

Pontifícia Universidade Católica do Rio Grande do Sul
Programa de Pós-Graduação em Biologia Celular e Molecular
Programa de Pós-Graduação em Pediatria e Saúde da Criança



Noções Básicas em Bioestatística

Prof Márcio Donadio
(mdonadio@pucrs.br)

Bioestatística

"Por serem mais precisos do que as palavras, os números são particularmente mais adequados para transmitir as conclusões científicas."

PAGANO e GAUVRE

No entanto tal como se pode mentir com palavras, pode-se fazer o mesmo com números.

"É fácil mentir com a estatística, mas é mais fácil mentir sem ela."



Bioestatística



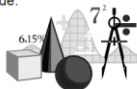
Bioestatística

• ESTATÍSTICA

É a ciência que tem por objetivo orientar a coleta, o resumo, a apresentação, a análise e a interpretação dos dados.

• BIOESTATÍSTICA

Aplicação da estatística nas ciências biológicas e da saúde.



População / Amostra

POPULAÇÃO: Universo. Conjunto de unidades experimentais que possuem pelo menos uma característica em comum.

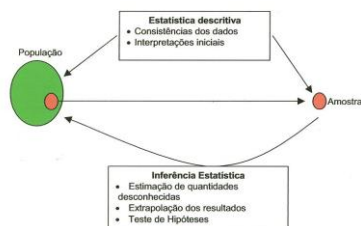
Amostra

É toda fração independente de seu tamanho obtida de uma população. São as pessoas ou unidades experimentais que são sorteados ou estão disponíveis para o estudo.



Bioestatística

Etapas da análise estatística.



Como se define o tamanho da amostra?

Cálculo do Tamanho Amostral

- Prévio e não retrospectivo
- Depende do desenho do estudo
- Necessita da opinião do especialista
- Deve basear-se em dados da literatura ou estudo piloto
- É uma estimativa

Cálculo do Tamanho Amostral

Porque calcular?

Viés - É qualquer tendência na coleta, análise, interpretação, publicação ou revisão de dados que leve a conclusões que sejam sistematicamente diferentes da verdade.

Acaso - Quando não é possível estudar toda a população, há chance de resultados ao acaso.

- O número não pode ser pequeno – acaso e falta de poder
- O número não pode ser grande demais – demanda tempo e recursos desnecessários - ética

Cálculo do Tamanho Amostral

O que é necessário para calcular?

- Significância

Nível de corte abaixo do qual rejeitamos a hipótese de nulidade

- Poder

Chance de detectar, como estatisticamente significante, um efeito especificado, se ele existir. Em geral, escolhemos um poder de, no mínimo, 80%

- Desvio padrão

- Menor diferença a ser detectada

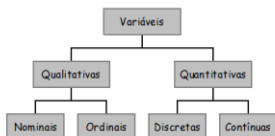
Cálculo do Tamanho Amostral

Exemplos



Variáveis

- Atributo mensurável que tipicamente varia entre indivíduos
- Tipos de variáveis



Variáveis

• Quantitativas:

– contínuas – espectro ordenado com intervalo quantificáveis em uma escala infinita de valores.

Peso, Altura

– discretas – possui uma escala com número finito de intervalos.

Peças produzidas, Pontos dado

• Qualitativas / categóricas:

– Nominal – categorias não ordenadas

- Duas categorias: dicotômica
- Três ou mais categorias: polinomial

Cor dos Olhos

– Ordinais – categorias ordenadas

Classe Social



Variáveis

Vantagens
da variável quantitativa

- Nível de informação é superior
- Pode ser transformada em qualquer outro tipo de variável
(Exemplo: peso ao nascer \Rightarrow baixo peso ao nascer)
- Aceita transformações matemáticas (log, raiz quadrada, inversão, etc.)
- Estudos com este tipo de variável necessitam tamanhos amostrais menores

