

DECIFRANDO O GRÁFICO DE CAIXA (BOX-PLOT) COM USO DO CODAP—RELATO DE UMA OFICINA

Rafael V. Bonangelo e Lisbeth K. Cordani
IME USP, SP, BR
rvieira@ime.usp.br

Gráficos fazem parte do ensino de estatística na escola básica e são ferramentas importantes na análise de dados. Um dos gráficos adequados, principalmente para comparação de grupos, é o chamado Gráfico de Caixa (Box-Plot). Sua construção foi apresentada em uma oficina online de duas sessões de 2 horas. As variáveis, coletadas em tempo real, foram idade, medida do palmo da mão (cm) e gênero. O aplicativo estatístico utilizado para análise foi o CODAP, online e gratuito. Foi feita análise do comportamento do palmo da mão globalmente e também segundo cada gênero e os resultados foram compatíveis com pesquisas anteriores. As reações dos alunos foram positivas tanto com o CODAP como com a construção do Gráfico de Caixa.

INTRODUÇÃO

A Estatística é ensinada no Brasil na escola básica junto com a disciplina de Matemática e esse conhecimento é cada vez mais necessário em nossa sociedade, coletora de dados, onde as informações são divulgadas e consumidas. Segundo Gould (2017), as democracias exigem um debate informado, e quase todas as discussões políticas requerem alguma compreensão estatística. Inserir Educação Estatística antes que a mente se cristalice em modelos determinísticos prepara os cidadãos para pensarem por eles mesmos, para fazerem perguntas apropriadas bem como terem confiança para se defender de imposições autoritárias. Paradoxalmente à abundância de dados que deveria proporcionar mais informação, algumas tendências atuais apontam para situações em que são deturpadas as evidências empíricas (Engel, 2017). Na maioria das vezes isso ocorre, segundo Frischemeier e Schnell (2021), pois as pessoas não estão preparadas para analisar evidências empíricas que diferem de suas expectativas iniciais em determinado contexto. Cabe aos educadores estatísticos a responsabilidade de prover o conhecimento para uma tomada de decisão consciente e responsável.

Existe no Brasil um documento normativo denominado Base Nacional Comum Curricular (BNCC), com o objetivo de definir o que é essencial para a aprendizagem em nível escolar a partir de competências e habilidades que devem ser adquiridas pelos alunos (Ministério da Educação, 2018). Ainda apresentada como parte da Matemática, o conteúdo da Estatística antes circunscrito a uma sequência de cálculos de natureza instrumental e sem motivação, possui diversos temas a serem explorados, que incluem medidas de posição e de variação, bem como suas representações gráficas, cuja importância reside no impacto que uma figura causa numa análise de dados para conhecimento de um fenômeno. A propósito, o objetivo deste texto será apresentar aos professores da escola básica (principalmente do Ensino Médio) a construção do Gráfico de Caixa (Box-Plot), o qual é citado no documento da Base Nacional como a sétima habilidade da Competência Específica 4:

Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos (histograma, de caixa (Box-Plot), de ramos e folhas, entre outros), reconhecendo os mais eficientes para sua análise. (Ministério da Educação, 2018, p. 539)

O Gráfico de Caixa (Box-Plot), criado por Tukey (1977), leva em conta a importância da representação de dados através de seu posicionamento de centralidade e a visualização da variabilidade (tamanho da caixa) e introduz a possibilidade de começar na escola básica a ideia de inferência, vista inicialmente sob o ponto de vista informal. O relatório GAISE (Franklin et al., 2007) comenta a importância do Gráfico de Caixa para comparação de grupos e é essa a abordagem que será feita aqui.

O artigo de Pfannkuch (2006) mostra as diferenças da iniciação de acordo com a faixa etária nesse tipo de gráfico no início do século XX (12 anos nos EUA, 17 anos na França, 14 anos na Nova Zelândia além de China e de Israel que não introduziam). Essa diferença de idade para início da aprendizagem pode ser uma consequência, em grande parte, de algumas dificuldades, como por exemplo o fato de o gráfico mostrar os pontos agregados e não individualizados (neste caso começar pelo Dot-Plot—Gráfico de Pontos facilita a familiarização com a visualização, pois mostra cada valor observado). Nessa direção, alguns autores, como Edwards et al. (2017) e Almeida et al. (2021) também sugerem construir o Dot-Plot primeiro e a partir dele construir o Box-Plot. Esta facilidade é disponível no

CODAP (Common Online Data Analysis Platform; The Concord Consortium, 2014), como veremos adiante.

