

## CARACTERIZAÇÃO GRÁFICA E A DISPERSÃO DA DISTRIBUIÇÃO DE DADOS UTILIZANDO O MÉTODO TRADICIONAL E O SOFTWARE “R”

Ailton Paulo de Oliveira Júnior – Flávia Helena Pereira

[drapoj@uol.com.br](mailto:drapoj@uol.com.br) - [fh\\_pereira18@yahoo.com.br](mailto:fh_pereira18@yahoo.com.br)

Universidade Federal do ABC, Brasil – Universidade Federal do Triângulo Mineiro, Brasil

Núcleo temático: Recursos para la enseñanza y aprendizaje de las matemáticas

Modalidade: Comunicación Breve

Nível educativo: Formación y actualización docente

Palavras chave: Dispersão; Software “R”; Ensino de Estatística; Ensino Superior.

### Resumo

*O presente trabalho tem como objetivo mostrar como alunos que cursaram ou que estejam cursando Probabilidade e Estatística no curso de Licenciatura em Matemática de uma universidade federal no Brasil, avaliam a sua aprendizagem em relação à representação gráfica das distribuições de frequência através dos conceitos de dispersão, através de duas metodologias de ensino (“utilizando papel milimetrado e lápis” e “utilizando o software estatístico R”). Desta forma, elaborou-se uma sequência didática para a apresentação dos conceitos básicos introdutórios de medidas que são necessárias para melhor entendimento da dispersão da distribuição dos dados através de duas metodologias, ou seja, a que consideramos como ensino tradicional (utilizando papel e lápis) e utilizando o software estatístico “R”. Para a avaliação foi aplicado um teste para identificação do domínio destes conteúdos, bem como um instrumento para identificar como os participantes perceberam as metodologias de ensino e solicitar que apresentem outras maneiras de aplicação dos conceitos em situações do cotidiano. Após a realização da pesquisa e análise dos resultados, pode-se verificar que o uso do software foi aceito pelos participantes, mas que o método tradicional também deve ser utilizado em sala de aula, ou seja, unir as duas metodologias para a apresentação dos conceitos.*

### Introdução

A sociedade de hoje exige que as pessoas sejam competentes para analisar, compreender e interpretar informações de vários eventos que acontecem ao seu redor, que envolvem elemento de incerteza ou variabilidade (por exemplo, eventos climáticos, eventos desportivos, jogos de azar, previsão financeira, seguros de vida ou danos). Como resultado, questões estatísticas têm mostrado presença no currículo da matemática escolar em muitos países, desde o ensino primário ao nível universitário.

Portanto, existem conceitos e medidas que nos permitem estabelecer maneiras mais práticas e fáceis de interpretar essas informações, e se forem estudadas adequadamente podem ser aplicados de alguma forma em situações em que vivemos, ou estudamos.

Desta forma, a variabilidade é o coração da estatística (Moore, 1990) e indispensáveis para sua existência (Watson, Kelly, Callingham & Shaughnessy, 2003).

Este conceito está relacionado com muitas ideias estatísticas fundamentais, sendo que Garfield e Ben-Zvi (2008, p. 203) observam que a compreensão das ideias de dispersão e variabilidade dos dados é componente chave para a compreensão do conceito de distribuição e é essencial para fazer inferências estatísticas.

Além destes aspectos, Ben-Zvi (2011) ao refletir sobre os elementos-chave para o ambiente de aprendizagem para o raciocínio estatístico, revela que um desses elementos é a integração de ferramentas tecnológicas no ensino. Para o autor, existe uma gama de ferramentas tecnológicas que podem auxiliar no desenvolvimento do raciocínio e da compreensão dos estudantes sobre a Estatística, como computadores, calculadoras gráficas, Internet, software estatístico e aplicativos da Web.

### **Procedimento Metodológicos**

O objetivo do trabalho foi o de indicar como os participantes da atividade avaliam a aprendizagem da variabilidade utilizando a representação gráfica das distribuições de frequência. Para tanto, foram utilizadas duas metodologias de ensino (“utilizando papel milimetrado e lápis” e “utilizando o software estatístico R”) em uma sequência didática para a apresentação dos conceitos básicos de medidas que são necessárias para o melhor entendimento da distribuição de frequência.

