

Processo Estatístico: Linguagem canônica

Statistical Process: Canonical language

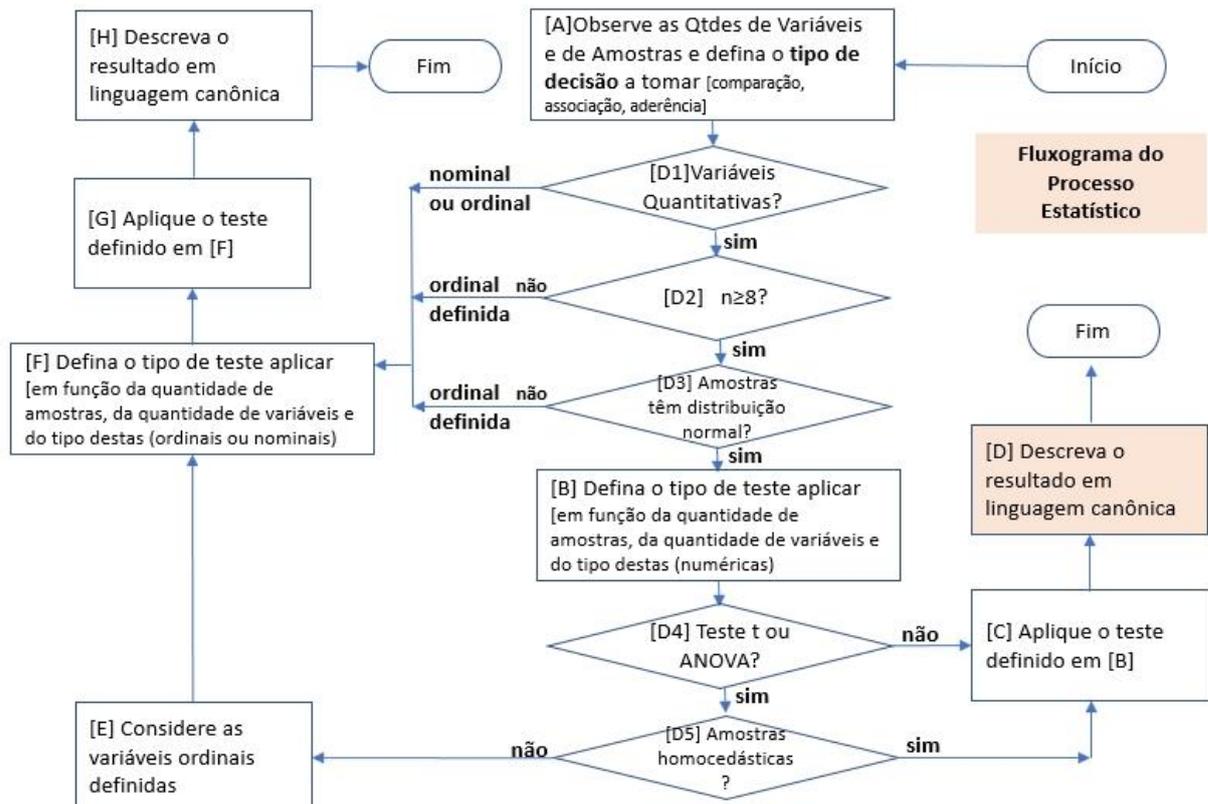
Manuel Meireles
UNIFACCAMP
meireles@faccamp.br

Este trabalho é continuação do publicado na edição anterior e tem por objetivo abordar a atividade [D] após ter sido aplicado o teste estatístico na atividade [C] 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 da atividade [D] que é descrever o resultado em linguagem canônica.

Figura 1: Localização da atividade D no Fluxograma do Processo Estatístico



O exemplo retirado de Meireles (2022), visto no número anterior, trata-se de um caso (o caso 2) referente a uma microempresa de Artefatos de Cimento que tem duas máquinas para fabricar blocos de cimento, designadas aqui por MaqA e MaqB. A Tabela 2.2 (no artigo

anterior) 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?

Seguindo-se o Processo Estatístico verificamos que a atividade [B] deu como resultado os testes sugeridos para o presente caso o Teste t ou ANOVA.

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> foi “sim” pelo que se precisou tomar a decisão <D5> logo a seguir: as amostras são homocedásticas?

