

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

PUC-SP

Marcelo Marcos Bueno Moreno

**Ensino e aprendizagem de estatística com ênfase na variabilidade:
um estudo com alunos de um curso de licenciatura em Matemática**

MESTRADO PROFISSIONAL EM ENSINO DE MATEMÁTICA

**São Paulo
2010**

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

PUC-SP

Marcelo Marcos Bueno Moreno

**Ensino e aprendizagem de estatística com ênfase na variabilidade:
um estudo com alunos de um curso de licenciatura em Matemática**

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de **MESTRE PROFISSIONAL EM ENSINO DE MATEMÁTICA**, sob orientação da **Profa. Dra. Cileda de Queiroz e Silva Coutinho**.

**São Paulo
2010**

Banca Examinadora

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta Dissertação por processos de fotocopiadoras ou eletrônicos.

Assinatura: _____ **Local e Data:** _____

“É preciso convencemo-nos da prioridade da ética sobre a técnica, do primado da pessoa sobre as coisas, e da superioridade do espírito sobre a matéria.

A causa do homem será servida se a ciência se aliar à consciência. O homem de ciência ajudará verdadeiramente a humanidade se conservar o sentido da transcendência do homem sobre o mundo e de Deus sobre o homem.”

Discurso do Papa João Paulo II à UNESCO em junho de 1980

AGRADECIMENTOS

À Providência Divina que, a partir de minha iniciativa, fez com que uma corrente de acontecimentos surgisse a favor de minha chegada até aqui. Senhor, que eu seja digno de sua graça!

A minha mãe Joana, guerreira incansável da causa de educar seus filhos nas circunstâncias difíceis da vida.

A minha esposa Giovana, pelo amor que tem me dedicado a todo momento.

À minha orientadora Professora Dra. Cileda Coutinho, pela paciência, pelo auxílio sempre presentes e relação tão respeitosa.

À professora Dra. Maria José F. da Silva, por fazer parte da banca e pelas diversas admoestações que tanto me ajudaram.

À professora Dra. Lisbeth Cordani, por participar da banca e ter estendido suas mãos, desde 2003, quando o mestrado era somente um sonho. Agradeço por sua luta incansável pela escola pública e pelo apoio àqueles que acreditam nela. Seus olhos sempre vão brilhar e acenderão essa chama em outras pessoas.

À professora Silmara Mattiazzo, Secretária Municipal de Educação de Caraguatatuba (2005-2008), por ter acreditado em meu trabalho, mesmo sem eu fazer parte do meio político de seu partido.

À CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, por ter concedido a Bolsa.

À professora Aracelia Pereira, coordenadora da Licenciatura em Matemática da Universidade Módulo, por ter aberto as portas da universidade para nosso experimento.

À direção da Universidade Módulo, por ter autorizado nossa pesquisa em um dos seus cursos, e aos alunos da Licenciatura em Matemática pela colaboração.

À professora Mr. Diva V. Novaes, por ter me auxiliado durante o experimento.

Ao amigo Jacinto Ordem, pelo companheirismo e determinação. Parabéns Jacinto, você deixou aqueles que ama em Moçambique e veio nos dar exemplos de luta em nosso meio.

RESUMO

Tendo em vista a novidade da estatística nos currículos da Educação Básica, a escassez de pesquisas sobre a variabilidade descrita por vários autores, a dificuldade dos alunos na compreensão do desvio-padrão e na articulação dos conhecimentos matemáticos e estatísticos para aplicá-los em situações novas, objetivou-se estudar as características das atividades de uma formação com alunos de um curso de licenciatura em Matemática, que favoreçam a apreensão da variabilidade em um conjunto de valores. Uma Sequência Didática foi elaborada, segundo a proposta de construção de cenários de aprendizagem que se baseia na articulação de diversas teorias de aprendizagem. A sequência foi dividida em três partes: introdução e construção do conceito de média aritmética; construção do conceito de variabilidade pelo uso de medidas de dispersão e aplicação do conceito de variação para tomada de decisões. Das tarefas aplicadas, escolhemos três que fossem representativas para a análise teórica e análise das produções dos participantes, seguindo os pressupostos da Engenharia Didática, que foi a metodologia utilizada nesta pesquisa. O resultado de outras pesquisas relacionadas ao presente tema, no que se refere à interpretação da variação, foi confirmado. A variação era considerada, apenas entre os dados e não em relação a uma medida de tendência central. Tal ocorrência consistiu em um fator de dificuldade para a interpretação das medidas de variabilidade. A abordagem dos intervalos com base em k desvio-padrão da média ajudou os participantes a darem sentido a esta medida de dispersão, e a comparação entre distribuições com a mesma média e diferentes dispersões auxiliou-os a perceberem a insuficiência da média, como medida-resumo e a importância das medidas de variação. A utilização de situações nas quais o conhecimento visado é a solução mais adequada para o problema e a estratégia da organização de tarefas cujos alunos pudessem mobilizar conhecimentos prévios para construir os novos, mostraram-se úteis para a construção dos conceitos. Ao final da formação, os alunos (futuros professores) mostraram uma visão mais ampla para fazer uma Análise Exploratória de Dados e preocupavam-se em considerar a variabilidade.

Palavras-chave: Variabilidade. Formação inicial de professores. Educação Estatística. Sequência Didática.

ABSTRACT

Considering that statistics has been newly included in the curriculum of primary and secondary education, the scarcity of research studies about variability described by various authors, and students' difficulties in understanding standard deviations and in articulating mathematical and statistical knowledge to apply them in new situations, the objective was to study the characteristics of teaching activities that promote the comprehension of variability in a set of values for students in a Mathematics teaching degree program. A Didactic Sequence was elaborated according to the proposal of constructing learning scenarios that were based on the articulation of the various theories of learning. The sequence was divided into three parts: introduction and construction of the concept of arithmetic mean, construction of the concept of variability through the use of dispersion measures, and application of the concept of variation for decision-making. Of the applied tasks, we chose three that were representative for theoretical analysis and analysis of the participants' work, according to the assumptions of Didactic Engineering, which was the methodology used in this study. The result of other research studies related to this subject, regarding the interpretation of variation, was confirmed. The variation was considered, only between the data, and not in relation to a measure of the central tendency. This occurrence consisted of a difficulty factor for interpreting the variability measures. The interval approach based on k standard deviations from the mean helped the participants give meaning to this measure of dispersion, and the comparison between distributions with the same mean and different dispersions helped them to perceive the insufficiency of the mean as a summary measure and the importance of measures of variation. The use of situations in which the knowledge sought is the most appropriate solution to the problem and the strategy for organizing tasks in which the students could mobilize prior knowledge to construct new knowledge were shown to be useful in constructing the concepts. At the end of the course, the students (future teachers) displayed a broader vision for performing an Exploratory Data Analysis and were concerned with considering variability.

Keywords: Variability. Pre-service teacher training. Statistics Education. Didactic Sequence.

SUMÁRIO

INTRODUÇÃO	10
CAPÍTULO I: REVISÃO BIBLIOGRÁFICA	
1.1 Pensamento e Raciocínio Estatístico	12
1.2 Estudo do desenvolvimento da concepção de variação em universitários.....	16
1.3 Estudo com alunos do Ensino Médio.....	19
CAPÍTULO II: PROBLEMÁTICA	
2.1 Justificativa.....	22
2.2 Objetivos e Questão de Pesquisa	24
2.3 Procedimentos Metodológicos	24
2.3.1 Os procedimentos adotados.....	25
2.3.2 A Engenharia Didática	26
CAPÍTULO III: QUADRO TEÓRICO	
3.1 As dimensões de análise de conteúdos a ensinar e a elaboração de cenários.....	30
CAPÍTULO IV: ANÁLISES PRÉVIAS	
4.1 Conceitos de estatística abordados.....	40
4.1.1 Medidas de posição	40
4.1.2 Medidas de dispersão	45
4.1.3 Desigualdade de Tchebichev.....	50
4.1.4 A Regra Empírica para determinar o percentual de dados contidos no intervalo criado com base na subtração e soma de desvios-padrão à média.....	51
4.1.5 Evento Raro.....	52
4.1.6 Z-escore.....	53
4.1.7 O Gráfico de Pontos (Dotplot).....	54
4.2 Os Parâmetros Curriculares Nacionais e a variabilidade	55
4.3 O Pensamento Estatístico	57

4.4 A Importância da Variabilidade	59
 CAPÍTULO V: A SEQUÊNCIA DIDÁTICA	
5.1 Definição de situação, situação didática, tarefa e atividade	66
5.2 Os sujeitos da pesquisa	67
5.3 Descrição da aplicação.....	67
5.4 Síntese e objetivos dos eixos em que foram divididas as situações	68
5.4.1 Tarefas agrupadas, conforme as noções visadas	70
5.5 Análise de algumas tarefas	74
5.5.1 Análise da tarefa 20	74
5.5.1.1 Análise teórica	76
5.5.1.2 Análise <i>a posteriori</i>	81
5.5.2 Análise da tarefa 38	86
5.5.2.1 Análise teórica	87
5.5.2.2 Análise <i>a posteriori</i>	90
5.5.3 Análise da tarefa 43	96
5.5.3.1 Análise teórica	97
5.5.3.2 Análise <i>a posteriori</i>	100
5.5.4 Comparação entre a avaliação do primeiro dia de formação com a do último dia.....	104
 CONSIDERAÇÕES FINAIS	 108
 REFERÊNCIAS.....	 114
 APÊNDICE A: TAREFAS PROPOSTAS NA FORMAÇÃO	 118
APÊNDICE B: QUESTIONÁRIO DIAGNÓSTICO	154
APÊNDICE C: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO....	156

INTRODUÇÃO

Este trabalho faz parte de um projeto maior intitulado Processos de Ensino e Aprendizagem da Estatística – PEA-ESTAT - coordenado pelo professor Dr. Saddo Ag. Almouloud, financiado pela FAPESP e desenvolvido pelo grupo de pesquisa PEA-MAT.

A presença da Estatística nos currículos escolares de diversos países vem aumentando nos últimos anos. No Brasil, esta presença ficou marcada de maneira mais acentuada com a implementação dos Parâmetros Curriculares Nacionais para o Ensino Fundamental (BRASIL,1997) e para o Ensino Médio (BRASIL, 2000).

Dessa forma percebemos a importância deste tema para a formação do cidadão, tendo em vista as exigências de uma sociedade marcada pela necessidade de ler, analisar, organizar e sintetizar dados presentes no cotidiano, para tomada de decisão. Para formar um cidadão crítico, é também necessário instrumentalizá-lo com ferramentas de análise quantitativa e, neste aspecto, a estatística exerce um papel capital.

Entretanto, em nossa prática docente em escolas de Ensino Fundamental e Médio, vemos o ensino de Estatística reduzido a gráficos e tabelas, com pouca ou nenhuma Análise Exploratória de Dados. Quando se usa alguma medida de tendência central, ela se resume ao uso do algoritmo da média¹ sem qualquer articulação com alguma medida de dispersão, tal como a amplitude, o desvio-médio, o desvio-padrão ou o intervalo interquartil, ou seja, uma abordagem estática das representações de um conjunto de dados. Faz-se necessária uma abordagem mais dinâmica e participativa, com desenvolvimento de projetos nos quais os alunos sejam protagonistas e pesquisadores, planejem suas pesquisas, coletem os dados, analisem e divulguem-nos, sob a orientação do professor.

Outra questão preocupante que percebemos, em nossa prática docente, mas também na bibliografia pesquisada, é o abandono pelos currículos e educadores matemáticos da alma da estatística: a variabilidade². Ela tem sido negligenciada.

¹ Toda vez que utilizarmos o termo média, estaremos nos referindo à média aritmética.

² Neste trabalho, utilizaremos os termos variabilidade e variação como sinônimos.

Pautados na investigação de Meletiou (2000), fazemos a hipótese de que tal omissão possa ser a origem das diversas dificuldades dos alunos quanto ao pensamento e raciocínio estatístico. Alguns dos problemas, apontados pela literatura, são as dificuldades dos alunos com interpretação e aplicação do desvio-padrão, a predominância de um tipo de pensamento determinista e, por consequência, a dificuldade para realizar uma análise global dos dados.

Para auxiliar a percepção da importância da variabilidade, sua medição e aplicação na análise de dados, organizamos uma sequência didática, apoiada nas propostas de cenários de ensino e aprendizagem de Matemática de Robert (1998) e aplicamos com alunos de licenciatura em Matemática, pois acreditamos que a boa formação dos futuros professores aumentará as chances de termos alunos críticos e preparados para os requerimentos da sociedade atual.

Assim, queremos estudar as características das atividades de uma formação de alunos de um curso de licenciatura em Matemática que favorecem a apreensão da variabilidade quando se analisa um conjunto de dados.

O primeiro capítulo foi reservado para a revisão bibliográfica sobre o pensamento, raciocínio e concepções estatísticas, com foco na variabilidade, entre professores, estudantes universitários e alunos do Ensino Médio. No segundo capítulo, dissertamos sobre a importância do tema, nossa questão de pesquisa, os procedimentos metodológicos e a metodologia de análise denominada Engenharia Didática. Em seguida, no terceiro capítulo, apresentamos o fundamento teórico sob o qual elaboramos nossa sequência: as dimensões de análise de conteúdos a ensinar e a elaboração de cenários de aprendizagem de Robert (1998). O quarto capítulo traz a definição dos conteúdos abordados e a análise didática dos mesmos, o estudo da variabilidade abordada nos Parâmetros Curriculares Nacionais (PCN) e uma reflexão sobre a importância da variabilidade na estatística. No quinto capítulo, analisamos três tarefas de nossa sequência, seguindo os pressupostos da Engenharia Didática e comparamos a avaliação feita no começo do curso com a realizada no último dia.

Finalmente, concluímos o texto com as considerações finais, com base na confrontação dos resultados observados com nossa questão de pesquisa.

CAPÍTULO I – REVISÃO BIBLIOGRÁFICA

Neste capítulo abordamos pesquisas realizadas com professores de matemática, alunos de licenciatura e do Ensino Médio.

1.1 Pensamento e Raciocínio Estatístico: um estudo com professores de matemática

O trabalho de Silva (2007) foi motivado pela dificuldade que os alunos de graduação sentiam na compreensão do conceito de desvio-padrão. Diante disto, a autora pesquisou o raciocínio sobre variação e variabilidade nas etapas do ciclo investigativo do pensamento estatístico³. Foram estudados nove professores de Matemática e dois graduandos na disciplina, por meio de uma formação continuada, ministrada com carga horária de 48h. Além da investigação com professores, o objeto variação/variabilidade foi analisado nos livros didáticos de Matemática para o Ensino Médio.

Tendo em vista a importância da variação no Pensamento, Raciocínio e Letramento Estatístico, a autora apresenta as definições desses temas que sintetizaremos nas próximas linhas.

Para Silva (2007), Pensamento Estatístico são as estratégias mentais utilizadas pelo indivíduo para tomar decisão em toda a etapa do ciclo investigativo. Os elementos do Pensamento Estatístico são: a necessidade de dados, a importância da produção de dados, a onipresença da variabilidade e sua medida e modelagem.

Assume também o raciocínio estatístico como sendo “a maneira com que as pessoas raciocinam com ideias estatísticas e como percebem a informação estatística”. (GARFIELD, 2002 apud SILVA, 2007, p.33).

Finalmente, a autora define o letramento estatístico nos termos de Gal (2002 apud SILVA, 2007), como a competência das pessoas para interpretar e avaliar criticamente a informação estatística, os argumentos relacionados aos dados ou a

³ A autora utiliza os ciclos propostos por Wild; Pfannkuch (1999) que são: o ciclo investigativo, os tipos de pensamento, o ciclo interrogativo e as disposições.

fenômenos estatísticos, é a competência das pessoas para discutir ou comunicar tais informações.

Outro esclarecimento importante para o trabalho da autora é a diferença entre os termos variabilidade e variação. Para ela, variabilidade é a característica da entidade que é observável, e variação é a descrição ou medida desta característica.

A questão de pesquisa principal do trabalho foi: Qual o nível de raciocínio de variação utilizado pelo professor de Matemática em diferentes etapas do ciclo investigativo do pensamento estatístico?

Em seguida, apresentaremos, sinteticamente, alguns resultados de Silva (2007) que serão úteis para este trabalho.

A autora identificou pontos frágeis na significação da estatística para os professores, pois observou em seus discursos que eles valorizam as pesquisas, como estratégia para ensinar estatística, mas os conteúdos abordados restringiam-se, praticamente, a gráficos e tabelas. Quando lhes foi solicitado resumir um conjunto de dados para futura divulgação, admitiram com facilidade a presença da variabilidade entre os dados, mas “não houve a percepção de que os valores poderiam variar em torno da média e não surgiu a necessidade de uma medida de variação.” (SILVA, 2007, p. 218).

Conforme a autora, a falta de percepção da variação em torno da média é reforçada pela associação do desvio-padrão pequeno com regularidade ou homogeneidade, encontrado na maioria dos livros didáticos analisados por ela.

Após a abordagem de alguns aspectos da variação, como: a existência da variação em um conjunto de dados, a limitação da média para representá-los, a necessidade de medir a variação e a utilização do desvio-padrão (DP)⁴, como um instrumento para medir a dispersão, Silva (2007) observa que a maioria dos professores não considerou tais aspectos em suas análises. Mas, assinala que quatro professores interpretaram o desvio-padrão como um intervalo em torno da média, mostrando assim uma evolução quanto ao nível de raciocínio de variabilidade, se considerado o estado inicial do conhecimento desses professores no início da formação. Para a autora, esta interpretação pode facilitar a

⁴ Utilizaremos o símbolo DP para representar o desvio-padrão

compreensão dos intervalos criados, a partir de k desvios-padrão em torno da média, o que é considerado por ela, como um grande avanço no raciocínio sobre variação.

O estudo de Silva (2007) aponta que os professores demonstraram inicialmente ausência de significação da noção de desvio-padrão e, em sua prática de ensino de estatística, limitavam-se à distribuição de frequência e sua representação gráfica. Portanto, não utilizavam como ferramentas de análise medidas de tendência central nem de dispersão. O raciocínio sobre variabilidade surgiu com naturalidade em contraste com o raciocínio de variação que apresentou dificuldades para ser apreendido.

A autora apresenta alguns aspectos que dificultaram a apreensão da noção de variação:

- A interpretação da média como moda foi um fator impeditivo para a percepção da necessidade de uma medida de variação;
- Interpretar variabilidade como regularidade, como diferença entre os dados ou homogeneidade (reforçada pelos livros didáticos); e
- Utilização de uma única variável que não permite a comparação entre grupos.

A questão central da pesquisa de Silva (2007) é a identificação dos níveis de raciocínio de variação utilizados pelo professor de Matemática que não é objeto do nosso trabalho, todavia a identificação de algumas limitações quanto à compreensão dos conceitos estatísticos de variação foram úteis para planejarmos nossa sequência didática, uma vez que nosso público constitui-se de alunos que serão professores. A leitura deste trabalho colaborou, para que construíssemos atividades enfatizando a importância da variação em torno de um ponto e não apenas a variação entre os dados, além de nos fornecer diversas referências bibliográficas que foram úteis em nossa pesquisa.

Outro estudo com professores foi a investigação de Canossa (2009), voltada à Estatística Descritiva, que teve como público alvo professores de Matemática do Ensino Médio. A justificativa da escolha desse público foi a falta de preparo de muitos deles para o ensino de estatística, visto que parte considerável dos

professores não teve esse conteúdo em sua formação inicial ou teve de forma superficial e tecnicista, conforme resultados de pesquisas analisadas pelo autor.

O objetivo da pesquisa era identificar as características didáticas de uma formação continuada para professores do Ensino Médio, visando ao trabalho de conceitos de mediana e quartis com seus alunos, auxiliados pelo Gráfico de Pontos (Dotplot) e pelo Box-Plot.

O autor assinala que a maioria dos professores, quando se propõem a ensinar estatística a seus alunos, limita-se à aplicação das fórmulas da média, variância e desvio-padrão, sem a abordagem dos conceitos de mediana e quartis, além de desconhecerem Dotplot e o Box-Plot. Para ele, uma consequência desse fato é que a compreensão do conceito de variabilidade e suas aplicações na análise de dados fica distante da realidade de muitas salas de aula.

Para responder à sua questão de pesquisa, o autor aplicou um questionário diagnóstico com um grupo de professores, elaborou uma formação continuada baseada nas constatações do questionário e observou a aula de uma professora para avaliar a evolução das concepções e das práticas docentes sobre o tema.

Canossa (2009) conclui seu trabalho, defendendo que, para uma boa formação continuada de professores, visando ao trabalho com conceitos de mediana e quartis, devemos ter uma organização didática que desenvolva o letramento e o raciocínio estatístico desses docentes. Para isso, assinala a importância dos professores (e alunos) construírem seu próprio banco de dados, sendo capazes de realizarem um tratamento adequado desses dados, utilizando várias maneiras de representá-los.

Em nosso trabalho, empregamos estratégias semelhantes às do autor no que se refere ao uso do Gráfico de Pontos, como fase intermediária para introdução da representação gráfica dos quartis. Tendo o Gráfico de Pontos (exemplo na figura 1) como referência, os quartis eram localizados com segmentos de retas verticais, depois era desenhada uma caixa e, em seguida contavam-se quantos elementos havia entre estas medidas de posição.

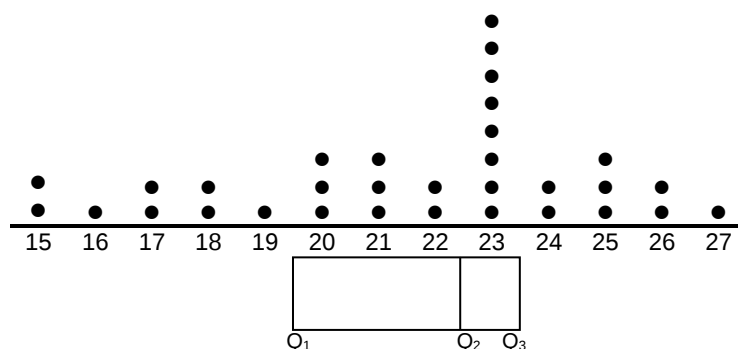


Figura 1 – Utilização do Gráfico de Pontos para representar os quartis.

1.2 Estudo do desenvolvimento da concepção de variação em universitários

Meletiou (2000) realizou uma pesquisa com 33 estudantes universitários, com pouco conhecimento de Matemática em um curso que durou 40 horas distribuído em 5 semanas. A investigação pretendia compreender os obstáculos para aprendizagem de estatística e utilizar este entendimento para encontrar formas de criar ambientes que facilitassem a aprendizagem.

A conjectura que estabeleceu o fio condutor de toda a pesquisa, baseou-se na concepção de que a ajuda fornecida aos alunos para aprimorarem suas intuições sobre a variação e sua relevância para a estatística melhorariam a compreensão dos conceitos que compõem esta ciência. Portanto, os currículos de estatística deveriam valorizar os conhecimentos que os estudantes possuem e dar mais ênfase à variação. O levantamento bibliográfico feito pela autora indicou que, apesar do papel central da variação, este tema era negligenciado.

A abordagem utilizada pela autora na formação baseou-se na concepção que entende a investigação estatística, como um processo dinâmico e não linear. A autora utilizou a estratégia de apresentar aos alunos situações que requeriam ferramentas estatísticas para resolver o problema. Diante dessas situações, eles eram instigados a dar suas opiniões, levantando hipóteses a respeito da melhor

solução aos problemas. No processo, os conceitos deveriam ser construídos, ajudando os alunos a expandirem, sobretudo suas compreensões sobre variação.

A análise da primeira etapa do curso identificou que os alunos percebiam a variação nos dados, mas apresentavam limitações quanto ao uso das ferramentas estatísticas para lidar com elas.

Meletiou (2000) relata que, embora os alunos reconhecessem a variação entre as amostras, tinham a tendência de subestimar seus efeitos, uma vez que, ao fazerem juízos sobre situações da vida real, ignoravam os perigos das amostras pequenas e não hesitavam utilizá-las, como base para inferências. Esta tendência era mais acentuada quando a questão proposta era contextualizada, pois no momento em que as questões eram substituídas por problemas típicos de probabilidade utilizados nas aulas (lançamento de moedas, jogo de dados, etc.), eles pensavam probabilisticamente, ou seja, se a questão referia-se a uma situação contextualizada, havia uma tendência ao pensamento determinístico⁵.

Quanto ao estudo da distribuição normal e do evento raro, a autora observou que a estratégia de relacionar esta distribuição, o desvio-padrão e a determinação do evento raro mostraram-se úteis para preparar os alunos à inferência estatística. Segundo a mesma, a forma como o instrutor⁶ aproximou os estudantes da distribuição normal não foi a abordagem típica utilizada nos cursos introdutórios de estatística, baseada no ensino formal das propriedades dessa distribuição, mas procurava ajudá-los a entender por que poderiam usar uma distribuição normal para modelar uma variável determinada e de que forma isso seria útil.

Na pesquisa de Meletiou (2000), a abordagem da distribuição normal era importante porque seria utilizada na inferência estatística que fazia parte do conteúdo a ser trabalhado com os alunos. Nossa pesquisa restringe-se à estatística descritiva, voltada para conteúdos a serem trabalhados por professores do Ensino Básico e, portanto, não abordaremos este conteúdo. A estratégia para analisar,

⁵ Segundo Ferreira (2003), “determinismo é um princípio segundo o qual todo fato tem uma causa e, nas mesmas condições, as mesmas causas produzem os mesmos fatos, o que implica a existência de leis específicas que regem fatos e causas.” Assim, entendemos o pensamento determinístico como aquele que busca uma explicação causal completa aos fenômenos naturais, excluindo a possibilidade do acaso.

⁶ A autora atuou como observadora no curso.

quando um evento é raro por intermédio do desvio-padrão, é útil para mostrar uma aplicação desta medida de dispersão e para preparar o aluno para a estatística inferencial e, portanto, será considerada em nossas atividades.

Em nossa sequência didática, elaboramos algumas tarefas baseadas nas ideias de evento raro, analisadas por Meletiou (2000).

A investigação feita pela autora permitiu observar que até alunos que haviam passado por formação em estatística básica, tinham poucas intuições sobre fenômenos aleatórios e tendiam a pensar deterministicamente. Isso a levou a concluir que a razão dessas dificuldades poderia estar no descuido de uma abordagem da variação nas aulas da disciplina.

A respeito dos princípios úteis para se construir conexões entre a Matemática formal e as intuições informais dos alunos, os seguintes pontos foram destacados:

- a) Complementaridade da teoria e da experiência: as ferramentas estatísticas não devem ser tratadas separadamente, mas, como uma síntese do conhecimento estatístico, do contexto e das informações sobre os dados, a fim de produzir implicações, discernimentos e conjecturas.
- b) Equilíbrio entre o raciocínio determinístico e aquele que leva em conta a aleatoriedade: a instrução deveria ver como um importante precursor do raciocínio estatístico, a tendência dos alunos de tentar explicar as causas dos fenômenos quando lhes são apresentadas situações cujo contexto é conhecido por eles. A partir dessas explicações, a instrução deveria possibilitar que o aluno percebesse que muitas ocorrências naturais são aleatórias. Portanto, é fundamental buscar o equilíbrio entre os aspectos determinísticos e aleatórios da realidade

A autora concluiu que a abordagem empregada com ênfase na onipresença⁷ da variação ajudou os alunos a desenvolverem o Pensamento Estatístico que vai além do conhecimento superficial de regras e procedimentos. A instrução também se mostrou eficaz para ajudar os alunos a se afastarem do pensamento unidimensional e a integrarem medidas de centro e variabilidade em suas análises.

⁷ Termo empregado por Meletiou (2000)

Os resultados do estudo apontam para os benefícios de uma abordagem que considere mais seriamente a intuição dos alunos, integre a teoria com a prática e situe a instrução dentro de contextos familiares aos aprendizes. Os esforços para apresentar o Pensamento Estatístico como um equilíbrio entre o raciocínio determinístico e aleatório, ajudaram os estudantes a melhor entenderem a relação entre acaso e regularidade. No final do curso os alunos estavam menos propensos a pensar que todo fenômeno tem uma causa determinada, começando a compreender que algumas ocorrências são aleatórias.

Meletiou defende que

Empregar uma abordagem que incentive os alunos a explorar toda a distribuição de uma estatística para ver o que é provável e não provável de ocorrer, é mais eficaz do que abordagens de simulação padrão de inferência estatística, em que há uma tendência de centrar-se em frequências relativas ou médias e ignorar a variação. (MELETIOU, 2000, p. 290).

A investigação da autora citada foi importante para nosso trabalho, pois foi útil na elaboração de algumas atividades e reforçou nossa convicção de que é importante valorizar e incentivar as hipóteses dos alunos de que é preciso o equilíbrio entre o pensamento determinístico e aleatório e, sobretudo, reafirmou nossa posição de que a variabilidade deve ser mais valorizada na Educação Estatística.

1.3 Estudo com alunos do Ensino Médio

Vieira (2008) pesquisou aspectos do ensino e da aprendizagem da Estatística Descritiva, voltada a alunos do Ensino Médio, com o apoio de um software de estatística dinâmica (Fathom). O objetivo da pesquisa foi estudar as interações entre o aluno e um ambiente de estatística dinâmica, com base no desenvolvimento de uma sequência didática, para favorecer a análise exploratória de dados.

Para construir, usar e analisar a sequência didática aplicada com os alunos, a autora pautou-se nos pressupostos da Engenharia Didática

Com esta pesquisa, a autora pretendia responder a três questões: em que medida as articulações entre os diferentes tipos de registros de representações

semióticas podem constituir o acesso à compreensão em estatística; o uso de um software de estatística dinâmica pode gerar/potencializar a necessidade de diversos registros de representações semióticas e permitir que o sujeito desenvolva a capacidade de organizar, analisar e interpretar dados? Quais tipos de articulações entre os registros de representações semióticas de um conjunto de dados podem ser identificados, com base no uso do ambiente computacional, visando à análise crítica dos dados?

Vieira (2008) constatou que a articulação entre diferentes registros em estatística como entre tabelas e gráficos, representação numérica e gráfica de uma medida como a média ou a mediana, foram importantes para a compreensão desses conceitos pelos alunos. Quanto ao uso do software, a autora verificou que esta ferramenta foi essencial para possibilitar a visualização dos diferentes tipos de registros de representação semiótica de uma distribuição de frequência, contribuindo para a compreensão dos conceitos de média e mediana.

No que se refere à terceira questão, a autora percebeu que o uso do ambiente computacional possibilitou aos alunos a articulação de várias representações de um mesmo objeto e a visualização concomitante destes diferentes registros, propiciando a visualização de diversos aspectos de um mesmo objeto.

A teoria dos Registros de Representações Semiótica está voltada originalmente ao estudo do acesso à compreensão de conceitos matemáticos. O trabalho de Vieira (2008) foi importante para o nosso, porque mostrou a viabilidade da aplicação desta teoria na apreensão de alguns conceitos da estatística.

CAPÍTULO II – PROBLEMÁTICA

Neste capítulo apresentamos a justificativa, os objetivos, a questão de pesquisa e os procedimentos metodológicos adotados.

2.1 Justificativa

Destacaremos quatro razões que mostram a relevância de nossa pesquisa: a necessidade crescente na sociedade atual, de um método para analisar a grande quantidade de dados aos quais os cidadãos são expostos diariamente; a recente introdução da estatística no currículo brasileiro, com a implantação dos Parâmetros Curriculares Nacionais para o Ensino Fundamental (PCN-EF) e Médio (PCN-EM) em 1997 e 2000, respectivamente; a necessidade de formar o professor para a abordagem de um tema novo e a importância da variabilidade na estatística.

Para Batanero (2001), existe relação entre o desenvolvimento de um país e o grau da produção de estatísticas completas e confiáveis, pois as informações produzidas por elas são necessárias para a tomada de decisões acertadas na economia, política, indústria, pesquisas, etc. Ora, de nada adianta a produção de ferramentas de análise de dados se a população não estiver preparada para compreendê-las e utilizá-las para tomadas de decisão, ou mesmo, muito simplesmente, para leitura de informações que permitam compreender fatos corriqueiros do dia a dia de sua comunidade, sua cidade, seu país e do mundo.

Diariamente estamos expostos a inúmeros dados e informações. O acesso a eles é facilitado pelas novas tecnologias e pela internet que oferecem as mais variadas fontes de pesquisa. Submersos nesta imensidão de dados e informações, faz-se necessário desenvolver a habilidade para organizar, selecionar, sintetizar, analisar criticamente, levantar hipóteses, inferir e comunicar as conclusões. Assim, o exercício da cidadania passa pelo desenvolvimento de tais habilidades e das competências relacionadas que podem ser construídas com o letramento estatístico⁸.

⁸ Entendemos letramento estatístico como a capacidade do indivíduo mobilizar os conhecimentos estatísticos adquiridos para interpretar, avaliar, decidir e comunicar-se sobre dados e informações que podem se apresentar em qualquer contexto.

A infinidade de informações às quais nos referimos, são oriundas da realidade física ou social e como tal podem estar sujeitas à aleatoriedade, à variação, à incerteza e ao erro. Neste contexto, o estudo da variabilidade e da probabilidade assumem um papel de destaque na análise estatística. Gal (2002) citando os trabalhos de Garfield e Gal (1999) e de Moore (1998), escreveu a respeito dos conhecimentos básicos de estatística, destacando a centralidade da variabilidade na análise de dados: “[...] os adultos deveriam possuir certos conhecimentos, ao menos informal, das ‘grandes ideias’ fundamentais que servem de base para as investigações estatísticas. A primeira na lista da maioria dos estatísticos é a existência da variação.” (GAL, 2002, p.12).

Nessa mesma direção, encontramos o trabalho de Silva (2007, p.32), ao afirmar que “variabilidade é elemento central do Pensamento Estatístico e conteúdo essencial para que um indivíduo seja letrado estatisticamente”.

Salientamos, também, a necessidade de pesquisas nessa área, pois sua incorporação ao currículo brasileiro é recente, fazendo parte de um dos quatro blocos de conteúdos de Matemática propostos pelos PCN-EF (BRASIL, 1997) – Tratamento da Informação - e integrando um dos três eixos estruturadores do ensino da Matemática no nível Médio, intitulado pelos PCN+ (BRASIL, 2002), como Análise de Dados .

Estes conteúdos exigem uma abordagem própria para a qual os professores não foram preparados em sua formação inicial. Para alguns destes, a estatística nem fez parte do currículo da graduação, outros a estudaram, mas com enfoque nos cálculos, fórmulas e resolução de exercícios, distantes de uma análise exploratória de dados que pudesse prepará-los para conduzir as situações de aprendizagem investigativa e autônoma.

Desta forma, torna-se necessário formar os atuais e futuros professores no que se refere aos conceitos e abordagem relativos ao bloco Tratamento da Informação e Análise de Dados. Seguindo esta linha de raciocínio, Silva (2007) afirma ser necessário letrar estatisticamente os professores, para que estes façam o mesmo com seus alunos.

Para formar bem professores e alunos em estatística, é preciso envolvê-los com as ideias de variabilidade que exercem um papel central na estatística. Moore

(1997) considera como núcleo do pensamento estatístico: a necessidade e a importância da produção de dados; a onipresença da variabilidade; a medição e a modelagem da variabilidade. Wild e Pfannkuch (1999) afirmam que o elemento central das definições publicadas sobre o Pensamento Estatístico é a variação. O levantamento bibliográfico de Meletiou (2000) identificou um abandono quase total de investigações referentes à variabilidade. Para ela, o abandono da abordagem de variabilidade no ensino pode ser uma explicação para a dificuldade dos alunos na compreensão de conceitos estatísticos.

Por tudo isso, consideramos que se faz necessário uma pesquisa sobre o ensino e a aprendizagem de estatística com ênfase na variabilidade.

2.2 Objetivos e questão de pesquisa

Tendo em vista a novidade da estatística nos currículos, a escassez de pesquisas sobre a variabilidade, descrita por vários autores (MELETIOU, 2000; READING; SHAUGHNESSY, 2004;), a dificuldade dos alunos para compreensão do desvio-padrão e articularem conhecimentos matemáticos e estatísticos para aplicá-los em situações novas, objetivamos estudar as características das atividades de uma formação de alunos de um curso de licenciatura em Matemática, que favoreçam a apreensão da variabilidade, quando se analisa um conjunto de dados.

