REVISTA DA MICRO E PEQUENA EMPRESA (RMPE)
Vol. 15, Nº 3, Set-Dez 2021, 1 de 5
ISSN: 1982-2537

http://dx.doi.org/10.48099/1982-2537/2021v15n3p15

O Processo Estatístico: tamanho e aderência à distribuição normal*

Manuel Meireles
UNIFACCAMP
meireles@faccamp.br

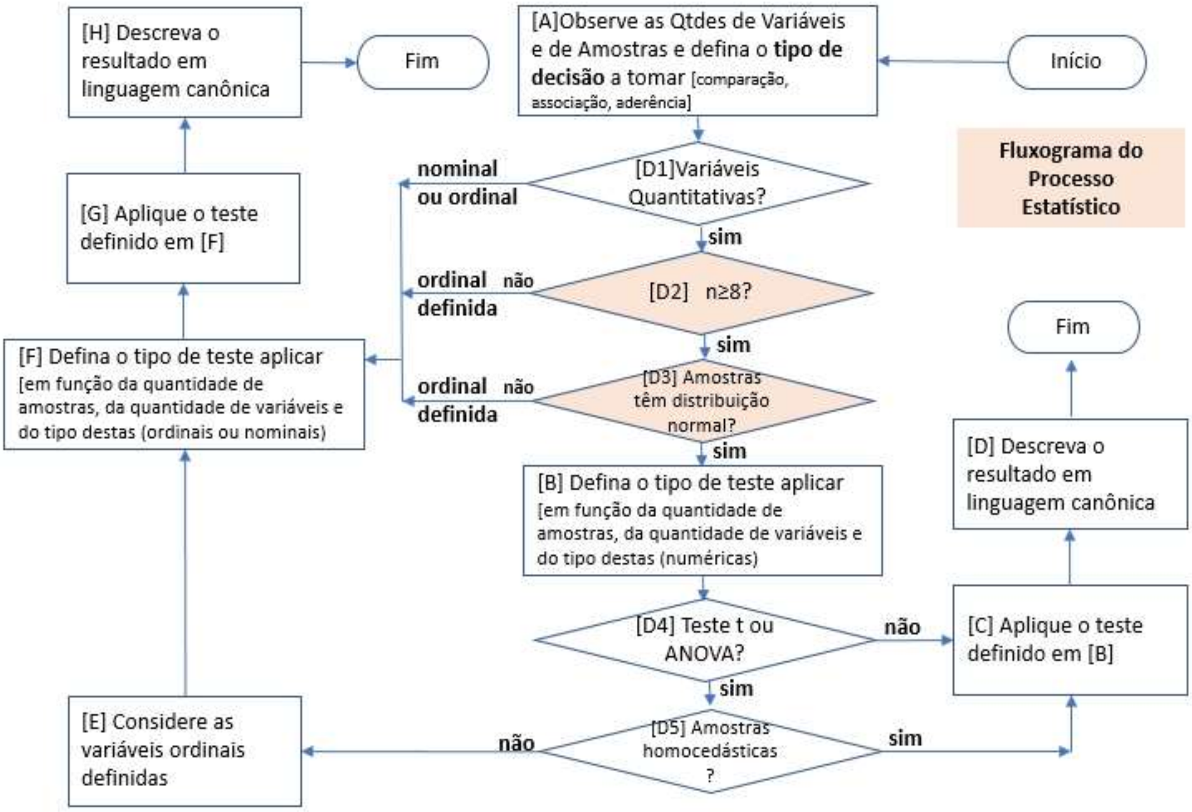
Este trabalho é continuação do publicado na edição anterior e tem por objetivo discorrer sobre o conceito de variável e definir seu tipo de forma correta, pois isso é vital para a boa condução do processo estatístico.

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 das decisões <d2> e <D3> (ver Figura 1) já que os passos A e <D1> foram objeto de estudo nos dois números imediatamente anteriores.

*Recebido em 04 de abril de 2022, aprovado em 04 de abril de 2022, publicado em 11 de abril de 2022.

