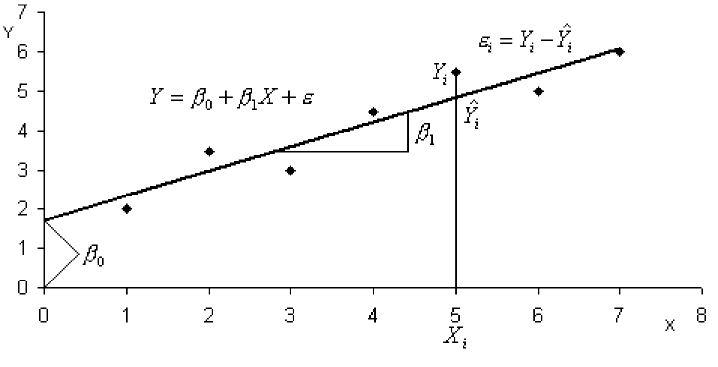
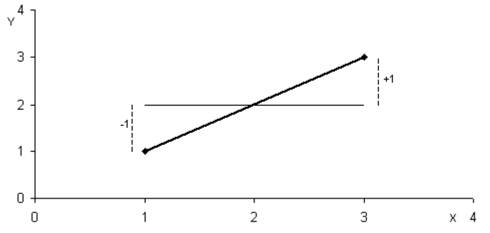
REGRESSÃO E CORRELAÇÃO

AJUSTE DE UMA RETA

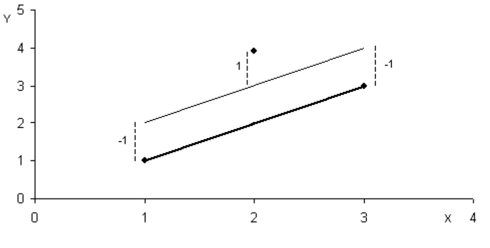


MINIMIZAÇÃO DOS DESVIOS



$$\sum\left(Y_i - \hat{Y}_i\right)$$

MINIMIZAÇÃO DOS DESVIOS ABSOLUTOS



$$\sum\left|Y_i - \hat{Y}_i\right|$$

EXEMPLO 1

- Considere o seguinte conjunto de pontos

X	Y
1	1
2	1
3	2
4	2
5	4

RETAS DE AJUSTE

R1	Y=-0.1+0.7X
R2	Y= 0.5+0.5X
R3	Y=-0.7+0.9X

RETAS

R1	R2	R3
0.6	1	0.2
1.3	1.5	1.1
2	2	2
2.7	2.5	2.9
3.4	3	3.8

DESVIOS

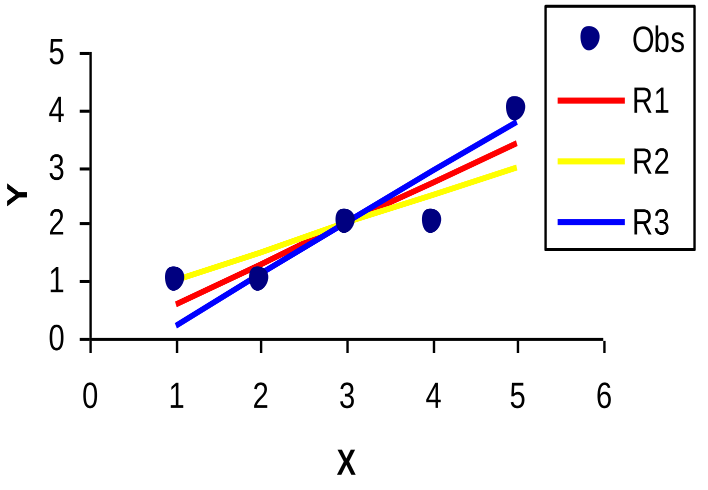
Desv1	Desv2	Desv3
0.4	0	0.8
-0.3	-0.5	-0.1
0	0	0
-0.7	-0.5	-0.9
0.6	1	0.2
0	0	0

DESVIOS ABSOLUTOS

|Desv1|	|Desv2|	|Desv3|
0.4	0	0.8
0.3	0.5	0.1
0	0	0
0.7	0.5	0.9
0.6	1	0.2
2	2	2

QUADRADO DOS DESVIOS

$(Desv1)^2$	$(Desv2)^2$	$(Desv3)^2$
0.16	0	0.64
0.09	0.25	0.01
0	0	0
0.49	0.25	0.81
0.36	1	0.04
1.10	1.50	1.50