Neste contexto, levantamos a seguinte questão de pesquisa:

Quais as características didáticas de uma formação continuada em Estatística que utilize o enfoque da Análise Exploratória de Dados para favorecer a percepção da variabilidade por alunos de um curso de licenciatura em Matemática?

2.3 Procedimentos Metodológicos

Como metodologia de pesquisa, seguimos alguns pressupostos da Engenharia Didática, tais como: a construção de situações que possam propiciar

situações adidáticas⁹, a análise *a priori* e análise *a posteriori* de três situações didáticas que compõe o conjunto. As demais tarefas serão analisadas apenas quanto aos objetivos, uma vez que seu papel é o de construir e/ou homogeneizar o conjunto de conhecimentos estatísticos necessários, para que os participantes possam construir uma estratégia para resolver os problemas propostos na situação didática em questão.

Neste trabalho, não faremos a análise epistemológica dos conceitos relacionados às medidas de variação, limitando-nos ao estudo da situação atual do ensino desses conceitos, pelo estudo das pesquisas publicadas na área e dos PCN.

2.3.1 Os procedimentos adotados

A sequência didática foi organizada inicialmente em 37 tarefas e, à medida que o curso era ministrado, algumas delas sofriam alterações para se adaptarem melhor ao objetivo da formação. Em alguns momentos, foi necessário introduzir atividades intermediárias, possibilitando o salto entre o conhecimento velho e o novo, perfazendo, assim, um total de 43 tarefas.

A sequência foi organizada de tal forma que os alunos pudessem utilizar seus conhecimentos anteriores, debater com os membros do grupo para assim construir um novo conhecimento. Este deveria ser utilizado na atividade seguinte ou em um conjunto delas, ou seja, ele seria mobilizado para a construção de um novo conhecimento. Após cada bloco de atividades, era feita a discussão com a turma toda e, pautado nessa interação, institucionalizava-se, o que foi aprendido.

Os alunos resolveram as tarefas organizados em grupos para propiciar a cooperação, a troca de ideias e o debate em torno das propostas de solução das questões. Assim, deveriam apresentar suas sugestões, defender suas posições, ouvir as críticas dos colegas e juntos formularem a solução. Em seguida, havia a socialização entre os grupos. Neste momento, o formador destacava alguns pontos

⁹ “Situação adidática é uma situação na qual a intenção de ensinar não é revelada ao aprendiz, mas foi planejada pelo professor para propiciar condições favoráveis para a apropriação do novo saber que se deseja ensinar”. (ALMOULOU, 2007, p.33)

que considerava relevantes para a noção trabalhada, fazia certos questionamentos a respeito dos aspectos do problema que não foram apreendidos pelos alunos, incentivava o debate e, com base na participação dos alunos fazia a institucionalização do conhecimento

A formação foi organizada em 11 encontros semanais de 3 horas cada ministrada pelo próprio pesquisador. Os 17 participantes realizavam as tarefas organizadas em grupos para incentivar a interação. A coleta de dados usou os escritos produzidos pelos alunos, gravações de áudio e observações feitas por uma pesquisadora colaboradora. As gravações só aconteceram nos últimos três encontros.

No primeiro encontro, os alunos responderam a um questionário que visava a traçar o perfil do grupo. Foram propostas também três questões simples de estatística para diagnosticar algumas concepções dos alunos. No último dia do curso, foram dadas as mesmas questões do primeiro encontro com a finalidade de comparar a evolução dos participantes em relação aos conceitos abordados nas três questões: mediana, média e variabilidade.

2.3.2 A Engenharia Didática

Machado (2008) citando Waldegg (1997) diz que a pesquisa em didática é um processo empírico, no sentido que deve extrair dados da realidade e compará-los com as hipóteses, podendo ser a sala de aula o laboratório de confrontação das hipóteses levantadas. Como as hipóteses são formuladas com base em um modelo teórico e são validadas ou não na complexidade da prática, é necessário buscar uma metodologia de pesquisa cuja “concepção contemple a dimensão teórica e experimental [...], articulando pesquisa e ação pedagógica.” (PAIS, 2002, p.99).

A Engenharia Didática é um instrumento metodológico que tem a dupla função: considerar o modelo teórico e a complexidade da realidade da sala de aula, servindo de controle sistemático para preservar as condições de confiabilidade da atividade científica. Segundo Artigue, o termo Engenharia Didática:

Surgiu no trabalho didático que é aquele comparável ao trabalho do engenheiro que, para realizar um projeto preciso, se apóia sobre

conhecimentos científicos de seu domínio, aceita submeter-se a um controle científico, mas, ao mesmo tempo, se vê obrigado a trabalhar sobre objetos bem mais complexos que os depurados da ciência [...]. (ARTIGUE, 1996, p. 193)

Artigue (1996, p. 196) caracteriza a Engenharia Didática “como um esquema experimental baseado em ‘realizações didáticas’ na sala de aula, isto é, a respeito da concepção, realização, observação e da análise de sequências de ensino.” De acordo com Almouloud e Coutinho (2008), esta metodologia é um estudo de caso que possui uma validação interna, apoiando-se na confrontação entre, o que se analisou antes do experimento, com os resultados obtidos após a aplicação das atividades.

Esta metodologia divide-se nas seguintes fases: análises preliminares, análise *a priori*¹⁰, experimentação, análise *a posteriori* e validação. Descreveremos resumidamente cada uma delas.

Para Machado (2008), as análises preliminares são feitas com base no quadro teórico geral da didática e apoiados nos conhecimentos didáticos específicos sobre o assunto já pesquisado. Almouloud e Coutinho (2008) destacam alguns aspectos dessa análise, a saber: a epistemologia dos conteúdos visados pelo meio; concepções dos alunos e os obstáculos e dificuldades encontrados por eles para apreensão do conteúdo; objetivos específicos da pesquisa e estudo da transposição didática do saber no meio ao qual se destina.

Para estes autores, as análises preliminares são importantes porque permitem que o pesquisador identifique as variáveis didáticas potenciais que serão explicitadas e manipuladas nas próximas fases.

A análise teórica é a fase na qual se explicitam as variáveis que consideramos pertinentes. Machado (2008) afirma que o pesquisador, baseando-se na fase anterior, delimita algumas variáveis do sistema sobre o qual o ensino poderá atuar. Para Artigue (1996), o objetivo da análise teórica é estabelecer as escolhas feitas que permitem controlar o comportamento dos aprendizes e o sentido desses comportamentos. Para isso, esta fase deve se basear em hipóteses que

¹⁰ Nos primeiros textos sobre Engenharia Didática, usava-se o termo análise *a priori* para designar todo o estudo teórico que se faz das situações didáticas que compõem a engenharia. No entanto, para evitar confusões de temporalidade, adotaremos, a partir de agora, o termo "Análise Teórica" para designar esse estudo.

serão relevantes na confrontação entre a análise teórica e a análise *a posteriori*, para validação interna da engenharia.

Conforme Almouloud e Coutinho (2008), em uma análise teórica são elementos importantes relatar as escolhas das variáveis locais e a característica da situação adidática desenvolvida; analisar a importância dessa situação para o aluno em função das possibilidades de ações, escolhas, estratégias de controle e validação que ele terá; antecipar comportamentos possíveis; fazer análise e tentar mostrar que se o comportamento previsto ocorrer, ele será resultado do surgimento do conhecimento visado.

A experimentação é a etapa da aplicação das atividades com os alunos. Almouloud e Coutinho (2008) esclarecem ser este o momento em que se coloca em funcionamento todo dispositivo construído, corrigindo-o se necessário, quando as análises locais do desenvolvimento experimental identificarem essa necessidade.

A última fase da Engenharia Didática é a análise *a posteriori* e validação. Segundo os autores citados no parágrafo anterior:

A análise *a posteriori* de uma sessão é um conjunto de resultados que se pode tirar da exploração dos dados recolhidos e que contribuem para a melhoria dos conhecimentos didáticos que se tem sobre as condições da transmissão do saber em jogo. [...] É uma análise feita a luz da análise *a priori*, dos fundamentos teóricos, das hipóteses e da problemática da pesquisa. (ALMOULOUUD e COUTINHO, 2008, p.68)

Artigue (1996) considera que é com base na confrontação das análises teóricas e *a posteriori* que se fundam essencialmente a validação das hipóteses envolvidas na investigação.

Assim, a Engenharia Didática é uma metodologia de pesquisa útil para confrontar hipóteses baseadas em modelos teóricos com a prática em sala de aula. Com esta dupla função, sua importância fica confirmada na validação das situações que propusemos aos alunos. Escolhemos algumas tarefas e fizemos a *análise* teórica, posteriormente aplicamos com os alunos e analisamos os resultados (*análise a posteriori*).

CAPÍTULO III – QUADRO TEÓRICO

A sequência didática foi preparada com base nas dimensões da análise de conteúdos a ensinar e na elaboração de cenários proposta por Robert (1998). Assim, esta teoria foi empregada tanto na avaliação dos níveis de funcionamento dos conhecimentos adquiridos pelos alunos como na organização de cenários que propiciassem a aprendizagem.

3.1 As dimensões de análise de conteúdos a ensinar e a elaboração de cenários

Robert (1998) apresenta quatro dimensões para tentar penetrar na complexidade das noções matemáticas. A intenção da autora foi propor instrumentos de diagnóstico dos conhecimentos dos estudantes ou propor exercícios que ajudassem a superar suas dificuldades.

As três primeiras dimensões referem-se às noções que se pretende ensinar tal como aparecem nos programas, sob o ponto de vista semiótico e ao salto conceitual que deve ser efetivado para que o aluno possa construir novos conceitos. A Dialética Ferramenta-Objeto e Mudança de Quadros e os Registros de Representação Semiótica contemplam parte significativa das primeira e segunda dimensões. A terceira dimensão está relacionada à Teoria dos Campos Conceituais, que não é objeto de nossa pesquisa. A quarta dimensão identifica as diferentes maneiras dos alunos colocarem em funcionamento suas noções em atividades, classificando o funcionamento dos conhecimentos deles em três níveis: técnico, mobilizável e disponível. Esta dimensão foi a utilizada como quadro teórico de nossa pesquisa.

Segundo a autora, estas quatro dimensões podem nos auxiliar na elaboração de cenários (no plano dos conteúdos) que ajudem na introdução de noções matemáticas.

As duas primeiras dimensões foram úteis na elaboração de nossa sequência didática. Na Dialética Ferramenta-Objeto, um conceito matemático pode assumir a condição de instrumento para resolução de problemas e ser o próprio objeto da aprendizagem. Assim, um saber matemático assume o papel de ferramenta quando é utilizado para se chegar a uma nova noção matemática que adquire o status de objeto, pois é o próprio conteúdo da aprendizagem. O processo é cíclico e contínuo, pois o aluno mobiliza conhecimentos antigos para construir um novo (objeto) que passa a fazer parte do repertório de conhecimentos do aprendiz, podendo ser utilizado como ferramenta para resolver problemas e gerar novos conhecimentos.

Nossa sequência didática foi organizada com diversas tarefas de tal forma que os alunos pudessem mobilizar seus conhecimentos prévios e chegar ao conhecimento visado. Nas atividades posteriores, os alunos deveriam utilizar as conclusões das anteriores, como instrumentos de resolução para construírem noções novas ou descobrirem outros aspectos das noções construídas anteriormente.

Recorremos também à teoria dos Registros de Representação Semiótica para formular tarefas que possibilitassem ao aluno transitar entre os diferentes registros de representação da noção estatística em estudo.

Damm (2008); Vieira (2008), baseando-se em Raymond Duval,¹¹ afirmam que os conceitos matemáticos necessitam da representação para sua apreensão, pois seus objetos são abstratos e não podem ser diretamente acessíveis à percepção. O estudo do ensino e da aprendizagem da Matemática implica a compreensão das diversas formas de representações, fundamentais para o desenvolvimento desta ciência.

Desta forma, percebemos a importância da diversidade de representações e articulação entre elas, pois:

A apreensão conceitual dos objetos matemáticos somente será possível com a coordenação, pelo sujeito que aprende, de vários registros de representação. Ou seja, quanto maior for a mobilidade com registros de representação diferentes do mesmo objeto matemático, maior será a possibilidade de apreensão desse objeto. (DAMM, 2008, p. 177).

¹¹ Filósofo e psicólogo francês que desenvolveu importantes trabalhos sobre a teoria de Registros de Representação Semiótica

Tendo em vista que a estatística necessita do ferramental matemático para realizar suas análises e que seus objetos, assim como os da Matemática são abstratos e carecem da representação para sua apreensão, defendemos que as conclusões da teoria dos Registros de Representação Semiótica podem ser aplicadas à estatística. Os resultados observados por Vieira (2008, p. 24) corroboram com nosso pensamento ao afirmar que “os objetos estatísticos não são facilmente acessíveis ou perceptíveis: o acesso a esses objetos passa necessariamente por representações semióticas [...]”

Na análise exploratória de dados, temos situações que podem ser representadas de diversas maneiras. Uma coleta de dados, por exemplo, pode ser realizada por meio de pesquisa na língua natural, e os dados brutos obtidos precisam ser registrados na linguagem escrita, quantificados e organizados. Uma ferramenta de organização muito comum é a tabela.

Outros instrumentos fundamentais nas análises estatísticas são os gráficos, pelo seu poder de sintetizar e comunicar a informação. Além deles, podemos utilizar várias medidas-resumo para analisar e comunicar informações importantes sobre o conjunto de dados pesquisado. Percebemos, assim, a importância do uso de diversos registros para a construção do letramento estatístico. No exemplo citado, utilizamos registro na língua natural, registro numérico, tabela e gráfico. Cada registro possibilita a exploração de um aspecto diferente do objeto de estudo, enriquecendo a análise e dando maiores subsídios à tomada de decisão.

A quarta dimensão proposta por Robert (1998) trata dos níveis de funcionamento dos conhecimentos dos alunos. São eles: técnico, mobilizável e disponível. O nível *técnico* se refere à aplicação direta de uma fórmula ou teorema sem a necessidade de adaptações e corresponde:

às mobilizações indicadas, isoladas, que explicitam aplicações imediatas de teoremas, propriedades, definições, fórmulas etc. Trata-se então de contextualização simples, locais, sem etapas, sem trabalho preliminar de reconhecimento, sem adaptações. (ROBERT, 1998, p. 165)¹²

Como exemplo, podemos citar a aplicação do algoritmo da média. Suponhamos que o procedimento do cálculo da média foi apresentado aos alunos e

¹² Traduzido com fins didáticos pelas professoras doutoras Cileda Q. S. Coutinho e Maria José Ferreira da Silva

foi solicitada a idade média de um grupo de indivíduos, com base nos valores fornecidos. Para resolver tal questão, só será necessária a aplicação do algoritmo. O aluno obterá o valor correto da média, mesmo sem saber onde aplicá-la, qual seu significado ou como utilizá-la para analisar um conjunto de dados.

O nível denominado *mobilizável* exige do aluno certo grau de mobilidade de conhecimentos, no sentido de adaptar, o que ele já conhece para aplicar um determinado teorema, algoritmo ou procedimento. Segundo a autora, em alguns casos, ele precisará mudar de domínio, de ponto de vista ou aplicar várias vezes o mesmo procedimento ou teorema. O que está em jogo é explícito, mas requer mais do que a aplicação direta do que foi ensinado. Tanto no nível técnico como no mobilizável há dicas para os alunos resolverem as tarefas.

Como exemplo, podemos utilizar novamente uma tarefa que envolve a média: uma turma de dez alunos tem idade média igual a 21 anos. Qual será a nova média se chegar um aluno com 32 anos de idade e se juntar ao grupo? A aplicação direta do algoritmo da média não resolve o problema. Ao aplicar a fórmula:

$$\frac{\sum_{i=1}^n x_i}{n} = \mu$$

(x_i corresponde ao valor de cada dado, n representa a quantidade de dados e μ a média aritmética) ele não obterá imediatamente a resposta solicitada, mas descobrirá o valor da soma dos dez dados:

$$\frac{\sum_{i=1}^n x_i}{10} = 21 \Rightarrow \sum_{i=1}^n x_i = 210.$$

Chegando mais uma pessoa com idade 32 anos, a nova soma será 242. Aplicando a fórmula novamente tem-se:

$$\mu = \frac{242}{11} \Rightarrow \mu = 22.$$

Portanto, a nova média será 22 anos. Este exemplo, embora simples pode representar o nível mobilizável, pois exige do aluno alguma adaptação de seus conhecimentos que vão além da aplicação direta da fórmula da média ou aplicação da mesma mais de uma vez.

Robert (1998) esclarece que, no nível *disponível*, que é o mais elevado, o aluno é capaz de resolver problemas sem nenhuma indicação do professor e possui condições de articular seus conhecimentos, buscando o que for necessário para se chegar à solução requerida. Espera-se que ele seja capaz de mudar de domínios, fornecer contraexemplos e aplicar métodos não previstos. Para a autora:

Este nível de funcionamento está ligado a uma familiaridade importante ao conhecimento de situações de referências diferentes, que o estudante sabe que os conhece, podendo servir de terreno de experimentação, ao fato de dispor de referências, de questionamentos, de uma organização. (ROBERT, 1998, p.167)¹³.

Quando a autora discute as práticas esperadas dos alunos e suas dificuldades, afirma que vários trabalhos têm mostrado que eles adotam com frequência modelos de resolução de problemas redutores, sendo tentados à limitarem-se ao uso de algoritmos. Assim, quando as condições são implícitas, eles tendem a um desempenho insatisfatório. Robert resume algumas constatações sobre as limitações dos alunos da seguinte maneira:

Os conhecimentos dos alunos são “atomizados”, sem organização, nem explícitos nem implícitos, há somente certa acumulação de saberes justapostos; são essencialmente de teoremas que são memorizados, com exceção de situações ou de exemplos, os alunos sabem aplicá-los se a aplicação é simples e indicada [...] e não mobilizam conhecimentos anteriores espontaneamente. (ROBERT, 1998, p.152).

No que se refere à interpretação dos enunciados dos problemas, a autora constatou que os estudantes fixam-se no que aparece explicitamente no texto e não possuem à sua disposição referenciais para retomar ideias já citadas.

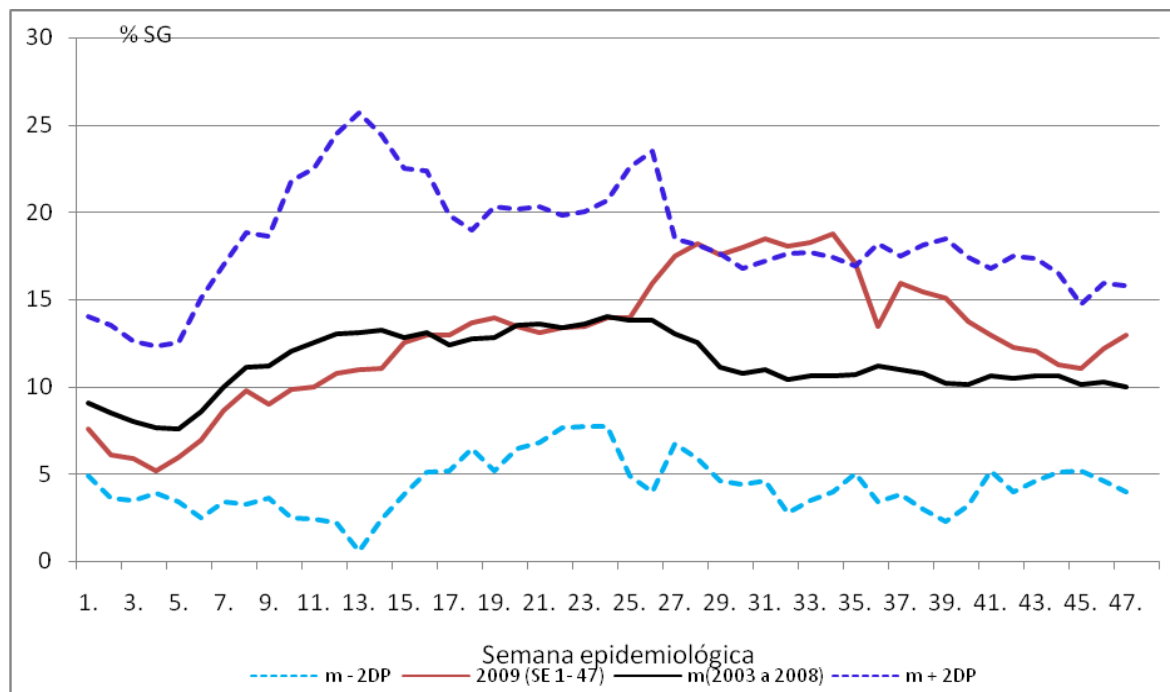
Pelo exposto, percebemos a necessidade das sequências didáticas promoverem a evolução do aluno do nível técnico para o mobilizável e deste para o disponível. Enfim, a autora sugere que nenhum desses três níveis seja negligenciado no ensino de Matemática, uma vez que a aprendizagem é progressiva.

Dependendo da forma como é proposta uma tarefa, dos dados expressos, dos questionamentos feitos, esta poderá exigir dos alunos diferentes níveis de conhecimentos.

¹³ Traduzido com fins didáticos pelas professoras doutoras Cileda Q. S. Coutinho e Maria José Ferreira da Silva

Na figura 2, tentaremos ilustrar os três níveis de conhecimentos em uma tarefa de análise de gráfico.

O informe epidemiológico do Ministério da Saúde, de dezembro de 2009, sobre a gripe H1N1, apresentou o gráfico a seguir que retrata a evolução dos atendimentos por síndrome gripal nas unidades sentinelas¹⁴. Vale ressaltar que os “[...] laboratórios de referência passaram a priorizar o processamento de amostras e diagnóstico de casos graves e óbitos”. (BRASIL, 2009, p.9).



Proporção de atendimentos por síndrome gripal em relação ao total de atendimentos nas unidades sentinela¹⁵. Brasil, até a semana 47 do ano de 2009.

Fonte: BRASIL. Informe epidemiológico influenza pandêmica. Dez/2009, p. 9.

Esclarecimento a respeito das legendas:

— **m (2003 a 2008)** significa a média aritmética dos atendimentos das semanas correspondentes entre os anos de 2003 a 2008;

----- **m - 2DP** indica média menos duas vezes o desvio-padrão;

----- **m + 2DP** indica média mais duas vezes o desvio-padrão;

— **2009 (SE 1-47)** corresponde ao percentual de atendimentos da primeira a quadragésima sétima semana de 2009.

a) Em qual semana de 2009 as unidades sentinela atingiram o ponto máximo de atendimento por síndrome gripal?

¹⁴ São unidades do Sistema Nacional de Vigilância de Síndrome Gripal do Ministério da Saúde.

- b) Compare o percentual de atendimento da 3ª semana de 2009 com o da 13ª.
- c) Compare o percentual de atendimento da 3ª semana de 2009 com o da 13ª, levando em consideração às respectivas médias de atendimentos, entre 2003 e 2008.
- d) Observando a linha do tempo em 2009, tendo em vista os registros históricos de 2003 a 2008, em qual momento as autoridades deveriam se preocupar? Qual a justificativa para isso?

Figura 2 - Exemplo de tarefa para analisar os níveis de conhecimento

Análise dos níveis de conhecimento

a) Em qual semana de 2009 as unidades sentinela atingiram o ponto máximo de atendimento por síndrome gripal?

Um aluno que estivesse no nível *técnico*, poderia responder a esta questão, bastando para isso identificar qual linha representa os atendimentos de 2009, qual seu ponto mais alto e saber localizar suas coordenadas. A informação está explícita, não sendo necessário adaptações.

b) Compare o percentual de atendimento, da 3ª semana de 2009, com o da 13ª.

Para resolver a questão, o estudante precisaria localizar os pontos que representam os atendimentos das 3ª e 13ª semanas, encontrar suas ordenadas e comparar os valores encontrados. Assim, bastaria estar no nível *técnico* para responder ao que foi solicitado.

c) Compare o percentual de atendimento da 3ª semana de 2009 com o da 13ª, levando em consideração as respectivas médias de atendimentos entre 2003 e 2008.

No item “b”, o aluno deveria concluir que houve aumento de atendimentos entre as 3ª e 13ª semanas. No entanto, ao fazer a comparação da linha do tempo do período de 2009 e a média dos atendimentos de 2003 a 2008 do mesmo período, poderia concluir que o aumento já era esperado, tendo em vista que a média da 13ª semana de 2003 a 2008, também, aumentou.

Para chegar a tais conclusões, não bastaria ler informações explícitas no gráficos, mas fazer inferências a partir delas, articulando as informações pontuais de 2009 com as fornecidas pelas médias de 2003 a 2009. Por isso, esta questão

exigiria que o aluno estivesse, pelo menos, no nível de conhecimento *mobilizável* quanto às noções estatísticas envolvidas (interpretação e integração dos dados do gráfico, e média)

Entretanto, para fazer uma análise mais consistente, seria necessário levar em consideração a variabilidade (ver desvio-padrão: tabela 1), pois a média é insuficiente para resumir dados. Com base no gráfico, podemos obter os seguintes valores aproximados:

Tabela 1 - Dados aproximados da 3ª e 13ª semanas epidemiológicas

	3ª semana	13ª semana
Atendimento em 2009	6%	11%
Média dos atendimentos de 2003 a 2008	8%	13%
Desvio-padrão ¹⁶	2,5%	6,5%

Para analisar os afastamentos das 3ª e 13ª semanas de 2009 em relação às respectivas médias históricas*, é fundamental considerar a variabilidade dos dados, pois isto tem consequências importantes na prática. Na 3ª semana, a variação foi menor indicando que os dados estavam mais concentrados em torno da média, na 13ª semana a variação foi bem maior indicando que os dados estavam mais espalhados em relação à média. Considerando tais diferenças de variação, os afastamentos de cada ponto em relação à média de seu grupo possuem significados diferentes, devendo ser levados em consideração nas análises.

No item “c”, não havia nenhuma indicação para o aluno utilizar o conceito de variação ao analisar as médias, mas tal utilização é fundamental. A compreensão do conceito de média, de variabilidade, a articulação entre os dois e a aplicação em uma situação de análise dos dados foram necessárias. Por isso, tais conhecimentos seriam classificados no nível *disponível*.

d) Observando a linha do tempo em 2009, tendo em vista os registros históricos de 2003 a 2008, em qual momento as autoridades deveriam se preocupar? Qual a justificativa para isso?

Para responder a esta questão, o aluno deveria observar que o momento crítico foi quando o percentual de atendimento ultrapassou o limite da média mais dois desvios-padrão, pois podemos considerar este evento incomum, observando

¹⁶ Para obter o DP, tomamos metade da diferença entre a média (linha preta) e os limites do intervalo criado a partir da soma e subtração de dois desvios-padrão da média (linha pontilhada azul).

* Médias das semanas analisadas de 2003 a 2008.

que a chance de ocorrência é pequena. Assim, o período mais grave seria das 28^a a 35^a semanas, a partir das quais a situação começa a retornar ao padrão esperado.

Neste item, o aluno precisaria mobilizar vários conhecimentos e articulá-los. Ele deveria saber ler o gráfico, relacionar dados entre si e interpretá-los. Para isso, o aprendiz precisaria ter habilidade com a leitura de gráfico de linhas, conhecer o conceito de média e desvio-padrão. Depois, seria necessário relacionar a linha que representa a média do atendimento à gripe, as linhas formadas com o desvio-padrão e a linha dos atendimentos de 2009. Para tomar a decisão, deveria saber que o intervalo criado com a soma e a subtração de dois desvios-padrão abarca mais de 75% dos valores, qualquer que seja a distribuição dos dados, segundo a desigualdade de Tchebichev¹⁷. Em distribuições simétricas, inclui mais de 95% deles. Assim, o trecho da linha vermelha que ultrapassa o limite desse intervalo, poderia ser considerado um período incomum e, neste exemplo, um caso mais grave.

Nesta questão, não há indicações para solução do problema, o aluno precisa procurar em seus próprios conhecimentos, o que pode intervir na solução e relacionar conceitos, por isso, classificamos no nível de conhecimentos disponível.

Os níveis de funcionamento do conhecimento proposto por Robert (1998), foram importantes ferramentas de análise das atividades realizadas pelos alunos neste trabalho. Procuramos diagnosticar se a sequência didática proposta possibilitou a transição dos alunos do nível técnico para o mobilizável e disponível e, conseqüentemente, maior mobilidade cognitiva para análise de dados para tomada de decisão.

Por fim, utilizamos elementos da construção de cenários para elaboração de situações de aprendizagem, propostos por Robert (1998). Segundo a autora as tarefas devem provocar certo desequilíbrio nos alunos, permitindo-lhes avançar em construções mais elaboradas. Os enunciados devem ser elaborados de forma a “forçarem” os alunos a construção de adaptações nos teoremas, articulações, etc. O professor deve respeitar o caráter adidático na primeira fase da resolução de problemas, mas caberá a ele fazer a institucionalização do novo saber; as tarefas

¹⁷ Apresentada no próximo capítulo

devem ser tais que os alunos se engajem na resolução; uma noção nova deve ser extensão de outra já construída, possibilitando ao aluno pensar sozinho; as tarefas devem propor a articulação de vários registros.

CAPÍTULO IV – ANÁLISES PRÉVIAS

Neste capítulo abordamos as noções estatística trabalhadas com enfoque didático, a variabilidade e os Parâmetros Curriculares Nacionais, o Pensamento Estatístico e a importância da variabilidade na estatística.

4.1 Noções Estatísticas abordadas

Nesta seção, apresentamos as noções estatísticas trabalhadas na formação. As medidas-resumo foram divididas em medidas de posição e de dispersão. Em seguida, apresentamos: a Desigualdade de Tchebichev; a Regra Empírica utilizada para determinar o percentual de dados contidos no intervalo criado, com base na subtração e soma de desvios-padrão à média; o Evento Raro e o Gráfico de Pontos

4.1.1 Medidas de posição

As medidas-resumo de posição são assim chamadas, pois podem ser posicionadas literalmente no eixo que representa o conjunto de dados. Dentre elas, destacamos a média, a moda, a mediana, o valor máximo, o valor mínimo e os quartis.

Média aritmética

A média aritmética é uma medida-resumo de posição e de tendência central que pode caracterizar um conjunto de valores, sendo um “ponto de equilíbrio” entre eles.

O seu algoritmo é:

$$\mu = \frac{x_1 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Onde μ representa a média, x_i os valores dos dados e n a quantidade de dados.

Ao citar Lavoie e Gattuso (1998) e Batanero (2000), Carvalho (2001) afirma que a origem da média remonta à antiguidade. Outrora, os astrônomos babilônicos faziam medições sucessivas da altura dos astros e precisavam encontrar uma forma de corrigir os erros. A maneira que encontraram para estimar o valor verdadeiro, após várias medições foi somar todas as medidas e dividi-las pela quantidade de dados.

Assim, percebemos que as primeiras ideias de média surgiram de uma necessidade real e estavam relacionadas com a tentativa de se obter uma estimativa do verdadeiro valor. É interessante notar que tal prática subsiste até hoje e vem sendo aperfeiçoada na Análise Exploratória de Dados, vinculando as medidas de tendência central com às de variação.

Além da importância da média na estatística, ela é largamente empregada socialmente, embora, muitas vezes, de forma inadequada.

Para estudar o significado de um objeto matemático, Batanero (2000) propõe que se parta de situações-problema. Assim, para compreender o conceito de média, podemos partir de diversos problemas e analisá-los sob diversos aspectos. Seja o seguinte problema:

1) Um grupo de oito alunos mediu o comprimento do palmo da professora deles e chegou aos seguintes valores: 20,2 cm, 20 cm, 20,3 cm, 20,1 cm, 20,5cm, 19,8 cm, 19,9 e 20,4 cm. Qual seria a melhor estimativa para a medida do palmo da professora?

Este é um exemplo de problema em que a média é uma possível solução para estimar o verdadeiro valor com base na análise de valores. Segundo Batanero (2000), se apresentarmos esta tarefa aos alunos, a maioria somará e dividirá por oito para obter o valor 20,15, o que é um indício de que a prática dos astrônomos babilônicos, de utilizar a média como estimativa, conserva-se até hoje.

A média, também, pode ser vista como uma divisão equitativa em problemas, como o seguinte:

2) Um pai deu R\$ 5,00 para o primeiro filho, R\$ 9,00 para segundo e R\$ 1,00 para o terceiro. Quanto cada filho deveria receber se o pai distribuísse o dinheiro igualmente para os três?

O problema 3 explora a média como elemento representativo.

3) A altura de um grupo de alunos é 157 cm, 159 cm, 160 cm, 161 cm, 163 cm. Qual altura você escolheria para representar o grupo?

Segundo Batanero

Para representar um conjunto de dados se utiliza a média por suas propriedades de localização central, por ser o “centro de gravidade” [...] Se a distribuição for muito assimétrica, o valor mais freqüente (moda) ou o valor central do conjunto de dados ordenados (mediana) poderia ser mais representativo. (BATANERO, 2000, p. 45, tradução nossa).

Para a autora, quando acrescentamos condições a um campo de problemas surgem conceitos relacionados com o objeto que está sendo estudado. Assim, percebemos que a construção de um conceito pode ser enriquecida com a escolha de problemas que abordem os diferentes aspectos do conhecimento visado.

Assinalamos algumas dificuldades e erros apontadas por Batanero (2000) que devem ser consideradas, tanto nas pesquisas sobre o tema como na elaboração de sequências didáticas. A autora identificou como erros, o emprego da média simples quando deveria ser utilizada a média ponderada, o uso do algoritmo de forma mecânica sem compreender seu significado, o cálculo da média das frequências sem considerar os valores, cálculo da média desprezando os valores iguais a zero; crença de que a média deve coincidir com algum dos dados; utilização da média, como sinônimo de valor mais frequente, valor aproximado ou como valor do meio. Destacou também as seguintes dificuldades: compreensão de que a soma dos desvios em relação à média é zero; encontrar um valor desconhecido de um conjunto pequeno de dados com base na média conhecida.

Moda

A moda é uma medida-resumo de posição e tendência central que pode caracterizar um conjunto de valores, sendo o valor mais freqüente entre eles.

O conceito de moda é de fácil compreensão e tem uma forte conotação social. Ao citar Watson e Moritz (1999), Carvalho (2001) afirma que a moda é a medida de posição mais utilizada pelos alunos quando precisam escolher entre

média, moda e mediana. Para estes autores, a moda deveria ser o primeiro conceito a ser trabalhado com os alunos, por sua baixa complexidade cognitiva.

Mediana

A mediana é uma medida-resumo de posição e tendência central que pode caracterizar um conjunto de valores, ocupando a posição central entre eles quando dispostos em ordem crescente (ou decrescente), ou seja, é o valor que ocupa a posição tal que o conjunto fique dividido em duas partes com o mesmo número de elementos.

Para identificar a posição central, primeiro devemos ordenar os dados e, depois, verificar se a quantidade de elementos do conjunto de dados é ímpar ou par. Se for ímpar, existirá um elemento que dividirá o conjunto em duas partes iguais que representará a mediana. Se a quantidade de elementos for par, a mediana será a média dos dois elementos centrais. Tal procedimento obedece a uma lógica:

Considere um conjunto com n elementos, sendo n par. Logo, a mediana deverá dividir o conjunto dado em dois outros com $\frac{n}{2}$ elementos cada.

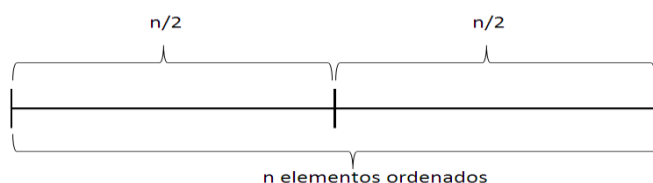


Figura 3 - diagrama representando a mediana.

A mediana não poderá ser nenhum dos valores já considerados em cada metade, pois isto faria com que o número de elementos em cada uma delas não fosse o mesmo. Assim, ela será qualquer valor entre os dois valores centrais (de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$). Convenciona-se, então, que seja a média aritmética entre eles.

Apesar dessa análise necessitar apenas das noções relativas à contagem e ordenação, observamos nos diversos materiais didáticos disponíveis ao professor que ela não é apresentada, restando apenas uma “regra operatória”: média entre os termos centrais.

