Elementos de Estatística

Lupércio F. Bessegato & Marcel T. Vieira

UFJF – Departamento de Estatística
2013

Análise Bivariada

Variável Qualitativa e Quantitativa

Variável Qualitativa *vs.* Quantitativa

Objetivo:

- representar graficamente as duas variáveis combinadas;
- definir e calcular uma medida de associação entre as variáveis.

Box-plot

- Pode ser utilizado para comparações entre diferentes grupos de dados
  - √ Variável quantitativa *vs.* variável categórica

Exemplo – Questionário

- Dados sobre:
  - √ Sexo
  - √ Idade
  - √ Altura
  - √ Peso
  - √ Qte filhos na família
  - √ etc.
  - √ Planilha: *BD_elementos_131.xls/questionario*

Análise de Peso e Altura

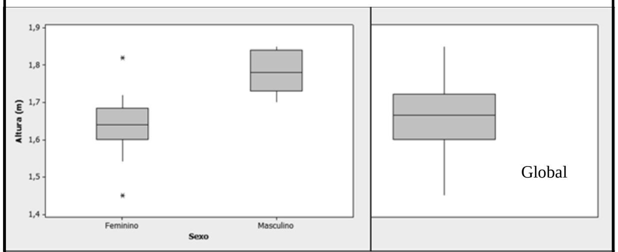
- Objetivo:

  Analisar o comportamento do peso e da altura com relação ao sexo

  √ Análise gráfica

  √ Análise de medidas resumo

- Box-plot



√ O padrão de altura é diferente para cada sexo?

√ Sexo ajuda a explicar a variabilidade de altura?

√ Simetria? Locação? Variabilidade? *Outliers*?

- Medidas resumo por Sexo:

Descriptive Statistics: Altura

Variable	Sexo	N	Mean	StDev	CoefVar	Median	IQR
Alt	F	37	1,6338	0,0665	4,07	1,6400	0,0850
	M	13	1,7808	0,0541	3,04	1,7800	0,1100

- Medidas resumo – Global:

Descriptive Statistics: Altura

Variable	N	Mean	StDev	CoefVar	Median	IQR
Alt	50	1,6720	0,0906	5,42	1,6650	0,1225

√ Simetria?

√ Locação?

√ Variabilidade?

Comentário

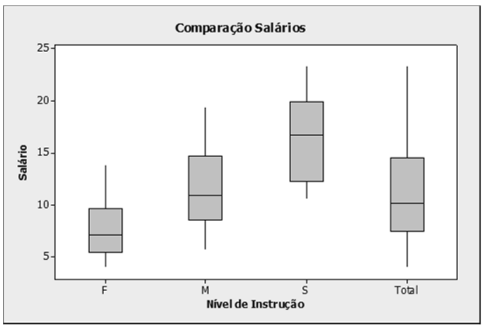
- É possível perceber, a partir dessas medidas e gráficos, uma dependência entre altura e sexo?

Exemplo – Dados de Empregados

- Dados sobre:
  - √ grau de instrução
    - F(Ensino Fundamental), M(Ensino Médio) e S(Ensino Superior)
  - √ salário
    - fração do Salário Mínimo
  - √ Região de procedência

    Planilha: *BD_elementos_131.xls/ciaMB*
  - √ Fonte: Bussab e Morettin, Cap. 2

- Box-plots para comparação global de salários



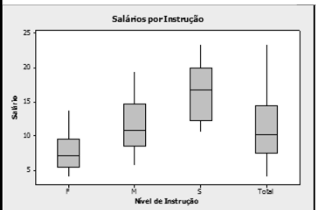
- Medidas resumo comparando-se com o Total

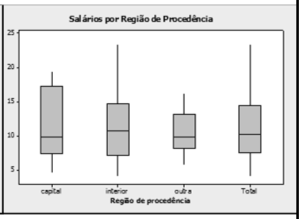
Descriptive Statistics: sal_1

Variable	instrucao_1	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
sal_1	F	7,837	2,956	4,000	5,503	7,125	9,588	13,850
	M	11,528	3,715	5,730	8,585	10,910	14,695	19,400
	S	16,48	4,50	10,53	12,23	16,74	19,89	23,30
	Total	11,122	4,587	4,000	7,478	10,165	14,480	23,300

Comentário

- Há indícios de uma associação entre salário e nível de instrução?
  - √ Há indícios amostrais de aumento do salário com o nível de escolaridade do empregado





Nível de instrução parece melhorar a capacidade de previsão de salário

Região de procedência não parece melhorar esta capacidade.