Variáveis

Tabela 1. Caracterização da amostra do estudo.	
Variáveis	N=24
<i>Sexo</i>	
Masculino (%)	15 (62.5)
<i>Variáveis Antropométricas</i>	
Idade (anos)	10.5 \pm 1.53
Altura (cm)	138.0 \pm 0.08
Peso (Kg)	34.6 \pm 9.07
IMC Absoluto	17.92 \pm 2.84
Percentil	54.25 \pm 31.14
<i>Espirometria</i>	
VEF ₁ Absoluto	1.80 \pm 0.62
Percentual	93.29 \pm 29.02
CVF Absoluto	2.19 \pm 0.65
Percentual	103.78 \pm 26.12
FEF _{25-75%} Absoluto	1.94 \pm 1.07
Percentual	78.04 \pm 39.63
<i>Manovacuometria</i>	
PIMAX (cmH ₂ O)	92.1 \pm 22.8
PEMAX (cmH ₂ O)	98.9 \pm 24.5

Variáveis

Tabela 1 – Características Clínicas e Epidemiológicas dos Pacientes		
Variáveis	Média \pm DP	n = 100 (%)
<i>Sexo</i>		
Feminino		53 (53)
Masculino		47 (47)
Idade (anos)	60 \pm 18,2	
Idade < 65 anos		56 (56)
Idade \geq 65 anos		44 (44)
<i>Tipo de paciente</i>		
Clinico		48 (48)
Cirúrgico		52 (52)
<i>Sítio da infecção</i>		
Pulmão		28 (28)
Trato urinário		11 (11)
Intra-abdominal		48 (48)
Pele e subcutâneo		09 (9)
Corrente sanguínea		02 (2)
Outros		02 (2)
Infecção nosocomial		33 (33)

Estatística Descritiva

Frequências:

- Frequência absoluta: número de ocorrências
- Frequência relativa: percentual de ocorrências



2
7
4
2
9
7
2
10
2

Categoria
"2"?

FA: 4,0
FR: 44,4%

Estatística Descritiva

Medidas de Tendência Central:

- média aritmética

- mediana

- moda



2 2 2 2
7 2 2 2
4 2 2 2
2 2 2 2
9 4 4 4
7 7 7 7
2 7 7 9
10 9 9 10
2 10 10

Média: 5,0
Mediana: 4,0
Moda: 2,0

Estatística Descritiva

Medidas de Tendência Central:

- A média aritmética é definida como a soma de todas as observações da variável X, dividida pelo número de elementos do conjunto de dados. Frequentemente a média aritmética é o valor que melhor representa um conjunto de dados.

- A mediana é o valor que divide o conjunto de dados ordenado em duas partes com igual número de observações.

- A moda é definida como o valor mais frequente de um conjunto de dados. É possível que o conjunto seja bimodal (duas modas) ou até mesmo multimodal (três ou mais modas).

Estatística Descritiva

Medidas de Variabilidade:

- Variância: a soma dos quadrados da diferença entre cada valor e a média aritmética, dividida pelo número de ocorrências.
- Desvio padrão: define-se como a raiz quadrada da variância.
- Erro padrão: obtém-se dividindo o desvio padrão pela raiz quadrada do tamanho da amostra.

Estatística Descritiva

Medidas de Variabilidade:

- Variância
- Desvio padrão
- Erro padrão



2
7
4
2
9
7
2
10
2

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Valid N (listwise)	9	8,00	2,00	10,00	5,0000	1,09291	3,27672

Estatística Descritiva

Funcionários	Quantidade de peças produzidas por dia				
	Segunda	Terça	Quarta	Quinta	Sexta
A	10	9	11	12	8
B	15	12	16	10	11
C	11	10	8	11	12
D	8	12	15	9	11



$$\text{var} = \frac{(segunda - \text{média aritmética})^2 + (terça - \text{média aritmética})^2 + \dots + (sexta - \text{média aritmética})^2}{\text{quantidade de dias}}$$

Funcionários	Média Aritmética (\bar{x})
A	$\bar{x}_A = \frac{10 + 9 + 11 + 12 + 8}{5} = \frac{50}{5}$ $\bar{x}_A = 10,0$