A Introdução da moda e mediana nos livros didáticos é bem mais recente que a média. Segundo Carvalho (2001), só depois que os três conceitos passaram a

ser ensinados nas escolas, foi possível fazer comparações entre eles, enriquecendo sua aprendizagem. No Brasil, os Parâmetros Curriculares Nacionais (BRASIL, 1998) sugerem que se inicie a abordagem da mediana no quarto ciclo (8º e 9º anos) do Ensino Fundamental, enquanto a média aparece no segundo ciclo (3º e 4º anos).

A vantagem da mediana em relação à média é a sua robustez, isto é, ela não é influenciável por valores extremos, diferente do que acontece com a média. Além disso, é útil em distribuições assimétricas.

Do ponto de vista da aprendizagem, a mediana tem a desvantagem de ser menos utilizada e conhecida socialmente do que a média, entretanto compreendê-la como termo “do meio” é intuitivo e, muitas vezes a própria média é interpretada erroneamente como o termo que ocupa a posição “do meio” da distribuição.

Embora seja fácil a apreensão da mediana como termo central, Batanero (2000) afirma que seu cálculo não é tão simples assim para o aluno, pois seu algoritmo é diferente quando a quantidade de elementos é par ou ímpar. Além disso, existem diferentes algoritmos que apresentam resultados distintos. Esta situação pode ser um complicador para os alunos que estão acostumados a métodos e resultados únicos em problemas de Matemática.

Batanero (2000) citando Carvalho (1988), apresenta alguns erros comuns no cálculo da mediana: obter o termo central sem ordenar os dados e calcular a moda em vez da mediana. A dificuldade dos alunos em ordenar os dados está relacionada com a falta de entendimento de que a mediana é uma medida estatística que se refere essencialmente a um conjunto de dados ordenados, razão pela qual é chamada de estatística de ordem.

Watson e Moritz (1999 apud CARVALHO 2001, p. 56) apresentam três sugestões didáticas para melhor distinguir as situações em que se aplicam a média, moda ou mediana:

- a) Utilizar gráficos de barras para dados agrupados por categorias onde os alunos facilmente compreendem o significado para valor modal e nos quais é possível verificar rapidamente a adequação de um parâmetro para resumir um conjunto de dados;
- b) apresentar tarefas com contextos diferentes, onde a utilização da moda ou da mediana é mais adequada;
- c) utilizar o algoritmo da média em contextos inadequados e evidenciar suas fraquezas.

Quartis

Os quartis são medidas-resumo de posição que dividem os valores, organizados em ordem crescente, em quatro partes com o mesmo número de elementos, a saber: primeiro quartil (Q_1) é o valor que faz com que 25 % das observações sejam menores ou iguais a ele, o segundo quartil (Q_2 ou mediana) é o valor que faz com que 50% das observações sejam menores ou iguais a ele e o terceiro quartil (Q_3) é o valor que faz com que 75% das observações sejam menores ou iguais a ele.

Uma estratégia bastante intuitiva para os alunos é calcular os quartis, utilizando o mesmo procedimento empregado para a mediana, ou seja, obtém-se a mediana (Q_2) que divide os dados em duas metades, em seguida, calcula-se a mediana da primeira metade (Q_1) e da segunda metade (Q_3), utilizando sempre os mesmos procedimentos para a quantidade de dados pares e ímpares indicados no cálculo da mediana.

Considerando os quartis como uma extensão da mediana, estendemos a eles as mesmas dificuldades e vantagens apresentadas para aquela medida.

Máximo e Mínimo

O máximo e o mínimo são medidas de posição que indicam os extremos de um conjunto de valores, ou seja, o maior e o menor valor respectivamente.

4.1.2 Medidas de dispersão ou de variação

Quando resumimos dados com medidas de tendência central sempre perdemos informações sobre o conjunto todo. Para contornar tal perda, precisamos considerar as medidas de dispersão. Segundo Ben-Zvi e Garfield (2008, p 203), “é difícil imaginar o resumo de um conjunto de dados usando apenas uma medida de centro ou usando apenas uma medida de dispersão.” Assim, não faz

sentido abordar uma e ignorar a outra. Os autores propõem que no processo de construção do conceito de média, por exemplo, já se aborde esta medida como um ponto de equilíbrio na distribuição dos dados, permitindo que o aluno perceba a variação dos valores em torno dela. A ideia de desvio em torno da média já prepara o estudante para a compreensão das medidas de variação como desvio-padrão, por exemplo, o que será explicado a seguir.

Amplitude

A amplitude é uma medida de variação obtida pela diferença entre os valores máximo e mínimo. Tem a vantagem de ser uma medida de fácil compreensão e rápida obtenção, permitindo uma primeira análise da dispersão em relação à média ou à mediana. Sua limitação é o fato de considerar apenas dois elementos do conjunto de dados.

Desvio-Médio

O desvio-médio é uma medida de dispersão obtida pela média dos valores absolutos dos desvios de cada valor em relação à média de todos os valores.

$$DM = \frac{\sum_{i=1}^n |x_i - \mu|}{n}$$

Na expressão, DM corresponde ao desvio-médio, x_i cada dado, μ a média e n a quantidade de elementos do conjunto.

A vantagem do desvio-médio em relação à amplitude é que seu algoritmo leva em conta todos os dados que estão sendo analisados, sendo, portanto, mais representativa dos mesmos. Quando comparado ao desvio-padrão, ele tem o benefício de ser mais intuitivo, pois é a média dos desvios absolutos, ou seja, toma estes desvios $|x_i - \mu|$ no sentido de “distância euclidiana”, já trabalhada com os alunos nas aulas de Geometria. Mas, tem a desvantagem de utilizar a função

módulo que, “por suas características matemáticas, torna mais difícil o estudo de suas propriedades”. (LIMA e MAGALHÃES, 2004, p.104).

Variância

A variância é uma medida-resumo de dispersão obtida pela média dos quadrados dos desvios de cada valor de um conjunto de dados em relação à média dos próprios valores.

$$Var = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Quando se trata de amostra, a variância é obtida:

$$Var = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N - 1}$$

Nas expressões, Var corresponde à variância, x_i cada valor fornecido, μ à média populacional, $\bar{x}$ à média amostral, n à quantidade de elementos do conjunto e N à quantidade de elementos da amostra.

Um inconveniente para a atribuição de significado à variância vem do fato dela ser representada por uma unidade de medida diferente dos dados originais, dificultando assim sua interpretação. Por exemplo, se os dados são expressos em centímetros, a variância será dada em centímetros quadrados. Resolve-se este problema cognitivo extraíndo a raiz quadrada da variância, que é denominada desvio-padrão.

Desvio-Padrão

Desvio-padrão é a raiz quadrada da variância.

$$DP = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad \text{ou} \quad DP = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N - 1}}$$

Reading; Shaughnessy (2004) afirmam que a complexidade dos cálculos do desvio-padrão é um obstáculo, para que o professor possa motivar os alunos para tal medida, pois ela exige que o estudante mobilize diversos conceitos como média, desvio, potência, raiz quadrada e fração. A compreensão desta medida não é intuitiva, e é difícil interpretá-la, pois é de natureza complexa, apesar de ser a medida de variação mais comumente utilizada.

A pesquisa de Meletiou (2000) também confirma esta constatação, pois dentre as medidas de dispersão trabalhadas pela autora, o desvio-padrão foi a que os alunos apresentaram maior dificuldade. Da mesma forma, Ben-Zvi e Galfield (2008) afirmam que, normalmente, os alunos aprendem a definição do desvio-padrão e como ele é calculado, mas raramente entendem de modo claro o significado desta medida e como interpretá-la.

Quando se pretende construir um conceito, é preciso ir além das fórmulas, dos cálculos e dos exercícios de aplicação de procedimentos, sendo indispensável dar ênfase ao significado do conceito visado. Para tal compreensão, DelMas e Liu (2005) conjecturam ser necessária a coordenação de outros conceitos estatísticos subjacentes, a partir dos quais o desvio-padrão é construído, ou seja, é preciso mobilizar a noção de distribuição, de média e de desvio em torno da média.

O conceito de distribuição requer a mobilização dos valores assumidos pela variável e de suas frequências. Portanto, a visualização gráfica facilita a compreensão da distribuição. É importante entender a média como um ponto de equilíbrio entre os dados e, finalmente, perceber os desvios de cada dado em relação à essa medida. Desta forma, notamos a importância de se trabalhar a visualização dos dados em representações gráficas e a exploração do significado da média, para a abordagem do desvio-padrão.

Conforme relata Meletiou (2000), muitos alunos associam variabilidade com variedade, com multiplicidade de valores, com diferença entre os dados, etc. Silva (2007) pesquisou livros didáticos e identificou que quase todos os analisados sugerem a interpretação do desvio-padrão, como uma medida de homogeneidade do grupo. Segundo Loosen, Lioen e Lacante (1985 apud Silva, 2007, p. 116) “os livros didáticos enfatizam a heterogeneidade entre as observações e não a heterogeneidade em relação à tendência central”.

A média é essencial na definição do desvio-padrão e quando abordamos a variação, estamos preocupados em estudar uma determinada medida-resumo em torno da qual existe ou não variação. Portanto, interpretar desvio-padrão como homogeneidade/heterogeneidade, concentração/dispersão sem se referir ao valor em torno do qual ocorre a variação, pode levar os alunos a interpretações equivocadas das medidas de variação.

O exemplo da figura 4 ilustra esta situação:

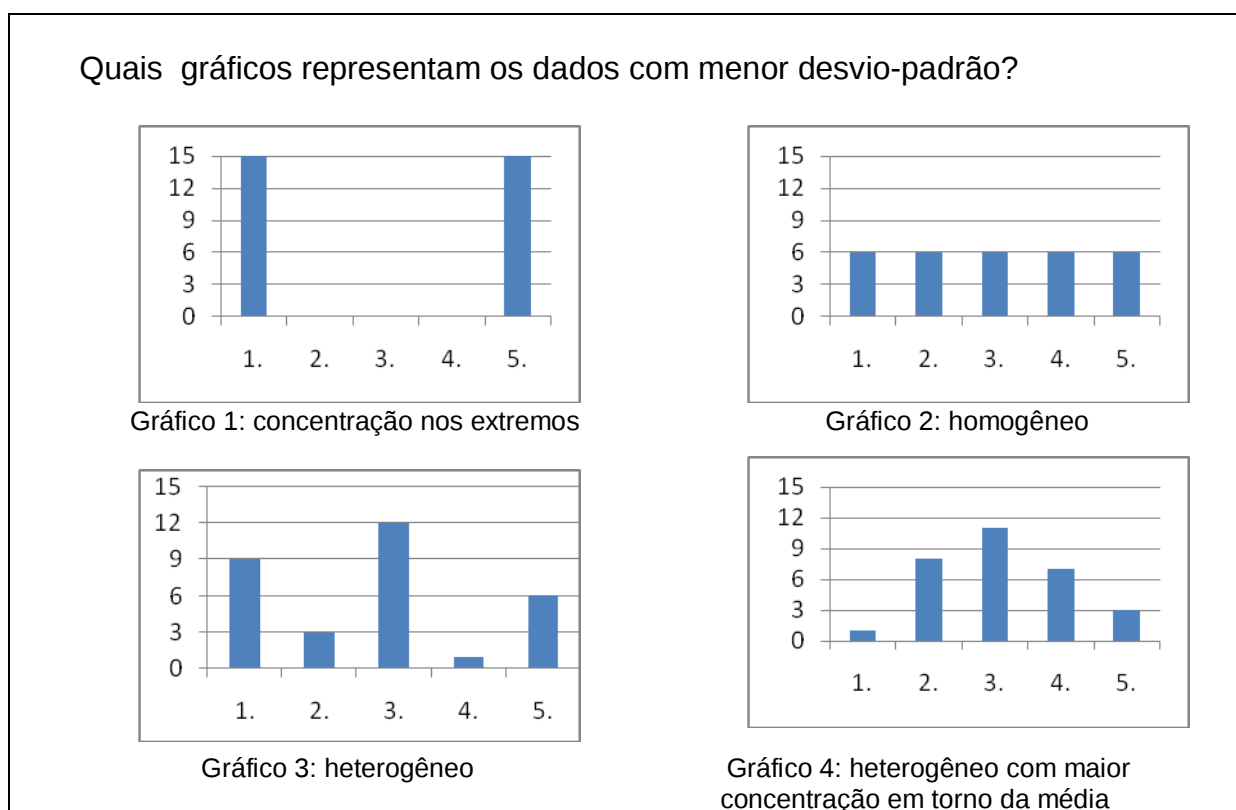


Figura 4 - Relação entre distribuição e variação dos dados

Se associarmos o valor do desvio-padrão, apenas com uma distribuição de maior ou menor concentração, o aluno poderá ser induzido a achar que os dados do Gráfico 1 da figura 4 possuem a menor dispersão, pois os dados possuem grande concentração nos extremos. Se relacionarmos o valor do desvio-padrão com homogeneidade entre os dados, os alunos serão induzidos a achar que o Gráfico 2 representa a menor variação e o Gráfico 3 a maior. Todavia, se relacionarmos a concentração, a homogeneidade com a dispersão dos dados em relação à média, eles terão mais chances de perceber que o menor desvio-padrão será encontrado nos dados representados pelo Gráfico 4.

Silva conclui seu trabalho mostrando que a percepção da variação entre os dados e não em relação a uma medida de tendência central, dificulta a construção de um raciocínio mais elaborado sobre variação:

[...] a utilização do termo variação levou alguns professores a perceber a variação entre as frequências de uma mesma categoria e escolher a categoria cujas frequências eram mais uniformes, ou seja, apresentavam menor variação entre elas, o que não faz sentido na análise de dados [...]. (SILVA, 2007, p.321).

Intervalo interquartil

O intervalo interquartil é uma medida-resumo de dispersão obtida pela diferença entre o terceiro (Q_3) e o primeiro quartil (Q_1). Este intervalo abarca 50% dos dados observados.

Segundo Bussab; Morettin (2005), o desvio-padrão e a média têm o inconveniente de serem medidas influenciadas por valores extremos. Por outro lado, o intervalo interquartil tem a vantagem de ser uma medida de dispersão robusta que não sofre tais influências, pois depende da posição ocupada pelos dados ordenados e não de seus valores. Enquanto o intervalo interquartil tem o benefício de ser uma medida robusta, ele tem a desvantagem de não considerar todos os dados. Assim, cabe ao pesquisador decidir qual medida é mais conveniente em cada caso. Logo, os alunos precisam construir um conjunto de conhecimentos que lhes permita tomar tal tipo de decisão.

4.1.3 Desigualdade de Tchebichev

De acordo com Triola (1999) o teorema de Tchebichev, afirma que:

a proporção (ou fração) de qualquer conjunto de dados a menos de k desvios-padrão a contar da média é sempre $1 - \frac{1}{k^2}$, onde k é um número positivo maior do que 1. Para $k = 2$, temos ao menos 75% de todos os valores no intervalo que vai de 2 desvios-padrão abaixo da média a 2 desvios-padrão acima da média. (TRIOLA, 1999, p. 43).

Este é um teorema muito importante para a análise de dados, pois garante que em qualquer distribuição, se criarmos um intervalo com base na média, considerando dois desvios-padrão de cada lado ($[\mu - 2 DP; \mu + 2 DP]$ ¹⁸), teremos aos menos 75% dos dados abrangidos pelo intervalo, não importando a forma da distribuição.

4.1.4 A Regra Empírica para determinar o percentual de dados contidos no intervalo construído com base na subtração e soma de desvios-padrão à média

Como já citado anteriormente, a noção de desvio-padrão é de difícil compreensão. Silva (2007) citando Hart (1984) e Loosen, Lioen e Lacante (1985) afirma que, para tornar esta noção significativa para o aluno, é fundamental apresentar a análise do desvio-padrão. Esses autores sugerem que se possibilite aos estudantes a oportunidade de se calcular a porcentagem de observações que estão a k desvios-padrão da média. A relação do desvio-padrão com a criação de intervalos de análise de dados ajuda o aluno a perceber a importância desta medida de dispersão e o sentido deste estudo.

Segundo Triola

Uma regra que auxilia na interpretação do valor de um desvio-padrão é a regra empírica, aplicável somente a conjuntos de dados aproximadamente em forma de sino. Nestas distribuições a média e o desvio-padrão estão relacionados com a proporção dos dados que se enquadram em determinados limites. Assim é que, com uma distribuição em forma de sino, temos 95% dos seus valores a menos de dois desvios-padrão da média. A regra empírica costuma ser designada abreviadamente como a regra 68-95-99 (TRIOLA, 1999, p. 81).

Esta regra simples pode ser verificada pelo aluno, bastando para isso que ele colete medidas antropométricas, calcule a média e o desvio-padrão, obtenha o intervalo com subtração e a adição de k desvios-padrão da média e conte quantos elementos estão contidos no intervalo. Geralmente, os resultados obtidos são: . cerca de 68% dos valores estão a menos de um desvio-padrão a contar da média;

¹⁸ μ representa a média e DP o desvio-padrão.

cerca de 95% dos valores estão a menos de dois desvios-padrão a contar da média; cerca de 99,7% dos valores estão a menos de três desvios-padrão a contar da média.

A razão de tomarmos medidas antropométricas deve-se ao formato de suas distribuições, ou seja, normalmente elas são, aproximadamente, simétricas e têm formato de sino ou triangular (Figura 5). Para distribuições assimétricas e com outros formatos, a regra 68-95-99 não é a mais adequada, mas ainda assim vale a Desigualdade de Tchebichev.

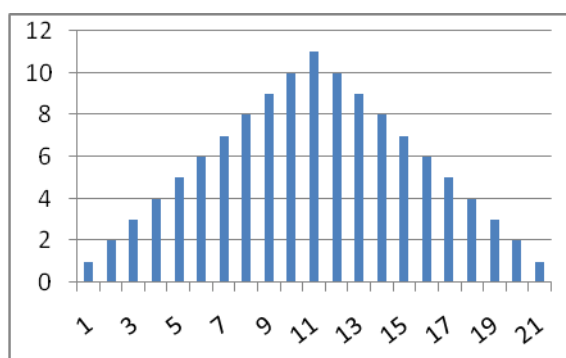


Figura 5 - Distribuição simétrica em formato triangular

Se tomarmos os dados representados na Figura 5 e calcularmos o percentual de valores contidos no intervalo criado com a subtração e adição de k desvios-padrão da média obteremos 65,3%, 95,4% e 100% para os intervalos criados com um, dois e três desvios-padrão da média, respectivamente. Tais valores aproximam-se das estimativas obtidas pela Regra Empírica: 68%, 95% e 99,7%.

4.1.5 Evento Raro

Segundo Meletiou (2000), um evento é considerado raro se ele estiver a uma distância acima de dois desvios-padrão da média.

A abordagem do evento raro mostra uma aplicação importante do desvio-padrão e possibilita uma articulação entre a Estatística Descritiva e a Inferencial. Com ela, podemos relacionar a regra empírica, a probabilidade e o formato da distribuição na análise dos dados, pois se uma distribuição for simétrica com formato

triangular, a probabilidade de um evento estar a mais de dois desvios-padrão da média é de, aproximadamente, 5%, ou seja, podemos considerá-lo um evento raro.

Como exemplo, podemos analisar a seguinte situação: a idade de um grupo de pessoas tem uma distribuição simétrica com formato triangular, idade média igual a 11 anos e desvio-padrão 4. Neste grupo, um valor de 20 anos pode ser considerado um evento raro?

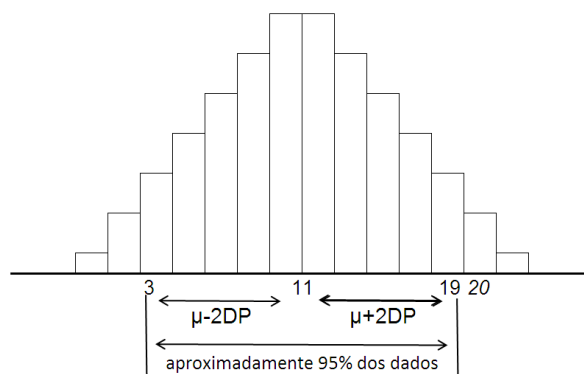


Figura 6 – Representação do intervalo criado com base em dois desvios-padrão da média.

Observando o gráfico, ilustrado na figura 16, percebemos que a idade 20 está situada fora do intervalo e que a chance de ocorrência desse evento é pequena. Assim, podemos considerar a idade 20 anos, como evento raro para esse grupo.

4.1.6 Z-escore

Segundo Callegari-Jacques

o z-escore pode ser interpretado como o número de desvios-padrão envolvidos no afastamento de um determinado valor x em relação à média. Em outras palavras, z é a diferença, em unidades de desvios-padrão, entre um valor x e a média. O z-escore também é conhecido como o valor de x padronizado. (CALLEGARI-JACQUES, 2003, p. 45).

O algoritmo que transforma x em z é:

$$z = \frac{x - \mu}{DP}$$

Onde μ e DP são a média e o desvio-padrão populacionais para a variável analisada.

4.1.7 O Gráfico de Pontos (Dotplot)

Segundo Vieira (2008, p. 62), o Gráfico de Pontos (Dotplot) “é um gráfico que representa cada observação obtida em uma escala horizontal, permitindo visualizar a representação dos dados ao longo desta”. Bussab e Morettin (2005) denominam-no como gráfico de dispersão unidimensional, cujos valores são representados ao longo de uma reta. Os valores repetidos podem ser representados por um número que indica as repetições, podem ser empilhados ou podem apresentar apenas o ponto mais alto da pilha.

Quando esse gráfico é representado por pontos empilhados para indicar valores repetidos, exige menor abstração do aluno quando comparado com o gráfico de barras, tendo em vista que cada ponto representa um valor como ilustrado na Figura 7. Esta representação torna-se bastante interessante quando o professor trabalha com dados colhidos dos próprios alunos e solicita que eles representem-nos por meio desse gráfico.

Deste modo, o gráfico representará a sala toda e cada ponto simbolizará o dado de um aluno, e este poderá se ver representado e comparar-se aos demais. Outra vantagem deste gráfico é a fácil percepção dos valores atípicos (outliers*) da distribuição dos dados, assim como da existência ou não da simetria dessa distribuição. Quando o objetivo da representação for as frequências de cada valor, este tipo de gráfico será menos adequado, pois exigirá a contagem dos pontos.

A Figura 7 ilustra um Gráfico de Pontos, representando a distribuição do número de irmãos de uma turma de alunos com valor atípico.

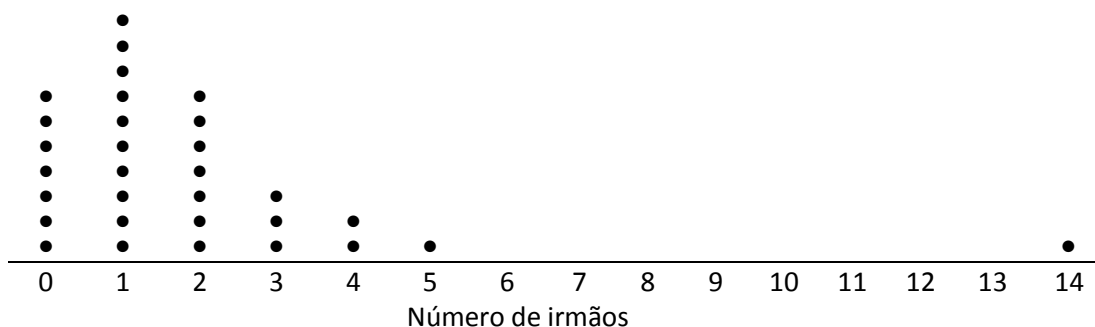


Figura 7 – Exemplo de Gráfico de Pontos com valor atípico

* Consideramos outliers os valores que estão fora do intervalo $[Q_1 - 1,5 \cdot (Q_3 - Q_1); Q_3 + 1,5 \cdot (Q_3 - Q_1)]$, sendo Q_1 o primeiro quartil, e Q_3 o terceiro quartil.

4.2 Os Parâmetros Curriculares Nacionais e a Variabilidade

Os Parâmetros Curriculares Nacionais do Ensino Fundamental (BRASIL, 1998) reservam um dos quatro blocos de conteúdos para o tratamento da informação. Este bloco envolve combinatória, probabilidade e estatística. A parte reservada à Estatística aborda coleta, organização (tabelas e gráficos) e descrição de dados; interpretação de dados apresentados por meio de tabelas e gráficos. Quanto às medidas de posição, há indicação para abordagem da média nos ciclos II e III¹⁹ e da média, moda e mediana no ciclo IV. Entretanto, não há nenhuma referência a qualquer estudo da variação dos dados em relação a tais medidas no Ensino Fundamental.

Nas orientações educacionais complementares aos Parâmetros Curriculares Nacionais do Ensino Médio – PCNEM+ (BRASIL, 2002), a Estatística compõe o eixo intitulado Análise de Dados. As medidas-resumo propostas neste eixo são média, moda, mediana, variância e desvio-padrão. Percebe-se, portanto, a introdução das medidas de variação apenas no Ensino Médio.

Quanto à distribuição dos conteúdos nas três séries do Ensino Médio, as medidas-resumo são sugeridas para a segunda série. Contudo, o documento adverte que a distribuição dos temas pode variar em função do número de aulas e do projeto da escola para aprofundamento de alguns tópicos ou inclusão de outros. Além disso, quando o professor tiver menos aulas de Matemática disponível, é proposto que ele trabalhe com as ideias centrais de cada tema. Por exemplo, quanto ao tema Estatística, é sugerida a descrição de dados, representações gráficas, média, moda e mediana. Assim, percebemos que o documento não considera a questão da variabilidade, como uma ideia central da estatística, o que contraria os resultados de pesquisas recentes na área, tendo em vista a importância capital da variabilidade na estatística, discutida na seção 4.4.

As Orientações Curriculares de Matemática para o Ensino Médio editadas pelo Ministério da Educação, em 2006, afirmam que o professor “deve possibilitar aos estudantes o entendimento intuitivo e formal das principais ideias matemáticas

¹⁹ Ciclo II corresponde aos 4º e 5º anos, ciclo III aos 6º e 7º anos e ciclo IV ao 8º e 9º anos.

implícitas em representações estatísticas, procedimentos ou conceitos” (BRASIL, 2006, p. 79). No exemplo proposto pelo documento, o aluno deve entender a influência que um valor extremo exerce sobre a média e sobre a mediana. Nós completaríamos o texto, sugerindo ao professor que apresentasse aos alunos atividades em que eles pudessem perceber qual medida de posição é mais adequada, relacionando o uso de tais medidas com a simetria da distribuição.

Quanto à apreensão das medidas-resumo, o documento sugere que:

Vale destacar a necessidade de se intensificar a compreensão sobre as medidas de posição (média, moda e mediana) e as medidas de dispersão (desvio-médio, variância e desvio-padrão), abordadas de forma mais intuitiva no Ensino Fundamental. (BRASIL, 2006, p. 79).

O documento como um todo insiste na importância da compreensão dos conceitos. Gostaríamos de enfatizar que, para o entendimento das medidas de dispersão propostas na citação acima, a relação da variação dos dados em torno da medida de tendência central é fundamental e as medidas de posição e de dispersão devem ser trabalhadas de forma integrada.

Em nossa avaliação, a inclusão do tema variabilidade nos PCN do Ensino Médio, apesar de discreta, é positiva. Entretanto, somos favoráveis que se dê maior ênfase à variabilidade na análise de dados, em razão de sua importância na Estatística. Defendemos que sejam considerados alguns aspectos da dispersão no Ensino Fundamental e que a abordagem da média seja acompanhada da amplitude, que é uma medida simples de variação com baixa complexidade sob o ponto de vista da cognição.

Não defendemos a inclusão do tema variabilidade nas ementas, simplesmente, como um tema a mais na lista de conteúdos existentes, antes uma abordagem da variação que “ajude o aluno a ultrapassar o conhecimento estatístico superficial ligado a procedimentos, terminologias e regras” (MELETIOU, 2000, p.89), pois, a autora acredita que a ênfase da variabilidade nos currículos poderá melhorar a compreensão dos conceitos estatísticos.

4.3 O Pensamento Estatístico

Silva (2007, p. 30), baseando-se em Snee (1990), Moore (1997) e em Wild; Pfannkuch (1999) sintetiza as definições de pensamento estatístico, como “as estratégias mentais utilizadas pelo indivíduo para tomar decisão em toda etapa do ciclo investigativo.” Tais decisões são tomadas frente à incerteza e, uma das importantes fontes dessa incerteza, é a variabilidade presente nos dados analisados. Assim, a variabilidade assume papel central no pensamento estatístico.

Para Snee (1990) o Pensamento Estatístico é constituído por processos de pensamento que reconhecem a presença da variação em tudo o que fazemos; Segundo Moore (1997), a onipresença da variabilidade e sua modelagem são elementos do Pensamento Estatístico que devem ser enfatizados no ensino; Wild; Pfannkuch (1999) afirmam que qualquer discussão séria a respeito do pensamento estatístico deve considerar a variação.

O Pensamento Estatístico envolve uma rica complexidade e, segundo Ben-Zvi; Garfield (2007) inclui: o conhecimento de como e por que se usa um determinado método, medida ou modelo; a compreensão profunda das teorias subjacentes aos processos e métodos estatísticos, bem como de suas limitações; reconhecimento de quando, como e por que os instrumentos de inferência existentes podem ser usados; a capacidade de compreender e utilizar o contexto de um problema para planejar, avaliar e tirar conclusões sobre as investigações.

Em suma, a descrição desses autores vai ao encontro da afirmação de Wild; Pfannkuch (1999), para quem o Pensamento Estatístico é a forma, como os profissionais de estatística pensam.

Ao discorrer a respeito das semelhanças e diferenças entre letramento, raciocínio e pensamento estatístico, DelMas (2002) afirma que este último vai além da aplicação dos conhecimentos básicos e do raciocínio no contexto:

Como tal, o pensamento estatístico é promovido quando a instrução desafia os estudantes a aplicar a sua compreensão em problemas do mundo real, para criticar e avaliar os projetos e as conclusões dos estudos ou para generalizar o conhecimento obtido em sala de aula a partir de exemplos e situações novas. (DELMAS, 2002, p. 6, tradução nossa).

Wild; Pfannkuch (1999) identificam quatro dimensões do processo de pensamento envolvido na resolução de problemas estatísticos: a investigativa, os tipos de pensamento, a interrogativa e as disposições.

A dimensão do *ciclo investigativo*: refere-se à forma como se age e aquilo que se pensa durante uma investigação estatística. Este ciclo é constituído por cinco fases: problema, planeamento, dados, análise e conclusões. A etapa denominada problema; refere-se à compreensão da dinâmica do sistema e definição do problema, na segunda etapa, ocorre o planeamento do sistema de medição, da amostragem e análise. Em um terceiro momento, vem a coleta e gerenciamento dos dados, seguidos pela exploração dos mesmos por meio de análises e formulação de hipóteses. O ciclo fecha-se na fase intitulada conclusão em que ocorrem as interpretações, o surgimento de novas ideias e a comunicação dos resultados.

Wild; Pfannkuch (1999) separam a dimensão *tipos de pensamento* em: tipos gerais de pensamento aplicado a um contexto estatístico e tipos fundamentais para o pensamento estatístico.

O Pensamento Geral refere-se às estratégias que incluem decidir, o que e como se vai realizar em uma atividade seu planeamento, a antecipação de problemas e a consciência das restrições. Outra forma geral de pensamento bastante utilizada na estatística é a modelagem, pois a construção de modelos visa a compreender e prever o comportamento de aspectos do mundo que nos dizem respeito. Também é incluída no pensamento geral a utilização de técnicas e a procura de explicações.

O Pensamento Fundamental, descrito pelos autores citados anteriormente, é importante para este trabalho, visto que inclui o reconhecimento da variação. Eles consideram pensamentos fundamentais da estatística: o reconhecimento da necessidade de dados, a transnumeração (diferentes representações dos dados que possibilitam melhor compreensão do sistema), a síntese entre o conhecimento do contexto e o estatístico, o uso de modelos estatísticos e a consideração da variação.

Ao discutir sobre a variação, Wild; Pfannkuch (1999) sugerem que a reflexão atual sobre o que é Estatística está relacionada ao conhecimento e à tomada de decisão sob a incerteza e, muito desta incerteza, decorre da onipresença da

variação. Vale ressaltar que a onipresença da variação é um dos elementos do pensamento estatístico proposto por Moore (1997). A importância da variação será abordada com maior detalhe na seção 4.4.

A terceira dimensão do Pensamento Estatístico proposta por Wild; Pfannkuch (1999) é a Interrogativa. Silva (2007) sintetiza esta dimensão, como sendo os questionamentos feitos pelo pensador enquanto resolve um problema. Complementa dizendo que, neste ciclo, são produzidas possibilidades, busca de informações e ideias para, posteriormente, interpretar o resultado, chegar à informação e tomar a decisão sobre o que deve ser mantido.

A dimensão intitulada *disposições* baseia-se nas qualidades pessoais do indivíduo que realiza a investigação. Dentre elas, destacamos a curiosidade, a conscientização, a imaginação que consiste na formação de modelos mentais para compreender a dinâmica essencial do problema, o ceticismo que significa a tendência para estar constantemente à procura de lógica e falhas nas novas ideias, o uso do raciocínio lógico para se chegar a conclusões válidas, a perseverança e a propensão para buscar um significado mais profundo, inclusive, para analisar novas ideias que estejam em conflito com as concepções do pesquisador.

As quatro dimensões descritas anteriormente visam organizar os elementos do pensamento estatístico. Ressaltamos que o indivíduo opera as quatro dimensões concomitantemente e quanto mais conhecimento de Estatística o aluno construir, melhor será a qualidade de seu pensamento estatístico.

A seguir, analisaremos o papel fundamental da variabilidade nas análises estatísticas.

4.4 A importância da variabilidade

A média aritmética parece ser a ferramenta estatística mais popular de todas, é utilizada por especialistas, profissionais das ciências humanas e exatas, da mídia, por políticos, estudantes, populares, etc., muitas das vezes, é confundida com moda ou com valor aproximado. Sendo utilizada adequadamente ou não, o fato é que esta medida-resumo é detentora de grande popularidade. O mesmo não se

pode dizer das medidas de variação, para as quais raramente se nota sua utilização. Alguns podem pensar que essas medidas, menos conhecidas, são simples caprichos de estatísticos ou conteúdos úteis para se cobrar em provas escolares, quando tais temas são abordados.

Vez ou outra ouvem-se piadas que, nas entrelinhas, colocam em dúvida a validade das conclusões da Estatística. Dentre elas, destacamos aquela que um estatístico é aquele que tendo a cabeça a arder de calor e os pés enterrados no gelo, ainda diz que na média está tudo bem! Outra anedota é a da alimentação média: uma pessoa alimenta-se com seis refeições diárias durante os 15 primeiros dias do mês e é privada de qualquer alimento durante os próximos 15 dias, mas, em média, teve três refeições diárias e está tudo bem! Qual seria a crítica subjacente a estas piadas?

Se a profundidade média de um lago é de 1,2 m, um adulto poderia caminhar de uma margem a outra tranquilamente sem correr o risco de ficar submerso? Suponhamos que pisos e revestimentos são melhor classificados, quando apresentam menor variação entre as medidas das peças. Um fabricante produz pisos retangulares cuja especificação é de 30 cm X 40 cm. Preocupado com a qualidade de seus produtos, ele solicita que seus funcionários meçam as peças e lhe informem a largura e o comprimento médios. Se as médias obtidas foram 30 cm X 40 cm, respectivamente, os pisos poderão receber uma boa classificação?

Com os dados fornecidos no parágrafo anterior, não poderemos responder às perguntas de modo afirmativo nem negativo, pois faltam elementos para melhor analisar as situações. Tais interrogações poderiam levar algumas pessoas a considerar a média uma ferramenta estatística que não tem utilidade prática, uma vez que pode levá-los a tomar decisões equivocadas ou não ajudar nas decisões, outros mais extremistas poderão dizer que a estatística só serve para manipular dados e enganar as pessoas.

Entretanto, a média é uma medida-resumo de grande utilidade e importância na análise de dados, útil para sintetizar um grupo, para fazer estimativas, obter o perfil ou representar um conjunto de dados, dentre outras utilidades.

