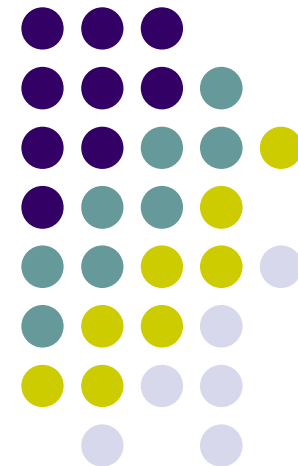


# Tipos de Estudo e Introdução à Análise Estatística

---

Airlane Alencar  
[www.ime.usp.br/~lane](http://www.ime.usp.br/~lane)





# MAE0261

## Programa Resumido

- A Estatística nos problemas científicos.
- Medidas de posição e dispersão.
- Gráficos e tabelas descritivas.
- Estimação por ponto e intervalo.
- Testes de hipóteses.



# MAE0261 - Programa

- 1. O método científico e o processo experimental.
- 2. O trabalho do Estatístico e sua participação nos problemas científicos.
- 3. Medidas e seus erros.
- 4. Escalas de medidas e tipos de variáveis.
- 5. Os conceitos de população e amostras.
- 6. Medidas de tendência central e suas estimativas: Média, Mediana e Moda.
- 7. Medidas de dispersão e suas estimativas: Variância, Desvio Padrão, Intervalo Interquartilício e Coeficiente de Variação.

# MAE0261 - Programa



- 8. Estatística Descritiva: Gráficos, Tabelas e Medidas Amostras.
- 9. Estimativas de parâmetros populacionais a partir das amostras.
- 10. Estimativas por intervalo: o conceito de erro padrão.
- 11. Planejamento estatístico e os protocolos experimentais.
- 12. Testes estatísticos de hipóteses.
- 13. A incerteza e suas medidas no trabalho estatístico: a inferência para a população a partir das amostras. A avaliação dos erros de primeira e segunda espécies.



# MAE0261 - Programa

- 14. A significância estatística e as diferenças biológicas significativas.
- 15. Críticas ao uso indevido de alguns testes estatísticos mais comuns.
- 16. Apresentação de problemas que motivam o uso de modelos mais sofisticados tais como: Análise de Variância (ANOVA), Análise de Regressão e Correlação, Modelos com medidas repetidas, Análise de Variância Multivariada (MANOVA), Estatística não paramétrica, Análise Discriminante, entre outras.
- 17. Cálculo de tamanho de amostras no planejamento de experimentos.

# Exemplos de pesquisas na área

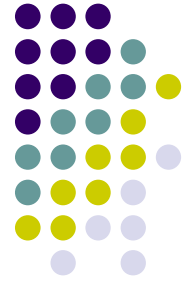


- Procurar trabalhos de pesquisadores do ICB.



# Artigo científico

- Título claro e específico
- Resumo
- Introdução com Objetivos
- Material e Métodos
  - Métodos Estatísticos
- Resultados
- Discussão
- Conclusão
- Bibliografia

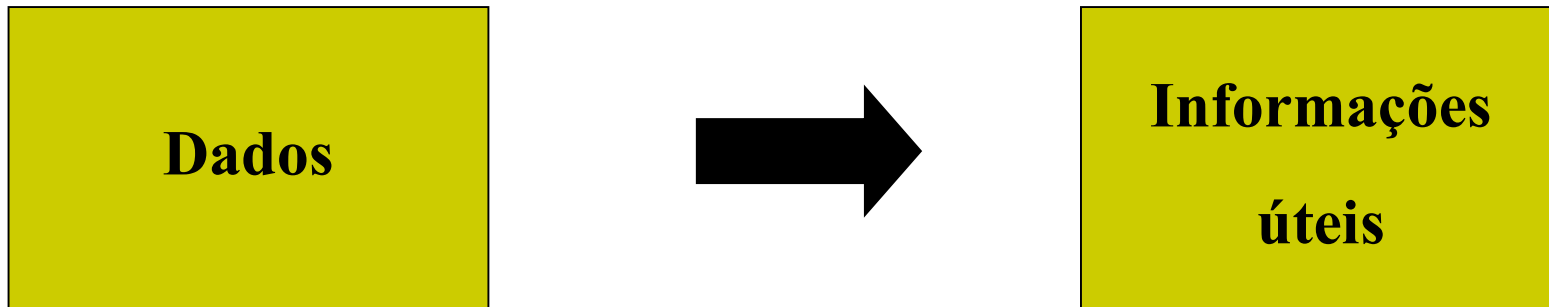


# Fases do Estudo

- Definir o objetivo
- Planejar o estudo e coletar dados
  - (Cód. ética e critérios de exclusão)
- Resultados – Análise Estatística
  - Análise Descritiva
  - Análise Inferencial



# Objetivo Geral



**Objetivo final: Tomada de Decisões**

# Tipos de estudo: Interferência



- **Estudo observacionais:** o pesquisador não impõe um tratamento para cada grupo de pessoas, mas usa as informações já disponíveis sobre o paciente. Mais comuns e viáveis.
- **Ensaio Clínico:** estudo experimental em que o pesquisador em geral aloca os pacientes a cada tratamento.



# População de estudo

- Deve ficar bem clara a população de estudo.
- Duarte, Luciane Simões; Fujimori, Elizabeth; Minagawa, Aurea Tamami; Schoeps, Flavia Antonini; Montero, Rosali Maria Juliano Marcondes ***Aleitamento materno e níveis de hemoglobina em crianças menores de 2 anos em município do estado de São Paulo, Brasil.*** Revista de Nutrição, 2007, vol.20, n. 2.
- **OBJETIVO:** Avaliar a relação entre aleitamento materno e níveis de hemoglobina em **crianças menores de 2 anos residentes na cidade de Itupeva, SP, Brasil.**



# Tipo de estudo: Tempo

- **Longitudinal:** ao longo do tempo.
- **Transversal ou cross-section:** realizado em um determinado instante de tempo.
- **MÉTODOS:** Estudo **transversal**, realizado com 573 escolares de 8 a 10 anos dos municípios de Dois Irmãos e Morro Reuter, Rio Grande do Sul. Foram coletados dados sobre insatisfação corporal e peso das crianças por meio de escala de imagem corporal (*Children's Figure Rating Scale*) e antropometria.

# Tipos de Estudos Longitudinais



- **Prospectivo**: acompanha pacientes ao longo do tempo – follow up. Em geral, são mais caros e há mais perda de dados, mas costumam ser mais precisos.
- **Retrospectivo**: baseado em dados de períodos passados.



# Tipos de estudo: Descritivo

- **Estudo descritivo:** apresentar de forma organizada informações sobre os pacientes atendidos ou dados produzidos por serviços de informação (estudo de caso).



# Estudo Descritivo

- Garófolo, Adriana; Boin, Susiane Gusi; Modesto, Patrícia Claudia; Petrilli, Antonio Sérgio *Evaluation of parenteral nutrition efficiency in supplying energy for pediatric oncology patients.* Revista de Nutrição, 2007, vol.20, n. 2, ISSN 1415-5273.
- **OBJETIVO:** Avaliar a eficiência da nutrição parenteral administrada a pacientes oncológicos pediátricos quanto à oferta de energia, indicações e principais dificuldades no procedimento.
- **MÉTODOS:** Estudo observacional descritivo, realizado entre julho de 2003 e julho de 2004 no Instituto de Oncologia Pediátrica da Universidade Federal de São Paulo. Critérios de inclusão: pacientes com câncer recebendo nutrição parenteral nas unidades de internação. A análise da adequação energética foi efetuada usando-se equações preditivas para o cálculo do gasto energético basal ajustado para a progressão inicial.



# Estudo Caso-control

- Temos 2 grupos de pessoas, casos (doentes) e controles, e verificamos se a exposição a um possível fator de risco difere nos dois grupos.
- Os controles devem ser comparáveis aos casos com relação à idade, sexo, e outras variáveis relevantes, mas sem a doença. É comum obter 1 ou 2 controles para cada caso (pareados com relação a variáveis relacionadas com o problema).





