



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Instituto de Física
Programa de Pós-Graduação em Ensino de Física
Mestrado Profissional em Ensino de Física

Medição, Incerteza e Evidência Empírica

Caio Marchon Ferreira

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ensino de Física, Instituto de Física, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Ensino de Física.

Orientadores:
Carlos Eduardo Aguiar
Roberto Affonso Pimentel Jr

Rio de Janeiro
Dezembro de 2024

Medição, Incerteza e Evidência Empírica

Caio Marchon Ferreira

Orientadores:

Carlos Eduardo Aguiar
Roberto Affonso Pimentel Jr

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Ensino de Física, Instituto de Física, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Ensino de Física.

Aprovada por:

Dr. Roberto Affonso Pimentel Jr, CAP-UFRJ, externo (Presidente)

Dra. Lúcia Helena Coutinho, IF-UFRJ, interna

Dr. Marcelo Martins Sant'Anna, IF-UFRJ, externo

Dra. Simone Coutinho Cardoso, IF-UFRJ, externo

Rio de Janeiro
Dezembro de 2024

FICHA CATALOGRÁFICA

F383m Ferreira, Caio Marchon
Medição, Incerteza e Evidência Empírica / Caio Marchon
Ferreira. – Rio de Janeiro: UFRJ/IF, 2024.
ix, *** f. : il. ; 30 cm.
Orientadores: Carlos Eduardo Aguiar; Roberto Affonso
Pimentel Jr.
Dissertação (mestrado) – UFRJ / Instituto de Física /
Programa de Pós-Graduação em Ensino de Física, 2024.
Referências Bibliográficas: f. ***-***.
1. Ensino de Física. 2. Incerteza de Medição. 3. Evi-
dência Empírica. I. Aguiar, Carlos Eduardo. II. Pimentel,
Roberto. III. Universidade Federal do Rio de Janeiro, Ins-
tituto de Física, Programa de Pós-Graduação em Ensino de
Física. IV. Medição, Incerteza e Evidência Empírica.

Dedico esta dissertação a todos os professores deste país.

Agradecimentos

Aos meus orientadores, Carlos Eduardo Aguiar e Roberto Pimentel Jr, por terem escutado minhas ideias iniciais e me ajudado a desenvolvê-las, culminando neste trabalho de mestrado. Admiro enormemente a sabedoria e capacidade de lidar com problemas que os dois carregam.

Ao professor Gustavo Rubini, por toda contribuição para este trabalho. Sua participação foi muito importante para a realização desta dissertação.

Aos meus pais, que me apoiam com vigor nos projetos acadêmicas em que me coloco.

À minha irmã, sempre me aconselhando sobre como lidar com as situações adversas da vida.

À minha amada Laryssa, que me acompanhou durante essa jornada e me trouxe paz em momentos difíceis.

E, de forma geral, à minha família e amigos, sempre me dando forças para atravessar os desafios.

“O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001”

RESUMO

Medição, Incerteza e Evidência Empírica

Caio Marchon Ferreira

Orientadores:

Carlos Eduardo Aguiar

Roberto Affonso Pimentel Jr

Resumo da Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Ensino de Física, Instituto de Física, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Ensino de Física.

Nesta dissertação discutimos as ideias que estudantes costumam manifestar sobre processos de medição e, com base nelas, apresentamos uma proposta para o ensino de noções sobre a coleta, sistematização e análise de dados experimentais. A proposta é dirigida ao ensino médio e está estruturada como um conjunto de atividades em sala de aula envolvendo a medição do tempo de reação dos alunos. Junto a isso, um texto de apoio é oferecido para aprofundar aspectos essenciais após a realização da aula. A importância de realizar repetidas medições é destacada durante as atividades, assim como maneiras de expor e sistematizar o conjunto dos resultados encontrados. Representações visuais como o histograma são apresentadas e os conceitos de valor médio e incerteza das medidas são introduzidos, com ênfase em seu significado intuitivo. O papel da incerteza é realçado, mostrando como é crítico avaliar a confiabilidade de resultados experimentais antes de tirar conclusões e tomar decisões baseadas neles. Na dissertação descrevemos também a realização de três aplicações dessa sequência didática e discutimos seus resultados, avaliados através de pré e pós-testes.

Palavras chave: Ensino de Física, Incerteza de Medição, Evidência Empírica.

Rio de Janeiro
Dezembro de 2024

ABSTRACT

Measurement, Uncertainty and Empirical Evidence

Caio Marchon Ferreira

Supervisors:

Carlos Eduardo Aguiar

Roberto Affonso Pimentel Jr

Abstract of master's thesis submitted to Programa de Pós-Graduação em Ensino de Física, Instituto de Física, Universidade Federal do Rio de Janeiro, in partial fulfillment of the requirements for the degree Mestre em Ensino de Física.

In this dissertation we discuss common ideas that students express about the measurement process, and present a proposal for the teaching of basic concepts related to the collection, organization, and analysis of experimental data. The proposal is designed for high school students and consists of a set of classroom activities that involve measuring their reaction times. In addition, a support text designed to deepen essential aspects after the class is offered. Throughout these activities we emphasize the importance of taking multiple measurements, as well as methods for presenting and organizing the resulting data set. Visual representations such as histograms are introduced, along with the concepts of average value and measurement uncertainty, with a focus on making these ideas intuitive. The role of uncertainty is emphasized, showing how critical it is to assess the reliability of experimental results in order to draw meaningful conclusions and make informed decisions. In the dissertation we also describe three classroom applications of this teaching sequence and discuss their outcomes, evaluated through pre- and post-tests.

Keywords: Physics education, Measurement Uncertainty, Empirical Evidence.

Rio de Janeiro
Dezembro de 2024

Sumário

1	Introdução	1
2	Concepções sobre medida e incerteza	5
2.1	Paradigmas de ponto e conjunto	5
2.2	Questionário de medidas físicas (PMQ)	10
2.3	Nossa adaptação do PMQ	13
3	Uma proposta para o ensino de medida e incerteza	18
3.1	Tempo de reação	19
3.2	Aplicativo, coleta de dados e histograma	21
3.3	A média de um conjunto de medidas	24
3.4	Dispersão das medidas e desvio-padrão	26
3.5	Flutuação da média e desvio-padrão da média	29
3.6	Comparação entre grupos de dados	32
3.7	Justificativa para as regras de bolso	37
4	Aplicações e resultados	44
4.1	Procedimento	44
4.1.1	Introdução	45
4.1.2	Medição do tempo de reação	46
4.1.3	Dispersão	46
4.1.4	Incerteza da média	47
4.1.5	Comparação entre grupos de dados	49
4.2	Aplicações e resultados	49
4.2.1	Primeira aplicação	50
4.2.2	Segunda aplicação	51
4.2.3	Terceira aplicação	53
5	Conclusões	56
A	Medidas e Incertezas	58

B	Questionário de Medidas Físicas	76
C	Classificação das respostas ao Questionário de Medidas Físicas	97
	Referências bibliográficas	102

Capítulo 1

Introdução

No ensino de ciências da natureza e engenharia, a discussão sobre medição e incerteza é bem estabelecida em disciplinas de laboratório, programadas para os períodos iniciais dos cursos universitários. Isso se deve, em parte, à relevância de introduzir os estudantes a práticas experimentais que os preparem para verificar a validade de uma informação e os iniciem nos procedimentos necessários para produzir conteúdos baseados em evidências. Noções corretas sobre medição são fundamentais para que o conhecimento produzido seja válido e confiável. Cursos de ciências sociais e da área da saúde também possuem disciplinas que abordam a obtenção e tratamento de medidas e avaliação de suas incertezas, não sendo esse um assunto tratado apenas nas ciências da natureza.

Para além da relevância no meio acadêmico, as medições estão presentes na rotina de todos. Elas aparecem quando se fala sobre o sucesso de uma política pública, o consumo médio de combustível de um veículo por quilômetro percorrido, a eficácia de um medicamento para o tratamento de alguma doença, entre muitos outros exemplos em que se deseja avaliar valores que foram atribuídos a uma grandeza. Geralmente esses valores estão associados à tomada de decisões, como, pensando nos exemplos acima, se devemos alterar determinada política pública, qual o veículo mais econômico para se comprar ou se um médico deve ou não receitar um medicamento. Logo, valores obtidos em medições influenciam diretamente as atitudes da

população, e é importante notar que todas essas medidas possuem incertezas que infelizmente são ignoradas em muitas situações.

As incertezas ocorrem porque nenhuma medida é infinitamente precisa em sua capacidade de representar a grandeza a que se refere. Por isso, estimar um valor para a incerteza de uma medida é algo relevante, que deve ser feito em situações envolvendo tomadas de decisão. A incerteza de medição evidencia a impossibilidade de obter o valor “verdadeiro” de uma grandeza [1, p. 19]; ela complementa o valor obtido na medição, indicando o quão confiável ele é. Uma medida com pequena incerteza oferece maior confiabilidade que uma medida com uma grande incerteza.

Apesar da noção de incerteza de medição ser familiar no meio científico e técnico, esse conceito não é tão disseminado na sociedade. Isso é problemático, dado que decisões são tomadas baseadas em medidas e medidas podem ser mais ou menos confiáveis devido a suas incertezas. Não é surpreendente que durante o episódio da pandemia de Covid-19 uma considerável parcela da população escolheu aderir a tratamentos sem eficácia comprovada, sustentando-se em observações pontuais e depoimentos de cura de pacientes que utilizavam algum tratamento não-convencional. A cura de um paciente que foi tratado por um método alternativo não é um indicador real da eficácia do tratamento, apesar de ser popularmente confundido com uma evidência suficiente. Essa confusão, recorrente na sociedade contemporânea, pode ser relacionada a uma concepção ingênua que ignora a necessidade uma grande quantidade de medidas, sob condições clínicas controladas, apontando sucessos e fracassos, para que seja possível estabelecer uma eficácia e sua incerteza. Além disso, parte importante desse processo costuma ser a comparação com os resultados em pacientes que receberam um tratamento placebo.

Cientes dessa compreensão inadequada sobre o significado de evidência empírica, propomos nesta dissertação uma sequência didática que aborda o processo de medição e estimação de incertezas, com ênfase na coleta de medidas, sistematização dos resultados e comparação entre conjuntos de dados. Tomamos como referencial os estudos das concepções de estudantes sobre o processo de medição em diferentes contextos de ensino. Dentre eles pode ser destacado o trabalho de Lubben et al. [2], em que se estabelecem os concei-

tos de “raciocínio de ponto” e “raciocínio de conjunto”, que categorizam a forma com que os estudantes concebem o processo de medição. O objetivo da nossa proposta é fazer com que os estudantes se afastem de uma perspectiva ingênua do processo de medição, correspondente ao raciocínio de ponto, segundo a qual uma única medida pode fornecer o valor verdadeiro de uma grandeza [3]. Procuramos aproximá-los de uma perspectiva em que a melhor estimativa do valor dessa grandeza possui necessariamente uma incerteza, estabelecendo um viés probabilístico no processo de medição, correspondente ao raciocínio de conjunto [4]. Além disso, enfatizamos que essas incertezas desempenham um papel fundamental na comparação entre diferentes conjuntos de dados e, portanto, são parte importante de um processo de tomada de decisão que tenha como base essa comparação.

A sequência didática desenvolvida utiliza a medição do tempo de reação para abordar processos em que a obtenção de várias medidas é essencial para a estimativa de incertezas [5, p. 3]. Para isso desenvolvemos um aplicativo para a medição do tempo de reação, que facilita a tarefa de coleta e sistematização dos dados. Também propomos procedimentos simplificados para a comparação de conjuntos de dados, baseados na inspeção de gráficos com a média e barra de erro desses conjuntos. A sequência didática é planejada para uma aula prática de 50 minutos e é acompanhada da disponibilização de um material textual para leitura complementar. Ela foi aplicada em três oportunidades diferentes a estudantes do ensino médio de uma escola pública.

