

# Aula 06

## Regressão Linear Simples

- A análise de regressão é geralmente feita sob um referencial teórico que justifique a adoção de alguma relação matemática de causalidade.



Variável independente  
ou  
Variável explicativa

Variável dependente  
ou  
Variável resposta

## Objetivos:

- Predizer valores de uma variável **dependente (Y)** em função de uma variável **independente (X)**.
- Conhecer o quanto variações de X podem **afetar Y**.

# Exemplos:

Variável Independente $X$	⇒	Variável dependente $Y$
Peso corporal	⇒	IMS
Peso ao nascer	⇒	Peso a desmama
Peso corporal	⇒	Rendimento de Carcaça
Quant// fertilizante	⇒	Produção de milho

# O Modelo matemático

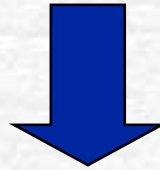
$$Y = \left[ \text{Predito por } X, \text{ segundo uma função} \right] + \left[ \text{Efeito aleatório} \right]$$

$$y = \alpha + \beta \cdot x + e$$

Parâmetros

Regressão  
Linear  
Simple

Amostra de observações de  $(X, Y)$



Conhecer o relacionamento entre  $X$  e  $Y$

# Construção da equação de regressão com base nos dados da amostra:

$$\hat{y} = a + b.X$$

$$\left( b = \frac{S_{XY}}{S_X^2} \right) \quad \text{(estimativa do beta)}$$

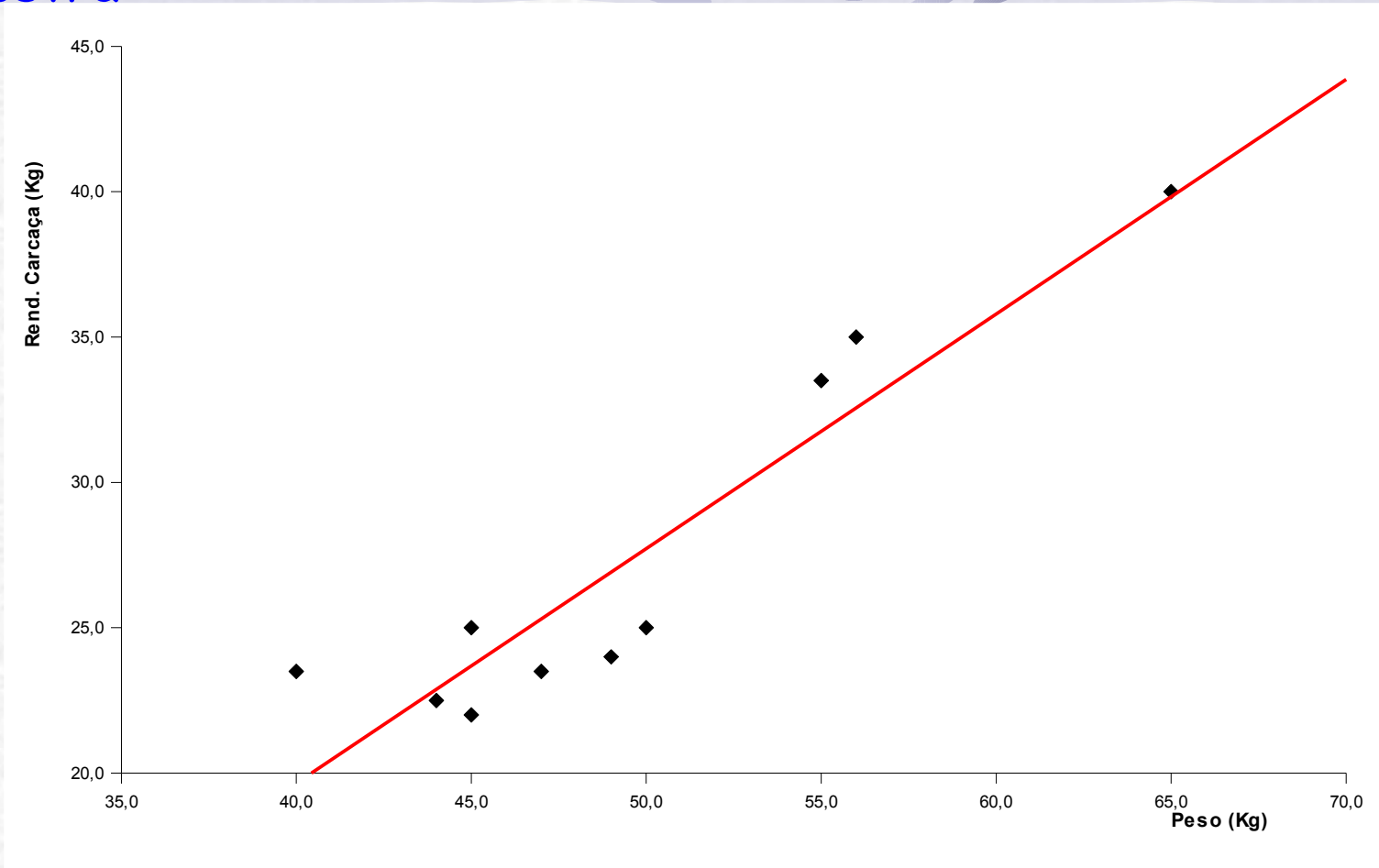
$$a = \bar{y} - b\bar{x} \quad \text{(estimativa do alfa)}$$

