

Regressão Linear

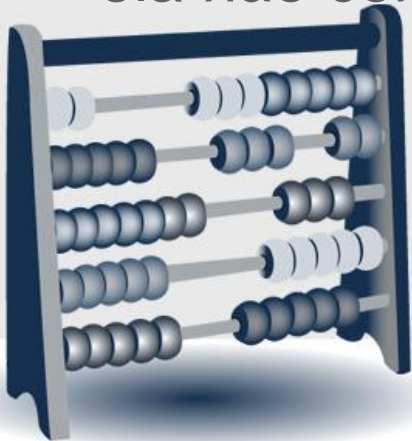
Reinaldo Madarazo - 2014

Covariância

- A covariância mede a força de uma relação linear entre duas variáveis numéricas (x e y).

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

- Uma vez que a covariância pode assumir qualquer valor, ela não consegue dizer a força relativa da relação. Por exemplo, um valor de 6,9377777 não diz se a relação entre x e y é forte ou fraca.

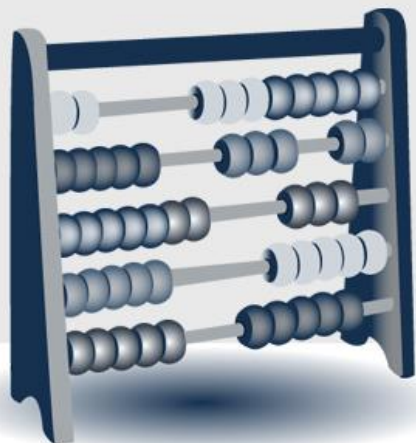


Exemplo

- A tabela fornece o custo de um hambúrguer em lanchonetes e o custo de dois ingressos da cinema em 10 cidades do mundo (K. Spors. “Keeping Up with... Yourself”, The Wall Street Journal, 11 de abril de 2005, p. R4).

Resp:

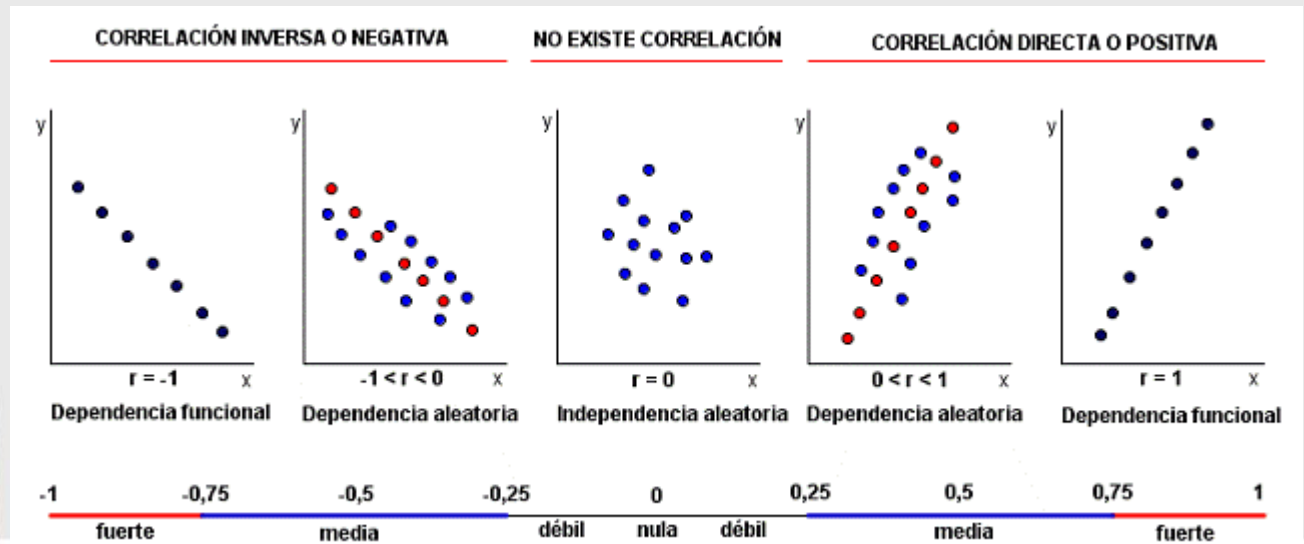
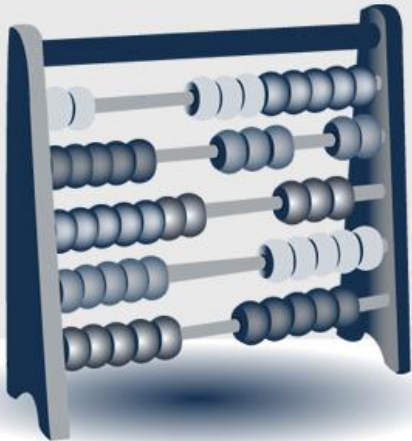
$$\text{Cov}(x,y)=61,53993/ \\ (10-1)=6,83777$$



Cidade	Hambúrguer	Ingressos
Tóquio	5,99	32,66
Londres	7,62	28,41
Nova York	5,75	20,00
Sydney	4,45	20,71
Chicago	4,99	18,00
San Francisco	5,29	19,50
Boston	4,39	18,00
Atlanta	3,70	16,00
Toronto	4,62	18,05
Rio de Janeiro	2,99	9,90

Coeficiente de Correlação

- O Coeficiente de Correlação mede a força relativa de uma relação linear entre duas variáveis numéricas. Os valores do coeficiente de correlação (r) vão de -1 , para uma correlação negativa (decrescente) perfeita, até $+1$, para uma correlação positiva (crescente) perfeita. *Perfeita* significa que todos os pontos estão contidos em uma reta.



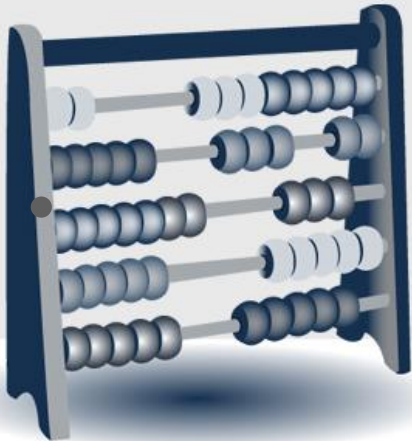
Coeficiente de Correlação da Amostra

$$r = \frac{\text{cov}(x, y)}{S_x S_y}$$

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$S_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}$$