Quanto ao CODAP, sua proposta é ajudar os alunos a explorar, visualizar e aprender com os dados que são inseridos de maneira fácil, propiciando ao aluno a construção de gráficos e tabelas e o correspondente cálculo de medidas estatísticas. Pode ser acessado em codap.concord.org onde se encontram dois tutoriais que orientarão os primeiros passos, bem como algumas atividades para a sala de aula. Importante dizer que é um aplicativo que foi projetado para estudantes da escola básica (fundamental e médio), gratuito e de código aberto, tendo sido financiado por doações da National Science Foundation dos EUA. A oficina foi apresentada com o suporte deste aplicativo.

ATIVIDADE

Com o propósito de discutir e explorar a construção do Gráfico de Caixa dentro de uma análise estatística passível de ser trabalhada por professores na educação básica, foi organizada em novembro de 2020 pelos autores uma oficina online (devido à pandemia) de modo síncrono, a mais de 40 professores, inspirada na medida do palmo da mão, descrita na Figura 1 (Cordani e Fontes, 2018). Essa oficina fez parte das atividades oferecidas anualmente pelo Centro de Aperfeiçoamento do Ensino de Matemática (CAEM) do Instituto de Matemática e Estatística da USP, que facilitou sua realização, preparando o ambiente virtual para quatro horas divididas em dois dias. A audiência foi constituída em sua maioria por professores da escola básica brasileira, livremente inscritos, e a participação efetiva na coleta de dados foi feita por 44 deles (19 mulheres e 25 homens). A idade média do grupo dos respondentes foi 38,8 anos e a idade mediana 39 anos. A amplitude foi de 42 anos (máximo 60 anos). A origem dos respondentes foi constituída por 70% de São Paulo, 11% de Minas Gerais e os restantes 19% provenientes dos estados Ceará, Maranhão, Paraná, Pernambuco, Rio de Janeiro, Rio Grande do Norte e Rio Grande do Sul.

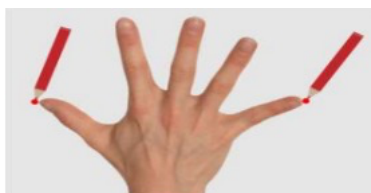


Figura 1. Medindo o palmo da mão direita (cm); Fonte: Cordani e Fontes, 2018

A programação foi inspirada no Ciclo Investigativo de Wild e Pfannkuch (1999), para quem o aprendizado da estatística se dá através da resolução de um problema real com a compreensão da necessidade de contemplar a variabilidade inerente à investigação de qualquer fenômeno. Esse ciclo investigativo propõe DEFINIR o problema, DETALHAR o plano de ação, COLETAR os dados, ANALISAR os dados e CONCLUIR fornecendo uma resposta ao problema inicial e, se for o caso, incluir novos questionamentos e retomar o ciclo. A sugestão apresentada aos professores foi de uma atividade para ser desenvolvida na sala de aula, com as etapas do ciclo, propondo responder às perguntas “O palmo da mão direita de meninos e meninas (homens e mulheres) tem o mesmo comportamento?” ou “Se formos fabricar luvas temos que fazer alguma distinção por gênero?” Foi enfatizado que os dados não devem ser coletados a esmo, só para ilustrar uma técnica, mas devem ter um propósito que estimule a execução da análise e motive a audiência (alunos e/ou professores).

Levando em conta o pouco tempo de atividade (4 horas divididas em 2 dias), optou-se por apresentar as duas “novidades” (que seriam o Gráfico de Caixa e o CODAP) logo no início da oficina para despertar a curiosidade dos professores. Assim no primeiro dia foi apresentado não só o link para o CODAP (como também o Gráfico de Caixa construído pelo CODAP com um conjunto pequeno (fictício) de 12 valores de tamanho de palmo de mão de homens e mulheres. Neste primeiro contato foi construído pelo CODAP o Gráfico de Pontos e acima deste o Box-Plot correspondente. A audiência ia trabalhando à distância, concomitantemente com o apresentador e tirando as eventuais dúvidas.

É importante ressaltar que uma amostra de 12 unidades é muito pequena para construir um Gráfico de Caixa informativo sobre qualquer fenômeno. Isso foi colocado na oficina para fins didáticos e a construção posterior foi feita com os dados da própria audiência, o que proporcionou um número

maior de dados (o ideal é que seja no mínimo 20 unidades). Além disso é importante enfatizar que o uso do Gráfico de Caixa é adequado para variáveis de natureza quantitativa (que é o caso da medida do palmo da mão). Foi então apresentada a construção passo a passo do Gráfico de Caixa, inicialmente na sua versão simplificada, construção essa que necessita somente de cinco medidas de posição, como será visto a seguir. Foi sugerido que os professores da oficina fizessem o gráfico à mão durante a oficina usando este pequeno conjunto de dados para sedimentarem sua construção e colocarem eventuais dúvidas ao apresentador.

