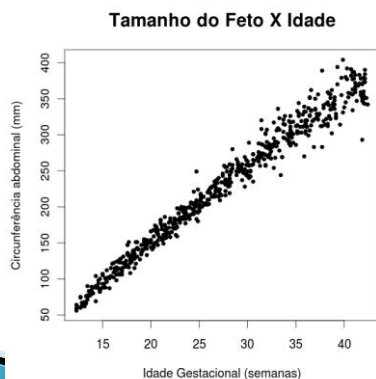


Correlação e Regressão Linear

Aula de hoje

Teste	Parâmetro	Qualitativa ordinal	Quantitativa	Nominal	2	Distribuição normal
Teste t / 2 amostras	Paramétrico	Quantitativa	Nominal	2	Independentes	Distribuição normal, variâncias iguais
Teste t / 2 amostras com variâncias diferentes	Paramétrico	Quantitativa	Nominal	2	Independentes	Distribuição normal, variâncias diferentes
Teste t pareado	Paramétrico	Quantitativa	Nominal	2	Dependentes	Distribuição normal, variâncias iguais
ANOVA	Paramétrico	Quantitativa	Nominal	3 ou mais	Independentes	Distribuição normal, variâncias iguais
ANOVA p / medidas repetidas	Paramétrico	Quantitativa	Nominal	3 ou mais	Dependentes	Distribuição normal, Esfericidade
Mann-Whitney (Wilcoxon não-pareado)	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	2	Independentes	
Wilcoxon (Wilcoxon pareado, teste dos sinais)	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	2	Dependentes	
Kruskal-Wallis	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	3 ou mais	Independentes	
Friedman	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Nominal	3 ou mais	Dependentes	
Teste p / 1 proporção	Paramétrico	Nominal	Nominal	1	-	
Teste p / 2 proporções	Paramétrico	Nominal	Nominal	2	Independentes	
Testes de associação						
Qui-quadrado	Não-Paramétrico	Nominal	Nominal	2 ou mais	Independentes	Células possuem valor esperado > 5
Teste exato de Fisher	Não-Paramétrico	Nominal	Nominal	2	Independentes	
Testes de correlação						
Regressão Linear Simple	Paramétrico	Quantitativa	Quantitativa	-	-	Distribuição normal
Correlação de Pearson	Paramétrico	Quantitativa	Quantitativa	-	-	Distribuição normal
Correlação de Spearman	Não-Paramétrico	Quantitativa / Qualitativa ordinal	Qualitativa ordinal	-	-	
Testes de Variância						
Teste F	Paramétrico	Quantitativa	Nominal	2	Independentes	Distribuição normal

Feto X Idade



Testes estudados até agora

Variável Qualitativa (grupos) X:

Qualitativa Nominal → χ^2

Qualitativa Ordinal → Mann-Whitney, Wilcoxon, etc.

Quantitativa → Teste t, ANOVA, Mann-Whitney, Wilcoxon, etc.

Testes estudados até agora

Variável Qualitativa (grupos) X:

Qualitativa Nominal → χ^2

Qualitativa Ordinal → Mann-Whitney, Wilcoxon, etc.

Quantitativa → Teste t, ANOVA, Mann-Whitney, Wilcoxon, etc.

Variável Quantitativa X Quantitativa → ?

Testes estudados até agora

Variável Qualitativa (grupos) X:

Qualitativa Nominal → χ^2

Qualitativa Ordinal → Mann-Whitney, Wilcoxon, etc.

Quantitativa → Teste t, ANOVA, Mann-Whitney, Wilcoxon, etc.

