

# HEP 57800 – Bioestatística

DATA		Aula	CONTEÚDO PROGRAMÁTICO
05/03	Terça	1	Níveis de mensuração, variáveis, organização de dados, apresentação tabular
07/03	Quinta	2	Apresentação tabular e gráfica
12/03	Terça	3	Apresentação gráfica; medidas de tendência central e de posição
14/03	Quinta	4	Medidas de tendência central e de posição; medidas de dispersão ou de variabilidade
19/03	Terça	5	Medidas de correlação, noções de regressão linear simples, estimando a equação da reta
21/03	Quinta	6	Medidas de associação
26/03	Terça	7	Consolidação de conteúdo - Exercícios
28/03	Quinta	8	Avaliação 1
09/04	Terça	9	Noções de probabilidade; noções de amostragem; distribuição binomial
11/04	Quinta	10	Distribuição normal, distribuição amostral da média
16/04	Terça	11	Teste de hipóteses de parâmetros populacionais – conceitos; teste de hipóteses de uma proporção populacional
18/04	Quinta	12	Teste de hipóteses de associação
23/04	Terça	13	Teste de hipóteses de uma média populacional
25/04	Quinta	14	Teste de hipóteses de duas médias com amostras independentes e dependentes
30/04	Terça	15	Consolidação de conteúdo – Exercícios
02/06	Quinta	16	Estimação de parâmetros por intervalo de confiança: média e proporção
07/05	Terça	17	Exercícios
09/05	Quinta	18	Exercícios
14/05	Terça	19	Avaliação 2

## Noções de correlação

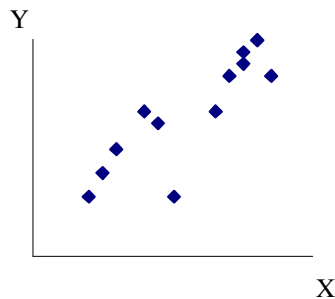
X e Y são variáveis aleatórias quantitativas

Gráfico de dispersão: deve ser feito antes da análise numérica dos dados.

É construído com conjuntos de pontos formados por pares de valores (x,y). Pode indicar correlação linear positiva, negativa ou inexistência de correlação.

Ex: X: coeficiente de mortalidade por câncer gástrico

Y: consumo médio de sal



correlação positiva

Correlation between High Salt Intake and Mortality Rates for Oesophageal and Gastric Cancers in Henan Province, China

