LAC-FMB

Liga Acadêmica de Ciências

## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>
1	0.805	0.648

### Model Coefficients - Pearson

Predictor	Estimate	SE	t	Pr >  t
Intercept *	48.943	1.59	30.78	<.0001
Doce (X/mês)	0.688	0.15	4.59	0.0001
Sexo:				
M - F	6.657	1.63	4.07	0.0001
Exercício:				
leve - intenso	0.694	4.466	2.394	0.0193
moderado - intenso	0.131	4.112	-0.320	0.7493
nenhum - intenso	0.243	4.243	0.057	0.9540

\* Represents reference level

# Demais coeficientes

O que são

os **resíduos**

e o **método**

**dos mínimos**

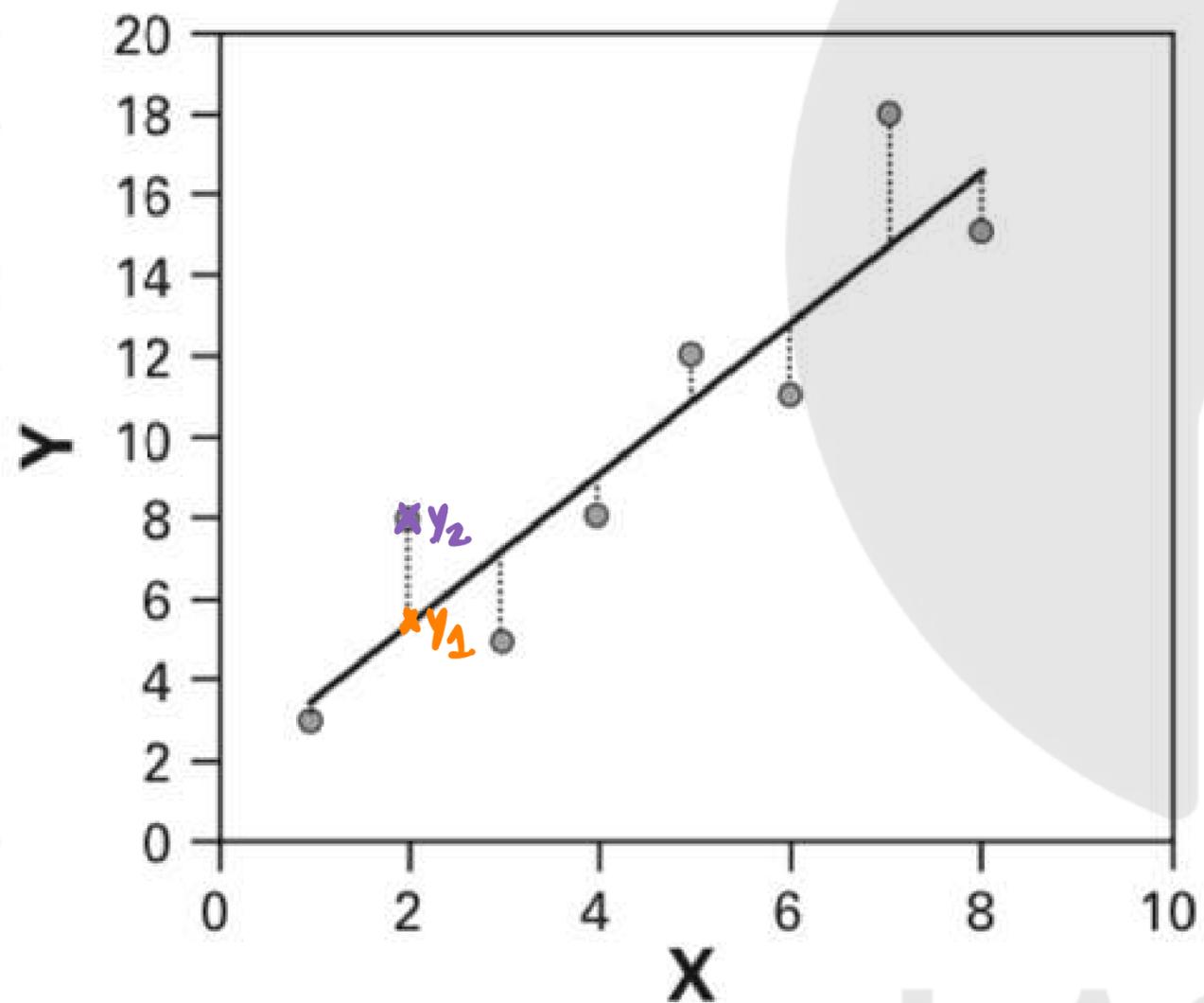
**quadrados?**



LAC-FMB

Liga Acadêmica de Ciências

# Resíduos



Sendo:

- $Y_1 = (b_0 + b_1 X_1)$

- $Y_2 \neq (b_0 + b_1 X_1)$

- $|Y_2 - Y_1| = \text{Resíduo}$

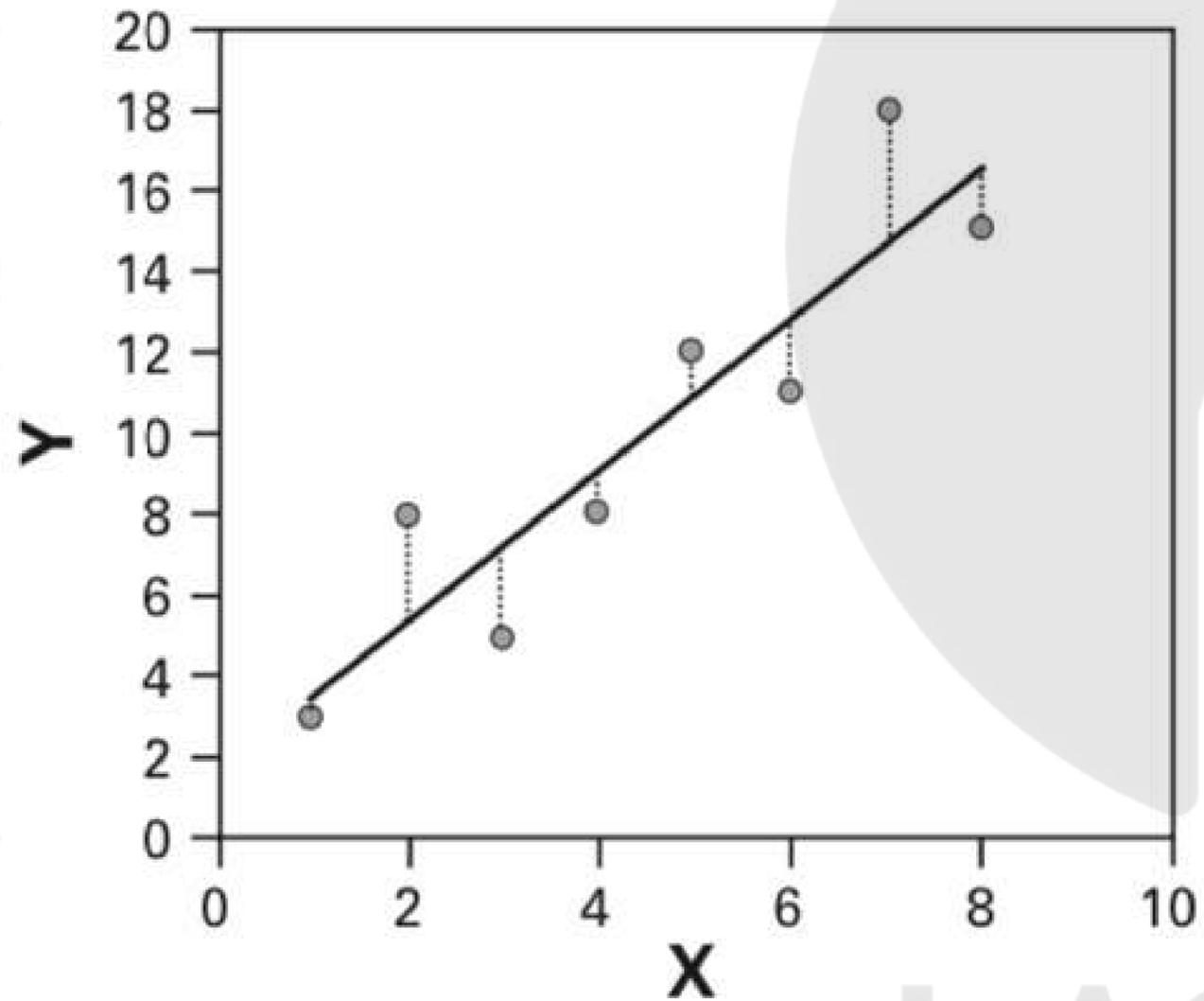
- Resíduo é a diferença entre o Y previsto pelo modelo e o Y encontrado na amostra



LAC-FMB

Liga Acadêmica de Ciências

# Método dos mínimos quadrados



- Cálculo dos quadrados dos resíduos

• Teste F  
• Modelo é significativo?

•  $p < 0,05$

• Coeficiente de correlação

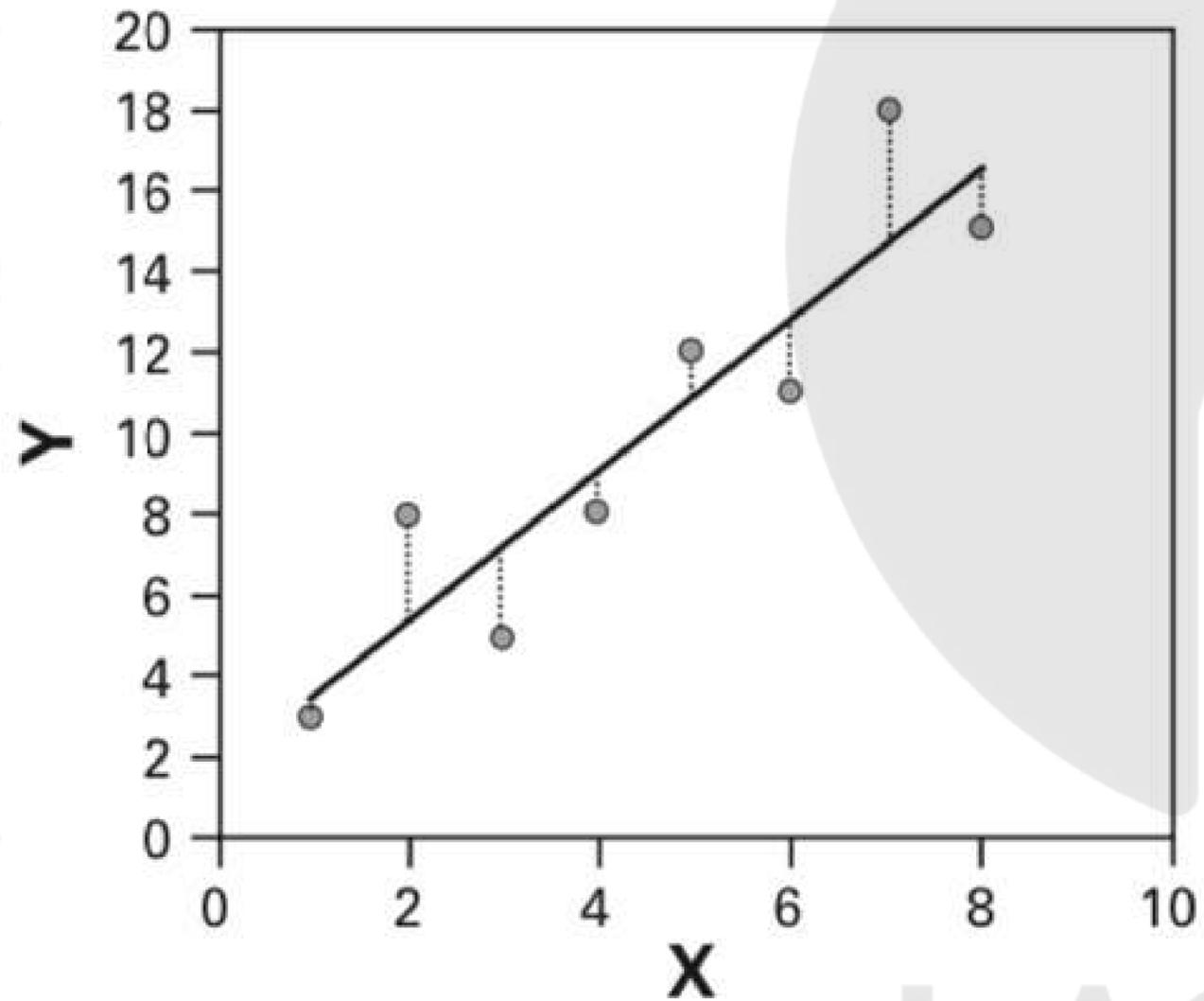
• Análise de correlação de Pearson



LAC-FMB

Liga Acadêmica de Ciências

# Método dos mínimos quadrados



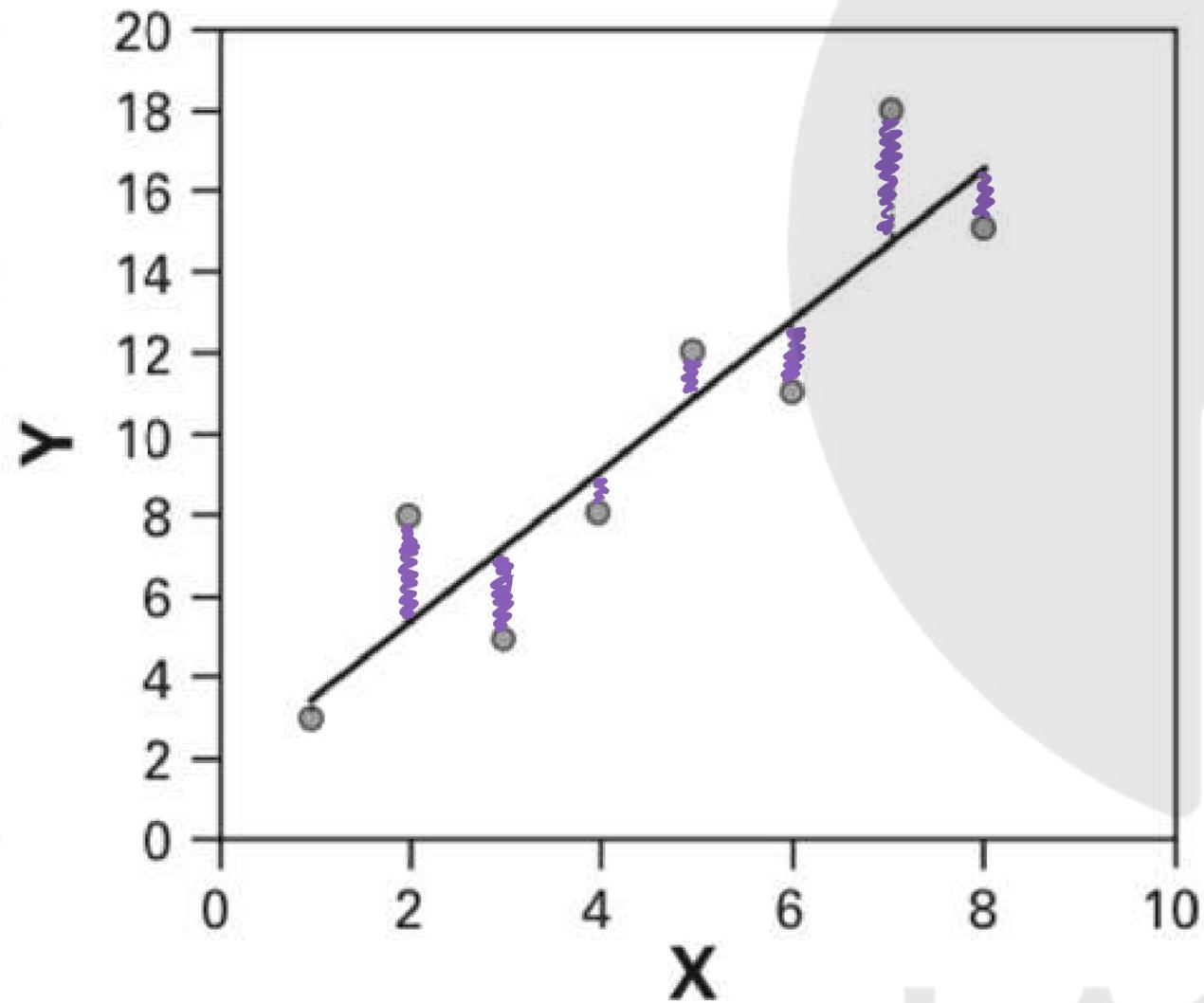
- Cálculo dos quadrados dos resíduos



LAC-FMB

Liga Acadêmica de Ciências

# Método dos mínimos quadrados



- Cálculo dos **quadrados dos resíduos**
- Soma das distâncias das observações para encontrar a reta que melhor explica o modelo.