Variável Quantitativa X Quantitativa →

Correlação e Regressão

Objetivos desta aula

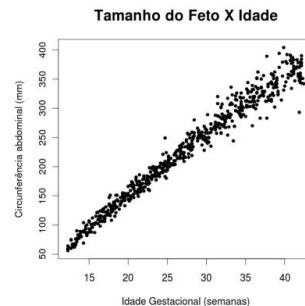
- Analisar as relações entre duas variáveis quantitativas:

Como prever valores de uma variável

Como medir a intensidade de relação

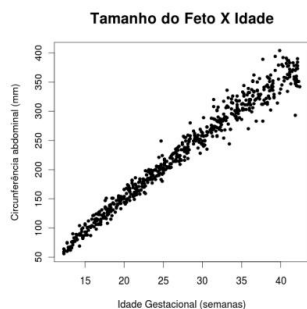
Como provar a existência dessa relação

Feto X Idade



Como utilizar o tamanho do feto para estimar a idade gestacional?

Feto X Idade



É possível que uma função linear de 1º grau (uma reta) explique esses dados?

Como desenhá-la?

Cálculo da reta de Regressão

- Equação da reta: $y = a + bx$

- Coeficiente angular: $b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$
- (inclinação)

- Coeficiente linear: $a = \bar{y} - b \bar{x}$
- (intercepto)

$\bar{x}$ = média de x
 $\bar{y}$ = média de y

- Esse é o Método dos Mínimos Quadrados ([animação](#))

Resultados Minitab

- Fetos X Idade

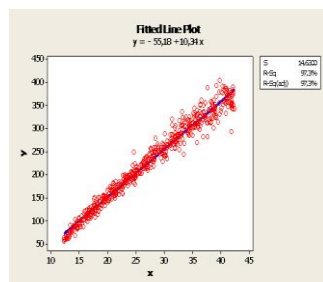
Regression Analysis: y versus x

The regression equation is
 $y = -55,18 + 10,34 x$

S = 14,6300 R-Sq = 97,3% R-Sq(adj) = 97,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	4655764	4655764	21752,14	0,000
Error	600	130135	214		
Total	609	4785899			



Resultados Minitab

- Peso x Coração (gatos)

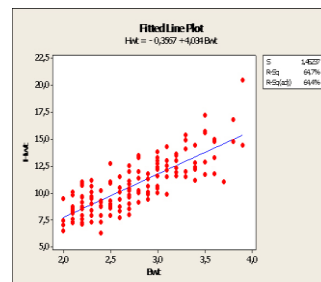
Regression Analysis: Hwt versus Bwt

The regression equation is
 $Hwt = -0,3567 + 4,034 Bwt$

S = 1,45237 R-Sq = 64,7% R-Sq(adj) = 64,4%

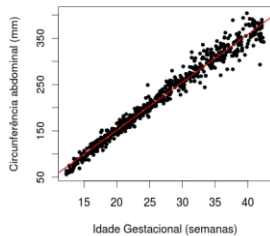
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	540,092	540,092	259,83	0,000
Error	142	299,533	2,109		
Total	143	847,626			

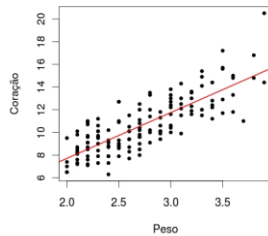


A precisão é igual?

Tamanho do Feto X Idade

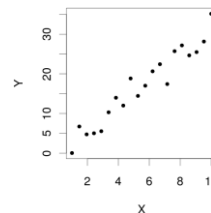


Coração X Peso de gatos



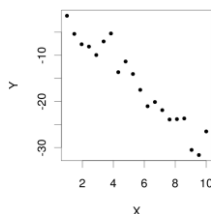
Correlação

- Correlação Positiva
- Se as variáveis X e Y crescem no mesmo sentido, i.e., quando X cresce, Y em média também cresce.



Correlação

- Correlação Negativa
- Se as variáveis X e Y variam em sentidos contrários, i.e., quando X cresce, Y em média decresce.