JIAN-BANG LU AND YU-MIN QIN

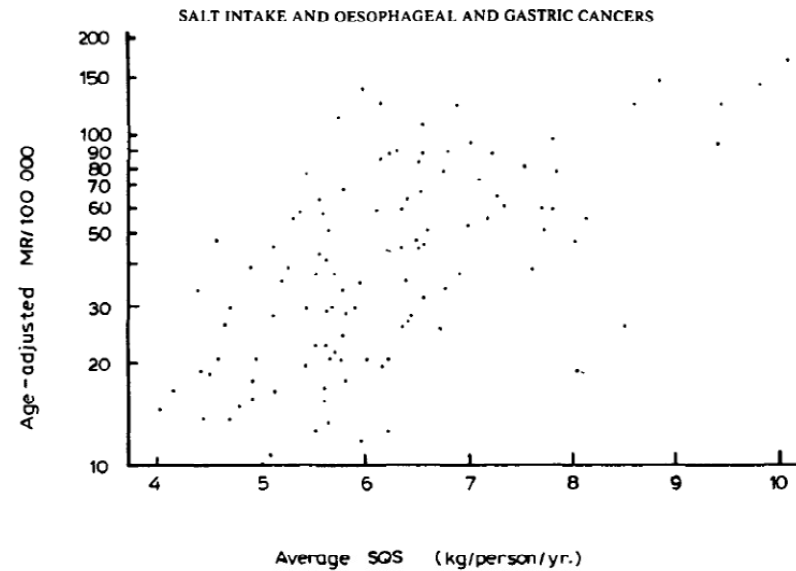
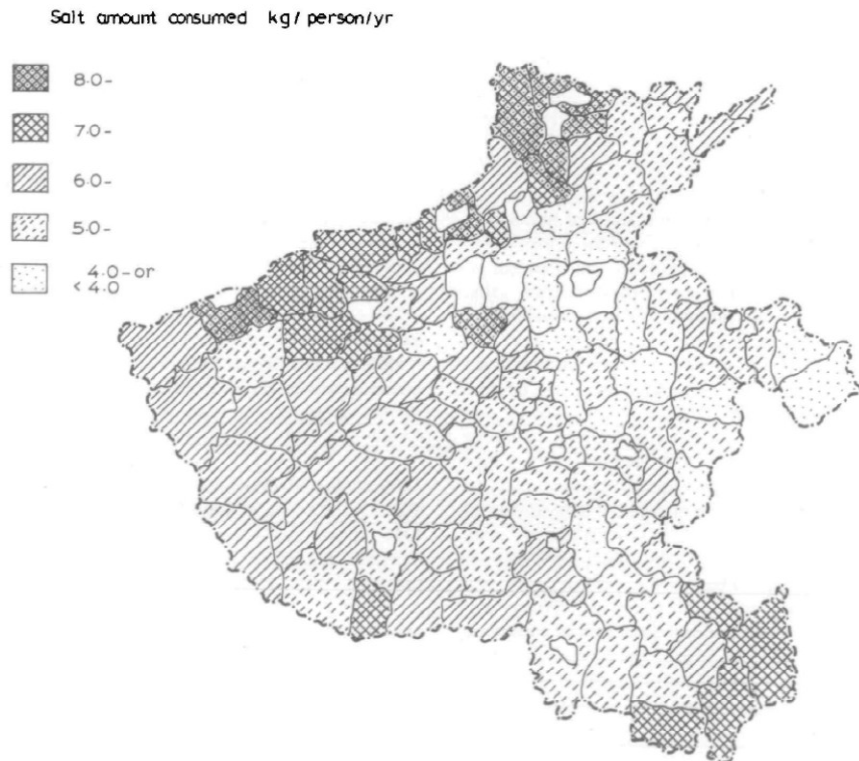
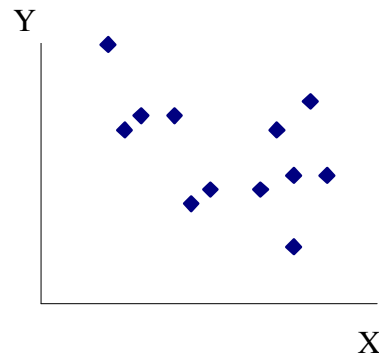


FIGURE 2 Graph of the correlation between salt quantity sold (SQS) during 1964-66, 1974-76 and the mortality rate of oesophageal cancer during 1974-76 in Henan Province, China.

TABLE 1 The rank correlation coefficient between the SQS in 1964-66, 1974-76 and mortality rate from malignant neoplasms selected in 1974-76 in Henan, China.

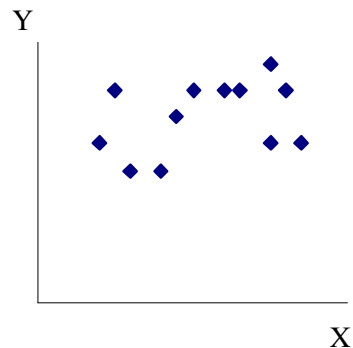
Cancer site	Sex	$\Sigma di^2$	$r_s$	p value
Oesophagus	M	81945.5	0.6097	<0.01
	F	11820.5	0.4674	<0.01
Stomach	M	77933.5	0.6288	<0.01
	F	96028.3	0.5426	<0.01
Liver	M	185273.5	0.1175	>0.05
Cervix	F	183543.3	0.1257	>0.05
Lung	M	185329.5	0.1172	>0.05
Leukaemia	M	216721.3	0.0323	>0.05

Ex: X: Peso ao nascer (gramas)  
Y: Aumento de peso entre 70 e 100 dias, como percentual de X



correlação negativa

Ex: X: coeficiente de mortalidade por câncer de colo de útero  
Y: consumo de sal



correlação inexistente

## Noções de correlação

Coeficiente de correlação de Pearson ( $\rho$ ): Mede o grau de associação entre 2 variáveis quantitativas X e Y.

### Definição:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

onde,  $\sigma_{XY}$  é a covariância de X e Y (dispersão conjunta de X e Y)

$\sigma_X$  é o desvio padrão de X (dispersão de X)

$\sigma_Y$  é o desvio padrão de Y (dispersão de Y)

## Noções de correlação

estimador ( $r$ )

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[ \sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right]}}$$