# Estudo Caso-control

- XIMENES, Ricardo Arraes Alencar, GALLO, Maria Eugênia Novinsk, BRITO, Maria de Fátima de Medeiros.  
**Retratamento de hanseníase: estudo de caso-control.**  
*Rev. Saúde Pública*, ago. 2007, vol.41, no.4, p.632-637.
- **OBJETIVO:** Analisar os fatores de risco para retratamento da hanseníase.
- **MÉTODOS:** Estudo de caso-control com pacientes de duas unidades de referência para tratamento da hanseníase, em Recife, Pernambuco, no de 2003. O grupo de casos incluiu pacientes retratados (N=155) e foi comparado com o grupo controle (N=155), pacientes não retratados pareados por ano-diagnóstico e unidade de saúde.



# Estudo Caso-Controle

- Gomes (1992).
- Casos= Pacientes portadoras de câncer de mama no HC da UFMG com idades entre 25 e 75 anos, diagnóstico entre 1978-1987, sem cirurgia de mama.
- Controles= Pacientes com idade igual ao caso e data de admissão no hospital igual à data de diagnóstico do caso. Dois controles para cada caso.
- Conclusão: A presença de histórico familiar de câncer de mama aumenta o risco da doença em 8,84 vezes (OR).



# Tipos de estudo: Coorte

- Temos 2 grupos de pacientes, expostos e não expostos a certo fator de risco, e queremos verificar se a proporção de pacientes que desenvolvem certa doença é a mesma nos 2 grupos ou não. As taxas de incidência são calculadas nos 2 grupos.
- A comparabilidade dos grupos deve ser analisada no início do estudo. Vantagens: Melhor qualidade dos dados.
- Podemos calcular o Risco Relativo.



# Tipo de estudo: Coorte

	Aspirina	Placebo
Amostra	11037	11034
Infartos	139	239
Proporção	1.26%	2.17%

- O risco relativo de ter infarto para quem toma placebo com relação a quem toma aspirina é 1,72 (= 2.17/1.26).
- Isso pode não ser estatisticamente diferente de 1.

# Tipo de estudo: Ensaio Clínico



- Ensaio clínico aleatorizado é um experimento médico, realizado em geral com o objetivo de verificar, entre 2 ou mais tratamentos, qual é o mais efetivo.
- Protocolo do estudo, Código de ética, Critérios de exclusão

# Ensaio clínico: Tamoxifeno



- Tamoxifeno e câncer de mama
- Pacientes com bom prognóstico, idade < 70 anos ao longo de 4 anos
- No grupo com tamoxifeno 83% estavam livres da doença no final do estudo enquanto no grupo placebo essa porcentagem foi de 73%.

# Ensaio clínico: Parar de fumar



- Otero, Ubirani Barros; Perez, Cristina de Abreu; Szklo, Moysés; Esteves, Gilcenira Ataliba; Pinho, Mariana Marques de; Szklo, André Salem; Turci, Silvana Rubano B. ***Ensaio clínico randomizado: efetividade da abordagem cognitivo-comportamental e uso de adesivos transdérmicos de reposição de nicotina, na cessação de fumar, em adultos residentes no Município do Rio de Janeiro, Brasil.*** Cadernos de Saúde Pública, 2006, vol.22, n. 2.

# Ensaio clínico: Parar de fumar



Participaram 1.199 voluntários, adultos, do Município do Rio de Janeiro, Brasil, alocados aleatoriamente em dez grupos: aconselhamento breve (GB) intensivo de 1 ou 2 sessões (G1-G2) e de 3 ou 4 sessões (G3-G4), com/sem terapia de reposição de nicotina.



# Ensaio clínico: Parar de fumar



- Proporções de abstinência foram estimadas durante os 12 meses. Essas proporções nos grupos sem adesivo foram 20% (GB), 17% (G1-G2) e 23% (G3-G4).
- Nos grupos com adesivo foram aproximadamente 30% (GBA), 34% (G1A-G2A) e 33% (G3A-G4A). Após ajuste múltiplo, as razões de proporção de abstinência parecem seguir padrão consistente com efeito "dose-resposta": comparadas com GB foram 0,85 (G1-G2); 1,13 (G3-G4); 1,51 (GBA); 1,66 (G1A-G2A) e 1,75 (G3A-G4A) ( $p < 0,05$ ).
- Os resultados sugerem que crescer terapia de reposição de nicotina aumenta a proporção de abstinência na cessação.



# Cross-over

- The clinical effect of a new infant formula in term infants with constipation: a **double-blind, randomized cross-over trial**. **MEJ Bongers, F Lorijn, JB Reitsma, M Groeneweg, JAJM Taminau and MA Benninga**, *Nutrition Journal* 2007, 6:8.
- Nutrilon Omneo (new formula; NF) contains high concentration of *sn*-2 palmitic acid, a mixture of prebiotic oligosaccharides and partially hydrolyzed whey protein. It is hypothesized that NF positively affects stool characteristics in constipated infants.
- Thirty-eight constipated infants, aged 3–20 weeks, were included and randomized to NF (n = 20) or a standard formula (SF; n = 18) in period 1 and crossed-over after 3 weeks to treatment period 2. Constipation was defined by at least one of the following symptoms: 1) defecation frequency < 3/week; 2) painful defecation; 3) abdominal or rectal palpable mass.

# Cross-over



- Results
- Period 1 was completed by 35 infants.
- A significant increase in defecation frequency (NF: 3.5 pre versus 5.6/week post treatment; SF 3.6 pre versus 4.9/week post treatment) was found in both groups, but was not significantly different between the two formulas ( $p = 0.36$ ).
- Improvement of hard stool consistency to soft stool consistency was found more often with NF than SF, but did not reach statistical significance (90% versus 50%; RR, 1.8; 95% CI, 0.9–3.5;  $p = 0.14$ ).
- No difference was found in painful defecation or the presence of an abdominal or rectal mass between the two groups. Twenty-four infants completed period 2. Only stool consistency was significantly different between the two formulas (17% had soft stools on NF and hard stools on SF; no infants had soft stools on SF and hard stools on NF, McNemar test  $p = 0.046$ ).

# Red wine consumption increases antioxidant status and decreases oxidative stress in the circulation of both young and old humans.

*Nutrition Journal* 2007, 6:27



- **Background**

Red wine contains a naturally rich source of antioxidants, which may protect the body from oxidative stress, a determinant of age-related disease. The current study set out to determine the in vivo effects of moderate red wine consumption on antioxidant status and oxidative stress in the circulation.

- **Methods**

20 young (18–30 yrs) and 20 older ( $\geq 50$  yrs) volunteers were **recruited**. Each age group was randomly divided into treatment subjects who consumed 400 mL/day of red wine for two weeks, or control subjects who abstained from alcohol for two weeks, after which they **crossed over** into the other group. Blood samples were collected before and after red wine consumption and were used for analysis of whole blood glutathione (GSH), plasma malondialdehyde (MDA) and serum total antioxidant status.

# Red wine consumption increases antioxidant status and decreases oxidative stress in the circulation of both young and old humans. *Nutrition Journal* 2007, 6:27



- **Results**

This study shows consumption of red wine induced significant increases in plasma total antioxidant status ( $P < 0.03$ ), and significant decreases in plasma MDA ( $P < 0.001$ ) and GSH ( $P < 0.004$ ) in young and old subjects. The results show that the consumption of 400 mL/day of red wine for two weeks, significantly increases antioxidant status and decreases oxidative stress in the circulation

- **Conclusion**

It may be implied from this data that red wine provides general oxidative protection and to lipid systems in circulation via the increase in antioxidant status.

# Exemplo de conjunto de dados

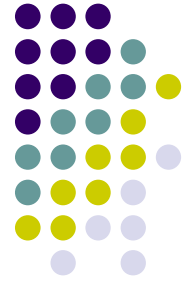


- A fim de estudar o perfil das pessoas que trabalham em certa empresa aplicamos um questionário.
- Variáveis: Idade, Profissão, Estado civil, Renda (em s.m.), sexo, colesterol, ...