Variância → Funcionário A:

$$\text{var}(A) = \frac{(10 - 10)^2 + (9 - 10)^2 + (11 - 10)^2 + (12 - 10)^2 + (8 - 10)^2}{5}$$

$$\text{var}(A) = \frac{10}{5} = 2,0$$

Desvio Padrão → Funcionário A:

$$\text{dp}(A) = \sqrt{\text{var}(A)}$$

$$\text{dp}(A) = \sqrt{2,0}$$

$$\text{dp}(A) = 1,41$$

Estatística Descritiva

Intervalo interquartil

Indica a dispersão de 50% da distribuição

Quartil Inferior \longrightarrow IQ 25%
 Quartil Superior \longrightarrow IQ 75% } 50%

Estatística Descritiva

IC95%

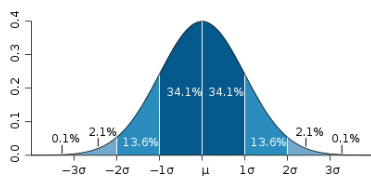
5,3 (IC95% 3,0-7,4)

Intervalo de confiança: *podemos afirmar que em 95 de 100 amostras hipotéticas, o resultado estará dentro deste intervalo.*

Normalidade

Paramétricos ou Não-paramétricos

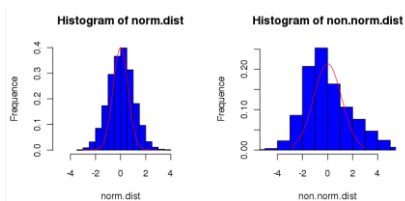
Distribuição Gaussiana



Normalidade

Paramétricos ou Não-paramétricos

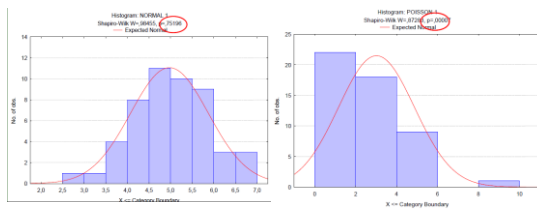
Distribuição Gaussiana



Normalidade

Testes: Shapiro-Wilk

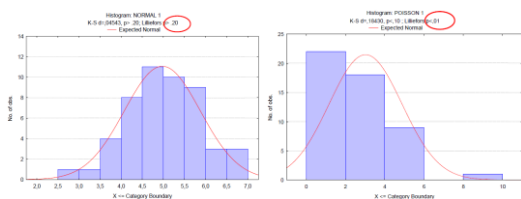
- normal quando $p > 0.05$
- amostras menores ou iguais a 30



Normalidade

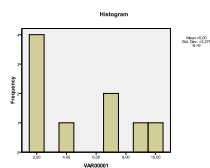
Testes: Kolmogorov-Smirnov

- normal quando $p > 0.05$
- amostras maiores do que 30



Normalidade

Testes: Shapiro-Wilk ou Kolmogorov-Smirnov?



	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
VAR00001	,264	9	,070	,829	9	,044

2
7
4
2
9
7
2
10
2

Estatística Inferencial



População / Amostra

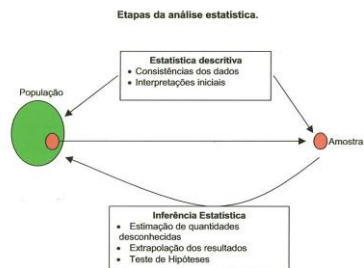
POPULAÇÃO: Universo. Conjunto de unidades experimentais que possuem pelo menos uma característica em comum.

Amostra

É toda fração independente de seu tamanho obtida de uma população. São as pessoas ou unidades experimentais que são sorteados ou estão disponíveis para o estudo.



Bioestatística



Estatística Inferencial

- Teste de Hipóteses

A verdade na População	Resultado de teste de hipótese	
	Acerta-se H_0 (Não existência de diferenças)	Rejeita-se H_0 (Existência de diferenças)
H_0 Verdadeira (não existência de diferenças)	Acerte-se correctamente	Erro tipo I (α)
H_0 Falsa (Existência de diferenças)	Erro tipo II (β)	Rejeita-se correctamente

Estatística Inferencial

- Teste de Hipóteses

O que significa o * ?

