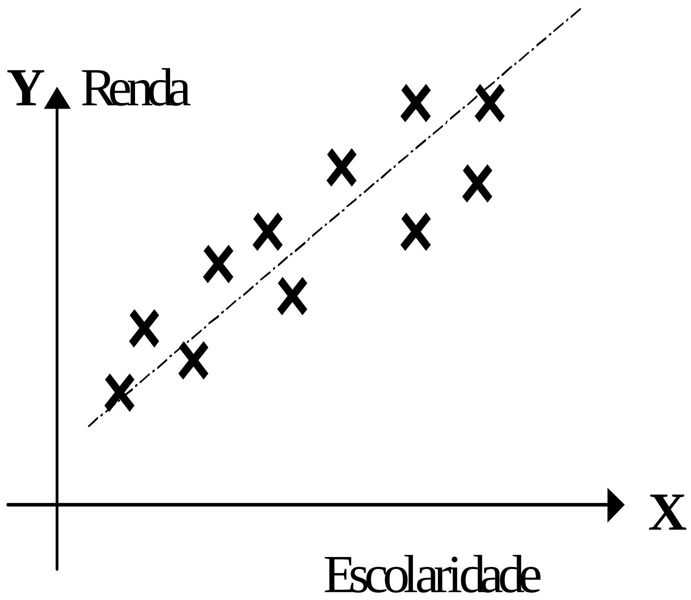
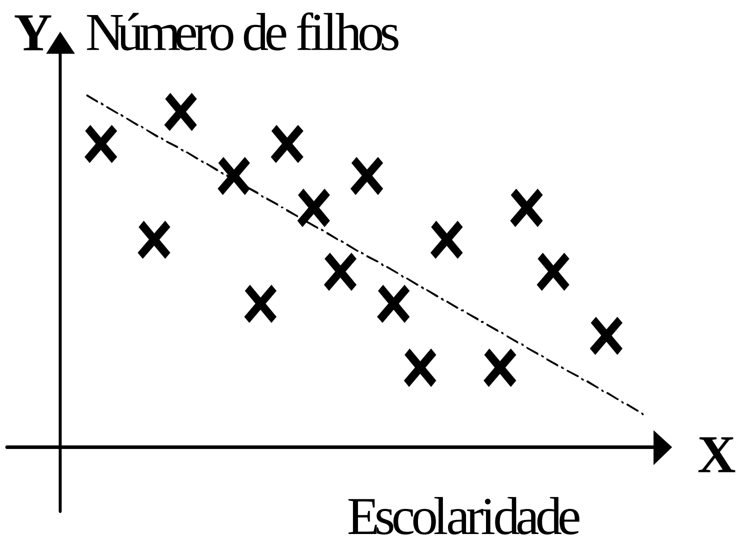
CORRELAÇÃO E REGRESSÃO

CORRELAÇÃO LINEAR SIMPLES

- Diagramas de Dispersão
- Variáveis explicativas e explicadas





COEFICIENTE DE CORRELAÇÃO DE PEARSON

- Correlações positiva e negativa
- Covariância

$$r = \frac{\sigma^2_{XY}}{\sigma_X \, \sigma_Y}$$

,

$$\overline{X} = \frac{\sum X}{N} \qquad \overline{Y} = \frac{\sum Y}{N} \qquad \overline{XY} = \frac{\sum XY}{N}$$

$$\overline{X^2} = \frac{\sum X^2}{N} \qquad \overline{Y^2} = \frac{\sum Y^2}{N}$$

$$\sigma^2_X = \overline{X^2} - \overline{X}^2 \qquad \sigma^2_Y = \overline{Y^2} - \overline{Y}^2 \qquad \sigma^2_{XY} = \overline{XY} - \overline{X}\,\overline{Y}$$

EXEMPLO

- Anos de estudos de pais e filhos

A(12, 12); B(10 , 8); C(6 , 6); D(16 , 11); E(8 , 10); F(9 , 8) e G(12 , 11)

Dupla	*Pais (**X**)*	*Filhos (**Y**)*
A	12	12
B	10	8
C	6	6
D	16	11
E	8	10
F	9	8
G	12	11
N = 7	73	66

EXEMPLO

- Anos de estudos de pais e filhos

Dupla	*Pais (X)*	*Filhos (Y)*	X^2	Y^2
A	12	12	144	144
B	10	8	100	64
C	6	6	36	36
D	16	11	256	121
E	8	10	64	100
F	9	8	81	64
G	12	11	144	121
N = 7	73	66	825	650

EXEMPLO

- Anos de estudos de pais e filhos

Dupla	*Pais* (X)	*Filhos* (Y)	X^2	Y^2	XY
A	12	12	144	144	144
B	10	8	100	64	80
C	6	6	36	36	36
D	16	11	256	121	176
E	8	10	64	100	80
F	9	8	81	64	72
G	12	11	144	121	132
N=7	73	66	825	650	720

