

M109

Estatística

Aula 8 – Testes de Normalidade

Prof^a. Rosimara Beatriz Arci Salgado

Créditos:

Prof^a. Karina Perez Mokarzel Carneiro

Prof. Rausley Adriano Amaral de Souza

Inatel

**CAMINHOS
QUE CONECTAM
COM O FUTURO**



Bibliografia básica:

Arango HG. Bioestatística: teórica e computacional. 3ªed. Rio de Janeiro: Guanabara Koogan; 2011.

SPIEGEL, Murray Ralph; FARIA, Alfredo Alves De Probabilidade e estatística. São Paulo, SP: McGraw-Hill, 1978.

TESTES PARA DADOS CATEGORIZADOS



→ Amostras independentes

→ $n > 40$ → Teste χ^2 Clássico.

→ $20 < n < 40$ → Teste χ^2 com Correção de Yates.

→ $n < 20$ → Teste Exato de Fisher

→ Amostras pareadas

→ Teste McNemar

TESTES PARA DADOS AMOSTRAIS



Testes de Normalidade:

TESTES PARAMÉTRICOS → Teste de Student (t)

- **Amostras Independentes** → Populações Homocedásticas e Populações Heterocedásticas
- **Amostras Pareadas**



Teste de Fisher

TESTES NÃO PARAMÉTRICOS

- **Amostras pareadas** → Testes dos sinais
OBS: Também utilizado na comparação dos resultados de uma amostra com a mediana de uma população e na comparação de dados qualitativos
- **Amostras independentes** → Teste de wilcoxon-mann-whitney

TESTE DE NORMALIDADE



→ **TESTE DE KOLMOGOROV-SMIRNOV (K-S)**

Média e desvio padrão populacional conhecidos.

→ **TESTE LILLIEFORS**

Média e desvio padrão estimados.

→ **TESTE DE SHAPIRO-WILKS**





→ Conceito e finalidade

Os testes baseados na comparação de valores (variáveis quantitativas) podem ser divididos em dois grandes grupos:

- Testes Paramétricos
- Testes Não-paramétricos

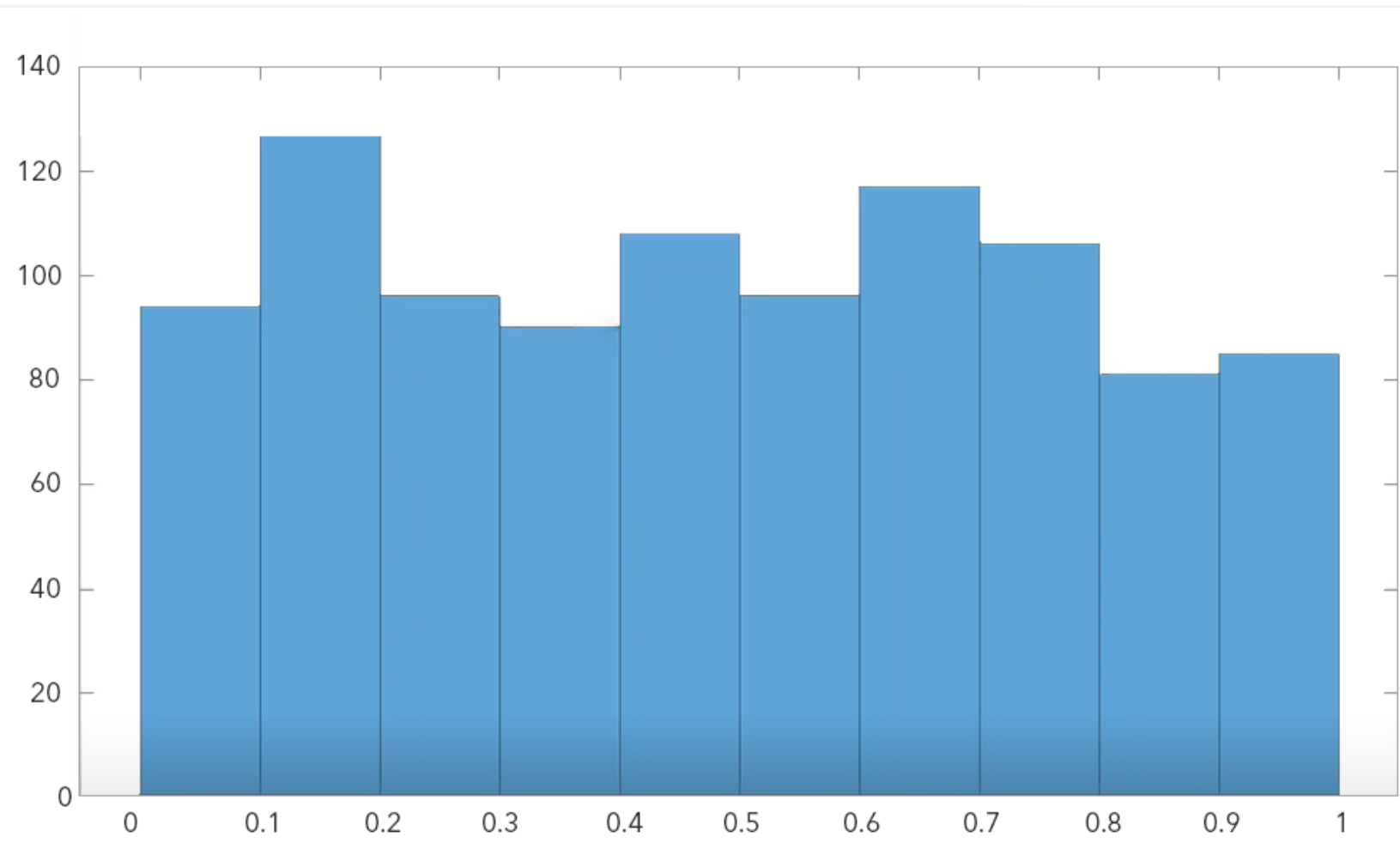
- Os **testes paramétricos** baseiam-se na hipótese de que as variáveis que representam os grupos, que estão sendo comparados, apresentam uma distribuição mais ou menos bem-comportada.
- Um exemplo de distribuição bem-comportada seria a **distribuição Normal**.
- Os testes paramétricos baseiam sua decisão na comparação de parâmetros: média e desvio padrão.