- Salários vs. Escolaridade:
  - √ Cálculo das variâncias:

```
MTB > Describe  'sal_1';
SUBC>   By  'instrucao_1';
SUBC>   Variance.

Descriptive Statistics: sal_1

Variable  instrucao_1  Variance
sal_1     F               8,741
          M              13,802
          S               20,27
          Total          21,045
```

As variâncias DENTRO de cada nível <u>são menores</u> que a variância global

- Salários vs. Região de Procedência:
  - √ Cálculo das variâncias:

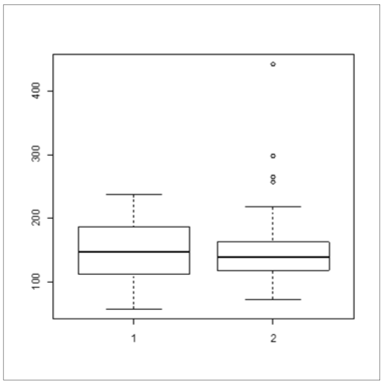
```
MTB > Describe  'sal_1';
SUBC>   By  'rp_1';
SUBC>   Variance.

Descriptive Statistics: sal_1

Variable  rp_1       Variance
sal_1     capital       29,99
          interior      28,05
          outra         9,894
          Total        21,045
```

As variâncias DENTRO de cada nível <u>não são menores</u> que a variância global

Exemplo – Doenças Cardiovasculares

- Universo:
  - √ Homens doentes com idade entre 45 e 67 anos
- Amostra:
  - √ 100 casos coletados em 1969
- Variáveis
  - √ nível de glicose no sangue, em mg percentuais
  - √ atividade física em casa
    - 1 = sedentário; 2 = moderada; 3 = alta

Box-plot



Associação entre Variáveis Quantitativas

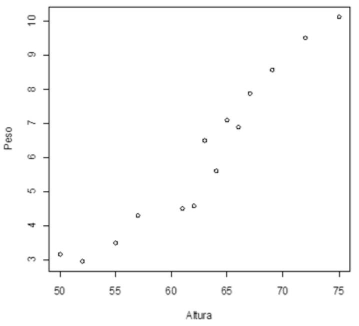
Diagrama de Dispersão

- Gráfico de pares ordenados por elementos da amostra (indivíduos)
- É a maneira mais simples de se estudar a relação entre duas variáveis quantitativas
- Objetivo:
  - ✓Ocorrência de tendências (lineares ou não)
  - ✓Agrupamentos de uma ou mais variáveis
  - ✓Mudanças de variabilidade de uma variável em relação à outra
  - ✓Ocorrência de valores atípicos (‘outliers’)

Exemplo

- Altura (cm) e peso (kg) de crianças até 1 ano

Altura	Peso
52	2,95
50	3,15
62	4,58
63	6,50
55	3,50
72	9,50
75	10,13
69	8,57
65	7,10
64	5,60
66	6,90
61	4,50
57	4,30
67	7,89

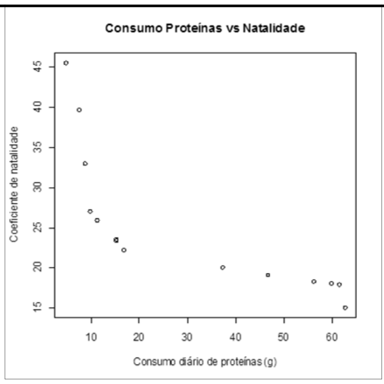


- Qual a relação entre o peso e a estatura das pessoas?
- Percebem-se ‘clusters’ no conjunto de dados?
- Há diferenças na variabilidade de uma variável, considerados os valores da outra?
- Há valores atípicos?

Relação entre consumo de proteínas e natalidade

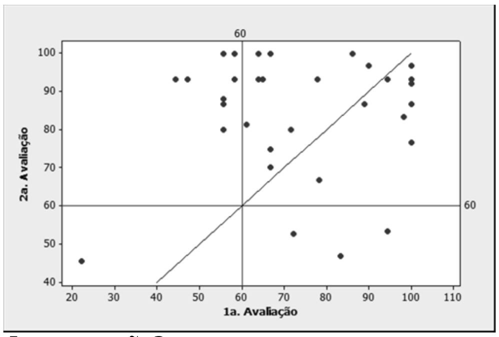
País	Consumo de Proteínas	Coeficiente de Natalidade
Formosa	4,7	45,6
Malásia	7,5	39,7
India	8,7	33,0
Japão	9,7	27,0
Iuguslávia	11,2	25,9
Grécia	15,2	23,5
Itália	15,2	23,4
Bulgária	16,8	22,2
Alemanha	37,3	20,0
Irlanda	46,7	19,1
Dinamarca	56,1	18,3
Austrália	59,9	18,0
Estados Unidos	61,4	17,9
Suécia	62,6	15,0

- Qual relação entre as variáveis?