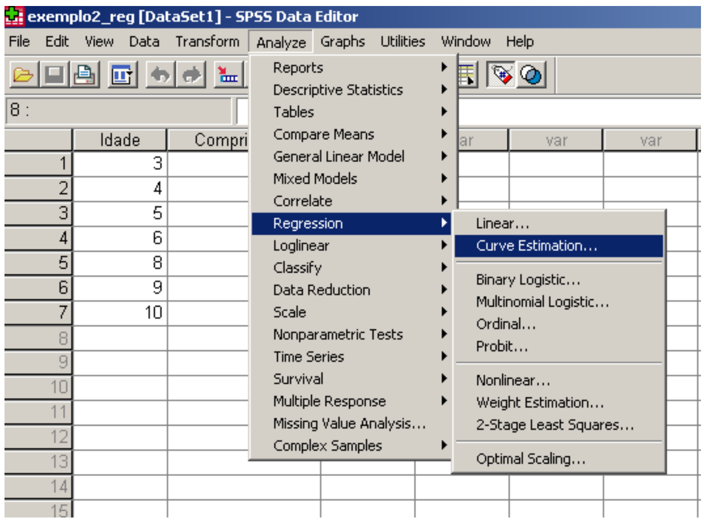


EXEMPLO 2

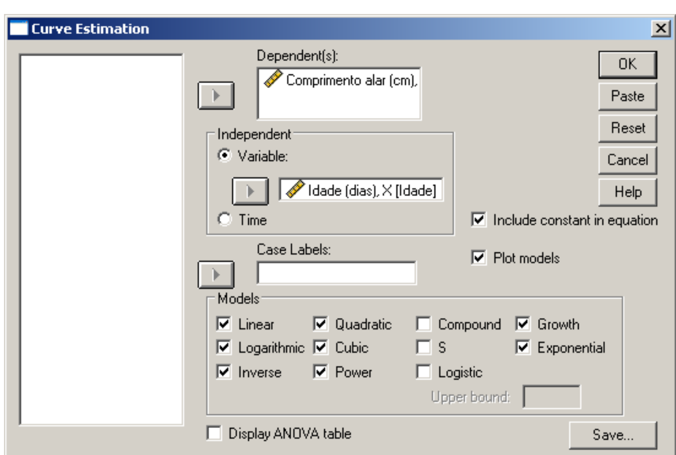
- Comprimento alar (cm) em função da idade (dias) para andorinhas

Dias	Comp.
3	1,4
4	1,5
5	2,1
6	2,4
8	3,1
9	3,2
10	3,3

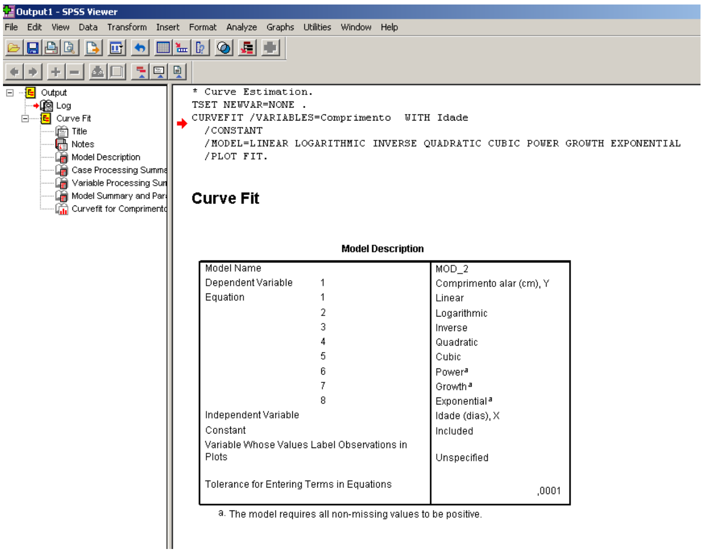
EXEMPLO 2



EXEMPLO 2



EXEMPLO 2




```
* Curve Estimation.
TSET NEWVAR=NONE .
CURVEFIT /VARIABLES=Comprimento  WITH Idade
  /CONSTANT
  /MODEL=LINEAR LOGARITHMIC INVERSE QUADRATIC CUBIC POWER GROWTH EXPONENTIAL
  /PLOT FIT.
```