**Exemplo:** Peso corporal e Rendimento de carcaça, aos 4 meses de idade, de 10 cordeiros da raça Hampshire Down.

Peso (Kg)	Rend. Carcaça (Kg)
49,0	24,0
65,0	40,0
45,0	25,0
40,0	23,5
55,0	33,5
45,0	22,0
44,0	22,5
47,0	23,5
50,0	25,0
56,0	35,0



# Construindo a equação de regressão com base nos dados da amostra:



$$\hat{y} = a + b.X \quad b = \frac{S_{XY}}{S_X^2} \quad a = \bar{y} - b\bar{x}$$

# Construindo a reta de regressão para os dados do exemplo

Peso (X)	Carcaça (Y)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x}) \times (y_i - \bar{y})$	
49,00	24,00	-0,60	-3,40	0,36	2,04	
65,00	40,00	15,40	12,60	237,16	194,04	
45,00	25,00	-4,60	-2,40	21,16	11,04	
40,00	23,50	-9,60	-3,90	92,16	37,44	
55,00	33,50	5,40	6,10	29,16	32,94	
45,00	22,00	-4,60	-5,40	21,16	24,84	
44,00	22,50	-5,60	-4,90	31,36	27,44	
47,00	23,50	-2,60	-3,90	6,76	10,14	
50,00	25,00	0,40	-2,40	0,16	-0,96	
56,00	35,00	6,40	7,60	40,96	48,64	
<b>Soma =</b>	<b>496,00</b>	<b>274,00</b>	<b>0,0</b>	<b>0,0</b>	<b>480,40</b>	<b>387,60</b>