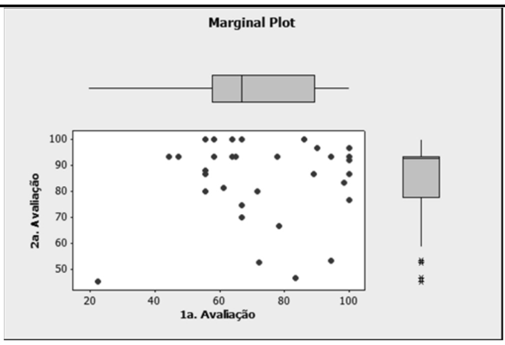


- Pode-se afirmar que há relação causal entre consumo de proteínas e natalidade ?
- Há indícios de clusters?

Exemplo



- Interpretação?



- Leituras gráficas

Correlação Linear

Correlação

- Correlação Positiva:
  - √ Se ambas as variáveis crescem no mesmo sentido
- Correlação Negativa:
  - √ Se as variáveis crescem em sentidos opostos
- Correlação significativa indica apenas associação entre as variáveis
  - √ NÃO INDICA RELAÇÃO DE CAUSALIDADE

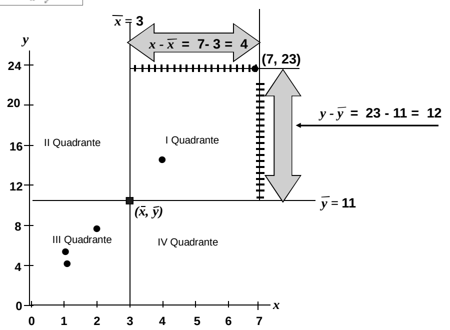
Coeficiente de Correlação

- Como quantificar a correlação entre as variáveis?
  - √ Grau de associação

Justificação para a Fórmula de *r*

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

$(\bar{x}, \bar{y})$ centróide da nuvem de dados



Coeficiente de Correlação de Pearson

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \, \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

- O numerador mede o total da concentração de pontos pelos quatro quadrantes
- Dá origem uma medida bastante usada

Notação

x_i y_i $\bar{x}$ $\bar{y}$

- : i-ésimo valor observado da variável *x*
- : i-ésimo valor observado da variável *y*
- : média dos valores observados da variável *x* (média amostral)
- : média dos valores observados da variável *y* (média amostral)

Soma de Quadrados – Notação

$$S_{xx} = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - n(\bar{x})^2$$

$$S_{yy} = \sum_i (y_i - \bar{y})^2 = \sum_i y_i^2 - n(\bar{y})^2$$

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - n(\bar{x}.\bar{y})$$

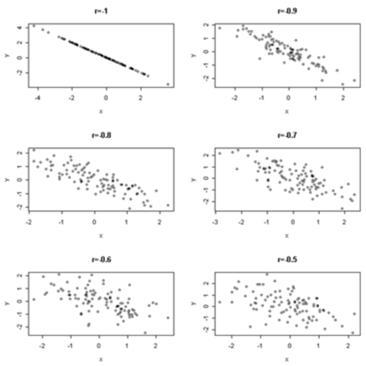
Propriedades de *r*

- Mede a intensidade de relacionamento <u>linear</u>
- *r* é adimensional e $-1 \leq r \leq 1$
  - √ r = 1 ou -1 ➔ correlação linear perfeita
  - √ r = 0 ➔ correlação linear nula
- O valor de *r* não é afetado pela escolha de x ou y.

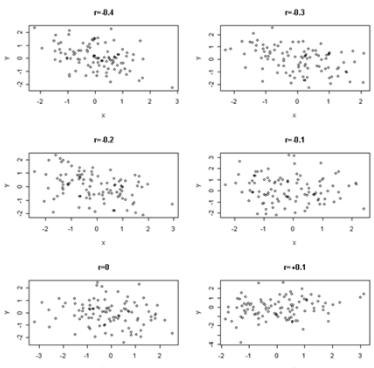
Propriedades de *r*

- A conversão da escala de qualquer das variáveis não altera o valor de *r*
- O valor de *r* não é alterado com a permutação de valores de x e y.
- Uma correlação baseada em médias de muitos elementos, em geral, é mais alta do que a correlação entre as mesmas variáveis baseada em dados para os elementos