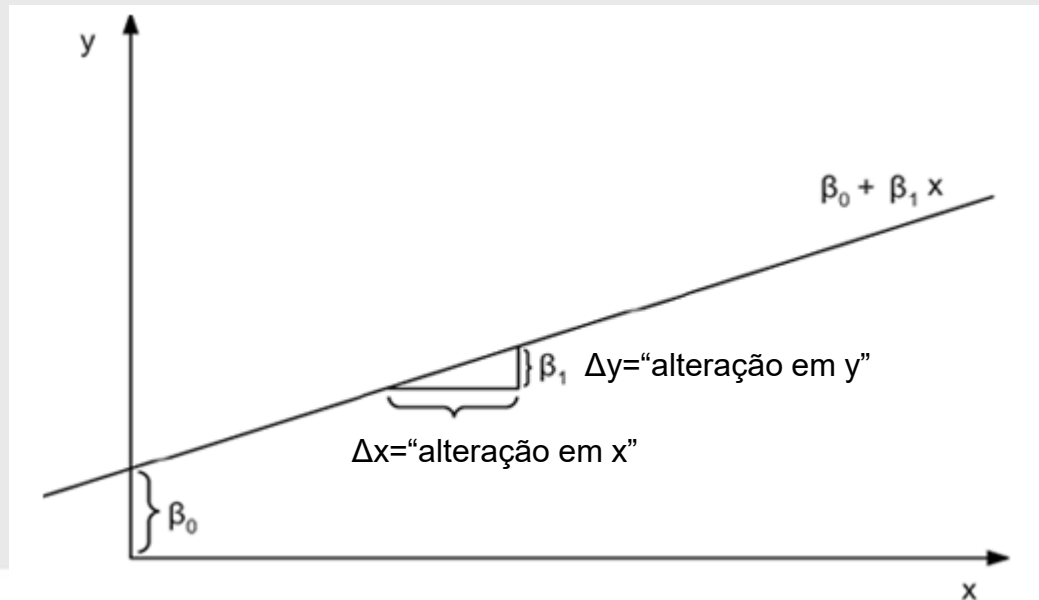
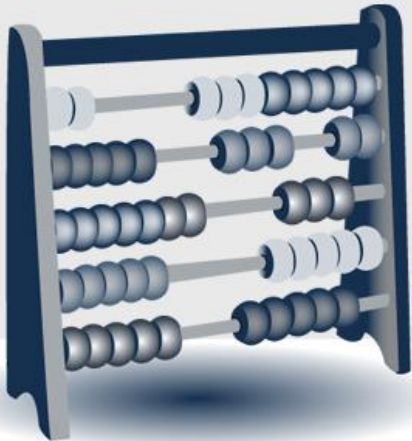


Exemplo: Calcule o coeficiente de correlação no problema do preço do hambúrguer e ingressos de cinema.

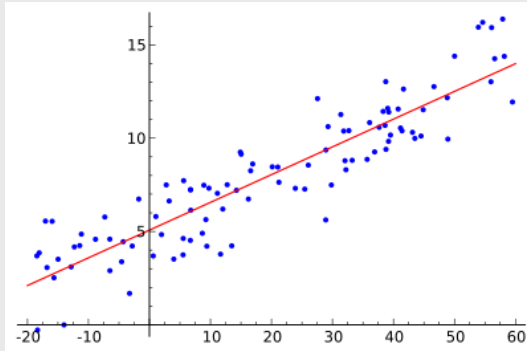
Resposta: $r = 6,83777 / [(1,2925) \cdot (6,337)] = 0,8348$

Regressão Linear Simples

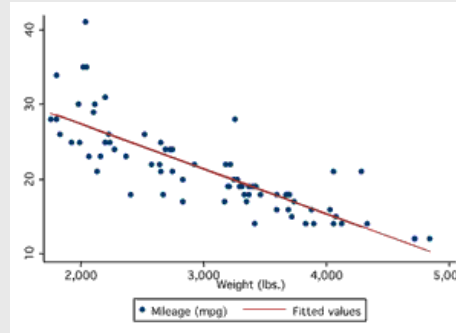
- Na Regressão Linear Simples, uma única variável independente numérica, x , é utilizada para prever a variável dependente numérica, y .
- *Gráfico ou Diagrama de Dispersão*: mostra a relação entre uma variável x no eixo horizontal e uma variável y no eixo vertical.



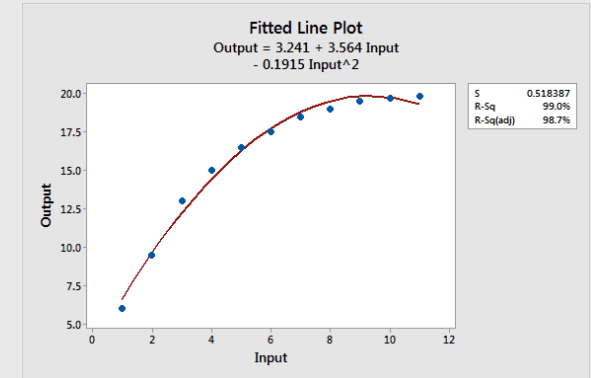
Tipos de Regressão



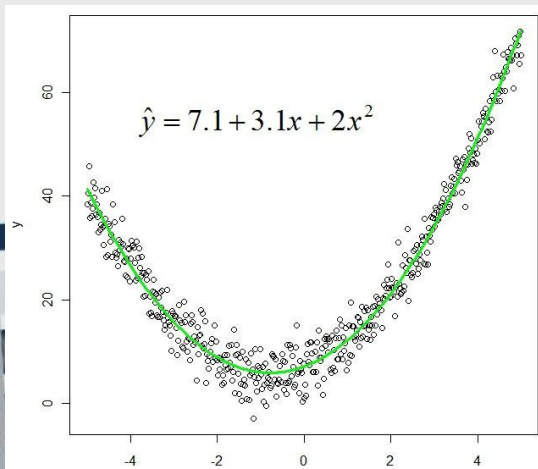
Relação Linear Positiva



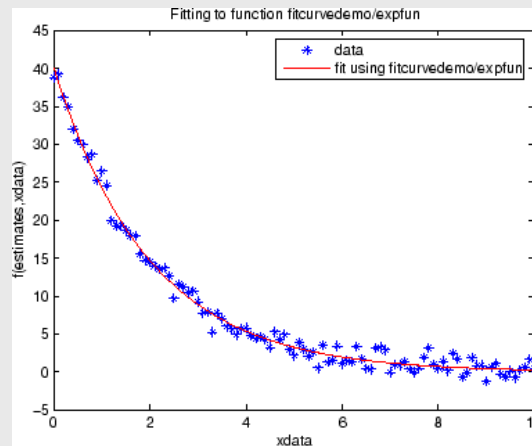
Relação Linear Negativa



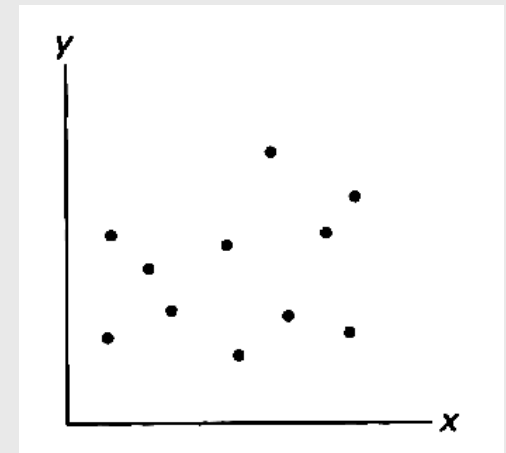
Relação Curvilínea Positiva



Relação Curvilínea em "U"