A decisão <D5> “Amostras homocedásticas?” foi “sim” e isso conduziu o fluxo do processo estatístico para a realização da Atividade [C] que é “aplicar o teste definido em [B]”. E os testes definidos na Atividade [B] foram, como visto acima para este exemplo, o Teste t ou ANOVA. Na Atividade [C] foi aplicado o Teste t: teste destinado a analisar a diferença estatística entre duas amostras independentes do mesmo tamanho ou desiguais. O número de escores de cada amostra deve ser igual ou inferior a 30. Esta prova é conhecida como teste de Student. Obteve-se o resultado mostrado na Figura 2.

Figura 2: Output do teste t

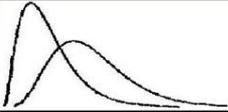
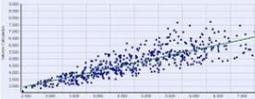
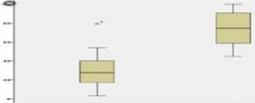
t - Teste: Amostras Independentes

Arquivo Editar Gráfico

	- 1 - MaqA	- 2 - MaqB
Tamanho =	10	10
Média =	127.7000	121.9000
Variância =	3.1222	5.2111
	Homocedasticidade	
Variância =	4.1667	---
t =	6.3536	---
Graus de liberdade =	18	---
p (unilateral) =	< 0.0001	---
p (bilateral) =	< 0.0001	---
Poder (0.05)	1.0000	---
Poder (0.01)	1.0000	---
Diferença entre as médias =	5.8000	---
IC 95% (Dif. entre médias) =	3.8821 a 7.7179	---
IC 99% (Dif. entre médias) =	3.1728 a 8.4272	---

Os resultados do teste t começam por mostrar o tamanho das amostras e a média e a variância. Adicionam a informação de que não há diferença significativa entre as variâncias já que as amostras são Homocedásticas. Informações relacionadas ao teste t são as que se apresentam na parte restante da figura devendo ser destacado o valor p (unilateral) <0.0001. Este é o chamado p-value que deve ser interpretado conforme a Figura 3. O p-value é a medida de quanta evidência se tem a favor da hipótese nula.

Figura 3: Interpretação do p-value.

Interpretação do p-value			
Tipo de decisão	Hipótese Alternativa	Confirma a Hipótese Alternativa	Mnemônico gráfico
ADERÊNCIA	Há diferença entre Observados e esperados	$p\text{-value} \leq 0.05$	
ASSOCIAÇÃO	Há associação entre as variáveis	$p\text{-value} \leq 0.05$	
COMPARAÇÃO	Há diferença entre as amostras	$p\text{-value} \leq 0.05$	

No exemplo, trata-se de uma comparação e a hipótese alternativa (oposta à hipótese neutra) é que há diferença entre as amostras considerando a variável em análise já que o p-value observado é menor do que 0.05. Ou seja, a produção de blocos da MaqA é significativamente maior do que a produção da MaqB.

A linguagem canônica é um conjunto de regras e terminologias específicas adotadas em um determinado campo ou área do conhecimento. No contexto estatístico, a linguagem canônica estatística é essencial para comunicar resultados de análises e testes de maneira clara, precisa e padronizada. Quando se trata de apresentar resultados estatísticos, é fundamental seguir a linguagem canônica apropriada para o teste realizado. Por exemplo, ao conduzir um teste de hipótese, é comum utilizar expressões como "Hipótese rejeitada" ou "Hipótese não rejeitada". Essas declarações indicam se há evidências estatísticas suficientes para aceitar ou rejeitar uma hipótese nula, que é uma afirmação inicial a ser testada.

Outra forma comum de apresentação de resultados estatísticos é por meio do uso de níveis de significância. Ao analisar a diferença entre duas variáveis ou grupos, é possível afirmar, a um determinado nível de significância, que uma variável é significativamente maior do que outra, ou que há uma diferença estatisticamente significativa entre dois grupos.

A linguagem canônica estatística não apenas garante a precisão e a objetividade na comunicação de resultados, mas também facilita a compreensão e interpretação por parte dos leitores e pesquisadores. Ela estabelece padrões claros e uniformes, permitindo que os resultados de estudos sejam comparáveis e replicáveis, o que é essencial para o avanço da ciência e da pesquisa.