R:

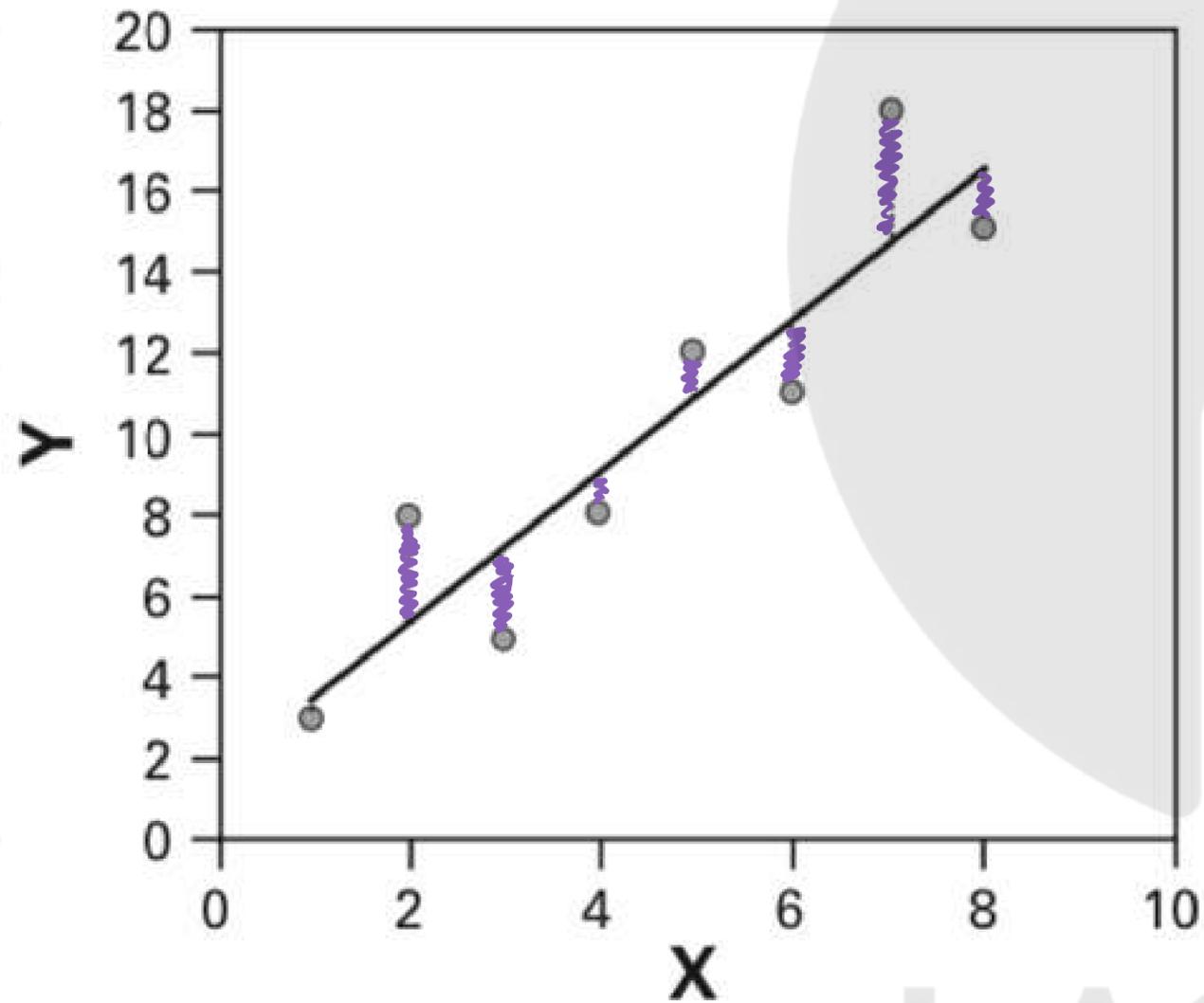
- Coeficiente de correlação
- Análise de correlação de Pearson



LAC-FMB

Liga Acadêmica de Ciências

# Método dos mínimos quadrados

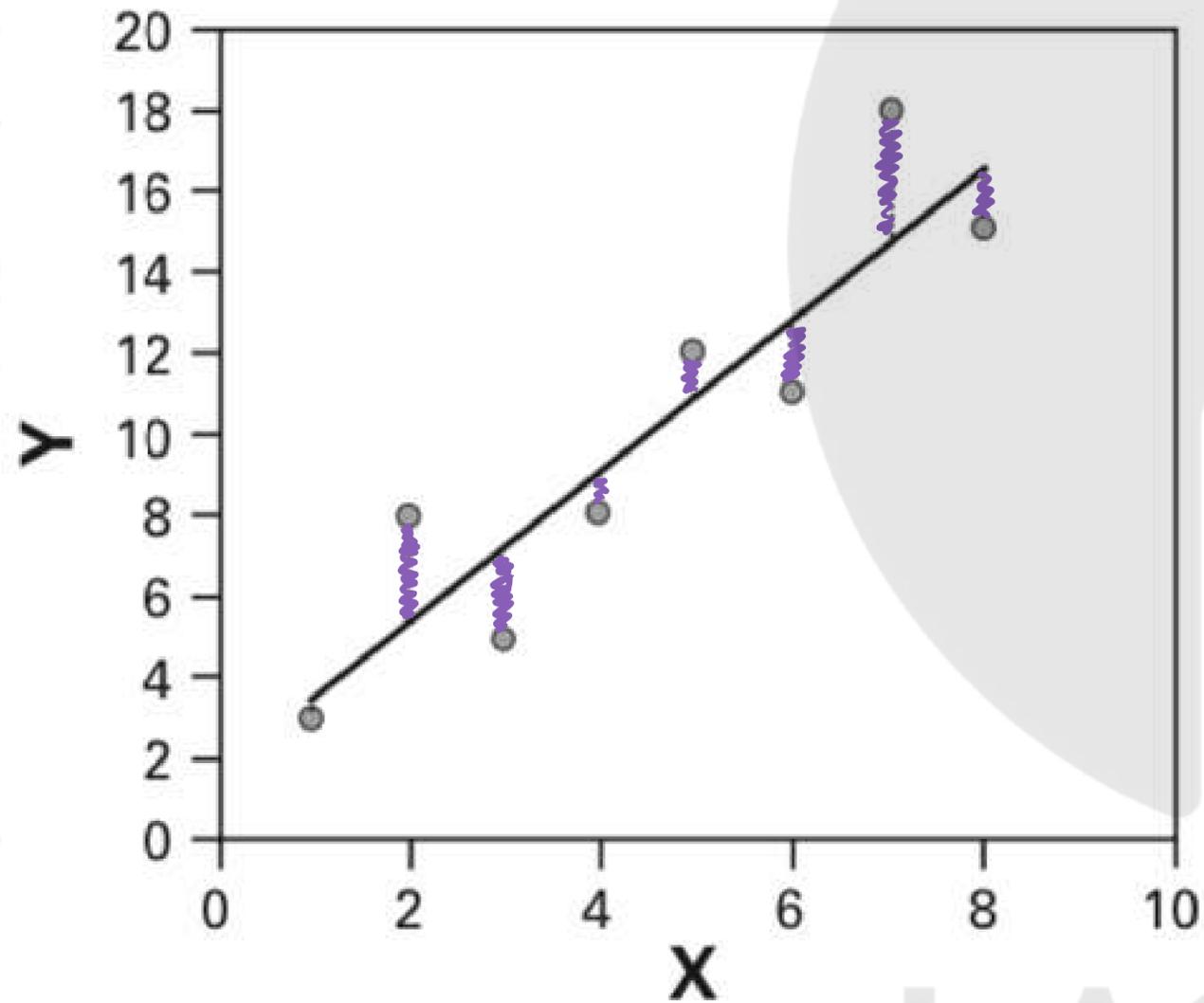


- ~~Cálculo dos quadrados dos resíduos~~
- Soma das distâncias das observações para encontrar a reta que melhor explica o modelo.

R:  
"Quanto **menor** a soma,  
**melhor** a representação!"



# Método dos mínimos quadrados

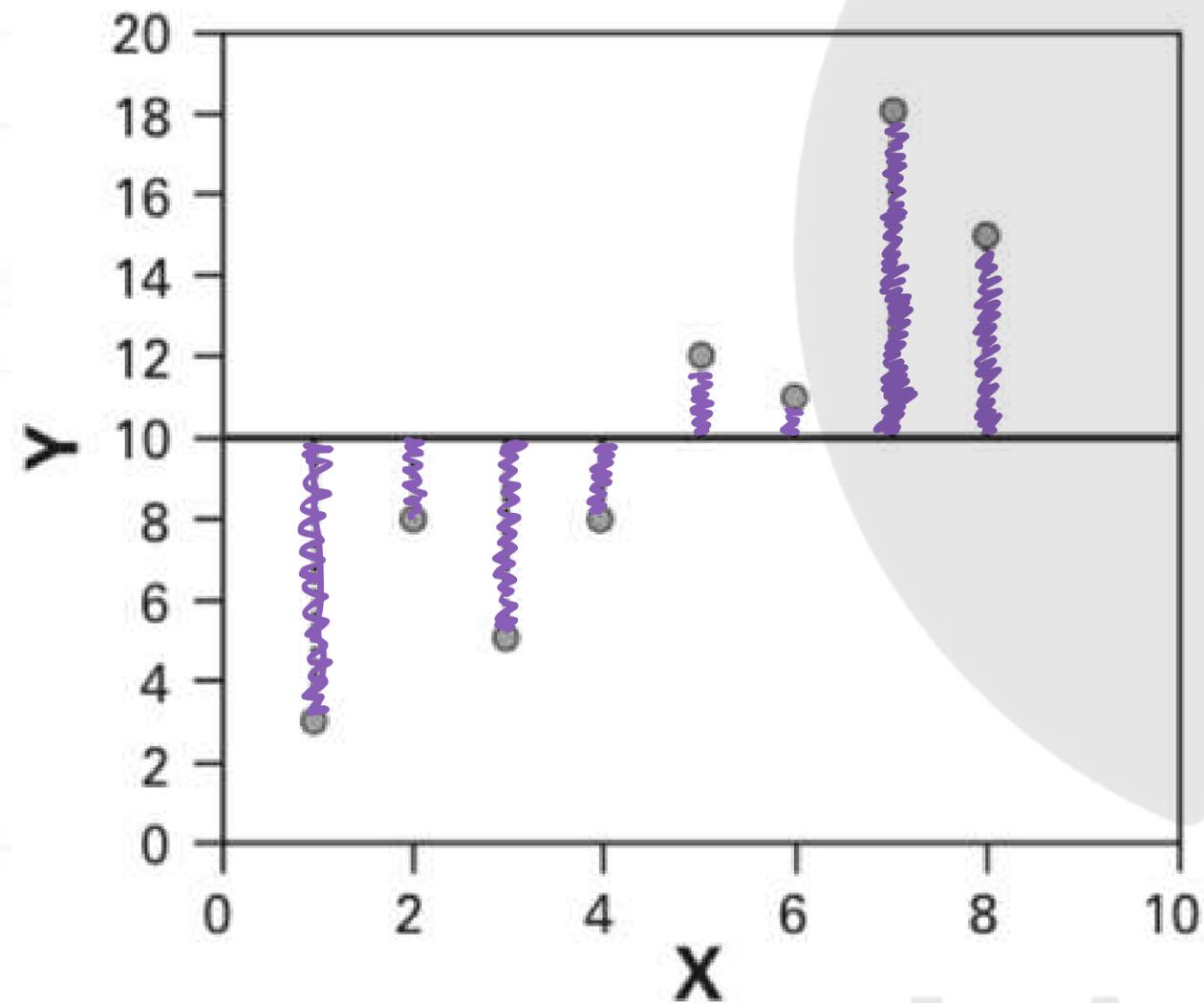


- Cálculo dos quadrados dos resíduos
- Soma das distâncias das observações para encontrar a reta que melhor explica o modelo.

R:  
"Quanto maior a soma, pior a representação!"



# Método dos mínimos quadrados



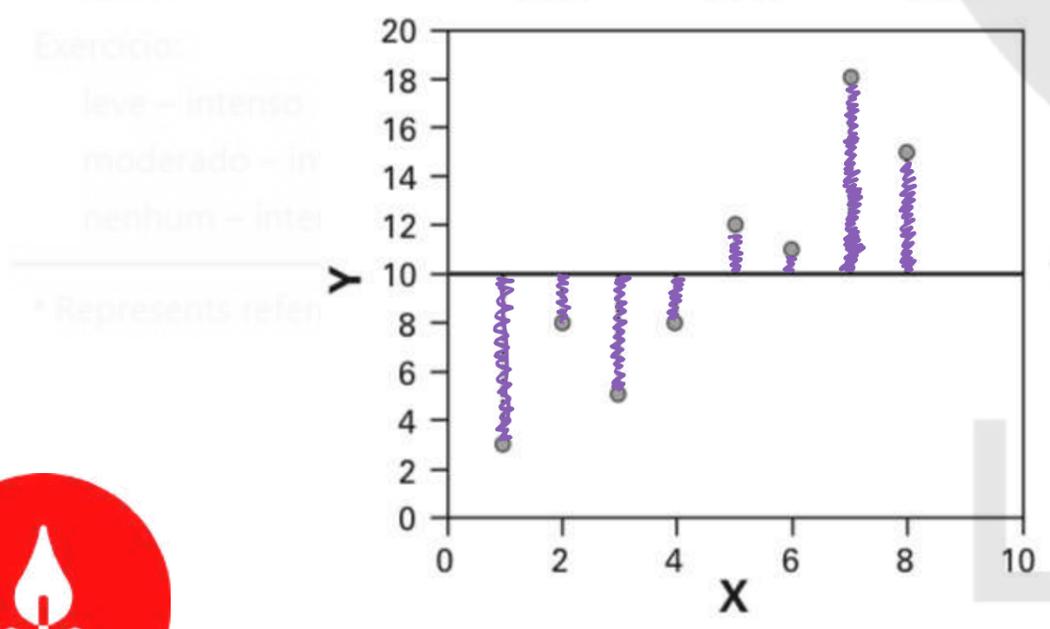
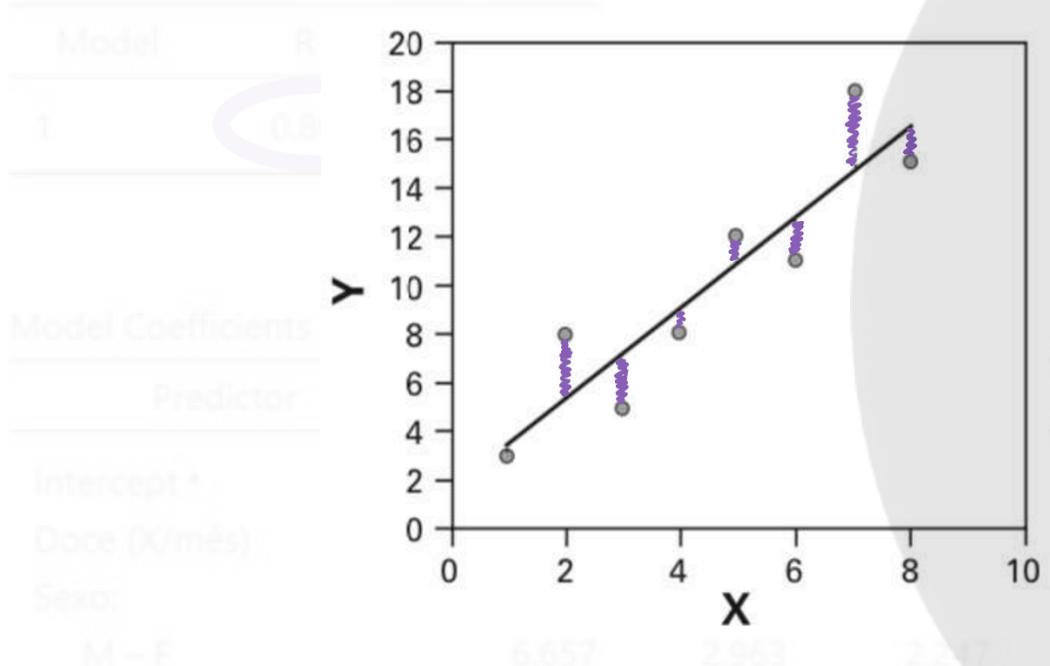
- ~~Cálculo dos quadrados dos resíduos~~
- Soma das distâncias das observações para encontrar a reta que melhor explica o modelo.

R:  
"Quanto **maior** a soma,  
**pior** a representação!"

$R^2$  → O que ocorre na **média**



# Método dos mínimos quadrados



- Melhor modelo
- Soma total dos Resíduos (SSR)

Model Coefficients

Predictor

Intercept *	6.657	2.963	2.247
Doce (X/mês)			
Sexo:			
M - F			
Exercício:			
leve - intenso			
moderado - in			
nenhum - intel			

\* Represents refer

p	< .001
	0.002
	0.027
	0.019
	0.743
	< .001

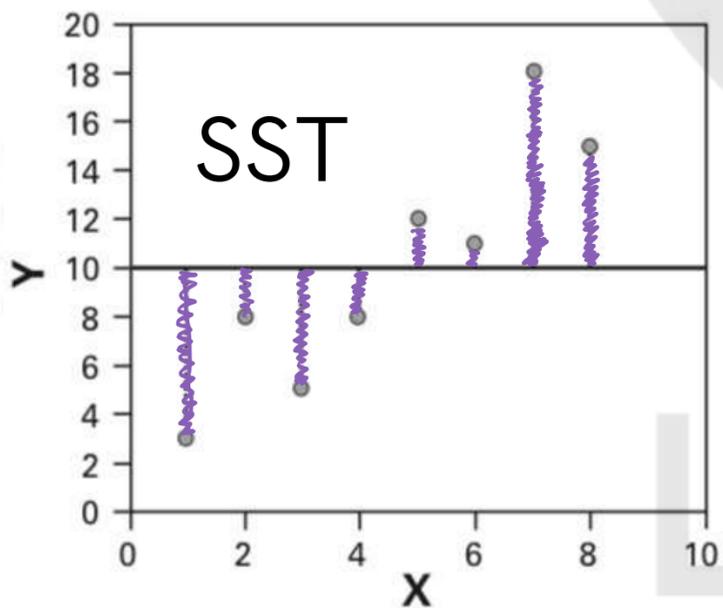
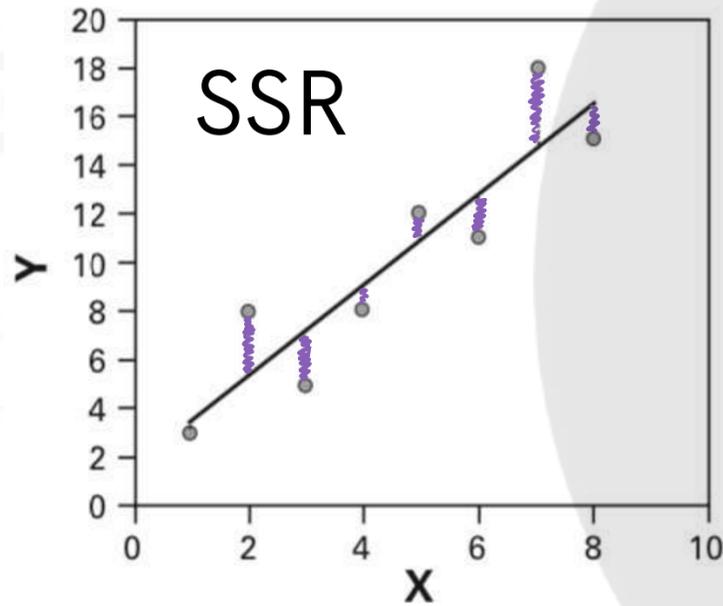
R:

- Coeficiente de correlação
- Análise de correlação de Pearson

- Média
- Soma total dos Quadrados (SST)



# Método dos mínimos quadrados



- Melhor modelo
- Soma total dos Resíduos (SSR)

$$R = \frac{SSR - SST}{SST}$$

- Média
- Soma total dos Quadrados (SST)



## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>
1	0.805	0.648

### Model Coefficients - Peso

Predictor	Estimate	SE	t	p
Intercept *	48.943	6.059	8.078	< .001
Doce (X/mês)	0.688	0.215	3.196	0.002
Sexo:				
M – F	6.657	2.963	2.247	0.027
Exercício:				
leve – intenso	10.694	4.466	2.394	0.019
moderado – intenso	-1.353	4.112	-0.329	0.743
nenhum – intenso	24.521	4.243	5.779	< .001

\* Represents reference level

# Demais coeficientes

p:

- Teste F
- Modelo é significativo?
- $p < 0,05$

R:

- Coeficiente de correlação
- Análise de correlação de Pearson

R<sup>2</sup>: ?