Figura 1: Decisões <D2> e <D3>



Antes de executar a atividade B na qual se define o tipo de decisão a tomar (se aderência, comparação ou associação) é necessário tomar as decisões <D2> e <D3> como se pode ver na Figura 1. A decisão <D2> responde à questão: “O tamanho da amostra é ≥ 8?”. Basta contar a quantidade de itens da amostra para responder a esta questão. Mas por que 8 é o limite? Ouçamos Vessereau (1965:68) a respeito de pequenas amostras:

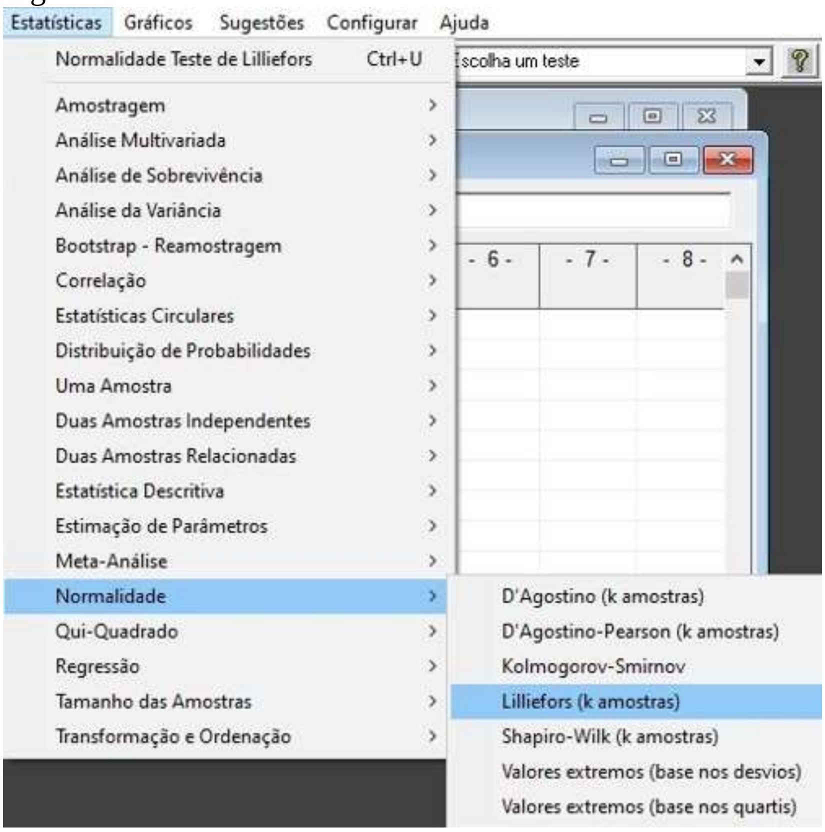
> Acreditou-se, durante muito tempo, que era impossível tirar proveito d amostras pequeníssimas de 5 a 10 medidas, por exemplo, e que o domínio da estatística era o das medidas muito numerosas. De fato, os testes adaptados às pequenas amostras podem pecar por falta de sensibilidade, mas corretamente interpretados, são tão rigorosos quanto os que se aplicam às estatísticas abundantes. É a Student (pseudônimo de um estatístico inglês cujo verdadeiro nome era Gosset) que cabe o mérito de ter estabelecido o método que permite comparar as médias de pequenas amostras.

Optou-se por estabelecer a média arredondada para cima de 5 a 10 medidas, para definir 8 como limite de amostra passível de ser executada por teste paramétrico. Desta forma na decisão <D2> pode-se ter duas respostas: “sim” e em seguida se responde à questão <D3>; ou “não” que transforma a variável quantitativa em uma variável ordinal, o que significa que nesta situação só se poderá aplicar testes não-paramétricos.

Uma vez resolvida a decisão <D2> surge a decisão <D3> que requer que se diga se a amostra tem uma distribuição que não difere significativamente da distribuição normal. Para testar esta condição é necessário fazer uso de um software estatístico. Faremos uso do software BioEstat que é *free*. Ver Figura 2.

