

Explorando os dados

Estatística Descritiva:
Medidas de tendência central e
medidas de dispersão



“Um leitão médio num rebanho bem manejado poderá apresentar um ganho médio de 230 a 300 gramas, do nascimento à desmama. Isto equivaleria a um peso aproximado de 11 a 14 quilos à desmama, aos 42 dias de idade, admitindo-se um peso médio de 1200 gramas ao nascimento.” (Sergito de Souza Cavalcanti)

Roteiro da aula

- Tabelas
- Gráficos
- Medidas de tendência central
- Medidas de dispersão

Estatística Descritiva - Tabelas

- Dicas para construção de uma tabela:
 - título (informativo, conciso, sem ambigüidade)
 - cabeçalho curto para cada linha e coluna
 - incluir unidades de medida
 - números com grau de acurácia adequado
 - é mais fácil olhar informações dispostas em colunas do que em linhas

Tabelas

Tabela 1 – Número e percentual de alimentos segundo a origem, incriminados em surtos de doenças transmitidas por alimentos (Paraná, 1998)

Origem do alimento	N	%
Mista	72	66,1
Animal	32	29,4
Vegetal	5	4,6
Total	109	100,0

(Zoli JA et al., *Higiene Alimentar*, vol. 16, n. 95, pp. 62-71 (2002))

Tabela 2 – Agentes etiológicos de surtos de doenças transmitidas por alimentos (Paraná, 1998)

Categoria do agente	Confirmados (A)		Suspeitos (B)		Total
	N	%	N	%	N (%)
1 – Bacteriana					
<i>Salmonella spp</i>	44	57,1	16	53,3	60 (56,1)
<i>Staphylococcus aureus</i>	24	31,2	6	20,0	30 (28,0)
<i>Clostridium perfringens</i>	-	-	4	13,3	4 (3,7)
<i>Bacillus cereus</i>	3	3,9	2	5,7	5 (4,7)
Coliforme fecal	6	7,8	1	-	1 (0,9)
Sub-total	77	100,0	29	96,7	106 (99,1)
2 – Química					
Agrotóxicos	-	-	1	3,3	1 (0,9)
Sub-total	-	-	1	3,3	1 (0,9)
Total	77	100,0	30	100,0	107 (100,0)

(A) Laboratorialmente; (B) Clínica e epidemiologicamente.

Gráficos

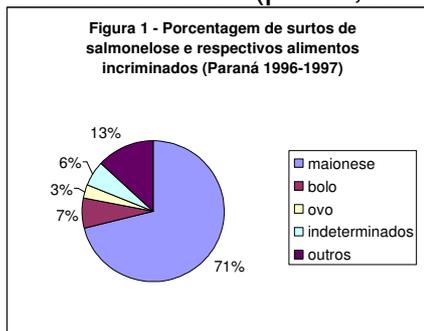
■ Dicas para fazer um gráfico:

- simplicidade
- título (informativo, conciso, sem ambigüidade)
- nomear eixos, segmentos, barras, usando legendas que descrevam os símbolos usados
- apresentar as unidades
- a escolha do tipo de gráfico deve assegurar que toda informação relevante seja mostrada

Gráficos - Dados qualitativos

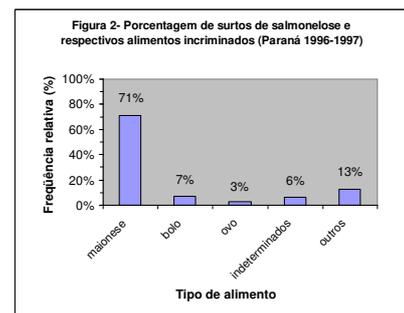
- Para dados qualitativos, cada observação pertence a uma dentre várias categorias ou classes distintas.
- São utilizados, entre outros:
 - gráfico de barras (comprimento)
 - gráfico de setores (ângulo)

Gráfico de setores (pizza, torta)



(Zoli JA et al., *Higiene Alimentar*, vol. 16, n. 95, pp. 62-71 (2002))