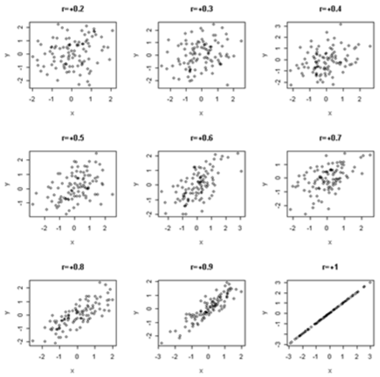
Diagramas de Dispersão (1)

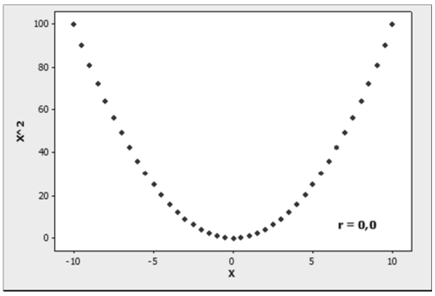


Diagramas de Dispersão (2)



Diagramas de Dispersão (3)





Existe uma relação de dependência <u>NÃO –LINEAR</u> entre as variáveis.

Exemplo 4 – Hábito de Fumar

- Dados sobre hábito de fumar entre homens e mortalidade por câncer de pulmão, na Inglaterra:
  - √ Dados distribuídos em 25 tipos de ocupação;
  - √ Variáveis:
    - Grupo: grupo de ocupação
    - Ifumo: índice de fumo
    - Imorte: índice de mortalidade

Planilha: *fumo*

Fonte: The Data and Story Library http://lib.stat.cmu.edu/DASL/

Exemplo 4 – Hábito de Fumar

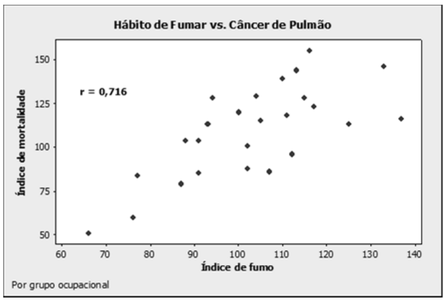
- ifumo: razão do número médio diário de cigarros fumados sobre a média global de cigarros.
  - √ Base: 100
  - √ ifumo = 100: número médio de cigarros por dia para o grupo é igual ao número médio global de cigarros fumados por dia
  - √ ifumo > 100: grupo fuma mais que o global
  - √ ifumo < 100: grupo fuma menos que o global

Exemplo 4 – Hábito de Fumar

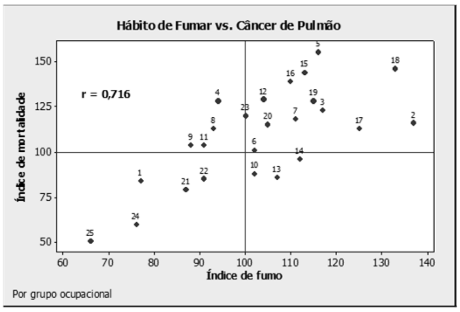
- imorte: razão da taxa de mortes sobre a taxa global de mortes (por câncer de pulmão).
  - √ Base: 100
  - √ imorte = 100: número médio de mortes por câncer de pulmão para o grupo é igual ao número médio global de mortes por câncer de pulmão
  - √ imorte > 100: grupo com incidência de mortes por câncer de pulmão maior que o geral
  - √ imorte < 100: grupo com incidência de mortes por câncer de pulmão menor que o geral

- Base de dados:

	Grupo	Ifumo	Imorte
1	Agricultores/Pescadores	77	84
2	Mineradores	137	116
3	Carvao.gas.quimicos	117	123
4	vidor.ceramica	94	128
5	fornalha/ferraria/fundicao	116	155
6	eletrica/eletronica	102	101
7	Engineeringandalliedtrades	111	118
8	madeira	93	113
9	couro	88	104
10	textil	102	88
11	roupas	91	104
12	comida.bebida.tabaco	104	129
13	papel.impressao	107	86
14	outros.produtos	112	96
15	construcao	113	144
16	pintores.decoradores	110	139
17	motoristas.de.máquinas.paradas	125	113
18	outros.trabalhadores	133	146
19	transporte.comunicacao	115	128
20	lojista.gerente.de.armazem.empacotadores	105	115
21	trabalhadores.de.escritorio	87	79
22	vendas	91	85
23	servicos.esportes.recreacao	100	120
24	administradores	76	60
25	profissionais.liberais.artistas	66	51



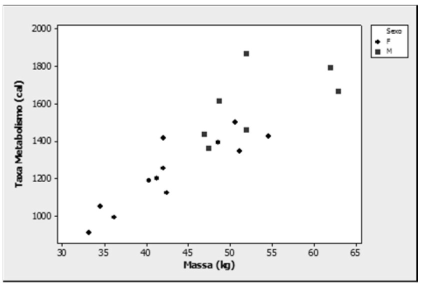
- Percebe-se uma correlação positiva entre as duas variáveis.