Pela <D3> se confirma 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%).

Figura 2: Testes de normalidade no BioEstat



Basicamente podem ser usados dois dos testes:

TESTE D'AGOSTINO (k amostras): trata-se de um teste para pequenas amostras, mas cujo valor de n deve ser igual ou superior a dez (10) unidades. Os resultados são comparados com os valores críticos previamente estabelecidos. É recomendável, ainda, que os resultados sejam fornecidos em cinco (5) decimais, uma vez que o valor estatístico do desvio (D) pode ser muito pequeno. Pode ser efetuado para várias (k) amostras simultaneamente.

TESTE DE LILLIEFORS (k amostras): prova não-paramétrica de aderência destinada a comparar o grau de concordância entre a distribuição acumulada de um conjunto de valores de K amostras (com n>3) com a distribuição normal teórica acumulada esperada.

A Figura 3 mostra exemplo do teste D'Agostino (para n≥10): Notar que uma das variáveis (a da coluna 2) difere significativamente da distribuição normal ao nível de significância de 0.05. Neste caso a variável originalmente quantitativa (<D1>) passa a ser definida como ordinal. Neste exemplo a resposta à questão <D3> é não pelo fato de uma das variáveis ser significativamente diferente da distribuição normal ao nível de significância de 0.05. A próxima ação a executar é a F, como mostra a Figura 1.

Figura 3: Exemplo de aplicação do teste D’Agostino.

Caso 01	
GCP	Vendas
10.6179	159.827
16.03521	176.9471
10.8428	170.1453
15.15681	189.5719
14.40552	168.489
16.14578	187.9311
13.09779	162.1699
15.50828	182.6738
13.94173	158.6136
12.25715	184.1674
18.83607	182.195
13.80307	112.4424

10.618	15
16.035	17
10.843	17
15.157	18
14.406	16
16.146	18
13.098	16
15.508	18
13.942	15
12.257	18
18.836	18
13.803	11

D'Agostino

Arquivo Editar Gráfico

Resultados	Coluna 1	Coluna 2
Tamanho da amostra =	12	12
D (Desvio) =	0.2798	0.2470
Valores críticos 5%	0.2544 a 0.2854	0.2544 a 0.2854
Valores críticos 1%	0.2420 a 0.2862	0.2420 a 0.2862
p =	ns	p < 0.05

A Figura 4 mostra exemplo do teste Lilliefors (para n≤10): Notar que nenhuma das variáveis difere significativamente da distribuição normal ao nível de significância de 0.05. Neste caso as variáveis originalmente quantitativas (<D1>) continuam como quantitativas. Neste exemplo a resposta à questão <D3> é sim pelo fato de nenhuma das variáveis ser significativamente diferente da distribuição normal ao nível de significância de 0.05. A próxima ação a executar é a B, como mostra a Figura 1.

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 será vista no próximo número.

Figura 4: Exemplo de aplicação do teste Lilliefors

Caso 07	
GCP	Vendas
13.35702	172.8105
13.53593	191.9662
15.58088	165.9153
15.34116	176.2393
14.82637	178.6661
15.08761	166.5655
14.08672	169.9649
15.87607	170.7085

Normalidade Teste de Lilliefors

Arquivo Editar Gráfico

	- 5 -	- 6 -
Tamanho da amostra =	8	8
Desvio máximo =	0.1733	0.1859
Valor crítico (0.05) =	0.2850	0.2850
Valor crítico (0.01) =	0.3310	0.3310
p(valor)	ns	ns

Embora os dois testes apontados acima sejam suficientes para testar se os dados aderem ou não significativamente, os demais testes referidos na Figura 2 podem ser estudados no Manual do BioEstat de Ayres, Ayres Jr. e Dos Santos (2007).

Referências

Assis, J. P.; Sousa, R. P., Dias, C.T.S. (2019). Glossário de estatística. Mossoró: EdUFERSA, 2019.

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

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

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