# Conjunto de dados - lay out\*



Obs	Estado Civil	Grau de Instrução	Filhos	Salário	Idade	Região de Procedência
1	Solteiro	1	0	4.00	26	2
2	Casado	1	1	4.56	32	1
3	Casado	1	2	5.25	36	1
4	Solteiro	2	0	5.73	20	3
5	Solteiro	1	0	6.26	10	3
6	Casado	1	0	6.66	28	2
7	Solteiro	1	0	6.86	41	2
8	Solteiro	1	1	7.39	43	1



# Tipos de variáveis

- Quantitativas: numéricas
  - Discretas: número de filhos
  - Contínuas: renda, concentração de alguma substância
- Qualitativas ou categóricas
  - Nominal: estado civil, sexo, presença de diabetes (s/n)
  - Ordinal: grau de instrução, dor forte, moderada ou fraca





# Análise Descritiva

- Tabelas
- Gráficos
- Medidas resumo
  - Posição: Média, Mediana, Máximo
  - Variabilidade: Desvio Padrão, Amplitude
  - Associação: Correlação, Risco Relativo

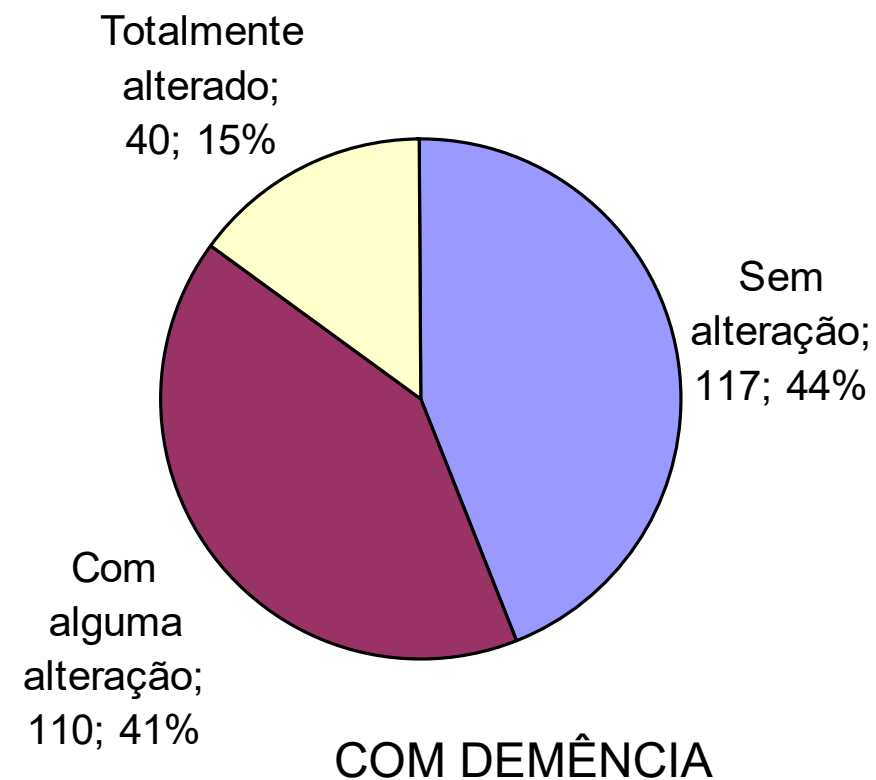
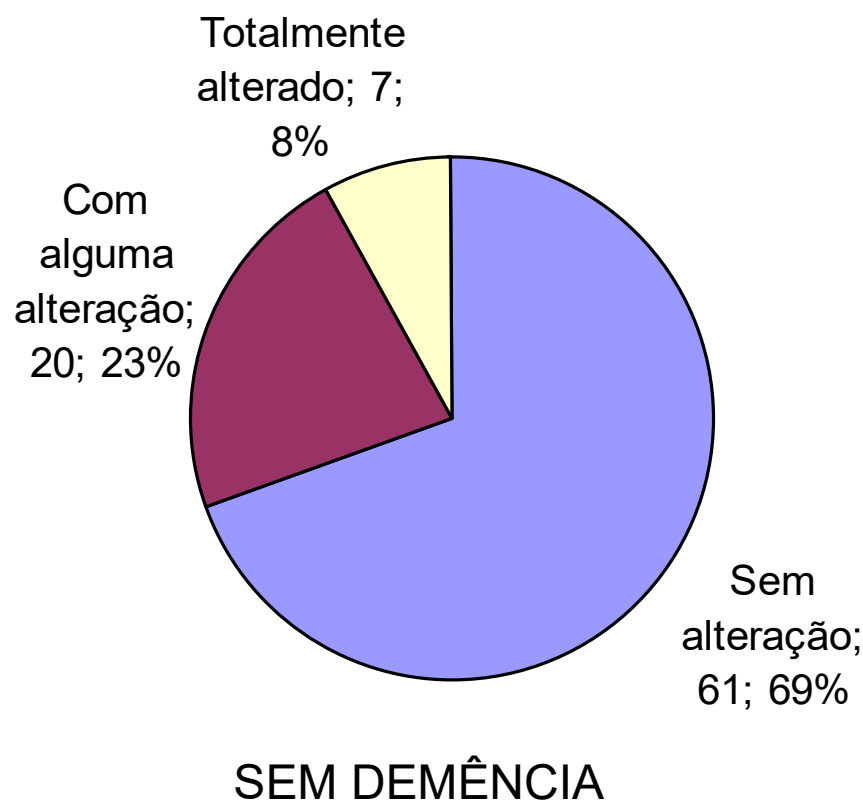


# Variáveis qualitativas

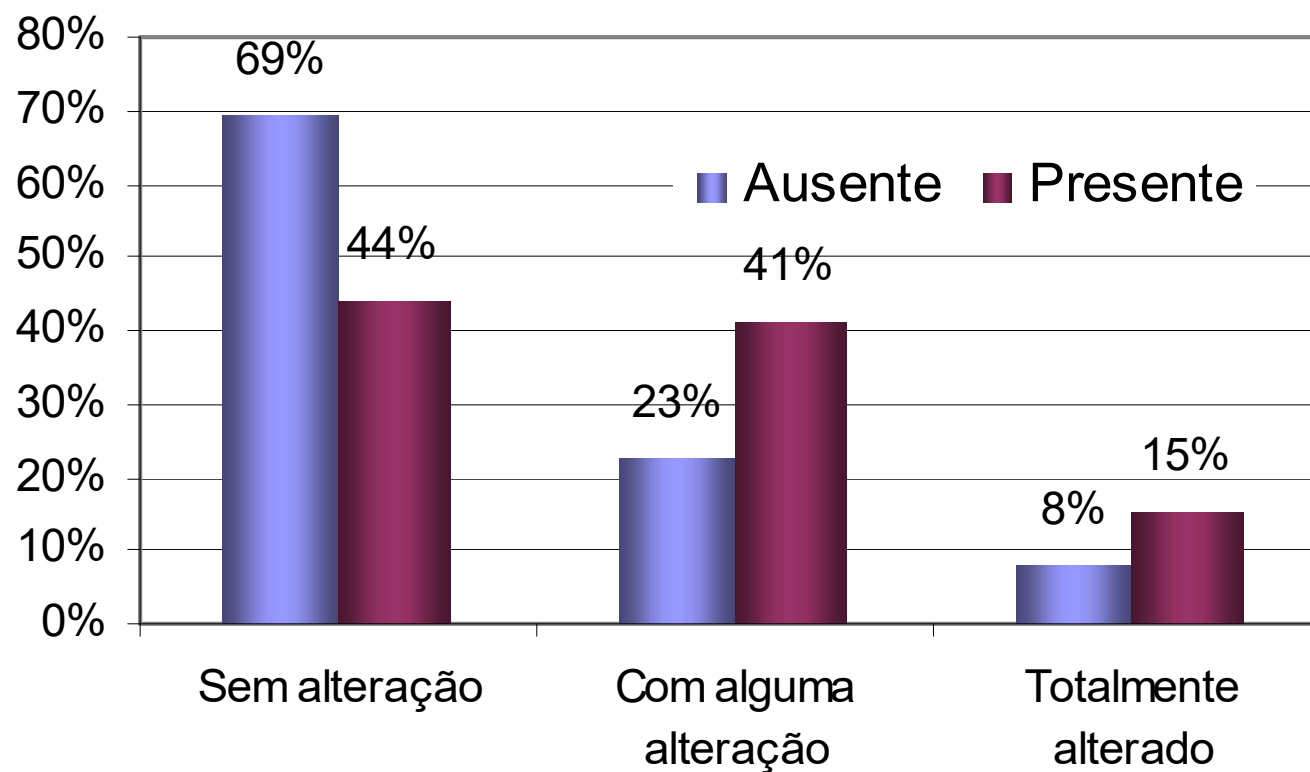
- Tabelas de frequências

Demência	Grau de incapacidade			Total
	Sem alteração	Com alguma alteração	Totalmente alterado	
Ausente	61 (69,3%)	20 (22,7%)	7 (8,0%)	88 (100,0%)
Presente	117 (43,8%)	110 (41,2%)	40 (15,0%)	267 (100,0%)