No contexto do exemplo faz sentido prever o índice de mortalidade por câncer de pulmão num particular grupo, dado o índice de fumo do grupo.

Atividade

- Relação entre taxa de metabolismo e massa



- Evidências empíricas:
  - √ Associação linear e positiva
  - √ Associação mais forte entre a mulheres

- Cálculo correlações por grupos

```
MTB > corr c3 c4

Correlations: Massa; Taxa

Pearson correlation of Massa and Taxa = 0,865
P-Value = 0,000

MTB > corr c3 c4;
SUBC> by c2.

Correlations: Massa; Taxa

Results for Sexo = F

Pearson correlation of Massa and Taxa = 0,876
P-Value = 0,000

Results for Sexo = M

Pearson correlation of Massa and Taxa = 0,592
P-Value = 0,161
```

Não há evidências de correlação significativa entre os homens

- Valores médios dos grupos

```
MTB > describe c3 c4;
SUBC> by c2;
SUBC> stdev;
SUBC> mean.

Descriptive Statistics: Massa; Taxa

Variable  Sexo    Mean  StDev
Massa     F      43,03   6,87
          M      53,10   6,69

Taxa      F     1235,1  188,3
          M     1600,0  189,2
```

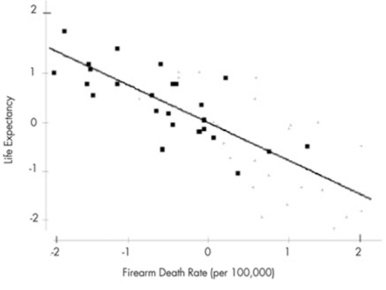
- Evidências empíricas:
  - ✓ Variabilidade semelhante entre os grupos;
  - ✓ Poucos homens com peso menor, poucas mulheres com peso maior
  - ✓ Possíveis influências na correlação:
    - Peso;
    - Sexo;
    - Variável não apresentada

Exemplo

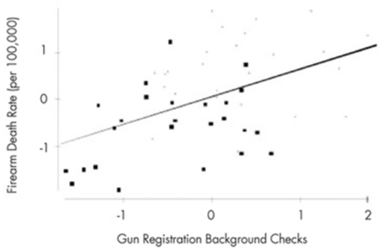
- Há conexão causal entre ignorância e crime?
  - √ Objetivo: evidência empírica para apoiar (rejeitar) essa afirmação
  - √ Análise com dados de estados dos EUA
  - √ Fonte: Chance Magazine/ASA
    http://chance.amstat.org/2013/04/plotting-evidence/

- Variáveis em estudo
  1. Número de armas
  2. Taxa de mortes por arma de fogo
     (mortes por 100.000 habitantes)
  3. Expectativa de vida (2010 – 2011)
  4. Habilidade em leitura (ignorância):
     1. Scores (8 níveis) para uma amostra da população, em todas as faixas etárias e em todos os assuntos (NAEP – *National Assessment of Educational Progress*)
  5. Renda per capita
- Dados estão padronizados

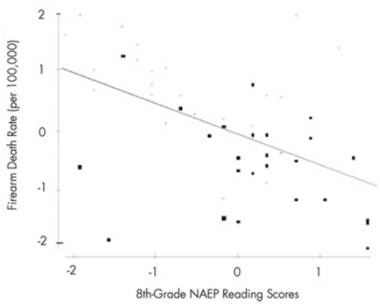
- Afirmação 1:
  - √ Quanto mais mortes por armas de fogo, menor será a expectativa de vida (em um estado)



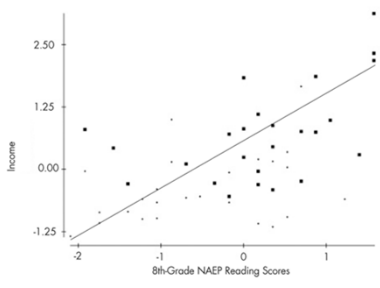
- Afirmação 2:
  - √ Quanto mais armas registradas, maior será o número de mortes por arma de fogo (por estado)



- Afirmação 3:
  - √ Quanto menor a habilidade em leitura, maior será o número de mortes por arma. (por estado)



- Afirmação 4:
  - √ Quanto maior a habilidade em leitura, maior será a renda. (por estado)



- O acesso a armas de fogo torna-se cada vez menos controlado, a vida se torna mais breve, a riqueza diminui e a ignorância aumenta
- É o que as evidências mostram?