$$S_{XY} = \frac{387,6}{9} = 43,07$$

$$S_x^2 = \frac{480,4}{9} = 53,38$$

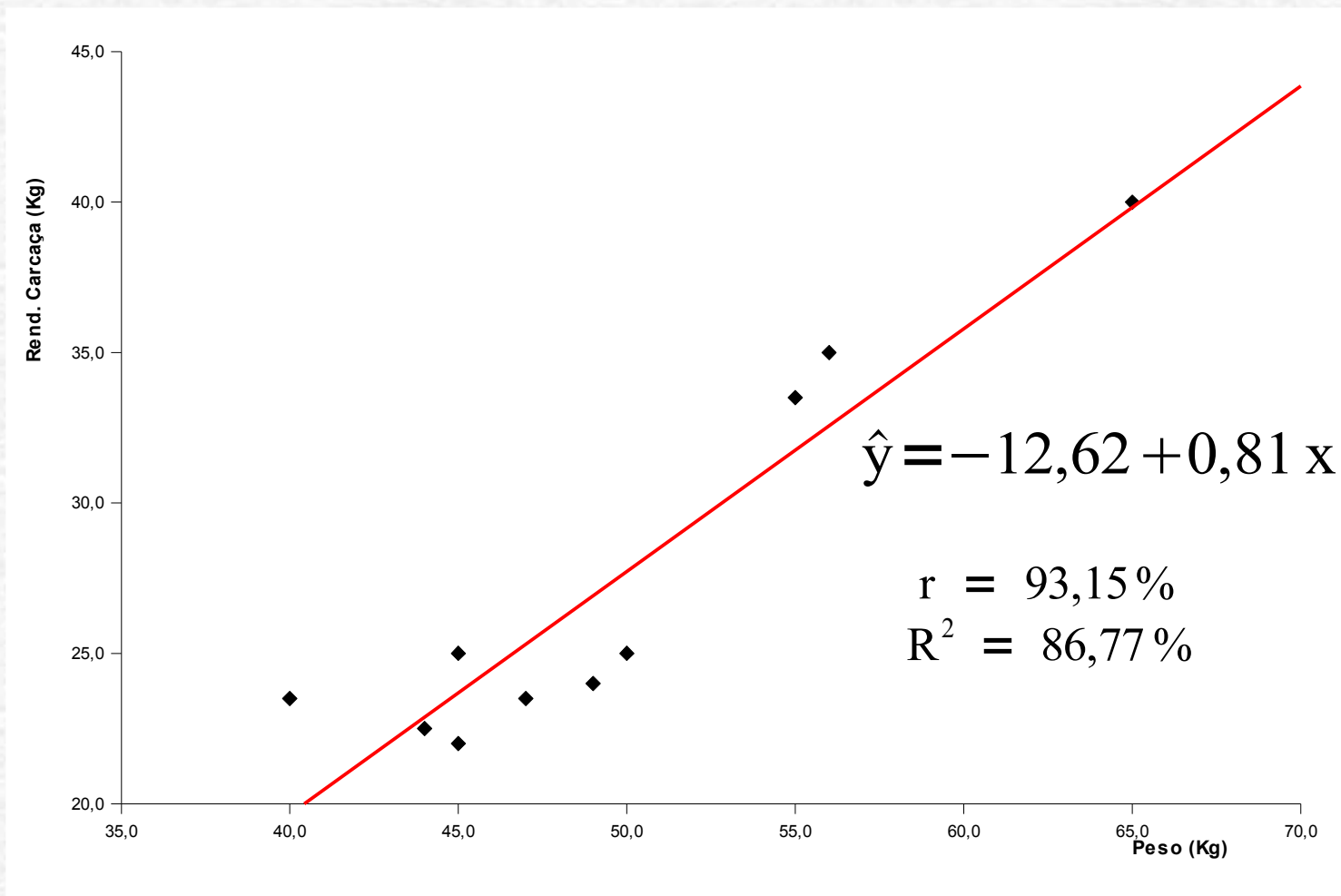
$$b = \frac{S_{XY}}{S_x^2} = \frac{43,07}{53,38} = 0,807$$