O ponto-chave da questão é saber o quanto a média é representativa de um determinado grupo, pois médias iguais podem ser menos ou mais representativas.

Como exemplo, podemos comparar duas salas de aula cujas notas médias dos alunos de cada uma foi cinco. Na primeira turma, metade dos alunos obteve conceito zero e a outra metade, dez; na segunda, todos ficaram com cinco. Nos dois casos, a nota média foi cinco, mas no segundo a média foi bem mais representativa. Assim, quanto maior a concentração dos dados em torno da média melhor ela poderá resumi-los. Portanto, ao utilizar a média é fundamental observar, como os dados se distribuem em torno dela e medir a dispersão deles em relação a mesma.

Com a indicação da média acompanhada de uma medida de variação, poder-se-á responder com mais propriedade à questão da profundidade do lago e da classificação dos pisos descritos anteriormente. Desta forma, fica claro que uma medida de variação sempre deveria acompanhar a média. Isto informaria melhor o leitor sobre a síntese apresentada e evitaria alguns equívocos oriundos da utilização da média como medida-resumo única. A necessidade deste laço estreito entre média e variação é defendida também por Konold; Pollatsek (2002) ao escreverem que a noção de média, entendida como medida de tendência central, é inerente à variabilidade, de tal forma que é inseparável da noção de dispersão.

Considerar a variabilidade tem uma importância capital quando se utiliza a média, porém a relevância do tema não se restringe a esta medida de posição, ela vai além e, segundo Silva (2007), é considerada por vários autores, como o coração da Estatística. É a respeito da centralidade da variabilidade na estatística que discorreremos a seguir.

Wind; Pfannkuch (1999) afirmam que a variação é o elemento central das definições publicadas sobre o Pensamento Estatístico e qualquer discussão séria a respeito desse pensamento deve considerar o papel da variação. Uma destas definições é dada por Snee (1990) ao escrever sobre as contribuições do pensamento estatístico para a Qualidade Total. Conforme o autor, o pensamento estatístico é um processo de pensamento que reconhece que a variação está em torno de nós e está presente em tudo o que fazemos. Completa afirmando que todo trabalho é uma série de processo interligados e identificar, caracterizar, quantificar, controlar e reduzir a variação proporciona oportunidades de melhorar o processo.

De fato, a variabilidade é onipresente em nosso mundo, pois como afirmam Wild; Pfannkuch :

A variabilidade afeta todos os aspectos da vida e tudo o que observamos. Nada do que pode ser fabricado pode ser idêntico. Não existem dois organismos idênticos ou que reagem da mesma forma. Os organismos individuais são na verdade, sistemas em fluxo constante. (WILD; PFANNKUCH, 1999, p. 235, tradução nossa).

Se observarmos fenômenos naturais e sociais, produtos produzidos pelo homem ou gerados pela natureza, veremos o quanto a variabilidade está presente. Muito daquilo que pensamos ser igual deixa de ser quando o examinamos com maior acuidade. Tais variações são inerentes aos dados, mas não são as únicas, pois, quando nos propomos a fazer pesquisas que envolvem coleta de valores, surgem mais variações que podem ser provenientes da medição, do acaso e da amostragem. Desse modo, sempre temos de tomar decisões em meio a incerteza e a variabilidade e, para isso, necessitamos de instrumentos que nos ajudem a decidir da melhor forma possível.

Se não houvesse variação, qual seria a importância das medidas-resumo como média, moda e mediana? Por que estaríamos preocupados com o tamanho de amostras, margem de erro, distribuição, etc.? Caso os dados de que dispuséssemos fossem idênticos, não haveria necessidade de resumi-los. Se não houvesse diferença entre eles, bastaria tomarmos um exemplar da população estudada e já teríamos ali todas as informações de que necessitaríamos sobre o grupo todo.

Diante disto, nossa amostra sempre seria composta de um único elemento, totalmente representativa e sem margem de erro. Mas, o contexto em que nos deparamos diariamente e sobre o qual temos de decidir o tempo todo, é complexo e heterogêneo. Wild; Pfannkuch (1999) escrevem sobre o impacto prático da variação na vida das pessoas e nos negócios, além da relação da variação com os métodos estatísticos:

É a variação que torna a questão de causa e efeito difícil de ser resolvida, o que dificulta a descoberta de mecanismos. A variação é a razão pela qual as pessoas tiveram que desenvolver métodos estatísticos sofisticados para filtrar quaisquer interferência nos dados que possam ser considerados como ruído subjacente. (WILD E PFANNKUCH, 1999, p. 235, tradução nossa).

Snee (1999, p. 257) deixa claro em suas reflexões sobre o Pensamento Estatístico que “se não houvesse variação, não haveria necessidade de estatística nem de estatísticos”.

Em 1997, Moore apresenta as recomendações do Conselho da Associação Americana de Estatística e do Comitê de Currículo, nas quais constam os quatro

elementos principais do pensamento estatístico, a saber: a necessidade de dados, a importância da produção de dados, a onipresença da variabilidade, a medição e sua modelagem. É interessante notar que, destes quatro elementos, dois referem-se à variabilidade, confirmando nossa proposição de papel fundamental deste tema para a Estatística que, também, está em consonância com os elementos chave do Pensamento Estatístico propostos por Snee (1999): processo, variação e dados.

Em seu artigo, Snee (1999) discute os três elementos do Pensamento Estatístico, trata da relação da tecnologia e do desenvolvimento com a estatística. Segundo ele, a globalização dos mercados obriga as empresas a melhorarem cada vez mais para permanecerem competitivas e, para isso, muitas têm implementado abordagens estatísticas para processar a melhoria organizacional.

No texto citado anteriormente, o autor defende que toda atividade é um processo e tal processo é definido, como qualquer atividade que converte entradas em saídas. Toda pesquisa empírica está associada a processos, e estes fornecem o contexto para o trabalho estatístico.

Destarte, cabe às empresas aperfeiçoarem os processos para a melhoria da qualidade, mas tal aperfeiçoamento torna-se complicado, porque os processos variam. Então, surge a necessidade de lidar com a variabilidade que nos leva a fazer medições, como forma de caracterizar o processo que está sendo estudado, criando uma base numérica para comparações. Percebemos que o resultado do processo de medição são os dados, e estes também variam. Para analisar tais dados, usamos ferramentas estatísticas. “Os elementos dos métodos estatísticos são, conseqüentemente, a variação, os dados e as ferramentas estatísticas” (SNEE, 1999, p. 257).

A Figura 8 resume a relação entre os elementos do Pensamento Estatístico e os elementos dos métodos estatísticos, mostrando a centralidade dos dados e da variação.

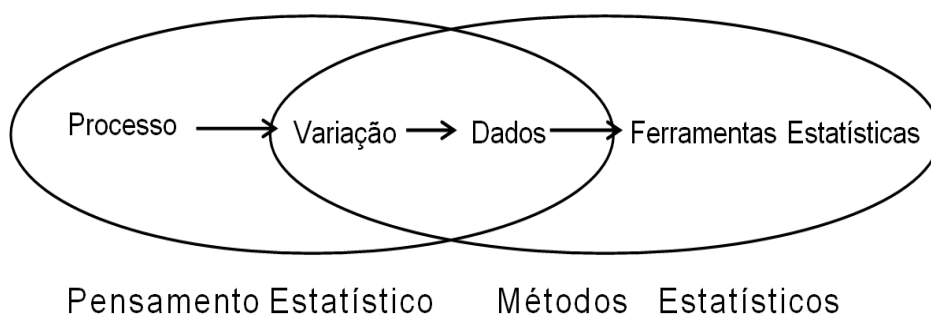


Figura 8 - Relação entre Pensamento Estatístico e métodos estatísticos
 Fonte: Snee (1999, p. 256).

Como já comentamos anteriormente, a variação pode ser inerente aos dados e/ou causada pelos seus processos de coleta, provocando impactos importantes na vida das pessoas e nos negócios. Wild; Pfannkuch (1999) oferecem respostas racionais para lidarmos com a variação, assim, podemos ignorá-la, permiti-la ou modificá-la.

Conforme relatam os autores citados, muitas vezes, consideramos cada objeto ou organismo como se fossem iguais ou diferentes de uma maneira deterministicamente conhecida. Desse modo, ao agirmos assim, estamos ignorando a variação. Uma outra atitude frente à variação é aceitá-la, ou seja, podemos investigar o modelo atual de variação e chegarmos a formas de trabalho em torno dela, como em nosso sistema de tamanho de roupas e calçados. Também costumamos aceitar a variação na fabricação de produtos que serão submetidos a uma diversidade de usos ou de condições, como a construção de estruturas metálicas com folga para dilatação quando expostas à mudança de temperatura. Em terceiro lugar, podemos mudar o padrão de variação para algo mais desejável. Como exemplo, podemos citar as tentativas de diminuição da variabilidade na fabricação de produtos que precisam ser produzidos com medidas exatas para garantir maior produtividade, como a fabricação de porcas e parafusos, com menor desperdício de peças refugadas por estarem fora dos limites estabelecidos. Assim:

Os estatísticos modelam a variação para as finalidades de predição, explicação ou controle. O controle é a modificação da variação para algo mais desejável. A previsão é a informação crucial de entrada para a “permissão,” [...] a explicação melhora nossa capacidade de fazer boas previsões, o que é necessário para o controle. (WILD; PFANNKUCH, 1999, p. 236, tradução nossa).

A variabilidade está presente em tudo, e a estatística fornece recursos que nos ajudam a trabalhar com esta característica inevitável dos dados. Por isso, desenvolver competências mínimas para perceber a variabilidade não deve ser uma habilidade exclusiva de especialistas, mas é uma necessidade para o cidadão comum, pois diariamente ele é bombardeado por informações e precisa saber analisar, questionar e criticar para aceitá-las ou não.

Terminamos esta seção apontando a importância da variabilidade para as organizações e, para isso, citamos os fundamentos do Pensamento Estatístico segundo a American Society for Quality: “todo trabalho ocorre em um sistema de processos interligados, a variação existe em todos os processos, o entendimento e a redução da variação são a chave para o sucesso” (1996 apud SNEE, 1999, p. 257). Seja para modificar ou aceitar a variação, é fundamental conhecê-la.

Tais ideias servirão de pano de fundo para nosso trabalho, visto que a sequência didática que organizamos, tem por objetivo fornecer condições para que futuros professores e, em consequência, seus alunos possam desenvolver o Pensamento Estatístico com base na apreensão da variabilidade.

CAPÍTULO V: A SEQUÊNCIA DIDÁTICA

A sequência é um dos componentes de nossa Engenharia Didática, constituindo a fase experimental, composta por 43 tarefas, organizadas em sete Situações Didáticas, nos termos de Brousseau (2008), ou seja, situações que permitiram aos participantes evoluírem de forma autônoma nas fases de ação, formulação e validação²⁰. Deste modo, a sistematização dos conhecimentos constituídos é feita pelo pesquisador ao final de cada situação didática.

Neste capítulo, abordamos os objetivos e os procedimentos utilizados em cada bloco de tarefas, descrevemos as aplicações, analisamos três atividades realizadas pelos alunos e apresentamos os critérios para escolha de cada uma destas.

5.1 Definição de situação, situação didática, tarefa e atividade

Situação é um modelo de interação entre um sujeito com um meio determinado. A ação do sujeito no meio para obter um estado favorável depende do conhecimento. Assim, o estudo da aquisição de conhecimento está relacionado com o estudo das situações. (BROUSSEAU, 2008).

Situação didática é o conjunto de relações estabelecidas entre aluno, ou grupo alunos, um certo meio e um sistema educativo (professor), para que adquiram um saber constituído. (BROUSSEAU, 1978 apud ALMOULOU, 2007).

Na concepção de Robert (1998), o termo *tarefa*, refere-se aos enunciados dos exercícios, atividades ou problemas, e, a expressão *atividade* está associada ao que fazem os alunos para resolver uma tarefa.

²⁰ Fase da ação é quando o aluno realiza ações imediatas, gerando um conhecimento de natureza mais operacional; formulação é a fase em que o sujeito deverá refletir sobre sua ação por meio do reconhecimento, identificação e reconstrução em um sistema linguístico; fase da validação é o momento da prova, no qual o sujeito deve justificar a exatidão e a pertinência de seus procedimentos.

Neste trabalho, consideramos uma situação didática como um conjunto de tarefas que colaboram na construção de um conceito, à medida que os alunos interagem com elas.

5.2 Os sujeitos da pesquisa

A formação foi realizada com um grupo de 17 alunos voluntários do quarto semestre da licenciatura em Matemática de uma universidade particular na cidade de Caraguatatuba, Estado de São Paulo. A idade mediana do grupo era de 29 anos, variando de 19 a 43 anos. Dos 17 participantes, sete já lecionavam. Dentre os que lecionavam, três eram professores eventuais (substitutos de qualquer disciplina), três ministravam aulas de Matemática e um informou apenas que lecionava. O tempo médio de magistério do grupo é de 12 meses, variando de 6 a 24 meses.

5.3 Descrição da aplicação

A formação constituiu-se de 43 tarefas²¹, interligadas entre si, visando à articulação de conhecimentos antigos com os visados, distribuída em 11 encontros, de 5 de setembro a 14 de novembro de 2009, perfazendo uma carga horária de 33h.

O conhecimento construído em uma atividade era utilizado como ferramenta de solução das tarefas seguintes, como proposto por Robert (1998) na primeira dimensão de análise de conteúdo. Em alguns momentos, confrontamos o conhecimento debatido em tarefas precedentes com questões que evidenciavam suas limitações, incentivando a busca de novos conhecimentos. Tal ideia é defendida por Brousseau ao afirmar que:

[...] Alguns outros problemas de introdução colocam em jogo um questionamento de concepções anteriores, ineficazes, com as questões que permitem aos alunos constatar essa ineficácia por eles mesmos, resultando

²¹ A Sequência completa está no apêndice A.

também em uma institucionalização* que se apóia nessas constatações.²²
(1991 apud ROBERT, 1998, p. 171)

Em dois encontros utilizamos a planilha eletrônica do software Excel, para que pudéssemos trabalhar com quantidade maior de dados reais, agilizando os cálculos e liberando os alunos para as análises dos resultados e conclusões.

5.4 Síntese e objetivos dos eixos em que foram divididas as situações

Tendo em vista que a ideia central do trabalho referiu-se às percepções de variabilidade e suas medidas e, considerando que optamos pela construção do conceito do desvio-padrão, a sequência iniciou-se pela abordagem da média aritmética, passando pela discussão de sua insuficiência como medida-resumo, e a necessidade de uma medida de variação relacionada a ela.

Em seguida, o estudante deveria notar a amplitude como medida de variação e depois perceber suas limitações. O desvio-médio foi utilizado com duas finalidades: a primeira, como instrumento mais eficaz do que a amplitude para medir a variação, tendo em vista que contempla todos os dados do conjunto e, a segunda foi para auxiliar o aluno na compreensão do conceito de desvio-padrão. Finalmente, a sequência abordou algumas aplicações do desvio-padrão, como a criação de intervalos por meio da subtração e adição de dois desvios-padrão à média para a análise de valores de variável e o z-escore para comparar afastamentos da média.

²²Traduzido com fins didáticos pelas professoras doutoras Cileda Q. S. Coutinho e Maria José Ferreira da Silva.

* A institucionalização referida pela autora é entendida por nós como institucionalização local.

Tabela 2 - Distribuição das tarefas em cada encontro com seus respectivos objetivos

Encontro	Tarefa	Objetivo(s)	Parte
1 ^o	1; 2; 3; 4; 5; 6; 7; 8; 9	Introduzir o conceito de média.	A
2 ^o	10; 11; 12	Perceber que a média pode resumir melhor os dados quando estes estão mais concentrado em torno dela.	B
3 ^o	13; 14; 15; 16	Ampliar a compreensão conceitual da média.	C
4 ^o	17; 18; 19; 20	Refletir sobre a insuficiência da média como medida-resumo única e utilizar a amplitude como medida de variação.	D
5 ^o	21; 22; 23; 24; 25;	Perceber as limitações da amplitude, como medida de variação e construir o algoritmo do desvio-médio.	E
6 ^o	26; 27; 28	Introduzir desvio-padrão e utilizá-lo na tomada de decisão entre dois conjuntos de dados	
7 ^o	29	Compreender o papel do desvio-padrão na criação de intervalos para análise dos valores da variável: a Regra Empírica.	F
8 ^o	30; 31; 32; 33; 34; 35; 36; 37		
9 ^o	38; 39	Utilizar o desvio-padrão para determinar um evento raro	G
10 ^o	40	Fazer estimativas com base na amostragem	
	41	Utilizar o desvio-padrão para padronizar notas e decidir qual delas é, relativamente, melhor.	
11 ^o	42	Utilizar o desvio-padrão para determinar se um evento é raro	
	43	Utilizar o desvio-padrão para padronizar notas e decidir qual candidato obteve o melhor desempenho.	

A figura 9 mostra um esquema que representa a síntese do encadeamento dos principais conceitos trabalhados na formação com os universitários. As setas mais largas indicam concatenação das ideias abordadas em uma sucessão temporal, enquanto as outras e os conectores estreitos representam as ligações das ideias principais com outras noções importantes para a percepção e medida da variabilidade. Representamos algumas das articulações possíveis entre os conceitos, porém existem outras conexões que não foram indicadas.

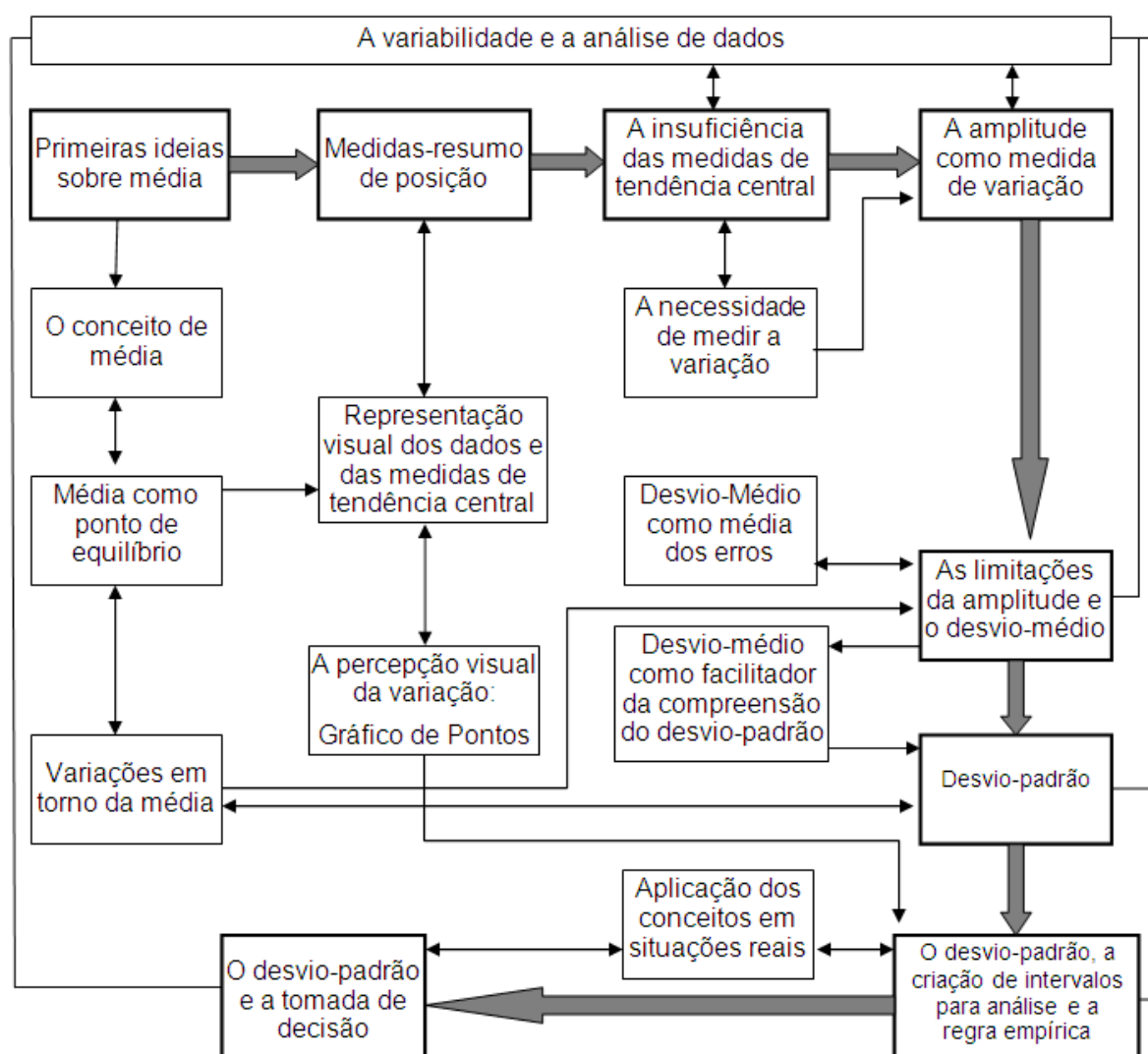


Figura.9. – Síntese do encadeamento dos principais conceitos trabalhados.

5.4.1 Tarefas agrupadas, conforme as noções visadas

Parte A - Primeiras ideias sobre média aritmética

Número de tarefas: 9 (da 1ª a 9ª)

Duração: 3 horas

O objetivo foi introduzir o conceito de média, como medida de partição equânime, como síntese e como instrumento de estimativa. As duas primeiras tarefas são questões de partição, as cinco seguintes sobre síntese e a oitava trata de estimativa. A última apresenta a média, como o conhecimento adequado para resolver os itens anteriores. É feita uma institucionalização local.

Parte B - Média, quartis e introdução à variabilidade.

Número de tarefas: 3 (da 10^a a 12^a)

Duração: 3 horas

O objetivo foi utilizar a média e a mediana para obter uma medida que resumisse os dados de um grupo, percebendo a importância da variabilidade. Para isso, utilizamos a representação gráfica (Gráfico de Pontos), o valor mínimo, o máximo e os quartis para complementarem a descrição dos dados, promovendo a percepção da variabilidade.

Após esta seção, esperávamos que os alunos percebessem que em um conjunto de dados dispersos, a média e a mediana são menos representativas do que em um grupo em que os dados estão concentrados em torno delas. Assim, a questão da variação começará a ter importância para os estudantes, levando-os a perceberem a relevância de analisá-la, quando se calculam a média e a mediana ou quando lemos alguma informação nas quais estas medidas foram informadas.

Parte C - Ampliação da compreensão do conceito de média

Número de tarefas: 4 (da 13^a a 16^a)

Duração: 3 horas

Os objetivos foram ampliar a compreensão conceitual da média; interpretar a média como ponto de equilíbrio entre os dados. Perceber que este equilíbrio ocorre quando a soma dos desvios de cada dado em relação à média é zero, isto é, a soma dos valores absolutos dos desvios anteriores à média são iguais à soma desvios posteriores; estimular a observação e comparação das diversas distribuições com a mesma média.

Parte D - A insuficiência da média e a amplitude como medida de variação.

Número de tarefas: 4 (da 17^a a 20^a)

Duração: 3 horas

O objetivo principal foi mostrar a insuficiência da média, como medida-resumo e utilizar a amplitude como medida de variação que a complementa. Para

isso, apresentamos conjuntos de elementos cujas médias eram iguais, mas as variações bem diferentes. O aluno deveria perceber a necessidade de se conhecer a variação dos dados para tomar decisão. Além da observação da variação, ele deverá medi-la por meio da amplitude, percebendo que este instrumento é um recurso simples de medida de dispersão.

PARTE E - A busca de instrumentos mais abrangentes para medir a variação em razão da limitação da amplitude.

Número de tarefas: 8 (da 21^a a 28^a)

Duração: 6 horas (dois encontros)

O objetivo desta seção foi levar o aluno a perceber a insuficiência da amplitude como medida de variação e encontrar medidas que considerassem todos os dados.

As tarefas iniciais foram elaboradas de maneira que o aprendiz precisasse considerar a variação para decidir entre dois grupos, mas não poderia se utilizar das amplitudes, pois elas eram iguais. Assim, foram motivados a encontrar duas medidas que satisfizessem estas condições: o desvio-médio e o desvio-padrão.

As quatro primeiras questões foram organizadas, visando à construção da noção de desvio-médio. Para resolver os problemas, eram necessários cálculos de distâncias médias em relação a um valor fixo e o cálculo de erros médios, de maneira que a utilização do desvio-médio estivesse implícita nas resoluções. Após este processo, o algoritmo do desvio-médio foi apresentado e feita a institucionalização.

Em seguida, discutimos as desvantagens da utilização do desvio-médio e propusemos aos alunos a busca de outra medida de variabilidade que não empregasse a função módulo, mas que resolvesse o problema da soma dos desvios em relação à média ser zero. O objetivo era que eles construíssem o algoritmo do desvio-padrão.

Parte F - O papel do desvio-padrão na criação de intervalos para análise dos valores da variável.

Número de tarefas: 9 (da 29^a a 37^a)

Duração: 6 horas (dois encontros)

A finalidade era descobrir experimentalmente o percentual de dados compreendidos em intervalos criados, a partir de um e dois desvios da média

Esta seção foi dividida em duas partes: na primeira, foi apresentado um jogo em que os alunos, em duplas, deveriam lançar dois dados, somar os resultados e marcá-los em um quadro, formando um Gráfico de Pontos. Cada aluno escolheria, antes do início da partida cinco somas, e uma ficava sem ser escolhida. Ganharia o jogo quem acertasse a maior quantidade de somas. O jogo termina quando uma das colunas é preenchida completamente. Em seguida, as duplas deveriam calcular a média das somas obtidas, o desvio-padrão, criar um intervalo com a subtração e adição de um e dois desvios-padrão e, finalmente, calcular o percentual de dados compreendidos pelo intervalo criado.

Na segunda parte desta seção, foi proposto que os alunos calculassem as médias, os desvios-padrão e os intervalos com o(s) desvio(s)-padrão das medidas antropométricas de um grupo de estudantes em idade escolar. Eles deveriam levantar hipóteses sobre a porcentagem de dados que deveriam estar contidos nos intervalos criados com base no desvio-padrão.

Parte G: Aplicação do intervalo criado com base no desvio-padrão na tomada de decisão.

Número de tarefas: 6 (da 38^a a 43^a)

Duração: 9 horas (três encontros)

Nesta última parte da formação, pretendíamos que os alunos fossem capazes de utilizar o intervalo criado baseado no desvio-padrão e na média para tomar decisões quando analisam dados.

Aqui serão exploradas as ideias de eventos raros relacionando-as com os elementos que estão fora do intervalo $[\mu - 2DP; \mu + 2DP]$.

Finalmente, esperávamos que os alunos aplicassem os conceitos construídos, em padronização de notas.

Estas sete partes podiam ser reagrupadas em três grandes situações didáticas, nas quais o aluno deveria se desenvolver de forma autônoma: Situação Didática 1 cujo objetivo principal era a introdução e a construção do conceito de média; Situação Didática 2 que visava a introdução e construção do conceito de variação pelo uso de medidas de dispersão – amplitude, desvio-médio e desvio-padrão; Situação Didática 3 que tratava da aplicação do conceito de dispersão para tomada de decisões.

Os conceitos estavam encadeados de tal forma que, para resolver a Situação 2, os alunos necessitariam dos conhecimentos construídos nas Situações 1 e, para resolver a Situação 3, necessitariam do conhecimento das anteriores. Assim, para fins de avaliação e validação, analisaremos apenas uma tarefa da Situação 2 e duas da Situação 3.

5.5 Análise de algumas tarefas

Para analisar, selecionamos três tarefas que nos permitissem avaliar e diagnosticar a construção e mobilização dos conhecimentos estatísticos trabalhados nas situações didáticas

Na escolha das tarefas, não consideramos aquelas com o objetivo de realizar a revisão dos conhecimentos necessários ao estudo da variabilidade, ainda que esses conhecimentos tenham sido abordados, considerando aspectos da dispersão, assim, procuramos selecionar tarefas que tratassem de conhecimentos significativos para a percepção e análise da variabilidade.

5.5.1 Análise da tarefa número 20.

A tarefa descrita faz parte da seção “E” que trata da busca de instrumentos mais abrangentes para medir a variação em razão da limitação da amplitude. Escolhemos essa tarefa porque é a primeira que leva o aluno a considerar todos os dados para medir a dispersão em relação a um ponto

determinado e utiliza noções trabalhadas anteriormente. Além disso, aborda aspectos da variação que serão úteis para a compreensão do conceito de desvio-padrão.

Voltemos às medidas antropométricas do aluno Karlo. O professor convidou 12 alunos da 6ª A e 12 da 6ª D e solicitou que medissem o comprimento do pé de Karlo. Os resultados foram expressos nos gráficos abaixo.

Gráfico 1:

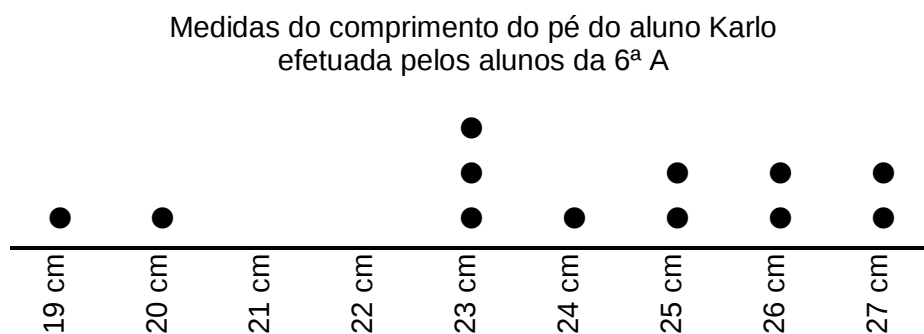
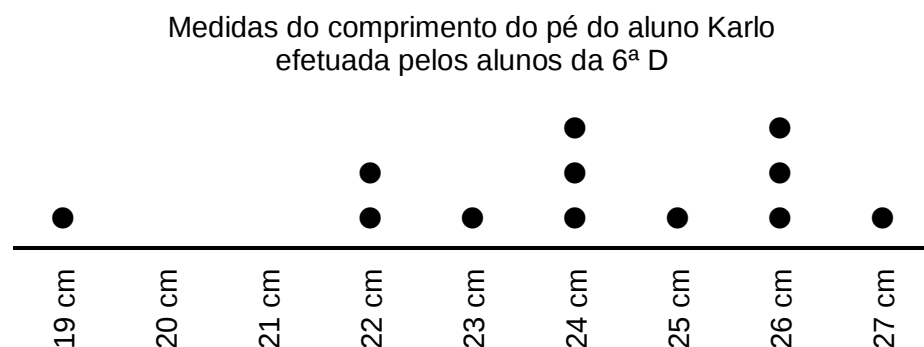


Gráfico 2:



Questão :

- Em qual conjunto (6ªA ou 6ªD) de dados, houve maior variação?
- Calcule a amplitude de cada conjunto. As amplitudes são iguais? As variações são iguais?
- Suponha que a medida real do pé de Karlo seja 24 cm, qual grupo se afastou mais da medida real (ou variou mais em relação à medida 24 cm)? Justifique.
- Em média, quantos centímetros erraram²³ os alunos da 6ª A? E os da 6ª D?²⁴

Figura 10 - Tarefa 20

²³ Os alunos foram informados que deveriam interpretar o termo *erro*, como o afastamento da medida efetuada em relação ao tamanho real.

²⁴ Para solução, o participante recebeu um esqueleto de uma tabela a ser preenchida

5.5.1.1 Análise teórica

O objetivo desta tarefa foi fazer com que o aluno percebesse a insuficiência da amplitude, como medida de variação e encontrasse uma forma de medir a dispersão em relação a uma medida exata que leve em consideração todos os dados. Esta comparação com a medida exata tinha por objetivo levar o aluno a perceber, em atividades posteriores que a média é um representante do grupo e os desvios podem ser encarados, como medidas de erro dessa representação. Ao final da atividade, o aluno deveria perceber que determinar o desvio-médio equivale a encontrar a média dos erros.

Quanto aos conhecimentos prévios necessários para a realização da atividade, destacamos a habilidade para leitura e interpretação de Gráfico de Pontos; cálculo de amplitude e de média aritmética para n valores.

Os conhecimentos visados são a percepção da insuficiência da amplitude, como medida de variação e a construção da noção de desvio-médio.

Para atingir os objetivos propostos, fizemos certas escolhas didáticas na elaboração da tarefa que passamos a descrever nas linhas seguintes. Nos itens “a” e “b”, os alunos foram questionados sobre a variação dos dados sem referência a nenhuma medida de tendência central. Eles ficaram livres para analisá-las, tomando alguma medida como referência ou simplesmente observando a variabilidade entre os dados. No item “c” os estudantes deveriam examinar a variabilidade em torno da medida 24 cm. Finalmente, no item “d”, foi solicitado o erro médio das medidas efetuadas em relação à medida exata.

Outra escolha refere-se ao Gráfico de Pontos e não de colunas, pois ele facilita a visualização dos dados individualmente, auxiliando o cálculo do desvio de cada ponto em relação a uma medida específica.

Os dois gráficos foram elaborados com amplitudes iguais para forçar o aluno a buscar outra medida de variação. A distribuição dos dados no gráfico foi feita de forma a dificultar a percepção da diferença da variabilidade existente, pois o objetivo era que os alunos a medissem. Esta forma de elaborar o enunciado do problema, forçando os alunos a construírem o conhecimento visado, está baseada na proposta

de Robert (1998), ao fornecer sugestões para elaboração dos enunciados dos problemas que objetivam a apropriação de noções matemáticas por parte dos alunos. Assumimos, assim, a aplicabilidade desta proposta também para a apropriação de noções estatísticas.

Segundo Almouloud (2007), a escolha de problemas para que o aprendiz adquira novos conhecimentos que estejam totalmente justificados pela lógica interna da situação, é uma das características da situação adidática. Assim, devemos propor uma tarefa cuja melhor solução seja o conhecimento a ensinar, seguindo-se os pressupostos da Teoria das Situações Didáticas.

Na tarefa em questão, o estudante recebe, inicialmente, apenas os dois gráficos e o item “a” para que os explore e perceba as limitações da análise visual. Espera-se que ele seja impelido a buscar uma medida de variação sem nenhuma indicação fornecida pelo enunciado do problema ou pelo professor. Vale lembrar aqui uma das etapas da dialética ferramenta-objeto, que se propõe que sejam apresentados aos alunos problemas cuja solução imediata não poderá ser totalmente obtida por eles, “porque o objeto de ensino é a ferramenta adequada para resolver o problema” (ALMOULOU, 2007, p. 63).

Em um segundo momento, o aluno recebe as questões dos itens “b”, “c” e “d”. No item “b”, espera-se que ele perceba que a amplitude é insuficiente para comparar as variações dos dois conjuntos, pois elas são iguais. A estratégia de solicitar a amplitude e a comparação das variações em dois gráficos, nos quais é difícil identificar a maior dispersão e que possuem amplitudes iguais, é uma variável didática que visa a levar o aprendiz a perceber a limitação dos conhecimentos anteriores e impulsioná-lo à construção de novos conceitos. Segundo Robert, alguns problemas utilizados para introduzir conceitos

colocam em jogo um questionamento de concepções anteriores, ineficazes, com as questões que permitem aos alunos constatar essa ineficácia por eles mesmos, resultando também em uma institucionalização que se apóia nessas constatações e sua resolução. (ROBERT, 1998, p. 171)²⁵.

No item “c”, é usada a ideia de afastamento em relação ao valor real, na tentativa de levar o aluno a associar esses afastamentos com a variação. Para

²⁵ Traduzido com fins didáticos pelas professoras doutoras Cileda Q. S. Coutinho e Maria José Ferreira da Silva

resolver o item “d”, é preciso calcular o erro médio de cada sala. Utilizamos a expressão “erro como distância”, para evitar o aparecimento de medidas negativas e, conseqüentemente, a necessidade do uso do valor absoluto. Após a resolução das duas últimas questões, esperamos que o aluno tenha calculado o desvio-médio em relação à medida 24 cm, antes mesmo desse conceito ter-lhe sido apresentado formalmente.