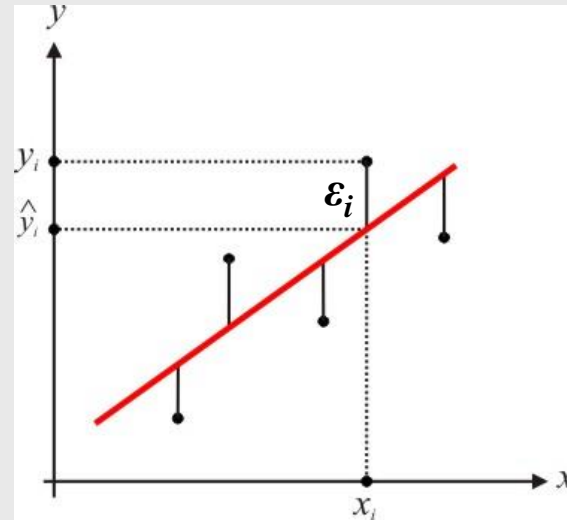
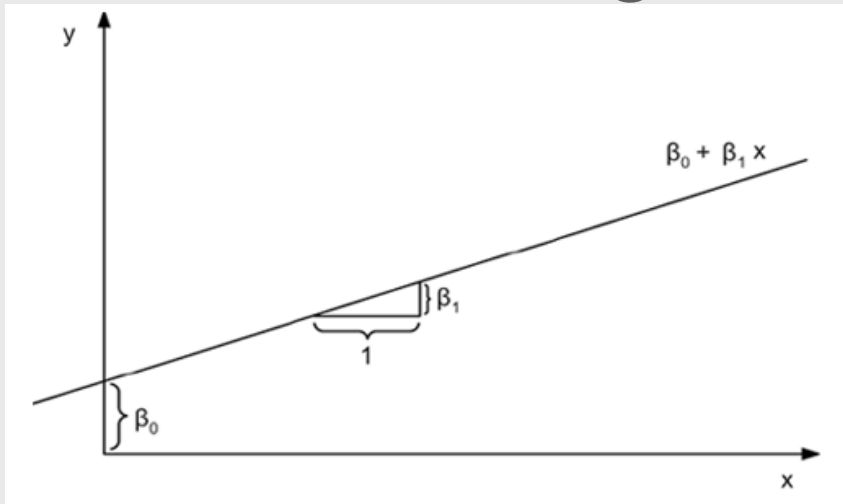


Relação Curvilínea Negativa



Nenhuma relação entre X e Y

Modelo de Regressão Linear Simples



$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

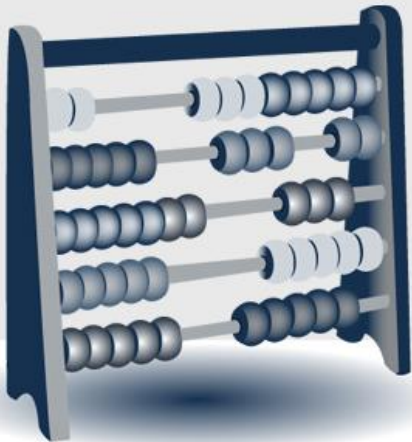
β_0 = intercepto de y (coef. linear)

β_1 = inclinação (coef. angular)

ε_i = erro aleatório em y para a observação i

y_i = variável dependente (ou variável de resposta)

x_i = variável independente (ou variável explanatória)



Método dos Mínimos Quadrados

- É uma forma de se determinar a *equação* da regressão linear simples, obtendo os valores do **intercepto**, **inclinação** e **correlação** entre os dados x e y .

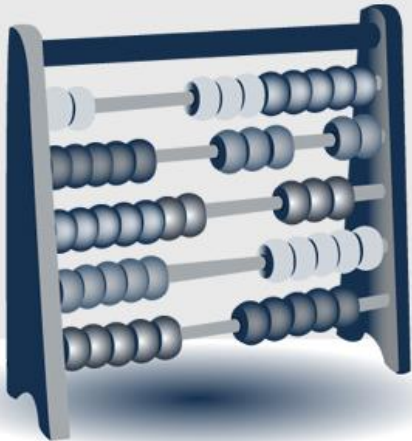
$$\hat{Y}_i = b_0 + b_1 X_i$$

\hat{Y}_i = *valor previsto de Y para observação i*

X_i = *valor de X para a observação i*

b_0 = *itercepto da amostra Y*

b_1 = *inclinação da amostra*



Calculo da Inclinação b_1

$$b_1 = \frac{SQXY}{SQX}$$

$$SQXY = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - \frac{\left(\sum_{i=1}^n X_i\right)\left(\sum_{i=1}^n Y_i\right)}{n}$$

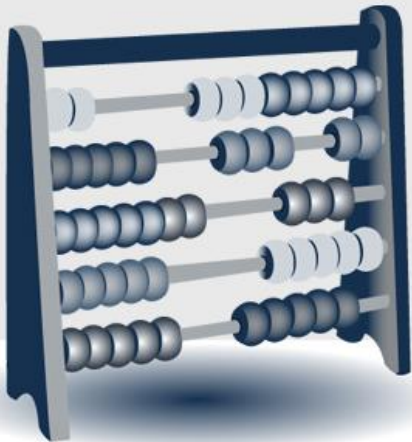
$$SQX = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}$$

Calculo do Intercepto b_0

$$b_0 = \bar{Y} - b_1 \bar{X}$$

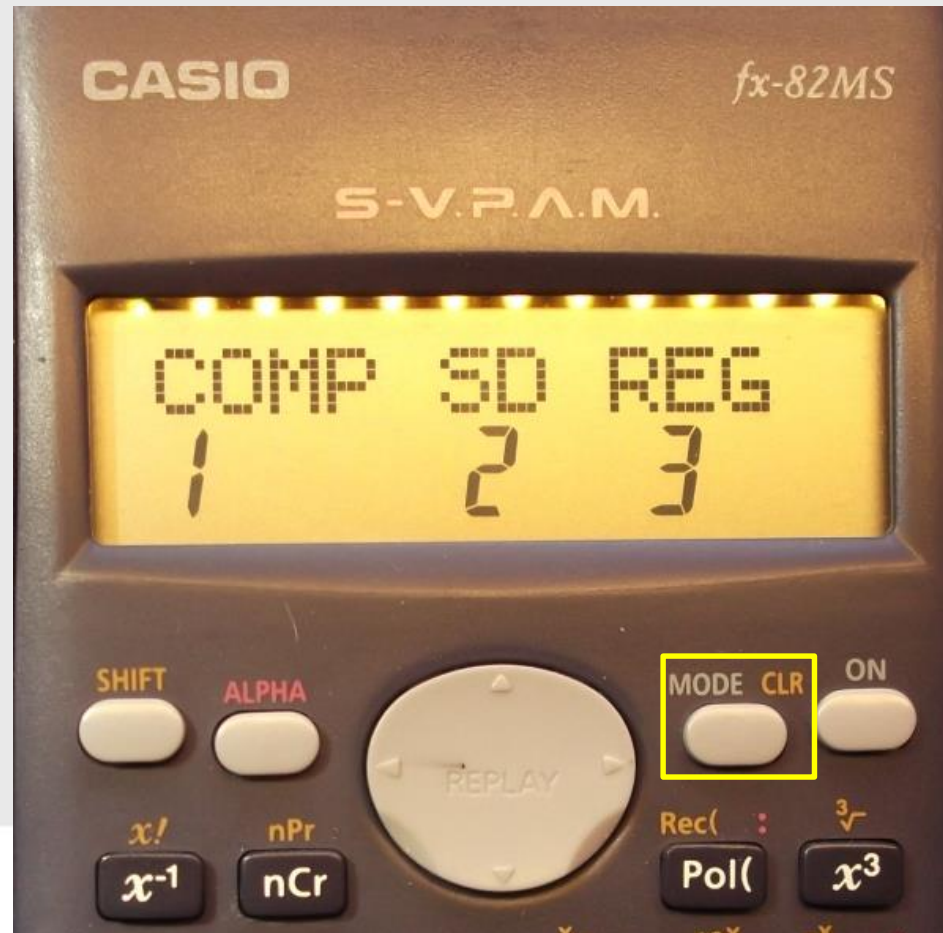
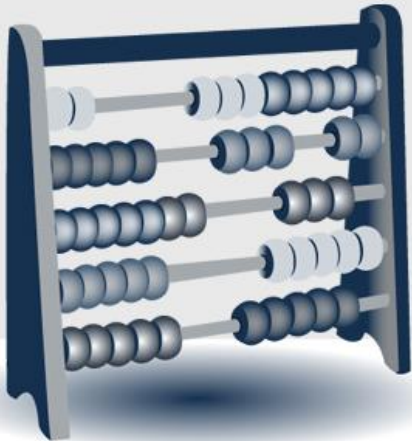
$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$