Os próximos capítulos desta dissertação estão organizados da seguinte forma. O capítulo 2 traz uma revisão da literatura que utilizamos para identificar e classificar as concepções dos estudantes sobre medida e incerteza. Ainda nesse capítulo, discutimos um teste conceitual que tem sido extensivamente utilizado como instrumento de avaliação dessas concepções e que adotaremos neste trabalho. O capítulo 3 é uma apresentação da sequência didática que desenvolvemos. Nele estão os conteúdos trabalhados e a fundamentação para a escolha desses em detrimento de outros. O funcionamento do aplicativo para a medição do tempo de reação também é explicado. O capítulo 4 descreve as aplicações da sequência didática realizadas em sala de aula e discute a evolução do procedimento adotado nas sucessivas in-

tervenções. Além disso apresentamos uma avaliação dos resultados dessas aplicações, analisados por meio do teste conceitual discutido no capítulo 2, administrado antes e após cada intervenção. Por fim, o capítulo 5 contém as conclusões deste trabalho. No apêndice A está o texto de apoio à sequência didática, que constitui o material instrucional associado a esta dissertação. Os apêndices B e C apresentam o teste conceitual que utilizamos nas avaliações.

Capítulo 2

Concepções sobre medida e incerteza

Neste capítulo falaremos inicialmente, na seção 2.1, sobre as referências que utilizamos na elaboração do nossa proposta instrucional. Será apresentada uma hierarquia de concepções dos estudantes sobre coleta e análise de dados empíricos e como essas concepções se encaixam em dois tipos de raciocínio: de ponto ou de conjunto. Em seguida, na seção 2.2 apresentaremos um teste conceitual sobre medidas físicas (o PMQ) que tem sido muito utilizado em estudos nessa área. Adaptamos o teste para avaliar o impacto de nossas intervenções e descreveremos essa versão do PMQ na seção 2.3. O presente capítulo não pretende apresentar uma revisão extensiva sobre o ensino de medidas e incertezas. Além das referências discutidas aqui, muitos outros trabalhos que abordam o processo de medição no ensino básico de física podem ser encontrados na revisão da literatura realizada por Pigosso e Heidemann [6].

2.1 Paradigmas de ponto e conjunto

Para desenvolver o material educacional apresentado nesta dissertação, foram utilizados trabalhos relacionados ao ensino e aprendizagem do processo de coleta, sistematização e análise de dados experimentais. Em particular, os

trabalhos que discutem as concepções dos estudantes sobre medidas e incertezas constituem a base conceitual de nosso estudo. Como ressaltado por Séré, Journeaux e Larcher [7]: “muitos artigos que estudam técnicas de análise de erro em trabalhos experimentais estão disponíveis. Entretanto, poucos artigos descrevem as concepções dos estudantes sobre essas técnicas”¹. Mesmo a pesquisa sobre ensino do processo de medição, sem ênfase na descrição de concepções, ainda é incipiente, como apontam Pigosso e Heidemann [6] em sua revisão sobre o assunto. Geralmente essa é a abordagem em trabalhos sobre os métodos de medição tratados em laboratórios didáticos de ensino básico. Artigos nesses dois temas – concepções e métodos – foram considerados na elaboração desta dissertação.

Uma abordagem das concepções sobre medição que desempenha papel fundamental em nosso trabalho foi realizada por Lubben e Millar [8]. Nesse artigo, os autores construíram uma hierarquia das concepções de estudantes sobre coleta e avaliação de dados empíricos. Mais de 1000 estudantes do Reino Unido, com idades entre 11 e 16 anos responderam a um teste, criado pelos próprios pesquisadores, com seis questões, as quais deviam ter respostas justificadas por extenso. Os pesquisadores buscaram padrões nas justificativas escritas pelos estudantes para detectar quais seriam as principais ideias utilizadas para fundamentar suas respostas. Os padrões que englobam as ideias utilizadas foram ordenados em níveis de A até H, sendo A o nível mais básico de entendimento e H o mais sofisticado.

Algumas das formulações utilizadas pelos estudantes para responder às questões foram apresentadas por Lubben e Millar. Como exemplo, há a questão em que devem escolher se várias medições são necessárias para conhecer o tempo de dissolução de uma quantidade de açúcar na água. Um argumento contrário à realização de várias medições foi: “se usarmos a mesma quantidade das substâncias, a resposta será a mesma” [8, p. 958]. Um argumento favorável foi: “é aconselhável fazer um experimento mais de uma vez, mas é ainda melhor fazê-lo algumas vezes. Assim, você pode ver como os resultados variam” [8, p. 959]. Respostas como essas indicam a existência de diferentes

¹O texto original se encontra em inglês. Salvo indicação em contrário, utilizaremos traduções nossas nas citações de trechos em língua estrangeira.

perspectivas sobre obtenção de medidas.

O artigo de Lubben e Millar proporcionou um entendimento inicial de como pessoas de fora do meio universitário e técnico concebem medições e critérios de confiança nas medidas. Além disso, os níveis de concepções mais sofisticados fornecem modelos das ideias que desejamos que os estudantes sejam capazes de compreender e utilizar, após terem contato com o material instrucional que desenvolvemos neste trabalho. Mesmo tendo origem na análise de respostas de estudantes em uma faixa de idade específica e no Reino Unido, os níveis conceituais têm servido para classificar as ideias acerca do processo de medição em outros contextos, como verificado por Allie et al. em artigo [9] no qual universitários da África do Sul, em seu primeiro ano de curso, foram avaliados com uma adaptação do teste de Lubben e Millar.

A tabela 2.1 é uma tradução da tabela encontrada no artigo de Allie et al., que resume os níveis originalmente propostos por Lubben e Millar. Uma concepção extra (nível I) foi acrescentada por Allie et al. à classificação original, pois os estudantes universitários utilizaram ideias que poderiam ser agrupadas em uma categoria mais avançada, envolvendo a comparação entre conjuntos de dados.

Nível	Perspectiva do estudante sobre o processo de medição
A	Medir uma vez e esse é o valor verdadeiro.
B	A medida é correta, a menos que você encontre um valor fora do esperado.
C	Faça medições para praticar. Em seguida, faça uma medida do que você precisa.
D	Faça medições até encontrar um valor recorrente, esse será o valor verdadeiro.
E	Você precisa tirar uma média de diferentes medidas. Varie ligeiramente as condições para evitar obter os mesmos resultados.
F	Calcular a média de várias medidas para lidar com a variação, causada pela imprecisão da medição. A qualidade do resultado só pode ser julgada por uma autoridade.
G	Calcular a média de várias medidas. A dispersão do conjunto de medidas indica a qualidade do resultado.
H	A consistência do conjunto de medidas pode ser avaliada e medições anômalas precisam ser rejeitadas antes de se tomar uma média.
I	A compatibilidade de conjuntos de dados pode ser julgada comparando as posições relativas de suas médias com suas dispersões.

Tabela 2.1: Classificação proposta por Lubben e Millar [8], com o nível I acrescentado por Allie et al. [9].

Visando criar uma classificação mais prática para o desenvolvimento de propostas de ensino, Lubben et al. [2] propuseram a existência de dois paradigmas sobre medição: “paradigma de ponto” e “paradigma de conjunto”, também chamados “raciocínio de ponto” e “raciocínio de conjunto”. Nessa classificação, as ideias utilizadas para abordar medições seriam geradas a partir de um dos dois tipos de raciocínio. Segundo Lubben et al. [2, p. 312-313]:

- “O raciocínio de ponto é caracterizado pela noção implícita de que cada medida pode ser o valor verdadeiro do objeto. Como consequência, cada

medida é independente das outras e elas não são combinadas de forma alguma”.

- Para o raciocínio de conjunto, “a melhor informação sobre o valor verdadeiro de uma grandeza é obtida pela combinação de medidas utilizando constructos teóricos visando descrever os dados coletivamente”.

Como exposto por Pollard et al. [10], o raciocínio de ponto está diretamente relacionado à ideia de obter o valor verdadeiro de uma grandeza, desconsiderando a existência da incerteza de medição. Desvios do valor verdadeiro são entendidos como erros no processo de obtenção de dados. Enquanto isso, o raciocínio de conjunto proporciona o entendimento de que cada medida traz informações sobre o mensurando, sendo necessário considerar toda a informação disponível e avaliar o grau de confiança dessa informação, dado pela incerteza de medição.

Essa classificação em raciocínio de ponto ou de conjunto divide a hierarquia de concepções proposta anteriormente em duas partes. Os níveis de A, B, C e D da tabela 2.1 pertencem ao raciocínio de ponto, enquanto os níveis E, F, G, H e I, ao raciocínio de conjunto. Lubben et al. [2] ainda sugerem a subdivisão do raciocínio de conjunto em básico e profundo. A classificação das ideias dos estudantes sobre o processo de medição em dois tipos de raciocínio possibilitou uma análise prática do modo pelo qual eles interpretam o assunto, e há vários exemplos de uso dessa divisão em avaliações de aprendizagem [2, 4, 10–13, 15]. Quase sempre isso é feito através de testes conceituais aplicados antes e depois da instrução correspondente (pré/pós-testes).

Um exemplo de utilização dos raciocínios de ponto e conjunto na análise da perspectiva dos estudantes sobre o processo de medição é o estudo feito por Pollard et al. [10]. Nele, dois cursos introdutórios de laboratório da Universidade do Colorado (o curso original e o transformado) foram avaliados segundo sua eficácia em promover o raciocínio de conjunto. A figura 2.1 mostra gráficos com a proporção de cada tipo de raciocínio utilizado no pré e pós-teste.

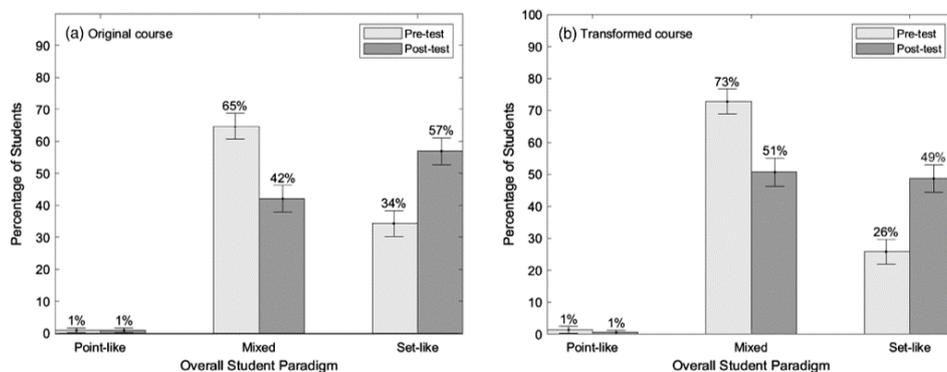


Figura 2.1: Resultados do pré-teste (barras claras) e pós-teste (barras escuras) em dois cursos de laboratório: original (à esquerda) e transformado (à direita). As barras representam a proporção dos paradigmas de ponto, misto e de conjunto. As barras de erro são intervalos de confiança de 95%. Fonte: [10, p. 9].

No estudo mostrado na figura 2.1, muitos estudantes utilizavam o “raciocínio misto” antes dos dois cursos. Esse conceito, para Pollard et al., indica a utilização de raciocínio de ponto em algumas situações e o de conjunto em outras. É interessante notar que poucos estudantes manifestaram o raciocínio de ponto em todas as situações avaliadas, mesmo antes do curso.

2.2 Questionário de medidas físicas (PMQ)

O instrumento mais utilizado em avaliações conceituais sobre medidas e incertezas tem sido, provavelmente, o Questionário de Medidas Físicas (*Physics Measurement Questionnaire* ou PMQ). Esse questionário foi criado com base no teste utilizado por Lubben e Millar [8]. Após várias alterações, foi consolidado um padrão para as questões, exposto na monografia “Teaching scientific measurement at university: understanding students’ ideas and laboratory curriculum reform” [13], de Campbell et al.