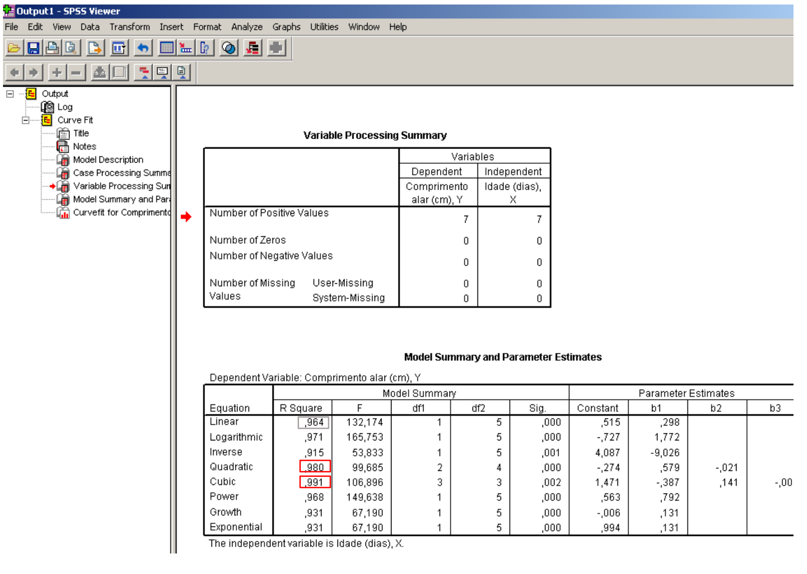
Curve Fit

Model Description

Model Name		MOD_2
Dependent Variable	1	Comprimento alar (cm), Y
Equation	1	Linear
	2	Logarithmic
	3	Inverse
	4	Quadratic
	5	Cubic
	6	Power[a]
	7	Growth[a]
	8	Exponential[a]
Independent Variable		Idade (dias), X
Constant		Included
Variable Whose Values Label Observations in Plots		Unspecified
Tolerance for Entering Terms in Equations		,0001

a. The model requires all non-missing values to be positive.

EXEMPLO 2



Variable Processing Summary

		Variables	
		Dependent	Independent
		Comprimento alar (cm), Y	Idade (dias), X
Number of Positive Values		7	7
Number of Zeros		0	0
Number of Negative Values		0	0
Number of Missing Values	User-Missing	0	0
	System-Missing	0	0