A atividade foi desenvolvida em dois encontros de 4 horas em junho de 2016 com 16 participantes, quais sejam: bolsistas do PET (Programa de Educação Tutorial) da UFTM; bolsistas do PIBID (Programa Institucional de Bolsas de Iniciação à Docência) da UFTM; e alunos do Mestrado em Educação da UFTM participantes do Grupo de estudos em Educação Estatística e Matemática - GEEM.

Foram elaborados e aplicados os seguintes instrumentos: (1) Sequência didática para a construção gráfica manual de uma distribuição de frequências e utilizando o software “R”, para mostrar a representação gráfica das distribuições de frequência para apreensão do

conceito de variabilidade; (2) Questionário para avaliação da atividade desenvolvida no tocante às sequências didáticas desenvolvidas.

Portanto, para avaliar e apresentar os resultados obtidos na sequência didática, tomamos, Garfield e Ben-Zvi (2005), que propõem um quadro de sete categorias ou "componentes de um modelo epistemológico" que organizou o conjunto de ideias e ações necessárias para uma compreensão completa de variabilidade. Seleccionamos as seguintes 2 primeiras categorias como as mais convenientes para analisar o nível cognitivo do conceito de variabilidade: (1) Desenvolver ideias intuitivas de variabilidade, que se referem a ela como uma característica onipresente nos fenômenos e dados obtidos a partir deles. Além disso, há fenômenos e conjuntos de dados que têm mais variabilidade do que outros e, finalmente, é uma característica da variabilidade de um conjunto de dados e não de os elementos isolados; (2) Descrever e representar a variabilidade: quando restrita a um conjunto de dados, a variabilidade é a dispersão destes, que pode ser representado graficamente para mostrar seus aspectos. A variabilidade pode ser medida de diferentes maneiras e expressa por um número (amplitude, desvio médio ou desvio-padrão).

### **Estudo do desenvolvimento das atividades**

Iniciamos a sequência didática propondo uma atividade “roteiro”, que foi entregue aos participantes para que resolvessem um problema com a participação dos pesquisadores.

A atividade apresentada na Figura 1, apresenta o problema proposto aos participantes.

Foram construídos o histograma (dados agrupados) e o seu polígono de frequências dos dados. Então, foram discutidos os resultados obtidos das medidas estatísticas e das representações gráficos. O objetivo desta discussão foi que os participantes pudessem apreender os conceitos que advém das representações gráficas.

**ATIVIDADE 1**

Em cerca de 2500 a.C., os gregos realizavam festivais esportivos em honra a Zeus, no santuário de Olimpia, o que originou o termo olimpíada. As Olimpíadas acontecem de quatro em quatro anos, onde atletas de centenas de países se reúnem num país sede para disputarem um conjunto de modalidades esportivas. A própria bandeira olímpica representa essa união de povos e raças, pois é formada por cinco anéis entrelaçados, representando os cinco continentes e suas cores. A paz, a amizade e o bom relacionamento entre os povos são os princípios dos Jogos Olímpicos que tiveram origem na Grécia Antiga. A primeira Olimpíada ocorreu no estádio Panathenáico em Atenas em 1896, na época da era moderna em Atenas, com a participação de 14 países. A Tabela 1 apresenta lista dos países com maior número de medalhas em toda história das Olimpíadas.