Gráfico de barras

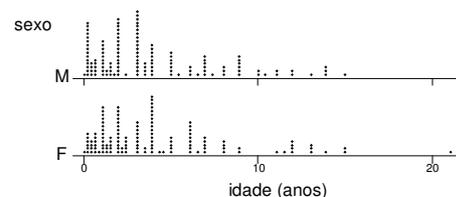


Gráficos - Dados quantitativos

- Diagrama de pontos
- Histogramas
- Boxplot
- Gráficos de dispersão

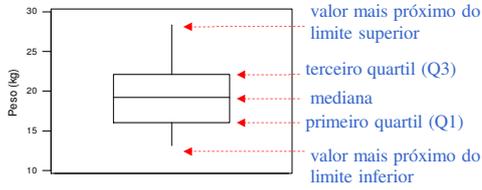
Diagrama de pontos

Figura 3(a): Idades de cães, segundo o sexo, obtidas em levantamento feito em uma escola (Shimozako, 2002)



“Boxplot”

Figura 4: Boxplot para peso de 48 cães



Limite inferior: $Q1 - 1,5 (Q3 - Q1)$

Limite superior: $Q3 + 1,5 (Q3 - Q1)$

“Outliers” (dados discrepantes): dados fora dos limites superior e inferior, indicados com asteriscos

“Boxplot”

Figura 3(b): Idades de cães, segundo o sexo, obtidas em levantamento feito em uma escola (Shimozako, 2002)

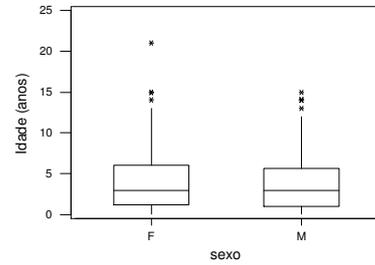
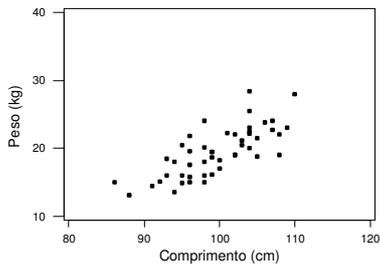
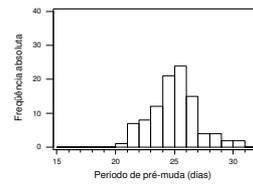


Diagrama de dispersão

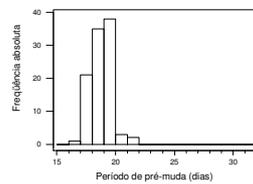
Figura 5: Diagrama de dispersão para comprimento e peso de 48 cães



Histogramas



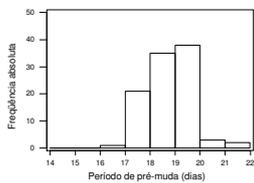
23 °C



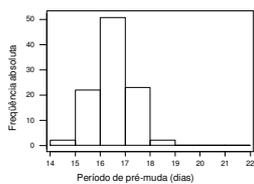
25 °C

Período de pré-muda de ninfa de carrapatos (dados hipotéticos)

Histogramas



25 °C

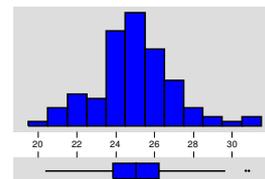


27 °C

Período de pré-muda de ninfa de carrapatos (dados hipotéticos)

Período de pré-muda de ninfa de carrapatos (23°C) (dados hipotéticos)

Descriptive Statistics



Variable: t23

Anderson-Darling Normality Test
A-Squared: 0.480
P-Value: 0.229

Mean 24.9955
StDev 2.0161
Variance 4.06458
Skewness 0.333140
Kurtosis 0.633287
N 100

Minimum 20.3755
1st Quartile 23.8634
Median 25.0360
3rd Quartile 26.2531
Maximum 30.8911

