

## O Processo Estatístico: definir o teste a aplicar\*

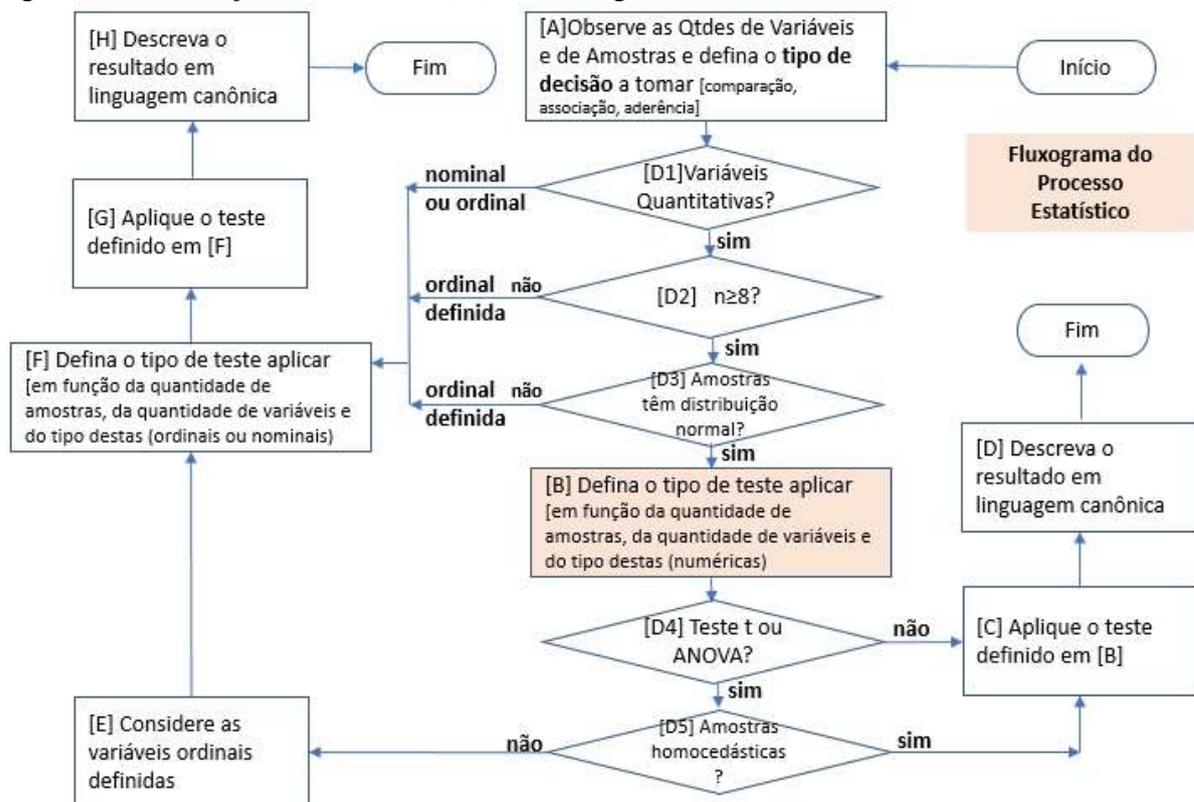
*Manuel Meireles*  
UNIFACCAMP  
[meireles@faccamp.br](mailto:meireles@faccamp.br)

Este trabalho é continuação do publicado na edição anterior e tem por objetivo abordar a ação [B] do Fluxograma do Processo Estatístico como mostra a Figura 1.

A análise estatística (de qualquer problema estatístico) requer algumas etapas:  
1-reconhecer o tipo de decisão a fazer;  
2-reconhecer os tipos de variáveis envolvidas;  
3-definir o teste a aplicar;  
4-aplicar o teste;  
5-interpretar o resultado pelo p-value;  
6-redigir a análise em linguagem canônica, isto é, em linguagem precisa dentro do estilo de linguagem da comunicação estatística.

Este trabalho ocupa-se da atividade [B] que é definir o tipo de teste a aplicar em função de quantidades de variáveis e do tipo de tais variáveis. Notar que os passos [A], e <D1> , <D2> e <d3> foram objeto de estudo nos números imediatamente anteriores.