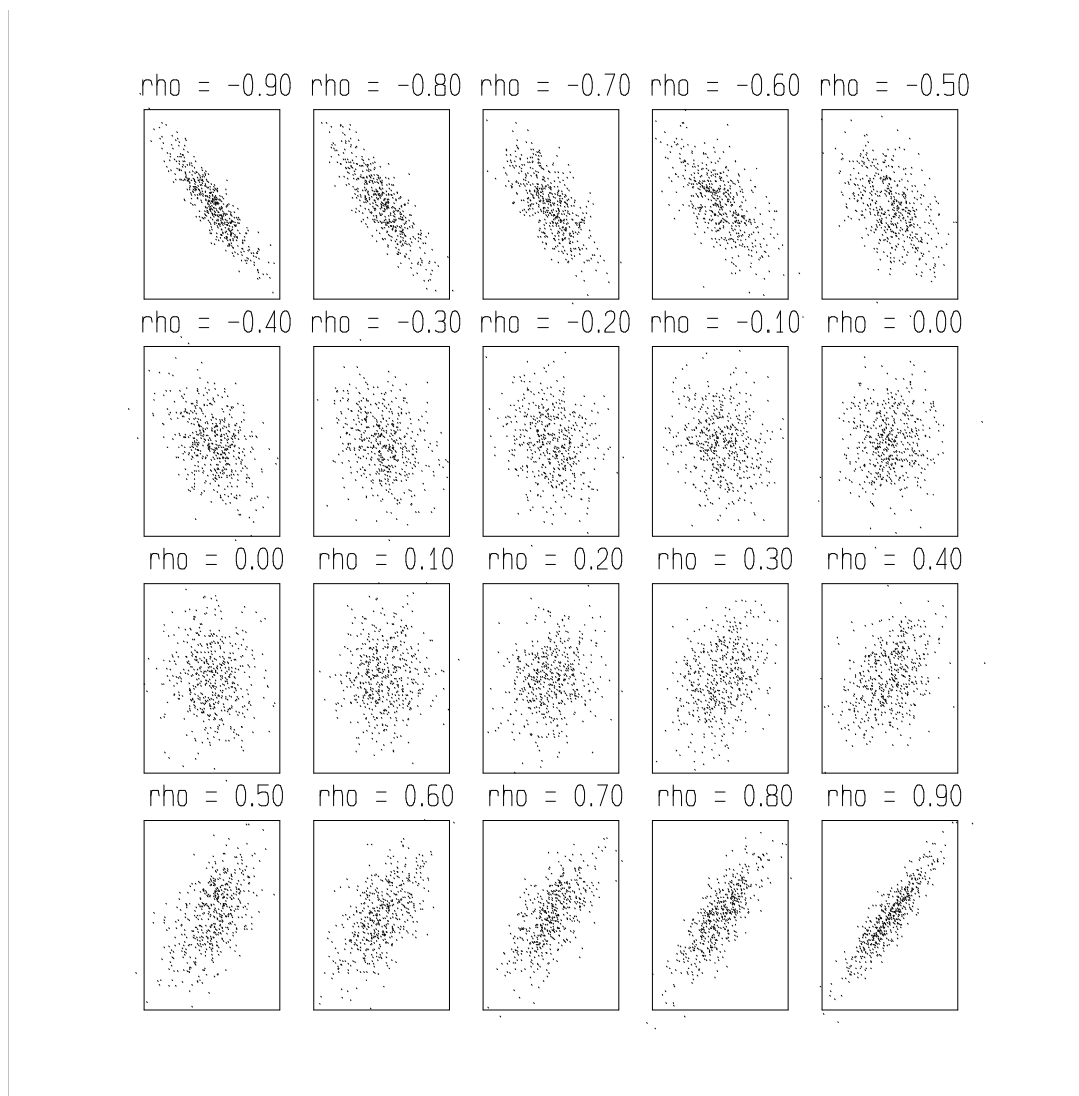
### Propriedades

$$-1 \leq \rho \leq +1$$

$\rho$  não possui dimensão, isto é, não depende da unidade de medida das variáveis X e Y

## Noções de correlação

Gráficos de dispersão para diferentes valores do coeficiente de correlação  $\rho$  (rho).



## Noções de correlação

### Exemplo:

Os dados a seguir são provenientes de um estudo que investiga a composição corporal e fornece o percentual de gordura corporal (%), idade e sexo para 18 adultos com idades entre 23 e 61 anos.