Todas as questões do questionário estão relacionadas com um experimento de rolamento de uma bolinha através de uma calha, que termina em uma situação de lançamento horizontal. A figura 2.2 contém a ilustração do

experimento utilizada originalmente.

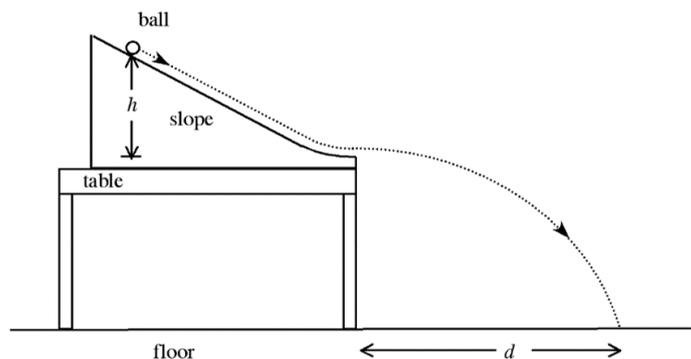


Figura 2.2: O experimento no qual se baseia o PMQ. Fonte: [13, p. 14].

Nas aplicações do PMQ, o experimento ilustrado na figura 2.2 não é realizado pelos estudantes; ele é apresentado pela figura no início do teste. As questões simulam situações de laboratório didático onde personagens, no papel de realizadores do experimento, apresentam suas ideias sobre como lidar com questões que surgem ao longo do desenvolvimento do trabalho. Há mais de dez questões disponíveis na monografia de Campbell et al. [13].

Nem sempre o experimento é apenas ilustrado. No trabalho de Laburú, Silva e Sales [14], a idealização do experimento da canaleta é materializada e proposta a um grupo de aproximadamente dez estudantes. Na atividade, eles lidam com tarefas que põem em prática questões do PMQ. A partir dos relatos dos estudantes o artigo apresenta uma análise das ideias utilizadas por eles durante as tarefas.

A figura 2.3 mostra um exemplo de questão do PMQ que traduzimos para o português e utilizamos como parte da avaliação da nossa proposta de ensino.

Os estudantes trabalham em grupos no experimento. A primeira tarefa deles é determinar d quando $h = 400$ mm. Um grupo lança a bolinha da altura $h = 400$ mm e, usando uma régua, eles medem que d é 436 mm.

A seguinte discussão ocorre então entre os estudantes.

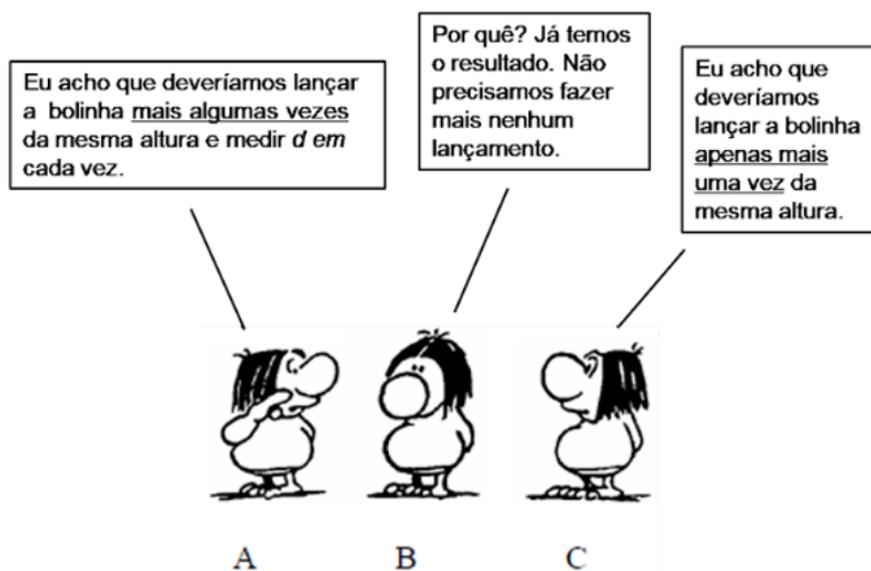


Figura 2.3: Exemplo de questão do PMQ que traduzimos para o português. Adaptado de [13, p. 15].

Para responder à avaliação, os estudantes geralmente devem escolher com qual das afirmativas feitas pelos personagens eles concordam mais. Após a escolha, deve ser dada uma justificativa por extenso.

O processo de avaliação das justificativas dos estudantes, explicitado na monografia de Campbell et al., se dá pela produção de categorias de justificativas mutuamente exclusivas para cada questão. Cada categoria possui uma ideia-chave que resume todas as respostas que foram atribuídas à categoria. Para validar essa classificação, os pesquisadores entrevistaram os estudantes avaliados para saber se as ideias-chave propostas para cada categoria estavam em acordo com o significado original de suas respostas. Em seguida, cada categoria foi classificada como pertencente ao raciocínio de ponto ou ao raciocínio de conjunto. Mesmo que o estudante marque que concorda com uma afirmação que remete ao raciocínio de conjunto, a categoria de sua jus-

tificativa pode indicar um raciocínio de ponto e vice-versa. A partir disso, os avaliadores puderam afirmar que tipo de raciocínio foi utilizado por cada estudante ao lidar com cada aspecto do processo de medição.

2.3 Nossa adaptação do PMQ

Cada questão do PMQ foi desenvolvida para avaliar um aspecto do processo de medição. Na versão do PMQ que aplicamos em nosso trabalho, escolhemos utilizar duas questões sobre coleta de dados (nossas questões 1 e 2), uma sobre sistematização de dados (questão 3) e duas sobre comparação entre conjuntos de dados (questões 4 e 5). A escolha das questões do nosso teste foi influenciada pela versão do PMQ utilizada por Volkwyn et al. [15]. Os autores utilizaram seis questões em comum em seus pré e pós-testes, das quais adotamos cinco, excluindo uma sobre análise de gráficos.

A nossa aplicação do PMQ foi realizada utilizando a plataforma Google Forms. Encaminhamos o link do formulário² aos estudantes que realizaram o teste. No formulário, as questões foram dispostas de forma que apenas após a resposta a uma questão a próxima seria liberada, seguindo a ordem apresentada por Volkwyn et al. [15]. Também orientamos os participantes a não retornar às questões após respondê-las. Na nossa versão da aplicação do PMQ, depois que o estudante seleciona a afirmativa com a qual mais concorda surge uma lista de justificativas possíveis para a escolha, relacionadas à alternativa tomada. Na figura 2.4 está a imagem que apareceria caso um aluno escolhesse a opção A para responder à questão 1 (mostrada na figura 2.3).

²<https://forms.gle/rmjpaVcPrLb8cX5Q8>

1-A

Eu acho que a bolinha deveria ser lançada mais algumas vezes da mesma altura, *
porque...

- com a prática faremos medidas melhores.
- devemos repetir as medidas até que elas fiquem próximas umas das outras.
- as novas medidas podem mostrar qual é o resultado correto.
- precisamos de mais medidas para achar a média.
- precisamos de mais medidas para achar a média e avaliar o grau de incerteza dos resultados.

Figura 2.4: Justificativas que aparecem após a escolha da afirmativa A na questão 1, mostrada na figura 2.3. As três primeiras representam o raciocínio de ponto; as duas últimas, o de conjunto.

Essas justificativas foram escolhidas para a nossa versão do questionário a partir da categorização disponível no apêndice da monografia de Campbell et al. [13]. Essa forma de aplicar o PMQ facilita o trabalho de determinar se os respondentes utilizaram raciocínio de ponto ou conjunto em suas respostas. Oferecendo as ideias-chave das principais categorias de justificativa disponíveis na literatura, eliminamos a necessidade da leitura individual das respostas e sua classificação em categorias. Essa disponibilização das ideias-chave para que os estudantes justifiquem a resposta escolhida, como alternativa para as respostas abertas, já foi utilizada com sucesso por Kung [16]. As questões incluídas na nossa versão do PMQ e as justificativas disponíveis para cada resposta estão no apêndice B. A classificação dessas justificativas segundo os paradigmas de ponto ou conjunto é dada no apêndice C.

Em alguns conjuntos de justificativas do nosso questionário, os estudantes poderiam escolher mais de uma opção entre as apresentadas. Nesses casos o tipo de raciocínio utilizado na questão foi considerado pela classificação da maioria de ideias-chave escolhidas. Na eventualidade de seleção da mesma quantidade de justificativas nos dois tipos de opção, o raciocínio utilizado na

questão foi classificado como misto.

O conceito de raciocínio misto que adotamos serve para lidar com a resposta dos estudantes a cada uma das questões. Isso é diferente do que Pollard et al. [10] chamam de raciocínio misto. Esse último descreve os estudantes que respondem a algumas questões do PMQ seguindo raciocínio de ponto e a outras seguindo o raciocínio de conjunto. O conceito de raciocínio misto no trabalho de Pollard et al. refere-se à forma como cada estudante responde a todo o questionário, enquanto o conceito de raciocínio misto que utilizamos analisa a resposta a cada questão individualmente.

Como exemplo de utilização do PMQ, a figura 2.5 apresenta resultados da análise de Volkwyn et al. no artigo “Impact of a conventional introductory laboratory course on the understanding of measurement” [15]. Na figura está a proporção de raciocínio de conjunto em um grupo de 53 estudantes universitários, antes e depois do curso de laboratório. A análise feita com o PMQ mostra como os estudantes lidam com o processo de medição em três aspectos principais: coleta de dados, sistematização dos dados e comparação entre conjuntos de dados (os mesmos que abordamos em nosso questionário). Para cada um desses três aspectos Volkwin et al. utilizam duas questões do PMQ.

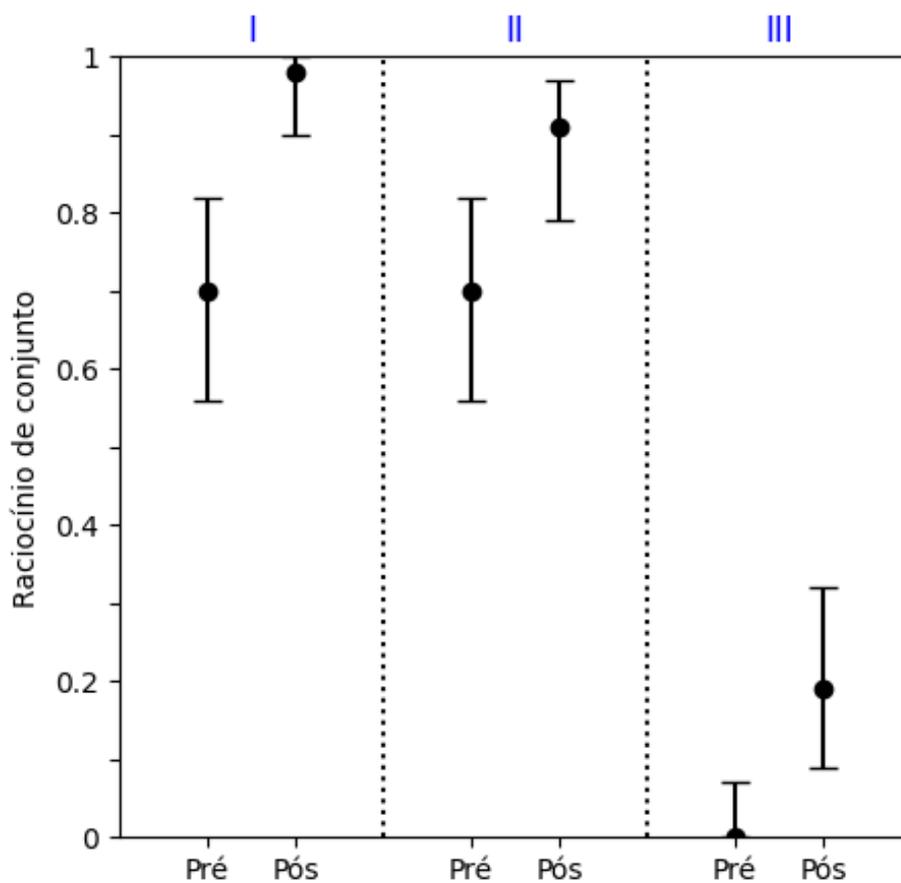


Figura 2.5: A proporção de raciocínios de conjunto em cada aspecto do processo de medição. A seção I corresponde a questões do PMQ sobre coleta de dados, a seção II a questões de sistematização dos dados e a seção III a questões de comparação entre conjuntos de dados. As barras indicam intervalos de confiança de 95%. Adaptado de [15].