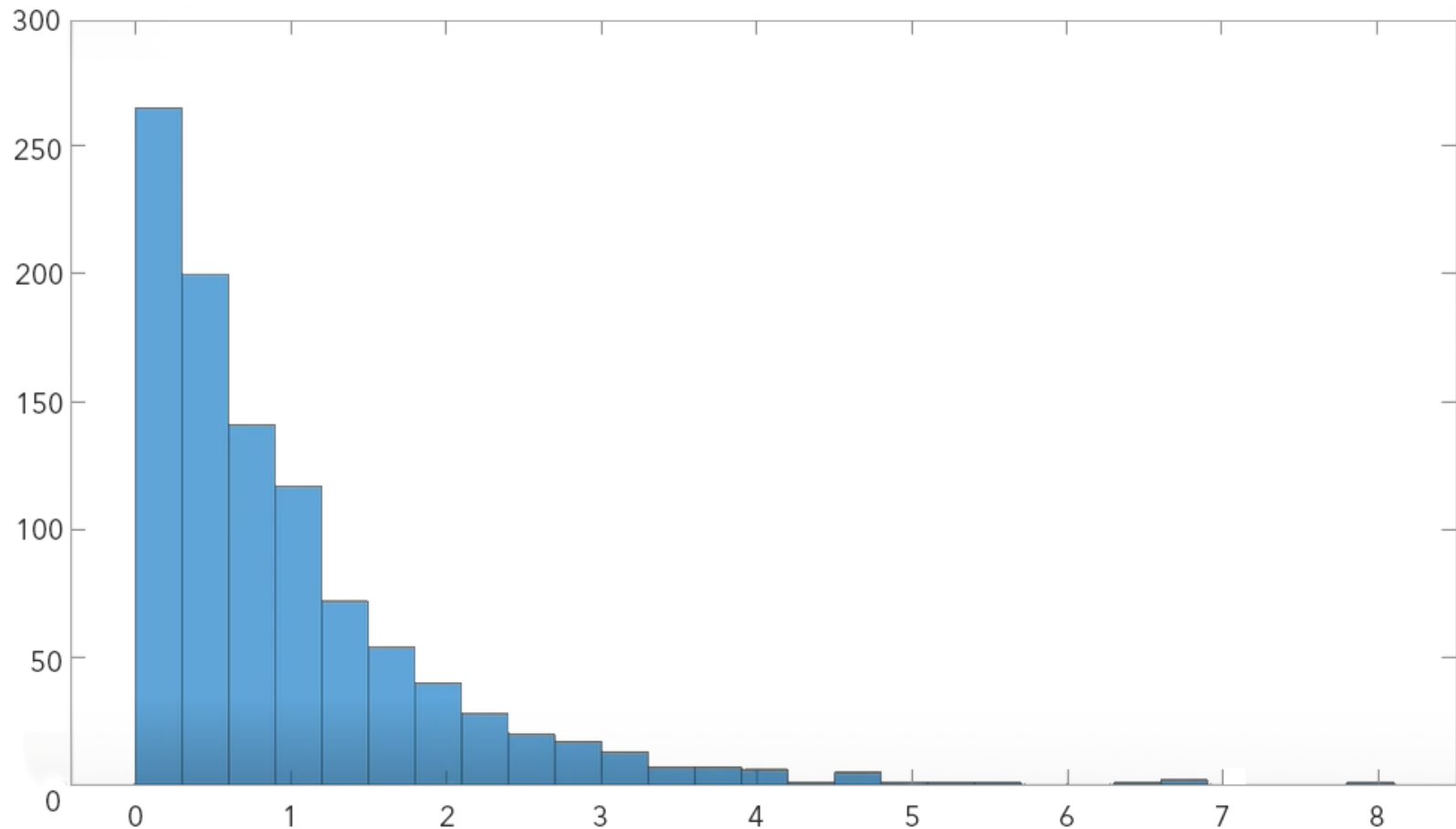


- A filosofia dos **testes não-paramétricos** é justamente evitar a utilização de parâmetros na comparação de grupos ou situações. Dessa forma, pode-se dizer que eles funcionam onde os testes paramétricos falham.
- Assim, antes de decidir entre os testes paramétricos ou não-paramétricos deve-se investigar a distribuição das variáveis que estão sendo comparadas.
- A prova de que uma variável possui determinada função de probabilidade é denominada **prova de aderência**.
- Quando a distribuição é a Normal, a prova da variável é denominada: **Teste de Normalidade**.

O que você pode dizer a respeito desse histograma referente a um conjunto de dados com 1000 valores (amostras)?

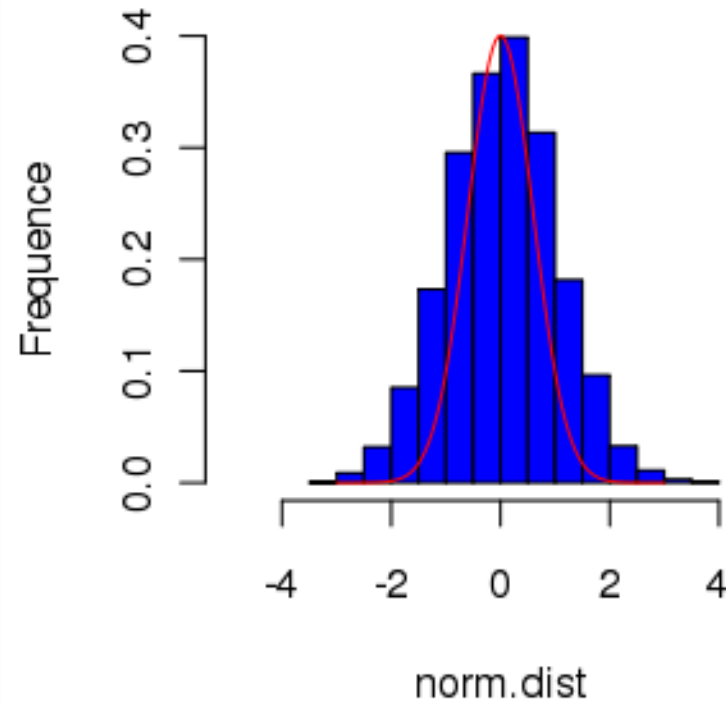


O que você pode dizer a respeito desse histograma, referente aos seus dados, visualmente falando?