# Gráficos de setores



# Comparação: Colunas



# Var. quantitativas: Medidas descritivas



- Posição:
  - Média, Mediana, Máximo
- Variabilidade:
  - Desvio Padrão, Amplitude

# Média e Mediana

- Média e Mediana

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{y_1 + y_2 + \dots + y_n}{n}$$

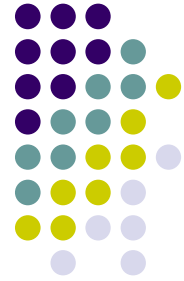


# Outras medidas de posição



- Outras medidas de posição
  - Mediana
  - Quantis

# Medidas de dispersão



- Medidas de dispersão
  - Variância
  - Desvio Padrão = Raiz da Variância
  - Amplitude = Máximo – Mínimo





# Variância

- Variância amostral

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{(y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n-1}$$

- Para dados populacionais, divide-se por n em vez de n-1.



# Desvio e Erro Padrão

- Desvio Padrão amostral

$$s = \sqrt{s^2}$$

- Erro Padrão

$$ep = \frac{s}{\sqrt{n}}$$

- Qual a diferença?

# Coeficiente de variação



- Coeficiente de variação é o desvio padrão dividido pela média

$$CV = \frac{S}{\bar{y}}$$

- Baixo se  $<0,1$       Médio de  $0,1$  a  $0,2$
- Alto de  $0,2$  a  $0,3$       Muito alto  $>0,3$

# Coeficiente de variação



- Colesterol em 2 grupos
- Grupo 1: Jovens médicos residentes:  
Média= 205 mg/dL e dp=22 mg/dL
- Grupo 2: Médicos especialistas  
Média= 244 mg/dL e dp=45 mg/dL
- A variabilidade é maior no grupo 2, bem como sua média.

$$CV = \frac{S}{\bar{y}}$$

# Coeficiente de variação



- Colesterol em 2 grupos
- Grupo 1: Jovens médicos residentes:  
Média= 205 mg/dL, dp=22 mg/dL => cv=0,107
- Grupo 2: Médicos especialistas  
Média= 244 mg/dL, dp=45 mg/dL => cv=0,184

$$cv = \frac{s}{\bar{y}}$$

# Linfócitos segundo Grupo

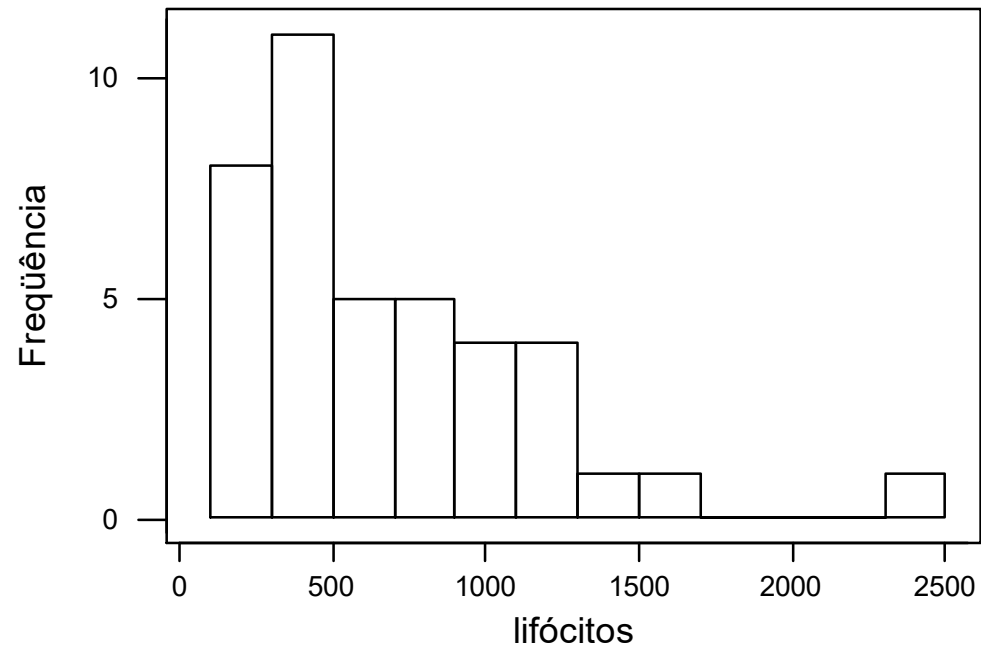


grupo	Hodgkin	Não Hodgkin
n	20	20
Média	823.0	522.1
Erro Padrão	127.0	65.5
Mediana	682.0	433.0
Mínimo	171.0	116.0
Máximo	2415.0	1252.0