Tabela 1 - Lista dos países com maior número de medalhas na história das Olimpíadas.			
Países	Medalhas	Países	Medalhas
Estados Unidos	2548	Coreia do Sul	260
União Soviética	1204	Alemanha Ocidental	243
Grã-Bretanha	802	Bulgária	218
França	730	Cuba	208
Alemanha	719	Dinamarca	171
Itália	628	Tchecoslováquia	168
Suécia	604	Bélgica	144
Alemanha Oriental	519	Equipe Alemã Unida	137
China	518	Equipe Unificada	135
Rússia	496	Espanha	133
Austrália	477	Grécia	110
Hungria	464	Brasil	108
Finlândia	455	Ucrânia	101
Noruega	447	Iugoslávia	94
Canadá	405	Nova Zelândia	87
Japão	397	Bielorrússia	75
Países Baixos	332	Quênia	75
Suiça	308	África do Sul	70
Romênia	293	Argentina	70
Áustria	287	Jamaica	67
Polónia	275	México	62

Fonte: [https://pt.wikipedia.org/wiki/Quadro\\_de\\_medalhas\\_dos\\_Jogos\\_Ol%C3%AADmpicos](https://pt.wikipedia.org/wiki/Quadro_de_medalhas_dos_Jogos_Ol%C3%AADmpicos).

1. Construa uma distribuição de frequência utilizando o método tradicional.
2. Esboce o histograma de frequências e o polígono de frequências utilizando o método tradicional e interprete o seu comportamento.
3. Interprete a dispersão dos dados e também as medidas de assimetria e curtose e interprete o seu comportamento.

Figura 1: Problema “roteiro” para desenvolvimento da atividade.  
 Fonte: Elaborado pelos autores.

A Figura 2 apresenta a solução da atividade em que foi utilizado o papel milimetrado para a construção das distribuições de frequência. A atividade foi resolvida com todos os dados da tabela e também foi desenvolvida com a remoção do número de medalhas conquistadas em Olimpíadas pelos EUA.

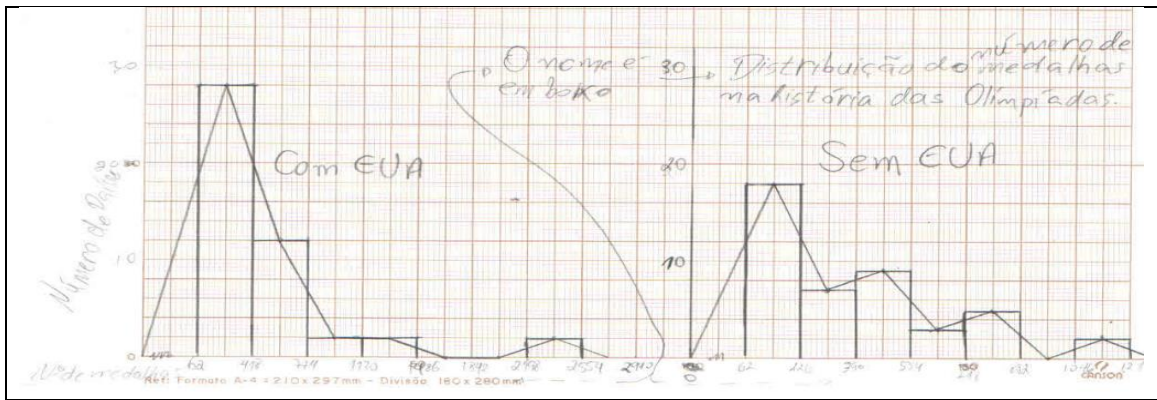


Figura 2: Gráficos comparativos da distribuição de frequências em relação à distribuição das medalhas por países na história das Olimpíadas.  
 Fonte: Elaborado pelos participantes da atividade.

Nesta atividade os alunos puderam observar a diferença entre as distribuições de dados e a forma do polígono de frequência quando utilizamos todos os dados e quando removido o número de medalhas referente aos EUA.

O desenvolvimento desta sequência possibilitou aos participantes a visualização do gráfico e percepção intuitiva de medidas estatísticas. E a partir desta atividade puderam perceber que o tratamento dos dados é de extrema importância para a apropriação do conhecimento do comportamento dos dados.

Após este momento iniciamos o processo de desenvolvimento das atividades com o *software R*, com a mesma atividade.

Consideramos aqui que com a análise gráfica dos dados, a tecnologia pode auxiliar através de uma variedade de gráficos que permitem que os alunos consigam investigar conceitos-chave através da visualização.