O que é o "p" ?

Estatística Inferencial

$$p < 0,05$$

- É um número estabelecido para evidenciar que existe diferença significativa entre grupos
- Chance dessa diferença ser ao acaso é $< 5\%$

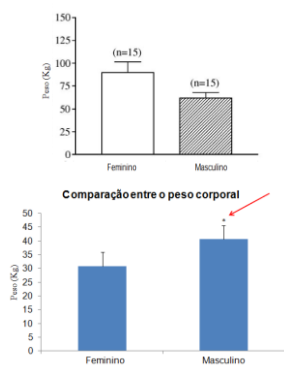
"PROBABILIDADE"

Estatística Inferencial

$$p < 0,05$$

- $p = 0,10$ 90% resultado ser verdadeiro
- $p = 0,05$ 95% resultado ser verdadeiro
- $p = 0,01$ 99% resultado ser verdadeiro

Estatística Inferencial



Estatística Inferencial

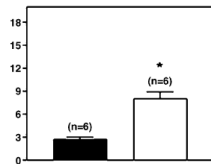
Tabela 2 - Comparação entre as características antropométricas e nutricionais entre o grupo controle e RNMBP.

Variáveis avaliadas	Controles N= 45	RNMBP N= 48	p
Idade (anos)	10,23 ± 1,27	10,18 ± 1,39	0,860
Altura (cm)	141,72 ± 10,29	138,53 ± 11,29	0,159
Altura/idade, escore-z	-0,10 ± 1,08	0,13 ± 1,22	0,323
Peso (kg)	37,64 ± 9,95	34,66 ± 10,36	0,161
Peso/idade, escore-z	-0,03 ± 0,89	0,27 ± 1,02	0,123
IMC, absoluto	18,49 ± 3,42	17,71 ± 3,32	0,260
IMC/idade, escore-z	-0,38 ± 1,15	-0,30 ± 1,27	0,740

Resultados apresentados em média ± desvio padrão. RNMBP= recém-nascidos de muito baixo peso.

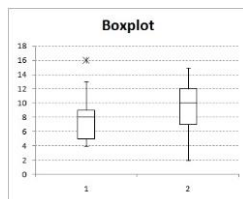
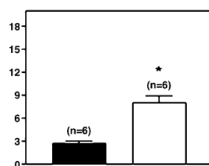
Estatística Inferencial

- Desenho Experimental
 - ✓ Ratos diabéticos e não-diabéticos
 - ✓ Medida de glicose no sangue

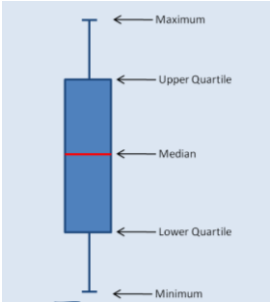


Estatística Inferencial

- Análise
 - ✓ Paramétrico - teste t de Student
 - ✓ Não-paramétrico - teste de Mann-Whitney

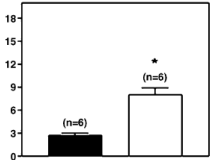


Estatística Inferencial



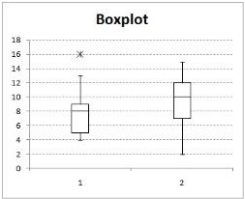
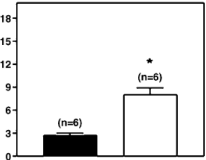
Estatística Inferencial

- Desenho Experimental
 - ✓ Pacientes diabéticos
 - ✓ Medida de glicose no sangue - intervalo de 30 dias



Estatística Inferencial

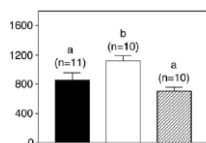
- Análise
 - ✓ Paramétrico - teste t de Student pareado
 - ✓ Não-paramétrico - teste de Wilcoxon