No artigo, Volkwyn et al. encontram uma diferença estatisticamente significativa ($p < 0,05$) entre os resultados do pré-teste e do pós-teste para cada um dos três aspectos avaliados.

A nossa análise dos resultados do PMQ no pré-teste e pós-teste foi conduzida de maneira um pouco diferente. Preferimos comparar os resultados do pré e pós teste em cada questão, individualmente. Desta forma identificamos melhor os aspectos em que os estudantes tiveram pior desempenho,

de modo a aperfeiçoar a sequência didática ao longo das três aplicações que realizamos.

Como exemplo de uso da nossa versão do PMQ, a figura 2.6 mostra o efeito da instrução na disciplina Física Experimental I de uma universidade pública do estado do Rio de Janeiro. O gráfico apresenta a fração dos estudantes que utilizou o raciocínio de conjunto para responder cada questão. Pode-se notar que o curso produziu uma melhora estatisticamente significativa na compreensão dos estudantes sobre medidas e incertezas.

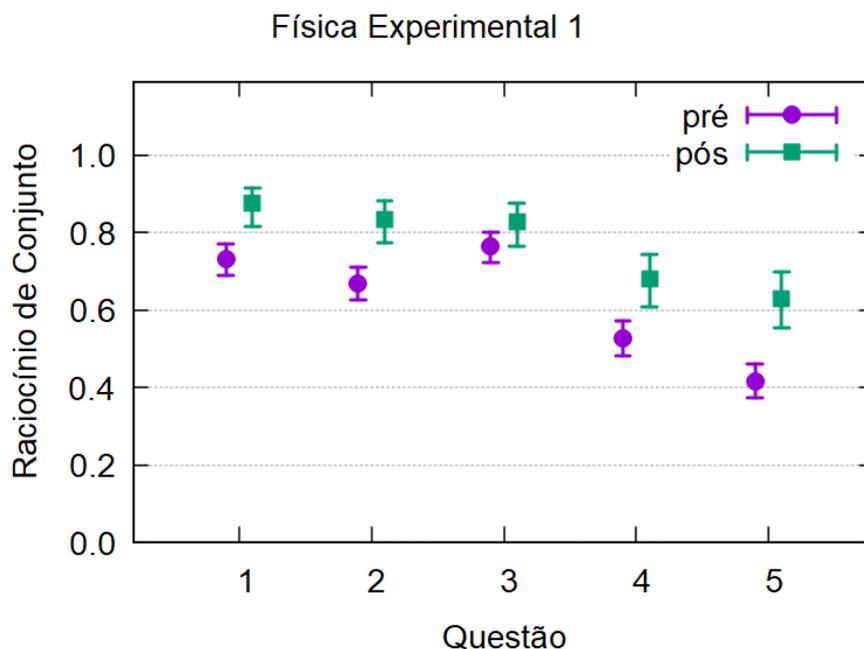


Figura 2.6: A proporção de raciocínios de conjunto no pré e pós-teste em cada questão. As barras representam intervalos de confiança de 95%.

No próximo capítulo apresentaremos nossa proposta para o ensino de medidas e incertezas, detalhando os conteúdos da sequência didática em sua forma final, correspondente à última aplicação em sala de aula. Discutiremos as motivações para a escolha de certas abordagens em detrimento de outras, além do funcionamento do aplicativo que desenvolvemos para a realização da atividade prática com os estudantes.

Capítulo 3

Uma proposta para o ensino de medida e incerteza

Nesta seção propomos uma sequência didática para o ensino de medidas, incertezas e seu uso em inferências e tomadas decisões. A proposta envolve discussões e atividades práticas a serem realizadas em uma sala de aula do ensino médio, tendo como base o material instrucional desenvolvido por nós. Um texto de apoio faz parte desse material e está disponível no apêndice A.

As primeiras aplicações da proposta foram realizadas em 2023, em turmas do segundo ano do ensino médio e em um clube de ciências de uma escola pública do Rio de Janeiro. Essas aplicações iniciais nos permitiram avaliar quais aspectos da sequência didática tinham sido bem sucedidos e quais deveriam ser aperfeiçoados. A partir dessa análise elaboramos o texto de apoio, que aprofunda os temas trabalhados nas atividades, inclusive detalhando as formulações matemáticas apresentadas aos alunos. Esse texto foi utilizado na segunda aplicação da proposta didática, realizada em 2024 em turmas do segundo ano da mesma escola.

A sequência didática está baseada na compreensão do processo de medição através do paradigma de conjunto, com ênfase na relação entre medidas e incerteza nos processos de coleta, análise e comparação de conjuntos de dados. Adotamos como exemplo de experimento a medição do tempo de reação a estímulos visuais. Utilizamos um aplicativo desenvolvido por nós

como meio de obter as medidas e apresentar algumas de suas características conjuntas, como média, desvio-padrão e histograma. O experimento é apresentado passo a passo, como forma de discutir aspectos inerentes ao processo de medição que poderiam passar despercebidos dentro da perspectiva de um raciocínio de ponto. Após abordarmos a obtenção de medidas e a sistematização delas por meio da média e de um histograma, discutimos como caracterizar numericamente a dispersão do conjunto de dados e o grau de confiança (ou a incerteza) que temos na representação desse conjunto pela sua média. Por fim, introduzimos algumas “regras de bolso” para a comparação entre conjuntos de medidas. O texto de apoio sempre utiliza dados vindos da medição do tempo de reação com o aplicativo, que pode ser usado pelos alunos tanto na sala de aula como em casa.

Nas próximas seções deste capítulo discutiremos os conteúdos que fazem parte da proposta de ensino, apresentando-os na ordem em que surgem na sequência didática. Apresentaremos também uma análise estatística que justifica as nossas “regras de bolso” para a comparação de conjuntos de medidas, embora essa justificativa não faça parte da sequência didática.

3.1 Tempo de reação

A escolha da medição do tempo de reação para estruturar a proposta didática está relacionada à ênfase que demos ao aspecto estatístico das incertezas de medição. Segundo o “Guia para a Expressão de Incerteza de Medição” (GUM) [5], as incertezas podem ser divididas em dois tipos: o tipo A, dado “pela avaliação estatística de séries de observações” [5, p. 3]; e o tipo B, dado “por outros meios que não a análise estatística de séries de observações” [5, p. 3]. Existem propostas de ensino que também consideram as incertezas do tipo B, como nos artigos de Buffler et al. [17] e Allie et al. [18] que discutem estimativas da incerteza de uma única medida do mensurando. Como desejávamos que os estudantes desenvolvessem um raciocínio de conjunto em sua perspectiva sobre o processo de medição, escolhemos lidar apenas com medições que envolvem a tomada de várias medidas.

Existem formas de realizar a medição do tempo de reação a estímulos vi-

suais sem a utilização de meios eletrônicos. A mais tradicional provavelmente é aquela em que uma pessoa (vamos chamar de medidor) segura uma régua entre o polegar e dedo indicador de outra pessoa (vamos chamar de reativo), que terá o tempo de reação medido. O medidor segura a régua de forma que a marca de zero centímetros é posicionada entre os dedos do reativo, estando o resto dela acima e alinhado na direção vertical. O medidor escolhe algum instante para largar a régua, que cairá verticalmente entre os dedos do reativo, que tentará segurá-la o mais rápido possível. Com a régua segura entre os dedos do reativo, pode-se utilizar a distância percorrida por ela em queda para medir quanto tempo foi necessário para a reação de segurá-la. Esse método para abordar o processo de medição é utilizado, por exemplo, no artigo de Kung [16].

O uso de um aplicativo para a medição do tempo de reação dos estudantes torna a coleta de medidas muito mais prática e rápida. O cálculo do tempo de reação a partir da distância de queda não é mais necessário, a sequência de medidas é automaticamente registrada e dezenas de medições podem ser realizadas em poucos minutos. Essa é uma característica essencial à nossa proposta de ensino, baseada nos aspectos estatísticos de um número razoável de medidas. O emprego desse método para o ensino do processo de medição também foi adotado por Pigosso e Heidemann [19], utilizando um site disponível na internet. Como desejávamos um programa adaptado à nossa proposta de ensino, confeccionamos nosso próprio aplicativo, pois, além da praticidade e rapidez no processo de obtenção de várias medidas, ele produz um histograma dos resultados e gera as descrições estatísticas dos dados que utilizaremos em nossas atividades práticas.

Como a resolução das medidas do tempo de reação feitas pelo nosso aplicativo está na casa dos milésimos de segundo, a flutuação entre cada medida tomada por uma mesma pessoa torna-se evidente. Em todas as utilizações do instrumento, poucas foram as vezes em que duas medidas consecutivas resultaram no mesmo valor. Esse caráter flutuante do mensurando favorece a argumentação de que é necessário realizar várias medidas para que seja possível estabelecer um valor que o represente. Segundo Lubben et al. [2] e Campbell et al. [13], os estudantes tendem a perceber de maneira intui-

tiva a necessidade de várias medidas para mensurar grandezas como tempo e temperatura, por conta de sua variabilidade perceptível em medições por métodos usuais.

3.2 Aplicativo, coleta de dados e histograma

O aplicativo, disponível na internet¹, funciona tanto em computadores quanto em tablets e smartphones. A seguir está a descrição do seu funcionamento, de forma semelhante à disponível no texto de apoio. É com a apresentação do aplicativo que tem início a discussão sobre a necessidade de se realizar várias medidas para se determinar o valor de um mensurando.

Acessando o link disponibilizado, pode-se observar uma área cinza contendo o botão “Fazer medida”, como na figura 3.1.



Figura 3.1: Aspecto inicial do aplicativo.

Quando esse botão é acionado, surge a mensagem “Clique na tela quando a cor mudar”. Após um intervalo escolhido ao acaso pelo aplicativo, a área cinza muda de cor, tornando-se laranja. O aspecto do aplicativo antes e logo depois desse intervalo está na figura 3.2.

¹http://pef.if.ufrj.br/producao_academica/dissertacoes/2024_Caio_Ferreira/tempo-reacao/tempo.html



Figura 3.2: Aplicativo iniciado.

No instante da mudança de cor, inicia-se uma contagem de tempo, que será interrompida por um toque na tela do celular ou clique do mouse na área laranja. O tempo entre a mudança de cor e o toque ou clique é o tempo de reação da pessoa, que será escrito na tela como se vê na medida mostrada na figura 3.3. A unidade de tempo utilizada é o segundo.

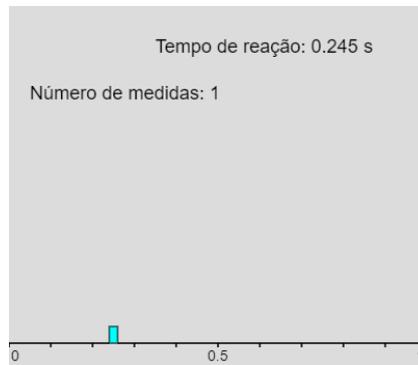


Figura 3.3: Medida do tempo de reação realizada com o aplicativo.