EXEMPLO

$$\bar{X} = \frac{\sum X}{N} = \frac{73}{7} = 10{,}429$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{66}{7} = 9{,}429$$

$$\overline{X^2} = \frac{\sum X^2}{N} = \frac{825}{7} = 117{,}857$$

$$\overline{Y^2} = \frac{\sum Y^2}{N} = \frac{650}{7} = 92{,}857$$

$$\sigma_X^2 = \overline{X^2} - \bar{X}^2 = 117{,}857 - 10{,}429^2 = 9{,}075$$

$$\sigma_Y^2 = \overline{Y^2} - \bar{Y}^2 = 92{,}857 - 9{,}429^2 = 3{,}951$$

$$\sigma_X = \sqrt{\sigma_X^2} = 3{,}012$$

$$\sigma_Y = \sqrt{\sigma_Y^2} = 1{,}988$$

$$\overline{XY} = \frac{\sum XY}{N} = \frac{720}{7} = 102{,}857$$

$$\sigma_{XY}^2 = \overline{XY} - \bar{X}\bar{Y} = 102{,}857 - 10{,}429 \times 9{,}429 = 4{,}522$$

$$r = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y} = \frac{4{,}522}{3{,}012 \times 1{,}988} = 0{,}755$$

REGRESSÃO LINEAR SIMPLES

- Ajustamento do modelo

$$b = \frac{\sigma^2_{XY}}{\sigma^2_X} \qquad\qquad a = \overline{Y} - b\overline{X}$$

$$\widehat{Y} = a + bX$$

- Poder explicativo do modelo: r^2

EXEMPLO

$$a = \overline{Y} - b\overline{X} = 9{,}429 - 0{,}498 \text{ x } 10{,}429 = 4{,}201$$

$$b = \frac{\sigma^2_{XY}}{\sigma^2_{X}} = \frac{4{,}522}{9{,}075} = 0{,}498$$

$$\widehat{Y} = a + bX = 4{,}201 + 0{,}498\,X$$

$$r^2 = 0{,}75^2 = 0{,}5625$$

REGRESSÃO LINEAR POR TRANSFORMAÇÃO

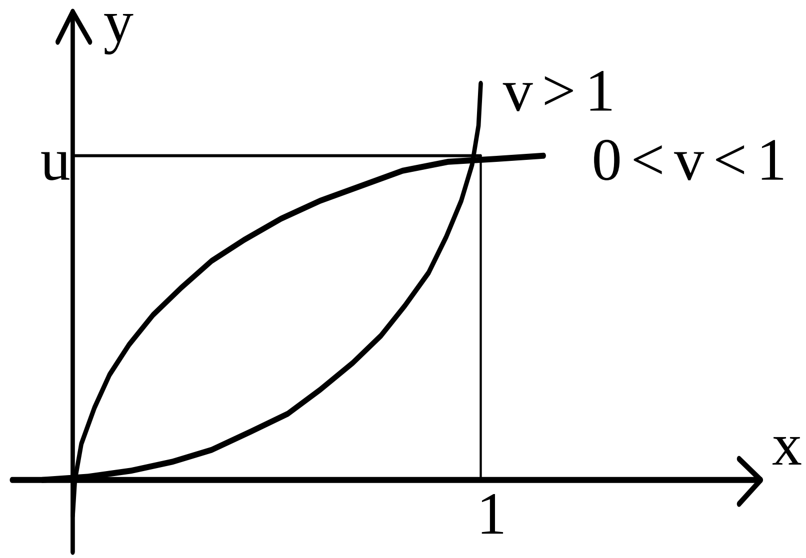
- Função potencial

A função teórica é do tipo:

$y = u\,x^{v}$ onde:

$u > 0$ e

$v > 0$



REGRESSÃO LINEAR POR TRANSFORMAÇÃO

- Função potencial

Aplicando a transformação abaixo:

$$\log y = \log\left(u\, x^{v}\right) = \log u + \log x^{v} = \log u + v \log x$$

Que, comparada à função linear estudada:

$$Y = a + bX$$

Leva às relações abaixo, que permitem a reversão da forma original para a transformada e vice-versa. Assim todo o tratamento dado à função linear pode ser aproveitado por essa função linearizável.