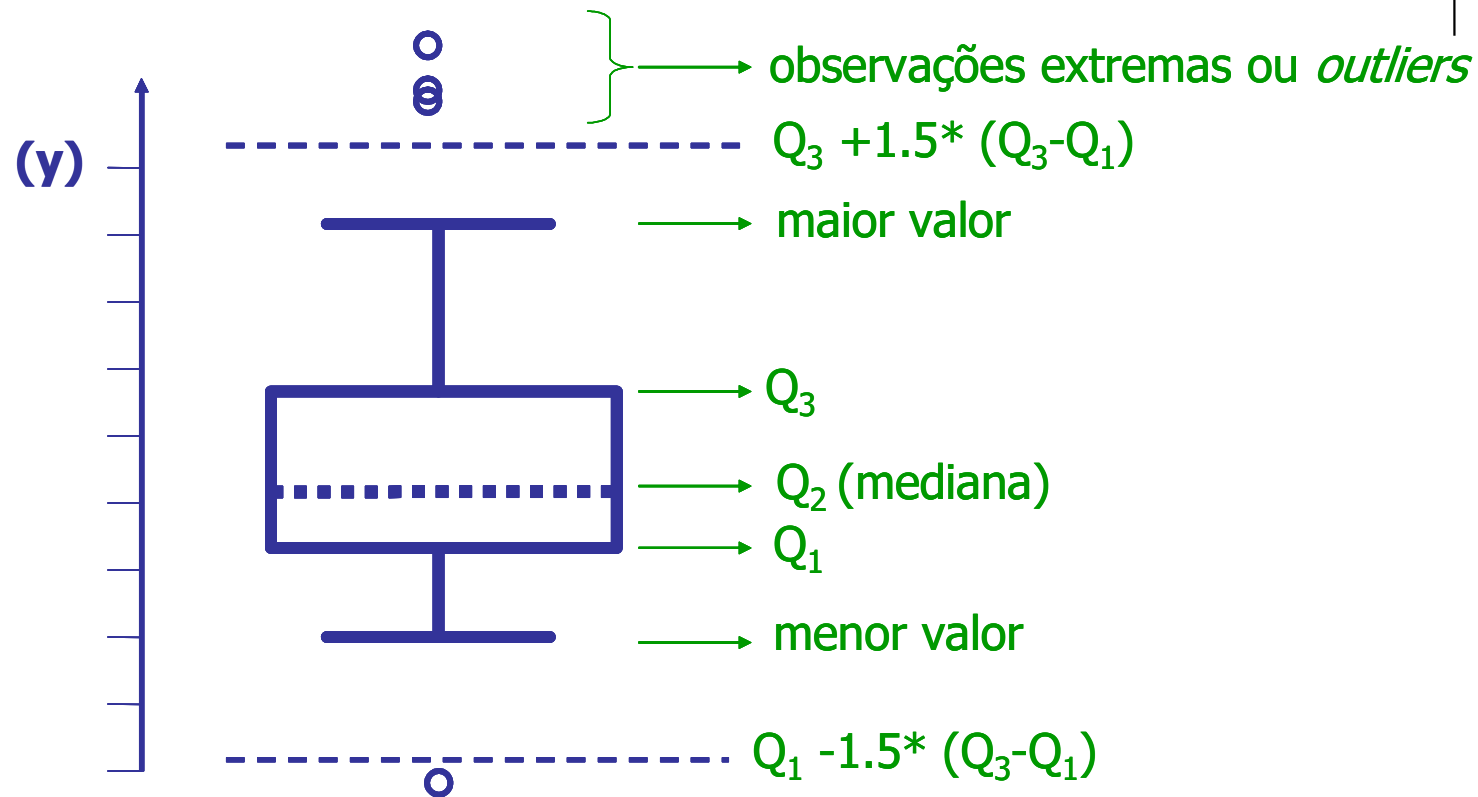
# Var.quantitativa: Histograma



Hodgkin



# Box-plot



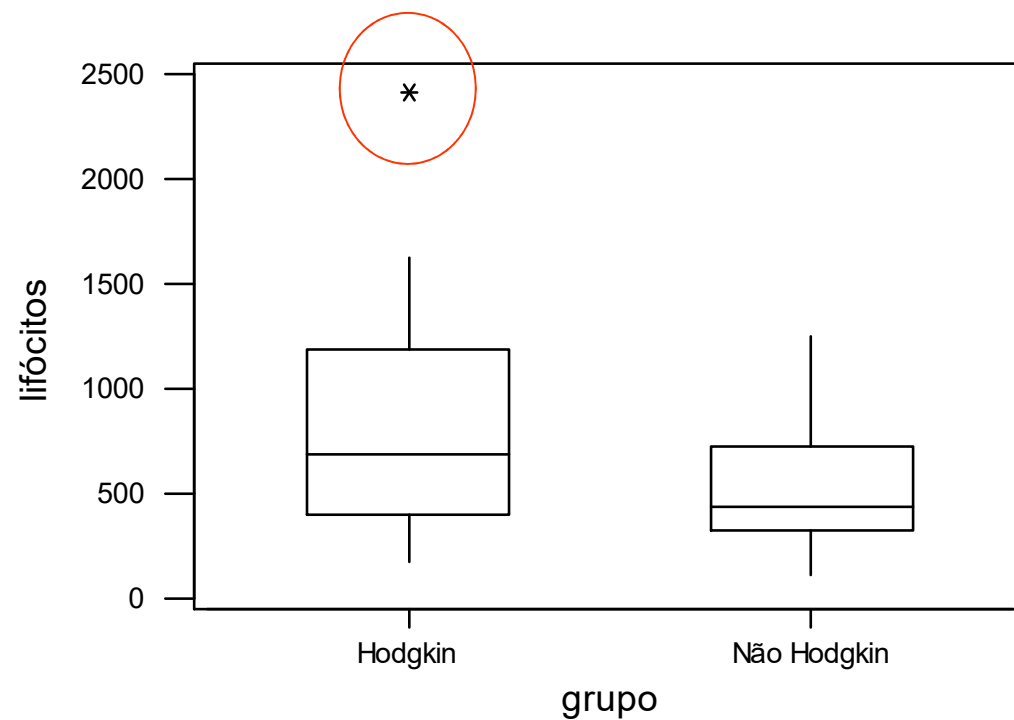
- Dados: {3, 4, 4, 5, 6, 6, 8, 9}



# Box plots



outlier





# Excluindo o outlier

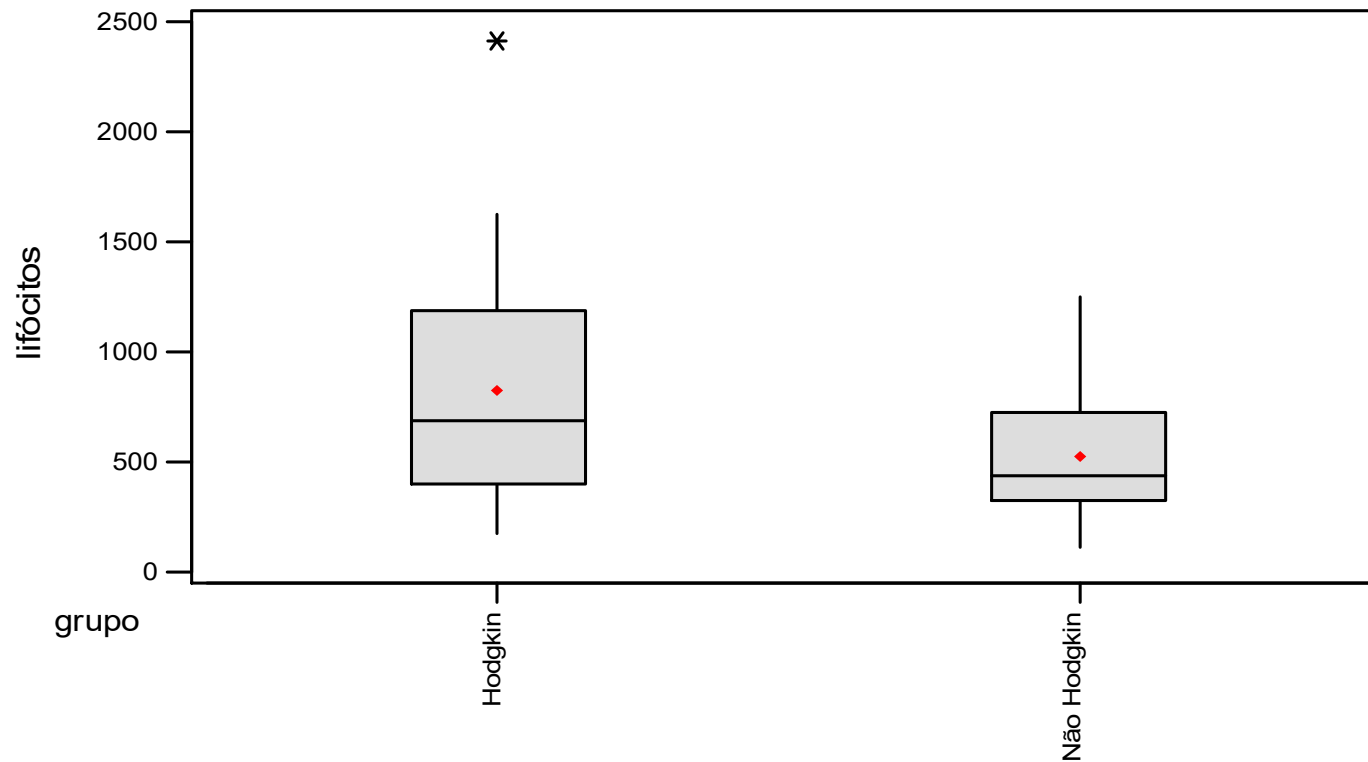
grupo	Hodgkin	Não Hodgkin
n	20	20
Média	823.0	522.1
Erro Padrão	127.0	65.5
Mediana	682.0	433.0
Mínimo	171.0	116.0
Máximo	2415.0	1252.0

grupo	Hodgkin	Não Hodgkin
n	19	20
Média	739.4	522.1
Erro Padrão	100.1	65.5
Mediana	568.0	433.0
Mínimo	171.0	116.0
Máximo	1621.0	1252.0

# Box plots e médias



Boxplots of lifócito by grupo  
(means are indicated by solid circles)