Correlação – Erros Comuns

- Causalidade:

  Uma correlação forte (r vizinho de $+1$ ou -1) não implica uma relação de causa e efeito.

  O fato de duas grandezas tenderem a variar no mesmo sentido não implica a presença de relacionamento causal entre elas.

Correlação e Causalidade

Perguntas pertinentes, no caso de correlação significante entre as variáveis:

- √ Há uma relação de causa e efeito entre as variáveis? (x causa y? ou vice-versa)
- √ Ex.: Relação entre armas e mortes
- √ É razoável concluir que mais armas resulta em mais mortes?

- É possível que a relação entre duas variáveis seja uma coincidência?
  - √ Obter uma correlação significante entre o número de espécies animais vivendo em determinada área e o número de pessoas com mais de 2 carros, não garante causalidade
  - √ É bastante improvável que as variáveis estejam diretamente relacionadas.

- É possível que a relação das variáveis tenha sido causada por uma terceira variável (ou uma combinação de muitas outras variáveis)?
  - √ Tempo dos vencedores das provas masculina e feminina dos 100 m rasos
  - √ Os dados tem correlação linear positiva

    Pode-se dizer que a diminuição no tempo masculino cause uma diminuição no tempo feminino?
  - √ A relação deve depender de outras variáveis: técnica de treinamento, clima, etc.

Correlação e Causalidade

- A flutuação de uma 3ª variável faz com que *X* e *Y* variem no mesmo sentido;

  Esta 3ª variável é chamada <u>variável intercorrente</u> (não-conhecida);

  A falsa correlação originada pela 3ª variável é denominada <u>correlação espúria</u>;

Noções de Regressão

- Objetivo:
  - √ Estudar a relação entre duas variáveis quantitativas
- Exemplos:
  - √ Idade e altura das crianças
  - √ Tempo de prática de esportes e ritmo cardíaco
  - √ Tempo de estudo e nota na prova
  - √ Taxa de desemprego e taxa de criminalidade
  - √ Expectativa de vida e taxa de analfabetismo

Regressão e Correlação

- Regressão:
  - √ Explicitando a forma dessa relação:
- Correlação:
  - √ Quantificando a força dessa relação
- Diagrama de dispersão:
  - √ Representação gráfica de duas variáveis quantitativas

Regressão e Correlação

- Regressão:
  - √ Usa variável(eis) explicativa(s) para explicar ou predizer comportamento de variável resposta (quando houver sentido).
- Correlação:
  - √ Trata simetricamente duas variáveis

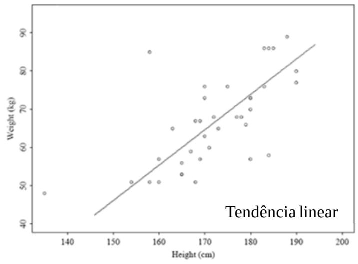
Regressão

- Variável resposta (Y):
  - √ Variável resposta cujo comportamento se quer explicar
- Variável(eis) explicativa(s) (X_i):
  - √ São de interesse caso ajudem a entender, explicar ou predizer o comportamento de Y.
  - √ O enfoque da regressão é natural quando Y é aleatória e X_i é controlada ou não-aleatória.

x	Y
• Variável explicativa • Variável independente • Regressor • Preditor • Variável exógena • Variável de controle ou estímulos	• Variável explicada • Variável dependente • Regredido • Predito • Variável endógena • Variável resposta

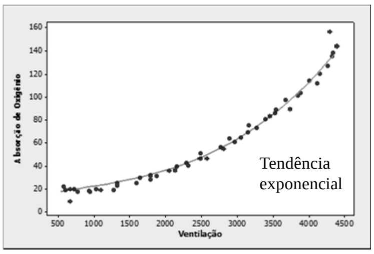
Exemplo 1 – Peso/Altura de Estudantes

- Variável resposta: Peso (kg)
- Variável explicativa: Altura (cm)