$$Y = \log y \qquad X = \log x$$

$$a = \log u \qquad b = v$$

REGRESSÃO LINEAR POR TRANSFORMAÇÃO

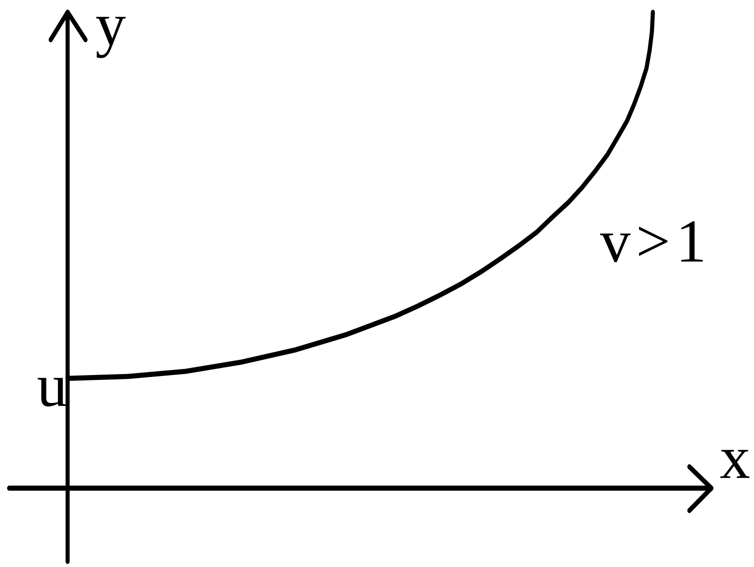
- Função exponencial

A função teórica é do tipo:

$y = u v^{x}$ onde:

$u > 0$ e

$v > 1$



REGRESSÃO LINEAR POR TRANSFORMAÇÃO

- Função exponencial

Aplicando a transformação abaixo:

$$\log y = \log\left(u\, v^{x}\right) = \log u + \log v^{x} = \log u + \left(\log v\right) x$$

Que, comparada à função linear estudada:

$$Y = a + bX$$

Leva às relações abaixo, que permitem a reversão da forma original para a transformada e vice-versa. Assim todo o tratamento dado à função linear pode ser aproveitado por essa função linearizável.

$$Y = \log y \qquad X = x$$

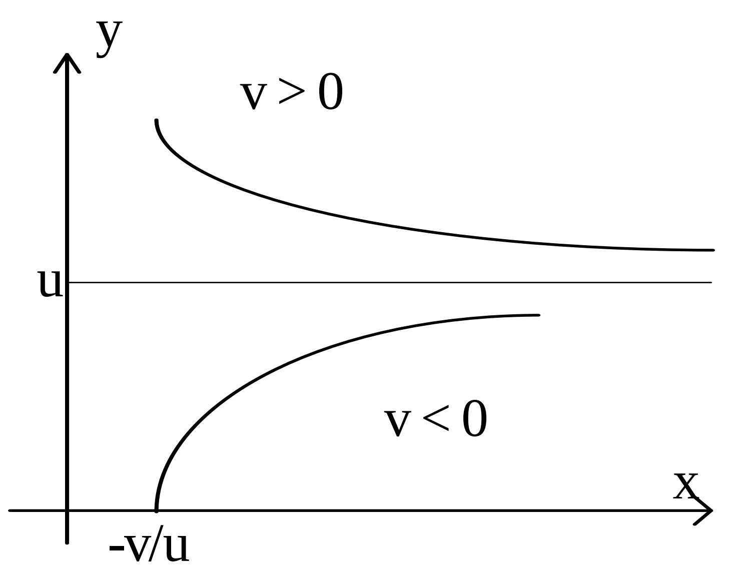
$$a = \log u \qquad b = \log v$$

REGRESSÃO LINEAR POR TRANSFORMAÇÃO

- Função hiperbólica (primeiro tipo)

A função teórica é do tipo:

$$y = u + \frac{v}{x}$$



REGRESSÃO LINEAR POR TRANSFORMAÇÃO

- Função hiperbólica (primeiro tipo)

Aplicando a transformação seguinte:

$$y = u + \frac{v}{x} = u + v\,x^{-1}$$

Que, comparada à função linear estudada:

$$Y = a + bX$$

Leva às relações abaixo, que permitem a reversão da forma original para a transformada e vice-versa. Assim todo o tratamento dado à função linear pode ser aproveitado por essa função linearizável.