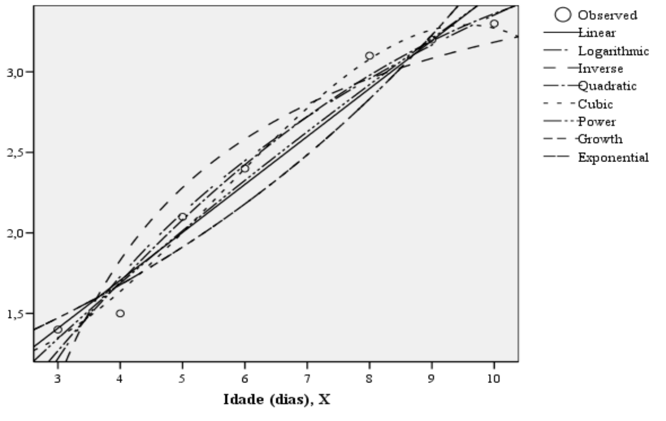
Model Summary and Parameter Estimates

Dependent Variable: Comprimento alar (cm), Y

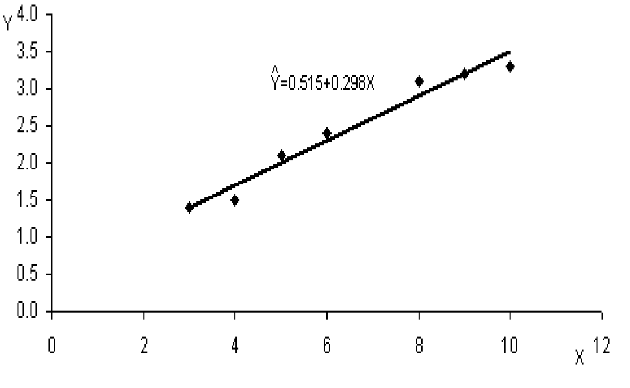
	Model Summary					Parameter Estimates			
Equation	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	,964	132,174	1	5	,000	,515	,298		
Logarithmic	,971	165,753	1	5	,000	-,727	1,772		
Inverse	,915	53,833	1	5	,001	4,087	-9,026		
Quadratic	$\boxed{,980}$	99,685	2	4	,000	-,274	,579	-,021	
Cubic	$\boxed{,991}$	106,896	3	3	,002	1,471	-,387	,141	-,00
Power	,968	149,638	1	5	,000	,563	,792		
Growth	,931	67,190	1	5	,000	-,006	,131		
Exponential	,931	67,190	1	5	,000	,994	,131		

The independent variable is Idade (dias), X.

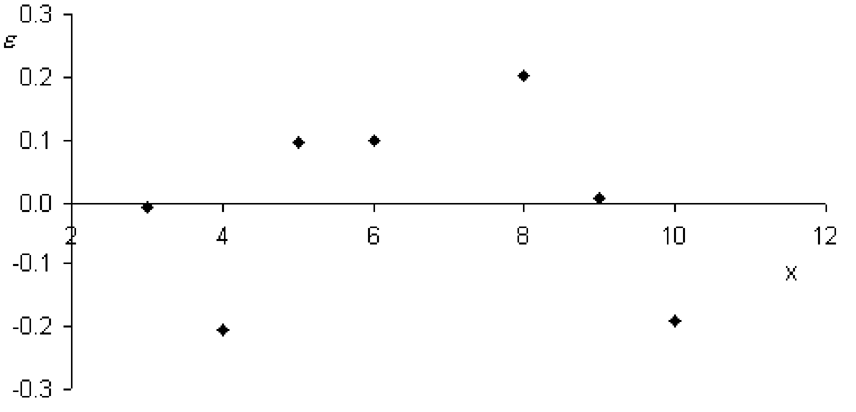
EXEMPLO 2



RETA DE MÍNIMOS QUADRADOS



RESÍDUOS



Estimadores

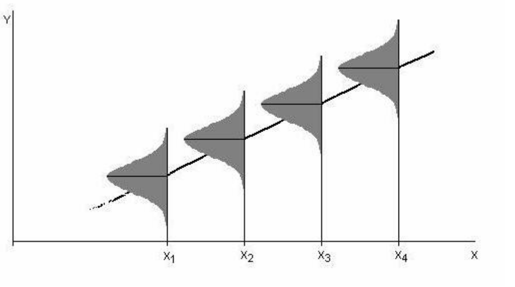
$$Y_i = \beta_0 + \beta_1.\left(X_i - \overline{X}\right) + \varepsilon_i \qquad \mathrm{i = 1,\ldots,n}$$

β_0 $\qquad \boxed{\hat{\beta}_0 = \frac{1}{n}\sum_i Y_i = \overline{Y}}$

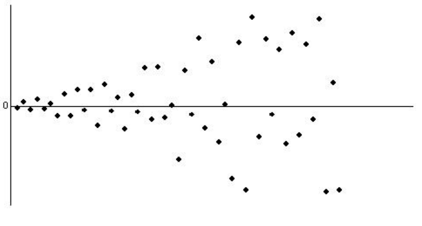
β_1 $\qquad \boxed{\hat{\beta}_1 = \frac{\sum_i \left(X_i - \overline{X}\right).\left(Y_i - \overline{Y}\right)}{\sum_i \left(X_i - \overline{X}\right)^2} = \frac{s_{XY}}{s_{xx}}}$

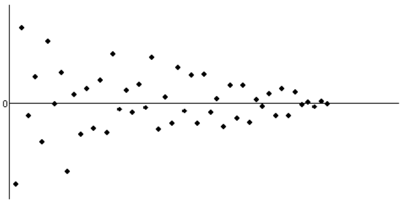
σ^2 $\qquad \boxed{s^2 = \frac{1}{n-2}\sum_i \hat{e}_i^2 = \frac{1}{n-2}\sum_i \left\{Y_i - \left[\hat{\beta}_0 + \hat{\beta}_1.\left(X_i - \overline{X}\right)\right]\right\}^2}$