Figura 1: Localização da atividade [B] no Fluxograma do Processo Estatístico



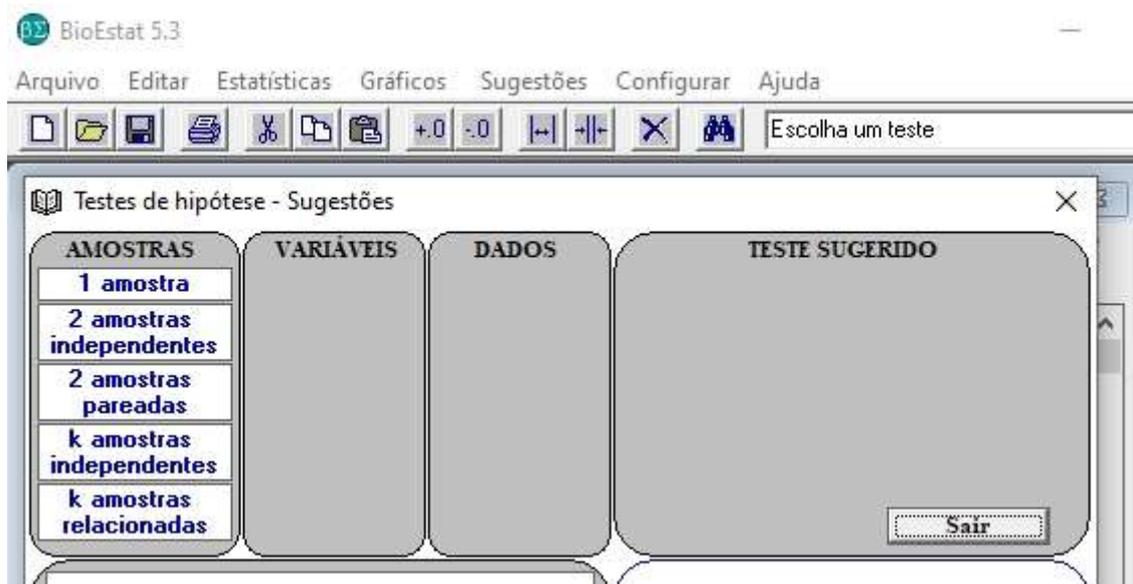
\* Recebido em 05 de setembro de 2022, aprovado em 05 de setembro de 2022, publicado em 05 de setembro de 2022.

Antes de executar a atividade [B] na qual se define o tipo de decisão a tomar (se aderência, comparação ou associação) foi necessário tomar as decisões <D1>, <D2> e <D3> como se pode ver na Figura 1. A decisão <D1> analisa se os dados originais são quantitativos ou não: caso sejam avança-se para a decisão <D2> ; caso pelo menos uma das variáveis não seja quantitativa é porque ele é nominal ou ordinal e se passa à atividade [F]. As decisões <B2> e <D3> foram discutidas no trabalho anterior. Para responder à questão da <D3> é necessário aplicar um teste de normalidade. Tal teste confirma (Ou não conforma) se as variáveis qualitativas não diferem significativamente da distribuição normal ao nível de significância de 0.05. Notar que nível de significância de 0.05 (ou 5%) equivale a nível de confiança de 0.95 (ou 95%).

Em resumo: se uma ou mais variáveis, no teste de aderência à normalidade for significativamente diferente da distribuição normal se passa para a ação [F] que consiste em definir o tipo de teste aplicar [em função da quantidade de amostras, da quantidade de variáveis e do tipo destas (ordinais definidas)]; caso todas as variáveis não difiram significativamente (ns) da distribuição normal se passa para a ação [B] para definir o tipo de teste aplicar [em função da quantidade de amostras, da quantidade de variáveis e do tipo destas (quantitativas ou numéricas)]. A definição do teste a aplicar é o que veremos agora.

Uma boa forma de definir o teste estatístico a aplicar é utilizando o software BioEstat e a sua função <Sugestões>. Clicando-se nesta função abre-se a janela “Teste de hipótese – Sugestões” como mostra a Figura 2.

Figura 2: Janela “Sugestões” no BioEstat



Como se pode ver o teste a sugerir depende das características do problema em análise: da quantidade de amostras, da quantidade de variáveis e do tipo de dados. Tomemos um exemplo retirado de Meireles (2022):

**Caso 2** → A microempresa Artefatos de Cimento tem duas máquinas para fabricar blocos de cimento, designadas aqui por MaqA e MaqB. A Tabela 2.2 mostra os dados referentes às produções de blocos nos últimos 10 dias úteis. Pode-se afirmar que a produção de blocos da MaqA é **significativamente** maior do que a produção da MaqB? (Dados, Caso 2)

Tabela 2.2: Dados para o Caso 2

CASO 2	
MaqA	MaqB
127	125
125	124
127	125
126	120
127	119
129	121
127	123
129	119
131	122
129	121

Quantas e quais são as amostras do caso 2?

Quantas e quais são as variáveis do caso 2?

Que tipo de decisão deve ser tomada no caso 2?