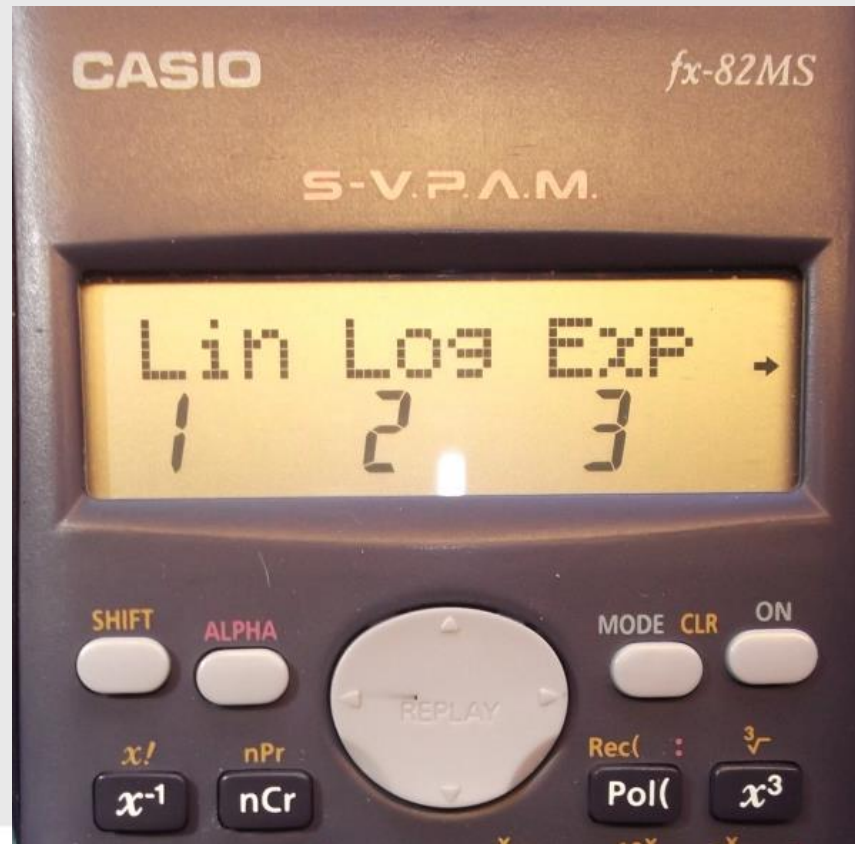
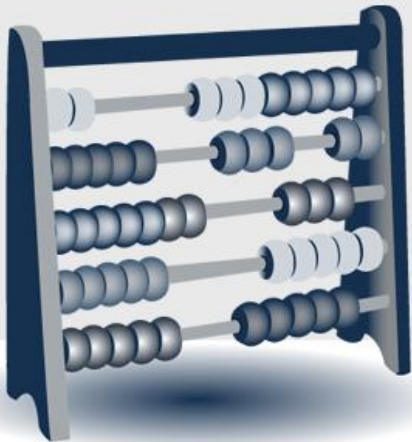
Usando a calculadora

- Selecionar o modo Regressão REG (menu “3”):



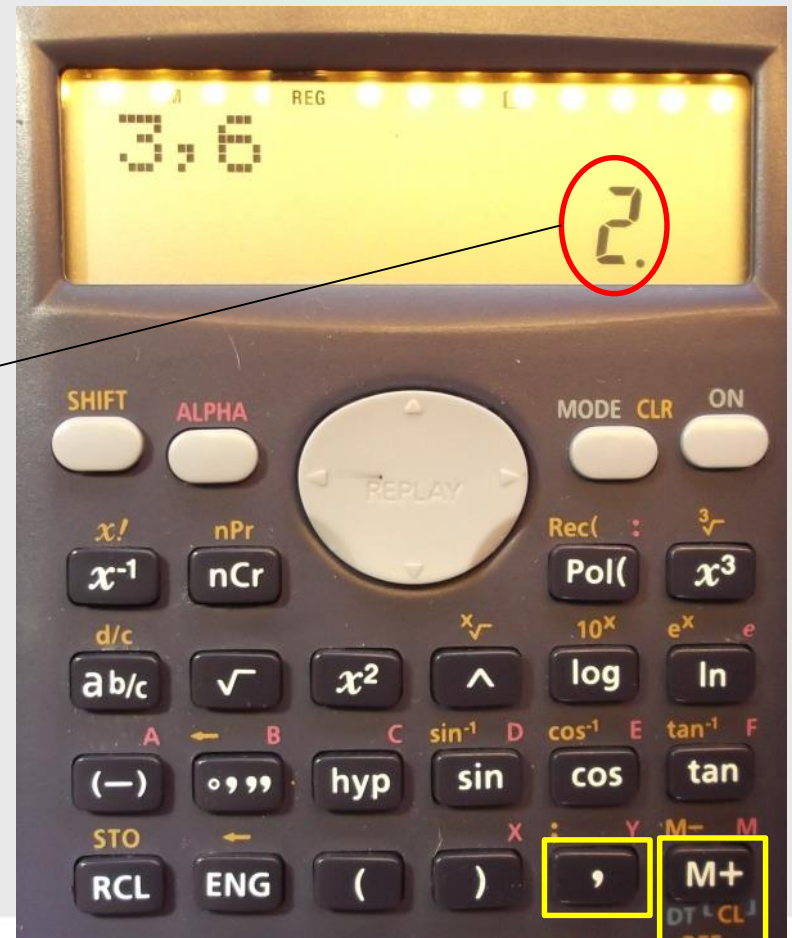
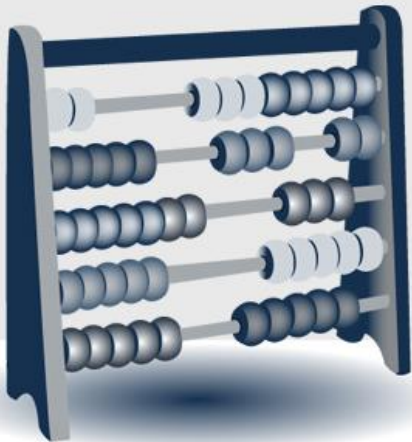
Usando a calculadora