Além do resultado indicado numericamente no alto da tela, uma caixa de 0,02 s de largura é posicionada em um eixo representando o tempo de reação, que vai de 0 a 1 s com marcações a cada 0,1 s. O desenho dessa caixa dá início à construção de um histograma. A partir da segunda medida, o aplicativo passa a apresentar a média e o desvio-padrão dos dados obtidos (a serem discutidos mais à frente) e coloca mais caixas sobre os “bins” que contém os os valores medidos, como se vê na figura 3.4.

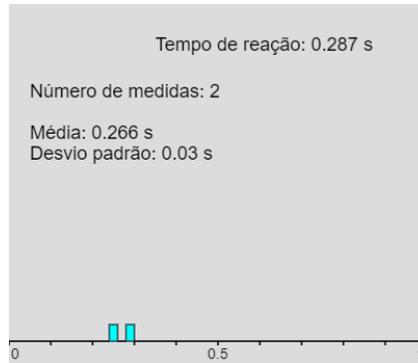


Figura 3.4: Duas medidas coletadas.

O resultado apresentado na figura 3.4, mostra uma situação típica: a segunda medida é diferente da primeira. Dada a divergência, é natural propor que a mesma pessoa realize mais medições para entender melhor o que está acontecendo com os resultados. Um conjunto de dez medidas tomadas com esse propósito está mostrado na tabela 3.1. Os mesmos resultados são apresentados no texto de apoio e um número semelhante de dados é coletado pelos estudantes durante a atividade prática.

Medida	Valor (s)
1	0,245
2	0,287
3	0,266
4	0,264
5	0,257
6	0,274
7	0,236
8	0,284
9	0,276
10	0,260

Tabela 3.1: Tabulação de um conjunto de dez medidas.

Registrar as medidas em uma tabela não facilita muito a análise de sua variação. Por isso o aplicativo apresenta um histograma das medidas já

realizadas. A figura 3.5 mostra esse histograma para as dez medidas listadas na tabela 3.1. A sequência didática prevê uma discussão mais detalhada acerca da construção do histograma, que pode ser encontrada no texto de apoio.

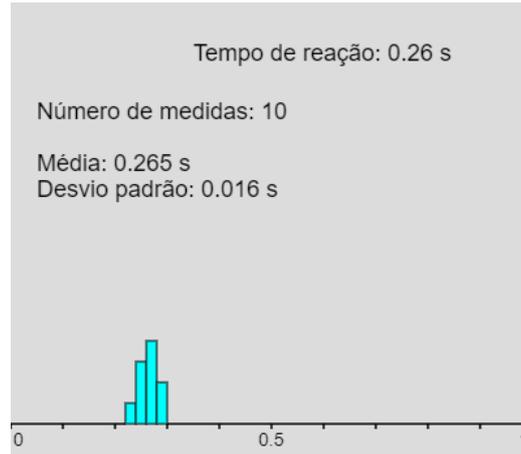


Figura 3.5: Histograma das dez medidas da tabela 3.1.

3.3 A média de um conjunto de medidas

Desde a apresentação da segunda medida, fica demonstrada a insuficiência da representação do tempo de reação de uma pessoa a partir de uma medida única. Por conta da variação do valor em cada medida, propomos que a média dos resultados encontrados deve fornecer uma representação mais útil desse tempo de reação, pois considera toda a informação obtida até então. A média representa razoavelmente bem o valor típico das medidas, mesmo que nenhuma medição tenha encontrado esse valor. Isso pode ser visto na figura 3.6 que mostra o valor médio das dez medidas da tabela 3.1 juntamente com seu histograma. O valor médio dá uma boa indicação sobre os tempos de reação encontrados.

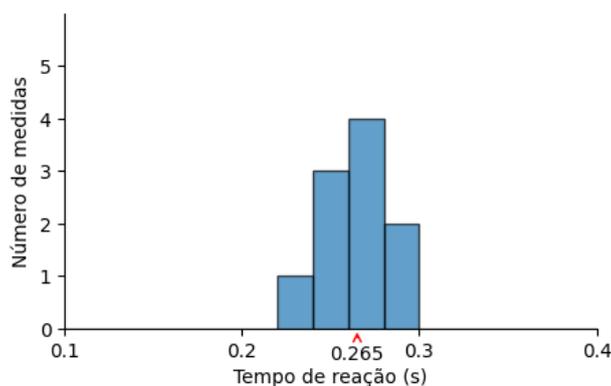


Figura 3.6: Histograma das dez medidas da tabela 3.1. A seta indica o valor médio (0.265 s).

A média aritmética é uma de diversas opções para representar o valor de um mensurando partindo de um conjunto de medidas. Durante o desenvolvimento da sequência didática, consideramos também a possibilidade de adotar a mediana das medidas como uma representação do conjunto. O uso da mediana é comum pois ela é menos sensível a valores extremos anômalos do que a média. Contudo, optamos por não abordar o conceito de mediana, pois esse vem associado a outros elementos (intervalos inter-quartil, por exemplo) que excederiam o tempo disponível para a apresentação da sequência didática.

Outra alternativa para a representação de um valor típico das medidas seria a moda. Em alguns trabalhos [4, 9, 13], processos que remetem à obtenção da moda de um conjunto são apresentados como uma proposta comum em estudantes que utilizam o raciocínio de ponto, mostrando um possível risco trazido pelo seu uso no ensino em nível introdutório. Esses trabalhos mostram que estudantes justificam a obtenção de varias medidas pela necessidade de encontrar um valor recorrente e que esse é o valor verdadeiro do mensurando. Outra versão desse argumento é que mais de uma medida é necessária para que um valor repetido possa ser encontrado, confirmando o valor verdadeiro. Por isso, sequer consideramos a utilização da moda em nossa proposta de ensino.

3.4 Dispersão das medidas e desvio-padrão

Na figura 3.7 são apresentados dois histogramas feitos com o aplicativo, mostrando dez medidas do tempo de reação de duas pessoas, *A* e *B* (à esquerda e direita, respectivamente). Em uma aula seguindo a sequência de ensino proposta, esses histogramas podem ser produzidos por dois estudantes.

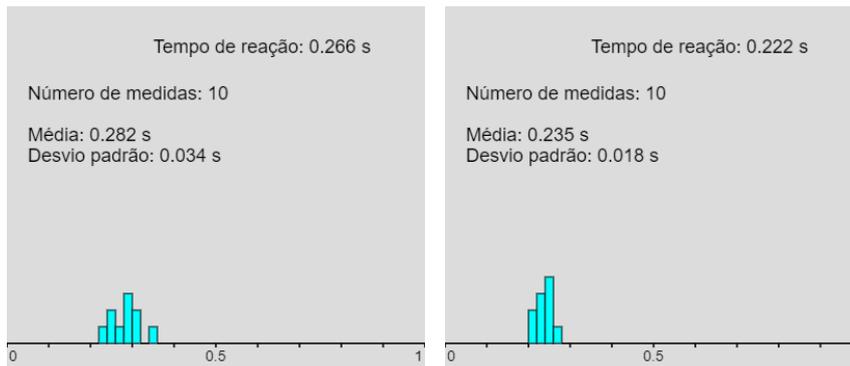


Figura 3.7: Histograma da pessoa *A* à esquerda e da pessoa *B* à direita.

A partir da imagem, fica claro que os dois conjuntos de medidas, possuem diferentes dispersões. Visualmente, pode-se notar que o histograma de *B* é mais estreito que o de *A*. Para tornar quantitativa essa noção de dispersão, o conceito de desvio-padrão é introduzido, estabelecendo um valor numérico (σ) para o grau de espalhamento das medidas em torno da média.

A sequência didática enfatiza que o desvio-padrão dá uma indicação da distância típica das medidas em relação à média. O texto de apoio constrói a fórmula matemática do desvio-padrão, partindo da diferença entre cada medida do tempo de reação e a média do conjunto. A fórmula final apresentada é o desvio-padrão amostral, caracterizado pela presença do número de medidas n subtraído por 1 em seu denominador:

$$\sigma = \sqrt{\frac{(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \dots + (t_n - \bar{t})^2}{n - 1}}$$

A escolha do desvio-padrão amostral em detrimento do desvio-padrão populacional, que possui apenas n em seu denominador, se dá por conta da pequena quantidade de medidas que normalmente coletamos nos experimen-

tos apresentados. Utilizamos esse fator de correção (a “correção de Bessel”) para que o desvio-padrão calculado não subestime a dispersão dos dados. A demonstração da correção de Bessel é complicada [20, p. 76]. Na apresentação da fórmula preferimos justificar a subtração por 1 no denominador como um indicador de que não é possível calcular o desvio-padrão de apenas uma medida. Apesar de ser verdade, essa não é a sua motivação real. Esse fator é necessário para corrigir um viés sistemático no cálculo sem a correção. Como a sequência didática é voltada para estudantes do ensino médio, optamos por não estender a discussão sobre fator de correção.

Após a apresentação da fórmula para o cálculo do desvio-padrão, um conjunto de medidas, como o da tabela 3.1, é apresentado em um novo histograma com a delimitação do intervalo de $\bar{t} - \sigma$ até $\bar{t} + \sigma$, como exposto na figura 3.8. Chamamos a atenção dos alunos para o fato desse intervalo conter boa parte das medidas obtidas (seis das dez, no caso ilustrado).

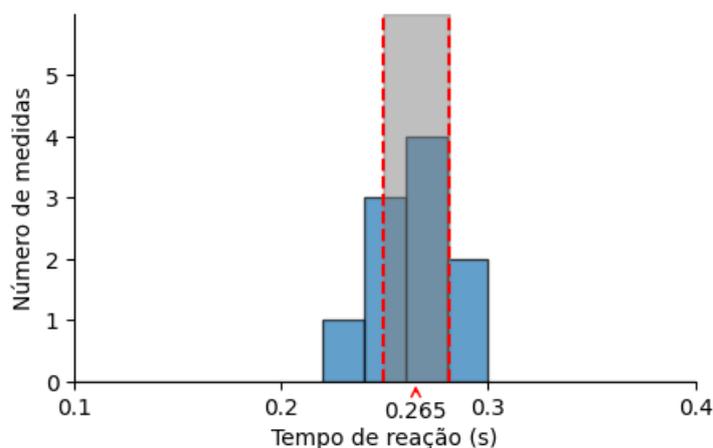


Figura 3.8: O intervalo de um desvio-padrão em torno da média indicado na figura contém seis das dez medidas do conjunto da tabela 3.1.

Numa situação em que os dados se distribuem de acordo com uma gaussiana, caracterizada por um histograma em forma de sino, aproximadamente 68% das medidas serão encontradas em até um desvio-padrão em torno da média. Na região que vai até 2σ estarão contidas aproximadamente 95% das medidas [21]. Entretanto, a distribuição dos dados do tempo de reação não

é simétrica, não seguindo o padrão gaussiano. Portanto, a afirmação de 68% das medidas estarem entre $\bar{t} - \sigma$ até $\bar{t} + \sigma$ não é válida nesse caso. Todavia, esse intervalo ainda contém uma grande proporção dos dados, sendo um bom indicador da região do eixo de tempo em que boa parte das medidas será encontrada, caracterizando a dispersão do conjunto.

Uma alternativa simplificada à utilização de histogramas, contendo a localização e dispersão do conjunto de medidas, é a representação gráfica pela média e barra de erro. Partindo da visualização deste tipo de gráfico, pode-se inferir sobre a região do eixo em que provavelmente surgirão novas medidas. Na figura 3.9 os dados de tempo de reação dos indivíduos *A* e *B*, apresentados anteriormente em histogramas, estão descritos dessa maneira. Os pontos representam a média dos tempos de reação de cada indivíduo e as linhas (barras de erro) marcam intervalos de $\bar{t} - \sigma$ a $\bar{t} + \sigma$. Essas barras permitem prever a faixa de valores na qual uma próxima medida provavelmente será encontrada.