Tendo em conta o Processo Estatístico começamos pela atividade [A] (ver Figura 1): [A]Observe as Quantidades de Variáveis e de Amostras e defina o tipo de decisão a tomar [comparação, associação, aderência]

Trata-se de uma **comparação** (comparar a produção da MaqA com a da MaqB) envolvendo duas amostras emparelhadas (MaqA e MaqB) sendo duas variáveis (quantidade de blocos produzidos por dia pela MaqA e quantidade de blocos produzidos por dia pela MaqB).

Continuando o Processo Estatístico temos a decisão <D1>: Variáveis são quantitativas? A Resposta é “Sim” pelo que se passa à etapa seguinte:  
Decisão <D2>:  $n \geq 8$ ? A resposta é sim, já que  $n$  de cada amostra=10. Passa-se à Decisão <D3>: Amostras possuem distribuição normal?

Para responder a esta questão é necessário proceder a um teste estatístico para testar a normalidade dos dados das amostras. Pode-se aplicar o teste D’Agostino como mostra a Figura 3.

Figura 3: Teste de normalidade D’Agostino para as amostras MaqA e MaqB

- 1 - MaqA	- 2 - MaqB	D'Agostino		
		Arquivo	Editar	Gráfico
		Resultados	Coluna 1	Coluna 2
127	125	Tamanho da amostra =	10	10
125	124	D (Desvio) =	0.2714	0.2840
127	125	Valores críticos 5%	0.2513 a 0.2849	0.2513 a 0.2849
126	120	Valores críticos 1%	0.2379 a 0.2857	0.2379 a 0.2857
127	119	p =	ns	ns
129	121			
127	123			
129	119			
131	122			
129	121			

O resultado do teste mostra que as amostras MaqA e MaqB têm distribuições que não diferem significativamente (ns) da distribuição normal ao nível de significância de 0.05 (Teste

D'Agostini, p-value não significativo). Isto quer dizer que os dados, originalmente numéricos possuem as características de dados numérico considerando suas distribuições.

Desta forma

Figura 4: Testes sugeridos para o caso em análise.

**AMOSTRAS**  
2 amostras pareadas

**VARIÁVEIS**  
1 variável

**DADOS**  
Numéricos

**TESTE SUGERIDO**  
Teste t  
ANOVA

<< Dados      Sair

**\* TESTE - t \***  
Teste paramétrico destinado a analisar a diferença estatística entre duas (2) amostras dependentes e do mesmo tamanho, onde cada indivíduo é seu próprio controle (antes e depois). Os dados devem ser medidos a nível intervalar ou de razões. O teste destina-se à pequenas amostras ( $N < 31$ ), mas pode ser empregado para grandes amostras.

$$\bar{D} = \sum D/N$$

$$s_D = \sqrt{\frac{\sum (D - \bar{D})^2}{N-1}}$$

$$s_{\bar{D}} = s_D / \sqrt{N}$$

$$t = \bar{D} / s_{\bar{D}}$$

$$gl = N-1$$

O resultado da função <Sugestões> sugerem os testes t ou ANOVA. Eis como isso ocorreu:

Figura 5: Janela que se abre após selecionar <Sugestões>

**AMOSTRAS**  
1 amostra  
2 amostras independentes  
2 amostras pareadas  
k amostras independentes  
k amostras relacionadas

**VARIÁVEIS**

**DADOS**

**TESTE SUGERIDO**

Sair

Quando se seleciona <Sugestões> abre-se uma janela que requer que se indique a Quantidade de amostras e se elas são independentes ou não. No presente caso seleciona-se “2 amostras pareadas” como mostra a Figura 5. Selecionado “2 amostras pareadas” surge a indagação da quantidade de variáveis (Figura 6). No presente caso a resposta é “1 variável”. Selecionada a quantidade de variáveis surge a necessidade, como mostra a Figura 7, de se selecionar o tipo de dados, que são “Numéricos” neste caso. Uma vez selecionado “Numéricos” aparece uma lista dos testes passíveis de serem aplicados como mostra a Figura 8. Teste t e ANOVA.

Passando-se o mouse sobre cada um dos testes sugeridos é possível ver as características específicas de cada teste e sua aplicabilidade.

Figura 6: Definida a quantidade de amostras surge “Variáveis”



Figura 7: Definida a quantidade de variáveis surge o tipo de “Dados”



Figura 8: Definido o tipo de dados surge a lista de testes sugeridos



Este resultado é interessante pois a decisão <Da> requer que se responda se o teste a aplicar é “Teste t ou ANOVA” como mostra a Figura 1. Veremos isso no próximo trabalho.

## Referências

Ayres, M., Ayres Jr, M. (2007) BioEstat- Manuel. Belém, Pará.

BioEstat <https://www.mamiraua.org.br/downloads/programas/>

Meireles, M. (2022). Processo Estatístico. Guarujá: Lifetools.