Coeficientes

Coeficiente de Correlação (r):

Medida da força da dependência linear entre duas variáveis

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad -1 \leq r \leq +1$$

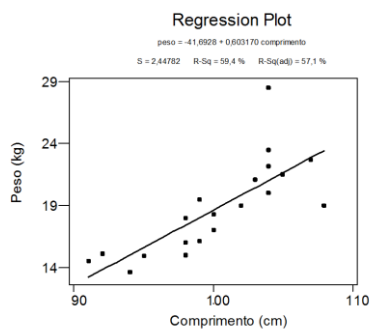
$$0 \leq r^2 \leq 1$$

Coeficiente de determinação (r²):

Proporção da variabilidade total em y que é explicada pelo grau de relação linear com os valores de x.

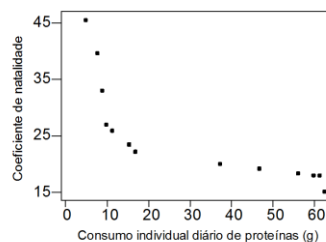
Correlação

Comprimento (cm) e peso (kg) de cães (Araújo e Hossne, 1977)



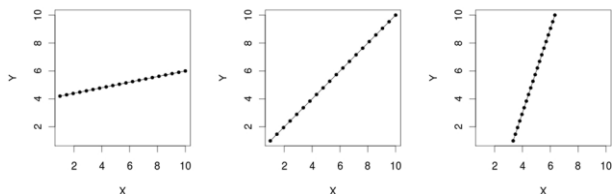
$r = 0,77$ $r^2 = 0,59$

Correlação



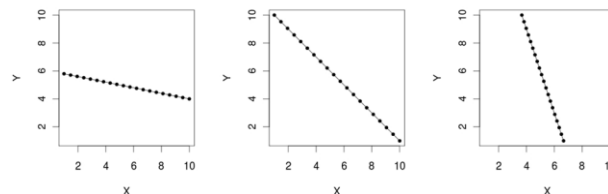
Exemplos

Se $r = 1$; $r^2 = 100\%$: Correlação positiva perfeita



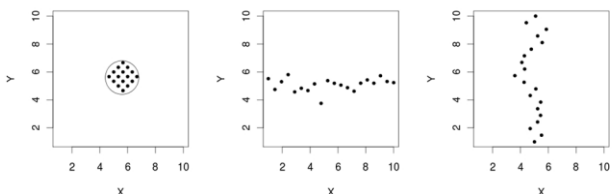
Exemplos

Se $r = -1$; $r^2 = 100\%$: Correlação negativa perfeita



Exemplos

Se $r = 0$; $r^2 = 0\%$: Não há correlação



Teste de hipóteses

- É possível realizar testes de hipótese sobre o coeficiente de correlação. Nesse caso:

$$H_0: r = 0$$

$$H_1: r \neq 0$$

- O que equivale a testar o coeficiente da reta.

$$H_0: b = 0$$

$$H_1: b \neq 0$$

- Em ambos os casos, a interpretação deste teste é:

$$H_0: \text{Não há relação linear}$$

$$H_1: \text{Há relação linear}$$

Resultados Minitab

Fetos X Idade

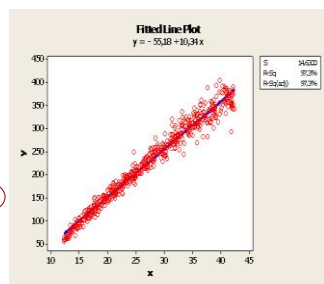
Regression Analysis: y versus x

The regression equation is
 $y = -55,18 + 10,34 x$

S = 14,6300 R-Sq = 97,3% R-Sq(adj) = 97,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	4655764	4655764	21752,14	0,000
Error	608	130135	214		
Total	609	4785899			



Resultados Minitab

Peso X Coração(gatos)

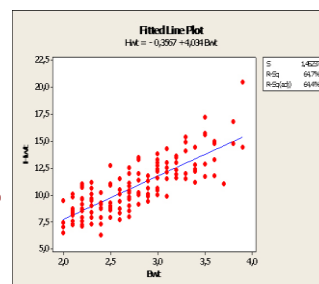
Regression Analysis: Hwt versus Bwt

The regression equation is
 $Hwt = -0,3567 + 4,034 Bwt$

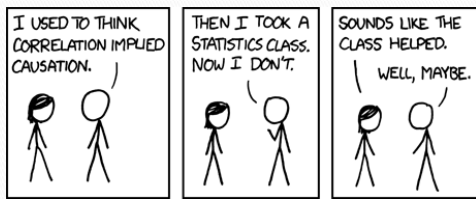
S = 1,45237 R-Sq = 64,7% R-Sq(adj) = 64,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	540,092	540,092	259,63	0,000
Error	142	299,533	2,109		
Total	143	847,626			