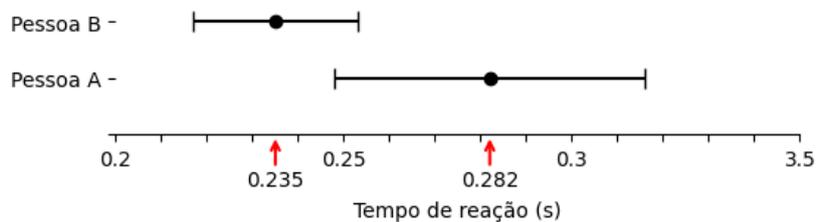


Figura 3.9: Representação gráfica da média e desvio-padrão do tempo de reação das pessoas *A* e *B*. As barras de erro representam o intervalo de $\bar{t} - \sigma$ a $\bar{t} + \sigma$ de cada conjunto.

A figura 3.9 permite responder à pergunta: “se as mesmas duas pessoas medirem mais uma vez seus tempos de reação, é possível que para essa medida extra o tempo de *A* seja menor que o de *B*?”. A resposta que pode ser dada olhando para o gráfico é que sim, apesar de não muito provável, dada a pequena sobreposição entre os intervalos.

A discussão trazida pelo desvio-padrão demonstra aos estudantes como a

média não é suficiente para descrever sozinha um conjunto de dados, sendo necessário estabelecer uma noção de como as medidas estão dispostas em torno dessa média. O desvio-padrão fornece essa noção, indicando quão precisamente podemos prever o resultado de uma próxima medição.

3.5 Flutuação da média e desvio-padrão da média

Não apenas as medidas individuais possuem incerteza. A média de um conjunto de medidas também tem sua incerteza; se realizarmos dois conjuntos de medidas, suas médias provavelmente serão diferentes. Isso está ilustrado na figura 3.10, onde estão dois histogramas correspondentes a dois conjuntos de dez medidas realizados pela mesma pessoa, com indicação de suas médias no eixo horizontal. Apesar de não ser surpreendente que as medidas sejam distintas, a comparação dos dois gráficos explicita que nem mesmo as médias são iguais.

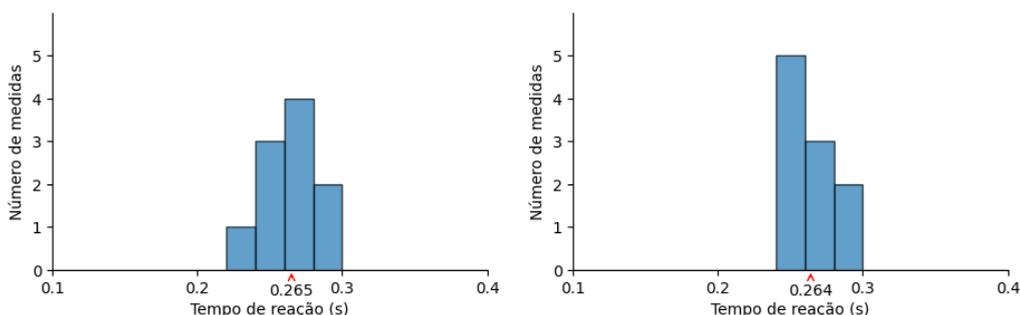


Figura 3.10: Dois grupos de dez dados obtidos pela mesma pessoa. As médias de cada grupo (0,265 s e 0,264 s) estão indicadas pelas setas.

Para analisar a variação das médias de medidas do mesmo mensurando, apresentamos na tabela 3.2 as médias de dez conjuntos de medidas do tempo de reação realizadas pela mesma pessoa. Cada conjunto possui dez medidas e os dois primeiros conjuntos da tabela são os representados na figura 3.10.

Conjunto	Média (s)
1	0,265
2	0,264
3	0,262
4	0,261
5	0,275
6	0,267
7	0,258
8	0,260
9	0,264
10	0,265

Tabela 3.2: Médias dos conjuntos de dez medidas. Os histogramas dos conjuntos 1 e 2 são aqueles mostrados na figura 3.10.

Na figura 3.11, essas médias são apresentadas em um histograma com a mesma escala e com *bins* de mesma largura dos histogramas de medidas presentes na figura 3.10. É fácil ver que a dispersão das médias é menor que a dispersão das medidas individuais.



Figura 3.11: Histograma das médias presentes na tabela 3.2.

Para quantificar a dispersão das médias, é usual introduzir o conceito de desvio-padrão da média (σ_M), também chamado de erro-padrão. Na sequência didática é ressaltado que o aumento do número n de medidas em um

conjunto deve fazer com que o valor da média seja menos incerto, já que isso leva a um aumento da informação que se tem sobre o mensurando. Matematicamente, a diminuição da incerteza da média com o aumento do número de medidas é dada pela fórmula

$$\sigma_M = \frac{\sigma}{\sqrt{n}},$$

onde σ é o desvio-padrão das medidas individuais, como antes. Por conta de sua relativa complexidade, na sequência didática essa fórmula não é demonstrada em tanto detalhe quanto foi o desvio-padrão. Uma demonstração simplificada pode se encontrada, por exemplo, em [20, p. 51].

Como exemplo, o erro-padrão do conjunto de dez medidas da tabela 3.1 é $\sigma_M = 0,005$ s. Esse valor indica que a média de um novo conjunto de dez medidas, obtidas pela mesma pessoa, provavelmente estará no intervalo que vai de $\bar{t} - \sigma_M = 0,260$ s até $\bar{t} + \sigma_M = 0,270$ s. Olhando para a tabela 3.2, podemos ver que oito das dez médias estão contidas neste intervalo.

Para verificar como o erro-padrão diminui com o número de medidas n , apresentamos na figura 3.12 o intervalo de um erro-padrão em torno da média para três conjuntos de medidas realizadas pela mesma pessoa com diferentes valores de n . O tempo de reação médio varia, como esperado, e as médias de conjuntos com mais medidas estão contidas dentro do intervalo de incerteza dos conjuntos menores.

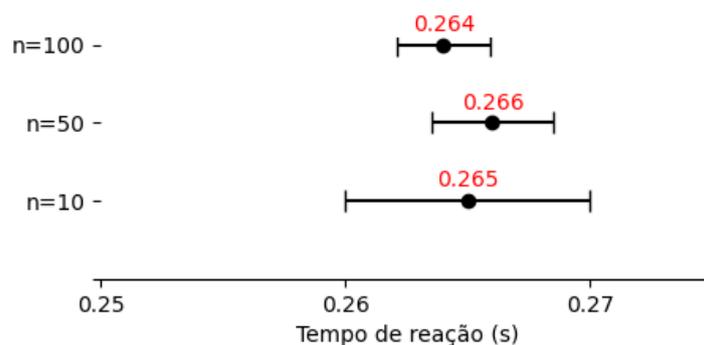


Figura 3.12: O tempo de reação médio de uma pessoa obtido com diferentes números de medidas. As barras representam o erro-padrão.

Embora isto não seja explicitamente abordado na sequência didática, é importante ressaltar que médias de conjuntos de n medidas têm uma distribuição gaussiana, para n suficientemente grande. A explicação é dada pelo teorema do limite central [20–22], que afirma que a distribuição das médias de um conjunto de dados converge para uma distribuição gaussiana, mesmo que as medidas originais não estejam distribuídas dessa maneira. Portanto, é possível estimar que a probabilidade de uma nova média de n medidas ser encontrada no intervalo de $\bar{t} - \sigma_M$ e $\bar{t} + \sigma_M$ é aproximadamente 68%.

3.6 Comparação entre grupos de dados

Com o conceito de erro-padrão podemos retornar à comparação entre os tempos de reação dos indivíduos A e B da seção 3.4 (figuras 3.7 e 3.9). O gráfico da figura 3.13 compara a média e erro-padrão do tempo de reação dessas pessoas. Essa figura permite responder à seguinte pergunta: “olhando para as informações no gráfico, é possível afirmar que uma das duas pessoas tem, em média, reação mais rápida?”

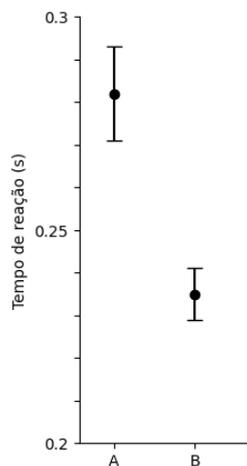


Figura 3.13: Tempo médio de reação dos indivíduos A e B . As barras de erro representam o erro-padrão.

A resposta intuitiva e, nesse caso, correta, é que a pessoa B reage mais rapidamente, em média. Gráficos como esse tornam possível realizar com-

parações partindo apenas de uma análise visual. Para isso, introduzimos na sequência didática a noção de “vão” entre os dados, que é a distância entre os limites mais próximos das barras de erro-padrão. Essa distância será fundamental para a formulação de nossa proposta de comparação entre grupos de dados.

Numa situação diferente da anterior, o tempo de reação médio de duas outras pessoas, C e D , é mostrado na figura 3.14. Embora C apresente um tempo de reação médio maior que D , diferentemente do gráfico anterior não há vão entre as barras de erro.

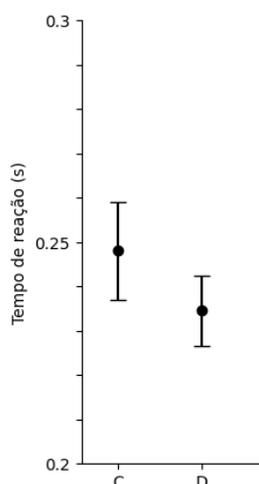


Figura 3.14: Tempo médio de reação dos indivíduos C e D . As barras de erro representam o erro-padrão.

Novamente, cabe a pergunta: “olhando para as informações no gráfico, é possível afirmar que uma das duas pessoas tem, em média, reação mais rápida?”. Nesse caso, a sobreposição de barras indica que, se novos grupos de medidas fossem feitos, as médias de C e D poderiam ter a ordem de valores invertida. Por isso é impossível garantir que os tempos de reação médios das duas pessoas são diferentes.

Exemplos como esses nos levaram a formular, na sequência didática, “regras de bolso” para a realização de comparações. Essas regras, baseadas no vão entre as barras de erro-padrão, são:

- Quando o vão entre as barras tem o comprimento de aproximadamente um erro-padrão, como na figura 3.15, **há uma probabilidade razoável de existir uma diferença entre as quantidades representadas**.

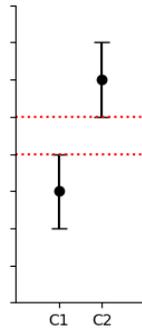


Figura 3.15: Situação em que há uma probabilidade razoável de existir uma diferença entre as quantidades representadas. A distância entre cada traço no eixo vertical equivale a um erro-padrão.

- Quando há uma interseção entre as barras, ou o vão entre elas é bem menor que o comprimento de um erro-padrão, como na figura 3.16, **os dados não permitem afirmar que há uma diferença entre as quantidades que comparamos** (o que não significa dizer que são necessariamente iguais).

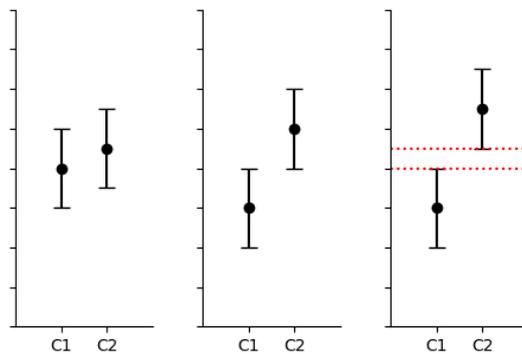


Figura 3.16: Situações em que os dados não permitem afirmar que há uma diferença entre as quantidades representadas pelas médias.