## Linear Regression

### Model Fit Measures

Model	R	R <sup>2</sup>
1	0.805	0.648

### Model Coefficients - Peso

Predictor	Estimate	SE	t	p
Intercept *	48.943	6.059	8.078	< .001
Doce (X/mês)	0.688	0.215	3.196	0.002
Sexo:				
M – F	6.657	2.963	2.247	0.027
Exercício:				
leve – intenso	10.694	4.466	2.394	0.019
moderado – intenso	-1.353	4.112	-0.329	0.743
nenhum – intenso	24.521	4.243	5.779	< .001

\* Represents reference level

# Demais coeficientes

**p:**

- Teste F
- Modelo é significativo?
- $p < 0,05$

**R:**

- Coeficiente de correlação
- Análise de correlação de Pearson

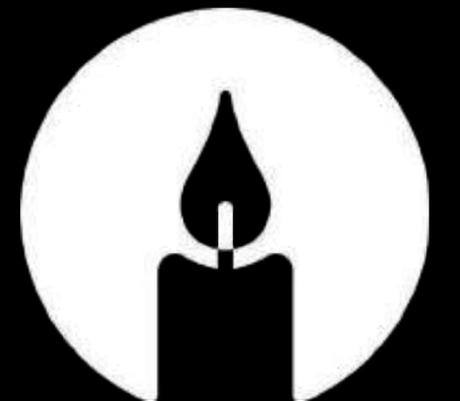
**R<sup>2</sup>:**

- Coeficiente de determinação
- Coeficiente de ajuste do modelo
- Quanto maior, melhor



# Regressão linear múltipla

**#02** Módulo de Bioestatística



# Regressão linear simples

X

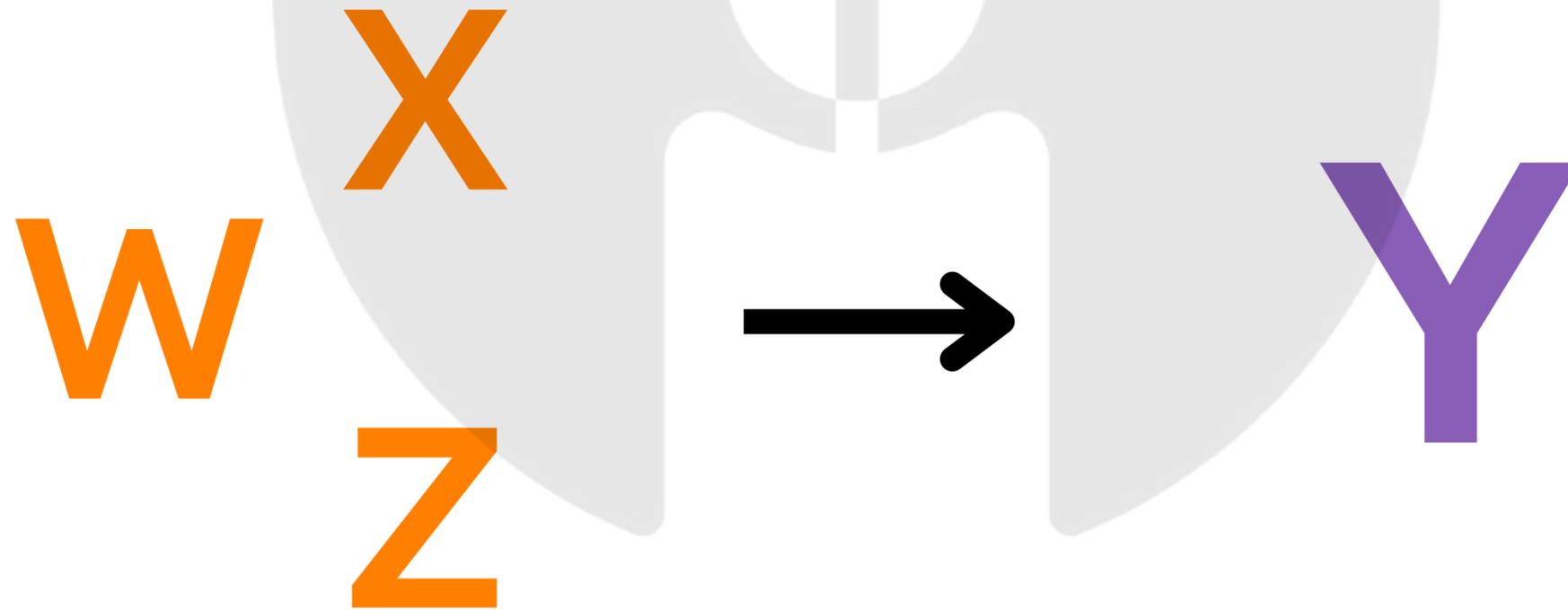


Y

LAC-FMB



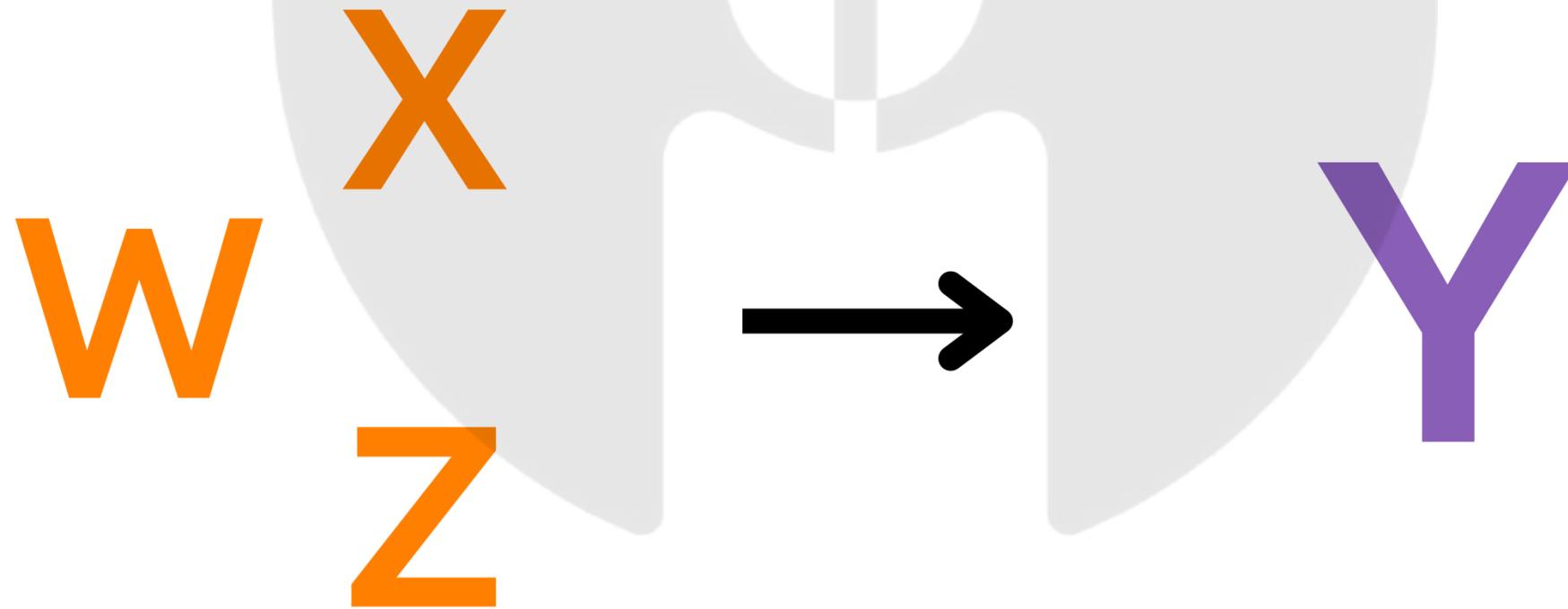
# Regressão linear múltipla



LAC-FMB



# Regressão linear múltipla



LAC-FMB

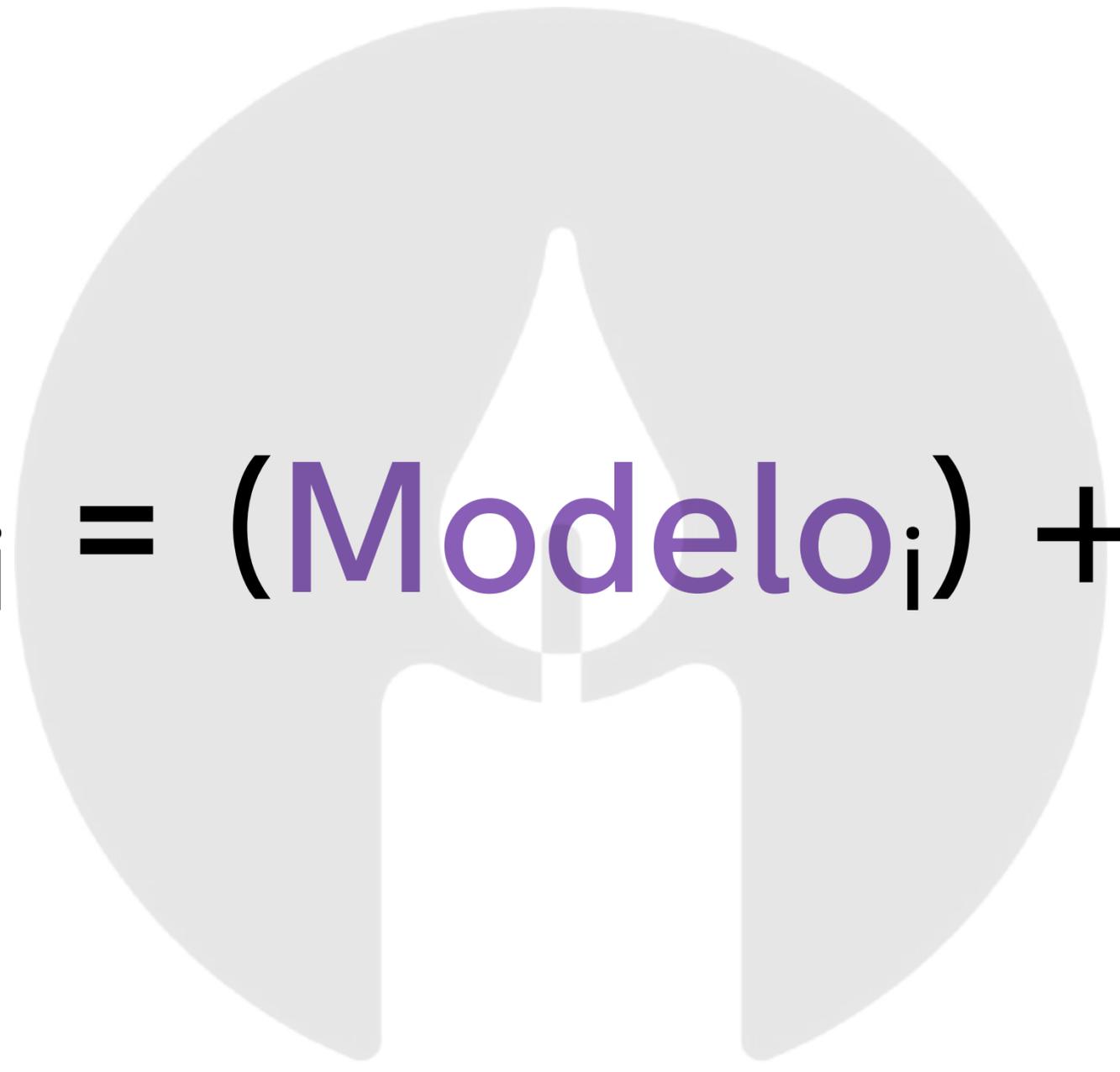




Guache

Como escrever a equação  
que determina uma  
regressão múltipla?




$$\text{Saída}_i = (\text{Modelo}_i) + \text{Erro}_i$$

LAC-FMB



$$\text{Saída}_i = (\text{Modelo}_i) + \text{Erro}_i$$

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n) + \varepsilon_i$$

LAC-FMB

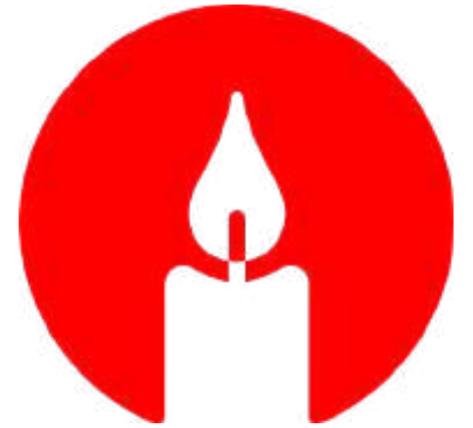


**NÃO ENTENDI NADA  
SOCORRO!!!!**

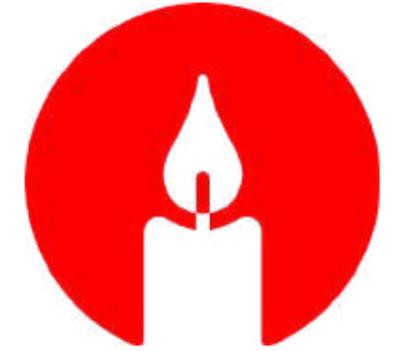
LAC-FMB



**Calma!**



LAG FMAR



Queremos prever a **quantidade de vezes que se fala sobre aviões** na LAC por ano...

LAC FMD

# Preditor 1

Quantidade de membros fãs de avião



# Preditor 2

Quantidade de voos pegos pelos membros



LAC-FMB



# Modelo de previsão

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

LAC-FMB



# Modelo de previsão

$$\underline{Y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Conversas =

LAC-FMB



# Modelo de previsão

$$Y = \underline{b_0} + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Conversas =  $b_0$

LAC-FMB



# Modelo de previsão

$$Y = b_0 + \underline{b_1 X_1} + b_2 X_2 + \dots + b_n X_n$$

$$\text{Conversas} = b_0 + \underline{b_1 \times \text{fãs de avião}}$$

LAC-FMB



# Modelo de previsão

$$Y = b_0 + b_1X_1 + \underline{b_2X_2} + \dots + b_nX_n$$

Conversas =  $b_0 + b_1 \times \text{fãs de avião} + \underline{b_2 \times \text{n}^\circ \text{ de voos}}$

LAC-FMB



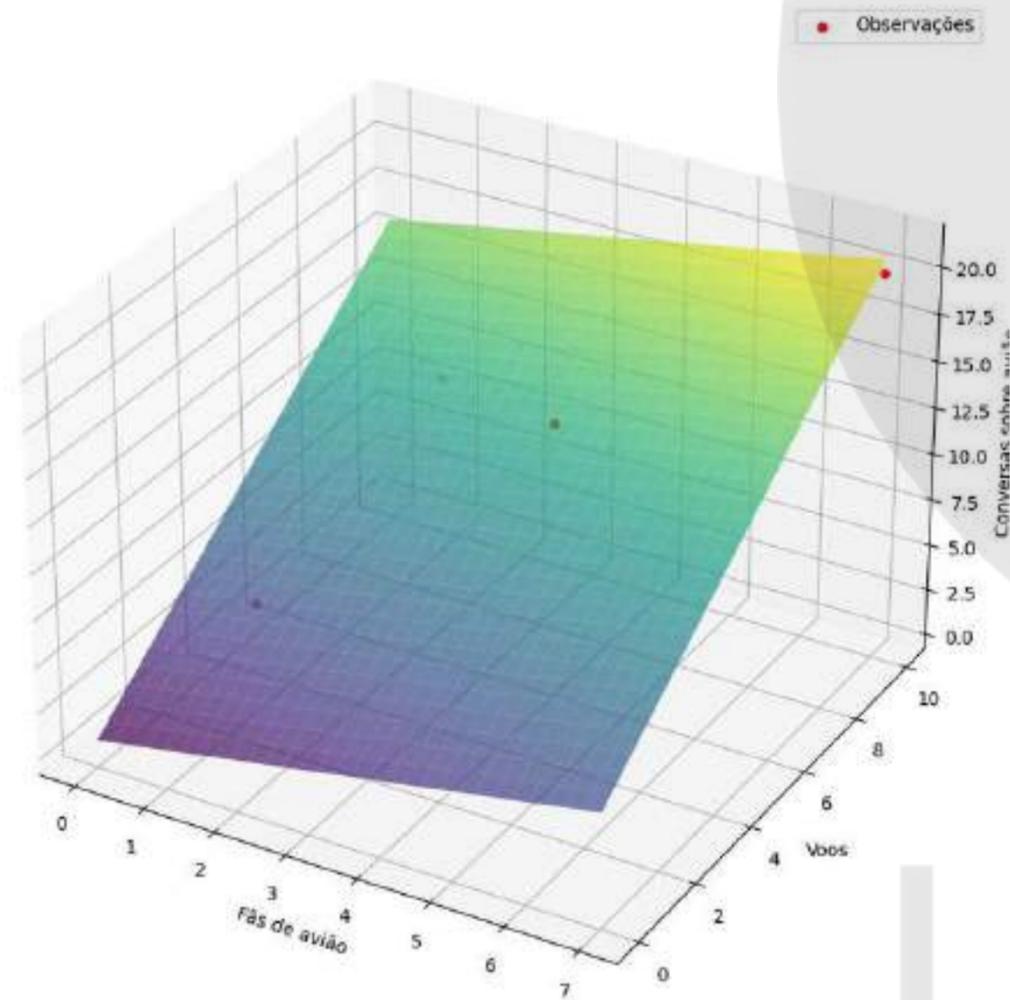
# Modelo de previsão

Ano	Fãs de avião	Voos	Conversas
2018	1	3	5
2019	2	7	13
2020	4	6	14
2021	7	10	20



# Modelo de previsão

Como prever a quantidade de conversar sobre aviões?