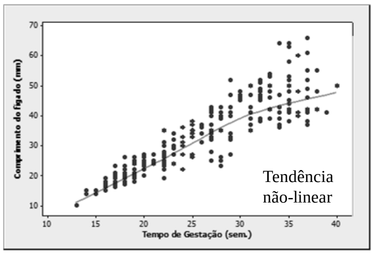
Exemplo 2 – Absorção de Oxigênio

- Variável resposta: Absorção de Oxigênio
- Variável explicativa: Ventilação

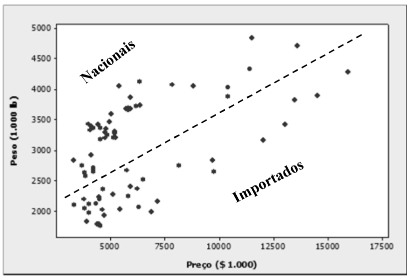


Exemplo 3 – Comprimentos de Fígados

- Variável resposta: Comprimento do fígado (mm)
- Variável explicativa: Tempo de gestação (sem.)

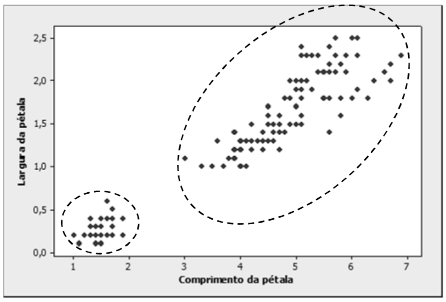


Outros Padrões (2)



Importante descobrir o que define os grupos

Outros Padrões (3)



Variedades diferentes de Flores

Modelo de Regressão

- Relação de regressão:
  - *Tendência + dispersão residual*
- Tendência:
  - √ Suavização dos dados
  - √ Explica a maior parte das diferenças de Y
- Valores atípicos:
  - √ Observações muito diferente do restante dos dados

Relações Fortes e Fracas

- Relação Forte:
- A dispersão é pequena em relação à amplitude dos valores da curva de tendência
- Em dados observacionais, relações fortes <u>não</u> são necessariamente causais

Ajuste de Funções

- Tendência linear: $Y = \beta_0 + \beta_1 X$
  - √ Para cada mudança de uma unidade em *X*, *Y* muda uma quantidade fixa.
- Tendência quadrática: $Y = \beta_0 + \beta_1 X + \beta_2 X^2$
  - √ Tendência levemente curva

Tipos

- Simples:
  - √ Uma variável independente (explicativa)
- Múltipla:
  - √ Duas ou mais variáveis independentes

Objetivos

- Encontrar equação matemática que permita:
  - √ Descrever e compreender a relação entre 2 ou mais variáveis aleatórias
  - √ Projetar ou estimar uma nova observação
- Ajustar uma reta a partir dos dados amostrais

Utilidades

- √ Busca de relações de Causa e Efeito;
- √ Predição de valores;
- √ Estabelecer explicação sobre população a partir de uma amostra

Regressão Linear Simples

- Busca-se a equação de uma reta que permita:
  - √ Descrever e compreender a relação entre duas variáveis
  - √ Projetar e estimar uma das variáveis em função da outra.

Regressão Linear Simples (2)

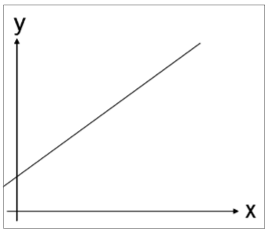
- A partir de valores observados de X e Y, modelar a tendência através de uma equação do tipo:

$$Y_i = \beta_0 + \beta_1 . X_i$$

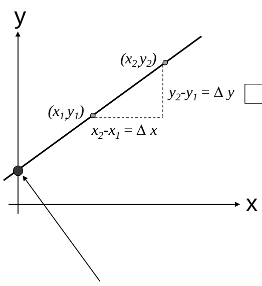
Função Linear

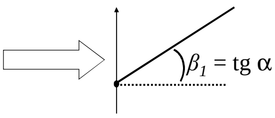
$$Y_i = \beta_0 + \beta_1 . X_i$$

- f(x) se modifica a uma taxa constante em relação à sua variável independente
- β_0 e β_1 são constantes
- β_0 : intercepto-*y*
- β_1 : coeficiente angular



Intercepto e Coeficiente Angular





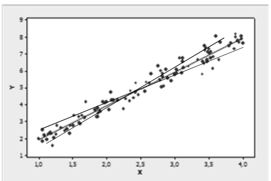
$$\beta_1 = \frac{\text{variação de y}}{\text{variação de x}}$$

β_0: intersecção da reta com o eixo y

Interpretação dos Parâmetros

- β_1 : declividade da reta
  - √ define o aumento ou diminuição da variável Y por unidade de variação de X
- β_0 = intercepto em y
  - √ define o valor médio de Y sem a interferência de X (com $X=0$).