- Quando o vão é bem maior que um erro-padrão, como na figura 3.17, **é praticamente certo que haja uma diferença entre as quantidades que estamos comparando.**

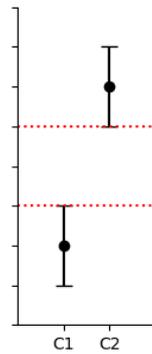


Figura 3.17: Situação em que é praticamente certo haver uma diferença entre as quantidades que estão sendo comparadas. O vão mostrado corresponde a dois erros-padrão.

- Numa situação em que lidamos com médias que possuem erros-padrão sensivelmente diferentes, aplicamos as regras de bolso anteriores comparando o vão entre as barras ao erro-padrão de maior valor.

Essas regras de bolso são baseadas nas categorias utilizadas por Holmes et al. [23]. Calculando a quantidade t' dada por

$$t' = \frac{\bar{A} - \bar{B}}{\sqrt{\sigma_A^2 + \sigma_B^2}},$$

onde \bar{A} e \bar{B} são as médias e σ_A e σ_B são os respectivos erros-padrão dos grupos de dados, Holmes et al. propõem as regras de comparação apresentadas na tabela 3.4.

valor t'	Interpretação
$0 < t' < 1$	Não é possível afirmar que há uma diferença entre os dois mensurandos.
$1 < t' < 3$	Probabilidade razoável de existir uma diferença entre os dois mensurandos.
$3 < t' $	Praticamente certo que há uma diferença entre os mensurandos.

Tabela 3.4: Regras de comparação de Holmes et al. [23].

O aspecto gráfico de nossas regras é inspirado na proposta de inferência visual de Cumming e Finch [24], na qual o vão entre as barras de erro é utilizado para apontar a compatibilidade entre medições, partindo da análise do valor-p.² A figura 3.18 contém a comparação proposta por Cumming e Finch para relacionar o vão com o valor-p da diferença de médias.

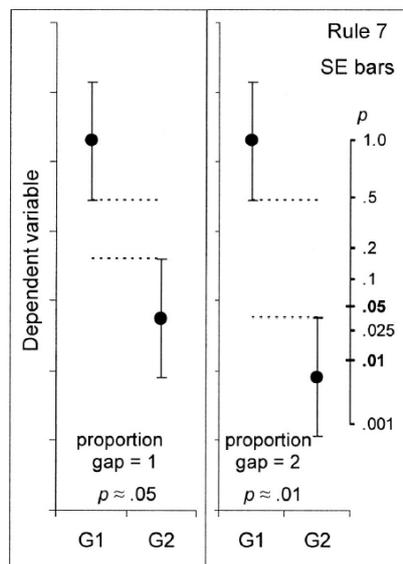


Figura 3.18: O valor-p como função do vão entre as barras de erro para erros-padrão iguais. Esse valor é dado pela projeção da média mais baixa na escala localizada à direita dos gráficos [24, p. 178].

²O valor-p mede a probabilidade da diferença entre as médias ser fruto apenas de uma flutuação estatística.

Para esclarecer o que chamamos de “pouco provável”, “probabilidade razoável” e “praticamente certo” nas nossas regras de bolso, realizamos uma análise do valor-p para diferentes situações experimentais. Essa análise é apresentada na próxima seção.

3.7 Justificativa para as regras de bolso

A relação entre o vão das barras de erro e a prescrição das regras de bolso está baseada nos valores-p calculados na condição de cada regra³. Para esse cálculo utilizamos o teste de Welch [21, p. 190], indicado para comparar conjuntos que podem ter erros-padrão diferentes (o teste de Student, em princípio, só se aplica a casos em que os erros-padrão são iguais).

A validade da primeira regra pode ser avaliada com auxílio do gráfico na figura 3.19. A curva indica o valor-p das médias de dois conjuntos A e B em função do número de medidas n , suposto igual para os dois conjuntos de dados ($n_a = n_b = n$). Esse número parte de um valor mínimo de $n = 5$, um valor razoável para que considerações estatísticas comecem a ter alguma significância. O gráfico corresponde às condições da primeira regra, na qual o vão é igual a um erro-padrão e os dois conjuntos têm o mesmo erro-padrão.

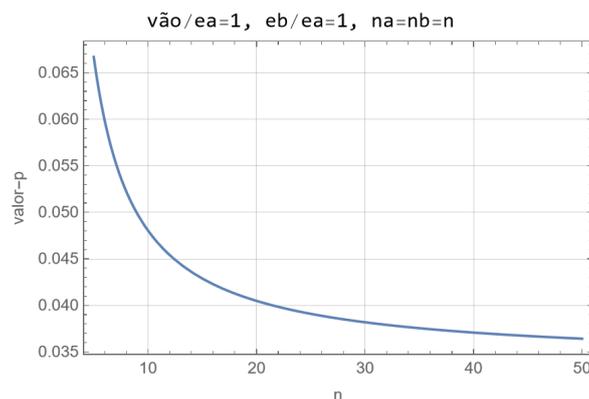


Figura 3.19: O valor-p para condições típicas da primeira regra (vão igual ao erro-padrão). O número de medidas dos dois conjuntos, n_a e n_b , é suposto igual. Os erros-padrão dos conjuntos, e_a e e_b , também têm valores iguais.

³Uma discussão sucinta sobre o valor-p e sua utilização está disponível em [25].

Observando o gráfico da figura 3.19, vemos que, mesmo na situação em que há apenas 5 medidas em cada conjunto, o valor-p está abaixo de 0,07. O valor máximo tradicionalmente utilizado para afirmar que existe uma diferença estatisticamente significativa é $p = 0,05$. A curva da figura 3.19 mostra que para $n \geq 8$ o valor-p está abaixo desse corte. Mesmo para $n = 5$, o valor-p não está muito acima do corte convencional. Vemos também que o valor-p tende a se estabilizar em aproximadamente 0,035 quando o número de medidas é grande. Desta forma, na afirmativa “há uma probabilidade razoável de existir uma diferença entre as quantidades representados”, essa probabilidade “razoável” é algo em torno de 95%.

A justificativa da segunda regra é encontrada na figura 3.20, semelhante à figura 3.19 com a única diferença que agora o gráfico indica o valor-p na condição em que o vão entre as barras de erro-padrão é nulo.

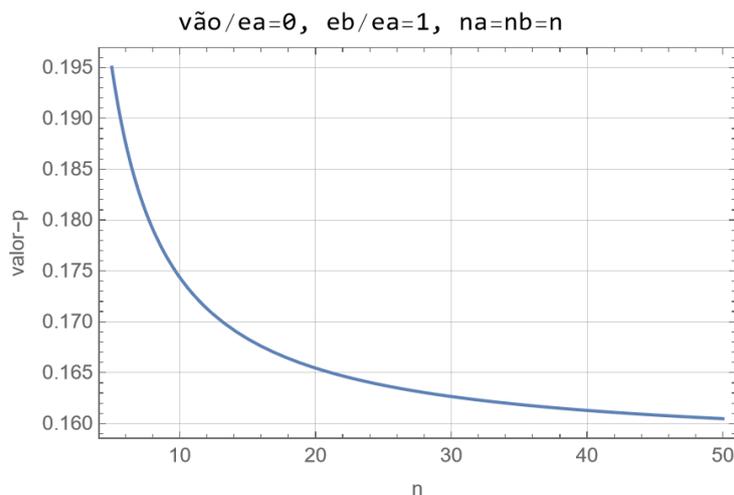


Figura 3.20: O valor-p para condições típicas da segunda regra (vão nulo). Como na figura anterior, $na = nb$ e $ea = eb$.

O gráfico da figura 3.20 mostra que, quando não há vão, o valor-p permanece bem acima do corte $p = 0,05$ convencionalmente utilizado para afirmar uma diferença estatisticamente significativa. Para um número muito grande de medidas, a situação não deve mudar, pois o valor-p tende a se estabilizar em um valor próximo a 0,16. Assim, vemos que a afirmativa da segunda

regra, “os dados não permitem afirmar que há uma diferença entre as quantidades que comparamos”, significa que há uma probabilidade de erro da ordem de 15% ou mais ao se afirmar que a diferença encontrada entre as médias garante que elas não são iguais.

Para justificar a terceira regra utilizamos a figura 3.21, semelhante às anteriores, em que o gráfico indica o valor-p para um $v\tilde{a}o$ igual a dois erros-padrão.

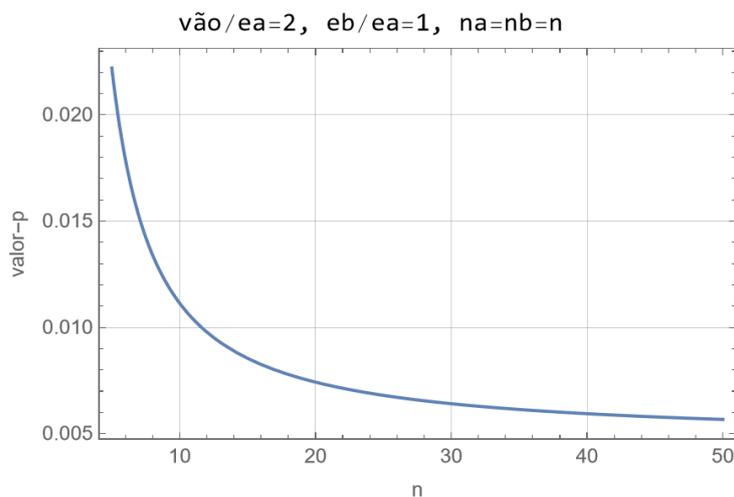


Figura 3.21: O valor-p para condições típicas da terceira regra ($v\tilde{a}o$ igual a dois erros-padrão). Como na figura anterior, $na = nb$ e $ea = eb$.

O gráfico da figura 3.21 indica que, mesmo na situação em que há apenas 5 medidas em cada conjunto, o valor-p está bem abaixo de 0,05. Além disso, para valores maiores de n o valor-p tende a aproximadamente 0,005. Portanto, vemos que a afirmativa da terceira regra, “é praticamente certo que haja uma diferença entre as quantidades que estamos comparando”, pode ser lida como a existência de uma probabilidade maior que cerca de 98% das médias serem diferentes.

A última regra de bolso é um complemento para as três regras anteriores, no caso em que os erros-padrão dos conjuntos comparados são notavelmente distintos. Nessa situação, a quarta regra recomenda que o $v\tilde{a}o$ seja comparado ao maior erro-padrão. Podemos analisar essa situação, com o auxílio dos

gráficos presentes nas figuras 3.22, 3.23 e 3.24. Neles, está indicado o valor-p em função da razão eb/ea entre os erros-padrão dos conjuntos A e B . Nessas figuras, ea é tomado como o maior dos erros-padrão, de modo que nos gráficos $0 \leq eb/ea \leq 1$. Cada conjunto possui cinco medidas, pois esse é o número mínimo razoável para a aplicação das regras de bolso anteriores. A figura 3.22, mostra o valor-p quando o vão tem o tamanho do maior erro-padrão. Vemos que a primeira regra (complementada pela quarta) pode ser aplicada mesmo que os erros-padrão estejam em uma razão de um para dois.

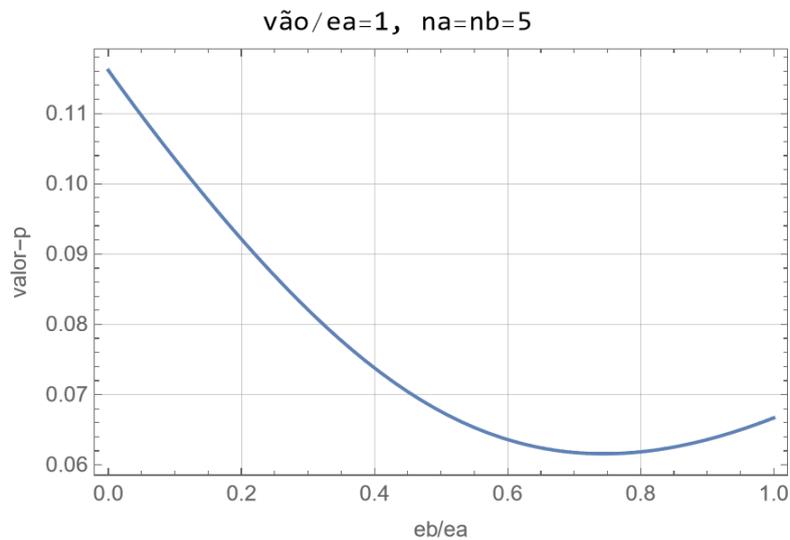


Figura 3.22: Valor-p em função de ea/eb , para o vão igual ao maior erro-padrão e $n = 5$.