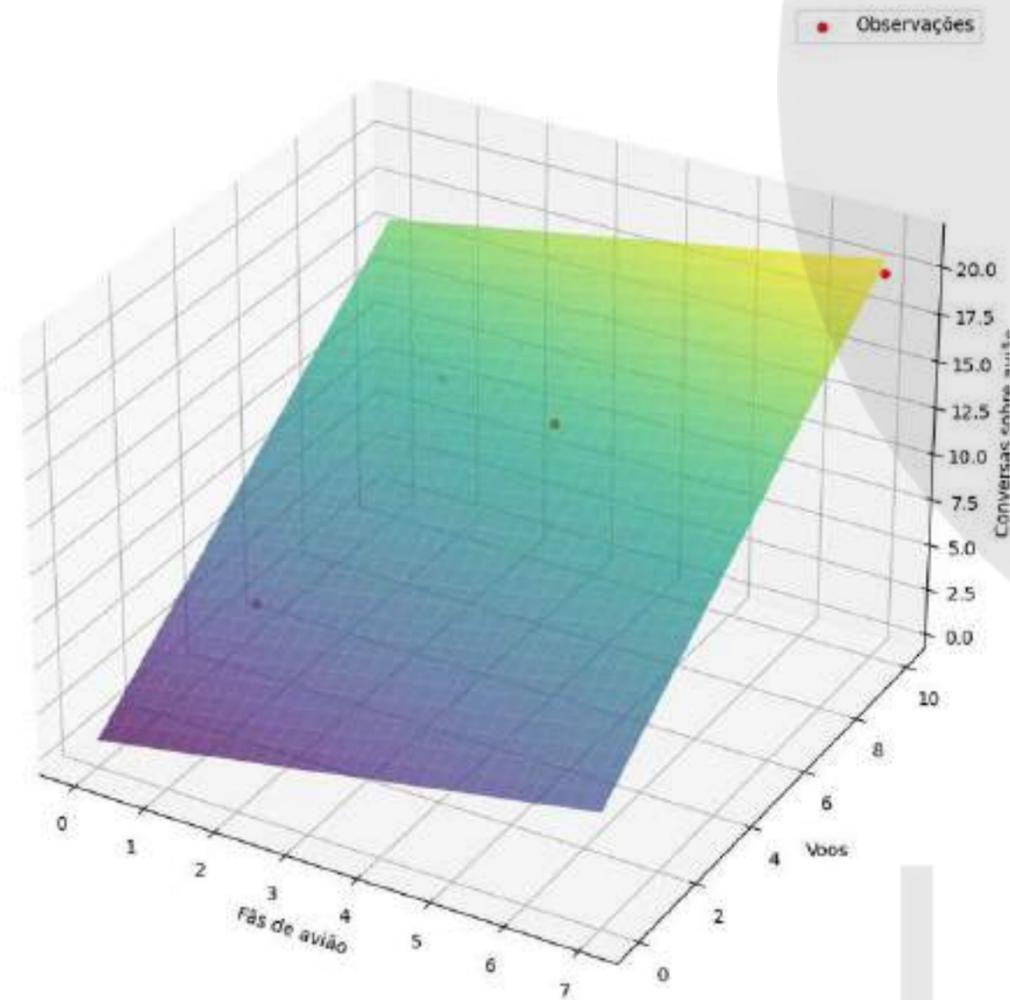


LAC-FMB



# Modelo de previsão

Como prever a quantidade de conversar sobre aviões?



$$\text{Conversas} = 0,6 + 0,7 \times \text{fãs de avião} + 1,5 \times \text{n}^\circ \text{ de voos}$$

LAC-FMB



**E agora?**



LAG FMD

# E agora?



Quais as formas de fazer uma regressão linear múltipla?



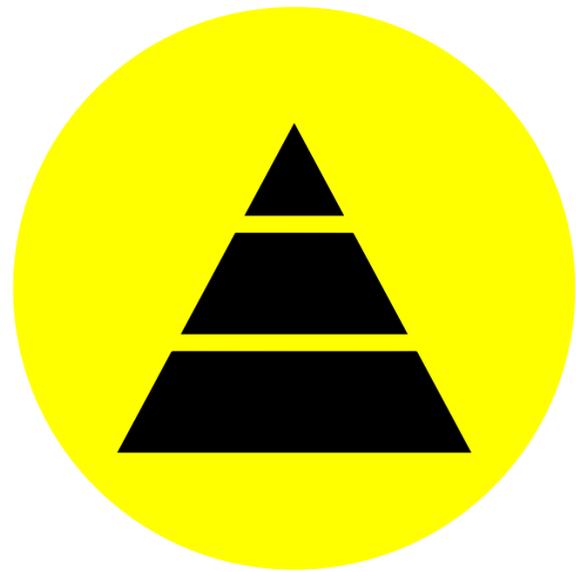
LAG FMD

# Métodos de regressão

LAC-FMB



# Métodos de regressão

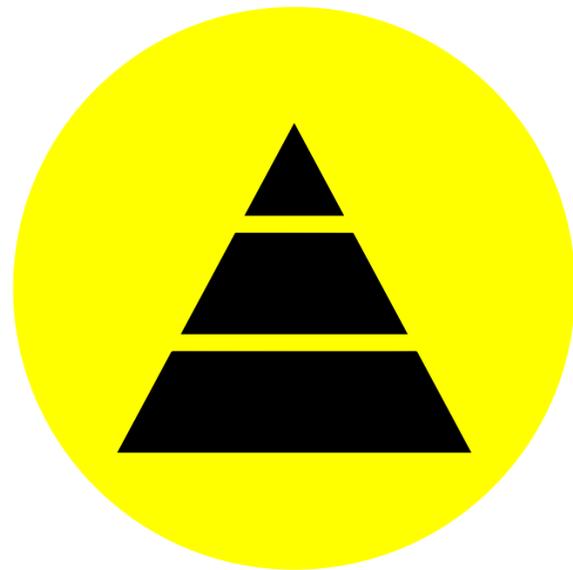


Hierárquico

LAC-FMB



# Métodos de regressão



Hierárquico

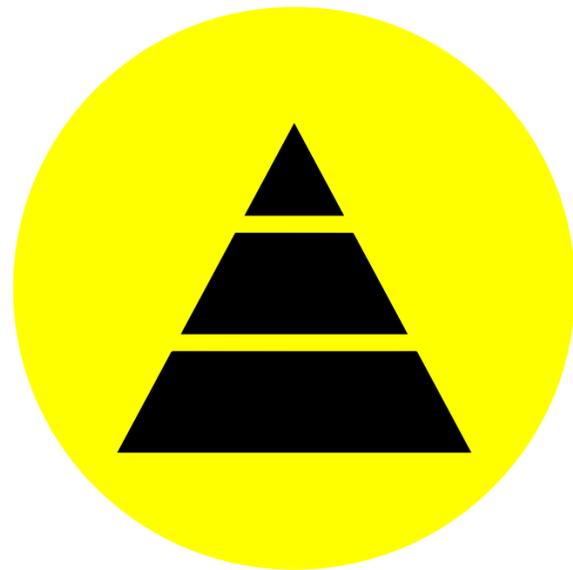


Entrada forçada

LAC-FMB



# Métodos de regressão



Hierárquico



Entrada forçada



Stepwise  
(passo a passo)

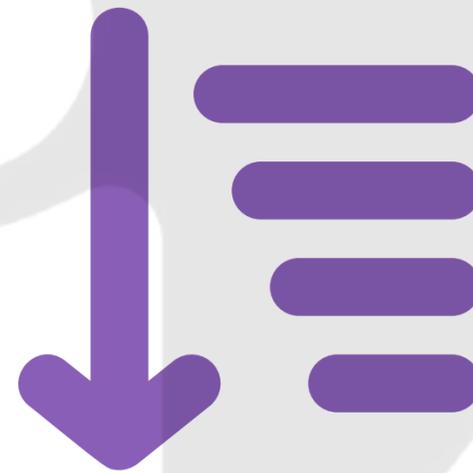
LAC-FMB





# Métodos de regressão

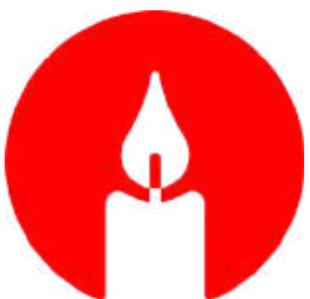
Hierárquico

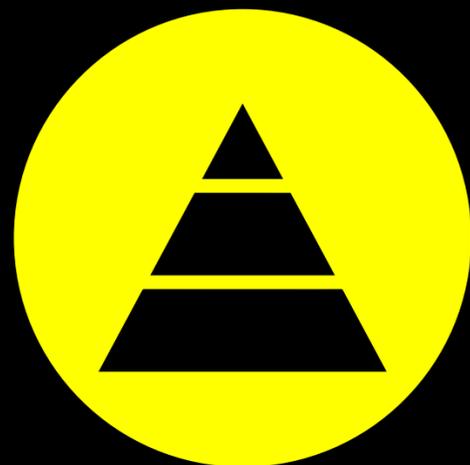


Preditores adicionados por  
ordem de **importância**

LAC-FMB

Liga Acadêmica de Ciências





# Métodos de regressão

Hierárquico

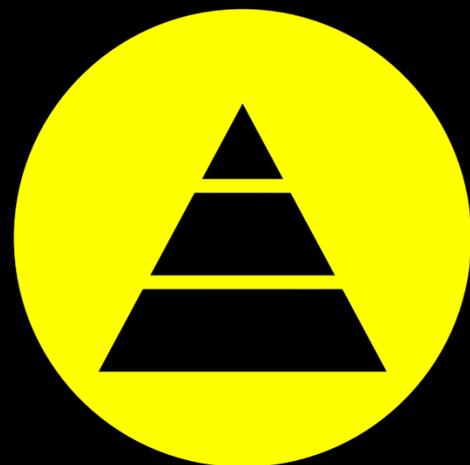


Baseado em estudos  
anteriores

LAC-FMB

Liga Acadêmica de Ciências





# Métodos de regressão

Hierárquico

## Ordem de inserção no modelo

1. Previsores conhecidos
2. Novos previsores

LAC-FMB

Liga Acadêmica de Ciências

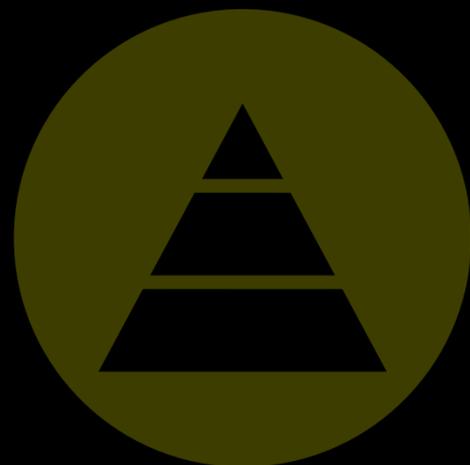


# Métodos de regressão

Hierárquico

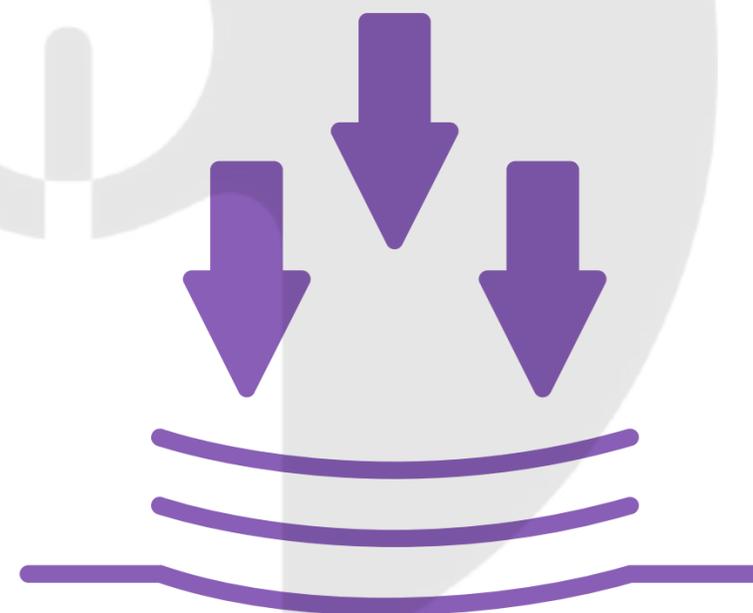
Novos previsores podem ser adicionados por **entrada forçada**, **stepwise** ou **hierarquicamente**.





# Métodos de regressão

Entrada forçada



Todos os previsores são forçados  
no modelo **ao mesmo tempo**

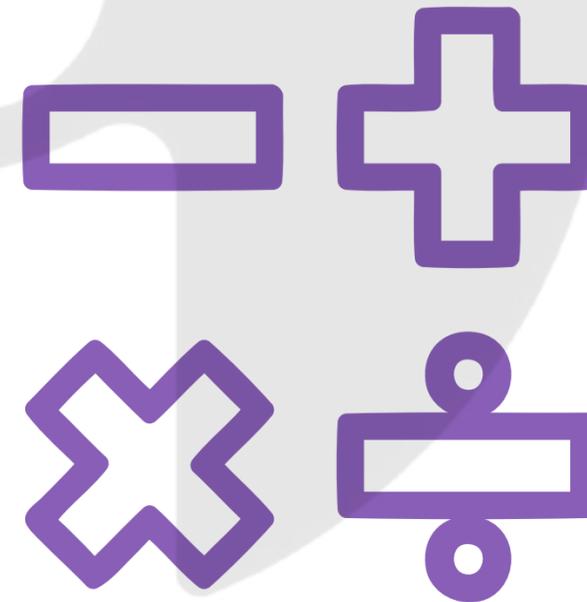
LAC-FMB

Liga Acadêmica de Ciências

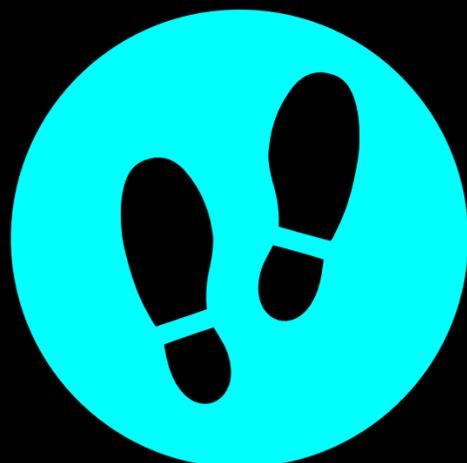
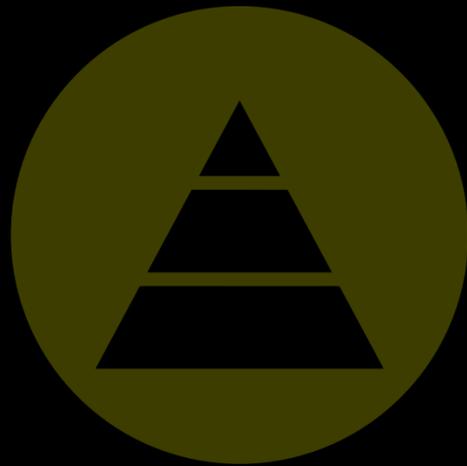
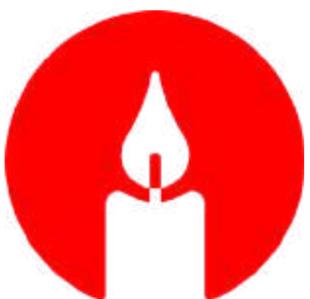


# Métodos de regressão

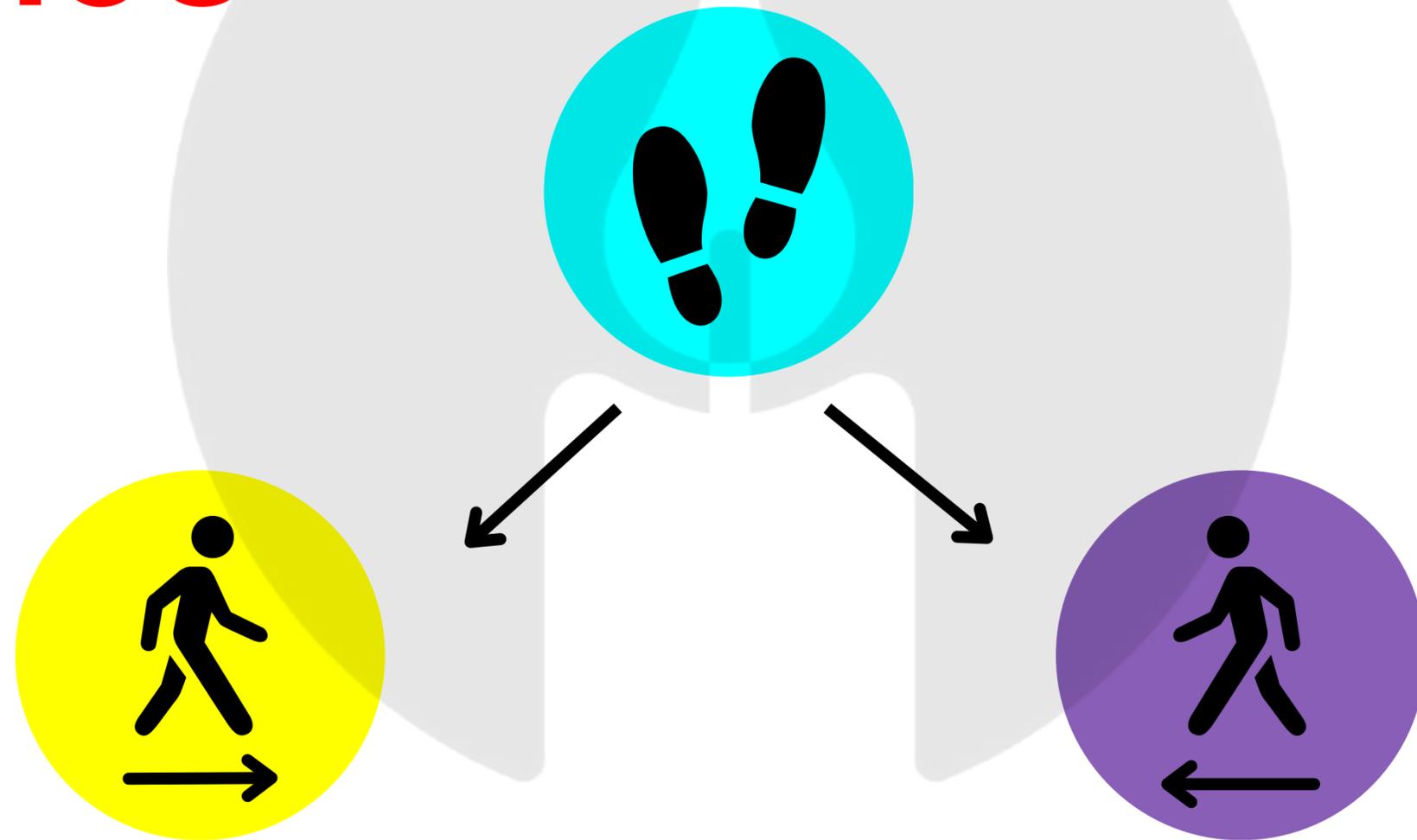
Stepwise



A ordem de entrada dos previsores é baseada em **critérios matemáticos**



# Stepwise



Forward  
(para frente)