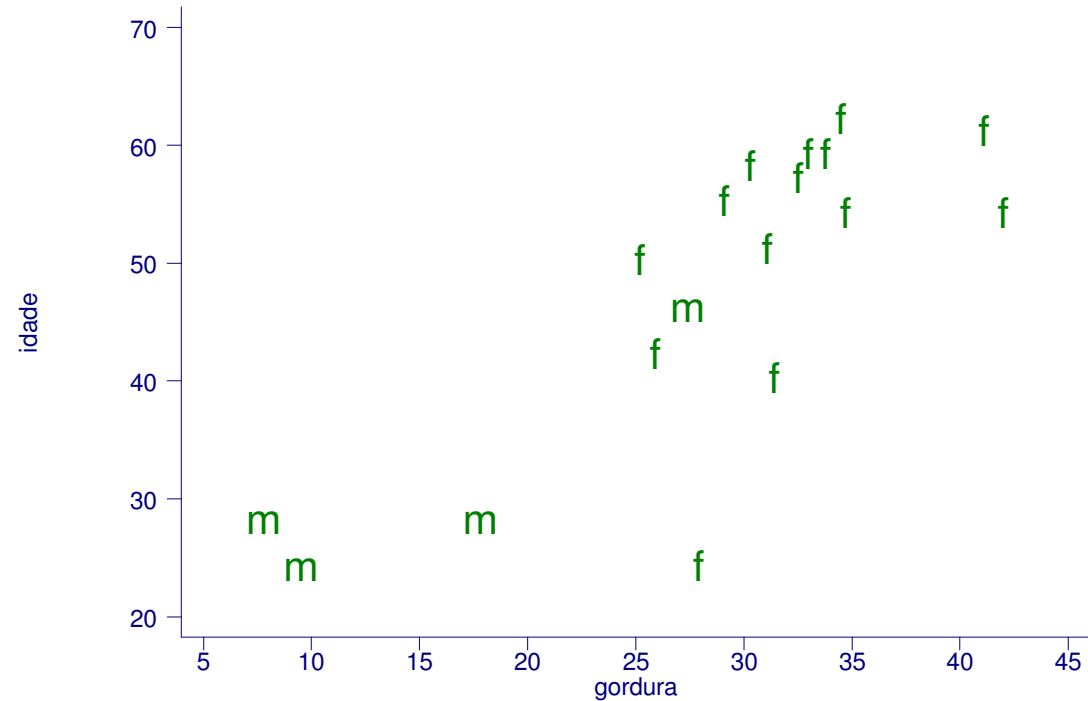
Idade	% gordura	sexo	Idade	% gordura	sexo
23	9,5	M	53	34,7	F
23	27,9	F	53	42,0	F
27	7,8	M	54	29,1	F
27	17,8	M	56	32,5	F
39	31,4	F	57	30,3	F
41	25,9	F	58	33,0	F
45	27,4	M	58	33,8	F
49	25,2	F	60	41,1	F
50	31,1	F	61	34,5	F

M=masculino ; F= feminino  
Fonte: Hand DJ *et al.*, 1994.



## Noções de correlação

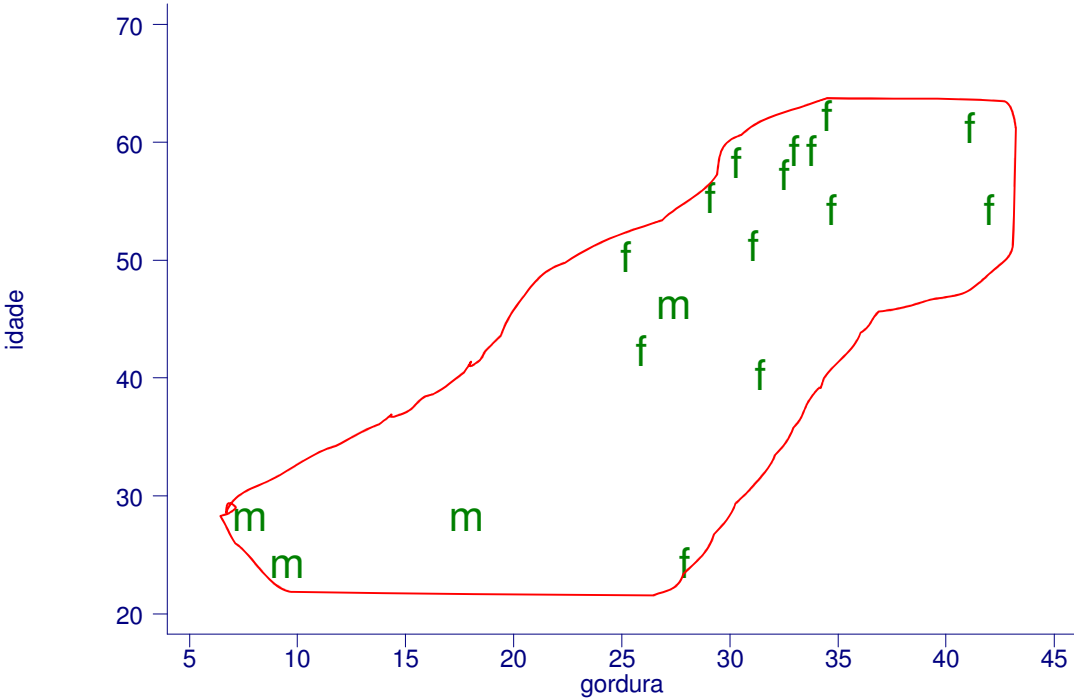
Dispersão entre % de gordura e idade



Fonte: Hand DJ *et al.*, 1994.