95% Confidence Interval for Mu

95% Confidence Interval for Mu

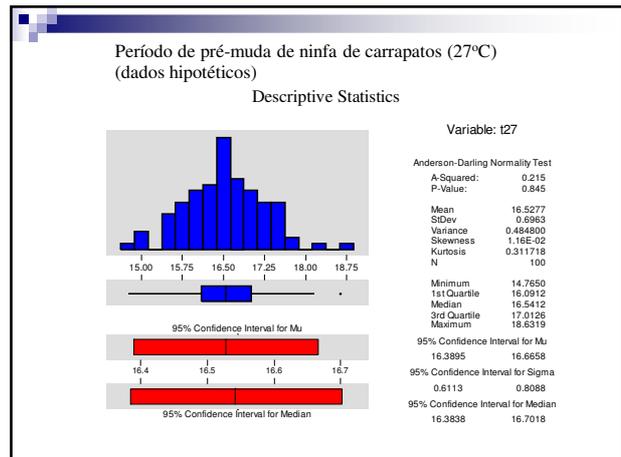
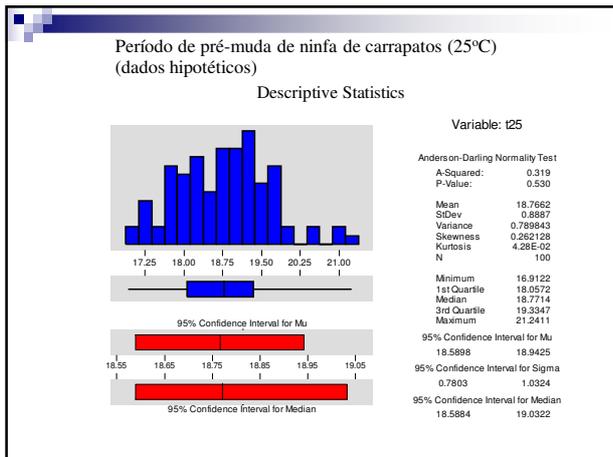
24.5955 25.3956

95% Confidence Interval for Sigma

1.7701 2.3420

95% Confidence Interval for Median

24.6352 25.3205



Medidas de tendência central (de localização ou de posição)

- média aritmética (ou simplesmente média)
- mediana
- moda

Média aritmética

- mais comum
- somam-se todos os valores e divide-se pelo número total de observações
- desvantagem: é influenciada por “outliers” (dados discrepantes)
- é apropriada se a distribuição dos dados é simétrica

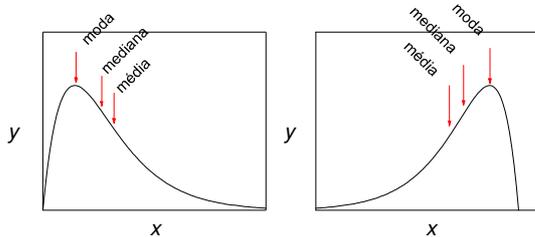
Mediana

- é o valor central em um conjunto de N observações colocadas em ordem de magnitude crescente (ou decrescente)
- mesmo número de observações acima e abaixo da mediana
- vantagem da mediana: não é afetada por “outliers” ou se a distribuição de dados é assimétrica

Moda

- Bastante conhecida, pouco utilizada
- É a observação mais comum (que ocorre com maior frequência) em um conjunto de dados
- Desvantagem: a moda é determinada sem se considerar a maioria das observações
- Há distribuições com mais de uma moda

Distribuições assimétricas



Medidas de dispersão (ou espalhamento)

- Amplitude
- Amplitude interquartil
- Variância
- Desvio Padrão

Amplitude

- Diferença entre o maior e o menor dado observado
- Simples de calcular
- Não é uma boa medida de dispersão, porque seu cálculo se baseia nos valores extremos da amostra e não em todos os dados

Amplitude interquartil

- $Q=Q3-Q1$: É a amplitude de valores que abrange os dados centrais (50%) das observações (diferença entre o terceiro e o primeiro quartil)
- não é influenciada pela presença de outliers
- ignora a maioria das observações

Variância

- Variância mede a dispersão dos dados em torno da média
- Utiliza todas as observações
- É uma medida sensível de dispersão

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

Desvio padrão (s)

- Raiz quadrada da variância
- apresenta a mesma dimensão que os dados originais
- muito útil para distribuições simétricas (p.ex. gaussiana)

Desvio padrão da média

- Também chamado de erro padrão
- É o desvio padrão de uma distribuição amostral de valores médios

$$S_m = \frac{S}{\sqrt{N}}$$

“A mediana não é a mensagem” (Stephen Jay Gould)

<http://www.phoenix5.org/articles/GouldMessage.html>

http://www.cancerguide.org/median_not_msg.html

“There are three kinds of lies: lies,
damn lies, and statistics”

(Mark Twain ou Benjamin
Disraeli?)

“ALL ANIMALS ARE EQUAL
BUT SOME ANIMALS ARE MORE
EQUAL THAN OTHERS”

(George Orwell, *Animal Farm*)