Backward  
(para atrás)

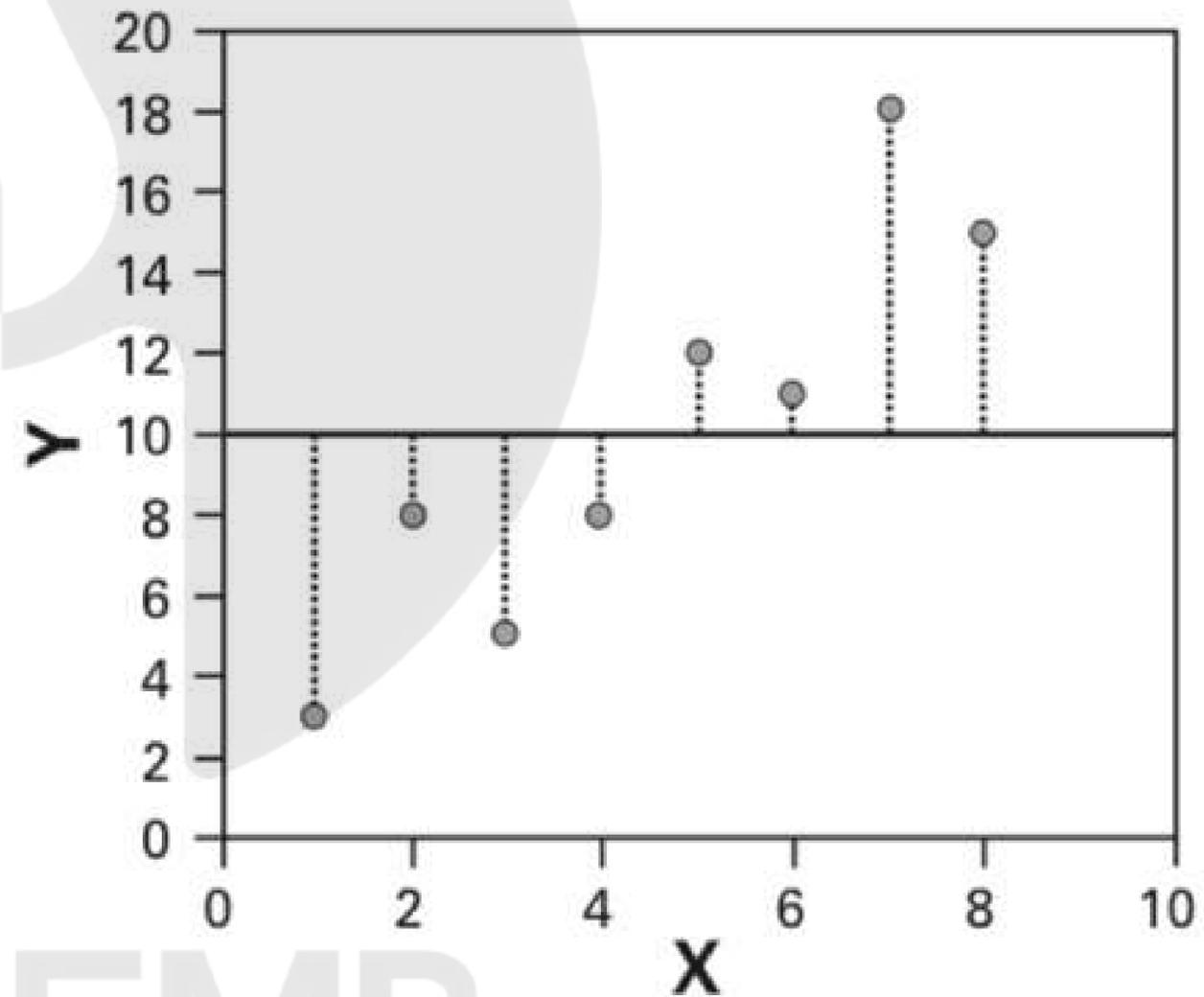


# Stepwise

## Forward

1. Modelo inicial apenas com  $b_0$  definido

$$Y = b_0$$



LAC-FMB

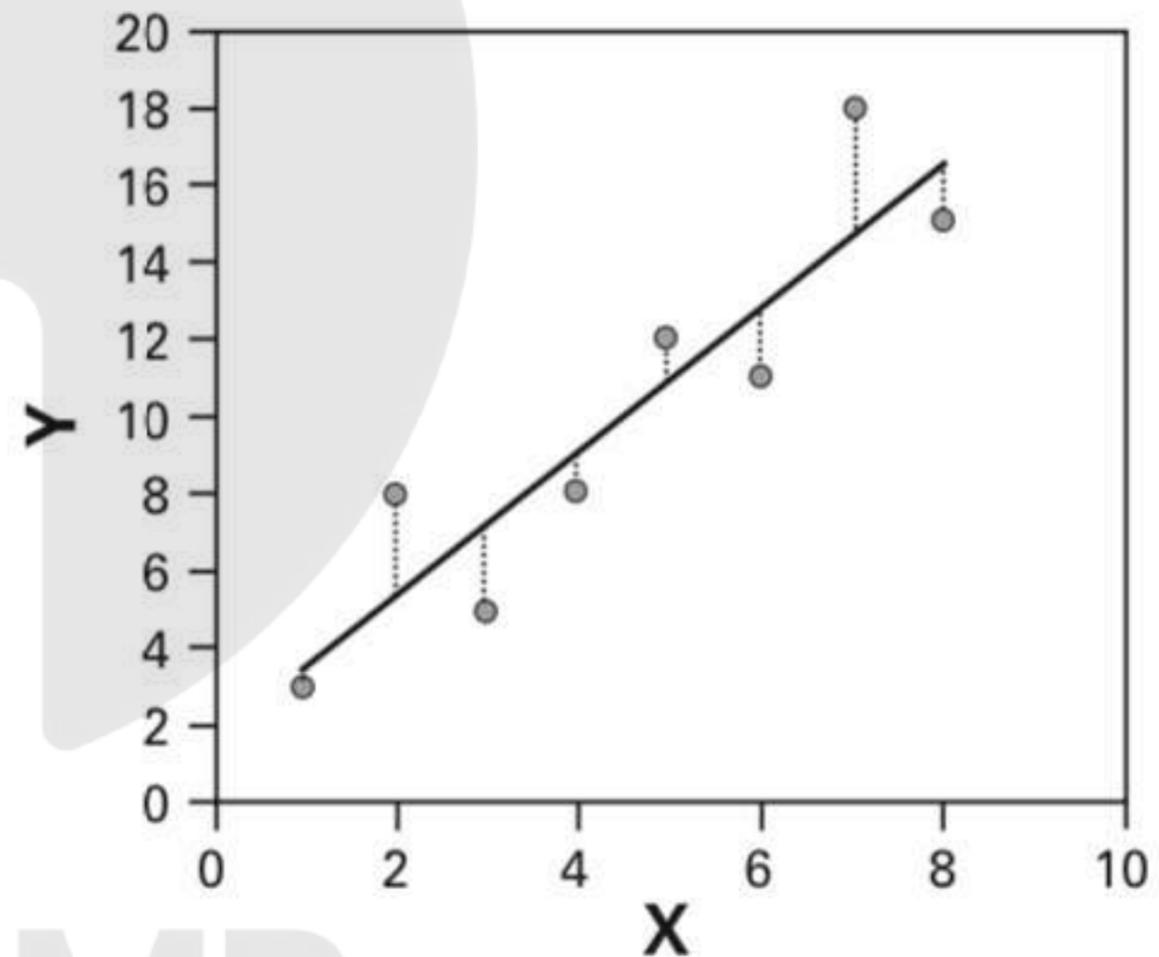


# Stepwise

## Forward

2. Previsor com maior  
coeficiente de correlação (R)  
entra no modelo

$$Y = b_0 + b_1 X_1$$



LAC-FMB

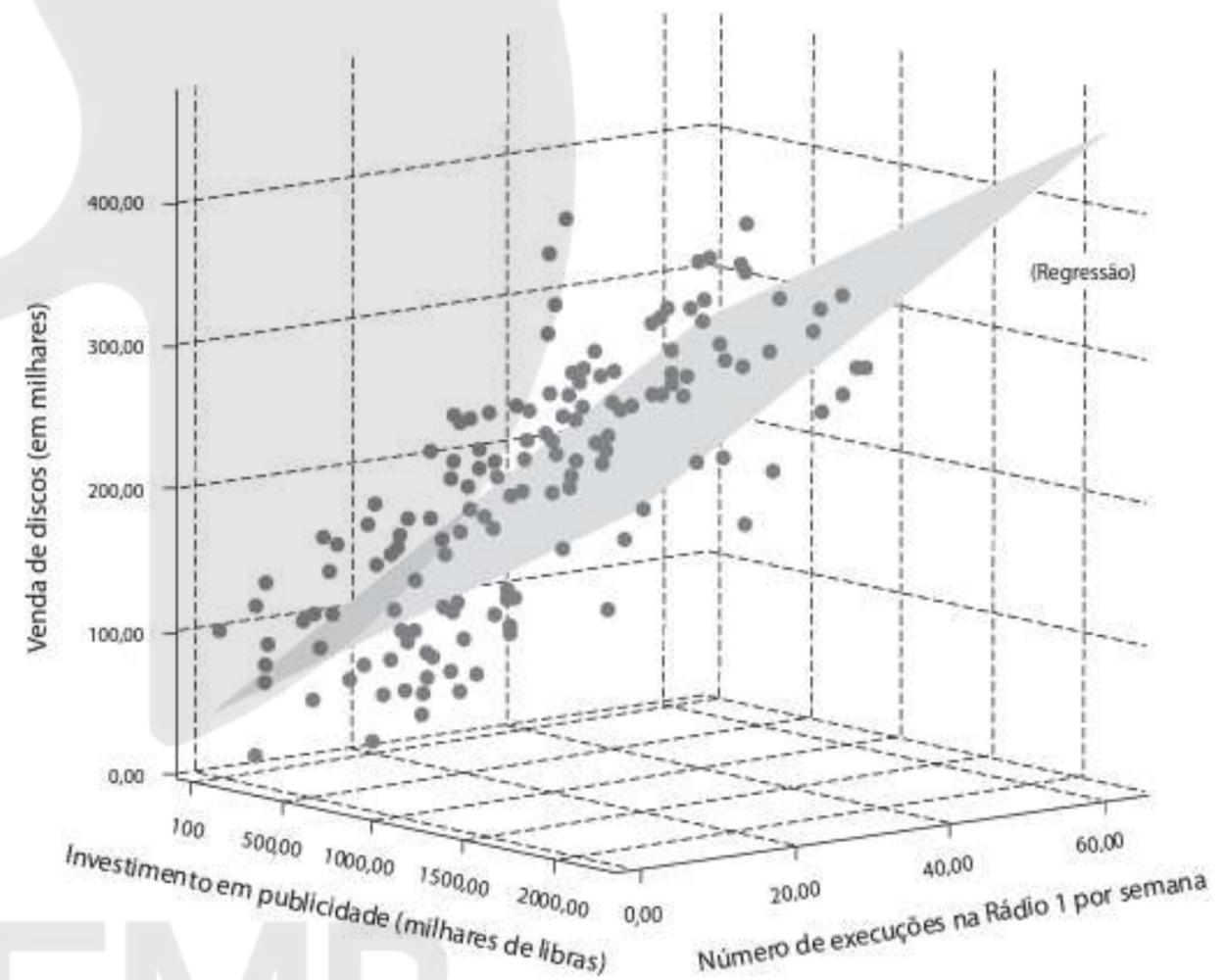


# Stepwise

## Forward

3. Previsor com **maior correlação semiparcial** entra no modelo

$$Y = b_0 + b_1X_1 + b_2X_2$$



LAC-FMB

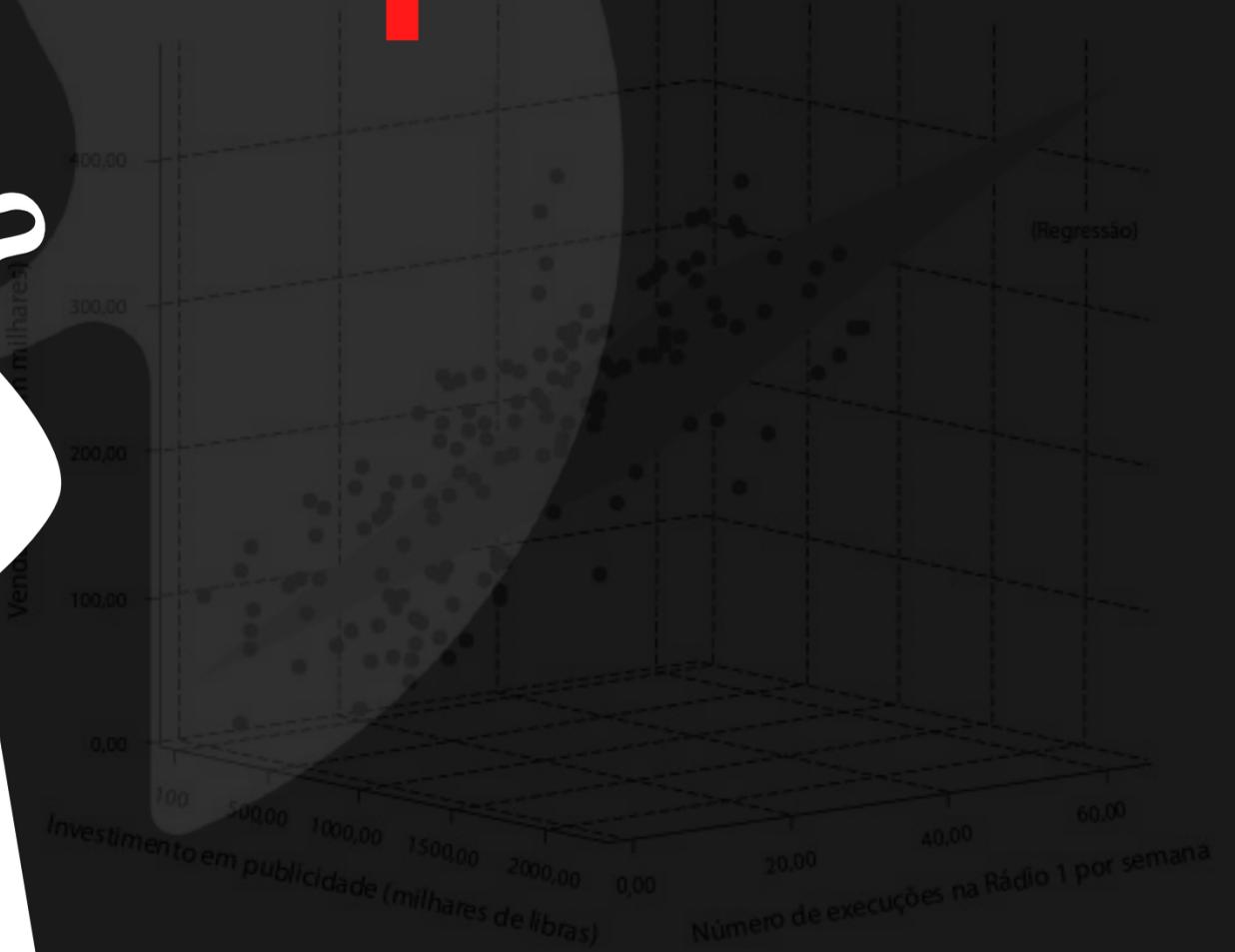


Stepwise

# Correlação

$Y = b_0 + b_1 X_1 + b_2 X_2$   
**semiparcial?!?**

3. Previsor com maior correlação semiparcial entra no modelo



# Stepwise

Forward

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

Digamos que o previsor 1

tem a maior correlação semiparcial entre

o modelo explicou **30%** do modelo



LAC-FMB



Stepwise

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

Precisamos que os outros

3. Previsor com maior

previsores expliquem os

no modelo

70% restantes



# Stepwise

## Forward

4. Previsor com maior correlação semiparcial entra no modelo

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

LAC-FMB



# Stepwise

## Backward

1. Insere **todas** as variáveis no modelo



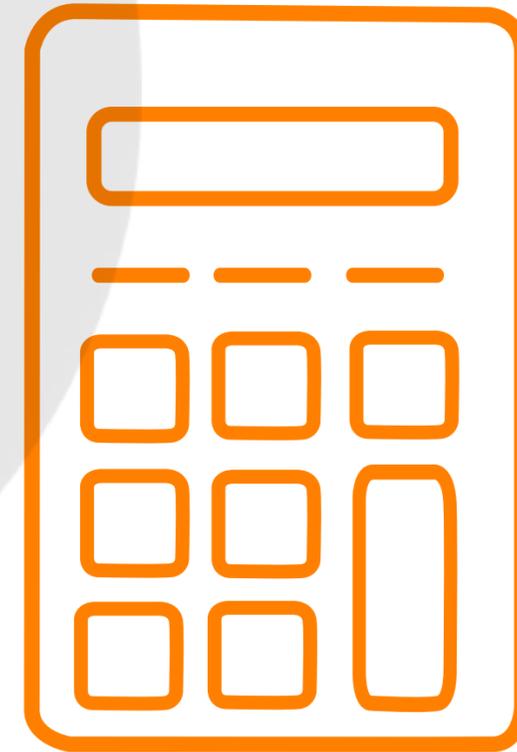
LAC-FMB



# Stepwise

## Backward

1. Insere **todas** as variáveis no modelo
2. Calcula a **contribuição** de cada previsor com um **teste t**



LAC-FMB



# Stepwise

## Backward

3. Se o previsor **não** faz uma **contribuição significativa**, ele é **removido**

LAC-FMB

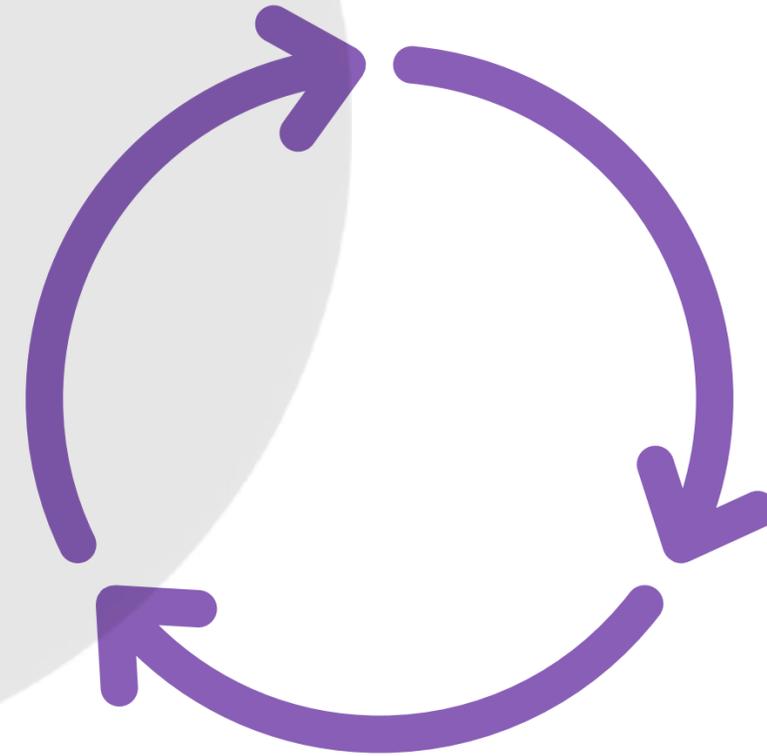


# Stepwise

## Backward

3. Se o previsor não faz uma contribuição significativa, ele é removido

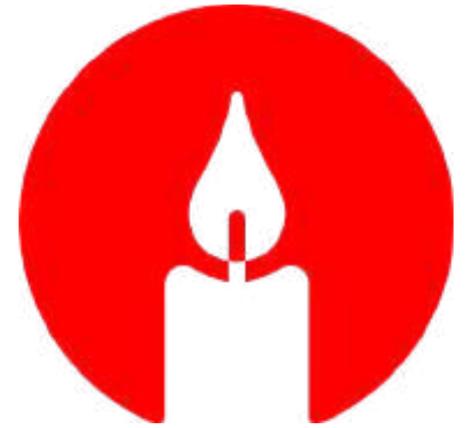
4. Modelo é **reestimado** com os previsores restantes



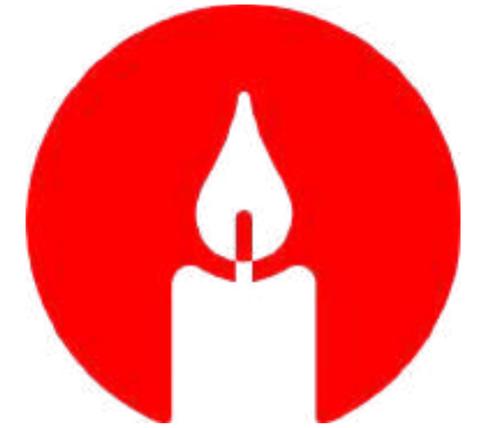
LAC-FMB



Foward ou  
backward?



LAC EMR



O método **backward** é preferível ao  
método forward

LAC FMD



O método **foward** aumenta a  
probabilidade de **erro tipo II**

LAC FMD



Samira

O que é erro  
tipo II?

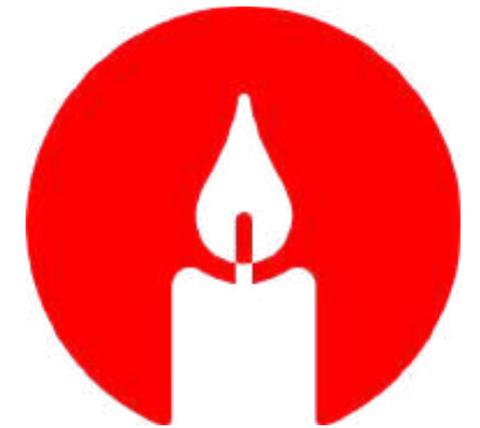