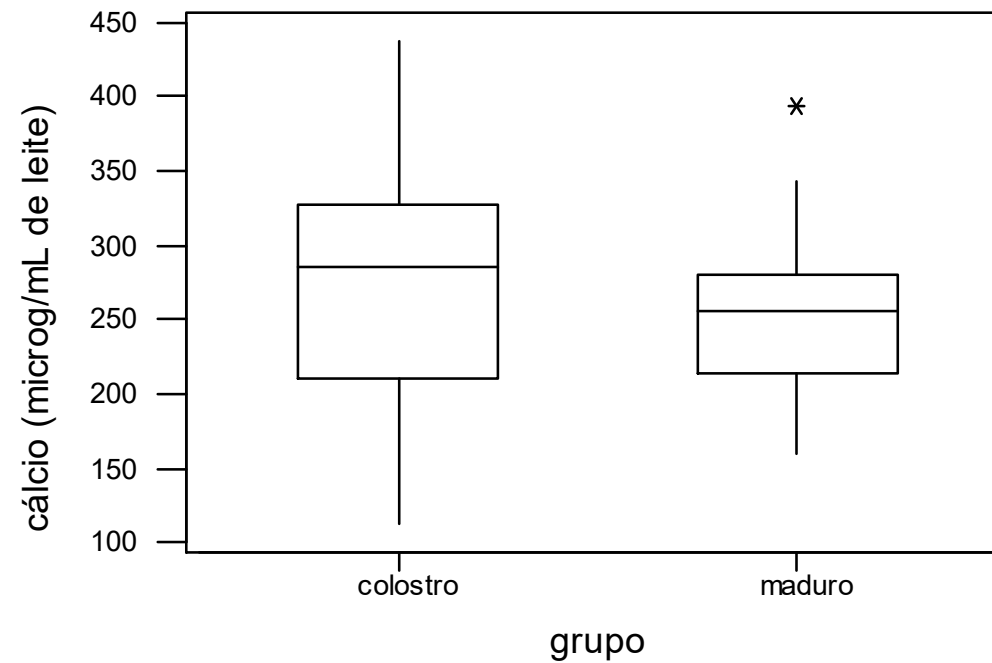


# Concentração de minerais do leite

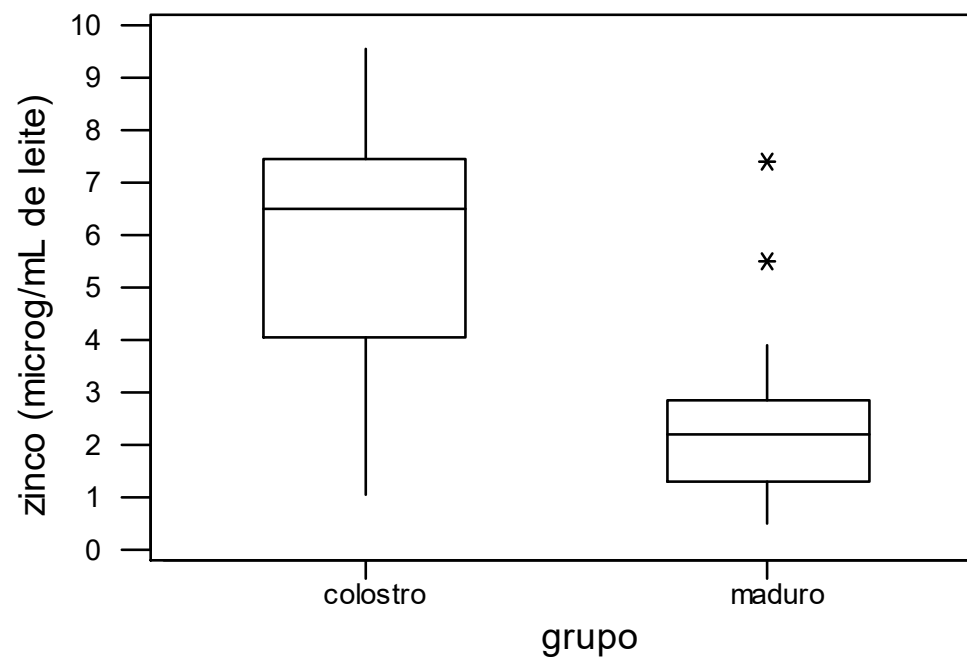


- Concentração de minerais do leite
- Amostra: 55 mães do hospital Odete Valadares em BH
- 2 grupos segundo o período de lactação: colostro (n=26) e leite maduro (n=29)
- Variáveis de interesse: concentração de cálcio e zinco