# Noções de correlação

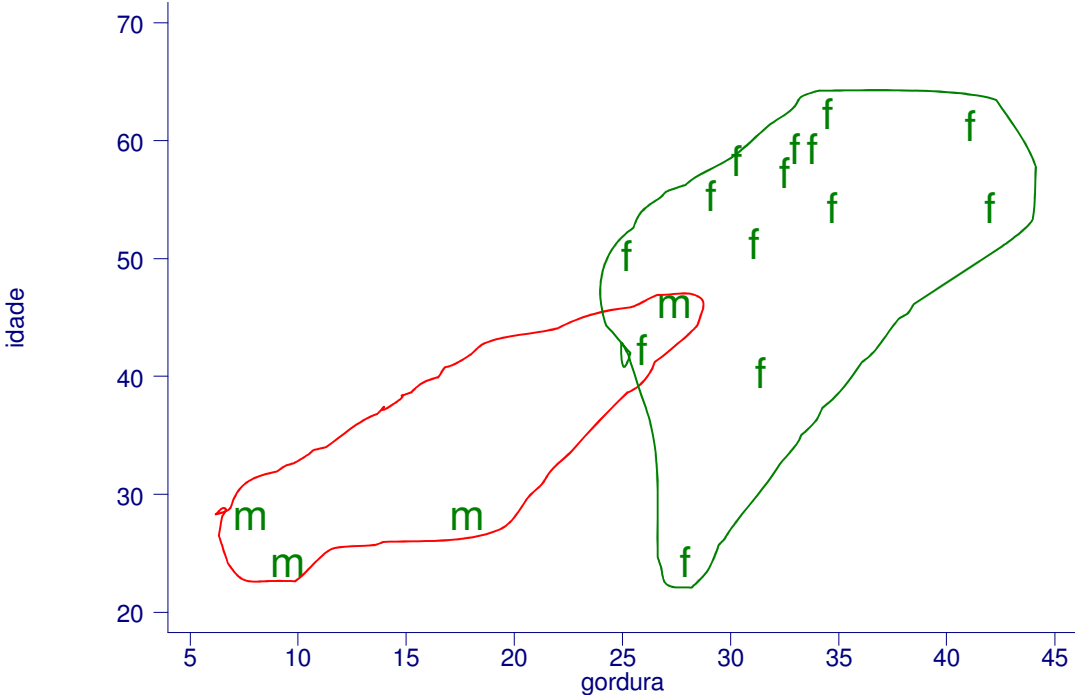
Dispersão entre % de gordura e idade



Fonte: Hand DJ *et al.*, 1994.

# Noções de correlação

## Dispersão entre % de gordura e idade



Fonte: Hand DJ *et al.*, 1994.

## Noções de correlação

### Cálculo do coeficiente de correlação de Pearson

Sexo: masculino

Idade	% gordura	$(y - \bar{y})$	$(x - \bar{x})$	$(x - \bar{x})(y - \bar{y})$	$(y - \bar{y})^2$	$(x - \bar{x})^2$
23	9,5	-7,5	-6,13	45,94	56,25	37,52
27	7,8	-3,5	-7,83	27,39	12,25	61,23
27	17,8	-3,5	2,18	-7,61	12,25	4,73
45	27,4	14,5	11,78	170,74	210,25	138,65
$\bar{y} = 30,5$	$\bar{x} = 15,63$	<b>Total</b>		<b>236,45</b>	<b>291,00</b>	<b>242,13</b>

Coeficiente de correlação (idade,%gordura) masculino:  $r = \frac{236,45}{\sqrt{291 \times 242,13}} = 0,89$