É possível construir, rapidamente, histogramas, gráficos de dispersão, polígonos de frequência, distribuições de probabilidade, distribuições de amostragem, etc. Também é possível manipular os dados, a fim de observar as mudanças nas representações gráficas, conseguindo assim remover o caráter abstrato para muitos conceitos.

Assim, os participantes receberam uma apostila com o passo a passo para a utilização do *software R* na resolução da atividade proposta. Também foi orientado que os participantes baixassem o *software* no computador e receberam as orientações sobre como utilizá-lo.

Ainda nesta atividade, agora utilizando o *software “R”*, os alunos puderam observar a diferença entre as distribuições de dados e a forma do polígono de frequência, inicialmente, quando foram utilizados todos os dados (Figura 3).

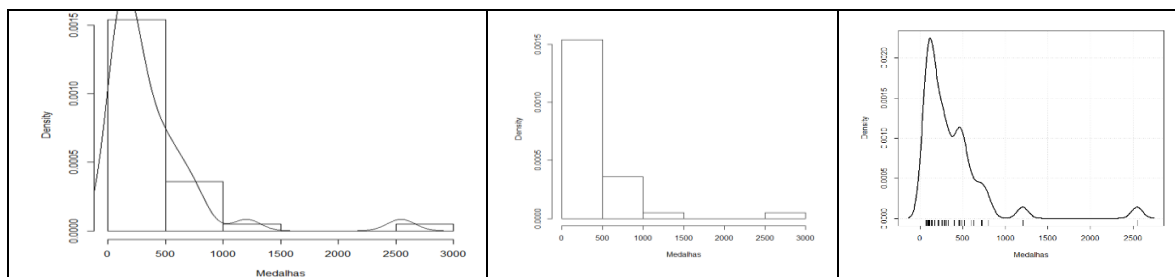


Figura 3: Gráficos das atividades com medalhas (todos os dados).

Fonte: Elaborado pelos autores da saída do software R.

É possível analisar o comportamento do polígono com base no histograma, a qual mostra o comportamento dos dados, indicando que há uma grande dispersão dos dados. O que motivou a continuidade da atividade com a retirada de observações que se mostram discrepantes em relação ao grupamento total dos dados. Assim, foi proposto retirar o número de medalhas dos Estados Unidos e da antiga União Soviética (Figura 4).

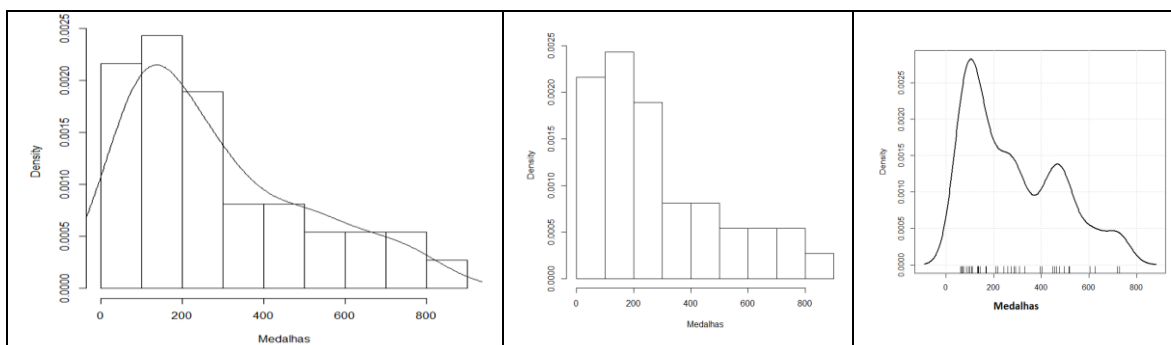


Figura 4: Gráficos Atividades com Medalhas (exceto Estados Unidos e União Soviética).

Fonte: Elaborado pelos autores da saída do software R.

Acreditamos que com esta atividade, pode-se examinar a variabilidade dentro de um grupo (observando como os dados variam dentro de um ou mais conjuntos de dados) e a variabilidade entre os grupos (a variabilidade das medidas usadas para resumir e comparar conjuntos de dados) e distinguir entre esses dois tipos da variabilidade.