LAC-FMB



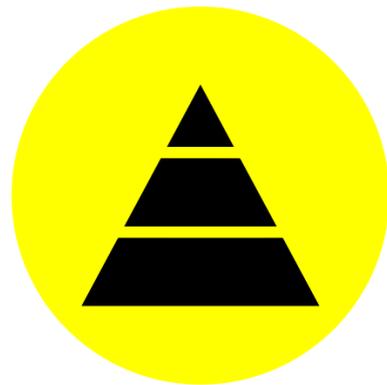
	Aceitar $H_0$	Rejeitar $H_0$
$H_0$ verdadeira	✓	Erro tipo 1
$H_0$ falsa	Erro tipo 2	✓

O método forward aumenta a probabilidade de erro tipo II

Forward tende a **excluir**  
**previsores que teriam um efeito**  
**significativo** quando outra variável  
fosse constante



# Qual método escolher?



Hierárquico

vs



Entrada forçada

vs



Stepwise  
(passo a passo)



LAC-FVIB



# Qual método escolher?

Hierárquico



VS

Entrada forçada



VS

Stepwise



Entrada forçada pode ser utilizada quando temos uma boa base teórica para todos os previsores

LAC-FMB



# Qual método escolher?

Hierárquico

VS

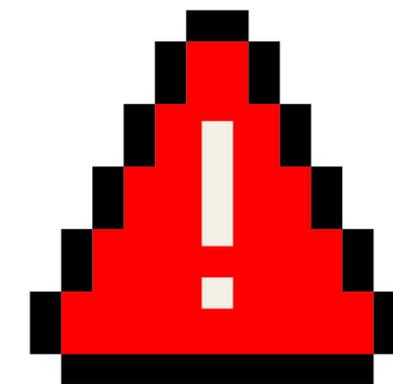
Entrada forçada

VS

Stepwise



Geralmente não temos  
uma boa base teórica para  
todos os previsores



LAC-FMB



# Qual método escolher?

Hierárquico



VS

Entrada forçada

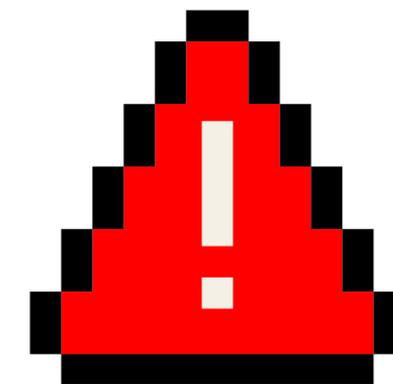


VS

Stepwise



Geralmente não temos  
uma boa base teórica para  
todos os previsores



LAC-FMB



# Qual método escolher?

Hierárquico



VS

Entrada forçada



VS

Stepwise



**Stepwise** tira do pesquisador decisões metodológicas importantes

LAC-FMB



# Qual método escolher?

Hierárquico



VS

Entrada forçada



VS

Stepwise



Pode incluir variáveis por pequenas diferenças nas correlações semiparciais

LAC-FMB



# Qual método escolher?

Hierárquico



VS

Entrada forçada



VS

Stepwise



Pode ser útil em  
análises exploratórias



LAC-FMB



# Qual método escolher?

Hierárquico



VS

Entrada forçada



VS

Stepwise



Sempre que possível, incluir os previsores **hierarquicamente** por **boas razões teóricas**

LAC-FMB



# Qual método escolher?



VS

Entrada forçada



VS

Stepwise



**GAME OVER**

LAC-FMB



# Próximo passo

Generalização do modelo





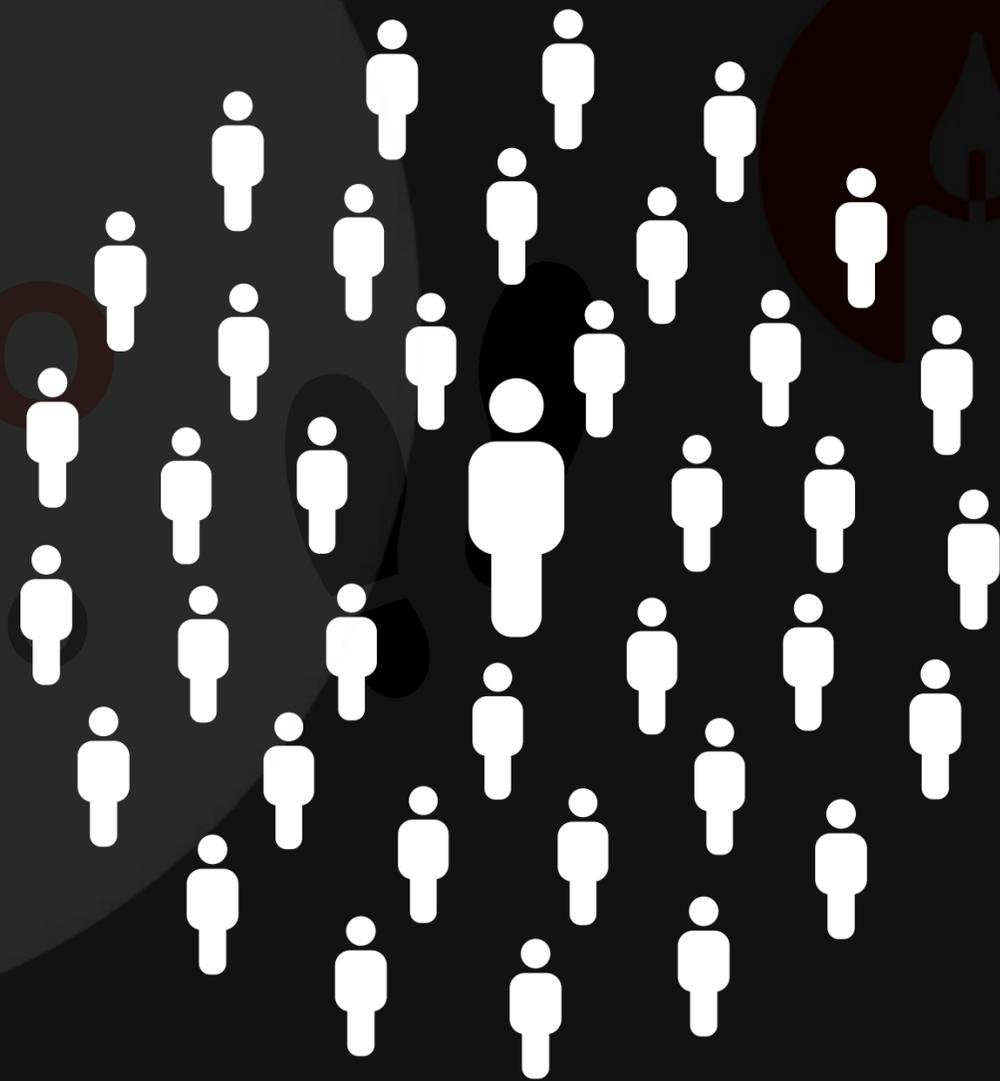
Amanda

O que seria  
generalizar o  
modelo?

# Extrapolação de resultados



Amostra



População

# Generalização



*"Para tirar conclusões sobre uma população com base em um modelo de regressão realizado sobre uma amostra, várias suposições (hipóteses) devem ser verdadeiras"*