As ferramentas utilizadas para resolução desta tarefa fazem parte dos conhecimentos prévios dos alunos. Alguns destes conteúdos, mantêm uma correspondência com as noções utilizadas no algoritmo do desvio-médio, a saber: os afastamentos em relação à medida 24 cm correspondem aos desvios absolutos em relação à média dos valores e a média dos erros corresponde à média dos desvios.

Destarte, é possível mobilizar os conceitos já construídos de média e de distância (afastamento), como ferramenta para resolver o problema e construir o conceito de desvio-médio. O conceito de média que era objeto nas tarefas anteriores de nossa sequência, deveria ser utilizado como ferramenta para construção do objeto desvio-médio.

Segundo Robert (1998), esta dialética ferramenta-objeto é útil na elaboração dos cenários que ajudam na introdução de noções matemáticas, segundo nossas premissas, também é adequada para introdução de noções estatísticas.

Soluções esperadas

a) Em qual conjunto (6ªA ou 6ªD) de dados, houve maior variação?

Uma solução correta para este item seria:

R: Sem efetuar cálculos é difícil precisar, pois a distribuição dos dados nos dois gráficos dificulta a comparação.

O trabalho de Silva (2007) mostra que a abordagem da variação como homogeneidade entre os dados e não em relação à tendência central tende a favorecer análises de variação equivocadas. Aqueles que associarem homogeneidade entre os dados com menor variação, poderão escolher o Gráfico 1

como o de menor dispersão, tendo em vista que as frequências das medidas 25 cm, 26 cm e 27 cm são iguais.

- b) Calcule a amplitude de cada conjunto. As amplitudes são iguais? As variações são iguais?

Solução correta:

Tabela 3 - Cálculo da amplitude

	Mínimo	Máximo	Amplitude
6ª série A	19 cm	27 cm	8 cm
6ª série D	19 cm	27 cm	8 cm

R: As amplitudes são iguais, mas as dispersões parecem ser diferentes. A observação dos gráficos e o cálculo das amplitudes são insuficientes para precisar em qual dos conjuntos há maior variação.

- c) Suponha que a medida real do pé de Karlo seja 24 cm, qual grupo afastou-se mais da medida real (ou variou mais em relação à medida 24 cm)? Justifique.

Uma solução correta para o item seria:

R: 6ª A, pois a soma dos erros da 6ª A foi maior do que a soma dos erros na 6ª B. Outra justificativa possível seria o cálculo da média dos erros (item d) de cada sala.

Todavia, alguns alunos podem considerar que as medidas da 6ª A afastaram-se mais, justificando que a frequência do valor 24 cm nesta série é um e na 6ª D é três. Aqueles que assim raciocinarem, estarão cometendo o equívoco de fazer observações localizadas sem considerar todos os dados, como salienta Silva (2007).

- d) Em média, quantos centímetros erraram os alunos da 6ª A e os da 6ª D?

Solução correta:

Tabela 4 - Cálculo do erro médio

Gráfico 1: 6A			Gráfico 2: 6D		
Medida	Tamanho real	Erro como distância ou variação em relação ao 24	Medida	Tamanho real	Erro como distância ou variação em relação ao 24
19	24	5	19	24	5
20	24	4	22	24	2
23	24	1	22	24	2
23	24	1	23	24	1
23	24	1	24	24	0
24	24	0	24	24	0
25	24	1	24	24	0
25	24	1	25	24	1
26	24	2	26	24	2
26	24	2	26	24	2
27	24	3	26	24	2
27	24	3	27	24	3
Soma		24	Soma		20
Erro médio		2	Erro médio		1,67

R: A sala com maior variação ou maior erro médio é a 6ª A, o que chamamos de erro médio pode ser interpretado como uma medida de variação média em relação ao valor correto.

Segundo Batanero (2000), um dos erros cometidos pelos alunos no cálculo da média é ignorar a frequência igual a zero. Aqueles que assim procederem obterão erro médio igual a 2,18 cm para a 6ª A e 2,22 cm para a 6ª D.

Análise, segundo os níveis de conhecimento de Robert (1998), para esta tarefa

Para esta atividade, identificaremos o conhecimento técnico, disponível e mobilizável de acordo com os seguintes critérios:

a) Técnico: se o aluno for capaz de calcular o mínimo, máximo, a amplitude e observar a variação entre os dados.

b) Mobilizável: se o aluno calcular a amplitude, perceber sua insuficiência, medir a distância de cada dado até o ponto 24 e conseguir determinar qual grupo obteve a maior variação, com auxílio de informações do enunciado ou do pesquisador.

c) Disponível: o aluno foi capaz de calcular o desvio-médio já no item “c”, pois não há dica até este momento (só no item “d”).

5.5.1.2 *Análise a posteriori*

A maior parte dos alunos observou a variação entre os dados e não em torno de alguma medida de tendência central. Muitos associaram variação com diferença de frequência entre medidas, justificando que a variação da 6ª A é menor em razão das frequências iguais das medidas 25 cm, 26 cm e 27 cm. O ocorrido com estes alunos está de acordo com o previsto na *análise teórica*: confusão entre variação em torno de uma medida de tendência central com variação entre os dados, confirmando os resultados observados na pesquisa de Silva (2007).

Os gráficos da figura 11 mostram o procedimento de um aluno ao traçar uma linha sobre os pontos. Ele se utilizou da imagem criada por essa linha para escolher o Gráfico 2 (6ª D), como o de maior variação, quando o correto seria escolher o Gráfico 1 (6ª A).

Gráfico 1:

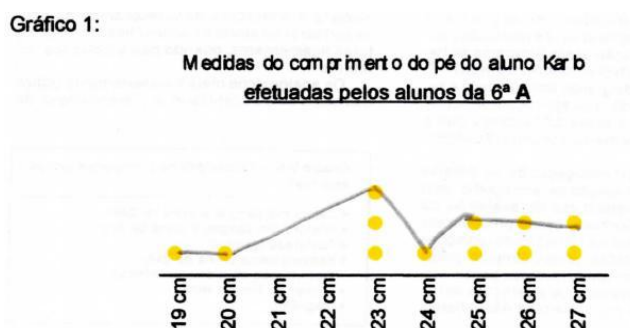


Gráfico 2



Questão:

- a) Em qual conjunto (6ª A ou 6ª D) de dados houve maior variação?

6ª D

Figura 11 - Interpretação de variação como diferença entre frequências.

Talvez a intenção do aluno fosse construir gráficos de linha para auxiliar na análise, mas tal estratégia pode ter levado-o a conclusões equivocadas, pois, como assinala Silva:

O objetivo do gráfico de linha é verificar se há uma regularidade da medida ao longo do tempo ou se há oscilação grande. Porém, essa comparação se faz entre as observações [...] que pouco contribui para o raciocínio em torno de uma medida de tendência central. (SILVA, 2007, p. 92).

No momento da socialização, uma aluna observou que analisar a variação entre os dados não era errada, já que a questão não indicava nenhuma referência por meio da qual se deveria analisar a dispersão dos dados. Concordamos com a observação da aluna no que tange ao enunciado do item “a”, pois a questão proposta só se referia à variação, sem indicar nenhuma medida como referência.

Diante disso, percebemos a necessidade de se deixar claro para os alunos que as medidas de variação como intervalo interquartil, desvio-médio e desvio-padrão referem-se às dispersões em relação às medidas de tendência central. Em nosso trabalho, quando nos referimos ao tema “variabilidade” na Estatística, não estamos fazendo alusão à variedade, mas, sim, a desvios em relação a um ponto. Conjecturamos que uma possível forma de minimizar este obstáculo didático seja a integração das medidas de tendência central, a visualização gráfica da distribuição e as medidas de variação.

Reading; Shaughnessy (2004) apresentam a integração das medidas de tendência central com as de dispersão, como uma das formas relevantes para o ensino da variabilidade:

Não se deve separar o estudo das medidas de tendência central das medidas de dispersão. É frequente, em situações de aprendizagem, os professores negligenciarem totalmente o estudo da propagação ou artificialmente separá-lo para o estudo de tendência central. Os educadores precisam incentivar discussões que vá além das medidas de centro, articulando, tanto quanto possível, raciocínio sobre variação com o de tendência central. (READING; SHAUGHNESSY, 2004, p. 223, tradução nossa).

Quanto ao item “c”, os alunos mostraram mais dificuldade na resolução quando comparado aos itens “a” e “b”. Uma das dificuldades foi considerar a frequência de cada ponto, ou seja, se a frequência da medida 23 cm era três, eles deveriam considerar três afastamentos medindo 1 cm, mas alguns alunos

consideravam esta medida apenas uma vez. Conjecturamos que esta dificuldade deve-se ao fato dos alunos olharem para os pontos como colunas.

Assim, sobre cada valor haveria uma coluna e eles teriam raciocinado que deveriam calcular as distâncias das colunas em relação ao valor exato e não as distâncias de cada ponto. Uma de nossas opções por utilizar o Gráfico de Pontos foi para incentivar o uso de todos os dados, mas em alguns casos a dificuldade persistiu. Este é um tema que pretendemos pesquisar futuramente: será que o uso do gráfico de colunas, desde as séries iniciais constitui-se em um obstáculo didático, para que os alunos considerem todos os pontos no cálculo da variação? Consideramos a hipótese de que seria mais adequado iniciar o ensino das representações gráficas pelo Gráfico de Pontos.

A escolha das duas variáveis didáticas - propor dois conjuntos de dados com amplitudes iguais e dois gráficos que dificultassem a percepção visual do grupo com maior variação - foram importantes a fim de criar uma situação fundamental²⁶ para introduzir uma medida de variação que considerasse todos os dados.

Embora alguns participantes mostrassem dificuldades para calcular o afastamento em relação a um ponto determinado, vários alunos conseguiram medir o desvio, tomando como referência o valor verdadeiro. Para resolver este item foi necessário ir além da aplicação imediata da amplitude e da observação dos gráficos, conhecimentos que já haviam sido abordados em tarefas anteriores.

Destarte, classificamos estes alunos no nível mobilizável, proposto por Robert (1998), pois na análise teórica apresentada no item 5.5.1.1, consideramos no nível mobilizável para esta atividade, os participantes que viessem a calcular a amplitude, percebessem sua insuficiência, medissem a distância de cada dado até o ponto 24 e conseguissem determinar qual grupo teve a maior variação, com auxílio de informações do enunciado ou do pesquisador (formador).

Diante da dificuldade de se saber qual conjunto variou mais, um aluno decidiu somar os desvios à direita e à esquerda da média e concluiu que os desvios da 6ªA somavam 12 cm de cada lado e da 6ª B, 10 cm de cada lado. Diante do

²⁶ Segundo Almouloud (2007) uma situação fundamental é a noção a ensinar que é a resposta considerada mais adequada ao problema.

constatado, concluiu que a maior variação era a da 6^a A. O aluno teve uma compreensão mais elaborada de variação ao associá-la com os afastamentos.

Nenhum aluno chegou a calcular o afastamento médio para tomar a decisão do item “c”, por isso, ninguém foi classificado no nível de conhecimentos disponível.

Após a análise deste item, fomos rever como Bussab e Morettin (2005) introduzem o desvio-médio e percebemos que poderíamos ter apresentado cada gráfico com quantidades diferentes de elementos, porque a soma dos desvios seria insuficiente para determinar qual grupo variou mais. Deste modo, a questão propiciaria uma reflexão sobre o uso da média dos afastamentos que seria o próprio cálculo do desvio-médio.

Para melhorar a tarefa, a sugestão seria retirar a medição correspondente a 20 cm do grupo da 6^a A, o que resultaria na seguinte distribuição:

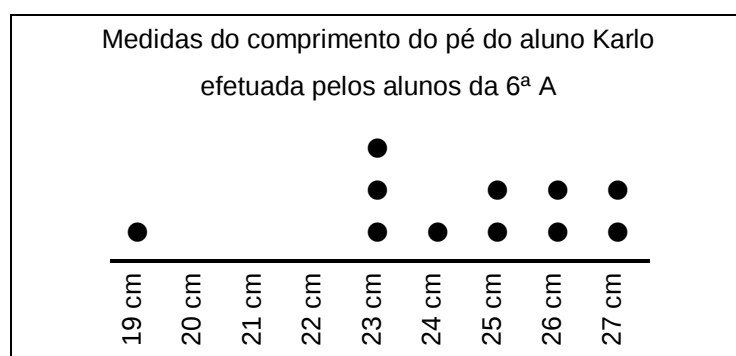


Figura 12 – Modificação do Gráfico 1 do enunciado da tarefa 20.

Assim, as somas dos desvios das 6^a A e 6^a D seriam iguais, apesar dos desvios médios serem diferentes, isto é, 1,82 cm e 1,67 cm, respectivamente.

Quanto ao item “d”, os alunos não apresentaram dificuldades para resolvê-lo, salvo alguns que efetuaram os cálculos, como se todas as medidas tivessem frequência igual a um, e outros que não consideram o desvio igual a zero para o cálculo da média.

O objetivo de se calcular o erro médio era aproximar o termo desvio-médio de uma palavra mais comum ao vocabulário do aluno. Mas o termo *erro* gerou dúvidas em um dos grupos, quando este interpretou (corretamente) *erro* zero, como acerto. Desta forma, surgiu a dúvida do grupo de se considerar ou não o zero, relativo a esse *erro* solicitado. O instrutor questionou com o grupo sobre o uso ou não de

valores iguais a zero no cálculo da média dos erros. Ao final, eles perceberam que deveriam considerar o zero

Acreditamos que a resolução do item foi facilitada pela apresentação da tabela para que os alunos pudessem completá-la.

Após a aplicação e análise da atividade número 20, constatamos algumas dificuldades dos alunos apontadas anteriormente nas pesquisas de Batanero (2000) e Silva (2007), conforme são mostradas na análise teórica deste texto. A apresentação de uma situação na qual os conhecimentos prévios dos alunos são ferramentas para a resolução de alguns aspectos do problema, mas, insuficientes para a solução completa, ajudaram os alunos a encontrar uma medida de variação que considerasse todos os dados.

A atividade não foi suficiente para a institucionalização do conceito de desvio-médio. Em seguida, foram apresentadas mais duas tarefas (n. 21 e n. 22) nas quais o aluno era desafiado a tomar decisão fundamentada na variação, sendo preciso para isso calcular o desvio de todos os dados em relação a um ponto fixo. No final das três tarefas, foi proposto que cada grupo elaborasse uma definição de desvio-médio para propiciar um debate com todos os participantes. A definição que mais apareceu, foi: desvio-médio é a média dos desvios em relação à média.

Com base nas definições criadas pelos participantes e no debate, apresentamos formalmente a definição de desvio-médio e o algoritmo para seu cálculo.

Escolhemos o valor 24 cm, como medida exata do comprimento do pé de Karlo, o que coincidiu com a média. Não fomos felizes com esta escolha didática, pois ela pôde ter induzido algum aluno a pensar que a média deve corresponder ao valor exato, gerando um complicador para a aprendizagem. Para melhorar a tarefa, sugerimos que os dados sejam escolhidos de forma a não coincidir a medida exata com a média.

5.5.2 Análise da tarefa número 38

Esta tarefa faz parte do último bloco²⁷ (situação 3) no qual reagrupamos as últimas 15 tarefas; trata da aplicação do intervalo criado com base no desvio-padrão na tomada de decisão. Ela traz a ideia do evento raro abordado na investigação de Meletiou (2000).

Segundo a autora citada, esta abordagem na Estatística Descritiva pode ser uma ponte e um facilitador para a Estatística Inferencial, além de ajudar os alunos a perceberem a finalidade e utilidade do desvio-padrão e do z-escore.

Silva (2007), citando Hart (1984) e Loosen; Liden; Lacante (1985), defende a análise do desvio-padrão pautada na porcentagem de observações que estão contidas no intervalo criado com base nos desvios em relação à média, pois, segundo estes autores, para tornar o conceito de desvio-padrão significativo é fundamental a apreensão da análise desta medida de variabilidade.

Esta tarefa foi escolhida em razão da importância dos intervalos no auxílio da compreensão do papel do desvio padrão e porque o aluno deverá mobilizar vários conceitos já abordados anteriormente para resolvê-la.

O histograma a seguir representa a distribuição das notas dos alunos de um curso de Matemática na disciplina de estatística. A nota média do grupo é 40, o desvio-padrão 11,4 e a porcentagem de dados contidos no intervalo criado pela soma e subtração de dois desvios-padrão da média é de 96,3%.

- a) Localize abaixo do gráfico a média e o intervalo criado com a subtração e soma de dois desvios-padrão.

²⁷ Explicitado no subcapítulo 5.4.1

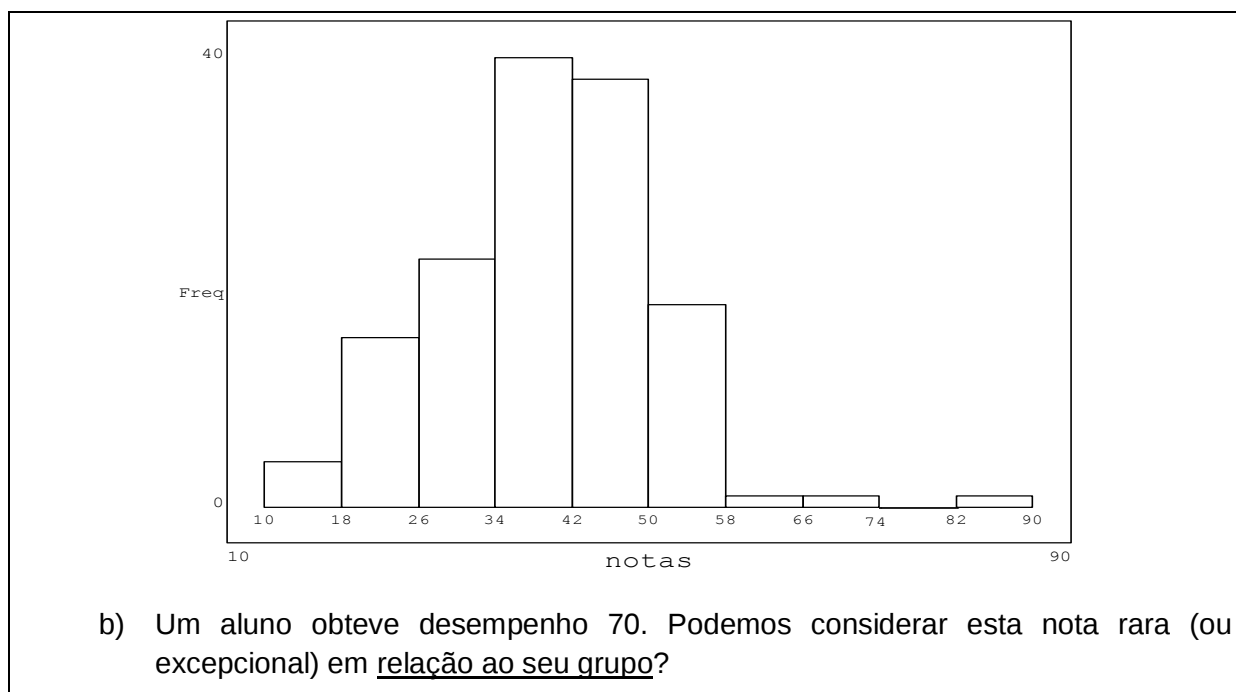


Figura 13 – Enunciado da tarefa 38: gráfico com concentração em torno da média

5.5.2.1 Análise teórica

O objetivo desta tarefa foi levar o aluno a utilizar o desvio-padrão para decidir se um determinado valor da distribuição pode ser considerado raro em relação aos demais. A reflexão principal da tarefa é saber se é suficiente a medida do afastamento em relação à média de um determinado elemento, para classificá-lo como evento raro. Além da média e do desvio-padrão, um conceito importante utilizado como ferramenta nesta atividade é o intervalo criado a partir de k desvios da média e a respectiva regra empírica²⁸.

Os conhecimentos prévios necessários para a realização da atividade, trabalhados nas atividades anteriores, são a leitura e a interpretação de gráficos, compreensão da média, do desvio-padrão, da regra empírica e dos intervalos criados a partir da média e do desvio-padrão.

O conhecimento visado é saber utilizar o intervalo $[\mu - 2DP, \mu + 2DP]$ na justificação de um evento raro.

²⁸ Ver sub-capítulo 4.1: o intervalo $[\mu - DP, \mu + DP]$ abarca, aproximadamente, 68% dos dados e $[\mu - 2DP, \mu + 2DP]$ abarca em torno de 95% dos dados.

No item “a”, trabalhamos com a localização da média e do intervalo no histograma. Para Vieira (2008), esta variável didática permite ampliar a compreensão da média, pois possibilita a visualização desta medida em relação à distribuição. Esta visualização também colabora, para que o aluno perceba que a média sozinha é insuficiente para resumir os dados. Segundo a mesma autora, a estratégia de representar os resultados numéricos (média e intervalos) no gráfico, trata-se de uma mudança de registro de representação semiótica que contribui para uma melhor apreensão do significado dos conceitos.

Conjecturamos que a localização da média e do intervalo no histograma ajudará o aluno a perceber a distribuição dos dados em torno da média e dentro do intervalo, ao mesmo tempo em que poderá notar, o que acontece com a quantidade de elementos fora do intervalo. Desta forma, a visualização pode se constituir em um elemento facilitador para a análise e tomada de decisão.

Uma vez localizada a média e os extremos do intervalo, no item “b” é solicitado que o aluno decida se 70 é ou não uma medida rara em relação a seu grupo. A visualização gráfica será uma aliada na resolução do problema, mas também será preciso lembrar o papel do intervalo criado, baseado no desvio-padrão e no significado dos elementos que estão fora dele.

Lembramos que, nas atividades da seção anterior, foram trabalhados o papel do desvio-padrão na criação de intervalos para análise, a regra empírica por meio de coleta de dados antropométricos, mas não foi abordada a questão do evento raro, introduzida na tarefa que está sendo analisada.

Soluções esperadas

- a) Represente, abaixo do gráfico, a média e o intervalo criado com a subtração e soma de dois desvios-padrão.

Uma solução correta:

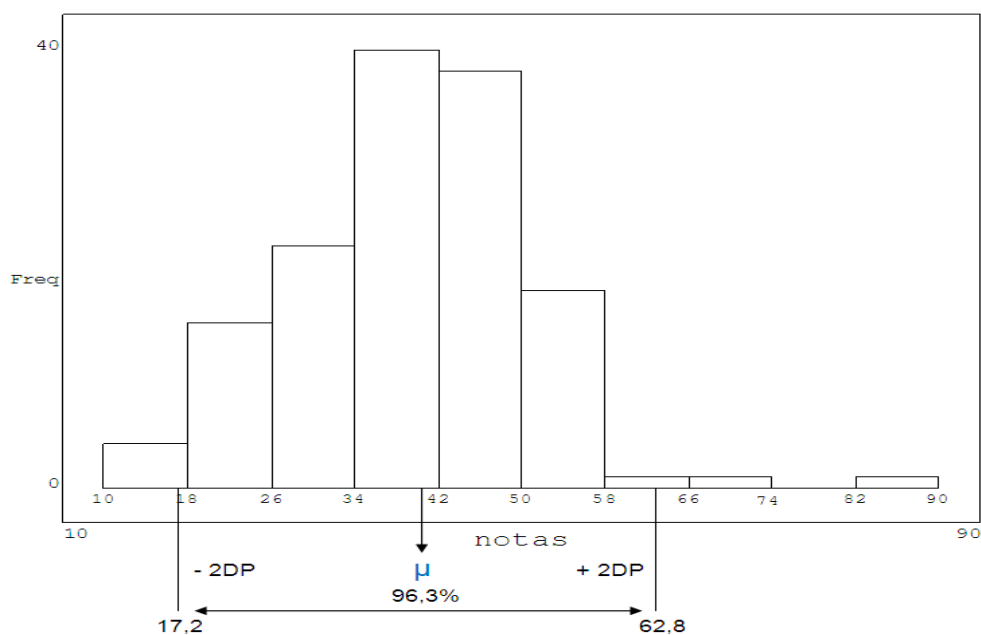


Figura 14 - Resolução do item "a"

- b) Um aluno obteve desempenho 70. Podemos considerar esta nota rara (ou excepcional) em relação ao seu grupo?

Uma solução correta:

R: o intervalo abarcou 96,3% dos dados e a nota 70 ficou fora deste intervalo. Assim, podemos concluir que as chances de ocorrência desta nota são pequenas e o evento pode ser considerado raro para este grupo.

Se o aluno não considerar o intervalo nem o desvio-padrão e decidir apenas pela observação das frequências do histograma, ele estará no nível *técnico* para os conhecimentos relativos à leitura gráfica, já que se limita à leitura dos eixos, sem qualquer outra adaptação necessária para a resolução do problema proposto. Neste caso, ele poderá chegar à resposta correta, argumentando que a frequência do intervalo que contém a nota 70 é baixa. Tal argumentação não é válida, pois não se trata de analisar a frequência, mas o quanto a nota se afasta da média do grupo e saber se este afastamento é ou não significativo.

Outra solução possível é considerar o afastamento em relação à média sem analisá-lo com a variação. Aqueles que, assim, raciocinarem também serão classificados no nível *técnico* para os conhecimentos relativos à variação (nesse caso, vista como a articulação entre média e desvio-padrão). Quem proceder desta forma, poderá avaliar que uma distância de 30 pontos em relação à média é uma distância grande, outros com a mesma informação poderão considerar a distância

pequena, fundamentando-se em uma apreciação pessoal para a decisão, que será assim subjetiva, pois não articula média e desvio-padrão.

O participante que resolver a questão corretamente, utilizando-se do intervalo $[\mu - 2DP, \mu + 2DP]$ será classificado no nível de conhecimento *mobilizável* a respeito da aplicação do desvio-padrão e do intervalo criado, a partir dela na análise de dados, pois a aplicação desse intervalo para justificar um evento raro exige do aluno um certo grau de mobilidade entre os conhecimentos anteriores e o novo, ou seja, a solução correta exige mais que uma aplicação direta do que foi aprendido.

Quanto ao nível de conhecimento *disponível*, não é possível avaliar se o aluno o atingiu ou não ao resolver esta tarefa, pois este nível exige que o aluno seja capaz de chegar à solução correta sem nenhuma indicação, mas logo no enunciado já há referência a um intervalo que contém 96,3% dos dados, indicando um caminho para a solução. Logo em seguida, no item “a”, o intervalo torna-se mais explícito na representação gráfica. Para avaliar o nível disponível, precisaríamos retirar o intervalo do enunciado e apresentar apenas o item “b” ao aluno. Todavia, esta situação mais complexa deixamos para as tarefas posteriores.

5.5.2.2 Análise a posteriori

De modo geral, os alunos mostraram dificuldades para decidir se o evento era raro, tomando por base o intervalo, então, o formador levantou vários questionamentos no intuito de fazê-los refletir sobre a atividade. A seguir, transcrevemos algumas das dificuldades e trechos dos diálogos entre o instrutor e os participantes e entre os próprios alunos.

Como foi previsto na análise teórica, alguns alunos analisaram se o evento era raro, baseando-se apenas na distância entre a média e a nota. Vejamos o diálogo:

Instrutor - Um aluno obteve nota 70 num grupo em que a média foi 40. Posso considerar o desempenho deste aluno excepcional (evento raro) em relação à sua turma?

Aluno 1 - Sim, porque no grupo a média é 40.

Aluno 2 - Ele não está dentro do intervalo.

Aluno 1 - Porque não está dentro do intervalo?

Aluno 3 - É, eu ia falar isso.

Aluno 1 - Ah! Agora entendi.

Instrutor - Têm duas questões aí: uma pessoa da sala falou que o desempenho foi excepcional porque a média é 40 (aluno 1), a outra (aluno 2) falou que é por causa do intervalo. Quem tem razão?

O aluno 1 considerou a nota 70 um valor excepcional em relação ao grupo, mas só estava baseando-se na média. Ele não fez referência à distribuição dos dados no gráfico nem ao intervalo. O aluno 2 também considerou a nota 70 rara, mas utilizou-se do intervalo para fundamentar sua argumentação.

Após ouvir os colegas, o aluno 1 mudou de opinião:

Aluno 1 - Agora já mudei de ideia, já mudei de opinião.

Instrutor - Por quê?

Aluno 1 - Eu falei sem pensar, agora eu pensei... porque, em relação ao intervalo, ... Ele está bem fora do intervalo. O maior número de alunos está lá dentro, quer dizer, a maior porcentagem do número de alunos está dentro deste intervalo que é menor que a nota 62,8. Quer dizer, se o aluno tirou 70, ele está bem acima.

Pela resposta do aluno 1, percebemos que ele compreendeu o papel do intervalo para justificar o evento raro. Até este momento, três alunos tinham se manifestado e estavam convencidos do papel do intervalo, enquanto os outros dez permaneciam em silêncio. Desconfiando que muitos ainda tinham dúvidas, o instrutor optou por insistir na questão da média e da variação. Para isso, propôs uma distribuição com dados concentrados nos extremos com média, aproximadamente, de 40 e a mesma nota para análise.

Instrutor: Suponhamos que um grupo obtenha nota média aproximadamente, 40⁽²⁹⁾, onde 50 alunos receberam nota zero na avaliação, 49 tiraram nota 80 e um aluno obteve nota 70. O aluno que obteve nota 70 teve desempenho excepcional em relação a seu grupo, isto é, muito melhor em relação a seu grupo?

Feito o questionamento o instrutor, representou a situação com o gráfico mostrado na figura 5.

²⁹ O valor exato é 39,9

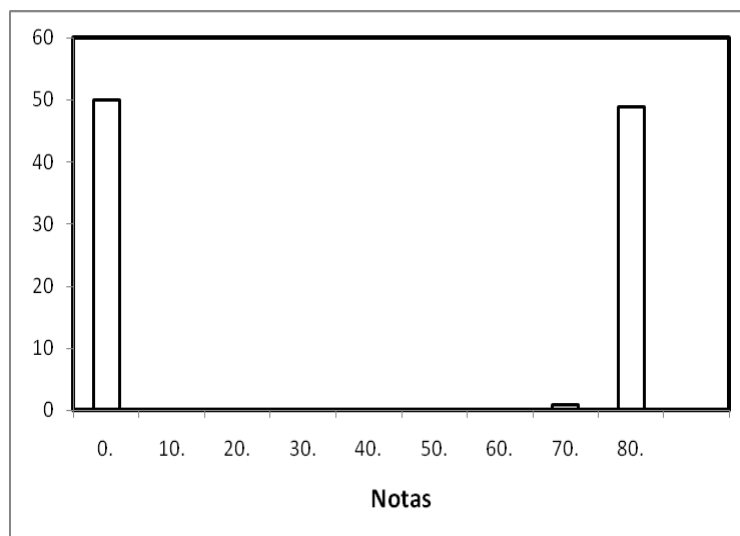


Figura 15 – Gráfico representando concentração nos extremos

Aluno 1 - Foi muito melhor.

Instrutor - Metade dos alunos tirou nota zero, metade nota 80.

Aluno 1 - Não, não!.

Pela resposta do aluno 1, ele voltou a olhar a nota 70 sem considerar a variação, mas, após o questionamento do instrutor, reconsiderou sua resposta.

Para tornar nossos comentários mais enxutos, chamaremos de Gráfico 1 aquele que representa a concentração em torno da média (gráfico do enunciado) e de Gráfico 2 aquele com concentração nos extremos (Figura 15)

Visando a levar os participantes a observarem o papel da variação na tomada de decisão em questão, o instrutor solicitou aos alunos a comparação dos dois gráficos, ambos em relação à nota 70.

Diante do silêncio da maioria dos alunos, o instrutor decidiu voltar ao desvio-padrão e discutir em qual gráfico essa medida era maior. Na comparação entre os gráficos, surge a associação equivocada entre homogeneidade e menor desvio, já identificado na pesquisa de Silva (2007). O Gráfico 2 possui maior concentração nos extremos e maior homogeneidade, embora seus dados tivessem o maior desvio em relação à média e, consequentemente, o maior desvio-padrão.

Instrutor - Aonde os dados variam mais em relação à média: Gráfico 1 ou Gráfico 2?

Aluno 1 - Gráfico 1.

Instrutor - No Gráfico 2 varia mais?

Aluna 1 - Não, no gráfico 1 varia mais em relação à média.

Instrutor - Você acha que é lá?(apontando para o Gráfico 1)

Aluno 1 falando com o aluno 3 - Não é isso que você tinha falado agora?

Aluno 3 - Não.

Aluno 1 - No gráfico 2 varia mais?

Aluno 3 - Varia mais porque a diferença é maior.

Aluno 1 - Ahhhh!

Apesar do instrutor insistir que a variação solicitada era em relação à média e não entre os dados, foi forte a associação entre concentração e desvio pequeno, pois quanto maior a concentração de dados em torno da média, menor será o desvio, mas se a concentração estiver nos extremos a relação não será válida.

Os alunos ficaram um pouco confusos ao tentarem identificar, por meio da observação dos gráficos, onde os dados variam mais em relação à média. A dificuldade para comparar variações em histogramas já havia sido identificada na pesquisa de Meletiou (2000). Na tentativa de facilitar a compreensão, substituímos a palavra variar pela expressão “estar longe” ou distância em relação à média

Instrutor - Vou trocar a variação por outra palavra, para ser mais fácil. O desvio-padrão mede a distância em relação à média. Considere que o Gráfico 1 é uma rua com 100 alunos, dos quais 50 estão na posição zero, 49 na posição 80 e um na posição 70. Todos virão com seus carros até a média (posição 40). Considere o mesmo para o Gráfico 2 onde os dados estão mais concentrados em torno da média. Onde haverá maior gasto de combustível?

Após esta comparação, outros alunos começaram a participar do debate e dar respostas corretas sobre o gráfico que representava maior variação. Neste momento, já havia sete alunos respondendo aos questionamentos do instrutor. Tendo reforçado as ideias sobre o desvio-padrão, o instrutor voltou à questão inicial.

Instrutor - Vamos voltar para ao gráfico do início: uma pessoa tirou nota 70 num grupo cuja média era 40. Posso considerar que seu desempenho sobressai-se em relação a seu grupo?

Aluno 2 - Podemos dizer que ele estava fora da média.

Aluno 8 - O intervalo pegou 96% dos dados?

Instrutor – Sim!

Aluno 8 - Ele está entre os 4%

Instrutor - Está. Então, a nota dele se sobressaiu em relação ao seu grupo?

Aluno 8 - Sobressaiu.

Instrutor - Agora vamos olhar para a distribuição em forma de duas torres (gráfico 2). O aluno que tirou nota 70 se sobressaiu positivamente em relação a seu grupo, teve ótimo desempenho em relação a seus colegas?

Aluno 8 - Não, porque ele não está próximo dos 49% que tiraram as melhores notas alunos. Eu não vejo como raridade.

Instrutor - Entretanto a distância da nota 70 até a média nos dois casos é igual.

Instrutor - O que te levou a tomar esta decisão?

Aluno 8 - Ele está próximo da metade dos alunos que tiraram as melhores notas.

O aluno 8 usando o intervalo para fazer a escolha correta do Gráfico 1, mas estava usando apenas a informação gráfica para decidir a respeito do Gráfico 2. Então, o instrutor decidiu ampliar a análise do aluno.

Instrutor - Isto porque você está olhando o gráfico, não é?

Aluno 8 - Olhando o gráfico.

Instrutor - E se você não tivesse o gráfico para olhar, tivesse só o intervalo? A nota 70 faria parte da minoria excepcional em relação a seu grupo?

Aluno 1 e 4 - Não

Instrutor - Como eu sei que ele não faz parte?

Aluno 3 - Ele está dentro do intervalo.

Instrutor - E daí?

Aluno 3 - Porque se ele está dentro do intervalo... (aluno 10 interrompe)

Aluno 9 - Ele é povão

Instrutor - porque ele é povão?

Aluno 9 - Não é diferente dos outros.

Instrutor - Quem dá esta dica que ele é povão?

Aluno 9 - Intervalo.

Aluno 3 - Era isso que eu ia falar.

Instrutor - Então tem hora que ele faz parte do povão e tem hora que ele faz parte da elite. Como que eu sei a qual grupo ele pertence?

Aluno 9 - Povão quando ele estiver dentro do intervalo e elite quando ele estiver fora do intervalo, fazendo parte daqueles 2%, 3%...