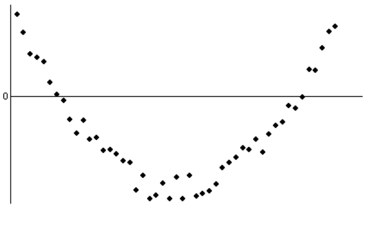
DISTRIBUIÇÃO DOS ERROS



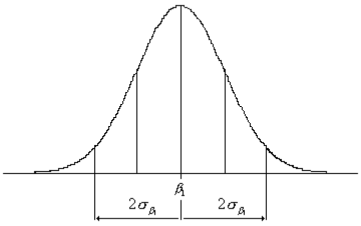
RESÍDUOS







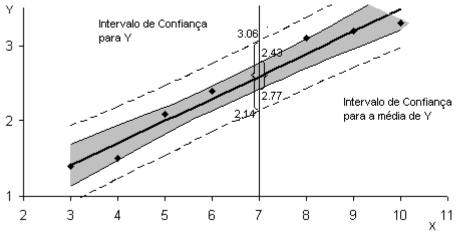
DISTRIBUIÇÃO DO DECLIVE



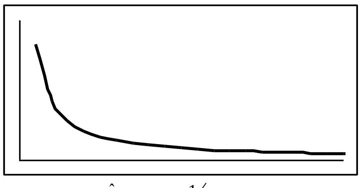
IC e Testes de hipóteses

	IC	TH
β_0	$\hat{\beta}_0 \pm t_{n-2,(\alpha/2)} \cdot \frac{s}{\sqrt{n}}$	$H_0: \beta_0 = b_0$ $H_1: \beta_0 \neq b_0, \beta_0 > b_0$ ou $\beta_0 < b_0$ $ET = \frac{\hat{\beta}_0 - b_0}{s/\sqrt{n}}$ H_0 verdadeira $\Rightarrow ET \sim t_{n-2}$
β_0'	$\left(\hat{\beta}_0 - \bar{X}.\hat{\beta}_1\right) \pm t_{n-2,(\alpha/2)}.s.\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{s_{XX}}}$	$H_0: \beta_0' = b_0'$ $H_1: \beta_0' \neq b_0', \beta_0' > b_0'$ ou $\beta_0' < b_0'$ $ET = \frac{\left(\hat{\beta}_0 - \bar{X}.\hat{\beta}_1\right) - b_0'}{s.\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{s_{XX}}}}$ H_0 verdadeira $\Rightarrow ET \sim t_{n-2}$
β_1	$\hat{\beta}_1 \pm t_{n-2,(\alpha/2)} \cdot \frac{s}{\sqrt{s_{XX}}}$	$H_0: \beta_1 = b_{10}$ $H_1: \beta_1 \neq b_{10}, \beta_1 > b_{10}$ ou $\beta_1 < b_{10}$ $ET = \frac{\hat{\beta}_1 - b_{10}}{\frac{s}{\sum_i (X_i - \bar{X})^2}}$ H_0 verdadeira $\Rightarrow ET \sim t_{n-2}$

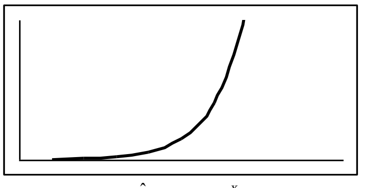
INTERVALO DE CONFIANÇA



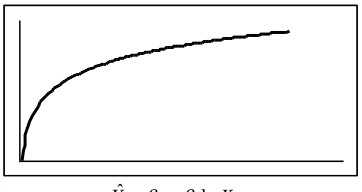
REGRESSÃO NÃO LINEAR



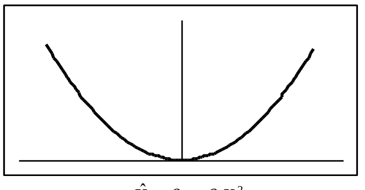
$\hat{Y} = \beta_0 + \beta_1 \frac{1}{X}$