# Pressupostos para generalização

## 1. Tipos de variáveis

**X**

Previsoras

Quantitativa

Categórica dicotômica

LAC-FMB



# Pressupostos para generalização

## 1. Tipos de variáveis

**X**

Previsoras

Quantitativa

Categórica dicotômica

**Y**

Saída

Quantitativa

Contínua

Não limitada

LAC-FMB



# Pressupostos para generalização

## 2. Variância não-nula

Os previsores devem ter alguma **variação** nos **valores observados**

LAC-FMB



# Pressupostos para generalização

## 2. Variância não-nula

Os previsores devem ter alguma **variação** nos **valores observados**

Fãs de avião 
2
2
2
2

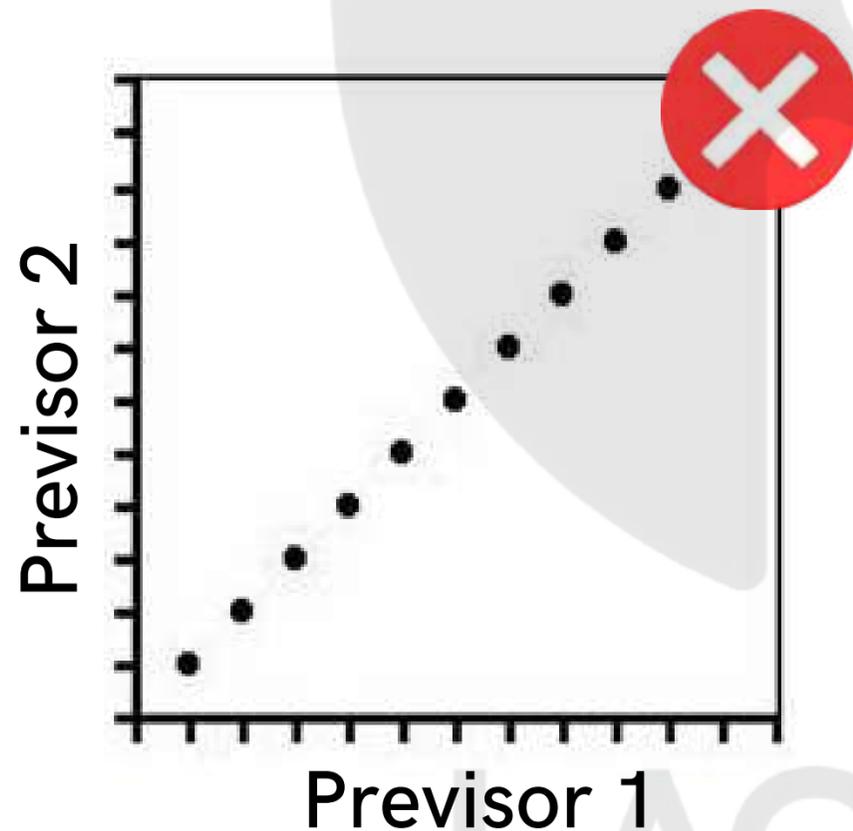
Fãs de avião 
1
2
4
7

LAC-FMB



# Pressupostos para generalização

3. Não existir relação linear perfeita entre dois ou mais previsores

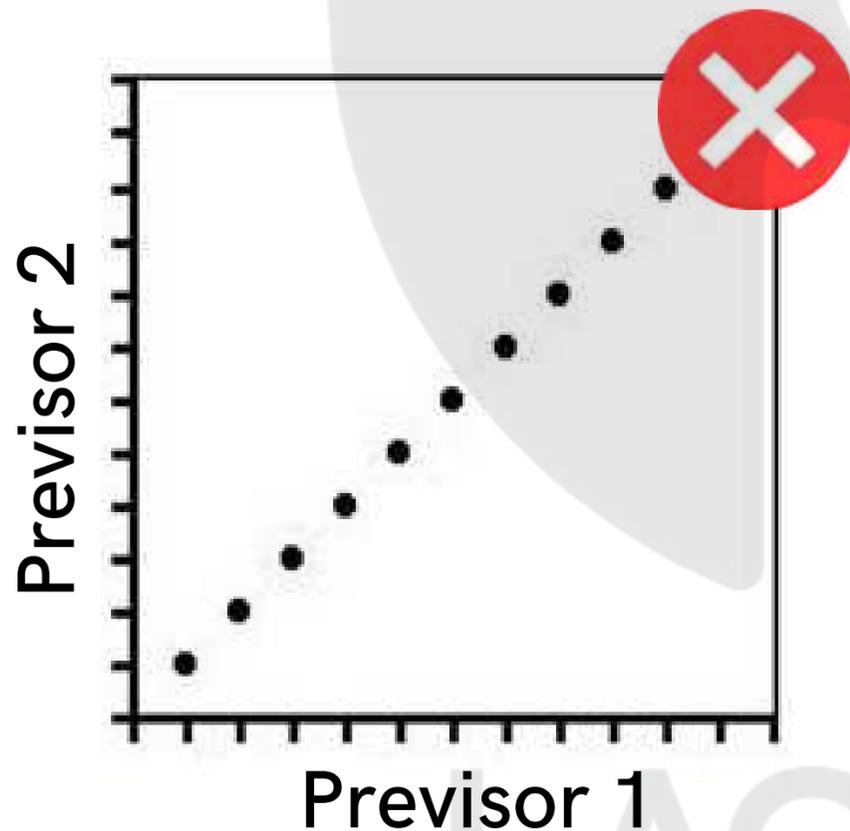


LAC-FMB



# Pressupostos para generalização

3. Não existir relação linear perfeita entre dois ou mais previsores



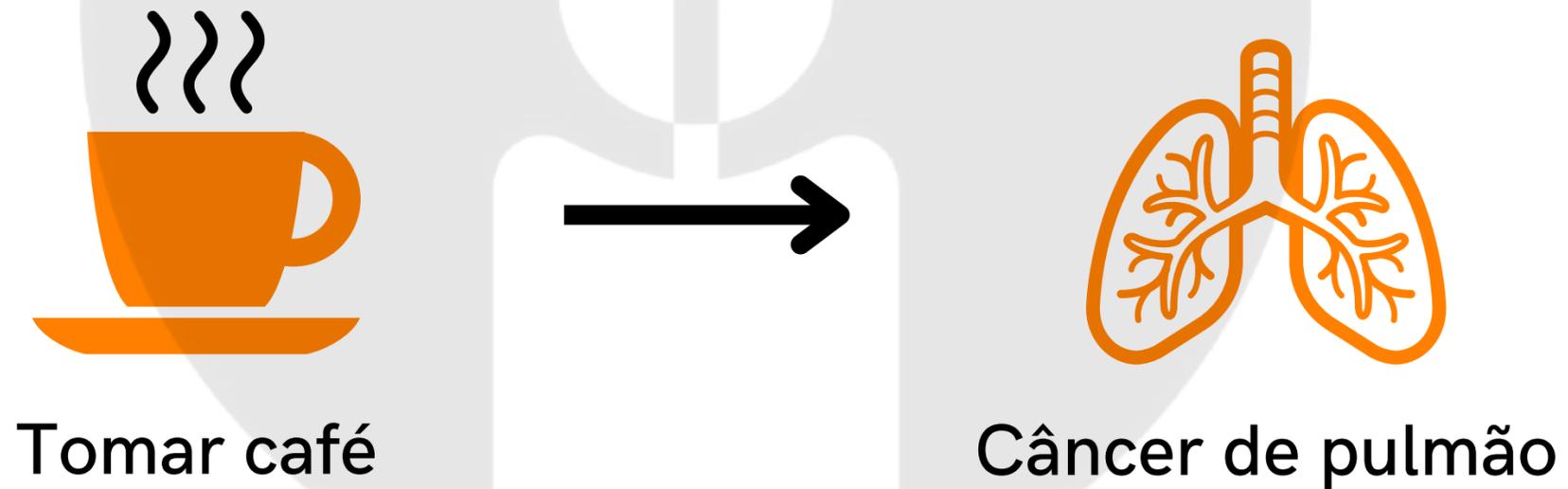
$$R = 1 \text{ X}$$

LAC-FMB



# Pressupostos para generalização

## 4. Previsores não relacionados com variáveis externas

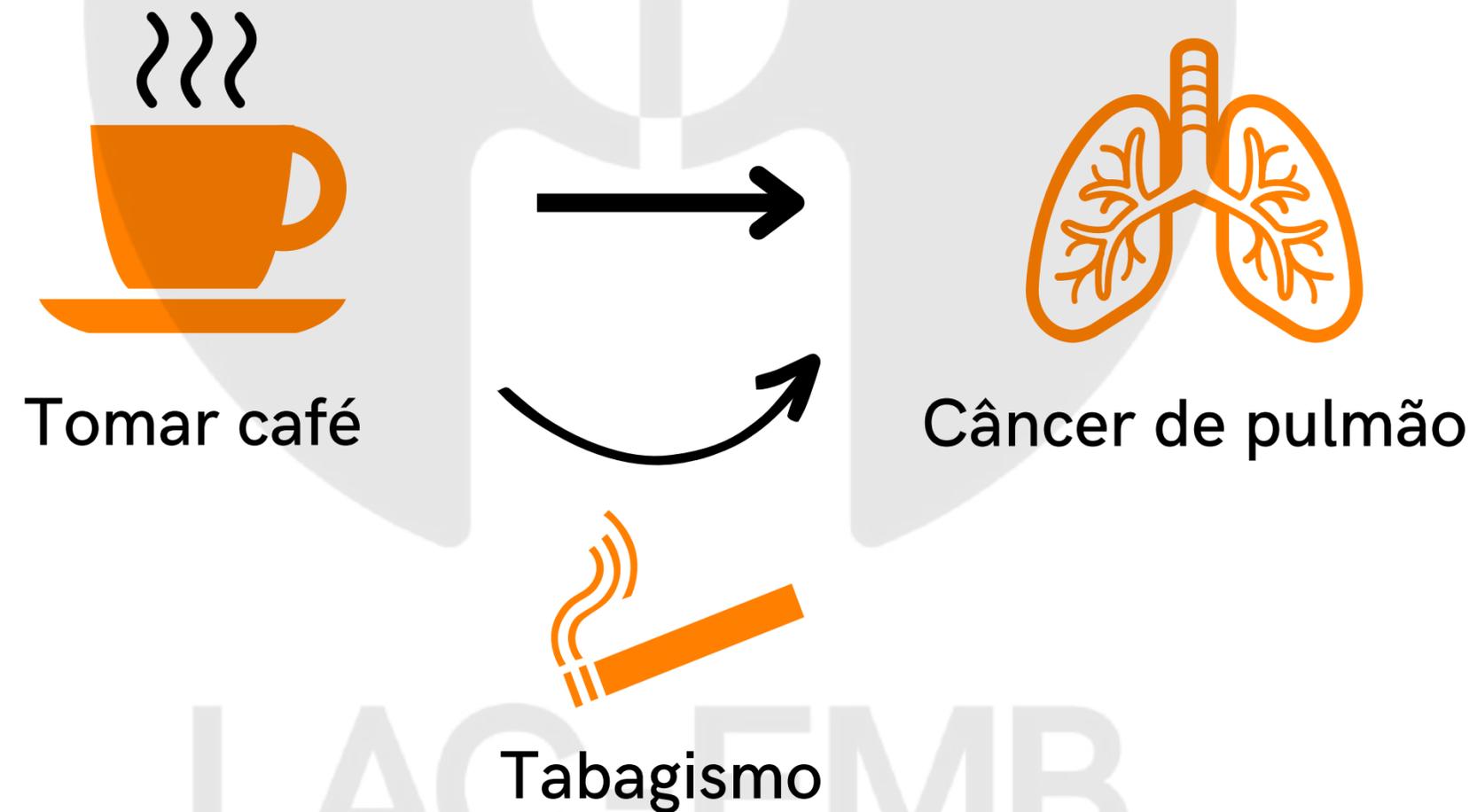


LAC-FMB



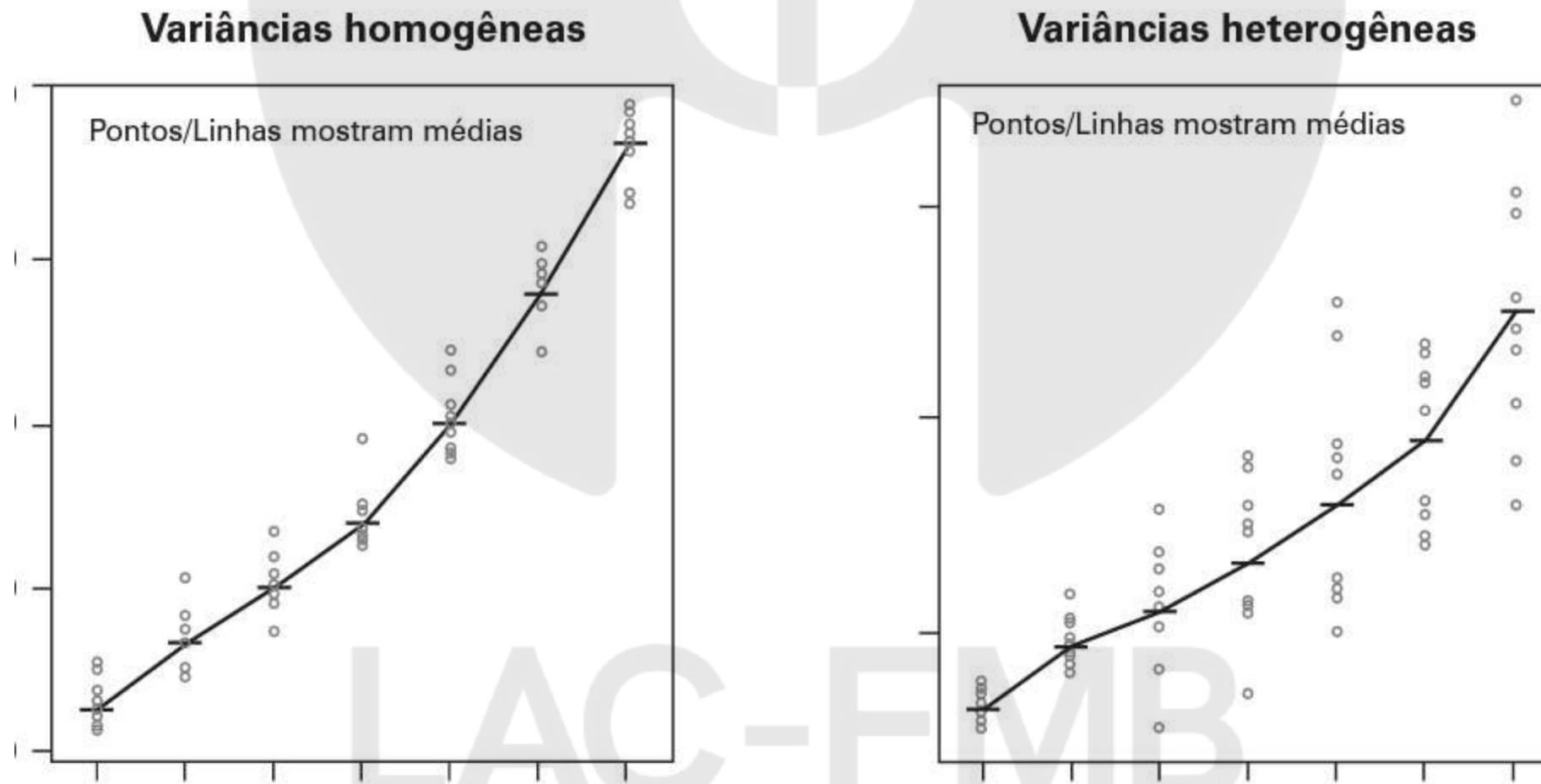
# Pressupostos para generalização

## 4. Previsores não relacionados com variáveis externas



# Pressupostos para generalização

## 5. Homocedasticidade: variâncias homogêneas dos resíduos



# Pressupostos para generalização

## 6. Erros independentes

Para **duas observações**,  
os erros devem ser  
**independentes**

LAC-FMB



# Pressupostos para generalização

## 6. Erros independentes

### Teste de **Durbin-Watson**

Verifica a correlação  
entre os erros

LAC-FMB



# Pressupostos para generalização

## 6. Erros independentes

0 - 4

## Teste de **Durbin-Watson**

Verifica a correlação  
entre os erros

LAC-FMB



# Pressupostos para generalização

## 6. Erros independentes

### Teste de Durbin-Watson

Verifica a correlação  
entre os erros

$$0 - 4$$
$$1 < D < 3$$

LAC-FMB



# Pressupostos para generalização

## 6. Erros independentes

### Teste de Durbin-Watson

Verifica a correlação  
entre os erros

$$0 - 4$$
$$1 < D < 3$$

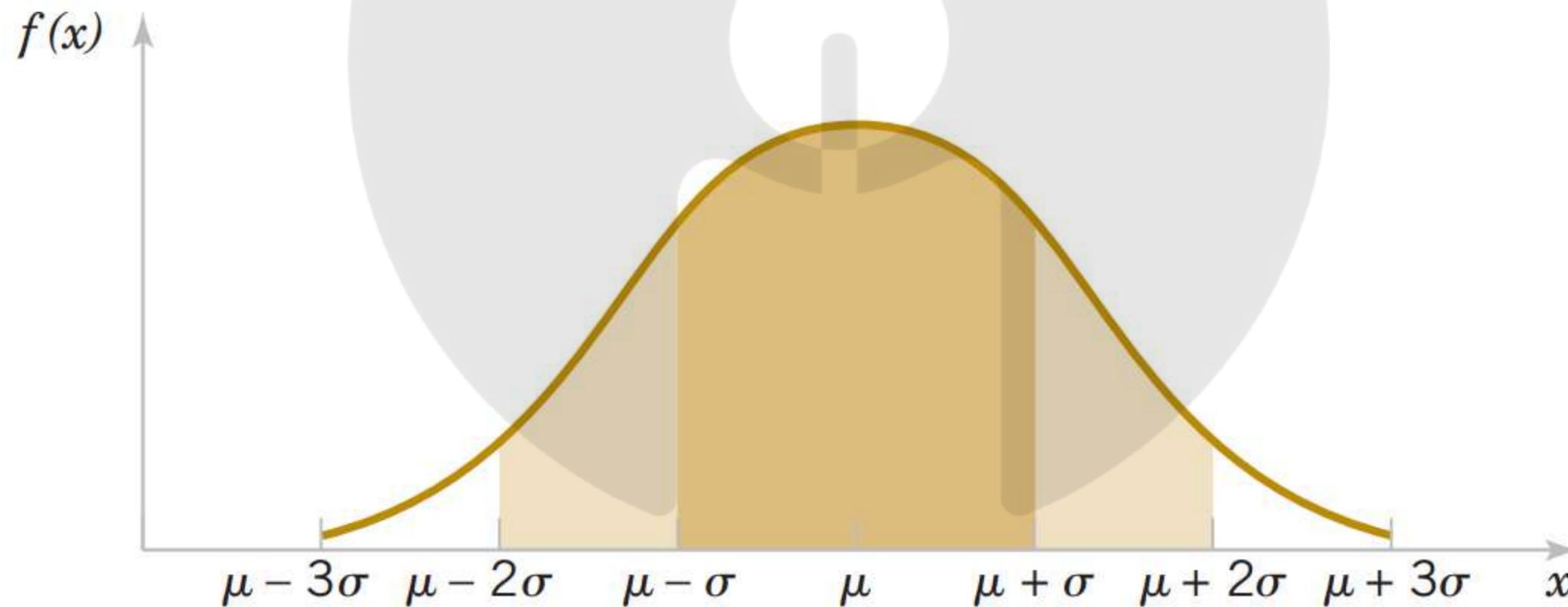
2: sem correlação 

LAC-FMB



# Pressupostos para generalização

## 7. Erros normalmente distribuídos



LAC-FMB



# Pressupostos para generalização

## 8. Independência

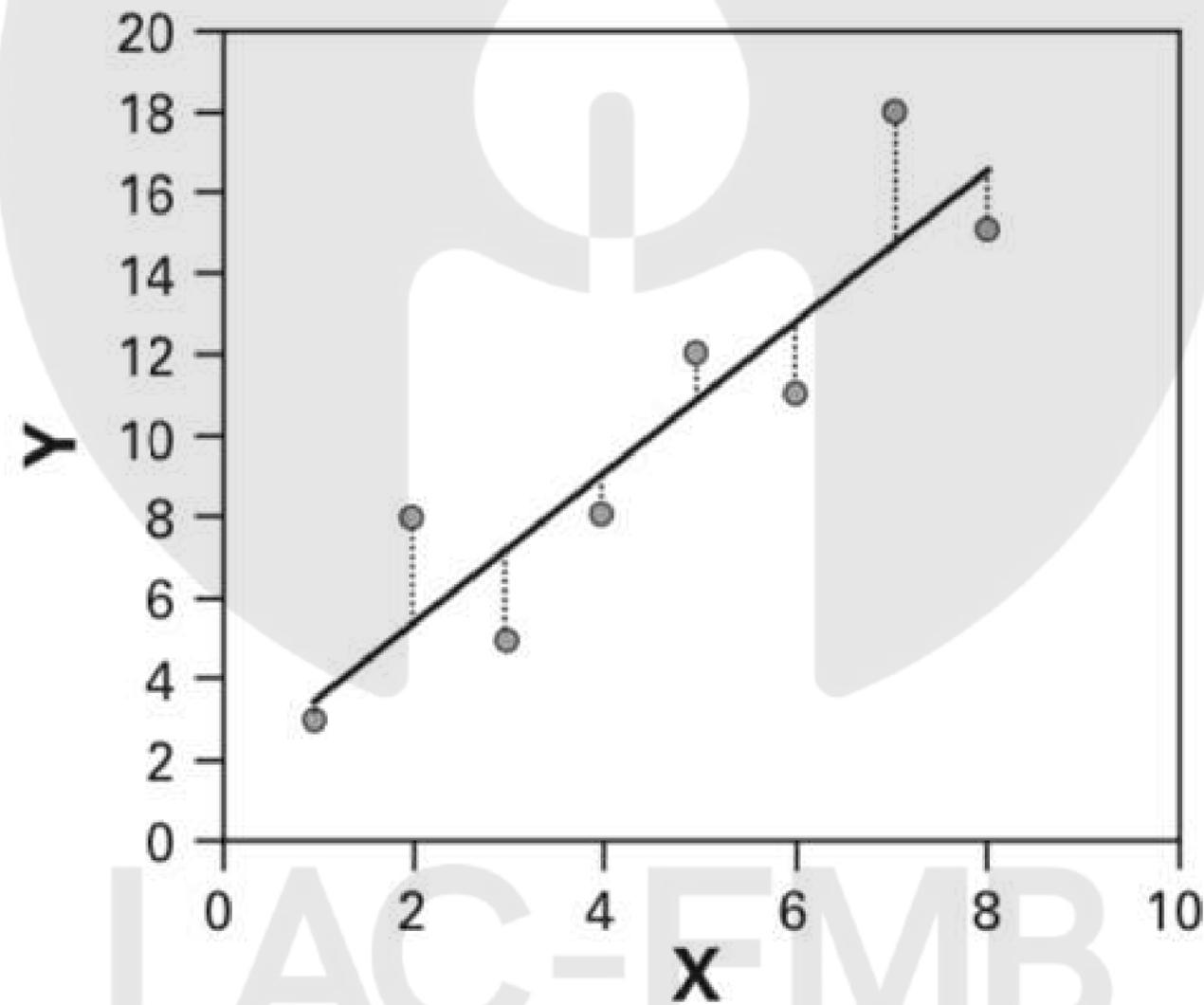
Todos os valores da variável de saída devem ser **independentes**