$$\bar{x} = \frac{496,0}{10} = 49,6$$

$$\bar{y} = \frac{274,0}{10} = 27,4$$

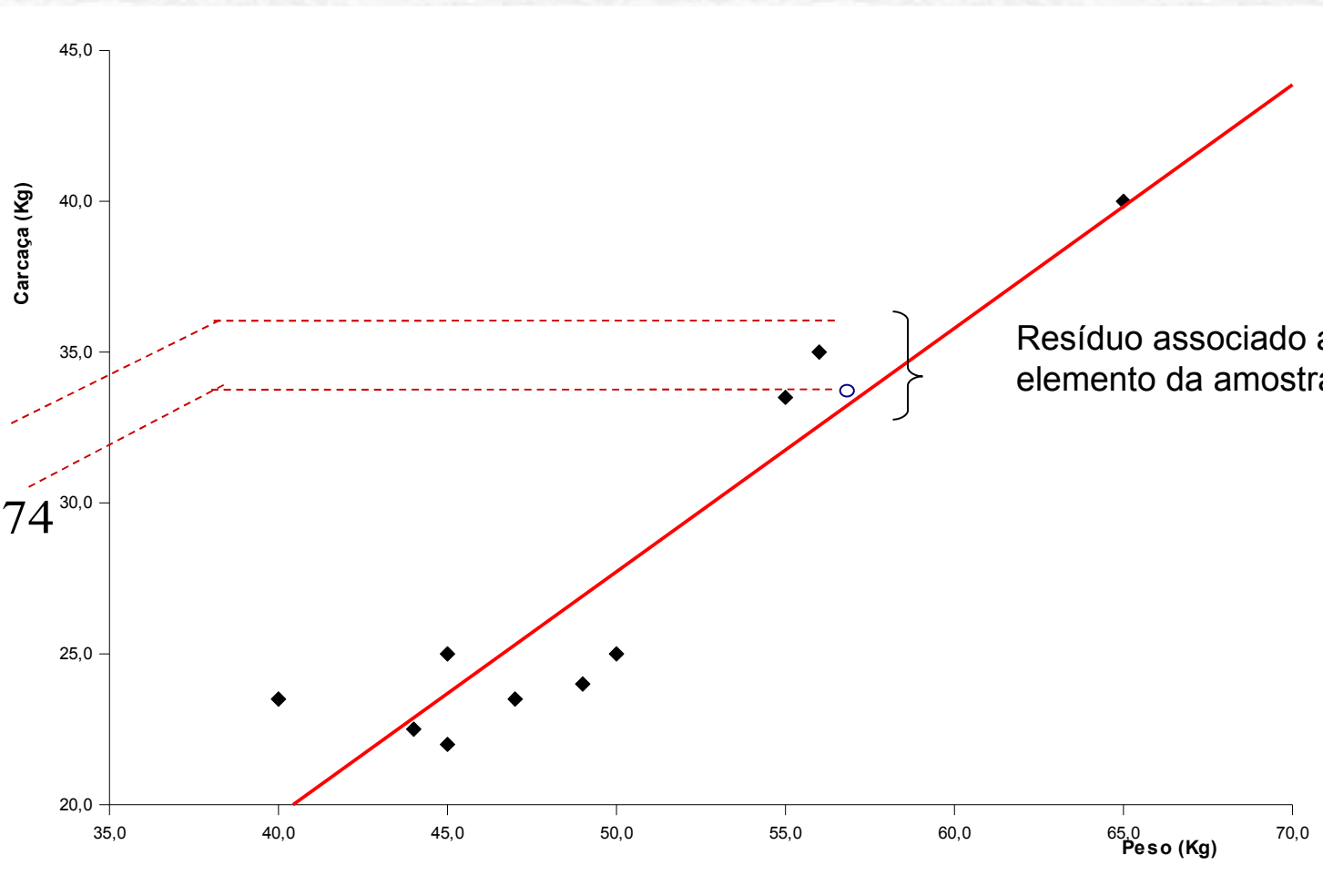
$$a = \bar{y} - b\bar{x} = 27,4 - (0,807 \cdot 49,6) = -12,62$$

**Exemplo:** Peso corporal e Rendimento de carcaça, aos 4 meses de idade, de 10 cordeiros da raça Hampshire Down.



# Valores preditos e resíduos

$$y_{10} = 35$$
$$\hat{y}_{10} = 32,74$$



Resíduo associado ao décimo elemento da amostra.