Pelos diálogos, percebemos que a participação dos alunos aumentou significativamente durante as reflexões. Iniciamos o debate com a participação ativa de três alunos e terminamos com nove. Do início das reflexões até este momento os participantes tiveram um salto conceitual no que se refere à compreensão do desvio-padrão, do papel da variação e do intervalo criado com base em k desvios da média na análise de eventos raros.

Após a realização dos itens “a” e “b”, os alunos permaneceram no nível técnico para o conhecimento visado: utilizar os intervalos criados com base no desvio-padrão na justificação de um evento raro. Esta classificação justifica-se por eles terem decidido apenas pela observação dos gráficos ou tomado como

referência a média sem considerar a variação em torno da mesma. Observamos que a atividade solicitava dos alunos apenas este nível

Durante o debate e após a introdução do Gráfico 2, com concentração nos extremos, os participantes perceberam a importância do intervalo para justificar se um evento é raro, além de aprofundarem suas compreensões da noção de desvio-padrão. Consideramos que a utilização do registro gráfico para representar a média e os desvios ajudou a ampliar a compreensão do desvio-padrão que, normalmente, é apresentado apenas na forma numérica. Assim, concordamos com Duval (2003 apud ALMOULOUD, 2007) quando afirma que a coordenação de vários registros (pelo menos dois) é uma condição necessária para a compreensão conceitual.

Os questionamentos feitos pelo instrutor no transcorrer dos diálogos forneceu várias pistas para solução do problema, mas fez com que os alunos concluíssem por si mesmos as razões pelas quais o valor 70 poderia ser considerado raro (ou não). Por isso, avaliamos que, ao final do debate, a maioria dos alunos atingiu o nível mobilizável para o conhecimento visado.

Como já relatamos na análise teórica, não pudemos avaliar o nível de conhecimento disponível para esta atividade.

Uma sugestão para uma próxima aplicação desta tarefa é apresentar primeiramente, o enunciado e o item “b”, pois aqueles que obterem êxito na resolução, utilizando-se para isso da medida de variação e da construção do intervalo fundamentado no desvio-padrão seriam classificados no nível disponível, haja vista que neste item não há nenhuma pista para resolução do problema. Caso os alunos tenham dificuldades, sugerimos que se apresente o Gráfico 1 e o item “a”. Finalmente, recomendamos que se apresente o Gráfico 2 (concentração nos extremos) e se refaça a mesma pergunta feita no item “b”, mas agora para nova distribuição.

Pelo que avaliamos nesta pesquisa, a comparação dos dois gráficos poderá colaborar com a maior compreensão do papel do intervalo criado com base em k desvios da média na análise de dados.

5.5.3 Análise da tarefa número 43

Esta foi a última tarefa da sequência didática e a que apresentou maior grau de complexidade, pois demandava que se mobilizasse a maioria dos conceitos abordados e não apresentava no enunciado nenhuma pista para a solução. Optamos por analisá-la porque nos possibilitaria avaliar se os alunos conseguiriam utilizar o desvio-padrão para medir afastamentos, padronizando as notas analisadas. Aplicar este procedimento, antes do professor apresentar o z-escore, demandaria domínio sobre a noção de desvio-padrão e sobre sua aplicação em situações concretas.

É importante salientar que a ideia de calcular o afastamento de um ponto em relação à média, em número de desvios-padrão, não é totalmente nova para os alunos, pois foi abordada intuitivamente na tarefa 41. Entretanto, esta atividade foi insuficiente para a construção do conceito. Na atividade que está sendo analisada (43), a exigência é maior, pois o participante, além de calcular o afastamento em desvios-padrão, deveria utilizar esta ideia para decidir entre os dois candidatos.

Suponhamos que em um processo seletivo para ingresso em um curso técnico bastante disputado, o critério de seleção seja o desempenho do candidato nas provas de Português e de Matemática. Disputando a última vaga, os candidatos abaixo obtiveram as seguintes notas:

	Nota em Português	Nota em Matemática
Candidato A	60	80
Candidato B	80	60

Notas de todos os candidatos que disputaram as vagas existentes

Notas na prova de Português				Notas na prova de Matemática			
60	0	20	100	50	50	50	100
10	60	20	60	30	20	60	80
60	60	70	80	40	20	30	70

Baseando-se apenas nas notas apresentadas, utilize as ferramentas estatísticas estudadas até agora e decida qual candidato obteve o melhor desempenho.

Figura 16 – Enunciado da tarefa 43

5.5.3.1 Análise teórica:

Os objetivos da tarefa foram levar o aluno a compreender o papel da variação na comparação de valores em relação à média, a utilizar o desvio-padrão para medir o afastamento das notas em relação às respectivas médias, visando decidir qual aluno afastou-se mais delas.

Os conhecimentos prévios necessários para a realização da atividade foram a leitura da tabela, os cálculos da média e do desvio-padrão, a compreensão do conceito de desvio-padrão e a percepção da importância da variação na análise de dados

O conhecimento visado é a construção da noção do z-escore e sua utilização em padronizações.

As escolhas didáticas feitas tiveram a intenção de fazer com que o aluno refletisse sobre os conhecimentos de análise de dados aprendidos anteriormente e, com base na mobilização dos mesmos, construísse uma solução para o problema proposto.

A soma das notas dos dois candidatos era igual para evitar que o aluno analisasse os dados sem considerar a média e a variação do grupo. O candidato A ficou abaixo de três concorrentes em Português e abaixo de um em Matemática, enquanto o candidato B ficou abaixo de três concorrentes em Matemática e abaixo de um em Português, ou seja, por este critério não é possível saber qual aluno obteve o melhor desempenho. As médias eram iguais para evitar que o aluno se limitasse a calcular o desvio bruto³⁰ de cada nota em relação às respectivas médias. As notas dos dois alunos estavam dentro do intervalo $[\mu - 2DP, \mu + 2DP]$, para impelir o aprendiz a utilizar o desvio-padrão como unidade de medida. A escolha de apenas 12 notas em cada disciplina foi para evitar muitos cálculos.

Soluções esperadas

³⁰ Consideramos desvio bruto a diferença de uma determinada medida em relação à média.

Para a construção do conhecimento visado, os participantes deveriam calcular o z-escore de cada candidato e verificar quem se afastou mais da média. Na tabela 5, apresentamos os cálculos.

Tabela.5.- Medidas-resumo dos candidatos que disputaram a prova

	Prova de Português	Prova de Matemática
Média	50	50
Desvio-Padrão	29,2	23,5

De posse do valor da média, do desvio-padrão e das notas de cada candidato (tabela 6), podemos calcular o afastamento de cada nota, em relação à média, utilizando como unidade de medida o desvio-padrão. Para isso, basta tomar o desvio bruto da nota em relação à média e dividi-lo pelo desvio-padrão do grupo.

Tabela 6.- Desempenho dos candidatos A e B em relação à média

	Português				Matemática			
	Nota	Nota média	Desvio da média	Afastamento em DP	Nota	Nota média	Desvio da média	Afastamento em DP
Candidato A	60	50	10	0,34	80	50	30	1,28
Candidato B	80	50	30	1,03	60	50	10	0,43

Na tabela 7 calculamos a média das notas de cada candidatos em unidades de desvio-padrão.

Tabela.7 – Notas padronizadas dos candidatos A e B

	Nota em Português	Nota em Matemática	Soma	Média
Candidato A	0,34	1,28	1,62	0,81
Candidato B	1,03	0,43	1,46	0,73

Os dois candidatos tiveram desempenho acima da média, todavia tomando como referência a variação dos dados, o candidato A foi o que mais se afastou da média. Por isso, classificamo-lo, como o candidato de melhor desempenho.

Se o aluno só considerasse as notas dos candidatos, ele poderia afirmar que o desempenho foi igual. Aqueles que compararem as notas dos candidatos A e B com os desempenhos médios de todos os concorrentes, perceberão que os dois candidatos afastam-se igualmente da média, podendo assim decidir pelo empate.

Esta atitude do aluno demonstra maior habilidade para resolver a tarefa, tendo em vista que leva em conta o desempenho do grupo, mas ainda demonstra limitação no que concerne à utilização da variação para a análise.

Segundo Jankovitz (2007) as notas brutas não podem ser simplesmente somadas, pois uma nota 90 em Português pode ser diferente de uma nota 90 em Matemática, ou seja, um 90 pode “valer” mais que o outro, se forem considerados o desempenho do grupo, a distribuição dos dados e a variabilidade. Tomemos como exemplo os dados da figura 17:

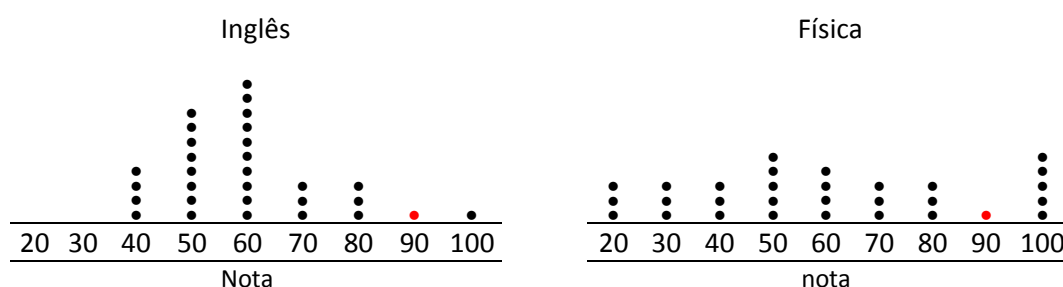


Figura 17 – Comparação da nota 90 em dois grupos com média 60: um com menor e outro com maior variação.

Ao observar os gráficos da figura 17, percebemos que os valores estão mais concentrados em torno da média 60 na prova de Inglês do que na prova de Física. Deste modo, o 90 de Inglês está mais “longe” da média do que o 90 de Física. Esta é uma medida de distância relativa à dispersão dos dados, pois em um grupo onde os dados estão “espalhados”, uma determinada distância é menos significativa do que a mesma distância em um grupo onde os dados estão mais próximos da média.

Para padronizar as notas e poder compará-las, basta tomar os afastamentos delas em relação à média e dividi-los pelos respectivos desvios-padrão. Este procedimento é chamado de z-escore e serve para deixar as notas na mesma escala. Calculando o z-escore para as notas em questão, chegaremos ao valor 2,1 para a nota 90 de Inglês e 1,2 para a de Física, confirmando assim que a nota 90 em Inglês afasta-se mais da média.

Análise segundo os níveis de conhecimento de Robert (1998), para esta tarefa

Consideramos nível *técnico* para esta tarefa, se o aluno calculou as médias, os desvios-padrão e representou os dados nos gráficos; nível *mobilizável* se o aluno articulou as ferramentas do nível anterior com o intervalo criado a partir de k desvios da média para decidir qual candidato tinha maior desempenho, utilizando assim a variação para analisar os dados.

Quanto ao nível de conhecimento *disponível*, consideramos aqueles que foram capazes de mobilizar os conhecimentos prévios levando em conta a variação, calcularam os afastamentos das notas em relação às respectivas médias em unidades de desvio-padrão e, com base neles, decidiram corretamente qual aluno obteve o melhor desempenho.

5.5.3.2 Análise a posteriori

No último encontro, 14 participantes realizaram a atividade. Na tabela 8, relacionamos as ferramentas estatísticas utilizadas.

Tabela 8 - Ferramentas estatísticas utilizadas na resolução da tarefa

Ferramentas estatística	Quantidade de alunos
Média	14
Mediana	08
Moda	08
Gráfico de Pontos	13
Gráfico de Colunas	01
Desvio-padrão	14
Intervalo criado com base em k desvios da média	10
Representação do intervalo numa reta	07
Z-escore	07

Os dados da Tabela 8 apontam para uma evolução da capacidade de análise dos participantes da formação. O primeiro aspecto positivo notado foi que nenhum aluno somou as notas brutas dos candidatos para comparar os desempenhos. Todos recorreram às médias dos grupos em estudo para fazer as comparações, além de calcularem a variação por meio do desvio-padrão. O uso do gráfico, também, foi uma atitude positiva, pois indicou a intenção de observar toda a distribuição e não apenas pontos isolados. A aplicação do intervalo criado com base na subtração e adição dos desvios-padrão à média na tentativa de resolver o problema, mostrou que os alunos compreenderam a importância desse instrumento ao resolverem as tarefas anteriores. Consideramos que o fato da metade dos participantes representar o intervalo e outras medidas em retas para fazer a análise, mostra uma evolução na habilidade de mudar de registro, o que favorece a apreensão de conceitos, segundo a teoria dos Registros de Representação Semiótica.

Além disso, de acordo com Wild; Pfannkuch (1999), o uso de diferentes representações dos dados possibilita melhor compreensão do sistema, sendo uma das características fundamentais do pensamento estatístico.

O uso do z-escore só surgiu após os questionamentos do instrutor com alguns grupos. As principais sugestões fornecidas referiam-se à tarefa 41, na qual o aluno precisava comparar duas notas em dois grupos com médias e desvios-padrão diferentes e a tarefa 38, em que o participante deveria decidir se um evento era raro comparando duas distribuições, uma com dados concentrados em torno da média e outra com dados concentrados nos extremos.

As discussões ocorridas durante estas duas atividades anteriores tiveram de ser retomadas diversas vezes no intuito da mobilização de alguns conceitos, para medir afastamentos em relação à média em unidades de desvio-padrão.

Dentre os alunos que calcularam o z-escore, apenas um somou as notas padronizadas, obtendo os valores de 1,62 para o candidato A e 1,46 para o candidato B. Apesar dos cálculos corretos, ele escolheu o candidato B como o de melhor desempenho, argumentando

Candidato B, pois o desempenho em Matemática e Português em relação à média foi melhor.

Ao observar a atividade deste aluno, levantamos a hipótese que se enganou ao optar pelo candidato B e compreendeu a ideia de medir afastamentos em desvio-padrão. Mas, precisaríamos tê-lo entrevistado para confirmar nossa conjectura.

O “aluno 2” não calculou o z-escore e cometeu alguns erros ao tentar resolver a tarefa (como mostra a figura 18), entretanto trechos de seu comentário apontam para uma evolução quanto à compreensão do uso da variação para decidir sobre afastamentos em relação à média. A seguir, a solução apresentada:

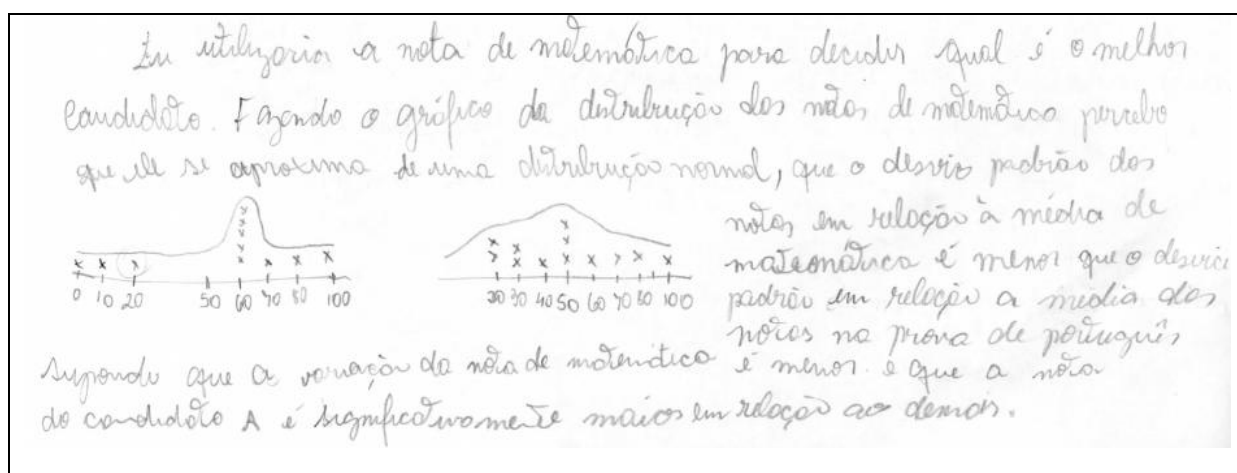


Figura 18 – Solução apresentada pelo aluno 2

Transcrevemos os trechos que consideramos indicarem uma evolução na compreensão do “aluno 2”

[...] Fazendo o gráfico da distribuição das notas de Matemática percebo que [...] o desvio-padrão das notas de Matemática em relação à média de Matemática é menor que o desvio-padrão em relação à média das notas de Português. Supondo que a variação da nota de Matemática é menor e que a nota do candidato A é significativamente maior em relação aos demais.

Vale ressaltar o destaque que o aluno fez da média, mostrando sua compreensão do desvio-padrão, como medida de variação em relação à média e não entre os dados. O texto indica que ele percebeu que desvios brutos iguais com variações diferentes têm significados distintos.

Quanto aos níveis de conhecimento de Robert (1988), classificamos cinco alunos no nível *técnico*, pois realizaram cálculos de medidas de tendência central e de dispersão, representaram os dados em gráficos, mas não demonstraram habilidades para relacionar a variação na análise da atividade.

Os nove alunos restantes não chegaram à conclusão correta, entretanto procuraram adaptar os instrumentos estatísticos aprendidos ao buscarem a solução do problema, preocupando-se em analisar todos os dados e sua variabilidade. Por esse motivo, classificamos estes alunos no nível *mobilizável* para aplicação da variação em análise de dados.

Nenhum aluno atingiu o nível de conhecimento *disponível* para utilização do z-escore na decisão do candidato com melhor desempenho. Alguns calcularam o z-escore, mas não conseguiram perceber que a soma das notas padronizadas³¹ (z-escore) poderia ajudá-los a diferenciar os candidatos.

Examinamos as atividades dos alunos e a gravação das discussões de um dos grupos para classificá-los nos níveis de conhecimento de Robert (1998). Todavia, consideramos que seria necessário entrevistar alguns participantes para entendermos melhor as estratégias utilizadas na resolução das tarefas, tendo mais subsídios para classificá-los.

Uma falha na proposição da tarefa foi a não solicitação da justificativa da resposta, pois com ela poderíamos ter mais elementos para entender como o aluno estava raciocinando e, conseqüentemente, analisar com mais elementos os níveis de conhecimentos atingidos por ele.

Depois que os alunos devolveram a atividade, discutimos com o grupo o significado de medir afastamentos da média em unidades de desvio-padrão e definimos z-escore. Apesar de não termos aplicado mais nenhum instrumento de avaliação, nossa percepção foi que os participantes compreenderam o z-escore e o papel da variação na análise exploratória de dados.

A resolução desta tarefa possibilitou a articulação de diversas noções anteriores, instigou os participantes a examinar dados sob diversos aspectos e a mudar de registros, gerando muita interação e debate de ideias. Toda esta dinâmica provocada pela atividade é importante para a construção de conhecimentos novos.

³¹ Neste contexto, o z-escore está sendo utilizado para padronizar notas.

5.5.4 Comparação entre a avaliação do primeiro dia de formação com a do último dia

No primeiro dia de formação, aplicamos um questionário para coletar dados a respeito dos participantes e três questões para avaliar a compreensão que tinham da média, mediana e variabilidade. No último dia do curso, aplicamos as mesmas questões para verificarmos a evolução dos alunos.

A primeira questão pretendia verificar a concepção dos alunos sobre média aritmética. A pergunta era: o que é média?

Tabela 9 - Concepção dos alunos sobre média no início e final da formação

Categoria da resposta	Avaliação antes da formação 20 alunos		Avaliação no final da formação 13 alunos	
	Nº de alunos	Percentual de alunos	Nº de alunos	Percentual ³² de alunos
Descreveu o algoritmo	17	95	8	61
Medida que representa os dados ou que os sintetiza	2	10	1	8
Define média como medida de equilíbrio entre os dados	0	0	1	8
Descreve o algoritmo e afirma que é uma medida de equilíbrio	0	0	1	8
Descreve o algoritmo e define como medida-resumo	0	0	1	8
Descreve o algoritmo e define como medida-resumo de tendência central	0	0	1	8
Termo que se encontra no meio	0	0	0	0

Segundo a literatura pesquisada, uma concepção equivocada sobre média é sua associação com o termo do meio (mediana). Pretendíamos verificar se tal associação ocorreria, mas nenhum aluno apresentou este erro. Mas, isto não significa que os participantes não pensem desta forma, pois eles se limitaram a descrever o algoritmo da média e, assim, não pudemos averiguar se tinham ou não

³² Valor arredondado para o inteiro mais próximo

esta concepção. Avaliamos que a forma como a questão foi proposta, dificultou a investigação.

Comparando as respostas dos alunos, notamos evolução na compreensão da média, tendo em vista que aproximadamente 38% dos alunos, além de descreverem o algoritmo associaram-no com medida-resumo, de tendência central ou de equilíbrio do dados, etc.

A segunda questão pretendia verificar a concepção dos alunos sobre mediana. A pergunta era: o que é mediana?

Tabela 10.- Concepção dos alunos sobre mediana no início e final da formação

Categoria da resposta	Avaliação antes da formação 20 alunos		Avaliação no final da formação 13 alunos	
	Nº de alunos	Percentual de alunos	Nº de alunos	Percentual de alunos
Não respondeu	2	10	0	0
Define mediana como valor do meio sem se referir a ordem	13	65	8	61
Define mediana como valor do meio em um conjunto ordenado	3	15	2	16
Média dos extremos	1	5	0	0
Mediana, como valor robusto	0	0	1	8
Mediana divide o conjunto deixando 50% dos dados de cada lado	0	0	1	8
Medida próxima à média	1	5	0	0
Divisão da média pelo desvio-padrão	0	0	1	8

A principal diferença entre os alunos no início e final do curso é que alguns, na segunda avaliação, acrescentaram que a mediana é uma medida-resumo, robusta ou que deixa 50% dos dados de cada lado da distribuição. Podemos observar que não houve muita diferença entre os dois momentos, isto pode ser justificado pelo fato de não termos dado muita ênfase para esta medida na formação.

Na literatura pesquisada, encontramos relatos de erros no uso da mediana, referentes a não ordenação dos dados. Apenas com a questão que propusemos,

não foi possível verificar se nossos alunos cometeriam este tipo de erro, mas o fato da maioria deles não se referir à ordem na definição, pode indicar uma tendência para ocorrência de tal equívoco.

Na última questão (figura 17), queríamos verificar se os alunos consideravam a variabilidade em suas análises.

Um aluno foi avaliado três vezes em cada uma das disciplinas descritas na tabela abaixo				
Disciplina		Notas		
Matemática	7,0	-	7,0	- 7,0
Português	10,0	-	1,0	- 10,0
Biologia	5,0	-	7,0	- 9,0
Compare quantitativamente o desempenho dos alunos nas três disciplinas.				

Figura 19 – Enunciado da última questão da avaliação final

Tabela 11 - Comparação das respostas dos alunos, antes e depois da formação no que se refere ao uso da variabilidade

Categoria da resposta	Avaliação antes da formação 20 alunos		Avaliação no final da formação 13 alunos	
	Nº de alunos	Percentual de alunos	Nº de alunos	Percentual de alunos
Não respondeu	2	10	0	0
Calculou a média	19	90	12	92
Utilizou a média para argumentar que o desempenho foi igual	8	40	1	8
Calculou a mediana	2	10	5	38
Calculou a moda	0	0	3	23
Além de calcular a média, fez comparação qualitativa.	1	5	3	23
Considerou a variabilidade	2	10	9	69

Quanto ao uso da média, já esperávamos que grande parte dos alunos se utilizasse dela, pois é uma medida comum em vários seguimentos sociais. Quanto à restrição de seu uso na argumentação de que o desempenho dos alunos foi igual, houve evolução na capacidade de análise dos alunos, pois antes da formação 40% deles restringiram-se à média na análise; no final do curso, este índice caiu para 8%.

A respeito do uso da variabilidade, percebemos progressos na capacidade de análise, pois, no início tínhamos 10% dos participantes, considerando esta característica dos dados e, no final, este índice subiu para 69%.

De modo geral, houve melhora na capacidade de análise dos alunos. Conjecturamos que a variável didática “nota dos alunos” conduziu alguns participantes a fazerem a análise qualitativa e não a quantitativa, pois vários já eram professores, e isto deve ter interferido, para que observassem se houve evolução nas notas, concluindo que se as notas finais fossem melhores, os alunos teriam melhor desempenho. O objetivo era verificar se eles utilizariam as noções de variabilidade na análise dos dados, mas parece que a intimidade dos alunos com o assunto motivou-os a fazer análises qualitativas. O fato confirmou a constatação de Meletiou (2000) de que a familiaridade com o contexto do problema pode ter aspectos positivos e negativos. O conhecimento do contexto pode ajudar na análise dos resultados, mas também pode desviar os estudantes dos dados, fazendo-os expressar opiniões demasiadamente fortes.

CONSIDERAÇÕES FINAIS

A consciência da importância da variabilidade para o Pensamento Estatístico fez com que nos motivássemos a optar pelo tema, por percebermos um certo abandono desse tópico nas pesquisas, currículos e escolas.

Conforme relatam Reading; Shaughnessy (2004), a presença da variação nos fenômenos foi o motivo para que as pessoas desenvolvessem estatísticas sofisticadas, ocupando posição central na análise de dados. Para eles, qualquer discussão séria sobre o Pensamento Estatístico deve examinar o papel da variação.

Autores como Moore (1997), Snee (1999), Meletiou (2000), Shaughnessy; Reading (2004) insistem no papel capital da variabilidade para o desenvolvimento da competência de análise de dados. Nesta linha, Silva (2007) afirma que muitos autores consideram que a variabilidade é o coração da estatística.

Por outro lado, percebemos que o assunto ainda é pouco abordado e necessita de pesquisas sobre seu processo de ensino e aprendizagem. Segundo Meletiou (2000), a variação é, muitas vezes, negligenciada e seu papel no raciocínio estatístico é sub-reconhecido. Para Reading; Shaughnessy (2004), há uma ênfase excessiva na medida de tendência central e falta de atenção à medição da variabilidade. Diante disso, procuramos estudar as características das atividades de uma formação de alunos de licenciatura em Matemática, que favoreçam a apreensão da variabilidade quando se analisa um conjunto de dados.

Neste contexto, levantamos a seguinte questão de pesquisa: quais as características didáticas de uma formação continuada em Estatística que utilize o enfoque da Análise Exploratória de Dados, para favorecer a percepção da variabilidade por alunos do curso de licenciatura em Matemática?

Elaboramos uma Sequência Didática, segundo a proposta de elaboração de cenários de aprendizagem de Robert (1998) que se baseia na Dialética Ferramenta-Objeto e Mudança de Quadros, na Teoria das Situações Didáticas e nos Registros de Representação Semiótica. A sequência foi dividida em três grandes situações didáticas, a saber: introdução e construção do conceito de média; construção do conceito de variação pelo uso de medidas de dispersão (amplitude,

desvio-médio e desvio-padrão) e aplicação do conceito de dispersão para tomada de decisões.

As situações didáticas estavam articuladas, para propiciar que os conhecimentos já incorporados fossem mobilizados para construção de conceitos novos. Nessa dialética, os conceitos eram descontextualizados, para que pudessem tornar-se ferramentas para a construção de novos conhecimentos.

As tarefas eram apresentadas aos participantes com o intuito de propiciar-lhes interação independente com o problema, sem que soubessem qual o conceito que o professor queria lhes ensinar. A noção a ensinar era a resposta considerada mais conveniente para solucionar o problema.

Após a aplicação da Sequência Didática, os participantes não se limitaram mais à aplicação direta do algoritmo da média, mas compreenderam que ela é limitada e que precisa estar acompanhada de outras medidas estatísticas. Uma das medidas que eles passaram a usar com mais frequência foi a mediana. Além dela, começaram a descrever os dados em gráficos de pontos e a observar como variam em torno da média.

A estratégia de criar um problema no qual a solução fosse o conhecimento visado, foi importante para introduzir o conceito de média e amplitude, para mostrar as limitações da amplitude e levar o estudante a procurar uma medida que considerasse todos os dados em relação a um determinado ponto fixo. A ideia de média dos erros foi importante para introduzir o conceito de desvio-médio.

Acreditamos que a compreensão do desvio-médio tenha facilitado o entendimento do algoritmo do desvio-padrão e ajudado a concebê-lo como a média dos desvios em relação à média. Embora essa concepção seja mais adequada para o desvio-médio, ela foi útil para auxiliar os alunos a apreenderem a noção de desvio-padrão.

Na resolução da tarefa 20, pudemos confirmar as conclusões do trabalho de Silva (2007) nas quais os participantes associavam pequena variação com homogeneidade e regularidade, observando a variação entre os dados e não em relação a uma medida de tendência central. Conforme a autora, esta atitude dos alunos é incentivada pelos livros didáticos que fazem tal relação. Percebemos, na tarefa 38, que a associação da palavra concentração com variação baixa, pode ser

um obstáculo didático à compreensão da variação, pois existem situações onde as concentrações dos dados estão localizadas nos extremos, gerando alta dispersão.

Assim, conjecturamos que muitos dos erros não se devem por falta de compreensão conceitual dos alunos, mas são incentivados pelo próprio enunciado que não deixa claro se a variação solicitada é entre os dados ou em relação a um determinado ponto. O próprio enunciado de nossa tarefa gerou esta dúvida nos participantes. Desta forma, percebemos que uma das características importantes a se destacar no desvio-padrão, desvio-médio e intervalo interquartil é que eles medem a variação tomando como referência uma medida de tendência central.

Na tarefa 38, o uso do desvio-padrão apoiado na porcentagem de observações que estão a k desvios da média, foi importante para tornar o conceito de desvio-padrão significativo para os alunos, indo ao encontro das ideias defendidas por Hart (1984) e Loosen, Lichen e Lacante (1985), citadas na pesquisa de Silva (2007). A articulação dos intervalos criados com base no desvio-padrão, na regra empírica e no evento raro proposto por Meletiou (2000) ajudou os alunos a perceberem a importância da variação na análise de dados e levou-os a dar significado ao desvio-padrão como instrumento de auxílio na tomada de decisão.

Na determinação do evento raro da atividade 38, inicialmente, houve dificuldade por parte dos alunos na determinação do solicitado, baseado nas análises do intervalo e do desvio padrão. Para ajudá-los, o instrutor resolveu propor a comparação de duas situações extremas, a do próprio enunciado com concentração em torno da média e, outra, cujos dados estavam concentrados nos extremos. Neste momento, surgiu outro complicador: os alunos tinham dificuldades para perceber o desvio-padrão com base no histograma, confirmando a constatação de Meletiou (2000). Foi necessário retomar esse conceito, para que os alunos percebessem que, na tarefa proposta, o evento poderia ser considerado raro, mas não naquela em que a variação era grande.

O uso dos gráficos associados ao desvio-padrão indicou que os alunos ainda não tinham uma compreensão completa dessa medida de variação. Desta forma, notamos ser necessário não se limitar ao cálculo algébrico das medidas e suas análises, mas buscar outras formas de representação para ampliar a compreensão do objeto de estudo, confirmando o que havíamos encontrado na literatura

(READING; SHAUGHNESSY 2004; VIEIRA, 2008). Nossa observação aponta para as vantagens na percepção da variabilidade quando nos utilizamos do cálculo algébrico aliado à representação gráfica da medida de variação e de tendência central.

Na última questão, os alunos não foram capazes de construir o z-escore apenas com a tarefa proposta, mas precisaram da interferência do instrutor fazendo-os relacionar, o que se pedia na tarefa com conceitos abordados anteriormente. Avaliamos que fomos muito arrojados ao propor esta questão, neste momento da sequência didática.

Apesar dos estudantes não terem conseguido construir o z-escore sozinhos e aplicá-lo para decidir qual candidato deveria ser classificado, a atividade propiciou momentos construtivos de reflexão entre os participantes, levando-os a mobilizar diversos conceitos abordados anteriormente, utilizando-se de diferentes representações. Toda esta dinâmica provocada pela atividade, indicou-nos que houve evolução conceitual quanto ao uso de instrumentos de análise estatística e quanto à percepção e emprego da variabilidade na análise exploratória de dados.

Nossa questão de pesquisa procurou as características das atividades de uma formação de alunos de licenciatura em Matemática que favoreçam a apreensão da variabilidade quando se analisa um conjunto de dados.

Robert (1998) propôs cenários de aprendizagem para o ensino e aprendizagem da Matemática e nós utilizamos os mesmos preceitos para a aprendizagem da Estatística. Como ela baseou-se na Dialética Ferramenta Objeto e na teoria dos Registros de Representação Semiótica, fomos buscar elementos destas teorias para montar uma sequência que favorecesse a aprendizagem dos conceitos de estatística relacionados à variabilidade.

Para o grupo pesquisado, pudemos observar que a elaboração de situações fundamentais, nas quais o saber visado é a resposta mais adequada para o problema, colaborou com a construção de conceitos. Nesta mesma linha, notamos que foi importante propiciar atividades que possibilitassem a mobilização de conhecimentos anteriores, visando à construção do saber visado, que, após descontextualizados, poderiam ser ferramentas para construção de outros conceitos. Concordamos com Reading; Shaughnessy (2004); Almouloud (2007) e

Vieira (2008), quando exaltam a importância da mudança de registros na compreensão ou descoberta de conceitos.

No que se refere aos conteúdos específicos de estatística, as atividades 14, 15 e 16 adaptadas de Ben-Zvi e Garfield (2008) possibilitaram a compreensão do conceito de média com vistas à variabilidade, pois abordavam a média como equilíbrio entre os dados. Notamos que a integração da abordagem da média com a variação favorece a apreensão da variabilidade e de sua medida.

Outra característica que se mostrou favorável à percepção da importância da variabilidade, foi a comparação entre dois conjuntos de dados com médias iguais e variação bem distintas. Os participantes puderam perceber as limitações da média e a necessidade de se considerar a variação na análise de dados.

Quanto aos níveis de funcionamento do conhecimento de Robert (1998), para a percepção da variabilidade em uma análise exploratória de dados, entendemos que a maioria dos alunos atingiu o nível disponível, pois, ao final da formação mostraram-se capazes de perceber a variação, de articular conhecimentos sobre o tema e mudar de representações para melhor analisar os dados.

No que se refere à utilização do intervalo criado, com base em k desvios da média para determinar um evento raro e a construção do conceito do z-escore, classificamos os participantes no nível mobilizável, pois tiveram de adaptar seus conhecimentos para aplicá-los à nova situação, mas fornecemos indicações.

Começamos abordar os quartis e sua representação associada aos gráficos de pontos nas tarefas 10 e 11, mas não demos prosseguimento ao estudo por falta de tempo. Assim, notamos em nossa pesquisa uma lacuna no que se refere ao estudo dos quartis e box-plot.

Quando os alunos foram calcular o desvio de cada dado em relação a um ponto determinado por meio do Gráfico de Pontos, alguns calcularam os desvios sem considerar as frequências, ou seja, não levaram em conta os vários pontos em cada “coluna”. Levantamos uma questão: será que se usássemos gráficos de colunas, a incidência de erros deste tipo seria maior, tendo em vista que a generalização é maior? Se a resposta fosse sim, teríamos outra questão: o uso do gráfico de colunas desde as séries iniciais é um obstáculo didático para que os alunos considerem todos os pontos no cálculo da variação?

Para futuras pesquisas, gostaríamos de examinar o potencial didático dos gráficos de pontos, frente ao de colunas, para facilitar a percepção da variabilidade e da distribuição dos dados e investigar as vantagens de se iniciar a análise descritiva pelo estudo da distribuição dos valores.

REFERÊNCIAS

ALMOULOUD, S. A. **Fundamentos da didática da matemática**. Curitiba: UFPR, 2007.

ALMOULOUD, S. A.; COUTINHO, C. Q. S. Engenharia Didática: características e seus usos em trabalhos apresentados no GT-19/ANPEd. **Revista Eletrônica de Educação Matemática**, v. 3.6, p.62-77, 2008. Disponível em: <http://www.redemat.mtm.ufsc.br/revemat/2008_pdf/revista_2008_06_completo.pdf> Acesso em 10 dez. 2009.

ARTIGUE, M. Engenharia Didática. In: BRUN, J. **Didáctica das Matemáticas**. Lisboa: Instituto Piaget, 1996, p. 193-218.

_____. **Didáctica de la Estadística**. Granada: Grupo de investigación en Educación Estadística da Universidad de Granada, 2001. Disponível em <www.ugr.es/~batanero>. Acesso em: 12 ago. 2008.