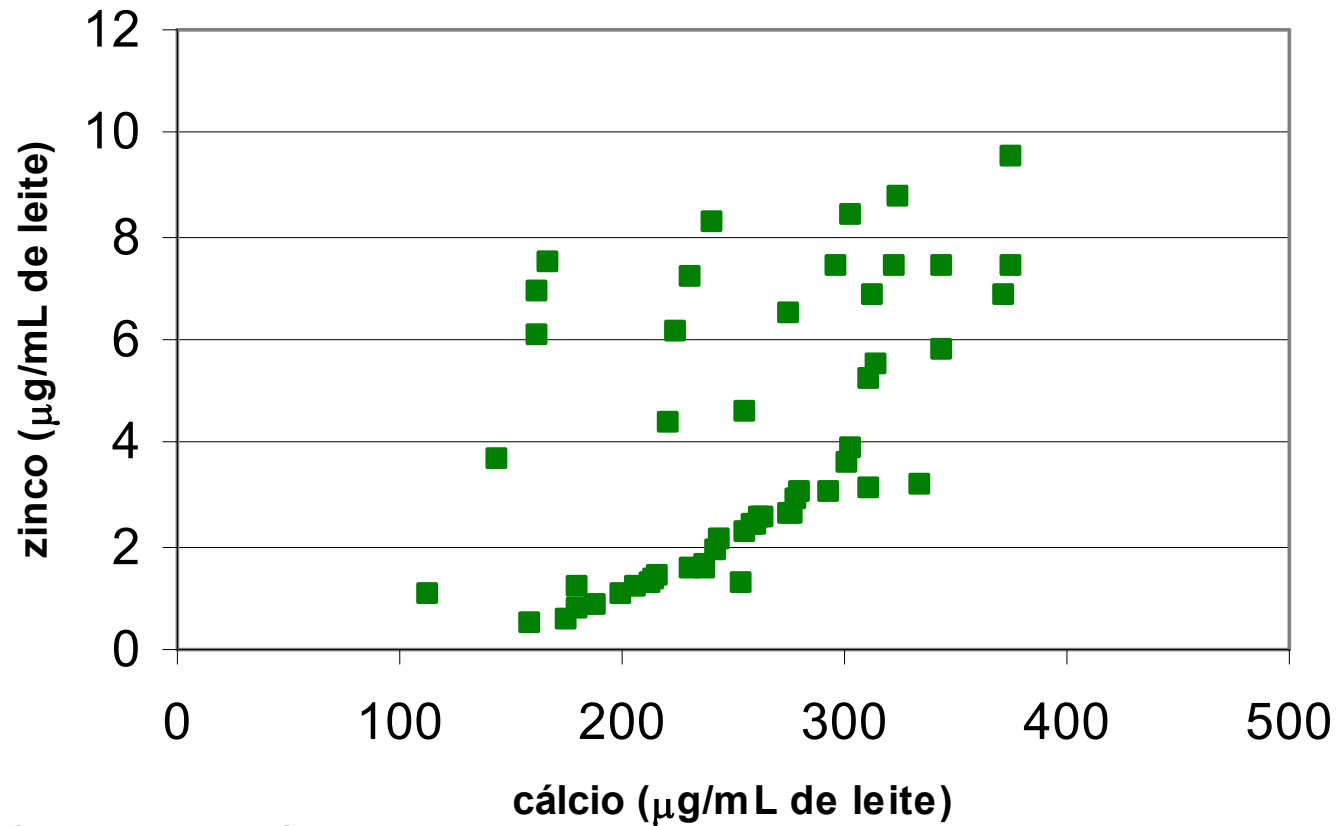
# Concentração de minerais do leite



# Concentração de minerais do leite

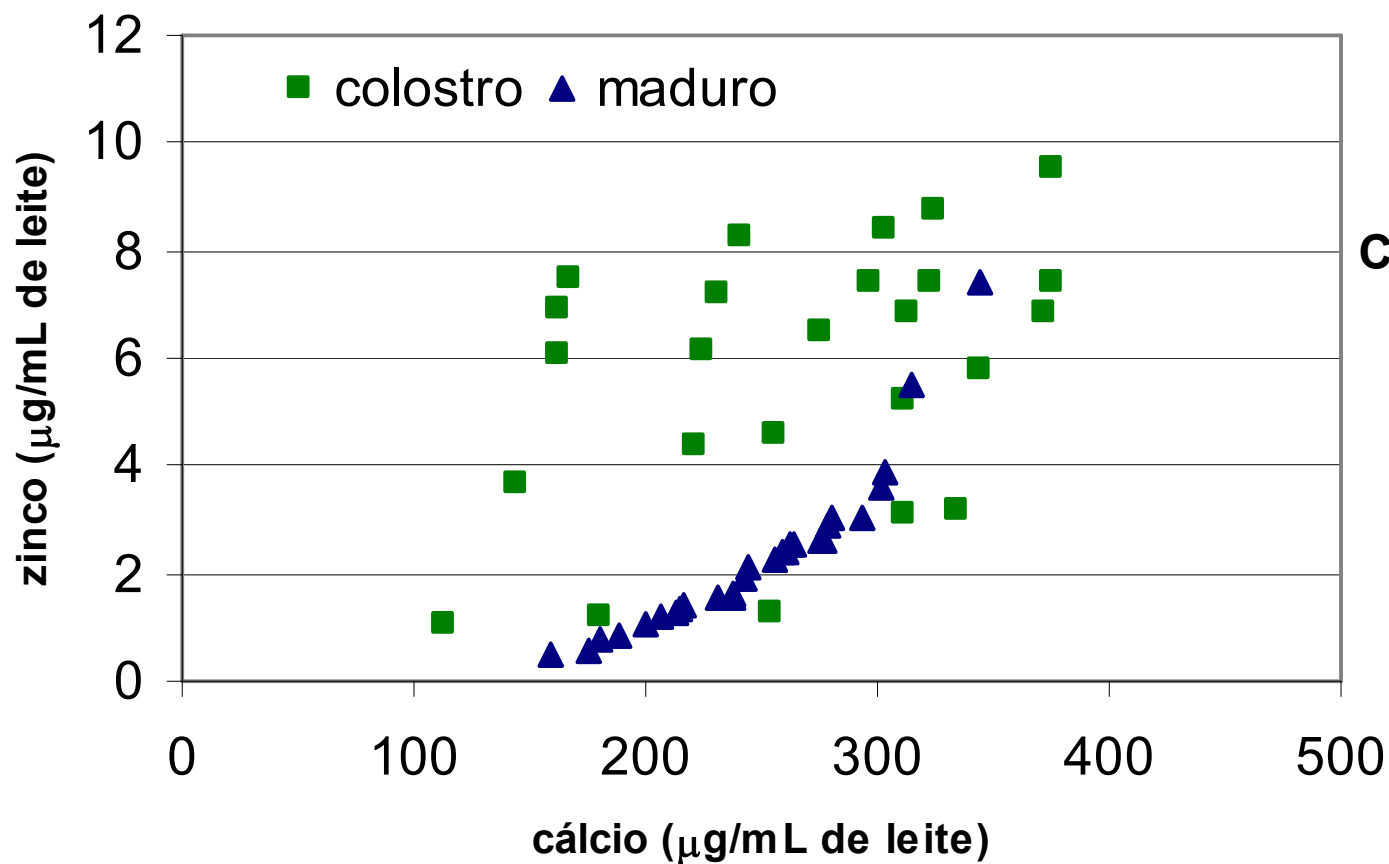


# Concentração de minerais do leite



Conclusões?

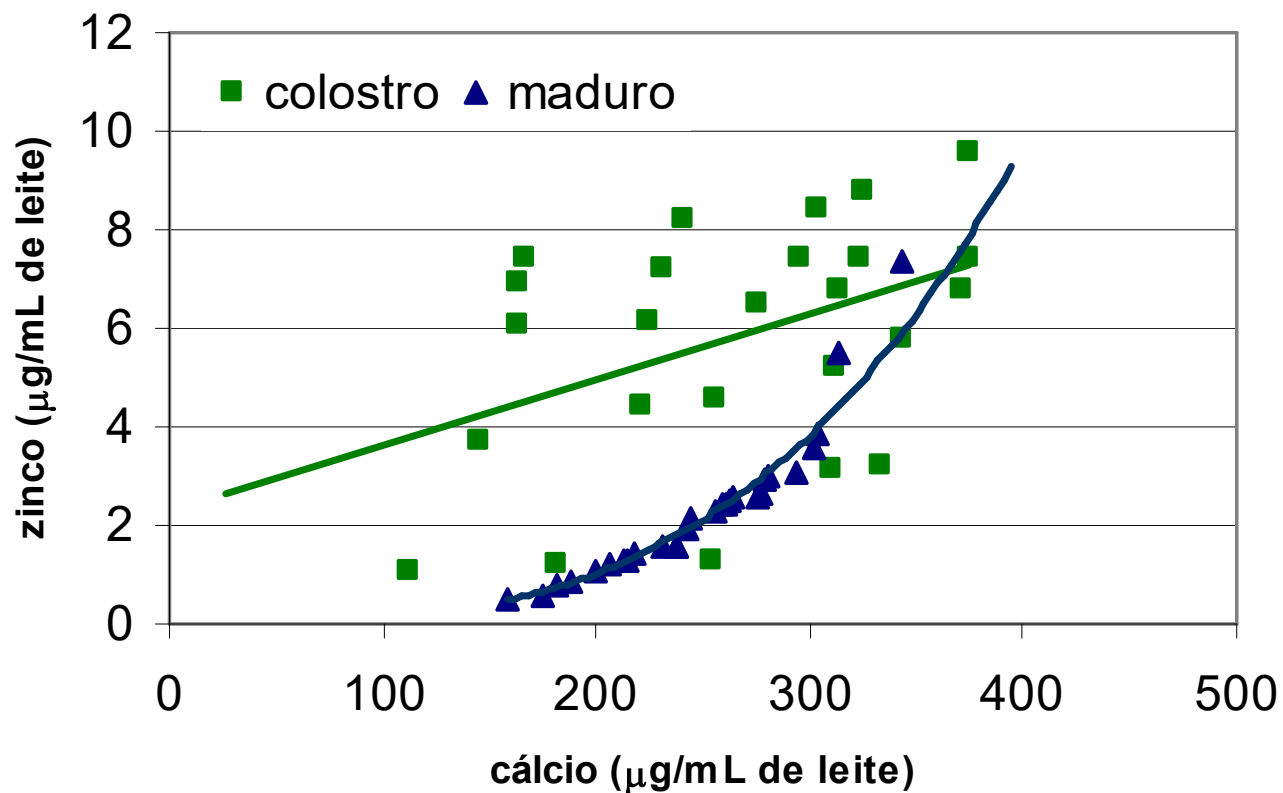
# Regressão



Correlações  
**0,4286**  
**0,9829**



# Regressão

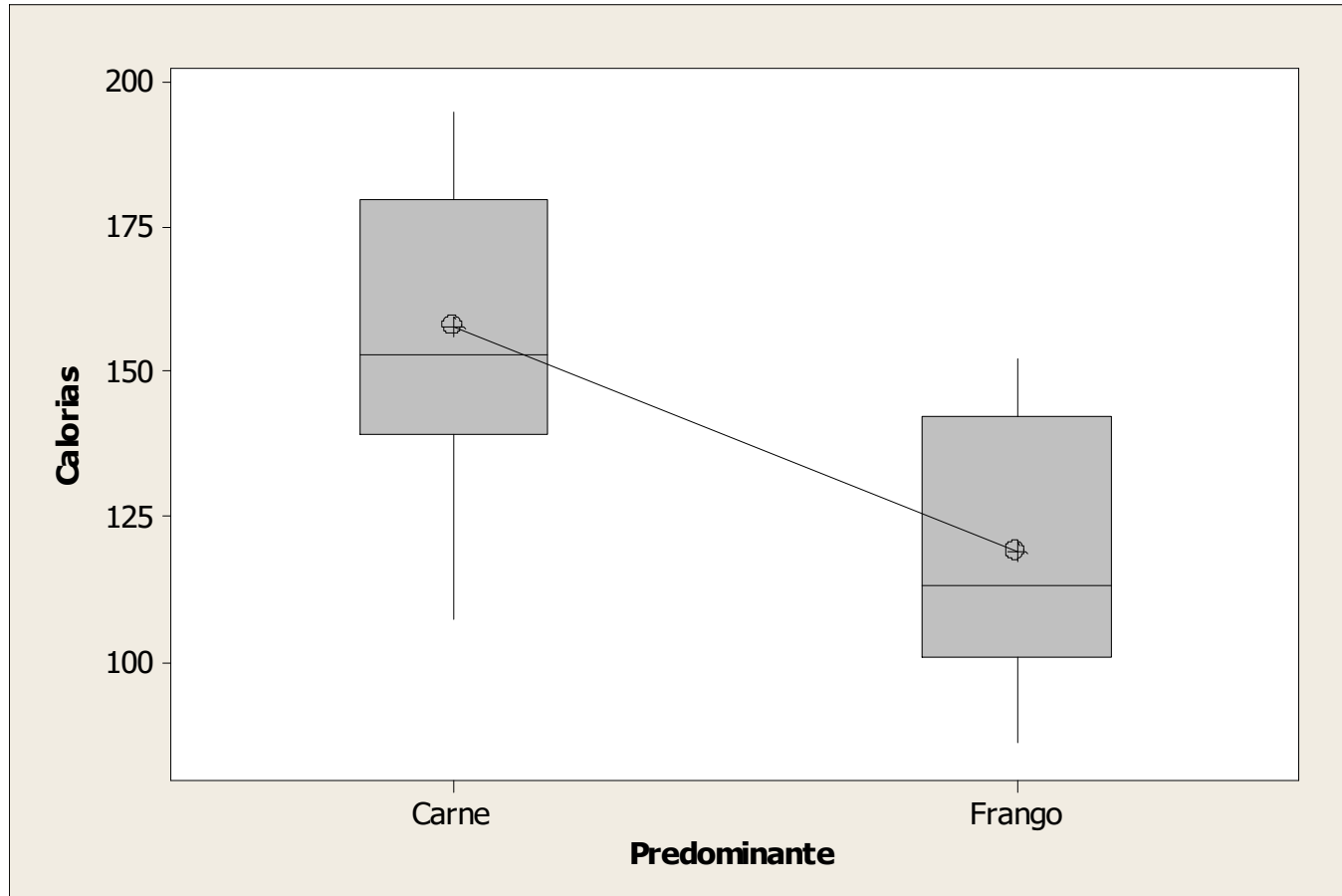
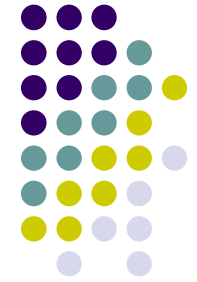