As facilidades do trabalho com o CODAP foram evidenciadas com a inserção dos dados na planilha e a geração de Gráfico de Pontos (Dot-Plot) com o correspondente Gráfico de Caixa (Box-Plot). No final do primeiro dia foi pedido aos professores que fizessem a medida da sua própria mão (Figura 1) e inserissem, numa planilha disponível online, os resultados arredondados para o inteiro mais próximo. O arredondamento proposto é no sentido de simplificar o tratamento de dados com a construção tanto do Gráfico de Pontos como do Gráfico de Caixa feitas à mão na sala de aula. Claro que ao trabalhar somente em ambiente tecnológico isso não seria necessário, embora apresentar somente valores inteiros facilita a visualização e a identificação dos valores individuais.

Com todos os dados coletados foram construídos no CODAP os correspondentes Gráficos de Caixa que foram visualizados por todos e analisados em tempo real. A oficina, gravada no Youtube, detalha o processo tanto de construção do Gráfico de Caixa como o uso do CODAP nesse contexto (Bonangelo e Cordani, 2020a, 2020b). Em resumo, esse gráfico trabalha somente com 5 medidas descritivas de posição obtidas através dos dados: mediana, mínimo, máximo, primeiro e terceiro quartis, o que é um ponto positivo pela simplicidade. A caixa é limitada pelos quartis, a mediana é posicionada dentro da caixa e as hastes vão de cada lado da caixa ao correspondente valor extremo (mínimo à esquerda e máximo à direita). Esta configuração é a que chamamos de simplificada, sugerida para os primeiros contatos com o gráfico.

A Figura 2, construída no CODAP com os dados dos professores coletados online, foi discutida na oficina e justapõe os Gráficos de Pontos e o de Caixa (simplificado), separando por gênero. É possível verificar que o grupo masculino apresenta valores em geral maiores do que o feminino e há uma certa assimetria em ambos os gráficos, bem como uma semelhança na variabilidade, indicada pelo tamanho das caixas ($Q3 - Q1$) que é o Intervalo Interquartil, que representa variabilidade. Esse comportamento é compatível com os encontrados em Cordani e Fontes (2018), sugerindo confecção separada de luvas para cada gênero.



Figura 2. Gráfico de Pontos e Gráfico de Caixa (simplificado) com dados dos professores
Fonte: Captura de tela do CODAP (Bonangelo e Cordani, 2020a)

A Figura 3, apresentada e discutida na segunda parte da oficina, mostra os Gráficos de Caixa para os mesmos dados da figura anterior com a diferença que não é simplificado e, nesse caso, contempla possíveis dados discrepantes (neste caso havia um dado discrepante, simbolizado por +). Na Figura 2 as hastes que saem da caixa do Gráfico vão de cada quartil até o valor extremo (mínimo ou máximo).

A Figura 3, apresentada e discutida na segunda parte da oficina, mostra os Gráficos de Caixa para os mesmos dados da figura anterior com a diferença que não é simplificado e, nesse caso, contempla possíveis dados discrepantes (neste caso havia um dado discrepante, simbolizado por +). Na Figura 2 as hastes que saem da caixa do Gráfico vão de cada quartil até o valor extremo (mínimo ou máximo). Mas

até onde a haste iria se um valor muito grande aparecesse? Na segunda parte da oficina isso é discutido e mostrado que valores muito afastados da caixa (maiores do que 1,5 x tamanho da caixa) são considerados como discrepantes (outliers) e representados com um sinal + (ou *), não sendo incluídos nas hastes. Este critério de verificação de valor discrepante (outlier) seria melhor explicado se houvesse tempo para inserir a distribuição normal (o que não foi o caso), mostrando o que o critério acarreta levando em conta as probabilidades das caudas. Esta construção levando em conta possíveis outliers deve ser abordada quando a versão simplificada estiver bem dominada pelos alunos, na perspectiva do que Franklin et al. (2007) preconizam, ou seja, seria trabalhado com alunos no estágio C de conhecimento de estatística (no atual cenário brasileiro isto se dará, na melhor das hipóteses—a nosso ver—nos dois anos finais do Ensino Médio).