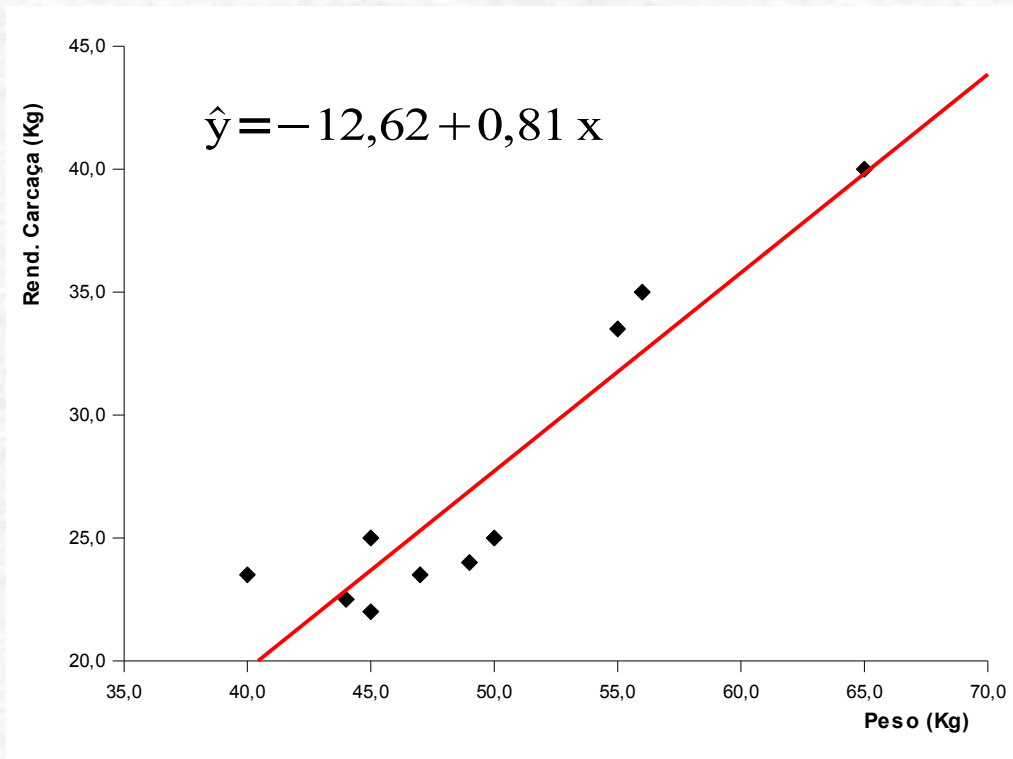
# Valores preditos e resíduos

$$\hat{y} = -12,62 + 0,81 x$$

$$\hat{e} = y - \hat{y}$$

Peso (X)	Rend. Carcaça (Y)	Previsto	Resíduo
49,0	24,0	27,07	-3,07
65,0	40,0	40,03	-0,03
45,0	25,0	23,83	1,17
40,0	23,5	19,78	3,72
55,0	33,5	31,93	1,57
45,0	22,0	23,83	-1,83
44,0	22,5	23,02	-0,52
47,0	23,5	25,45	-1,95
50,0	25,0	27,88	-2,88
56,0	35,0	32,74	2,26

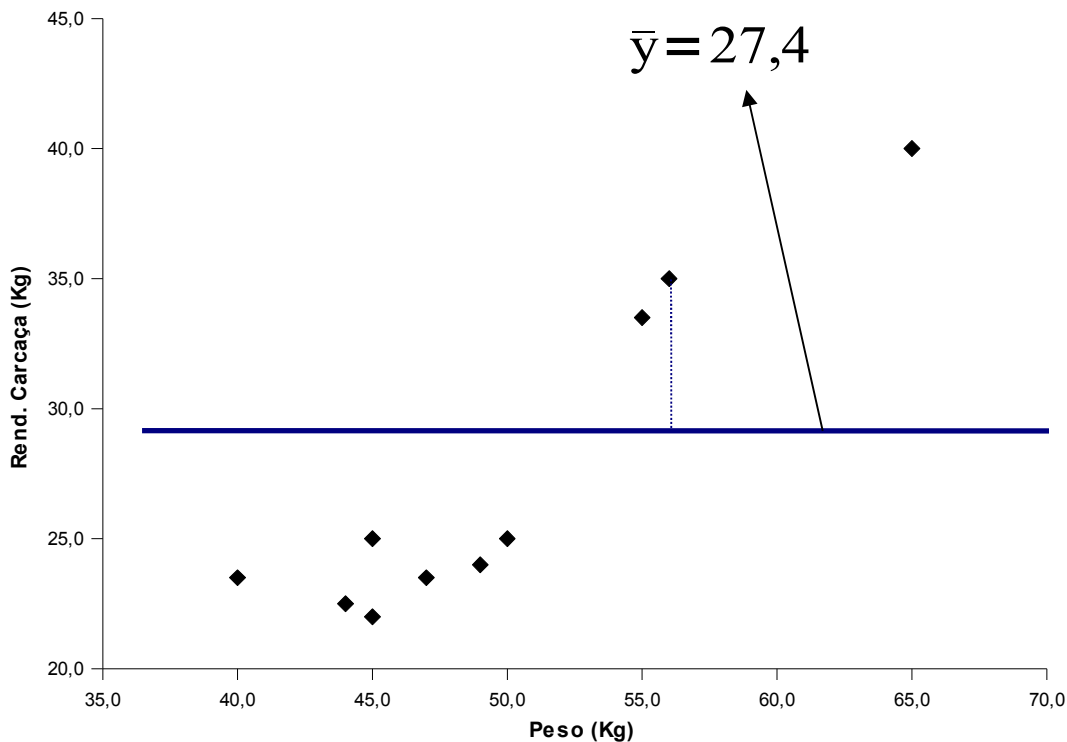
# Varição explicada e não explicada pelo modelo de regressão