Estatística Inferencial

• Desenho Experimental

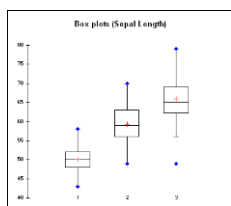
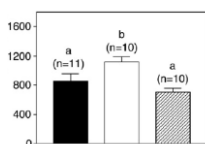
- ✓ Ratos não-diabéticos, diabéticos e diabéticos tratados
- ✓ Medida de glicose no sangue



Estatística Inferencial

• Análise

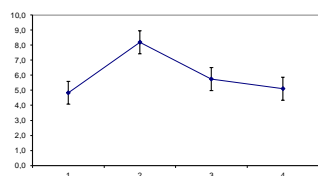
- ✓ Paramétrico - Anova de 1 via
- ✓ Não-paramétrico - teste de Kruskal-Wallis



Estatística Inferencial

• Desenho Experimental

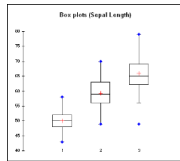
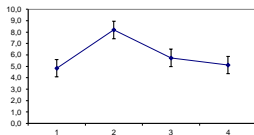
- ✓ Ratos diabéticos
- ✓ Medida de glicose no sangue em 0, 15, 30 e 60 min



Estatística Inferencial

• Análise

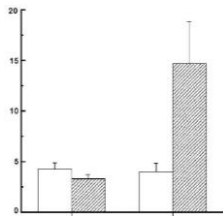
- ✓ Paramétrico - Anova de 1 via para medidas repetidas
- ✓ Não-paramétrico - teste de Friedman



Estatística Inferencial

• Desenho Experimental

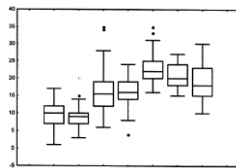
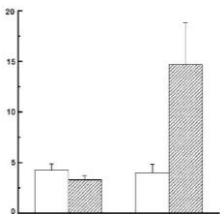
- ✓ Ratos não-diabéticos e diabéticos - efeito do tratamento
- ✓ Medida de glicose no sangue



Estatística Inferencial

• Análise

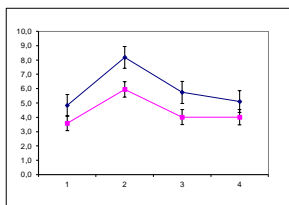
- ✓ Paramétrico - Anova de 2 vias
- ✓ Não-paramétrico - teste de Scheirer-Ray-Hare



Estatística Inferencial

• Desenho Experimental

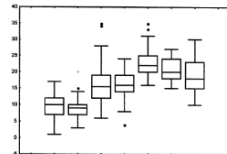
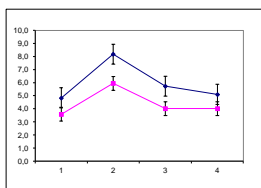
- ✓ Ratos não-diabéticos e diabéticos
- ✓ Medida de glicose no sangue em 0, 15, 30 e 60 min



Estatística Inferencial

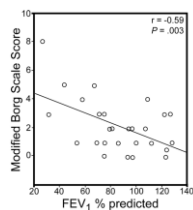
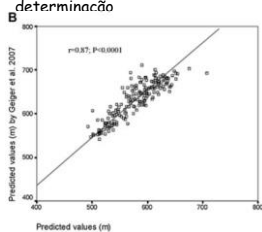
• Análise

- ✓ Paramétrico - Anova de 2 vias com medidas repetidas
- ✓ Não-paramétrico - teste Q de Cochran



Correlações

- Pearson ou Spearman
- “r” - coeficiente de correlação / “r²” - coeficiente de determinação



Regressões Lineares

- Estimativa do comportamento de variáveis
- Bastante utilizada para valores de referência
- Geração de equações de predição

$$\begin{aligned}\text{Distance walked(m)} &= 145.343 + [11.78 \times \text{age(years)}] \\ &+ [292.22 \times \text{height(m)}] + [0.611 \times \text{dif.HR(bpm)}] \\ &- [2.684 \times \text{body weight(kg)}] \\ \text{Standard error of estimate} &= 54.81 \text{ m/R}^2 = 36.6\%\end{aligned}$$