Além disso, podemos listar os seguintes aspectos para dar suporte à importância desta atividade: (1) Os gráficos de dados mostram como estes variam e podem revelar padrões para ajudar a descobrir o comportamento da distribuição dos dados e identificar aspectos que possam causar problemas para a sua identificação; (2) Diferentes gráficos podem revelar diferentes aspectos da variabilidade de um conjunto de dados, por isso é importante estudar mais do que um único gráfico de um conjunto de dados.

Acreditamos ser essencial solicitar aos alunos para desenhar gráficos de distribuições para o mesmo conjunto de dados e/ou diferentes conjuntos de dados, já que com este procedimento, permite-se visualizar a variabilidade dos dados.

Em particular, esta atividade mobilizou recursos, como diferentes objetos estatísticos ou matemáticos por meio de representações numéricas (medidas de tendência central e de dispersão), tabular (representação resumo dos dados), gráficas (fotografia do comportamento dos dados) e simbólicas (conceitos).

E que por sua vez pode gerar ambientes favoráveis para o surgimento de variadas experiências de aprendizagem através de uma abordagem concreta e visual, permitindo, intuitivamente, iniciar o contato com conceitos e também a descoberta de relações complexas entre esses tipos de objetos.

Após a aplicação das sequências didáticas foi aplicado instrumento para avaliar a aceitação das metodologias de ensino. Foi perguntado qual método foi o preferido no desenvolvimento da atividade e solicitado que fosse justificada a opção.

Assim, a metade dos participantes das atividades que responderam ao instrumento de pesquisa consideram que a utilização da tecnologia, ou seja, utilizar o *software* “R” para a solução dos problemas propostos, contribui de forma efetiva para o desenvolvimento da apreensão dos conteúdos abordados. Alguns destes depoimentos são apresentados a seguir:

Utilizando o recurso do *software* R é uma aula diferente, em outro ambiente. Esta locomoção atrai a curiosidade do aluno, no meu caso, facilita a construção do gráfico a partir dos dados encontrados. Outro fator interessante foi referente ao final da atividade onde se calculava o exercício usando o “R” e então podíamos desenhar o gráfico na folha, utilizando os dados encontrados no *software*. **(PET 1 – 21 anos)**

Pela rapidez com que você consegue obter o gráfico e as informações necessárias para calcular os coeficientes, também pela exatidão que você consegue o gráfico. Acredito que também seria interessante calcular os coeficientes no *software* e mostrar a localização da média, moda, no gráfico. **(PET 2 – 21 anos)**

Cabe ainda considerar que a média de idade deste grupo é de 20,5 anos, com desvio padrão 1,5 anos, ou seja, de acordo com o coeficiente de variação, a dispersão das idades é considerada baixa (7,3%). Observa-se que o grupo, em média, é jovem.

Outro fator é que a composição deste grupo é de 4 participantes do PIBID Matemática e 2 (dois) do PET Matemática, ou seja, alunos em formação inicial. A distribuição do sexo é 50% para homens e mulheres.

O depoimento dos participantes vai ao encontro do que afirma Lévy (2008), ou seja, que a era atual das tecnologias da informação e comunicação estabelece uma nova forma de pensar sobre o mundo que vem substituindo princípios, valores, processos, produtos e instrumentos que mediam a ação do homem com o meio.

Observa-se ainda que a outra metade dos participantes declarou que utilizar a tecnologia, no caso, o *software* “R” contribui de forma positiva para o desenvolvimento da apreensão dos

conteúdos abordados, mas destacam também ser importante trabalhar os conceitos de forma “tradicional”. Alguns destes depoimentos são apresentados a seguir:

O uso do software facilitou bastante, com ele calculando todos os valores e gerando o gráfico, mesmo se tiver que fazer o gráfico manualmente, o entendimento da questão já está mais clara. Mesmo para quem ainda não cursou a disciplina, o entendimento foi facilitado com o uso do software. **(PIBID 5 – 36 anos)**

A visualização é mais fácil, os dados são calculados mais rapidamente, mas a explicação dos conceitos pode ser feita do método tradicional. Concluindo, é interessante trabalhar os dois métodos em conjunto. **(PET 3 – 25 anos)**

Observamos que a média de idade deste grupo é de 29,3 anos, com desvio padrão 6,5 anos, ou seja, de acordo com o coeficiente de variação, a dispersão das idades é considerada moderada (22,2%).