Varição não explicada pelo modelo de regressão

Soma de quadrados devida ao erro aleatório:

$$SQE = \sum (y - \hat{y})^2$$



Varição em relação à média aritmética (variação total)

Soma de quadrado total:

$$SQT = \sum (y - \bar{y})^2$$

Varição explicada pelo modelo de regressão



$$SQR = SQT - SQE$$

# Varição explicada e não explicada

Soma de quadrado total:

$$SQT = \sum (y - \bar{y})^2$$

Soma de quadrados do erro:

$$SQE = \sum (y - \hat{y})^2$$

Soma de quadrados da regressão:

$$SQR = SQT - SQE$$

Coefficiente de determinação:

$$R^2 = \frac{SQR}{SQT} = \frac{\text{variação explicada}}{\text{variação total}}$$

$$0 \leq R^2 \leq 1$$



## No Exemplo:

Peso (X)	Carcaça (Y)	Média de Y	$y - \bar{y}$	$(y - \bar{y})^2$
49,0	24,0	27,4	-3,4	11,56
65,0	40,0		12,6	158,76
45,0	25,0		-2,4	5,76
40,0	23,5		-3,9	15,21
55,0	33,5		6,1	37,21
45,0	22,0		-5,4	29,16
44,0	22,5		-4,9	24,01
47,0	23,5		-3,9	15,21
50,0	25,0		-2,4	5,76
56,0	35,0		7,6	57,76
Soma =			0	360,4


$$SQT = \sum (y - \bar{y})^2$$

## No Exemplo:

Peso (X)	Carcaça (Y)	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
49,0	24,0	27,07	-3,07	9,425
65,0	40,0	40,03	-0,03	0,001
45,0	25,0	23,83	1,17	1,369
40,0	23,5	19,78	3,72	13,838
55,0	33,5	31,93	1,57	2,465
45,0	22,0	23,83	-1,83	3,349
44,0	22,5	23,02	-0,52	0,270
47,0	23,5	25,45	-1,95	3,803
50,0	25,0	27,88	-2,88	8,294
56,0	35,0	32,74	2,26	5,108
Soma =				47,922


$$SQE = \sum (y - \hat{y})^2$$

## No exemplo:

Fonte de variação	Soma dos quadrados
Explicada por X pelo modelo de regressão (variação explicada)	SQR = 312,48
Devida ao erro aleatório (variação não explicada)	SQE = 47,92
<b>Variação total</b>	<b>SQT = 360,40</b>

$$R^2 = \frac{SQR}{SQT} = \frac{312,48}{360,40} = 0,867 \text{ ou } 86,7\%$$

## Interpretar

# Pressupostos do modelo de regressão

$$y = \alpha + \beta \cdot x + e$$

- Os erros ( $e$ 's) são independentes e variam aleatoriamente segundo uma distribuição (normal) com média zero e variância constante.