# Análise Inferencial

- Testes de hipótese
  - Testar se a quantidade média de calorias por hot dog é a mesma para salsichas de frango e predominantemente de carne.
  - Testar se Peso e Pressão arterial são variáveis independentes.
- Intervalos de confiança

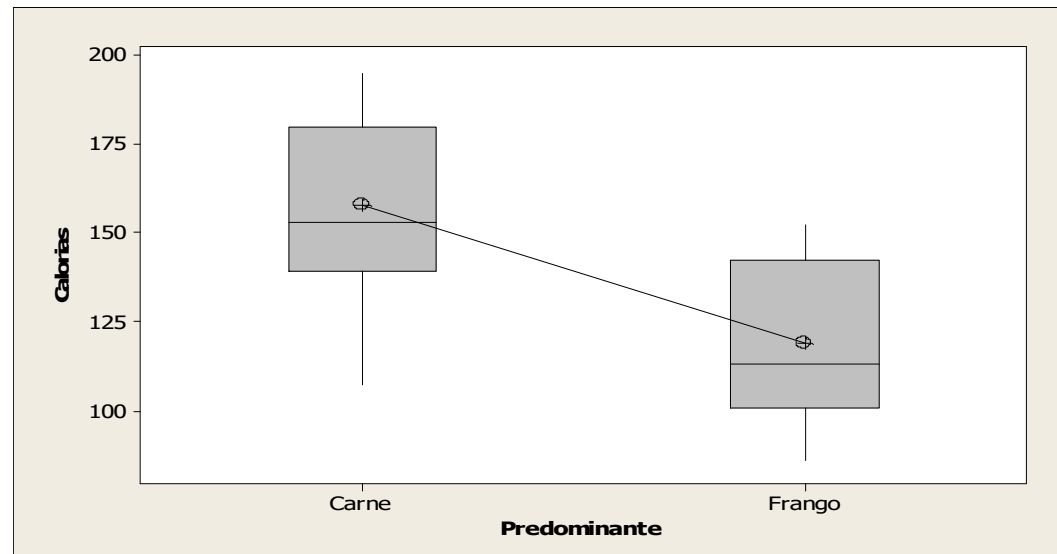
# Hot Dog



# Hot Dog



Tipo de Carne	N	Média	D.Pad	Erro Padrão
Carne	37	157,7	23,5	3,9
Frango	17	118,8	22,6	5,5



# Intervalo de Confiança para média



Tipo de Carne	N	Média	D.Pad	Erro Padrão
Carne	37	157,7	23,5	3,9
Frango	17	118,8	22,6	5,5

- Intervalos de Confiança para a médias com coeficiente de confiança 95%:

Tipo de Carne	
Carne	[149,8; 165,6]
Frango	[107,2; 130,4]

# Teste de hipóteses



- Idéia: Supomos que uma hipótese ( $H_0$ ) seja verdadeira, por exemplo, igualdade das quantidades médias de calorias para salsichas de frango e de carne.
- Calculamos as médias, os erros padrões em cada grupo e uma estatística que mede a diferença das médias e leva em conta os desvios padrão.

# Teste de hipóteses



- Se a estatística que calculamos assume um valor grande com relação aos valores que nós esperamos sob a hipótese  $H_0$ , então rejeitamos a hipótese.
- Caso contrário, não rejeitamos a hipótese.
- Não é questão de aceitar  $H_0$ !



# Teste de hipóteses

- Comparar as médias de 2 grupos
- Hipótese nula –  $H_0: \mu_1 = \mu_2$
- Hipótese alternativa –  $H_1: \mu_1 \neq \mu_2$
- Podemos errar de 2 modos:
  - Erro tipo I: rejeitar  $H_0$  |  $H_0$  verdadeiro
  - Erro tipo II: não rejeitar  $H_0$  |  $H_0$  falso
  - Temos que controlar as probabilidades desses erros.
- Na prática, fixamos a probabilidade de Erro Tipo I. Podemos calcular tamanhos de amostra, fixando a priori probabilidade do Erro Tipo II.



# Teste de hipóteses



- $\alpha = P(\text{Erro tipo I}) = P(\text{rej } H_0 \mid H_0 \text{ verdadeiro})$
- $\beta = P(\text{Erro tipo II}) = P(\text{n\~{a}o rej. } H_0 \mid H_0 \text{ falso})$
- Fixamos o  $\alpha = 5\%$ . Se diminui  $\alpha$ , aumenta  $\beta$ .
- Poder =  $1 - \beta = P(\text{rej. } H_0 \mid H_0 \text{ falso})$

# Teste de Hipóteses – Hot Dogs



- Difference =  $\mu$  (Carne) -  $\mu$  (Frango)
- Estimate for difference: 38,9380
- 95% CI for difference: (25,2887; 52,5873)
- T-Test of difference = 0 (vs not =): T-Value = 5,81  
P-Value = 0,000 DF = 32

# Teste – Independência entre Peso e Pressão Arterial



- Amostra de 100 pessoas.

Pressão arterial	Peso			Total
	Excesso	Normal	Deficiente	
Elevada	10	8	2	20
Normal	15	45	20	80
Total	25	53	22	100

Pressão arterial	Peso			Total
	Excesso	Normal	Deficiente	
Elevada	50%	40%	10%	100%
Normal	19%	56%	25%	100%
Total	25%	53%	22%	100%

# Teste – Independência entre Peso e Pressão Arterial



Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

	C1	C2	C3	Total
1	10	8	2	20
	5,00	10,60	4,40	
	5,000	0,638	1,309	
2	15	45	20	80
	20,00	42,40	17,60	
	1,250	0,159	0,327	
Total	25	53	22	100

Chi-Sq = 8,684; DF = 2; P-Value = 0,013

1 cells with expected counts less than 5.

# Tamanho de amostra – Comparação de médias



- Nível de significância =  $\alpha$
- Poder do teste =  $1-\beta$
- Diferença detectável entre as médias =  $\Delta$ , ou seja, se  $\mu_1-\mu_2=\Delta$ , quero que meu teste rejeite que são iguais.
- Supondo desvios iguais nos 2 grupos, quero o desvio padrão =  $\sigma$ .



# Tamanho de amostra

- Fórmula aproximada (melhor para n grande)

$$n = 2 \left[ \frac{(z_{\alpha/2} + z_{\beta})\sigma}{\Delta} \right]^2$$

# Referências



- Soares, J. F. e Siqueira, A. L. (2002).  
Introdução à estatística médica. UFMG.
- Bussab, W.O. e Morettin, P. A. (2002).  
Estatística Básica. Editora Saraiva.