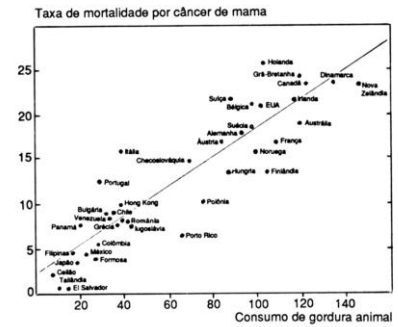
Correlação X Causalidade



"Correlation doesn't imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'."

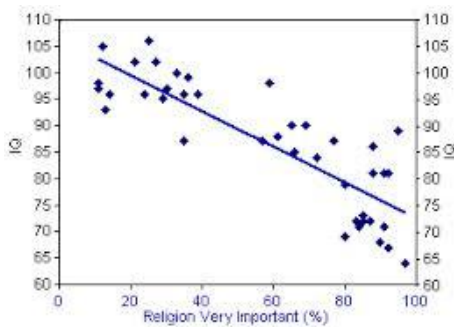
Tradução: "Correlação não implica em causalidade, mas mexe as sombrancelhas de maneira sugestiva e gesticula furtivamente enquanto sussura 'olha aquilo'."

Câncer X Consumo de carne

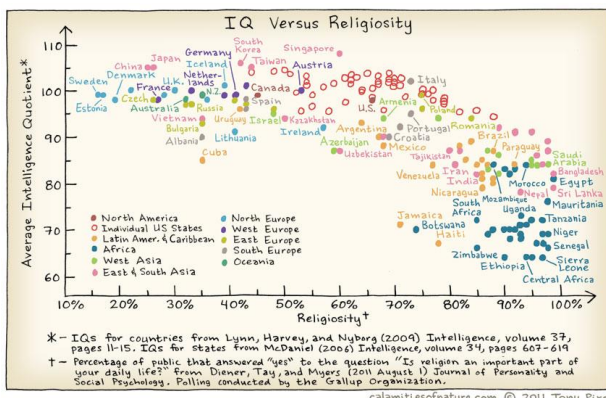
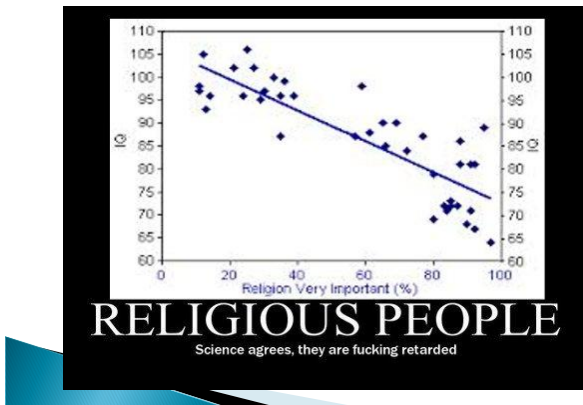


Fonte: CAROLL (1975)
Dados de 1964-1966

IQ x % de religiosos (por país)