- Selecionar a opção Linear LIN (menu “1”)



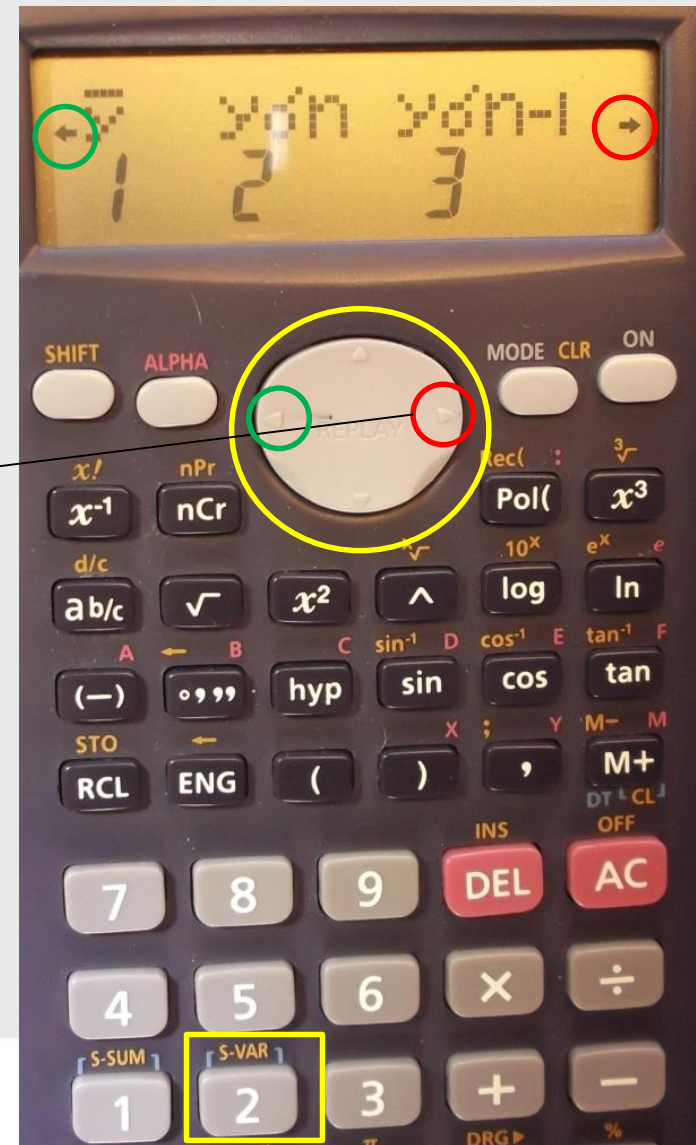
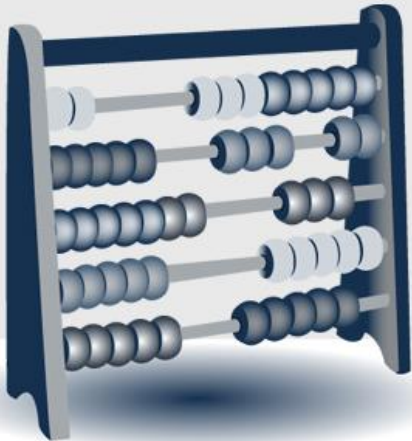
Usando a Calculadora

- Inserir os dados, colocando primeiro o valor de X, depois o valor de Y, separados por uma vírgula.
- Em seguida, pressionar DT.
- O display irá acusar quantos pares foram introduzidos.



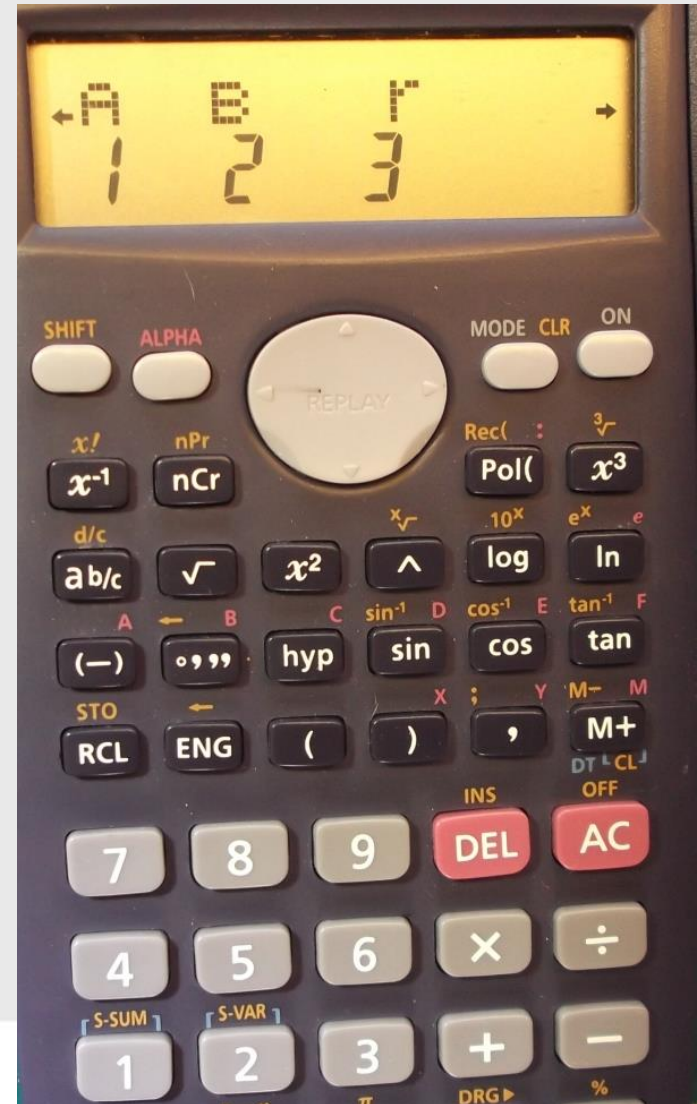
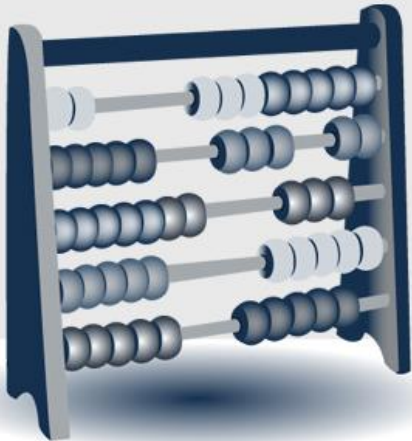
Usando a Calculadora

- Pressionar S-VAR (Shift 2).
- Essa opção possui vários menus, que são indicados pela seta do lado direito do display.
- Para navegar pelos menus, devemos usar o botão de navegação, pressionando >.