$$Y = y \qquad X = \frac{1}{x} = x^{-1}$$

$$a = u \qquad b = v$$

REGRESSÃO LINEAR POR TRANSFORMAÇÃO

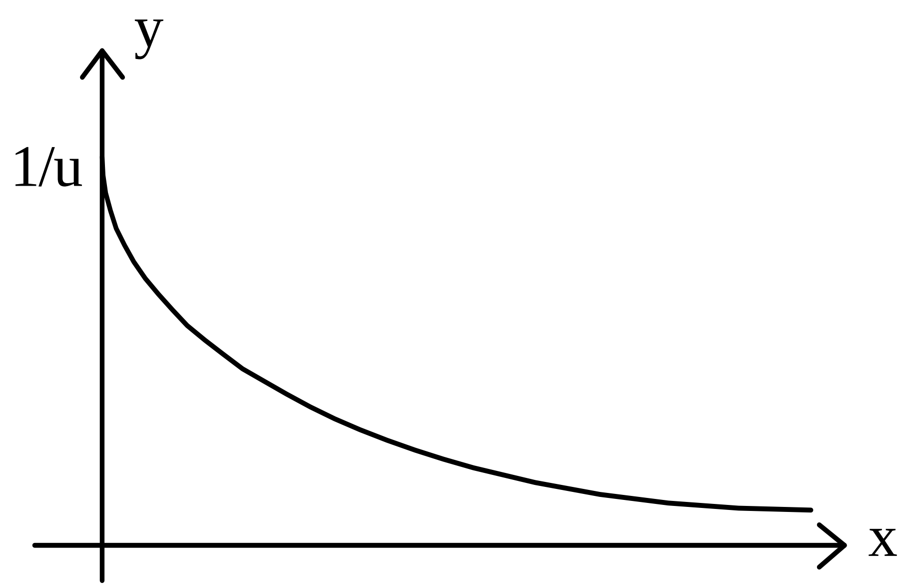
- Função hiperbólica (segundo tipo)

A função teórica é do tipo:

$$y = \frac{1}{u + v\,x}$$ onde

$u > 0$ e

$v > 0$



REGRESSÃO LINEAR POR TRANSFORMAÇÃO

- Função hiperbólica (segundo tipo)

Aplicando a transformação abaixo:

$$\frac{1}{y} = u + v\ x$$

Que, comparada à função linear estudada:

$$Y = a + bX$$

Leva às relações abaixo, que permitem a reversão da forma original para a transformada e vice-versa. Assim todo o tratamento dado à função linear pode ser aproveitado por essa função linearizável.

$$Y = \frac{1}{y} = y^{-1} \qquad X = x$$

$$a = u \qquad b = v$$

CORRELAÇÃO ORDINAL

- Coeficiente de Spearman

$$r_S = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

EXEMPLO

- Habilidade e QI das mecanógrafas

Nomes	*Q.I.*	*Destreza (**X**º)*
A	100	5
B	90	2
C	80	6
D	130	1
E	100	8
F	100	9
G	110	7
H	120	3
I	90	11
J	120	10
K	110	12
L	140	4
12		

EXEMPLO

- Habilidade e QI das mecanógrafas

Nomes	*Q.I.*	*Destreza (**Xº**)*	*Q.I. (**Yº**)*
A	100	5	8
B	90	2	10,5
C	80	6	12
D	130	1	2
E	100	8	8
F	100	9	8
G	110	7	5,5
H	120	3	3,5
I	90	11	10,5
J	120	10	3,5
K	110	12	5,5
L	140	4	1
12			

EXEMPLO

- Habilidade e QI das mecanógrafas

Nomes	*Q.I.*	*Destreza (**Xº**)*	*Q.I. (**Yº**)*	*Desvio (**D**)*	***D***^{2}
A	100	5	8	-3	9
B	90	2	10,5	-8,5	72,25
C	80	6	12	-6	36
D	130	1	2	-1	1
E	100	8	8	0	0
F	100	9	8	1	1
G	110	7	5,5	1,5	2,25
H	120	3	3,5	-0,5	0,25
I	90	11	10,5	0,5	0,25
J	120	10	3,5	6,5	42,25
K	110	12	5,5	6,5	42,25
L	140	4	1	3	9
12					215,5

EXEMPLO

- Habilidade e QI das mecanógrafas

$$r_S = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6\,(215{,}5)}{12\,(12^2-1)} = -0{,}247$$

CORRELAÇÃO ORDINAL

- Coeficiente Gama de Goodman e Kruskal

$$G = \frac{f_A - f_I}{f_A + f_I}$$

EXEMPLO

- Discriminação racial no emprego

Cidades	*Tamanho da população negra (**Xº**)*	*Nível de discriminação no emprego (**Yº**)*
A	1	2
B	2	3
C	3	1
D	4	6
E	5	5
F	6	4

EXEMPLO

- Discriminação racial no emprego

Cidades	*Tamanho da população negra (**Xº**)*	*Nível de discriminação no emprego (**Yº**)*	*Acordos*
A	1	2	0
B	2	3	1
C	3	1	0
D	4	6	3
E	5	5	3
F	6	4	3
			10

EXEMPLO

- Discriminação racial no emprego

Cidades	*Tamanho da população negra (**Xº**)*	*Nível de discriminação no emprego (**Yº**)*	*Acordos*	*Inversões*
A	1	2	0	0
B	2	3	1	0
C	3	1	0	2
D	4	6	3	0
E	5	5	3	1
F	6	4	3	2
			10	5

EXEMPLO

- Discriminação racial no emprego

$$G = \frac{f_A - f_I}{f_A + f_I} = \frac{10 - 5}{10 + 5} = +0,33$$