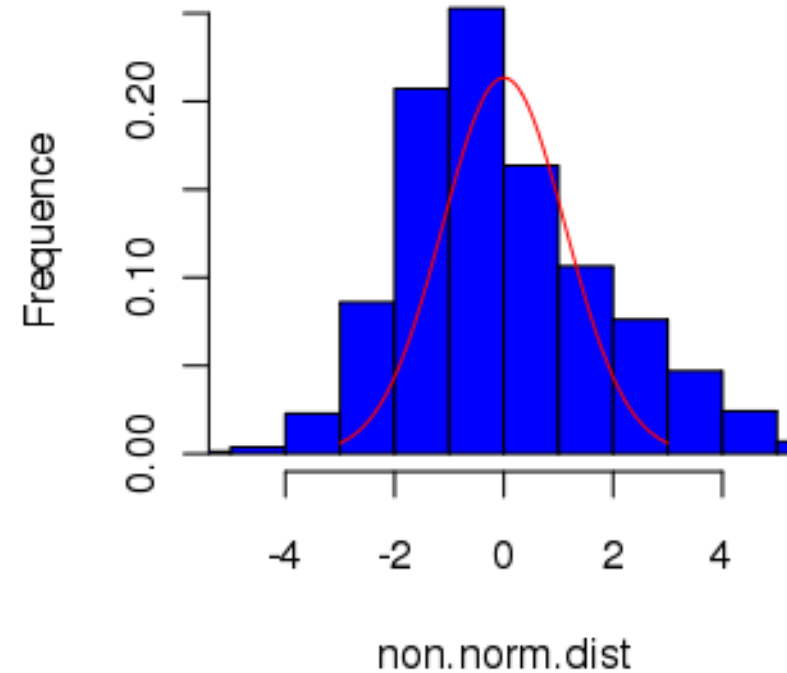


TESTE DE NORMALIDADE

Histogram of norm.dist



Histogram of non.norm.dist



https://fr.wikipedia.org/wiki/Fichier:Normality_histogram.png





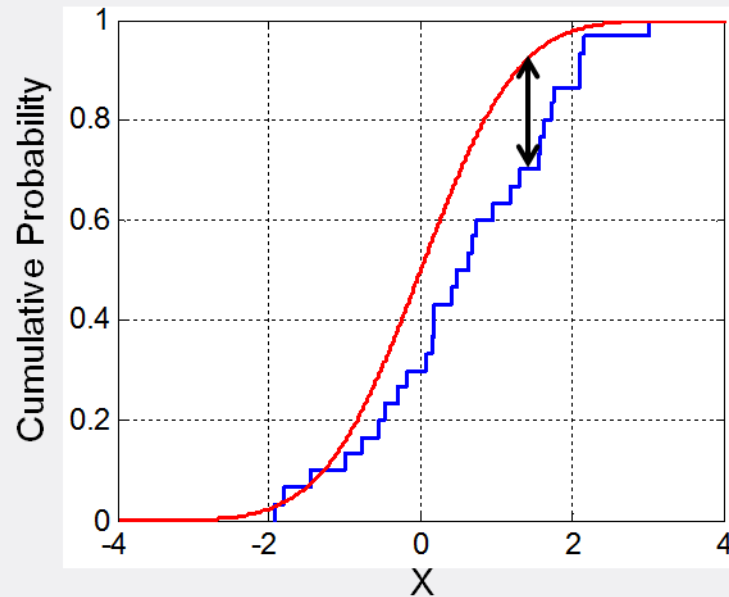
Dois testes bastante empregados para provar (ou rejeitar) a normalidade de uma distribuição são:

- O teste de Kolmogorov-Smirnov, ou teste K-S (tem uma variante deste teste chamado Lilliefors).
- O teste de Shapiro-Wilks, ou teste S-W.

Utiliza os parâmetros da população

Utiliza os parâmetros amostrais.