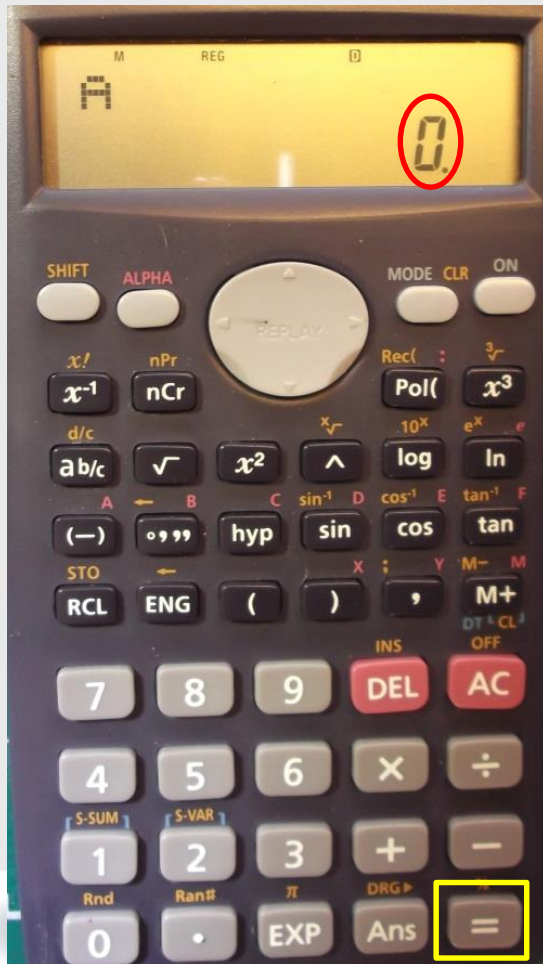


Usando a Calculadora

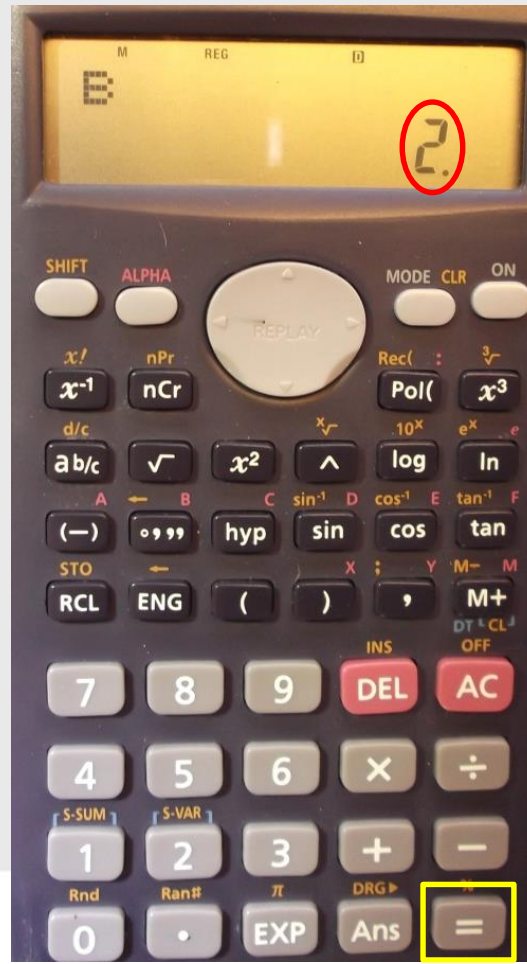
- Navegar até encontrar o menu exibindo A, B e r.
- Opção **3**: Obtém-se o **coeficiente de correlação** após pressionar =.
- Opção **1**: Obtém-se o coeficiente A, o **intercepto**, após pressionar =.
- Opção **2**: Obtém-se o coeficiente B, a **inclinação**, após pressionar =.



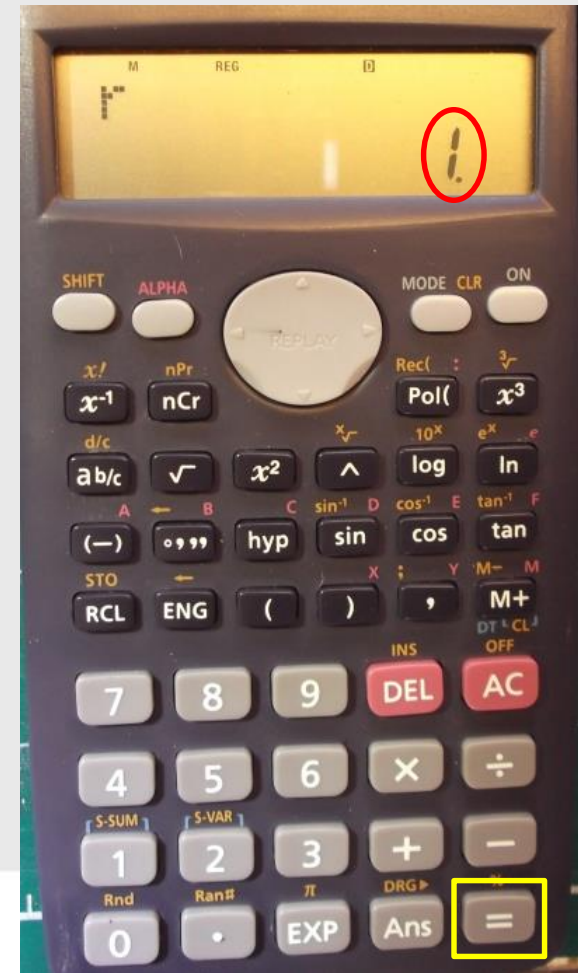
Obtendo o **Intercepto**:
Selecionar o menu "1"
e pressionar "="



Obtendo a **Inclinação:**
Selecionar o menu “2” e pressionar “=”



Obtendo a **Correlação**:
Selecionar o menu “3” e
pressionar “=”



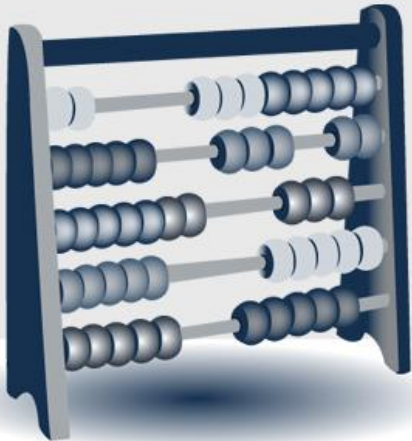
Linha de Previsão

- Conhecidos b_0 e b_1 , é possível prever o valor de Y , conhecendo X , usando a seguinte expressão:

$$\hat{Y} = b_0 + b_1 X$$

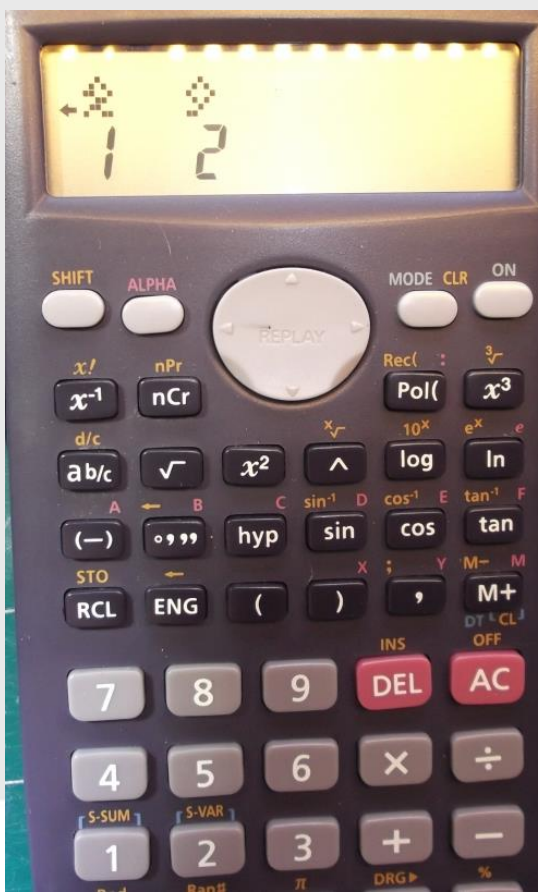
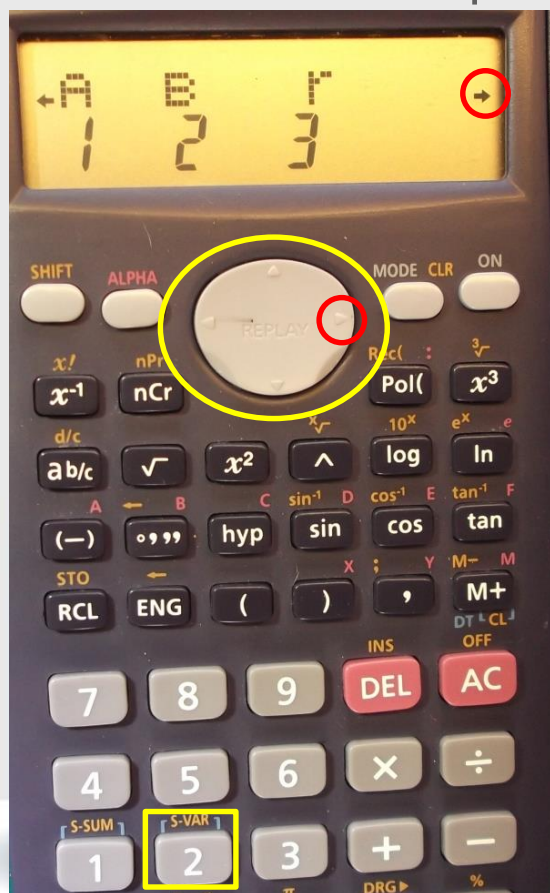
- Também é possível prever o valor de X , conhecendo Y , usando a seguinte expressão:

$$\hat{X} = \frac{Y - b_0}{b_1}$$



Usando a Calculadora

- Usando o menu obtido pressionando S-VAR, navegar até a opção de calculo das previsões:

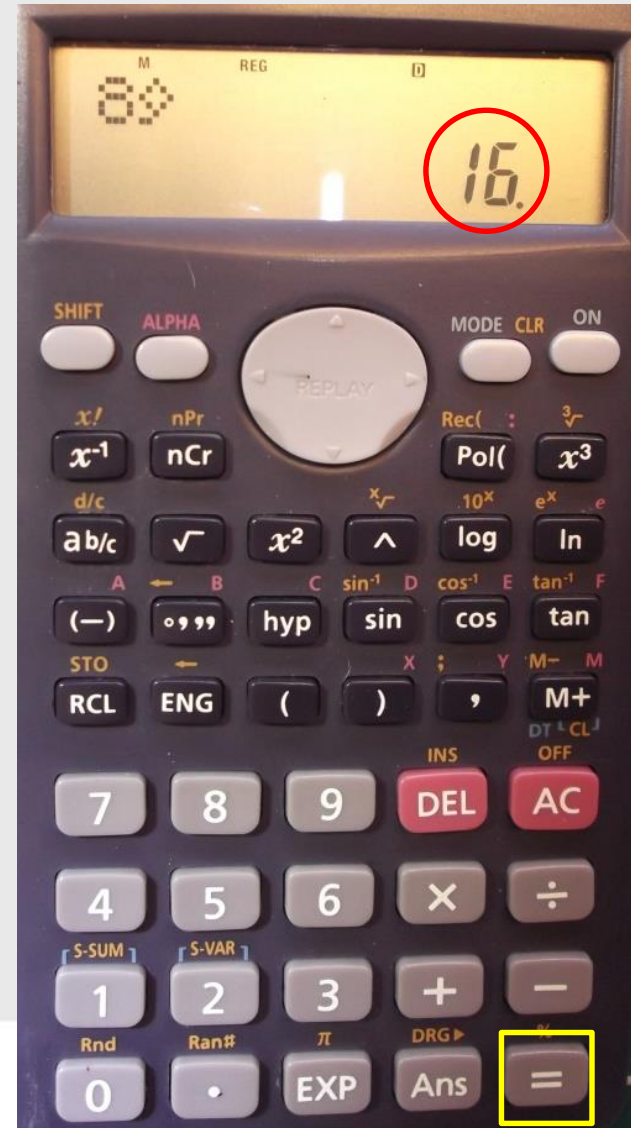
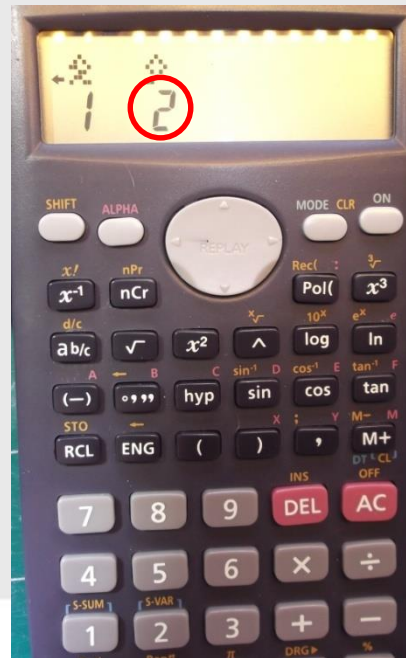
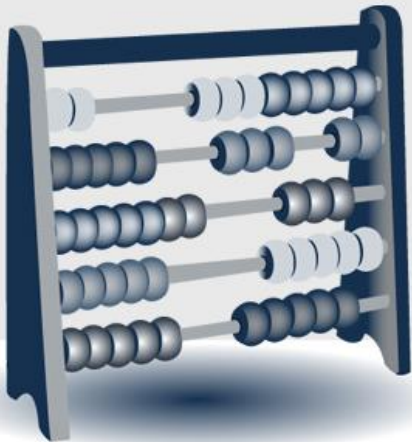