## Noções de correlação

Sexo: feminino

Idade	% gordura	$(y - \bar{y})$	$(x - \bar{x})$	$(x - \bar{x})(y - \bar{y})$	$(y - \bar{y})^2$	$(x - \bar{x})^2$
23	27,9	-27,86	-4,42	123,17	776,02	19,55
39	31,4	-11,86	-0,92	10,93	140,59	0,85
41	25,9	-9,86	-6,42	63,30	97,16	41,23
49	25,2	-1,86	-7,12	13,23	3,45	50,71
50	31,1	-0,86	-1,22	1,05	0,73	1,49
53	34,7	2,14	2,38	5,10	4,59	5,66
53	42	2,14	9,68	20,74	4,59	93,67
54	29,1	3,14	-3,22	-10,12	9,88	10,38
56	32,5	5,14	0,18	0,92	26,45	0,03
57	30,3	6,14	-2,02	-12,42	37,73	4,09
58	33	7,14	0,68	4,85	51,02	0,46
58	33,8	7,14	1,48	10,56	51,02	2,19
60	41,1	9,14	8,78	80,26	83,59	77,06
61	34,5	10,14	2,18	22,10	102,88	4,75
$\bar{y} = 50,86$	$\bar{x} = 32,32$		<b>Total</b>	<b>333,64</b>	<b>1389,71</b>	<b>312,12</b>

Coefficiente de correlação (idade,%gordura) feminino:  $r = \frac{333,64}{\sqrt{1389,71 \times 312,12}} = 0,51$

### Coeficiente de correlação considerando o grupo todo (homens e mulheres)

Idade (X)	% gordura (Y)	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
23	9,5	-23,33	-19,11	445,93	544,44	365,23
27	7,8	-19,33	-20,81	402,35	373,78	433,10
27	17,8	-19,33	-10,81	209,01	373,78	116,88
45	27,4	-1,33	-1,21	1,61	1,78	1,47
23	27,9	-23,33	-0,71	16,59	544,44	0,51
39	31,4	-7,33	2,79	-20,45	53,78	7,78
41	25,9	-5,33	-2,71	14,46	28,44	7,35
49	25,2	2,67	-3,41	-9,10	7,11	11,64
50	31,1	3,67	2,49	9,13	13,44	6,19
53	34,7	6,67	6,09	40,59	44,44	37,07
53	42	6,67	13,39	89,26	44,44	179,26
54	29,1	7,67	0,49	3,75	58,78	0,24
56	32,5	9,67	3,89	37,59	93,44	15,12
57	30,3	10,67	1,69	18,01	113,78	2,85
58	33	11,67	4,39	51,20	136,11	19,26
58	33,8	11,67	5,19	60,54	136,11	26,92
60	41,1	13,67	12,49	170,68	186,78	155,97
61	34,5	14,67	5,89	86,37	215,11	34,68
			Soma	1627,53	2970,00	1421,54

$$\bar{x} = 46,33; \bar{y} = 28,61; s_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n - 1}} = \sqrt{\frac{1421,54}{17}} = 9,14\%; s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{2970,0}{17}} = 13,22anos$$

## Noções de correlação

Abaixo temos o peso (kg) e a altura (cm) de 30 meninas de 11 anos de idade atendidas na escola *Heaton Middlel* em *Bradford*, Inglaterra.

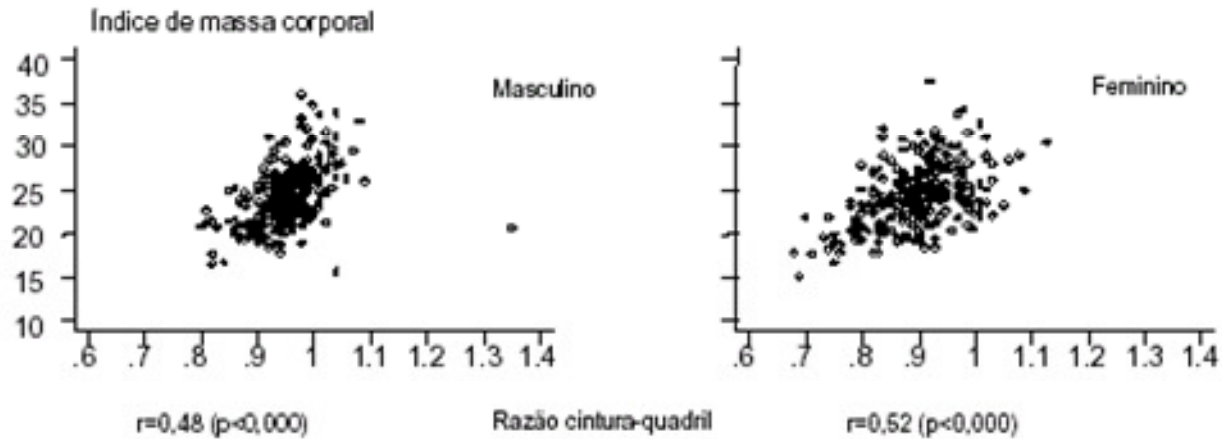
- Fazer o diagrama de dispersão e investigue como é a relação entre as variáveis.
- Calcular o coeficiente de correlação de Pearson.