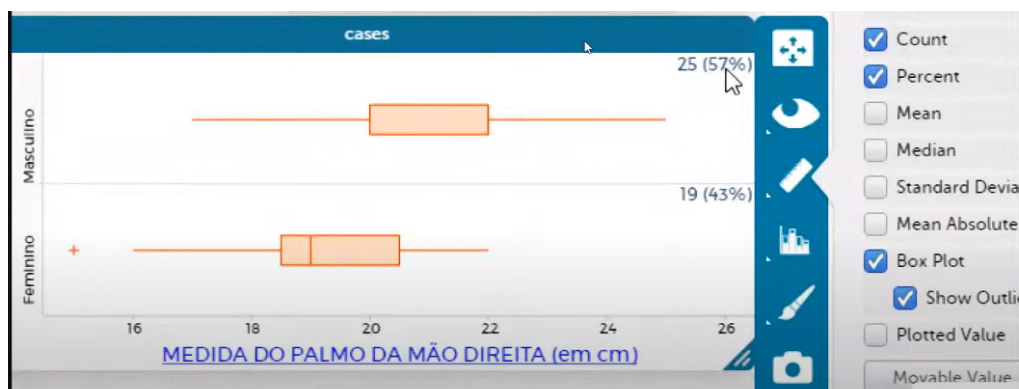


Figura 3. Gráfico de Caixa com dados dos professores e valor discrepante
Fonte: Captura de tela do CODAP (Bonangelo e Cordani, 2020a)

IMPORTÂNCIA DA ATIVIDADE

O Gráfico de Caixa, construído a partir de medidas facilmente calculadas, apresenta algumas vantagens para “aproximar” os alunos dos dados coletados. A saber:

- (a) Encaminha o aluno para ordenar os dados originais, fazendo com que ele tome contato com cada elemento coletado. Além disso, os alunos gostam de se verem retratados nas medidas trabalhadas, desde que a mesma não provoque discriminação (a medida do palmo é uma boa candidata sob esse aspecto, além de ser facilmente coletada). A superposição do Gráfico de Caixa ao Dot-Plot (apresentada pelo CODAP) facilita esse aspecto.
- (b) Introduce o aluno às cinco medidas de posição, muito informativas, necessárias e suficientes para a construção básica do gráfico, a saber: Mínimo (menor valor observado), Q1 (1º quartil, separa o primeiro quarto do conjunto ordenado de dados dos demais), Mediana (valor que divide o conjunto ordenado de dados ao meio), Q3 (3º Quartil, separa o último quarto do conjunto ordenado de dados dos demais) e Máximo (maior valor observado), todas denominadas “medidas de posição” pois podem ser posicionadas na reta associada aos dados.
 - A construção é simples, com a caixa limitada pelos quartis Q1 e Q3, o traço vertical dentro da caixa indicando a Mediana e com as hastes saindo de cada lado da caixa indo como primeira abordagem (simplificada) até os valores extremos (mínimo ou máximo).
 - Apesar de a literatura apresentar diferentes maneiras de aproximar o cálculo dos quartis, sugerimos inicialmente calculá-los de modo intuitivo, como medianas dos dois conjuntos de valores definidos pela Mediana de todos os dados, ou seja, Q1 (primeiro quartil) foi considerado como sendo a mediana da primeira metade dos dados (ordenados) e Q3 (terceiro quartil) foi considerado como sendo a mediana da segunda metade dos dados (ordenados).
- (c) Habitua o aluno a pensar no conceito de densidade (quantidade de unidades por área) uma vez que essas medidas, dispostas convenientemente no gráfico, mostram as porcentagens associadas a cada segmento (ou seja, 25%, 50%, 25%).
- (d) Informa (na sua versão tradicional) rapidamente se o conjunto apresenta ou não valores discrepantes.
- (e) Ilustra a variabilidade maior ou menor dos dados, quer seja através da Amplitude (medida elementar de variação, obtida da diferença entre o Máximo e o Mínimo), quer seja através do Intervalo

Interquartil (medida mais informativa de variação) que, como o nome indica, é a distância entre Q3 e Q1 (ou seja, o tamanho da caixa). Essas duas medidas são mais adequadas para iniciar a discussão sobre variabilidade dos dados, deixando para mais tarde a inclusão de desvio padrão, por exemplo.