Menu 1: Previsão do valor de X

Menu 2: Previsão do valor de Y

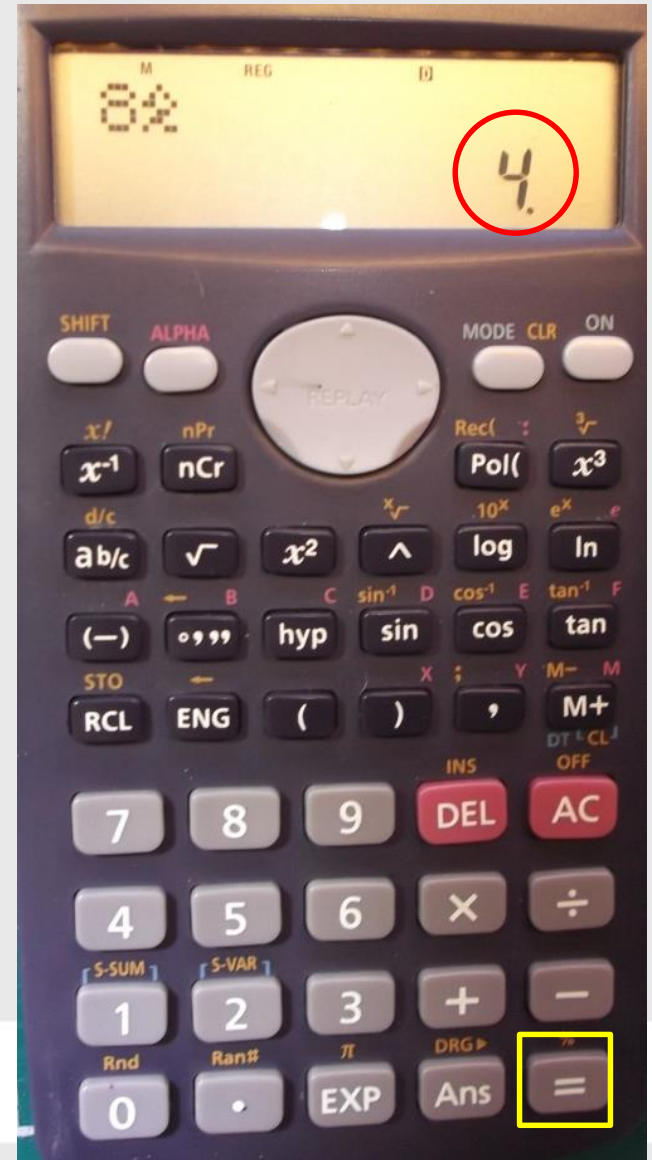
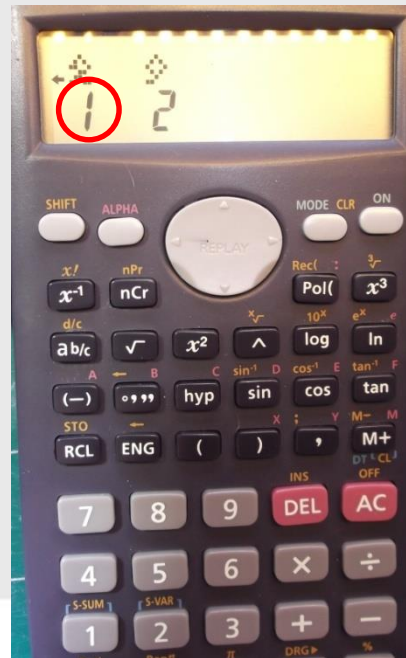
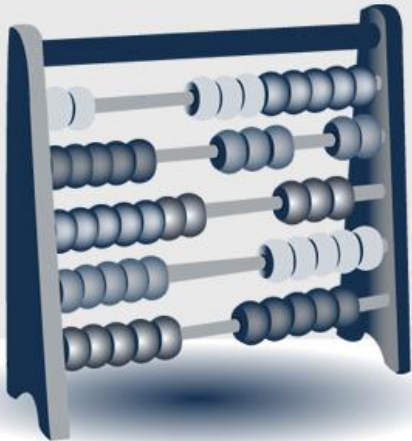
Usando a calculadora

- Inserir o valor de X (conhecido) e em seguida selecionar o menu 2 e pressionar "=", obtendo o valor de Y.



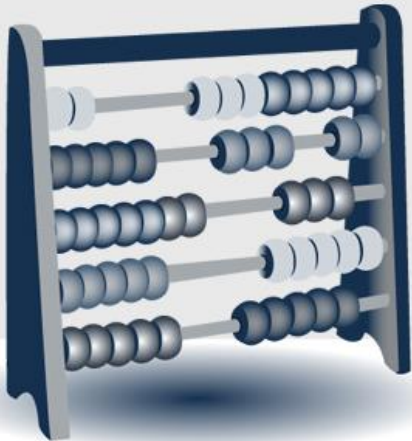
Usando a Calculadora

- Inserir o valor de Y (conhecido) e em seguida selecionar o menu 1 e pressionar "=", obtendo o valor de X.



Exercícios

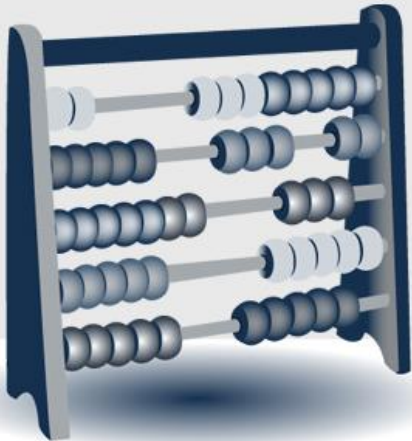
- 1. (a) Calcular os coeficientes de Regressão Linear Simples do problema dos Hambúrgueres e Ingressos de cinema.
- (b) Se em uma determinada cidade o hambúrguer custar \$ 6,00, quanto custariam dois ingressos de cinema?
- (c) Se em outra cidade os ingressos de cinemas custarem \$ 15,00, quanto custaria o hambúrguer?



Exercícios

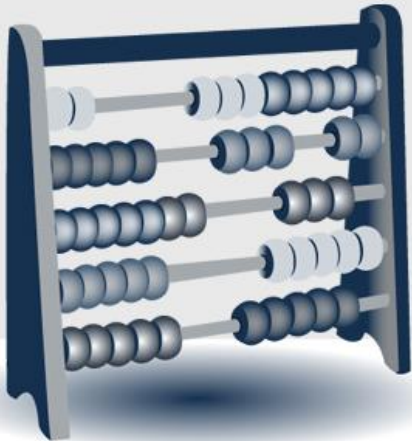
- 2. Queremos prever as vendas anuais para todas as novas filiais de uma grande cadeia de lojas com base no tamanho da loja. Para examinar a relação entre o tamanho de uma loja medido em milhares de pés quadrados, e as suas vendas anuais em milhões de dólares, foram coletados dados em 14 filiais e estão na tabela abaixo:

Loja	Área	Vendas Anuais	Loja	Área	Vendas Anuais
1	1,7	3,7	8	1,1	2,7
2	1,6	3,9	9	3,2	5,5
3	2,8	6,7	10	1,5	2,9
4	5,6	9,5	11	5,2	10,7
5	1,3	3,4	12	4,6	7,6
6	2,2	5,6	13	5,8	11,8
7	1,3	3,7	14	3,0	4,1



Exercícios

- (a) Existe alguma relação linear entre a área da loja e o volume de vendas?
- (b) Se uma nova loja for construída, que área deveria ter pra obter um volume de vendas de US\$ 6 milhões?
- (c) Se uma loja tem área de 5 mil pés quadrados, qual seria seu total de vendas em milhões de dólares?



Bibliografia

- Levine, Stephan, Krehbiel e Berenson.
Estatística Teoria e Aplicações usando o
Microsoft Excel. Quinta Edição. Ed. LTC.
- Imagens da Internet.