É importante ressaltar que a linguagem canônica estatística é desenvolvida com base em princípios estatísticos sólidos e métodos standardizados. Por isso, é crucial que pesquisadores e profissionais da área estatística estejam familiarizados com essa linguagem e a utilizem corretamente em suas análises e relatórios.

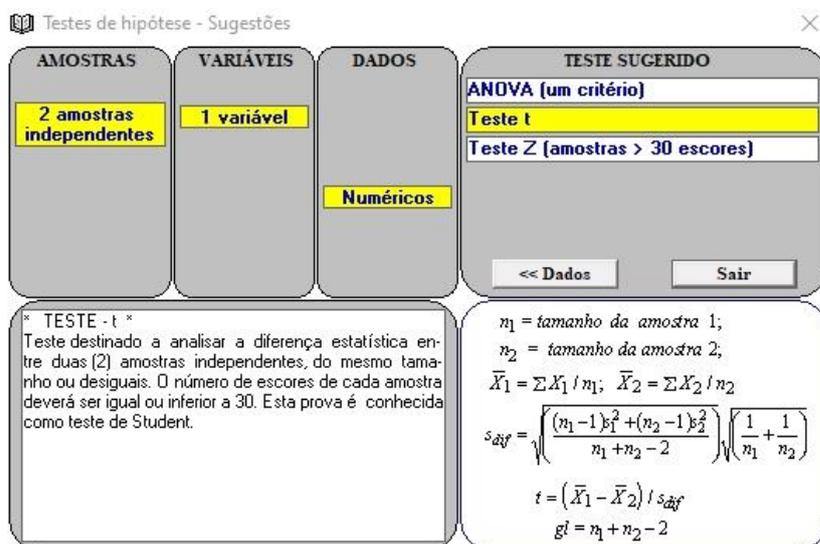
Em resumo, a linguagem canônica estatística é uma ferramenta poderosa para comunicar resultados estatísticos de forma precisa e universalmente compreensível. Ela desempenha um papel fundamental na disseminação do conhecimento científico e na garantia da confiabilidade dos resultados obtidos por meio de análises estatísticas.

O Manual do estilo APA, na sua sétima edição (2020) recomenda que não se adicione fórmulas estatísticas comuns, a menos que sejam novas, raras ou essenciais para a pesquisa. Ao relatar resultados de análises estatísticas (como testes t e F, qui-quadrado, tamanho do efeito e intervalos de confiança), forneça informações suficientes para entender a análise, mesmo que o tamanho do efeito possa ser calculado a partir dos dados no artigo. Para dados multiníveis,

apresente estatísticas para cada nível. Se estatísticas descritivas forem apresentadas em uma tabela ou figura, não repita no texto, apenas faça referência à tabela.

No presente exemplo se utilizaria a seguinte linguagem para escrever o resultado: Hipótese não rejeitada*: pode-se afirmar ao nível de significância de 0.01** que a MaqA produz significativamente *** mais do que a MaqB (teste t; p-value(unilateral)<0.0001; t=6.3536; gl=18)**** .

Figura 4: Informações mínimas no output do “teste sugerido”



Notas: *: adota-se a expressão Hipótese não rejeitada em vez de Hipótese aceita; **: embora o teste tenha sido feito ao nível de significância 0.05, dado que o p-value foi <0.0001 muda-se o nível de significância para 0.01 (isto deve ser feito sempre que o p-value for <0,01; ***: a expressão “significativa” deve ser sempre utilizada, pois em estatística o que se busca demonstrar é que há (ou não) diferença, associação ou aderência significativa; ****: as informações que devem constar variam de teste para teste. A figura 4, mostra que a aba “Sugestões” do BioEstat fornece o tipo de informação mínima que deve ser utilizada ao difundir o resultado do teste; no exemplo, além do p-value, o valor t e gl (graus de liberdade).

No próximo número encerraremos esta série.

Referências

APA Manual. Edição 7 em pdf do estilo APA de referência. (2020) <https://psy-journal.hse.ru/data/2021/11/14/1444762739/APA%202020%207th%20Ed.pdf>

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>