

O Processo Estatístico: amostras homocedásticas?*

The Statistical Process: homoscedastic samples?

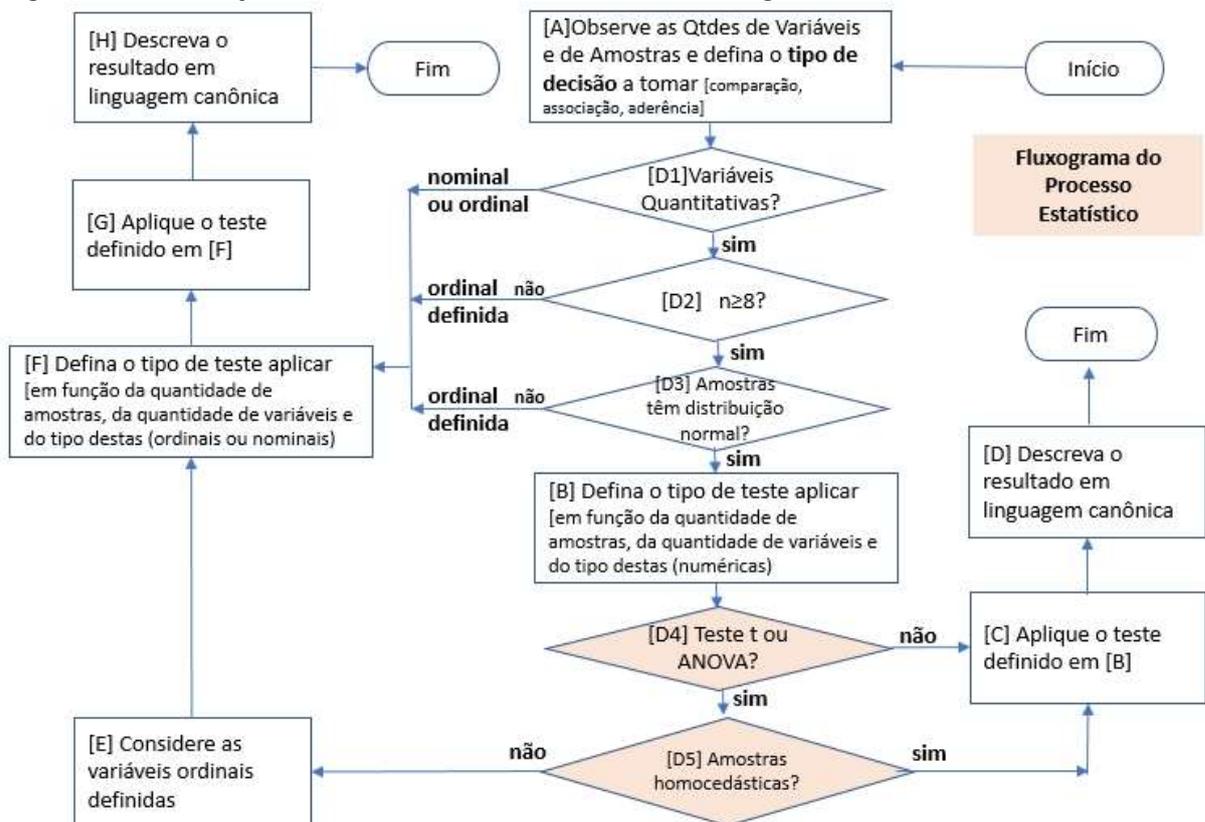
Manuel Meireles
UNIFACCAMP
meireles@faccamp.br

Este trabalho é continuação do publicado na edição anterior e tem por objetivo abordar as decisões <D4> e <D5> do Fluxograma do Processo Estatístico como mostra a Figura 1.

Temos visto que 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; e 6-redigir a análise em linguagem canônica, isto é, em linguagem precisa dentro do estilo de linguagem da comunicação estatística.

Este trabalho ocupa-se das decisões <D4> e <D5> que são subsequentes à atividade [B].

Figura 1: Localização das decisões <D4> e <D5> no Fluxograma do Processo Estatístico



Retomemos ao exemplo retirado de Meireles (2022) visto no número anterior:

* Recebido em 17 de novembro de 2022, aprovado em 17 de novembro de 2022, publicado em 30 de novembro de 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?

Seguindo-se o Processo Estatístico verificamos que a atividade [B] deu como resultado os testes aplicáveis (Ver Figura 2).

Figura 2: Detalhe do Processo Estatístico objeto do presente trabalho



Os testes sugeridos para o presente caso são o Teste t ou ANOVA, como se pode ver pela Figura 3 (que no trabalho anterior é a Figura 8).

Figura 3: Testes sugeridos pela atividade [B] do Processo Estatístico

Testes de hipótese - Sugestões

AMOSTRAS	VARIÁVEIS	DADOS	TESTE SUGERIDO
2 amostras pareadas	1 variável	Numéricos	Teste t ANOVA
			<< Dados Sair

A decisão <D4> é necessária para atender aos pressupostos dos testes paramétricos t e ANOVA que, de acordo com Cooper e Schindler (2003) devem ser atendidos: 1. As observações devem ser independentes; 2. As observações devem ser retiradas de populações normalmente distribuídas; 3. Essas populações devem ter variâncias iguais; 4. As escalas de mensuração devem estar no menor intervalo para que se possam efetuar operações aritméticas com as mesmas.

No presente caso a resposta decisão <D4> é “sim” pelo que se precisa tomar a decisão <D5> logo a seguir: as amostras são homocedásticas?

Duas amostras são ditas homocedásticas se as suas variâncias não diferirem significativamente uma da outra a um determinado nível de significância. Então é preciso tomar a decisão <D5>: as amostras são ou não homocedásticas?