LAC-FMB



# Pressupostos para generalização

## 9. Linearidade

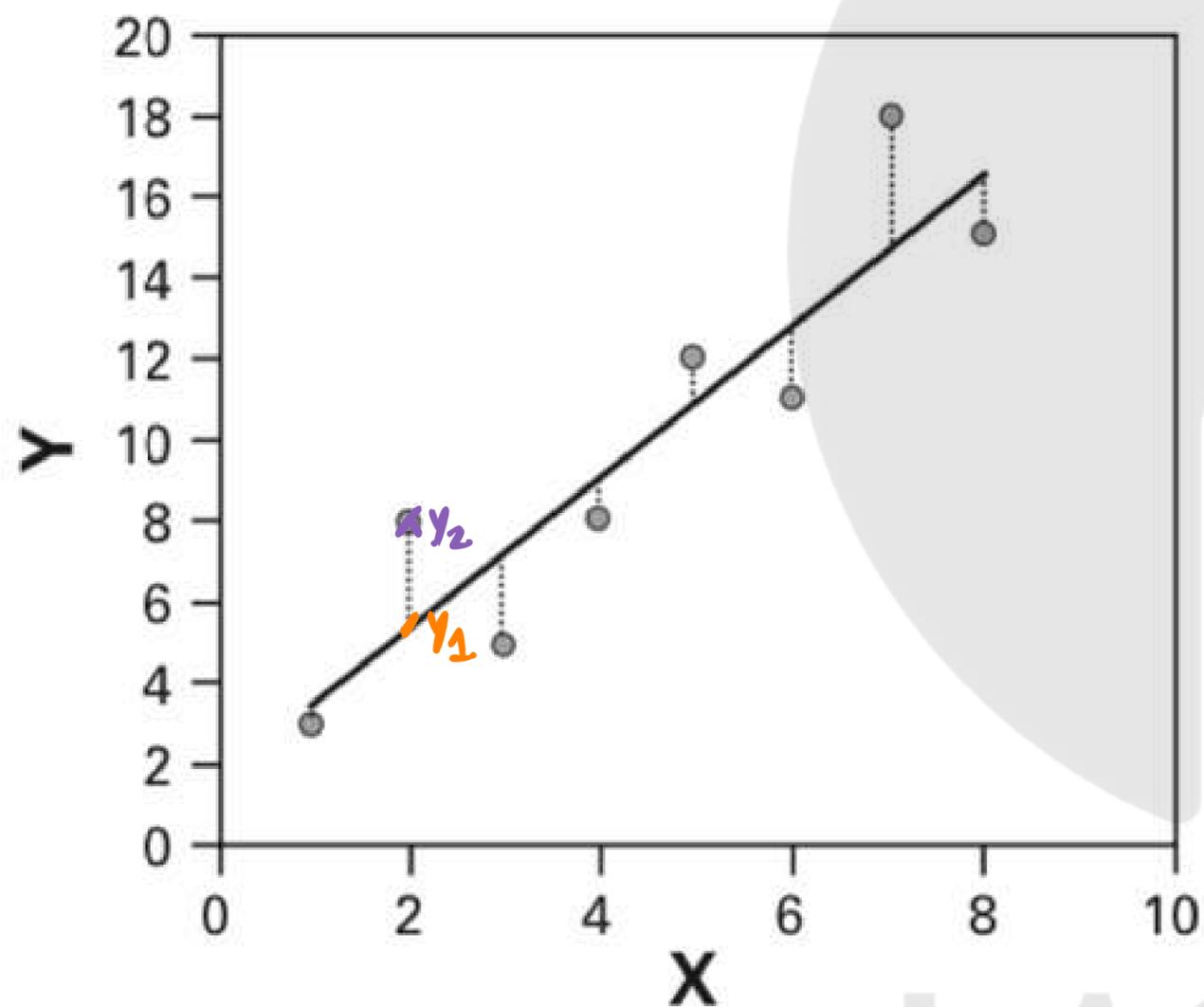


# Como avaliar meu modelo?

**#03** Módulo de Bioestatística



# Resíduos



Sendo:

- $Y_1 = (b_0 + b_1X_1)$

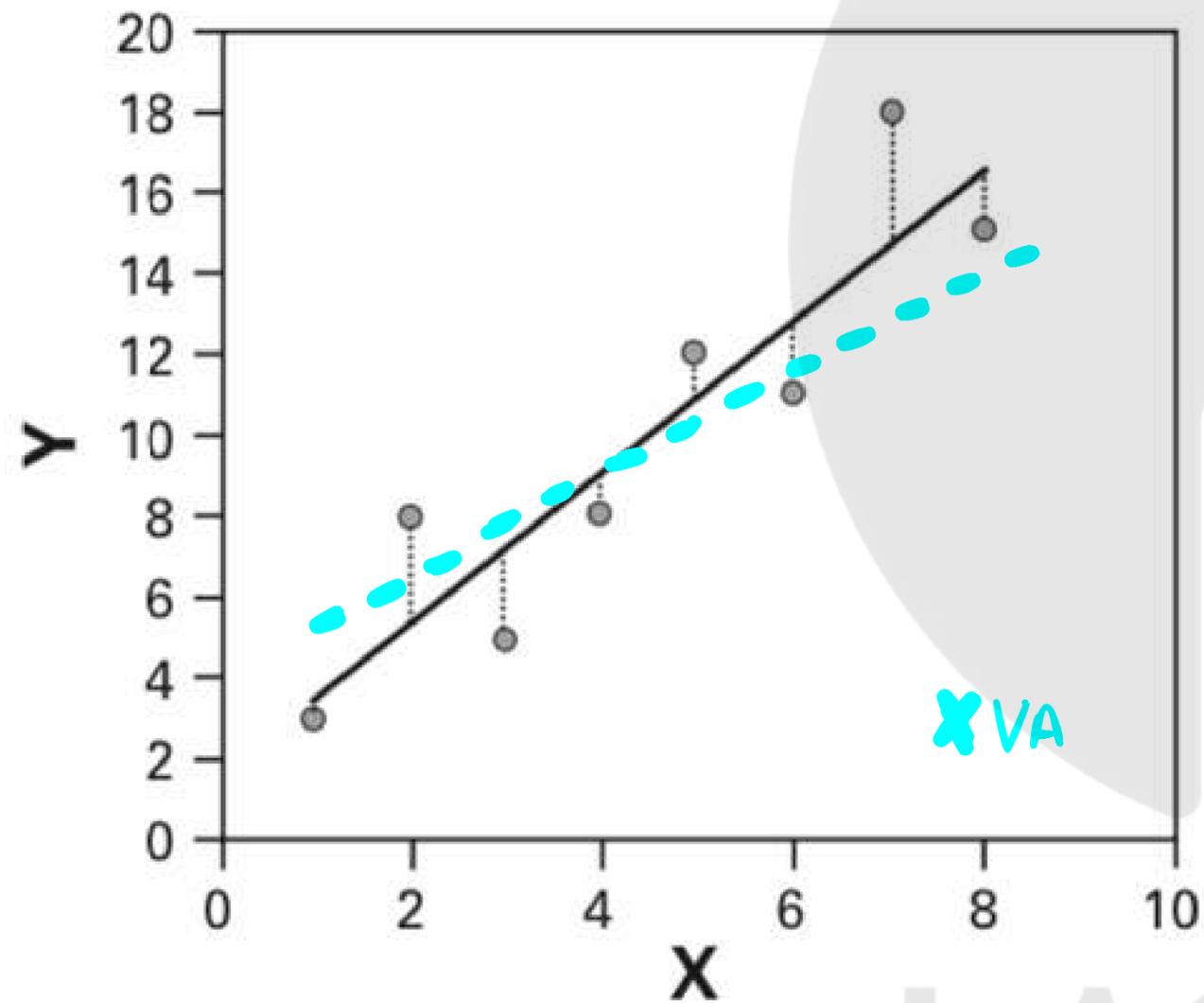
- $Y_2 \neq (b_0 + b_1X_1)$

- $|Y_2 - Y_1| = \text{Resíduo}$

- Resíduo é a diferença entre o Y previsto pelo modelo e o Y encontrado na amostra



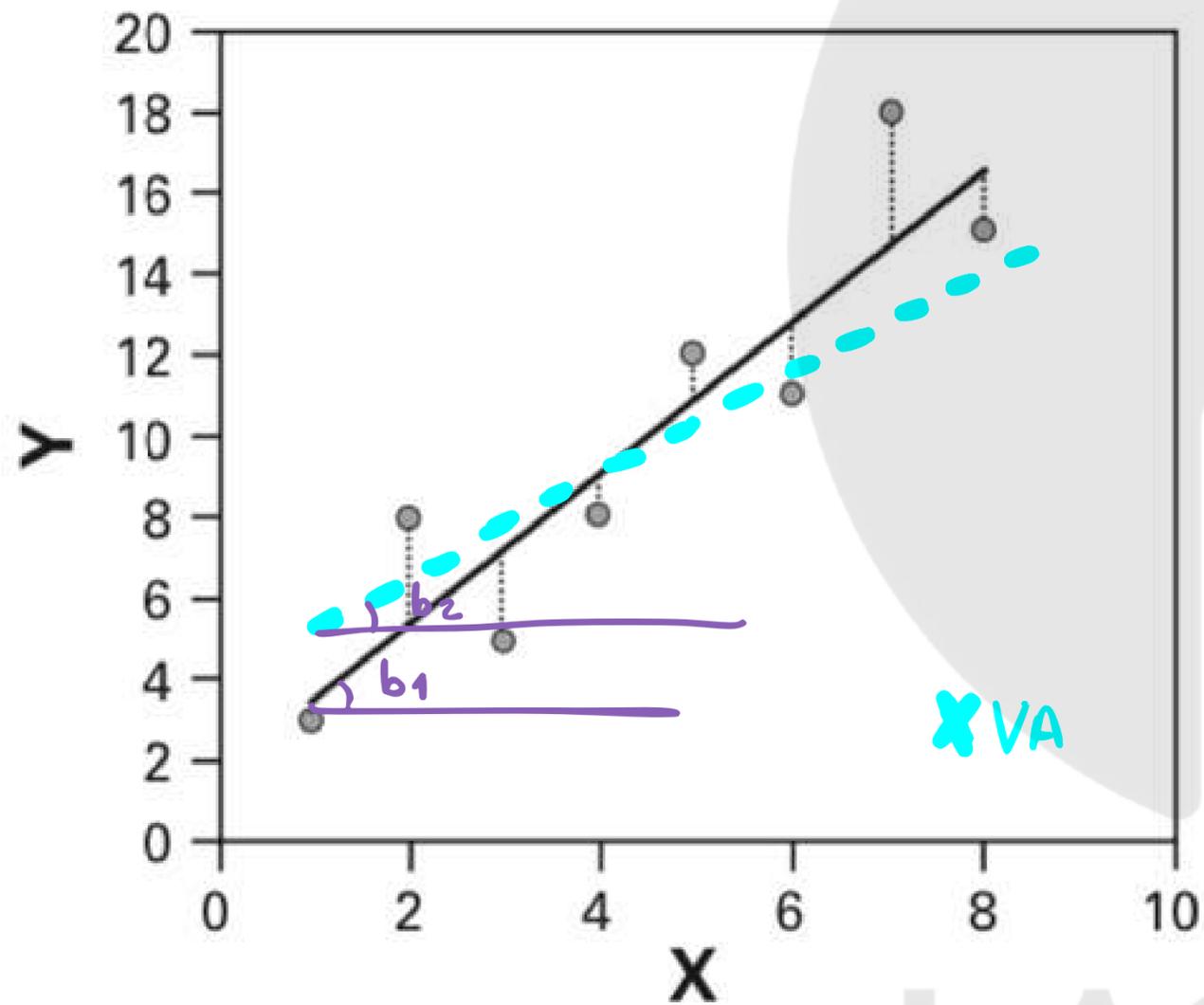
# Valores atípicos (Outliers)



Casos que diferem em grande quantidade dos demais



# Valores atípicos (Outliers)

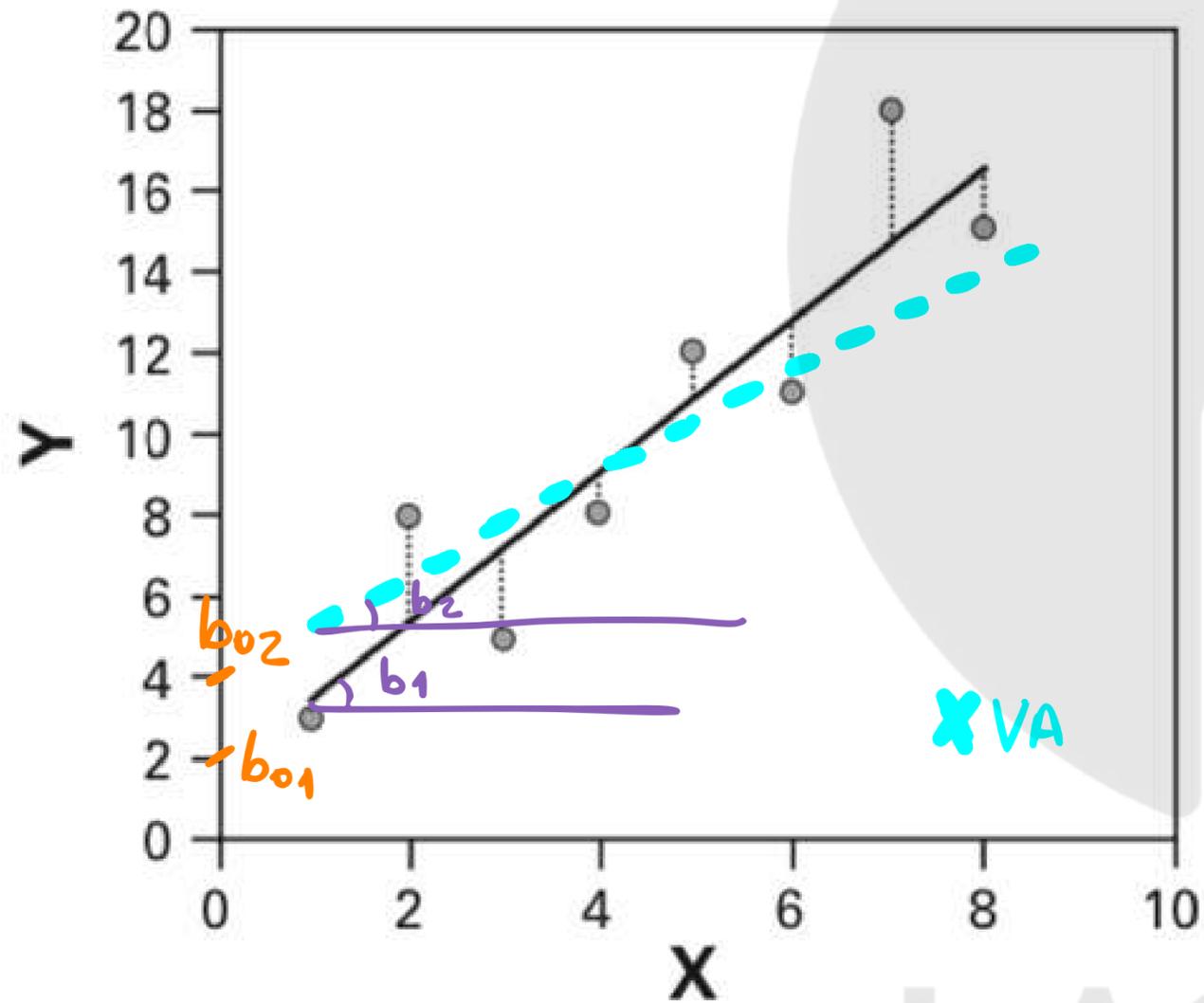


Casos que diferem em grande quantidade dos demais

A inclinação é reduzida



# Valores atípicos (Outliers)



Casos que diferem em grande quantidade dos demais

A inclinação é reduzida

O intercepto aumenta



# Casos influentes

↳ Se retirássemos determinados casos, teríamos coeficientes de regressão muito diferentes?

LAC-FMB



# Casos influentes

↳ Se retirássemos determinados casos, teríamos coeficientes de regressão muito diferentes?

**DFFit:** Diferença entre o **valor previsto** ajustado e o valor previsto original

LAC-FMB



# Casos influentes

↳ Se retirássemos determinados casos, teríamos coeficientes de regressão muito diferentes?

**DFFit:** Diferença entre o **valor previsto** ajustado e o valor previsto original

**DFBeta:** Diferença entre um **parâmetro** estimado utilizando todos os casos e estimado quando um caso é excluído

LAC-FMB



# Casos influentes

↳ Se retirássemos determinados casos, teríamos coeficientes de regressão muito diferentes?

**DFFit:** Diferença entre o **valor previsto** ajustado e o valor previsto original

**DFBeta:** Diferença entre um **parâmetro** estimado utilizando todos os casos e estimado quando um caso é excluído

Não são uma forma de justificar a remoção de dados!



# Generalização de resultados

- Os diagnósticos são recursos que permitem ver o quão bom ou ruim é o seu modelo em termos da **amostra**
- Não garantem uma boa generalização quando extrapolamos o grupo amostral
- Uma forma de ter um pouco mais de segurança é através do:

LAC-FMB



# Generalização de resultados

- Os diagnósticos são recursos que permitem ver o quão bom ou ruim é o seu modelo em termos da **amostra**
- Não garantem uma boa generalização quando extrapolamos o grupo amostral
- Uma forma de ter um pouco mais de segurança é através do:

**$R^2$  ajustado**

LAC-FMB



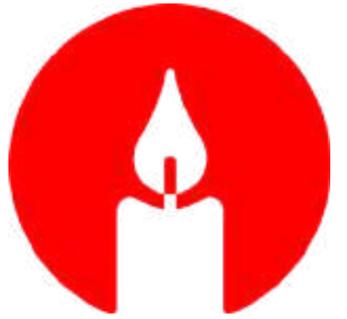
# Prática



**#04** Módulo de Bioestatística



# Referências



1. FIELD, Andy ; VIALI Lorí. Descubriendo a estatística usando o SPSS (2a. ed.).  
[s.l.]: Grupo A - Bookman, 2000.

# Obrigado!

Avalie a sessão Encontre-nos aqui



@lacfmb



Academicast



academiafmb.com



linkme.bio/lacfmb



**LAC-FMB**

Liga Acadêmica de Ciências

LAC-F

Liga Acadêmica de