$\hat{Y} = \beta_0 + \beta_1 e^X$



$\hat{Y} = \beta_0 + \beta_1 \ln X$



$\hat{Y} = \beta_0 + \beta_1 X^2$

REGRESSÃO NÃO LINEAR

Modelo	Transformação
• $Y_i = \alpha' + \frac{\beta}{X_i} + e_i$	$U_i = \frac{1}{X_i}$ $Y_i = \alpha' + \beta.U_i + e_i$
• $Y_i = e^{\alpha' + \beta.X_i + e_i}$	$Z_i = \ln Y_i$ $Z_i = \alpha' + \beta.X_i + e_i$
• $Y_i = e^{\alpha' + \frac{\beta}{X_i} + e_i}$ com $\alpha' > 0, \beta < 0$	$U_i = \frac{1}{X_i}$ $Z_i = \ln Y_i$ $Z_i = \alpha' + \beta.U_i + e_i$

COEFICIENTE DE CORRELAÇÃO

Coeficiente de correlação de Pearson

$$R = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = \frac{s_{XY}}{\sqrt{s_{XX}} \cdot \sqrt{s_{YY}}}$$

TESTES DE ASSOCIAÇÃO

Unilateral à direita	Unilateral à esquerda	Bilateral
$H_0 : \rho = 0$	$H_0 : \rho = 0$	$H_0 : \rho = 0$
$H_1 : \rho > 0$	$H_1 : \rho < 0$	$H_1 : \rho \neq 0$

Estatística de teste $t = \dfrac{r.\sqrt{n-2}}{\sqrt{1-r^2}}$

Região de Rejeição:

$t > t_{n-2,(\alpha)}$	$t < -t_{n-2,(\alpha)}$	$\|t\| > t_{n-2,(\alpha/2)}$

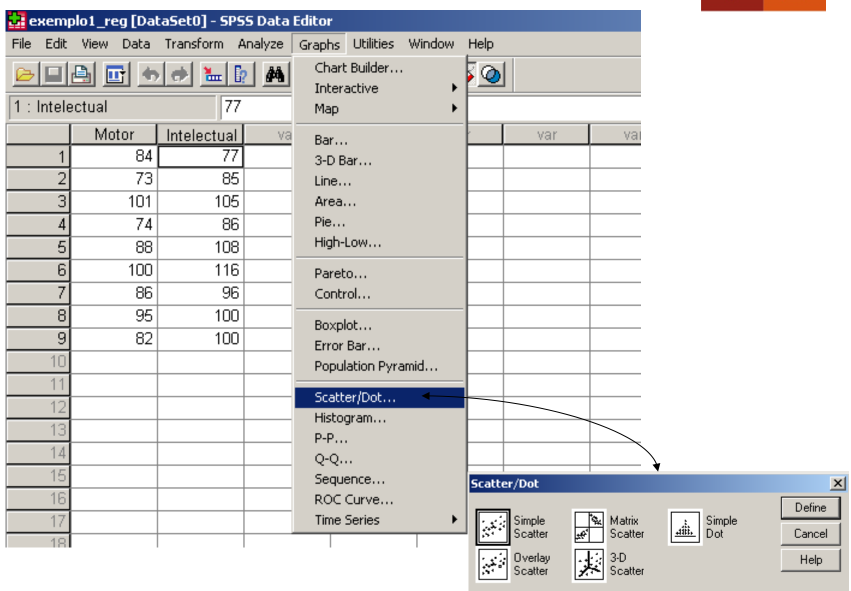
29

EXEMPLO

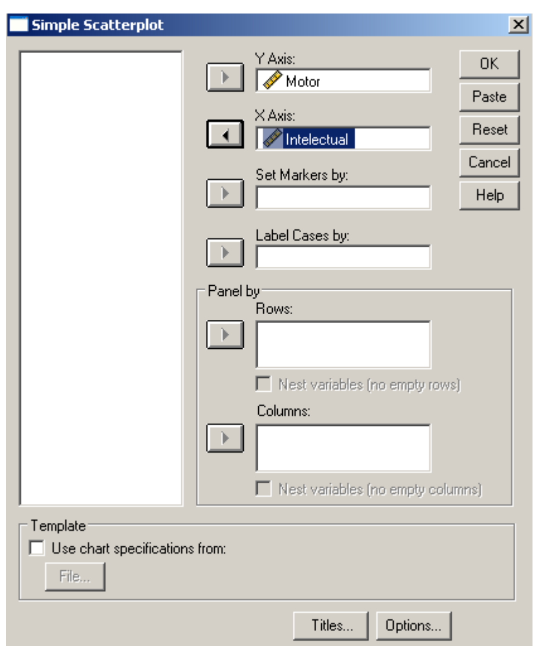
- Índice de Desenvolvimento de Griffiths
  - avaliações motora e intelectual para 9 crianças com a idade de 4 anos