Ajuste da Reta



- Qual a reta que se ajusta melhor aos dados?
- ou seja quais os valores de β_0 e β_1?
- Escolher β_0 e β_0 de maneira a tornar mínima a distância entre a reta e os pontos

Método dos Mínimos Quadrados

- Critério:
- Valores dos parâmetros que minimizam a soma dos quadrados dos desvios

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Método dos Mínimos Quadrados (3)

- Resultados das derivadas parciais:

$$\hat{\beta}_1 = \frac{n.\sum(x_i y_i) - (\sum x_i)\cdot(\sum y_i)}{n.\sum x_i^2 - (\sum x_i)^2} \qquad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \frac{\sum y_i - b\sum x_i}{n} \qquad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

- Calculando por medidas estatísticas :

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} = r_{XY}\frac{s_Y}{s_X}$$

Exemplo

- Resultado de Avaliação de Conhecimento em função do tempo dispendido para estudo
- Variável resposta (Y): *Resultado em TVC*
- Variável explicativa: *Tempo de estudo* (em horas)

Tempo (horas)	Resultado TVC
3	4,5
7	6,5
2	3,7
15	4,0
12	9,3

Resultado TVC (ptos)	Tempo (horas)			
Y	X	X^2	XY	Y^2
4,5	3	9,0	13,5	20,3
6,5	7	49,0	45,5	42,3
3,7	2	4,0	7,4	13,7
4,0	15	225,0	60,0	16,0
9,3	12	144,0	111,6	86,5
28,0	39	431,0	238,0	178,7

$$\bar{y} = 5,60 \quad \bar{x} = 7,80$$

$$S_{xx} = 431 - 5(7,80)^2 = 126,80$$

$$S_{xy} = 238 - 5(7,80)(5,60) = 19,60$$

$$\hat{\beta}_1 = \frac{19,60}{126,80} = 0,15$$

$$\hat{\beta}_0 = 5,60 - (0,15)(7,80) = 4,43$$

$$\hat{Y} = 4,43 + 0,15X$$

$$S_{yy} = 178,7 - 5(5,60)^2 = 21,90$$

$$r = \frac{19,60}{\sqrt{126,80}\sqrt{19,60}} = 0,37$$

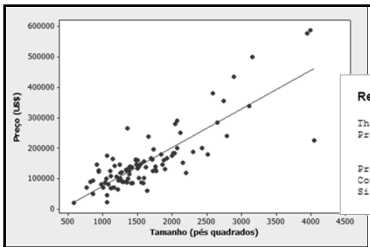
Interpretação

$$\hat{Y} = 4,43 + 0,15X$$

- Inclinação:
  - √ Taxa aumento do resultado do TVC por hora de estudo
  - √ *Quando o tempo aumenta 1 hora, estima-se um aumento de 0,15 no resultado do TVC*
  - √ As estimativas são válidas dentro da classe amostrada (tempos entre 2 e 15 horas)
- Intercepto-*y*
  - √ A reta indica 4,43 pontos no TVC para zero horas de estudo.
  - √ Esta interpretação é válida?
  - √ Há pontos amostrais próximos ao tempo igual a zero suficientes para suportar esta afirmação?

Exemplo

- O que afeta preço de uma casa?
  - √ Preços na Flórida, no outono de 2006
  - √ variáveis:
    - preço de venda (US$)
    - tamanho da casa (pés quadrados)
    - taxas anuais (US$)
    - Número de quartos
    - Número de banheiros
    - Condições da casa (recém construída ou usada)
  - √ Tamanho amostra: 100 casas vendidas



Regression Analysis: Price versus Size

```
The regression equation is
Preço = - 50926 + 127 Tamanho

Predictor      Coef  SE Coef      T      P
Constant     -50926    14896  -3,42  0,001
Size        126,594    8,468  14,95  0,000
```

- Preço = – 50926 + 126,6 Tamanho
  - √ Aumento de US$ 126,6 no preço para cada unidade de área de aumento
  - √ Intercepto não tem interpretação (neste caso)

- Comentários:
  - √ As 100 observações vêm da mesma cidade
    - Não podem ser usadas para fazer inferências entre x e y em geral

Referências

Bibliografia

- Wild, C.J. e Seber, G.A.F. (LTC) *Encontros com o Acaso: um Primeiro Curso de Análise de Dados e Inferência*
- Moore, D.S. e McCabe, G.P. (LTC) *Introdução à Prática da Estatística*
- Agresti, A. e Finlay, B. (Penso) *Métodos Estatísticos para as Ciências Sociais*