BATANERO, C. Significado y comprensión de las medidas de posición central. **UNO**, n. 25, p41-58, 2000. Disponível em <www.ugr.es/~batanero>. Acesso em: 12 dez. 2009.

BEN-ZVI, D.; GARFIELD, J. B. How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics. **International Statistical Review**, v. 75, n. 3, p. 372-396, dez. 2007. Disponível em: <<http://www3.interscience.wiley.com/journal/118698329/issue>>. Acesso em: 20 jun. 2009.

_____. **Developing students' statistical reasoning: connecting research and teaching practice**. New York: Springer, 2008.

BRASIL. Ministério da Educação. Secretaria de Educação Fundamental. **Parâmetros Curriculares Nacionais 3: Matemática (1ª a 4ª séries)**. Brasília: MEC/SEF, 1997.

_____. Secretaria de Educação Fundamental. **Parâmetros Curriculares Nacionais 3: Matemática (5ª a 8ª séries)**. Brasília: MEC/SEF, 1998.

_____. Secretaria de Educação Média e Tecnológica. **Parâmetros Curriculares Nacionais (Ensino Médio) – Parte III: Ciências da Natureza, Matemática e suas Tecnologias**. Brasília: MEC/SEB, 2000.

_____. Secretaria de Educação Média e Tecnológica. **PCN+ Ensino Médio: orientações educacionais complementares aos Parâmetros Curriculares Nacionais**. Brasília: MEC/SEB, 2002.

_____. Secretaria de Educação Média e Tecnológica. **Orientações curriculares para o Ensino Médio: Ciências da Natureza, Matemática suas Tecnologias**. Brasília: MEC/SEB, 2006.

_____. Ministério da Saúde, Secretaria de Vigilância em Saúde. **Informe Epidemiológico Influenza Pandêmica (H1N1)**. Brasília, n. 10, nov. 2009. Disponível em: <http://portal.saude.gov.br/portal/arquivos/pdf/boletim_influenza_se_44.pdf>. Acesso em 14/11/2009.

BROUSSEAU, G. Introdução ao estudo das situações didáticas. São Paulo: Ática, 2008.

BUSSAB, W.; MORETTIN, P. A. **Estatística Básica**. São Paulo: Saraiva, 2005.

CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artmed, 2003.

CANOSSA, R. **O professor de matemática e o trabalho com medidas separatrizes**. 2009. Dissertação (Mestrado Profissional em Ensino de Matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo.

CARVALHO, C. **Interações entre pares: contributos para a promoção do desenvolvimento lógico e do desempenho estatístico no 7º ano de escolaridade**. 2001. 628f. Tese (Doutorado em Educação) - Universidade de Lisboa, Lisboa.

CORDANI, L. K. “Estatística para todos”: oficina para professores. In: pre-International Conference on Teaching Statistics, 7, 2006, Salvador. Disponível em: <http://redeabe.org.br/OFICINA%20site_educacao.pdf> Acesso em 10 dez. 2007.

DAMM, F. D. Registros de representação. In: MACHADO, S. D. A. (Org.). **Educação Matemática: uma (nova) introdução**. São Paulo: Educ, 2008. p. 167-188.

DELMAS, R. C. Statistical Literacy, reasoning and learning: a commentary. **Journal of Statistics Education**, v.10, n.3, nov. 2002. Disponível em: <http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html>. Acesso em 13 dez. 2009.

DELMAS, R.; LIU, Y. (2005). Exploring students' conceptions of the standard deviation. **Statistics Education Research Journal**, v.4, n.1, p. 55-82, mai. 2005. Disponível em: <[http://www.stat.auckland.ac.nz/~iase/serj/SERJ4\(1\)_delMas_Liu.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(1)_delMas_Liu.pdf)>. Acesso em 31 nov. 2009.

FERREIRA, A. B. **Novo dicionário da língua portuguesa - século XXI**. Rio de Janeiro: Nova Fronteira, 2003.

GAL, I. Adult's statistical literacy: meanings, components, responsibilities. **International Statistical Review**, v. 70, n. 1, p. 1-25, 2002.

IMENES, L; LELLIS, M. **Matemática para todos: 5ª série**. São Paulo: Scipione, 2005.

INTERNATIONAL STATISTICAL LITERACY PROJECT - Fase 2 – Brasil: não publicado, fase 2, mar. de 2009.

JANKOVITZ, H. Access explica a nota padronizada. In: UNIFESO. **Manual do candidate – vestibular 2007**. Terezópolis: Unifeso, 2007, p. 11. Disponível em: <http://www.access-selecao.com.br/download/manuais/feso072_manual.pdf> Acesso em 30 jul. 2009

KONOLD, C.; POLLATSEK, A. Data Analysis as the Search for Signals in Noisy. **Journal for Research in Mathematics Education**. V. 33, n.4, p. 259-289, jul 2002.

LIMA, A. C. P.; MAGALHÃES, M. N. **Noções de probabilidade e estatística**. São Paulo: Edusp, 2004.

MACHADO, S. D. A. Engenharia didática. In: MACHADO, S. D. A. (org.). **Educação Matemática: uma (nova) introdução**. 3. ed. São Paulo: Educ, 2008. p. 233-247.

MELETIOU, M. M. **Developing student's conceptions of variation: an untapped well in statistical reasoning**. 2000. Tese (Doutorado em Filosofia) – Universidade do Texas, Austin, 2000.

MOORE, D. New pedagogy and new content: the case of statistics. **International Statistical Review**, v. 65, n.2, p.123-137, 1997.

PAIS, L. C. **Didática da Matemática, uma análise da influencia francesa**. 2. ed. Belo Horizonte: Autentica, 2002.

PAIVA, M. **Matemática**. São Paulo: Moderna, v. 1, 2004.

READING, C.; SHAUGHNESSY, M.. Reasoning about Variation. In: BEN-ZVI, D e GARFIELD, J. **The challenge of developing Statistical Literacy, Reasoning and Thinking**. Netherlands: Kluwer, 2004. p. 201-226.

ROBERT, A. Outils d'analyse des contenus Mathématiques à enseigner au lycée et à l'université. **Recherches en Didactique des Mathématiques**. vol. 18. N..2, pp.139-190, 1998.

SILVA, C. B. da. **Pensamento estatístico e raciocínio sobre variação: um estudo com professores de matemática**. 2007.354f. Tese (Doutorado em Educação Matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo.

SNEE, R. D. Statistical Thinking and its Contribution to Quality. **The American Statistician**, v.44, n. 2, p. 116-121, 1990.

SNEE, R. D. Development and use of statistical thinking: a new era. In: Statistical thinking in empirical enquiry. **International Statistical Review**, v. 67, n. 3, p. 223-265, 1999.

TRIOLA, M. F. **Introdução à estatística**. 7^a ed. Rio de Janeiro: LTC, 1999.

VIEIRA, M. **Análise exploratória de dados: uma abordagem com alunos do ensino médio**. 2008. Dissertação (Mestrado em Educação Matemática) – Pontifícia Universidade Católica de São Paulo, São Paulo.

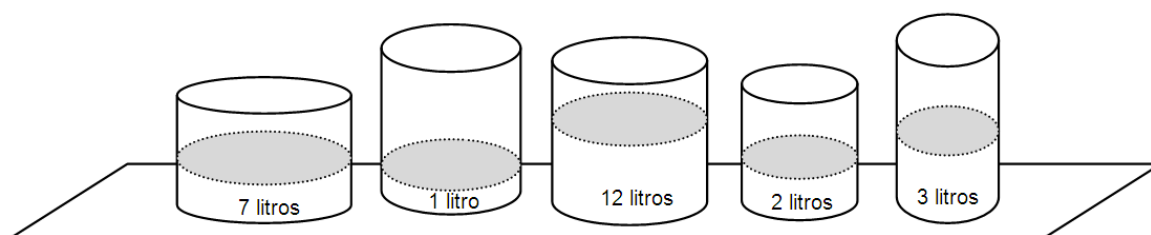
WILD, C; PFANNKUCH, M. Statistical thinking in empirical enquiry. **International Statistical Review**, v. 67, n. 3, p. 223-265, 1999.

APÊNDICE A: TAREFAS PROPOSTAS NA FORMAÇÃO

Parte A: Introdução à média

As tarefas 1 até a 8 foram entregues juntas.

- 1) Uma pessoa possui 5 baldes contendo quantidades de água, de acordo com o indicado nas figuras abaixo:



Se a pessoa deseja distribuir a mesma quantidade de água em cada balde, quantos litros ela deverá colocar?

(Atividade adaptada de Paiva, 2004)

Objetivo: levar o aluno a usar a média, como distribuição equitativa, para resolver o problema.

- 2) Um pai fez um testamento distribuindo os bens que possuía para seus três filhos da seguinte forma:

- Tiago, o mais novo, ficaria com R\$ 5.000,00
- Cristian receberia R\$ 10.000,00 e
- Pedro, o mais velho, receberia R\$ 15.000,00.

Os filhos começaram a brigar porque não achavam justo receberem quantias diferentes. Para evitar desentendimentos, o pai resolveu distribuir o dinheiro igualmente. Quanto cada filho recebeu na nova distribuição?

Objetivo: levar o aluno a usar a média, como distribuição equitativa, para resolver o problema.

- 3) Um professor deve entregar na secretaria da escola a síntese do desempenho dos alunos em cada bimestre. João obteve as seguintes notas em um bimestre: 8,0 na primeira prova e 4,0 na segunda. Qual nota de João, o professor deverá entregar na secretaria, sabendo-se que são atribuídos pesos iguais às duas notas? Justifique sua resposta.

Objetivo: utilizar a média, como medida síntese para resolver o problema

- 4) Outro professor deu quatro provas e as notas de João foram: 4,0; 5,0; 6,0 e 1,0. Que nota o professor deverá entregar na secretaria? Por quê?

Objetivo: utilizar a média, como medida síntese para resolver o problema.

- 5) Baseando-se nas notas, descubra quem teve o melhor desempenho geral:

Nome	Nota	Nota	Nota
Mirian	3,0	7,0	8,0
Carmen	2,0	9,0	7,0

Como você chegou a esta conclusão?

Objetivo: utilizar a média, como medida síntese para resolver o problema

- 6) Uma professora resolveu dar um prêmio para seu melhor aluno. Na 6ª série A, ela deu cinco provas, e na 6ª série B deu quatro provas. Veja as notas na tabela:

Nome	Nota1	Nota2	Nota3	Nota4	Nota5
Celso da 6ªA	8	7	10	9	6
Marta 6ªB	10	6	9	9	* * *

- a) Baseando-se no desempenho geral dos dois alunos, ajude a professora a decidir.
b) Qual critério você utilizou para decidir quem ganharia o prêmio? Por quê?

Objetivo: Objetivo: utilizar a média, como medida síntese para decidir.

- 7) Marcos e Celso são jogadores de basquete. Marcos jogou cinco partidas, nas quais fez 25, 32, 30, 27 e 40 pontos. Celso jogou 4 partidas nas quais fez 32, 29, 40 e 42 pontos.

- a) Qual dos dois jogadores fez maior número de pontos?
b) Qual dos dois apresentou melhor desempenho por partida?

(Adaptado de Imenes; Lellis, 2005)

Objetivo: utilizar a média, como medida síntese para decidir.

- 8) Uma dona de casa vai trocar o piso de seu apartamento. Para isso, chamou seis pedreiros para fazerem o orçamento do serviço. Como o preço depende da área a ser revestida, todos os pedreiros foram ao local para efetuar as medições, mas houve variação de medida entre os profissionais. As áreas calculadas foram: 150 m^2 , 152 m^2 , 149 m^2 , 153 m^2 , 148 m^2 e 154 m^2
- a) Faça uma estimativa para a área real a ser revestida.
- b) Por que você escolheu este valor?

(Atividade adaptada de Batanero, 2001, p.90)

Objetivo: Utilizar a média como estimativa para resolver o problema

A tarefa 9 só foi entregue após os alunos terem recebido e respondido as anteriores

- 9) Quando você calculou quantos litros de água seriam colocados em cada balde (atividade 1), caso fosse distribuída a mesma quantidade de água em cada um deles, você estava calculando a **média**. Assim, diríamos que a quantidade **média** de água nos baldes é: litros.

Para descobrir quantos reais cada filho receberia (atividade 2) se o dinheiro fosse dividido em partes iguais, você calculou a média. Assim, quando o pai deu R\$ 5.000,00 para Tiago, R\$ 10.000,00 para Cristian e R\$ 15.000,00 para Pedro, dizemos que elei deu em **média**: R\$...... para cada filho.

Objetivo: Introduzir o termo média, como resposta dos problemas anteriores, associando seu algoritmo com operações já conhecidas pelos alunos: adição e divisão.

Discussão

- a) Quais as medidas estatísticas utilizadas para resolver as atividades anteriores?
- b) Qual foi a importância da média e/ou de outras medidas que você utilizou?
-

Parte B: Medidas-resumo de tendência central, quartis e variabilidade

- 10) Observe as idades de um grupo de universitários participantes de um curso de estatística:

19	20	32	27	36	31	20	25	35	28
20	21	32	27	40	43	30	42	8	31

Você deve sintetizar os dados e traçar um perfil do grupo quanto à idade.

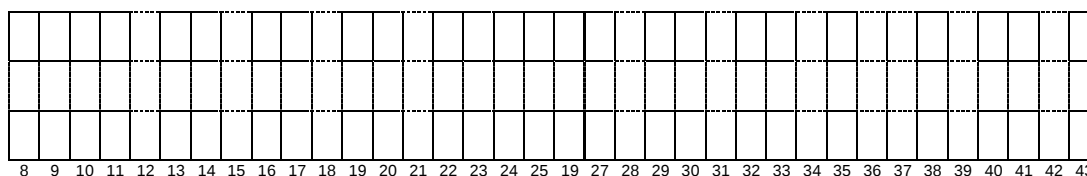
Objetivo: sintetizar dados coletados entre a própria turma (exceção do 8), utilizando-se das ferramentas que tiver disponível. Espera-se que eles se utilizem da média e mediana, observem onde há maior concentração de dados, o valor máximo e o mínimo e resumam tais informações em um texto.

O valor oito foi incluído para que os alunos refletissem a respeito da influência ou não de valores extremos na média e mediana.

Só após os alunos terem sintetizado os dados da tarefa 10, foram entregues as questões a seguir:

Sobre as idades do grupo de universitários:

- b) Construa o esquema gráfico das idades, colocando um ponto • sobre cada valor correspondente. Caso algum valor se repita, coloque na linha superior, formando uma “torre”.



- Qual a idade média?
- Qual a idade mediana?
- Obtenha o 1º quartil.
- Obtenha o 3º quartil.
- Quantos dados estão entre os 1º e 3º quartis?
- Qual o percentual de idades presentes entre os 1º e 3º quartis?
- Abaixo do diagrama localize, com um segmento de reta vertical os 1º e 3º quartis e a mediana (2º quartil). Desenhe um retângulo (caixa) a partir do 1º segmento até o 3º.
- O retângulo ficou dividido em duas partes. Qual o percentual de dados em cada parte?
- Em qual parte os dados estão mais concentrados?

- j) Qual a relação entre o tamanho de cada parte do retângulo e a concentração de dados?

Objetivo: perceber que o intervalo interquartil contém em torno de 50% dos dados, que a quantidade de elementos compreendida entre os 1º e 2º quartis e entre o 2º e 3º são iguais e concluir que quanto menor as distâncias entre os quartis, menor a variação. A representação por meio do Gráfico de Pontos, também, objetiva a percepção da variabilidade dos dados.

A tarefa 11 foi entregue em dois momentos; a primeira parte contém o enunciado e o item “a”, a segunda inclui do item “b” até o “h”.

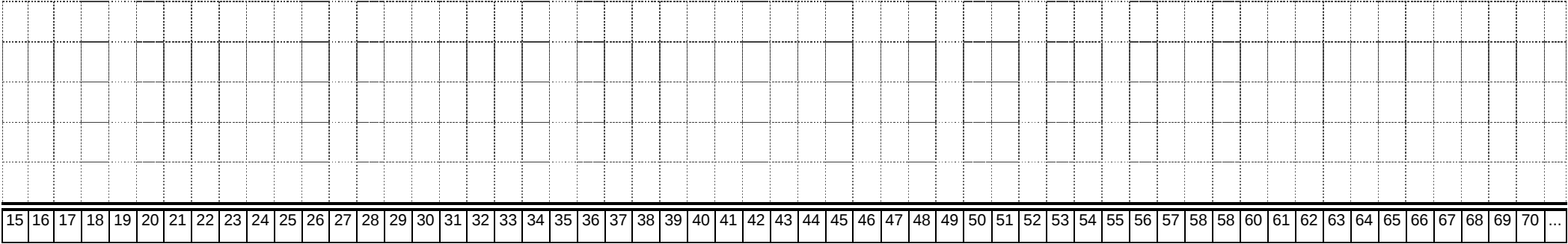
- 11) O Ministério da Educação deseja conhecer o perfil de cada sala de Educação de Jovens e Adultos (EJA). Para atendê-lo, a escola deverá enviar dados que representem algumas características de cada grupo de alunos. Fazendo um levantamento das idades da turma do T-1 (EJA), ela obteve os seguintes dados:

Idade em anos do T-1 EJA													
15	16	16	16	23	25	27	28	28	30	31	44	47	53
34	33	34	35	37	39	39	40	40	41	42	62	47	50

- a) Qual é a idade “típica”³³ a ser enviada ao Ministério da Educação?

³³ Um valor típico é um número único que resume determinada característica do grupo

b) Construa o esquema gráfico das idades, colocando um ponto • sobre cada valor correspondente. Caso algum valor repita-se, coloque na linha superior, formando uma “torre”.



Idade (anos)

- c) Calcule a idade média.
- d) Obtenha a idade mediana.
- e) Observe o gráfico e diga em torno de quais idades há maior concentração de pontos (dados)?
- f) No gráfico faça uma marca sobre a idade média.
- g) Os dados sempre se concentram em torno da média?
- h) Qual a menor e a maior idade? Qual a variação das idades?
- i) Olhando para o gráfico, você modificaria a idade “típica” ?
- j) Abaixo do diagrama localize, com um segmento de reta vertical os 1º e 3º quartis e a mediana (2º quartil). Desenhe um retângulo (caixa) a partir do 1º segmento até o 3º.

Objetivo: utilizar a média, como medida “típica” de um grupo e perceber a distribuição dos dados em torno dela.

A tarefa 12 foi entregue em dois momentos; a primeira parte contém o enunciado e o item “a”, a segunda inclui do item “b” até o “m”.

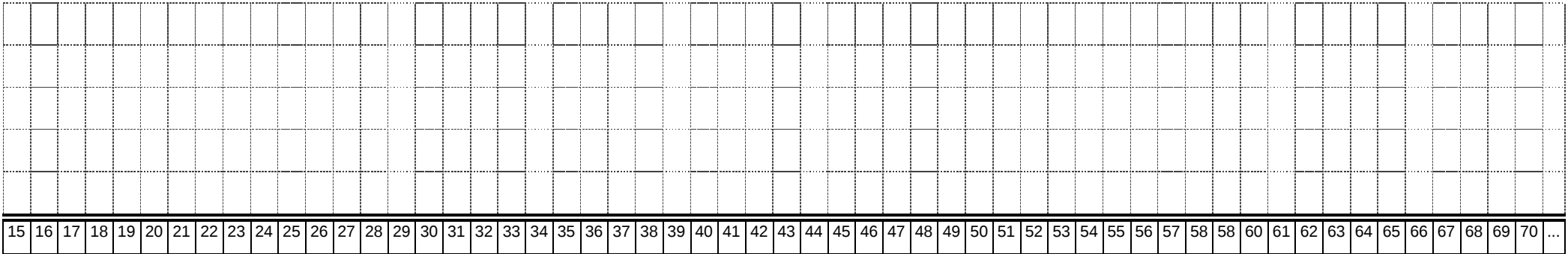
12) A tabela abaixo contém as idades dos alunos da 3ª série B (Ensino Médio)

Idade em anos da 3ª série B do Ensino Médio												
16	16	16	16	16	16	16	16	17	17	17	18	22
17	17	17	17	17	17	18	18	18	19	21	19	17

- a) Qual a idade “típica” dos alunos do 3º B?

Objetivo: utilizar a média como medida “típica”, perceber a variação dos dados em torno dela e relacionar a representatividade desta medida de tendência central com a dispersão dos dados.

b) Construa o esquema gráfico das idades, colocando um ponto • sobre cada valor correspondente. Caso algum valor repita-se, coloque na linha superior, formando uma “torre”.



Idade (anos)

- c) Calcule a idade média.
- d) Observe o gráfico e diga em torno de qual idade há maior concentração de pontos (dados)?
- e) No gráfico faça uma marca sobre a idade média.
- f) Os dados sempre se concentram em torno da média?
- g) Quais a menor e a maior idade? Qual a variação das idades?
- h) Olhando para o gráfico, você modificaria a idade “típica”?
- i) Abaixo do diagrama localize, com um segmento de reta vertical os quartis. Desenhe um retângulo (caixa), a partir do 1º primeiro quartil até o 3º.
- j) Observe os gráficos e diga em quais salas (3ºB ou T1), você percebe maior variação das idades?
- k) Em quais das salas, foi mais fácil obter a idade “típica”?
- l) Em quais das turmas, a média é mais representativa?
- m) Para analisar a idade representativa do grupo (idade “típica”), é importante observar a variação dos dados no gráfico, ou seja, verificar se eles estão mais próximos ou afastados? Comente.

Objetivo: utilizar a média como medida “típica” de um grupo, perceber a importância de alguma medida de variação.

Discussão

Parte C: Reflexões sobre a média e sobre a variação dos dados

13) Coisas para pensar!

- a) Suponha que a idade média de uma turma de alunos é de 21 anos. Então o que nós sabemos sobre a distribuição das idades dos alunos desta classe (por exemplo, todos eles têm cerca de 21 anos?)? Explique.
- b) Como se chega a esse número (21)? Explique.

(Retirado de Ben-Zvi e Garfield. (2008))

14) O Jogo das Médias: foram dadas dez caixas de fósforos e cada uma representa uma pessoa. Você deverá distribuí-las sobre a reta que está numerada de 10 a 32 (anos), empilhando-as ou não para formar um Gráfico de Pontos de acordo com as instruções fornecidas. A posição de cada caixa na reta indicará a idade da pessoa representada, ou seja, uma caixa colocada na posição 12 indicará que existe uma pessoa

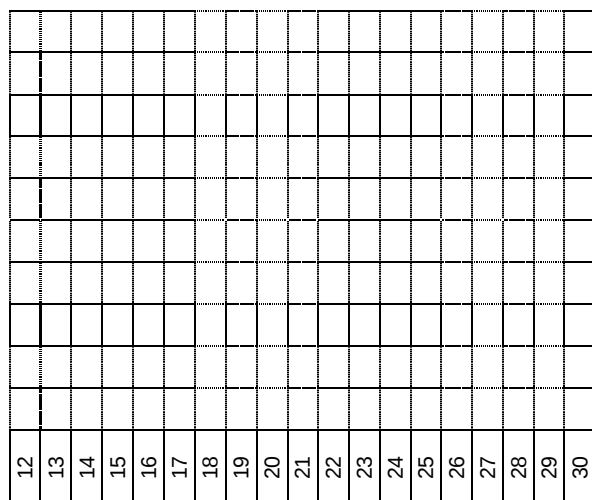
com 12 anos. Cada gráfico deverá ser criado de forma que a média seja sempre 21 anos.

- a) A maneira mais fácil de fazer isso, é empilhar todas as caixas sobre o valor 21. Se fizer isso, qual será o formato do gráfico? Como os dados estarão distribuídos: espalhados ou concentrados?
- b) Retire uma caixa de fósforo da pilha da posição 21 anos e passe para a posição 24 (anos) da reta.
 - b.1 A média continua sendo 21 anos?
 - b.2 Mova mais uma caixa da pilha, de forma que a média das idades seja 21 anos.
 - b.3 Como você pode mover outras caixas da pilha maior, de forma que a média se mantenha em 21 anos?
 - b.4 Há mais de uma maneira de fazer isto? Explique.
- c) Em seguida, coloque uma caixa de fósforo na posição 17 anos. Descreva o que deve ser feito para manter a média em 21 anos.
- d) Agora, mova toda as caixas da posição 21 anos, mas faça com que a média do grupo continue sendo 21. Descreva o que você fez para conseguir isto.

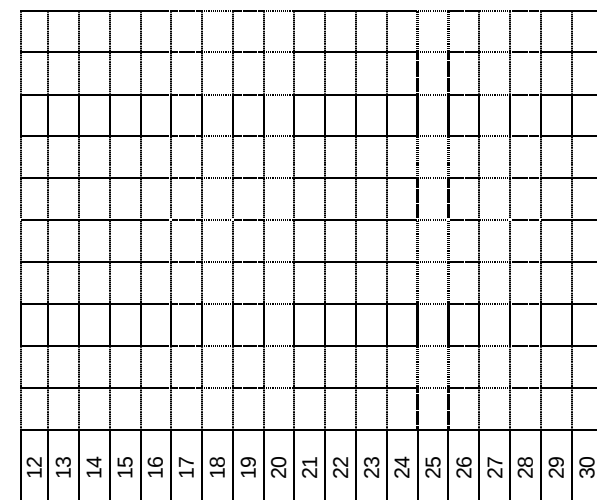
(atividade adaptada de Ben-Zvi e Garfield. (2008))

15) Desenhar dois gráficos de pontos diferentes de acordo com as orientações abaixo:

- Cada gráfico deve ter dez valores (10 pontos).
- A média dos valores deve ser 21 anos.
- Em um Gráfico de Pontos (Dotplot), as idades deverão estar concentradas (pouca dispersão/propagação) e no outro mais espalhadas em relação à média (maior dispersão/propagação).



Idade



Idade

Objetivo: representar a tarefa 14 em Gráfico de Pontos para analisar a distribuição sob o aspecto gráfico.

As duas questões seguintes foram entregues após a confecção dos gráficos

Ideias que você deve considerar:

- a) Que estratégias você usou para manter a média de 21 em cada gráfico?
- b) Foi importante dar atenção à distância de cada ponto até o 21? Como é que isso ajuda a manter a média de 21 anos?

(atividade adaptada de Ben-Zvi e Garfield (2008))

Objetivos das tarefas 13, 14 e 15: ampliar a compreensão conceitual da média; interpretar a média como ponto de equilíbrio entre os dados; perceber que o equilíbrio se dá quando a soma dos desvios de cada dado em relação à média é zero, isto é, os valores absolutos dos desvios anteriores à média são iguais aos posteriores; estimular a observação de diversas distribuições com a mesma média.

- 16) Nos itens “a” até “d”, considere o termo “desvio” como a diferença entre cada valor e a média. Isto, também, pode ser visto como o número de unidades para a esquerda ou para a direita da média, em relação ao valor de cada dado.

- a) Observe o Gráfico de Pontos abaixo e anote os desvios de cada ponto em relação à média.



Idade observada x_i	Média $\bar{X}$	Desvio ($x_i - \bar{X}$)
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
Soma		

- b) Se você excluir uma idade que tenha um desvio de -3, o que você precisa fazer com um ou mais valores (idades), para manter a média em 21 anos?
- c) E se você retirar uma pessoa de 24 anos e colocar uma pessoa de 25 anos, o que você precisa fazer com um ou mais valores (idades) para manter a média em 21 anos?
- d) Qual o valor da soma dos desvios? Por que deu este resultado?

Objetivos: calcular os desvios e perceber que a soma dos desvios sempre será igual a zero, preparar o aluno para o cálculo do desvio-médio.

Discussão:

Qual a relação entre média e equilíbrio?

PARTE D: Reflexão sobre o uso da média e a amplitude como medida de variação

17) Oficina das Placas

(Atividade adaptada de Cordani, 2006)

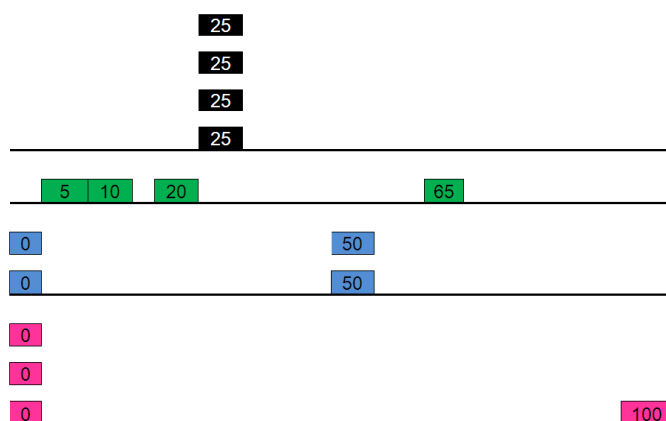
Descrição da atividade: foram distribuídas, para 16 alunos da turma, cartelas coloridas (verde, azul, rosa e preta), representando os valores das horas-aula de quatro escolas, de acordo com o indicado a seguir:

Escola	Valores nas cartelas (hora-aula)			
Verde	R\$ 20,00	R\$ 65,00	R\$ 5,00	R\$ 10,00
Azul	R\$ 50,00	R\$ 0,00	R\$ 0,00	R\$ 50,00
Rosa	R\$ 100,00	R\$ 0,00	R\$ 0,00	R\$ 0,00
Preta	R\$ 25,00	R\$ 25,00	R\$ 25,00	R\$ 25,00

Os participantes são orientados a encontrar um eixo de referência na classe (0 a 100) para se posicionarem: pode ser a lousa, a mesa, a parede etc .

- 1- Chamamos uma cor: cada aluno coloca-se no referencial. A turma deveria copiar os valores na ordem do menor para o maior, na linha correspondente a cada cor. Solicitamos que encontrassem o valor médio.
- 2- Chamamos a próxima cor e pedimos novamente que se colocassem no referencial, mantendo a cor anterior no referencial. São pedidos a média e o valores máximo e mínimo.
- 3- Fizemos o mesmo com as duas cores restantes, uma de cada vez.

Disposição das 16 placas, após a participação dos alunos.



As tabelas abaixo foram entregues para que os participantes pudessem preenchê-las

Escola	Valores nas cartelas (hora-aula)	Média
Verde		
Azul		
Rosa		
Preta		

Questões:

a) Suponhamos que foi divulgado apenas o valor médio das horas-aula de cada escola. Baseando-se somente nesta informação, qual escola você escolheria para trabalhar?

b) Obtenha em cada escola os valor máximo, mínimo e a amplitude

Escola	Média	Máximo	Mínimo	Amplitude
Verde				
Azul				
Rosa				
Preta				

c) Qual a diferença entre escolher uma das escolas tomando-se como referência apenas a média e escolhê-la com base na média e na amplitude?

Objetivos: perceber a insuficiência da média como medida-resumo e utilizar a amplitude como medida de variação.

- 18) Observe as temperaturas entre duas cidades medidas em um dia: San Pedro do Atacama, cidade localizada numa região desértica e úmida no norte do Chile a 2 400 metros de altitude com população de 3.000 habitantes e Caraguatatuba, localizada no Litoral Norte de São Paulo/Brasil com clima tropical úmido e 93.000 habitantes

Algumas prováveis temperaturas em um dia:

Temperatura marcada durante um dia em San Pedro do Atacama		Temperatura marcada durante um dia em Caraguatatuba	
Horário	Temperatura ³⁴	Horário	Temperatura ³⁵
6h	10°C	6h	20°C
9h	30°C	9h	20°C
12h	39°C	12h	23°C
16h	35°C	16h	21°C
20h	6°	20h	18°C
24h	0°	24h	18°C

- Calcule a temperatura média de cada uma das cidades.
- Calcule a temperatura mediana de cada uma das cidades.
- Baseando-se apenas na média, qual cidade você escolheria para viver?
- As figuras abaixo representam hastes em equilíbrio sobre pontos de apoio na posição 20°C. Cada ponto - ● - representa um objeto com massa fixa. Sobre a haste 1, marque com ● as temperaturas de San Pedro do Atacama e na haste 2 as temperaturas de Caraguatatuba/SP. (Quando as temperaturas se repetirem, coloque um ponto sobre o outro, montando uma pilha).
- Acima de cada haste localize, com um segmento de reta vertical, os quartis. Desenhe um retângulo (caixa), a partir do 1º primeiro quartil até o 3º.

³⁴ Dados fictícios

³⁵ Dados fictícios

Haste 1

0°C
1°C
2°C
3°C
4°C
5°C
6°C
7°C
8°C
9°C
10°C
11°C
12°C
13°C
14°C
15°C
16°C
17°C
18°C
19°C
20°C
21°C
22°C
23°C
24°C
25°C
26°C
27°C
28°C
29°C
30°C
31°C
32°C
33°C
34°C
35°C
36°C
37°C
38°C
39°C
40°C



Haste 2

0°C
1°C
2°C
3°C
4°C
5°C
6°C
7°C
8°C
9°C
10°C
11°C
12°C
13°C
14°C
15°C
16°C
17°C
18°C
19°C
20°C
21°C
22°C
23°C
24°C
25°C
26°C
27°C
28°C
29°C
30°C
31°C
32°C
33°C
34°C
35°C
36°C
37°C
38°C
39°C
40°C



- e) Após a colocação dos pontos, qual haste deveria permanecer em equilíbrio?
- f) Por que foi escolhida a posição 20, para colocar o ponto de apoio?
- g) Encontre a temperatura máxima e mínima de cada cidade.
- h) Calcule a amplitude das temperaturas de cada cidade.
- i) Compare as amplitudes com as variações de temperatura das duas cidades. Qual relação você percebe entre amplitude e variação?
- j) Em um mesmo sistema de eixos cartesianos, represente com gráficos de linha, as variações das temperaturas das duas cidades. Trace uma reta horizontal na linha que contém a média.

Objetivos: perceber as limitações da média, como medida-resumo, reconhecer a amplitude como medida de dispersão, interpretar a média como ponto de equilíbrio e utilizar o gráfico, para a visualização da variabilidade.

[illegible]

e) Represente as medidas da tabela acima em um Gráfico de Pontos.

No quadro abaixo, represente a medida de cada aluno com um círculo - ● -, fazendo uma “torre” quando os valores repetirem-se.

Observação: pinte com cores diferentes os círculos que representam os homens e as mulheres; faça um X sobre o círculo que representa seu palmo.

Cada aluno deverá ir ao quadro e colocar o ponto que representa a medida do seu palmo

16	17	18	19	20	21	22	23	24	25	26	27
----	----	----	----	----	----	----	----	----	----	----	----

f) Qual sugestão você daria para resumir estes dados por meio de medidas estatísticas?

g) Abaixo do diagrama de pontos, localize os 1º e 3º quartil e a mediana (2º quartil) com segmentos de retas verticais e desenhe um retângulo a partir do 1º segmento até o 3º.

h) Observando os itens anteriores, faça um pequeno relatório das medidas dos palmos da turma.

i) Como está sua medida em relação a dos outros?

j) Discussão sobre os relatórios.

Objetivo: observar a distribuição dos dados em torno da média e visualizar graficamente os quartis.

20) Medida do palmo de um único aluno.

Um professor solicitou a seus alunos do T1-EJA e T2-EJA que medissem o comprimento do palmo da mão do aluno Karlo. Com as medidas obtidas, foram construídos os gráficos abaixo.