Motor	Intelectual
84	77
73	85
101	105
74	86
88	108
100	116
86	96
95	100
82	100

30

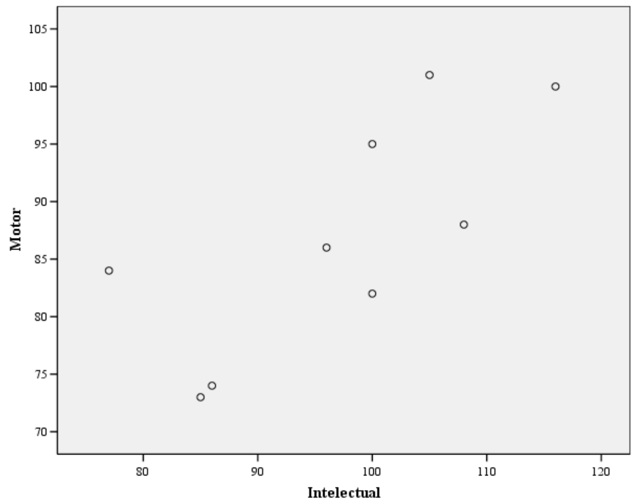

exemplo1_reg [DataSet0] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : Intelectual 77
Motor Intelectual
1 84 77
2 73 85
3 101 105
4 74 86
5 88 108
6 100 116
7 86 96
8 95 100
9 82 100
Chart Builder...
Interactive
Map
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Pareto...
Control...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
P-P...
Q-Q...
Sequence...
ROC Curve...
Time Series
Scatter/Dot
Simple Scatter
Matrix Scatter
Simple Dot
Overlay Scatter
3-D Scatter
Define
Cancel
Help