TESTE KS



$$D = \sup_x |\hat{F}_X(x) - F_X(x)|$$

em que $\hat{F}_X(\cdot)$ e $F_X(\cdot)$ são as CDFs das distribuições empíricas e teóricas

https://pt.wikipedia.org/wiki/Teste_Kolmogorov-Smirnov#/media/Ficheiro:KS_Example.png





H_0 : as amostras apresentam normalidade em suas distribuições

H_1 : as amostras não apresentam normalidade em suas distribuições

Ou

H_0 : a amostra provém de uma população Normal

H_1 : a amostra não provém de uma população Normal

Na prática, se for provado que o valor do teste K-S ou do teste S-W é significativo (**$p < 0,01$**), a hipótese de normalidade da distribuição deve ser rejeitada.

Exemplo: Peso, altura e idade

BioEstat 5.0 - [Dados 1]

Arquivo Editar Estatísticas Gráficos Sugestões Configurar Ajuda

Escolha um teste

1.75

	- 1 - Peso	- 2 - Altura	- 3 - Idade	- 4 -	- 5 -	- 6 -	- 7 -	- 8 -
1	56	1.65	17					
2	63	1.53	18					
3	54	1.62	18					
4	60	1.54	19					
5	48	1.60	19					
6	75	1.70	19					
7	74	1.91	19					
8	74	1.70	20					
9	62	1.63	20					
10	50	1.68	20					
11	65	1.76	20					
12	67	1.80	21					
13	57	1.68	24					
14	60	1.58	16					
15	58	1.66	17					
16	69	1.72	18					
17	48	1.56	18					
18	46	1.63	18					
19	67	1.64	19					
20	52	1.60	19					
21	67	1.80	19					
22	50	1.56	19					
23	60	1.66	20					
24	47	1.68	20					
25	48	1.58	21					
26	43	1.56	22					
27	55	1.71	23					

Arquivo Editar Estatísticas Gráficos Sugestões Configurar Ajuda

Último Teste Ctrl+U ha um teste

Amostragem

Análise Multivariada

Análise de Sobrevivência

Análise da Variância

Bootstrap - Reamostragem

Correlação

Estatísticas Circulares

Distribuição de Probabilidades

Uma Amostra

Duas Amostras Independentes

Duas Amostras Relacionadas

Estatística Descritiva

Estimação de Parâmetros

Meta-Análise

Normalidade

Qui-Quadrado

Regressão

Tamanho das Amostras

Transformação e Ordenação

D'Agostino (k amostras)

D'Agostino-Pearson (k amostras)

Kolmogorov-Smirnov

Lilliefors (k amostras)