- (f) Favorece a comparação de grupos, evidenciando diferenças ou não tanto em centralidade como em variabilidade, ou em comportamento simétrico ou ainda notando se os grupos apresentam discrepâncias dignas de nota.
- (g) Permite ao aluno a passagem de um gráfico mais simples que é o Gráfico de Pontos (Dot-Plot), onde cada elemento representa uma unidade de coleta, para o Gráfico de Caixa que sintetiza o comportamento dos dados coletados, através de medidas facilmente construídas.

O uso do CODAP (com suas facilidades) é apresentado em Bonangelo e Cordani (2020a, 2020b), através da criação de tabela de dados coletados, do trabalho com colunas bem como criação dos gráficos. O software possui recursos importantes para o ensino da estatística como a interatividade entre representações, o registro de procedimentos e a possibilidade de compartilhamento do que está sendo desenvolvido. Muitos dos procedimentos são facilitados com o estilo de arrastar e soltar e logo os professores se familiarizaram com os procedimentos.

COMENTÁRIOS FINAIS

Pelos resultados encontrados foi possível verificar que o grupo masculino apresenta valores em geral maiores do que o feminino (aqueles visivelmente deslocados à direita destes), com variabilidade semelhante (tamanhos das caixas semelhantes), e que a sugestão seria fazer luvas levando em conta o gênero como visto em Cordani e Fontes (2018), sendo eventualmente dois tamanhos para cada gênero. Pode-se dizer que o comentário final é uma introdução, ainda que primitiva, do que se denomina de inferência informal, que não foi objeto da oficina.

A atividade, apesar de ter sido curta, foi bem avaliada pelos professores assistentes, que deixaram algumas opiniões, como por exemplo: (a) trabalhar esse método de introduzir o aluno na prática é muito interessante; (b) ADOREI! Nunca foi tão simples fazer um Box-Plot; (c) queria ter visto isso antes de fazer estatística; (d) CODAP salva; e (e) curso pertinente e completamente aderente ao solicitado pela BNCC etc.

AGRADECIMENTOS

Os autores agradecem a leitura cuidadosa dos referees, bem como as sugestões apontadas.

REFERÊNCIAS

- Almeida, C. R., Sousa, H. J., e Cazorla, I. M. (2021). Letramento estatístico na educação básica: Os desafios de ensinar o diagrama da caixa (box-plot) em contexto. *Educação Matemática Pesquisa*, 23(1), 499–529. <https://doi.org/10.23925/1983-3156.2021v23i1p499-529>
- Bonangelo, R. V., e Cordani, L. K. (2020a, novembro 4). *Decifrando o gráfico boxplot (diagrama de caixa) no Ensino Médio (Parte 1)* [Video]. YouTube. <https://www.youtube.com/watch?v=2cpfvtZ2D5o>
- Bonangelo, R. V., e Cordani, L. K. (2020b, novembro 6). *Decifrando o gráfico boxplot (diagrama de caixa) no Ensino Médio (Parte 2)* [Video]. YouTube. <https://www.youtube.com/watch?v=FIo3YMLU5Eg>
- Cordani, L. K., e Fontes, D. S. (2018). Planejando abrir um negócio? Use estatística! *Ensino da Matemática em Debate*, 5(1), 85–102. <https://revistas.pucsp.br/index.php/emd/article/view/37180>
- Edwards, T. G., Özgün-Koca, A., e Barr, J. (2017). Interpretations of box-plots: Helping middle school students to think outside the box. *Journal of Statistics Education*, 25(1), 21–28. <https://doi.org/10.1080/10691898.2017.1288556>
- Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, 16(1), 44–49. <https://doi.org/10.52041/serj.v16i1.213>
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., e Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k–12 curriculum framework*. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseprek-12_full.pdf

- Frischemeier, D., e Schnell, S. (2021). Statistical investigations in primary school—the role of contextual expectations for data analysis. *Mathematics Education Research Journal*. <https://doi.org/10.1007/s13394-021-00396-5>
- Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1), 22–25. <https://doi.org/10.52041/serj.v16i1.209>
- Ministério da Educação. (2018). *Base nacional comum curricular–BNCC*. Brasil. http://basenacionalcomum.mec.gov.br/images/historico/BNCC_EnsinoMedio_embaixa_site_110518.pdf
- Pfannkuch, M. (2006). Comparing box-plot distributions: A teacher’s reasoning. *Statistics Education Research Journal*, 5(2), 27–45. <https://doi.org/10.52041/serj.v5i2.498>
- The Concord Consortium. (2014). *Common Online Data Analysis Platform* [Computer software]. <https://codap.concord.org/>
- Tukey, J. (1977). *Exploratory data analysis*. Addison-Wesley.
- Wild, C. J., e Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>