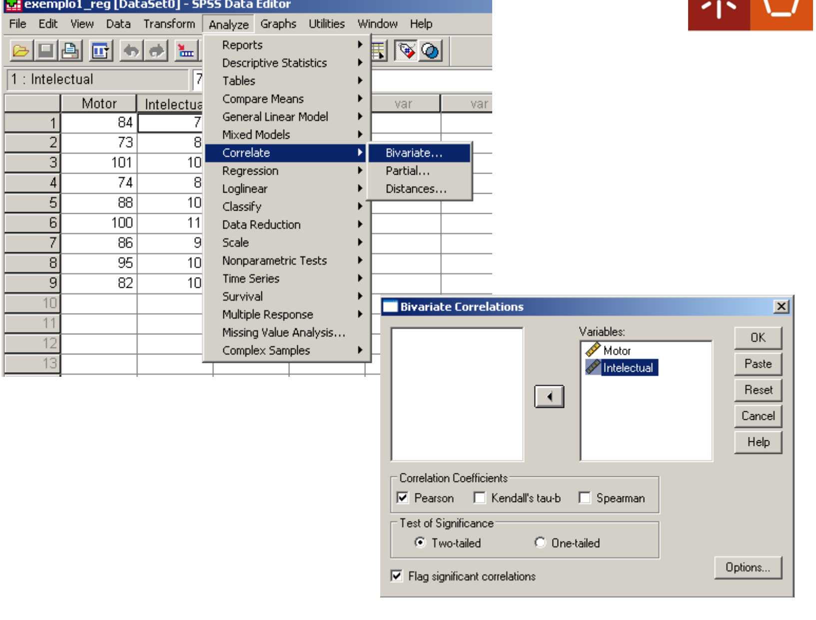
31

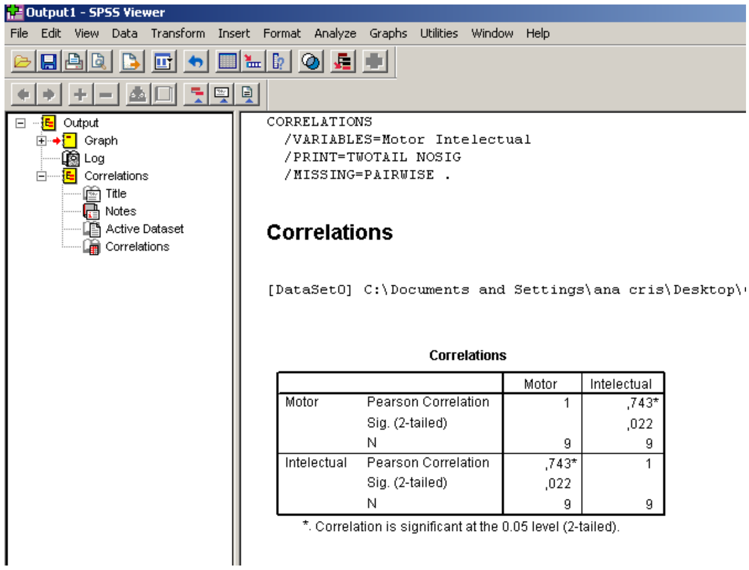

Simple Scatterplot
Y Axis:
Motor
X Axis:
Intelectual
Set Markers by:
Label Cases by:
Panel by
Rows:
Nest variables (no empty rows)
Columns:
Nest variables (no empty columns)
Template
Use chart specifications from:
File...
OK
Paste
Reset
Cancel
Help
Titles...
Options...

32

DIAGRAMA DE DISPERSÃO






Output1 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Output
Graph
Log
Correlations
Title
Notes
Active Dataset
Correlations

```
CORRELATIONS
  /VARIABLES=Motor Intelectual
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE .
```

Correlations

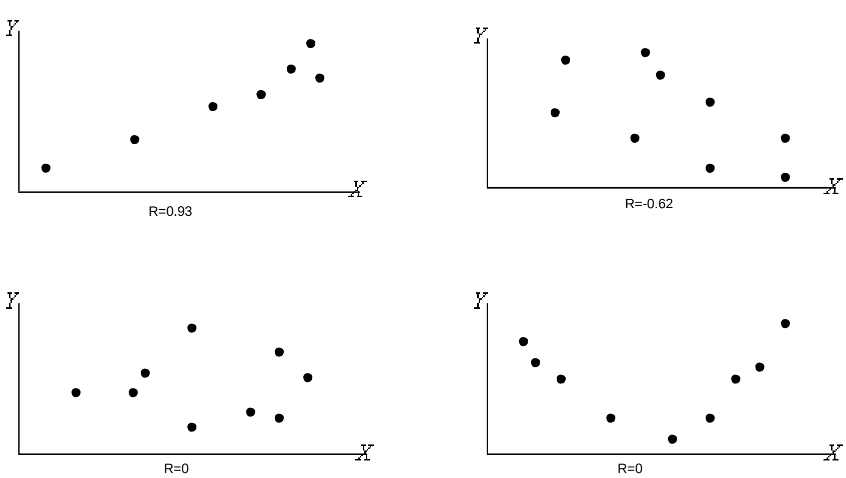
[DataSet0] C:\Documents and Settings\ana cris\Desktop\

Correlations

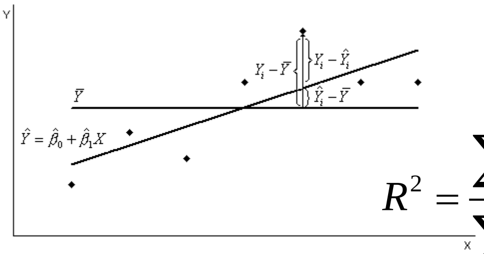
		Motor	Intelectual
Motor	Pearson Correlation	1	,743*
	Sig. (2-tailed)		,022
	N	9	9
Intelectual	Pearson Correlation	,743*	1
	Sig. (2-tailed)	,022	
	N	9	9

*. Correlation is significant at the 0.05 level (2-tailed).

CORRELAÇÃO



COEFICIENTE DE DETERMINAÇÃO



$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

Coeficiente de determinação (r^2), representa a proporção da variação de Y que é explicada pela regressão

$$r^2 = \frac{\hat{\beta}_1^{\,2}.s_{XX}}{s_{YY}} = \frac{\hat{\beta}_1^{\,2}.\sum_i\left(X_i - \bar{X}\right)^2}{\sum_i\left(Y_i - \bar{Y}\right)^2} = \frac{\text{variação de } Y \text{ explicada pela regressão}}{\text{variação total de } Y}$$