Para melhor ilustrar este caso vamos considerar a existência de uma outra máquina designada MaqBx cujos dados estão na Figura 4. Se fizermos as estatísticas descritivas dessas três amostras de dados podemos observar que a variância desta última é bem superior às duas originais como mostra a Figura 5.

Figura 4: Dados das três amostras

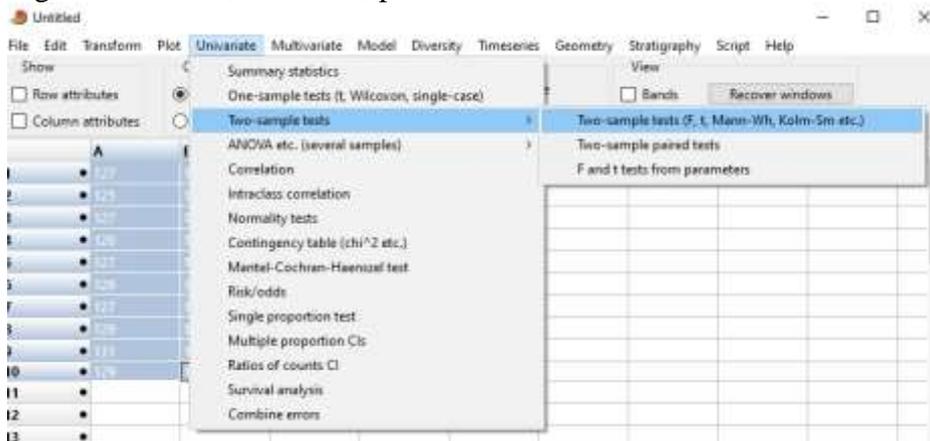
- 1 - MaqA	- 2 - MaqB	- 3 - MaqBx
127	125	132.0
125	124	132.0
127	125	127.0
126	120	136.0
127	119	128.0
129	121	119.0
127	123	125.0
129	119	117.0
131	122	121.0
129	121	119.0

Figura 5: Estatísticas descritivas das amostras

	- 1 - MaqA	- 2 - MaqB	- 3 - MaqBx
Tamanho da amostra =	10	10	10
Mínimo	125.0	119.0	117.00
Máximo	131.0	125.0	136.00
Amplitude Total	6.0	6.0	19.00
Mediana	127.0	121.5	126.00
Primeiro Quartil (25%)	127.0000	120.2500	119.5000
Terceiro Quartil (75%)	129.0000	123.7500	131.0000
Desvio Interquartilico	2.0000	3.5000	11.5000
Média Aritmética	127.7	121.9	125.60
Variância	3.1	5.2	42.27

Pode-se observar que a variância da MaqBx de 42.27 é bem superior à da MaqA (3.1) e da MaqB (5.2). Para tomarmos a decisão <D5> precisamos fazer um teste para testar se as variâncias das amostras MaqA e MaqB são estatisticamente iguais (homocedásticas) ou estatisticamente diferentes (heterocedásticas) ao nível de significância de 0.05.

Figura 6: Teste F, no PAST, para testar a homocedasticidade das variâncias MaqA e MaqB



Para fazer este teste é recomendável utilizar o software PAST que está disponível gratuitamente em <https://past.en.lo4d.com/download>. Copie e cole os dados das amostras MaqA e MaqB na planilha PAST e selecione o teste como mostra a Figura 6: selecione com o cursor as colunas que quer analisar, depois selecione “Univariate” → “Two-sample Tests” → “Two-sample Tests (F,t, Mann-Wh, Kolm-SM, etc)” e pressione “Enter”.

O resultado pode ser visto na Figura 7. Veja que na janela que se abre deve selecionar a aba “F test” para ter acesso ao “Tests for equal variances” (Hammer, Harper e Ryan, 2001). Note que na coluna A tem uma amostra com $n=10$ e variância 3.1222 e na coluna B tem outra amostra com $n=10$ e variância de 5.2111. O resultado mostra um $F=1.669$ e um $p\text{-value (same var.)}=0.45721$. Outras informações são também fornecidas tal como o Critical F value ($p=0.05$)= 4.026 . A interpretação mais simples é que as amostras são homocedásticas, isto é, não diferem significativamente ao nível de significância de 0.05 porquanto o valor de $F=1.669$ é menor do que o valor crítico de $F=4.026$.

Também se observa que as amostras são homocedásticas pelo $p\text{-value}$ que é maior do de 0.05.

Figura 7: Resultado do Teste F para testar a homocedasticidade das variâncias MaqA e MaqB

Two-sample tests			
Tests for equal variances			
A		B	
N:	10	N:	10
Variance:	3.1222	Variance:	5.2111
F:	1.669	p (same var.):	0.45721
Critical F value (p=0.05):	4.026	p (same var.):	0.301
Monte Carlo permutation:		p (same var.):	0.30603
Exact permutation:			

A decisão <D5> “Amostras homocedásticas?” é “sim”.

Substituindo os valores da amostra MaqB pelos valores MaqBx, simplesmente a título de exemplo, os resultados mostram MaqBx com uma variância de 42.267 e o valor de $F=13.537$ que é superior ao valor crítico de $F=4.026$, o que significa que as amostras, neste exemplo possuem variâncias significativamente diferentes ao nível de significância de 0.05. Isso também é exibido pelo $p\text{-value}=0.0006354$ que é menor do que 0.05.

Referências

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

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

Cooper, D. R.; Schindler, P. S.(2003). Métodos de pesquisa em administração. 7.ed. Porto Alegre: Bookman.

Hammer, Ø., Harper, D. A., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia electronica*, 4(1), 9.

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

PAST <https://past.en.lo4d.com/download>

Vessereau, A. (1965) A Estatística. São Paulo: Difusão Europeia do Livro.