Gráfico A:

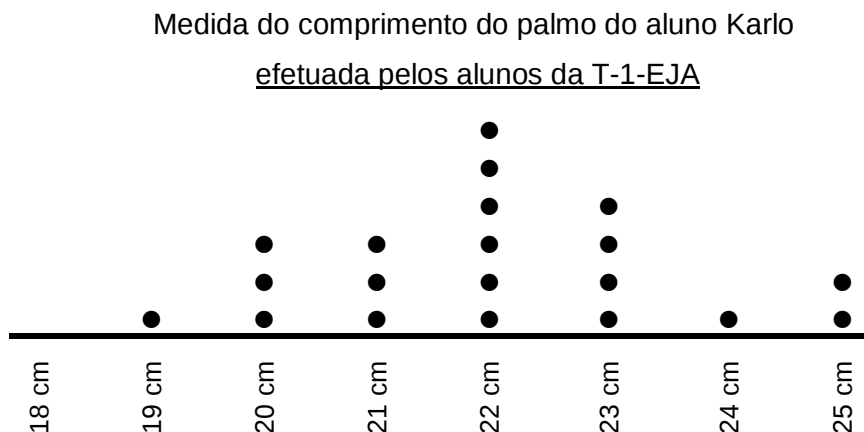
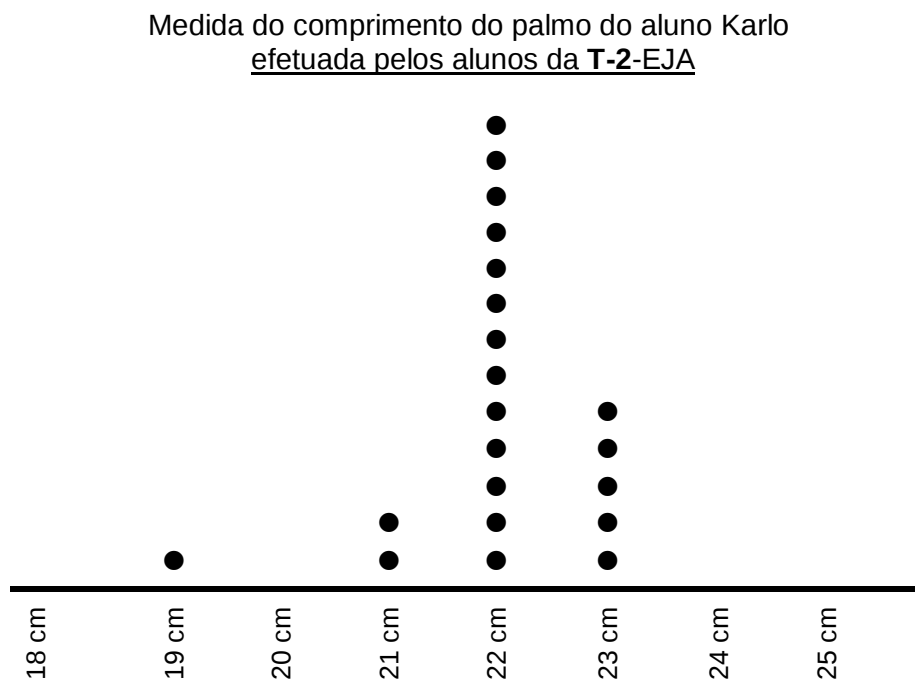


Gráfico B:



- a) Suponhamos que haverá uma disputa entre as turmas T-1 e T-2 e vencerá aquela que tiver realizada a melhor medição do palmo de Karlo. Você apostaria na turma T-1 ou na T-2? Por quê?
- b) Escolha um dos gráficos e estime um valor para a medida do comprimento real do palmo do Karlo. Justifique sua escolha e sua estimativa.

- c) Em qual conjunto (T-1 ou T-2) de dados, houve maior variação dos dados?
- d) Encontre o máximo, o mínimo e a amplitude de cada conjunto? Você percebe alguma relação entre amplitude e variação? Qual?

	Máximo	Mínimo	Amplitude
T-1			
T-2			

- e) Você foi encarregado(a) de contratar uma equipe para fazer medições. Qual das duas equipes você contrataria? Existe alguma relação entre a confiança que você depositará em um dos grupos e a variação das medidas encontradas?

Objetivo: perceber o papel da variabilidade na tomada de decisão, utilizar a amplitude, como medida desta variabilidade.

PARTE E: A insuficiência da amplitude como medida de variação, a busca de um instrumento mais abrangente.

Objetivos: o aluno deverá perceber a insuficiência da amplitude como medida de variação e encontrar uma forma de medir a variação que leve em consideração todos os dados.

21) Voltemos às medidas antropométricas do aluno Karlo. O professor convidou 12 alunos da 6ª A e 12 da 6ª D e solicitou que medissem o comprimento do pé de Karlo. Os resultados foram expressos nos gráficos abaixo.

Gráfico 1:

Medidas do comprimento do pé do aluno Karlo
efetuada pelos alunos da 6ª A

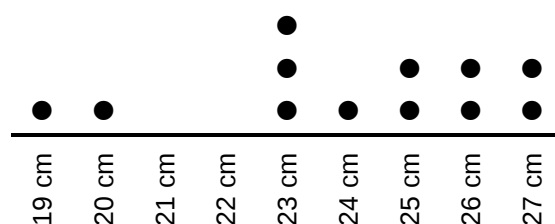
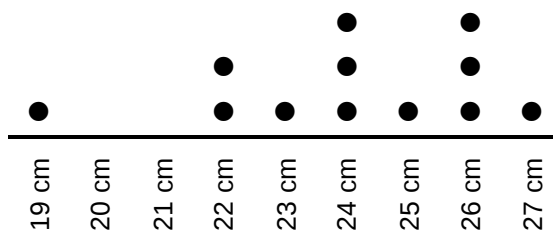


Gráfico 2:

Medidas do comprimento do pé do aluno Karlo
efetuada pelos alunos da 6ª D



Questão :

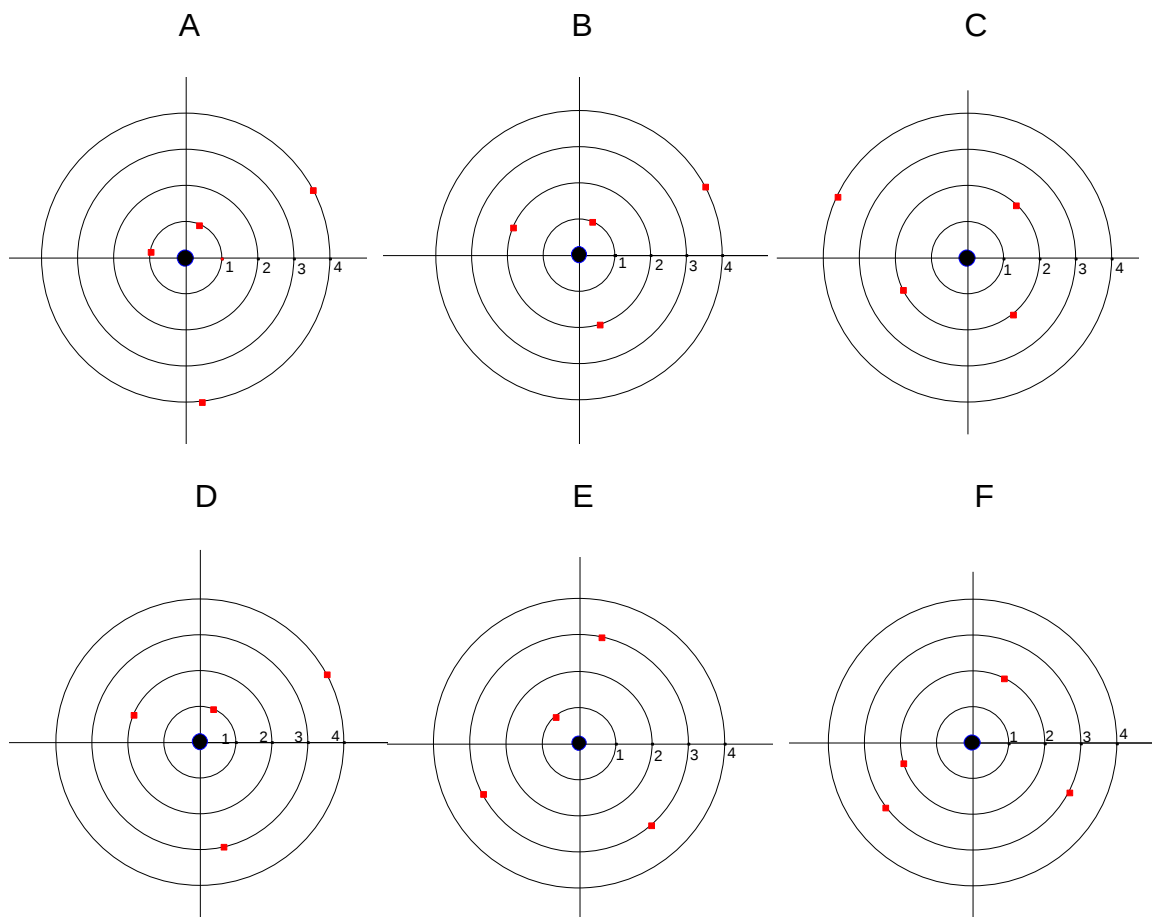
- Em qual conjunto (6ªA ou 6D) de dados, houve maior variação ?
- Calcule a amplitude de cada conjunto? As amplitudes são iguais? As variações são iguais?
- Suponha que a medida real do pé de Karlo seja 24 cm, qual grupo se afastou mais da medida real (ou variou mais em relação à medida 24 cm)? Justifique.
- Em média, quantos centímetros erraram³⁶ os alunos da 6ª A? E os da 6ª D?

Objetivo: introduzir o desvio-médio por intermédio do cálculo do erro médio.

³⁶ Entenda-se erro como o afastamento da medida efetuada em relação ao tamanho real.

22) Seis jogadores disputaram um torneio de tiro ao alvo, no qual cada participante podia atirar quatro vezes. Considerando que cada letra representa o resultado de um jogador, verifique qual deles obteve o melhor desempenho geral.

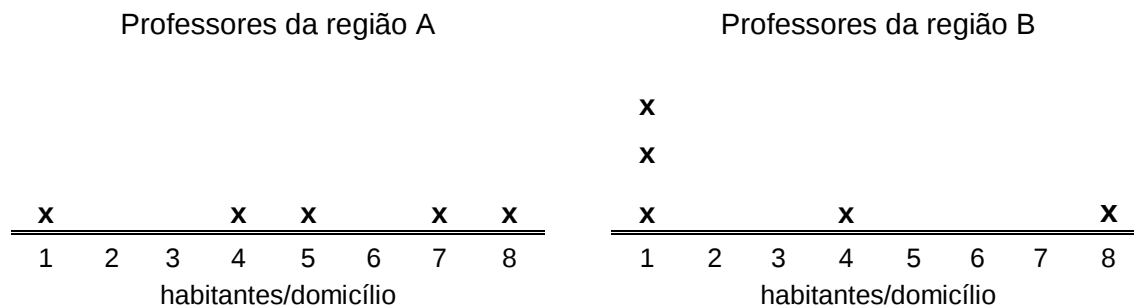
Obs: os números indicam as distâncias de cada circunferência até o alvo.



Em média, qual atirador desviou-se menos do alvo? De quanto foi esse desvio?

Objetivo: levar o aluno a perceber a diferença entre variabilidade entre os dados e variação em relação a um ponto determinado; introduzir o desvio-médio.

23) Foi feita uma pesquisa entre cinco professores da região A e cinco da região B, a respeito do número de moradores de seus domicílios. Os resultados estão indicados nos gráficos abaixo.



- Em qual gráfico, você percebe maior variação dos dados em relação à média?
- A amplitude é suficiente para diferenciar a variabilidade do número de habitantes por domicílio nos dois casos? Por quê?
- Proponha uma forma de medir a variação nas duas situações acima.

Objetivos específicos da atividade 23: perceber as vantagens e desvantagens da amplitude e desvio-médio, como medida de variação e institucionalizar o desvio-médio.

Discussão

24) Pelo que foi estudado até agora:

- Quais as diferenças e semelhanças entre amplitude e desvio-médio?
- Quais as vantagens ou desvantagens de se utilizar amplitude como medida de variação?
- Quais as vantagens ou desvantagens ao se usar o desvio-médio?
- Procure uma definição para o desvio-médio

Objetivo: definir desvio-médio

25) Volte à atividade (19), da medida dos palmos da turma e calcule o desvio-médio dos dados obtidos.

Objetivo: exercitar o cálculo do desvio-médio

26) Escolha três medidas de palmo da mão esquerda (atividade 19): o seu e o de mais dois colegas (com, pelo menos, uma diferente da sua)

(Adaptado de Cordani, 2006)

.....

a) Qual sugestão você daria para resumir estes três dados?

b) Vamos calcular o desvio-padrão destas medidas:

1º Passo: calcular os desvios em relação à média

	Medida observada x_i	Média $\bar{X}$	Desvio $(x_i - \bar{X})$	
1				
2				
3				
		Soma		

c) O que deverá ocorrer com a soma dos desvios?

d) Se não quisermos usar o módulo dos desvios, qual outra alternativa teremos para torná-los positivos?

e) Obtenha a nova soma para os desvios quadrados.

f) Calcule a média dos desvios.

g) Compare o valor obtido no item anterior com os desvios na tabela. O que você percebe em relação ao tamanho dos desvios? Dê uma sugestão para resolver esta situação

Discussão

No item **f**, foi calculada a Variância dos dados e no item **g**, o desvio-padrão entre eles.

27) Volte à atividade 17 e calcule a variância e o desvio-padrão dos dados da escola verde. Para isso, utilize a tabela auxiliar abaixo.

Valor observado (x_i)	Média $\bar{X}$	Desvio ($x_i - \bar{X}$)	($x_i - \bar{X}$) ²
5			
10			
20			
65			
Soma:			
Média:			

Objetivo: exercitar o cálculo do desvio-padrão

28) Duas fábricas produzem pisos quadrados e vendem-nos especificando na caixa as medidas: 400 mm X 400 mm. Entretanto, nem todos os pisos possuem as medidas especificadas.

Suponhamos que os quadros abaixo representem as medidas das larguras³⁷ dos pisos de uma determinada fábrica.

Fábrica A: medidas das larguras dos pisos em milímetros				
400	398	400	399	404
402	402	403	397	395

Fábrica B: medidas das larguras dos pisos em milímetros				
400	398	399,5	401	401
399	396	400	399	401,5

- Calcule a média das larguras dos pisos de cada fábrica.
- De qual fabricante você compraria? Justifique?
- Utilize o desvio-padrão para calcular a variação das medidas dos pisos de cada fábrica. Sua opinião a respeito do item b modificou? Justifique.

³⁷ O correto seria fornecer mais de uma medida para cada piso, mas tomamos apenas a largura para simplificação dos cálculos.

2	3	4	5	6	7	8	9	10	11	12	

- De todas as somas obtidas, calcular a soma média.
- Calcular desvio-padrão das somas obtidas na partida.
- Criar um intervalo de tal forma que o limite inferior seja a média menos um desvio-padrão e o limite superior seja a média mais um desvio-padrão.
[Média – 1DP; Média + 1DP]
- Observe o quadro e conte quantos valores estão compreendidos pelo intervalo [Média – 2DP; Média + 2DP].
- Dentre todos os valores obtidos, qual o percentual dos que estão compreendidos pelo intervalo?
- Crie outro intervalo, de tal forma que o limite inferior seja a média menos dois desvios-padrão e o limite superior seja a média mais dois desvios-padrão. [Média – 2DP; Média + 2DP]
- Quantos valores estão compreendidos pelo intervalo [Média – 2DP; Média + 2DP]?
- Dentre todos os valores obtidos, qual o percentual dos que estão compreendidos pelo intervalo [Média – 2DP; Média + 2DP]?
- Complete a tabela com os resultados obtidos pelas duplas da sala.

Duplas	Percentual de dados compreendidos pelo intervalo [Média – 1DP; Média + 1DP]	Percentual de dados compreendidos pelo intervalo [Média – 2DP; Média + 2DP]
D1		
D2		
D3		
D4		
D5		
D6		
D7		
D8		

- j) Resuma os resultados de cada coluna da tabela anterior, utilizando-se da média e da mediana.
- k) Observando os resultados da tabela anterior, qual hipótese você levantaria sobre o percentual de dados contidos em um intervalo criado com um desvio-padrão, no jogo das somas dos dados? Levante outra hipótese para o percentual de dados contidos em um intervalo criado com dois desvios-padrão.

- 30) Colete entre seus alunos a medida das alturas³⁸ deles em centímetros.
- a) Calcule a altura média e o desvio padrão.
 - b) Obtenha um intervalo³⁹ em torno da média, somando e subtraindo um desvio-padrão da mesma.
 - c) Calcule qual o percentual de medidas foram abarcadas por este intervalo.
 - d) Obtenha um intervalo em torno da média somando e subtraindo dois desvios-padrão da mesma.
 - e) Calcule qual o percentual de medidas foi abarcado por este intervalo.

- 31) Repita a atividade anterior, utilizando outra medida antropométrica dos alunos

32) Nas atividades 29, 30 e 31, você construiu um intervalo somando e subtraindo dois desvios-padrão à média e calculou o percentual de dados limitados por este intervalo. Levante uma hipótese sobre a porcentagem dos dados que, normalmente, estão compreendidos neste intervalo.

Objetivo das tarefas 30, 31 e 32: obter experimentalmente o percentual de dados, contidos em intervalos criados a partir de um e dois desvios da média.

- 33) Repita a mesma análise da tarefa 32, mas agora levantando uma hipótese sobre a

³⁸ Podem ser outras medidas antropométricas, ou seja, medidas do palmo, do perímetro da cabeça, tamanho do pé, etc.

³⁹ Não confundir o intervalo obtido por este procedimento com intervalo de confiança, que é obtido com a adição e subtração à média amostral do desvio-padrão, dividido pela raiz do tamanho da amostra, ou seja, $I_c = \bar{X} \pm z \cdot \frac{DP}{\sqrt{N}}$, onde z é a quantidade de desvio-padrão.

porcentagem de dados que, normalmente, estão compreendidos no intervalo construído com um desvio-padrão.

Objetivo: inferir sobre o percentual de dados contidos em intervalos criados, a partir da subtração e soma de um e dois desvios-padrão da média.

Discussão e institucionalização da Regra Empírica

34) Observando a tabela abaixo, obtida a partir de dados reais, na qual estão registrados os percentuais de valores contidos nos intervalos criados com a soma e subtração de um e dois desvios-padrão à média, verifique se os percentuais obtidos assemelham-se às informações teóricas descritas anteriormente.

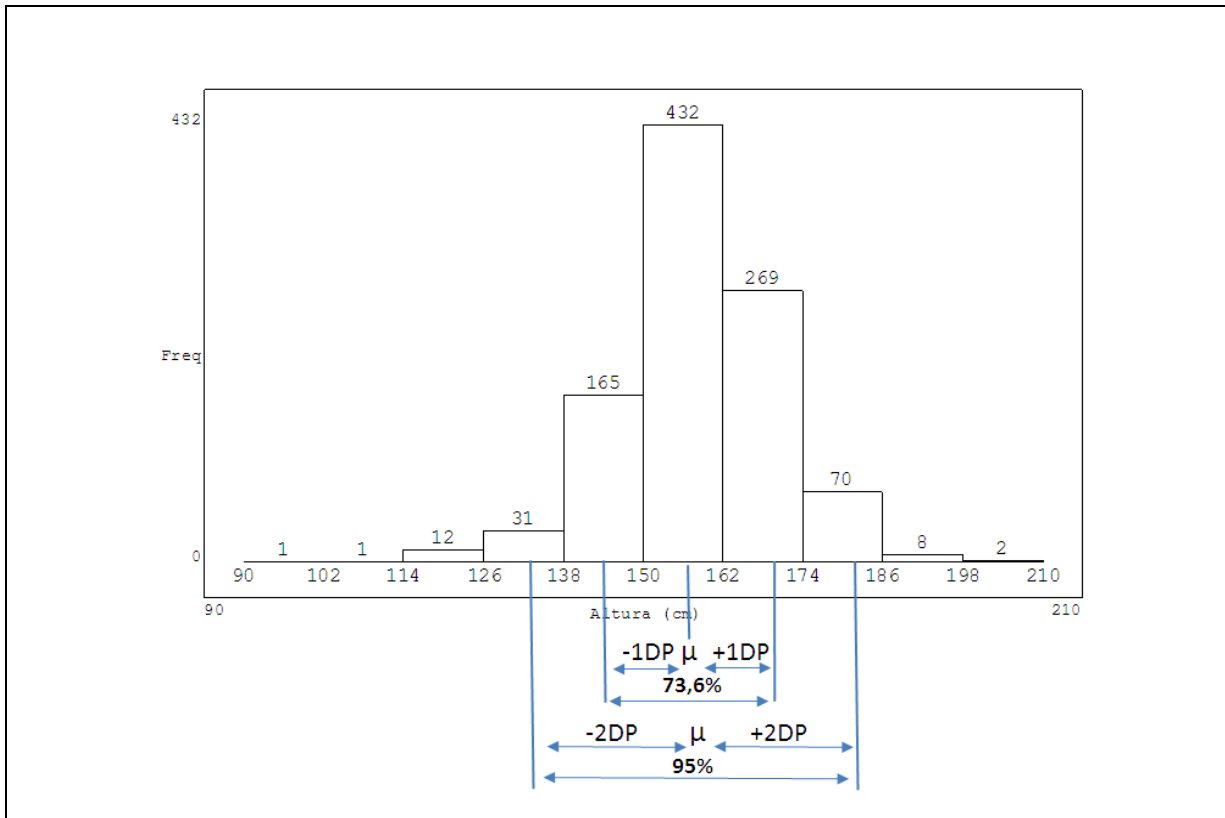
Tipo de medida	Percentual de dados compreendidos por [Média – 1DP; Média + 1DP]	Percentual de dados compreendidos por [Média – 2DP; Média + 2DP]
Alturas de 104 alunos do Reino Unido	83%	95%
Alturas de 100 alunos da África do Sul	75%	95%
Alturas de 140 alunos de 5 ^{as} séries no Brasil	75%	95,7
Comprimento dos pés de 100 alunos da África do Sul	72%	96%
Comprimento dos pés de 76 alunos de 5 ^a séries no Brasil	76,6	97,4
Comprimento dos pés de 104 alunos do Reino Unido	78%	93%
Perímetro dos punhos de 104 alunos do Reino Unido	66%	97%
Número de moradores por domicílio de 100 alunos da África do Sul	76%	96%
Tempo para ir da casa à escola de 100 alunos da África do Sul	89%	97%
Distância da casa à escola de 100 alunos da África do Sul	78%	100%
Altura de 318 alunos brasileiros	69%	97,7

Objetivo: inferir sobre o percentual de dados contidos em intervalos criados a partir da subtração e soma de um e dois desvios-padrão da média.

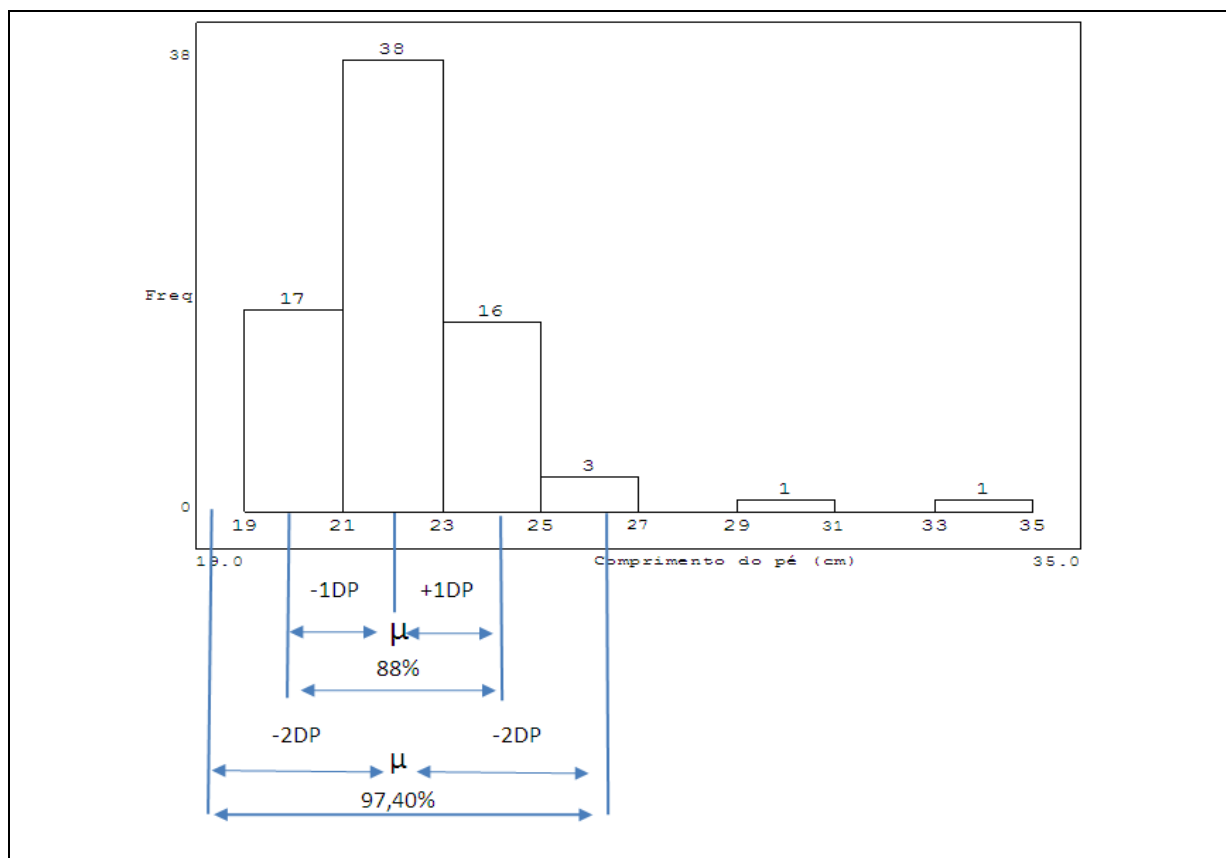
35) Construa⁴⁰ os histogramas dos dados das atividades 30 e 31, localize as respectivas médias e represente os intervalos com segmentos de reta.

Objetivo: desenvolver a percepção visual da variação e do intervalo

Exemplo de resolução com alguns dados coletados pelo instrutor



⁴⁰ A proposta inicial era que os alunos construíssem os histogramas, mas por falta de tempo foi-lhes fornecido o histograma pronto para a realização da tarefa.



36) Compare e comente os resultados dos itens anteriores (atividade 29 a 35), sugerindo o papel desempenhado pelo desvio padrão na criação de intervalos para análise de dados?

Objetivo: refletir sobre a importância da Regra Empírica na análise de dados.

37) Observe os gráficos das somas dos dados (tarefa 29) e os histogramas da aula de hoje (tarefa 35). Você percebe alguma semelhança entre o formato deles? Comente.

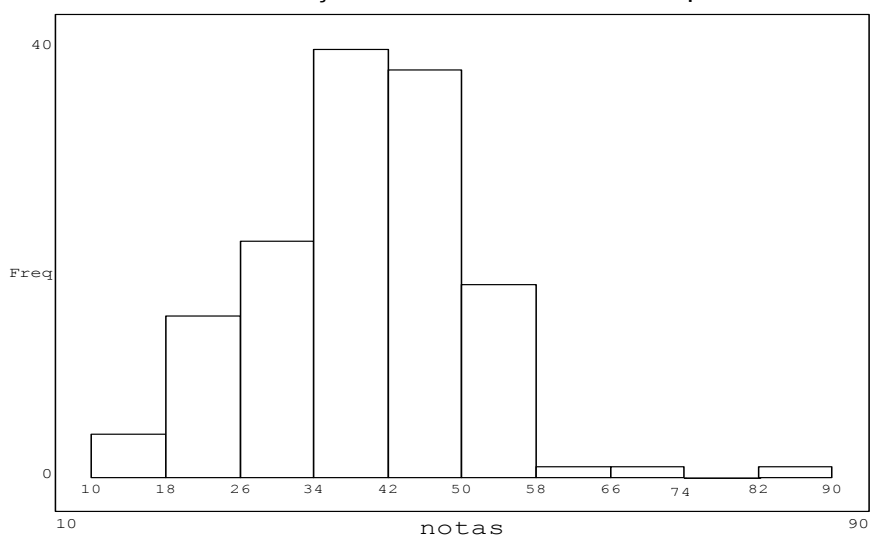
Objetivo: observar a forma como os dados se distribuem e relacionar a regra empírica com distribuições aproximadamente simétricas.

Discussão

Parte G: O desvio-padrão e o evento raro

38) O histograma abaixo representa a distribuição das notas dos alunos de um curso de Matemática na disciplina de Estatística. A nota média do grupo é 40, o desvio-padrão 11,4 e a porcentagem de dados contidos no intervalo [média – 2DP; média + 2DP] é de 96,3%.

- a) Represente com segmentos de retas verticais, abaixo do gráfico, a média e o intervalo criado com a subtração e soma de dois desvios-padrão.



- b) Um aluno obteve desempenho 70. Podemos considerar esta nota rara (ou excepcional) em relação a seu grupo?

Objetivo: utilizar o intervalo para considerar um evento raro

A tarefa 39 foi entregue em quatro momentos: primeiramente o enunciado e, só após os alunos tentarem resolver uma questão, era entregue a outra.

39) No rádio do carro, geralmente, escutamos a informação sobre o congestionamento naquele horário. Em um dia da semana passada, ouvi que o congestionamento em uma grande cidade naquele horário era de 100 km. Será que este valor era exagerado ou poderia ser considerado normal?

- Sabendo-se que a média para o horário é de 80 Km, devo considerar o congestionamento de 100 km exagerado? Explique.
- Considerando que o desvio-padrão dos congestionamentos para o horário é de 8 km, posso considerar 100 km como um valor exagerado? Explique.
- Se o desvio-padrão dos congestionamentos for 30 km, posso considerar 100 km um acontecimento raro?

(Adaptado da International Statistical Literacy Project - Fase 2 – Brasil, março/2009)

Objetivo: perceber a importância de se considerar a média vinculada ao desvio-padrão para tomar a decisão. Medir afastamento da média em unidades de desvio-padrão relacionado com a regra empírica.

40) Variabilidade em Pesquisas

(Adaptado de Cordani, 2006)

Imagine o seguinte cenário:

Uma cidade possui 300 eleitores e dois candidatos a prefeito: Sr. Bruno Branco e Sra. Vera Vermelha. O candidato Bruno contratou um instituto de pesquisa para saber a preferência dos eleitores.

a) Temos um saco opaco com 300 bolinhas, entre brancas e vermelhas, representando a intenção de voto de cada eleitor da população, isto é, bolas brancas representam as pessoas que votarão no Sr. Bruno e vermelhas, os eleitores da Sra. Vera. Considerando que você não pode contar todas as bolinhas do saco, qual sugestão você daria para estimar a porcentagem de votos no Sr Bruno?

Tendo colhida as respostas dos alunos e promovido o debate das sugestões, apresentamos a seguinte questão:

b) As amostras preservam as mesmas porcentagens?

Simulação proposta

1º Passo: dividir a turma em sete grupos, fornecendo-lhes um saco com 300 bolinhas, entre brancas e vermelhas, e uma pá com 50 orifícios. As equipes devem utilizar a pá para retirar, aleatoriamente, uma amostra de 50 bolinhas, simbolizando uma entrevista com 50 eleitores consultados. Por meio do percentual de bolinhas brancas na pá, fazer uma estimativa para toda a população.

2º Passo: Um representante de cada grupo deve ir ao quadro anotar sua estimativa

Grupo	Votos Bruno	Percentual ⁴¹
1		
2		
3		
4		
5		
6		
7		

⁴¹ Bolinhas brancas em relação ao total da amostra

- c) A tabela exibe sete diferentes porcentagens de eleitores do candidato Bruno. Qual percentual é o “mais” verdadeiro ?
- d) Sugira uma forma de obter uma estimativa da verdadeira porcentagem, usando os resultados obtidos pelos grupos.
- e) Baseando-se nas amostras, como poderemos criar um intervalo para estimar a porcentagem de votos do Sr. Bruno?
- f) Utilizando a média e dois desvios-padrão, crie um intervalo para estimar a porcentagem de votos do Sr. Bruno?
- g) Qual foi a margem de erro (empírica), para mais ou para menos em relação média?
- h) Quando utilizamos o intervalo acima (item f) para estimar os votos do Sr. Bruno, qual a chance (aproximada) que temos de acertar?

Objetivo: introduzir os intervalos, como medida (empírica) de erro de estimativa.

41) Um professor de Matemática aplica duas avaliações diferentes a duas turmas. Os resultados foram:

Turma A: Média 75 e Desvio-padrão 14

Turma B: Média 40 e Desvio-padrão 8

Que nota é relativamente melhor: 82 na avaliação da turma A ou 47 na turma B.

(Questão adaptada de Triola, 1999, p.46)

Objetivo: Utilizar o desvio-padrão como unidade de medida para decidir

42) As alturas da população de homens adultos têm média 169 cm, desvio-padrão 6,9 cm e distribuição simétrica com formato triangular. O jogador de basquete Michael Jordan ganhou reputação de gigante por suas proezas no jogo, mas, com 191 cm, ele poderia ser considerado excepcionalmente alto, comparado com a população geral de homens adultos?

(Questão adaptada de Triola, 1999, p.46)

Objetivo: utilizar o intervalo para decidir se o evento é raro

43) Suponhamos que no processo seletivo para ingresso em um curso técnico disputado, o critério de seleção seja o desempenho do candidato nas provas de Português e de Matemática. Disputando a última vaga, os candidatos abaixo obtiveram as seguintes notas:

	Nota em Português	Nota em Matemática
Candidato A	60	80
Candidato B	80	60

Notas de todos os candidatos que disputaram as vagas existentes

Notas na prova de Português				Notas na prova de Matemática			
60	0	20	100	50	50	50	100
10	60	20	60	30	20	60	80
60	60	70	80	40	20	30	70

Baseando-se apenas nas notas apresentadas, utilize as ferramentas estatísticas estudadas até agora e decida qual candidato obteve o melhor desempenho.

Objetivo: aplicar o z-escore na escolha do candidato

APÊNDICE B: QUESTIONÁRIO DIAGNÓSTICO⁴²

Seu nome não será divulgado e seus dados só serão utilizados nesta pesquisa com o seu consentimento.

1) Nome: _____

2) Curso: _____ 4) Qual semestre? _____

4) Data de nascimento: _____ 5) Altura: _____

5) Quantas pessoas vivem em sua casa (incluindo você)? _____

As questões de 7 a 9 deverão ser respondidas apenas pelas pessoas que lecionam

6) É professor? () Sim () Não 7) () Ed. Infantil; () EF(1-5^o); () EF(6-9^o); () EM

8) Qual disciplina leciona? _____ 9) A quanto tempo leciona? _____

Algumas questões:

10) O que é média?

11) O que é mediana?

12) Um aluno foi avaliado três vezes em cada uma das disciplinas descritas na tabela abaixo

Disciplina	Notas
Matemática	7,0 - 7,0 - 7,0
Português	10,0 - 1,0 - 10,0
Biologia	5,0 - 7,0 - 9,0

Compare quantitativamente o desempenho dos alunos nas três disciplinas.

⁴² As questões 10, 11 e 12 foram aplicadas no primeiro e último dia da formação.

APÊNDICE C: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Participação em Pesquisa

Eu, _____,
com _____ anos de idade, portador do RG _____. residente na _____, com número de telefone _____ email _____, abaixo assinado, dou meu consentimento livre e esclarecido para participar como voluntário(a) da pesquisa, sob responsabilidade de Marcelo Marcos Bueno Moreno, aluno do curso de Mestrado em Educação Matemática da PUC-SP, de Diva Valério Novaes, aluna do curso de Doutorado em Educação Matemática da PUC-SP e da Profa Dra. Cileda de Queiroz e Silva Coutinho, orientadora da pesquisa de ambos e docente do Programa de Pós-Graduação da PUC-SP.

Assinando este Termo de Consentimento estou ciente de que:

- 1) O objetivo da pesquisa é desenvolver e aplicar uma sequência de ensino para conteúdos de Estatística.
- 2) A realização da pesquisa é fundamental para a produção de material didático que apóie os professores de matemática no ensino de Estatística na Escola Básica.
- 3) Assim que for terminada a pesquisa terei acesso aos resultados globais do estudo.
- 4) Estou livre para interromper, a qualquer momento, minha participação nesta pesquisa.
- 5) A participação nesta pesquisa é voluntária.
- 6) Os dados pessoais dos integrantes serão mantidos em sigilo e os resultados obtidos com a pesquisa serão utilizados apenas para alcançar os objetivos do trabalho, incluindo a publicação na literatura científica especializada.
- 7) Obtive todas as informações necessárias para poder decidir conscientemente sobre minha participação na referida pesquisa.
- 8) Este termo de consentimento é feito em duas vias, de maneira que uma permanecerá em meu poder e a outra com os pesquisadores responsáveis.

Caraguatatuba, 19 de setembro de 2009

Assinatura do participante: _____

Marcelo Marcos Bueno Moreno: _____

Diva Valério Novaes: _____