Destacamos que a diferença média das idades entre os grupos é de  $(29,3 - 20,5) = 8,8$  anos. Assim, percebe-se que grupo que considera a associação das duas metodologias de ensino é 8,8 anos mais velho. Percebe-se que é um grupo mais experiente e com vivências maiores no dia a dia e em sua formação.

O depoimento do segundo grupo nos leva a pensar que há que existir um bom senso na relação ensino dito tradicional e ensino com suporte tecnológico. Não podemos levá-la a extremos, onde, por exemplo, todos os alunos tendo computadores para utilizar em sala de aula, não sendo obrigatório o ensino da letra cursiva. É preciso haver uma mescla entre o tradicional método de ensinar, ou seja, utilizando papel e lápis, e os avanços da tecnologia, como a utilização de um *software* estatístico.

Além dessa avaliação, perguntamos se após a realização das atividades os participantes compreenderam os conceitos de dispersão de dados, considerando 1 para uma menor aceitação até 5 que significa uma maior aceitação. Os resultados mostram que 91,67% marcaram “4” ou “5”, ou seja, indicando que estes consideram que entenderam o conceito de variabilidade a partir do desenvolvimento das atividades propostas.

Devemos também considerar como ideal um ensino usando diversos meios, um ensino no qual todos os meios deveriam ter oportunidade, desde os mais modestos até os mais elaborados: desde o quadro, os mapas e as transparências de retroprojetor até as antenas de satélite de televisão. Ali deveriam ter oportunidade também todas as linguagens: desde a



palavra falada e escrita até as imagens e sons, passando pelas linguagens matemáticas, gestuais e simbólicas (Sancho, 2001, p. 136).

Acreditamos que devemos aplicar um diferencial tecnológico em sala de aula, sem deixar de ser educativo, mantendo assim o aluno interessado e o professor com o controle, não só da sala de aula, mas também das ferramentas à disposição para levar conhecimento.

### **Considerações Finais**

O grau de compreensão dos participantes sobre a variabilidade convergiu para a maneira pela qual perceberam a distribuição de dados; no entanto, o problema proposto também foi importante para a compreensão do comportamento dos dados.

A atividade pode mostrar o auxílio da informática no ensino de Estatística. Por um lado, o seu apoio na redução dos cálculos para ensinar de uma maneira tradicional, e por outro lado, a visualização de conceitos. E também a possibilidade de simulações, que se mostrou vital na geração de situações de ensino que se tornaria difícil realizá-los em sala de aula por outros métodos.

Assim, após a realização da pesquisa e análise de resultados, verificou-se que o uso do *software* foi aceito pelos participantes. Mas, não devemos considerar a possibilidade de descartar o uso do método tradicional (utilizando papel milimetrado e lápis) em sala de aula, uma vez que, é possível unir as duas metodologias para a explicação dos conceitos.

### **Referências**

- Ben-Zvi, D. (2011). Statistical reasoning learning environment. *EM TEIA – Revista de Educação Matemática e Tecnológica Ibero-americana*, 2(2), 1-13.
- Garfield, J. & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92-99.
- Garfield, J. & Ben-Zvi, (2008). D. Developing students' statistical reasoning: Connecting research and teaching. New York: Springer.
- Levy. P. (2008). *Cibercultura*. Trad. Carlos Irineu da Costa. São Paulo: Editora 34.
- Moore, D. (1990). Uncertainty. In: Steen, L. A. (Ed.). *On the shoulders of giants: New approaches to numeracy*. pp. 95-137. Washington, DC: National Academy Press, 1990.
- Sancho, J. M. (2001). (Org.). *Para uma tecnologia educacional*. Porto Alegre: Artmed.
- Watson, J., Kelly, B., Callingham, R., & Shaughnessy, M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, 34(1), 1-29.