Altura (cm)	Peso (kg)	Altura (cm)	Peso (kg)	Altura (cm)	Peso (kg)	Altura (cm)	Peso (kg)
135	26	133	31	136	28	146	35
146	33	149	34	154	36	143	42
153	55	141	32	151	48	148	32
154	50	164	47	155	36	143	36
139	32	146	37	149	32	140	33
131	25	149	46	141	29	141	28
149	44	147	36	137	34		
137	31	152	47	135	30		

$$\sum (x - \bar{x})(y - \bar{y}) = 1275; \quad \sum (x - \bar{x})^2 = 1716,80; \quad \sum (y - \bar{y})^2 = 1718,17$$

## Noções de correlação

Os gráficos abaixo foram extraídos do artigo: Excesso de peso e gordura abdominal para a síndrome metabólica em nipo-brasileiros de LERARIO DG *et al.* (*Rev. Saúde Pública*, 2002;36(1):4-11). Interprete as figuras apresentadas no artigo.



**Figura 1** - Correlação entre os valores do índice de massa corporal ( $\text{Kg}/\text{m}^2$ ) e da razão cintura-quadril de nipo-brasileiros segundo sexo.



# Análise simultânea de duas variáveis quantitativas

## REGRESSÃO LINEAR

ADMITINDO-SE Y COMO FUNÇÃO LINEAR DE X, AJUSTA-SE A “MELHOR RETA” AO CONJUNTO DE DADOS.

EQUAÇÃO DE RETA:  $\hat{y} = a + bx$ , onde

$\hat{y}$  = valor ajustado (valor médio predito)

$x$  = valor escolhido de X

$a = \bar{y} - b\bar{x}$ ; a é denominado intercepto; é o valor predito para  $x=0$

$b = r_{xy} \frac{s_y}{s_x}$ ; b é denominado coeficiente angular (*slope*). Indica quantas unidades de Y mudam para a mudança de uma unidade de X.

Aplicando-se aos dados do exemplo:

$$a = 28,61 - b 46,33 ;$$

$$b = 0,79x \frac{9,14}{13,22} = 0,548$$

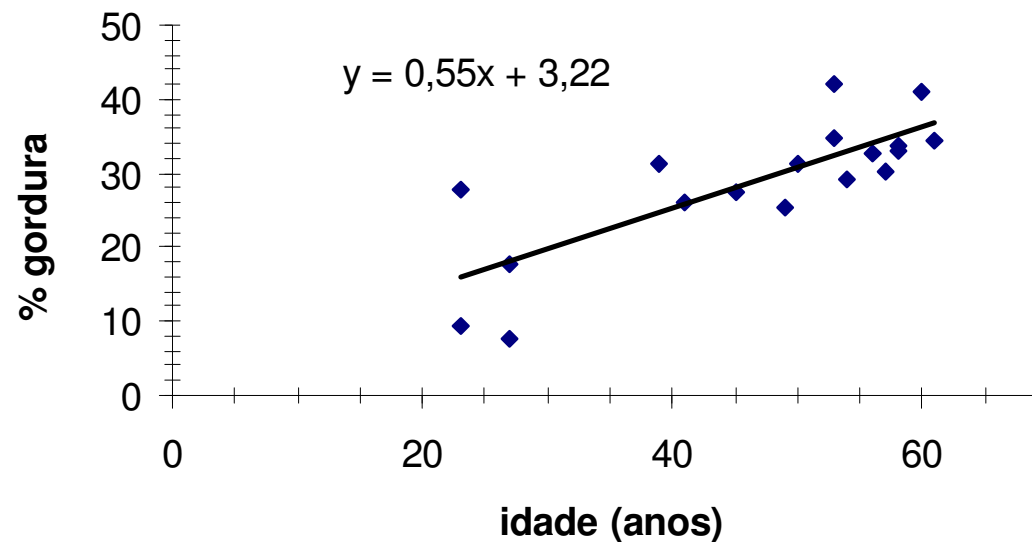
Para aumento de 1 ano, o percentual de gordura aumenta 0,55%.

Substituindo-se o valor b em a, obtém-se  $a=3,221$ .