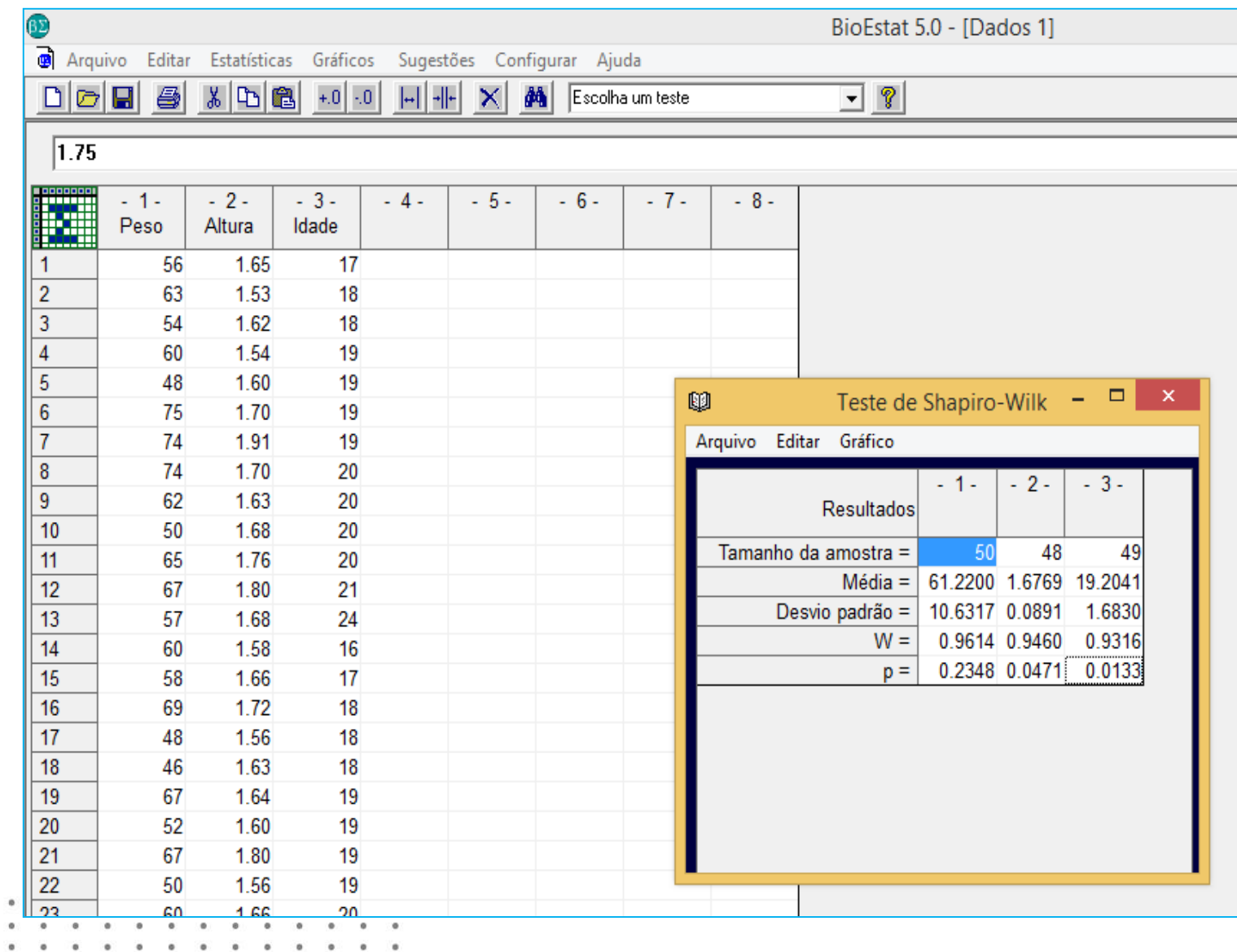
Shapiro-Wilk (k amostras)

Valores extremos (base nos desvios)

Valores extremos (base nos quartis)

	- 1 -	
	Peso	
1	56	
2	63	
3	54	
4	60	
5	48	
6	75	
7	74	
8	74	
9	62	
10	50	
11	65	
12	67	
13	57	
14	60	
15	58	
16	69	
17	48	
18	46	
19	67	1.64 19
20	52	1.60 19
21	67	1.80 19

TESTE SHAPIRO-WILK



CONCLUSÃO

Peso:

Aceita $H_0 \rightarrow$ aceita a normalidade

Altura:

Rejeita $H_0 \rightarrow$ rejeita a normalidade

Idade:

Rejeita $H_0 \rightarrow$ rejeita a normalidade

Ao nível de significância de 5%



EXERCÍCIO 1

Você e o seu sócio tem um provedor de acesso à Internet. Para auxiliar na configuração de equipamentos, o conhecimento do comportamento dos tempos de acesso de seus clientes é importante, mas vocês têm divergência a esse respeito. Você crê que o tempo de acesso segue uma distribuição Normal com média de 16 minutos e desvio padrão de 4 minutos, enquanto seu sócio, suspeita que o tempo de acesso segue uma distribuição Exponencial com média de 16 minutos. Para diminuir a dúvida, vocês coletaram uma amostra de tempos de acesso em minutos. Carregue o arquivo **Aula8_E1_Aluno_L1.csv** no BioEstat, faça o teste K-S e conclua se a distribuição é normal ou não.



EXERCÍCIO 2

Carregue o arquivo **Aula8_E2_Aluno_L1.csv** no BioEstat e verifique se as variáveis **Peso**, **Altura** e **Idade** têm aderência a uma distribuição normal. Aplique o teste S-W e conclua se as amostras apresentam uma distribuição normal ou não.



Rosimara Salgado

Professora
Coordenadora do NEaD

rosimara@inatel.br



Inatel

CAMINHOS
QUE CONECTAM
COM O FUTURO