Típica interpretação errônea



Curiosidades

Sobrevivência X dose de radiação

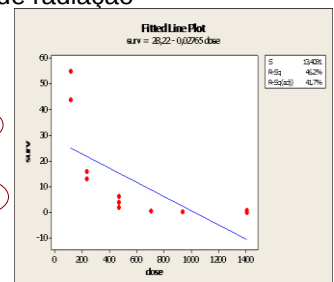
Regression Analysis: surv versus dose

The regression equation is
surv = 28,22 - 0,02765 dose

S = 13,4091 R-Sq = 46,2% R-Sq(adj) = 41,7%

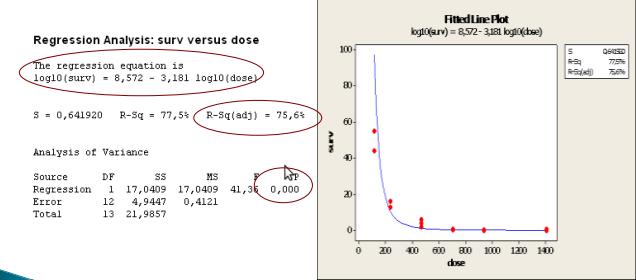
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1851,30	1851,30	10,30	0,008
Error	12	2157,64	179,80		
Total	13	4008,93			



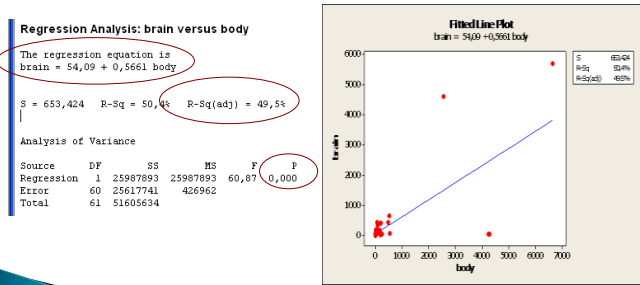
Curiosidades

Sobrevivencia X dose de radiação (log)



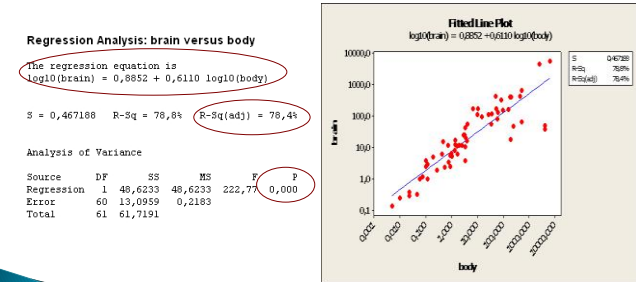
Curiosidades

Peso X Cérebro de mamíferos



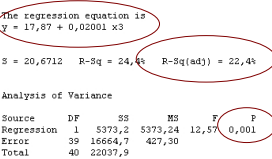
Curiosidades

Peso X Cérebro de mamíferos (log)



Curiosidades

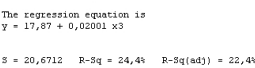
Regression Analysis: y versus x3



Poluição X População

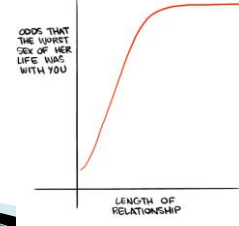
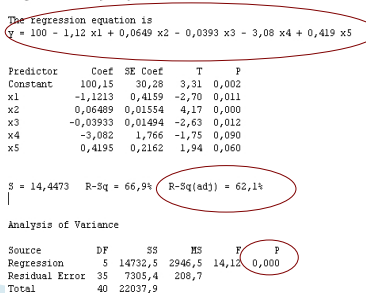
Curiosidades

Regression Analysis: y versus x3



- Poluição X:
- População
- Temperatura média
- Número de fábricas
- Velocidade do vento
- Chuva

Regression Analysis: y versus x1: x2: x3: x4: x5



Mais dados

- ▶ Dados sobre consumo de gordura animal e câncer de mama são de 1964–1966. O artigo é esse:
http://cancerres.aacrjournals.org/content/35/11_Part_2/3374.full-text.pdf
- ▶ Religião Vs. QI:
Os dados da aula são daqui:
http://hypnosis.home.netcom.com/iq_vs_religiosity.htm
- ▶ Gráfico mais bonito (e com mais pontos) aqui:
<http://www.calamitiesofnature.com/archive/?c=619>
- ▶ Ambos dão a fonte de dados que eles utilizaram.