Equação ajustada  $\% \text{ gordura} = 3,22 + 0,55 (\text{idade})$

Com base nesta equação é possível traçar a reta que passa pelos pontos.

Para  $x = 30$ ;  $y = 19,7$ ; para  $x = 50$ ,  $y = 30,7$



OBS: o coeficiente angular depende das unidades de medida de X e Y. Isto deve ser considerado na decisão da importância do coeficiente angular.

O coeficiente angular da equação de  $Y=f(X)$  é diferente do coeficiente angular de  $X=f(Y)$ , a menos que os desvios padrão de X e Y sejam iguais.

Usos da reta de regressão:

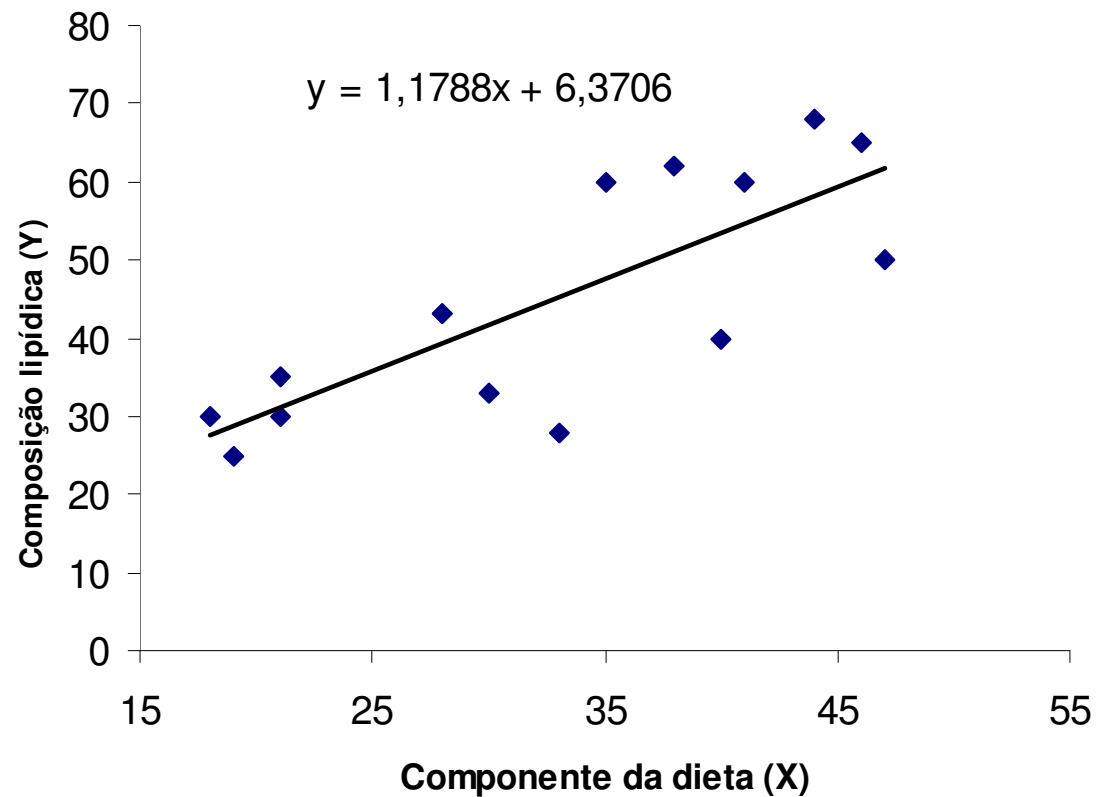
- Predição - utilizar X para prever Y; quando a correlação for forte, melhor é a predição;
- Correlação – mede o grau de relacionamento linear entre X e Y;
- Resumir os dados – cada valor de X tem um valor médio de Y

**Exemplo:**

Em um estudo sobre o efeito dos componentes de uma dieta (X) sobre a composição lipídica (Y) foram obtidos os seguintes dados em uma amostra de 15 animais.

Componente da dieta (X)	Composição lipídica (Y)
18	30
21	35
28	43
35	60
47	50
33	28
40	40
41	60
28	43
21	30
30	33
46	65
44	68
38	62
19	25

- Apresente os dados em um diagrama de dispersão
- Calcule o coeficiente de correlação de Pearson entre X e Y.
- Calcule a reta de regressão da composição lipídica como função do componente da dieta.
- Desenhe a reta de regressão
- Interprete os coeficientes da reta de regressão



Dispersão e reta de regressão entre componentes de uma dieta (X) e composição lipídica (Y). Local X, ano Y