A figura 3.23, semelhante à figura anterior, mostra o valor-p em função de eb/ea quando o vão é nulo. O gráfico mostra que a segunda regra é válida para qualquer razão entre os erros-padrão, ou seja, se o vão for muito pequeno ou houver uma superposição entre as barras de erro, não podemos afirmar que há uma diferença entre as médias.

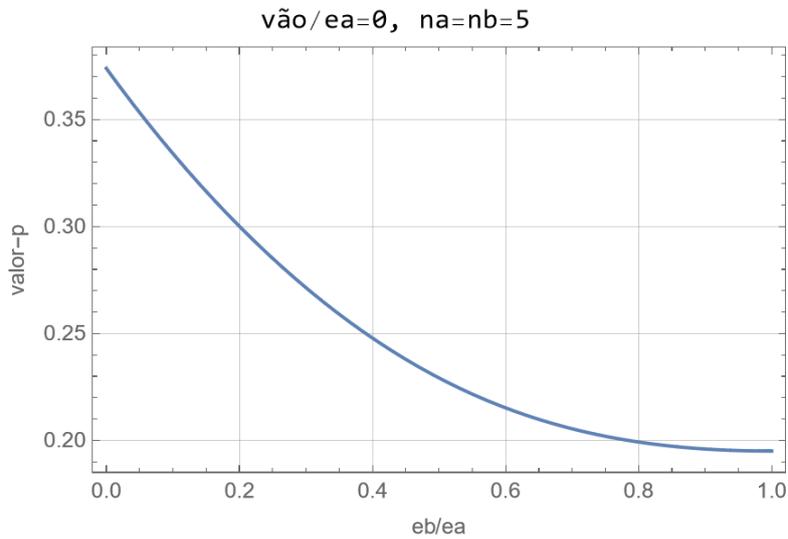


Figura 3.23: Valor-p em função de ea/eb , para o ṽo nulo e $n = 5$.

A figura 3.24 mostra o caso em que o ṽo vale duas vezes o maior erro-padrão. De forma semelhante ao que vimos para a primeira regra, a segunda regra pode ser aplicada mesmo quando os erros-padrão estão em uma razão de um para dois.

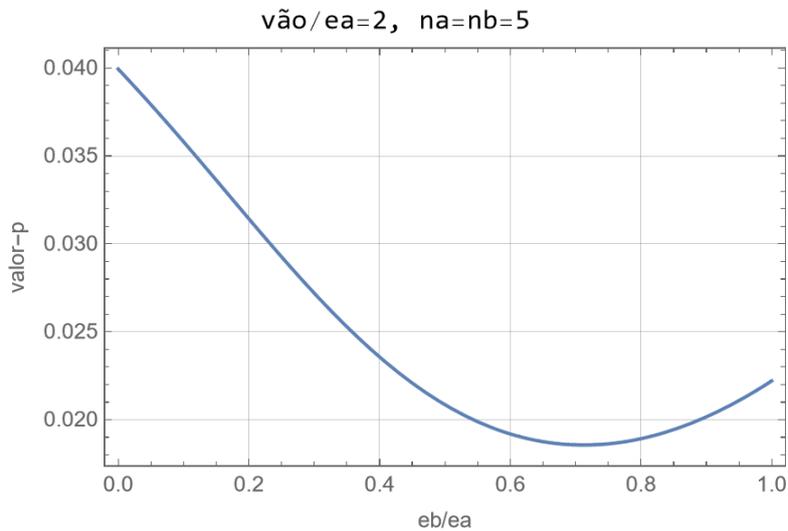


Figura 3.24: Valor-p em função de ea/eb , para o ṽo igual a duas vezes o maior erro-padrão e $n = 5$.

Os gráficos já mostrados nessa seção sugerem que as regras de bolso que propusemos têm uma fundamentação sólida. Entretanto, é importante ter em mente que o valor-p é uma função de muitas variáveis: o $v\tilde{a}o$, ea , eb , na , nb . Para verificar que os gráficos em duas dimensões já apresentados não restringem demais essas variáveis, mostramos também curvas de nível em que valor-p é dado como função de duas variáveis: $n = na = nb$ e eb/ea . Essas curvas de nível estão mostradas nas figuras 3.25, 3.26 e 3.27. O exame das figuras mostra que as conclusões a que chegamos com os gráficos mais simples já discutidos se mantém inalteradas.

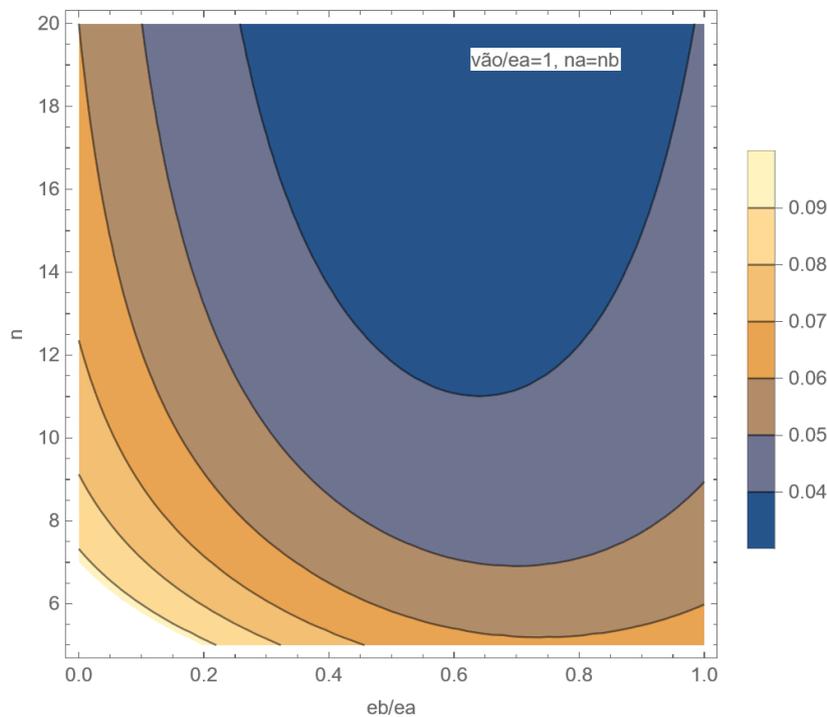


Figura 3.25: Curvas de nível do valor-p em função de $n = na = nb$ e eb/ea para o $v\tilde{a}o$ igual ao maior erro-padrão.

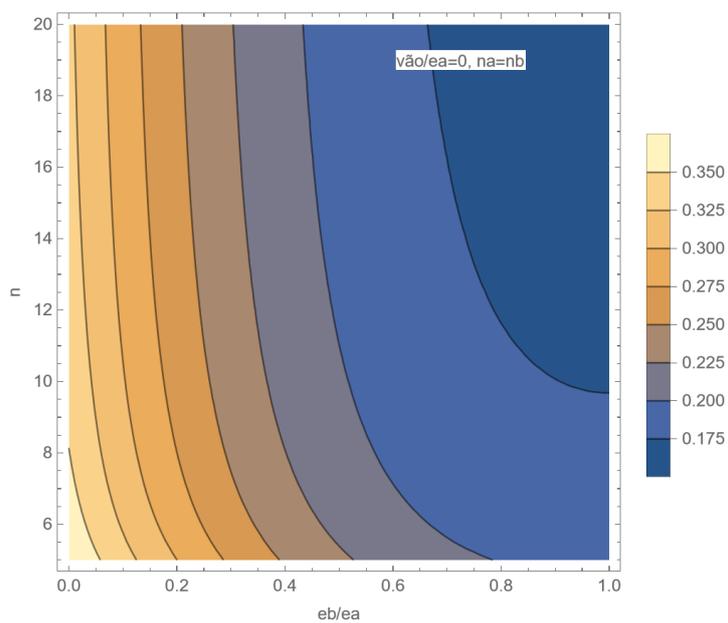


Figura 3.26: Curvas de nível do valor-p em função de $n = na = nb$ e eb/ea para o vão nulo.

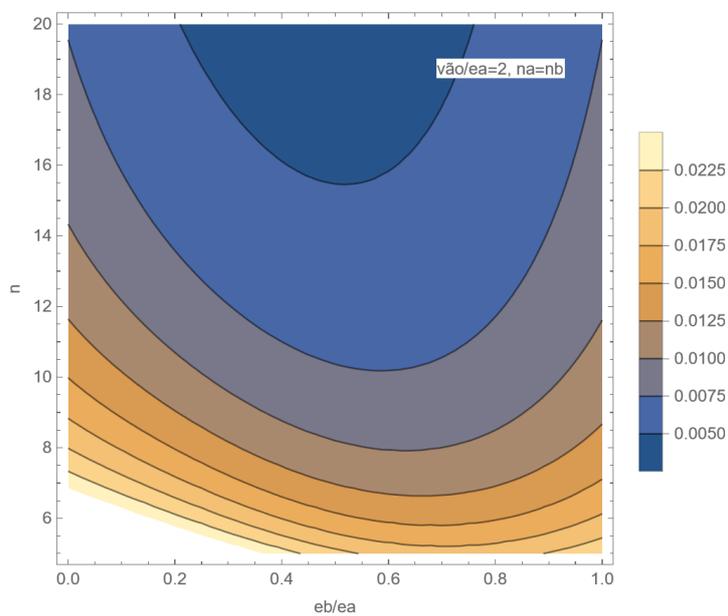


Figura 3.27: Curvas de nível do valor-p em função de $n = na = nb$ e eb/ea para o vão igual a duas vezes o maior erro-padrão.

Capítulo 4

Aplicações e resultados

Neste capítulo descreveremos as aplicações em sala de aula da nossa sequência didática, os procedimentos que adotamos e os resultados dos pré e pós-testes em cada uma delas. Foram realizadas três aplicações, duas em 2023 e uma em 2024, todas em uma escola pública federal de ensino médio da cidade do Rio de Janeiro.

4.1 Procedimento

O procedimento adotado nessas aplicações está descrito a seguir. Mais especificamente, apresentaremos o procedimento adotado na aplicação de 2024 e explicitaremos as principais diferenças em relação às aplicações anteriores. Todas as aplicações foram planejadas para uma aula de 50 minutos com atividades práticas. As atividades consistiram na medição e análise do tempo de reação de estudantes voluntários. Para isso utilizamos o aplicativo que desenvolvemos (descrito na seção 3.1), apresentado aos alunos em um notebook com acesso à internet e ligado a um projetor. Com esse arranjo, os voluntários puderam realizar a coleta de medidas, enquanto os resultados ficavam visíveis para todos os colegas de classe. O projetor apontava para um quadro branco, onde era possível realizar anotações sobre a imagem projetada.

4.1.1 Introdução

A intervenção se inicia por uma discussão sobre medições comuns no dia a dia, que geralmente envolvem a coleta de uma medida única; por exemplo, quando se mede o espaço disponível para a colocação de um móvel. Em seguida, apresentamos uma situação em que uma única medida não fornece informação suficiente. Para isso, consideramos a determinação da maior massa que um fio pode suportar. Comentamos que, para se ter noção dessa massa limite, pode ser realizado um experimento em que uma amostra do fio presa ao teto suporta massas que são gradualmente aumentadas de 10 em 10 gramas. Perguntamos à turma se, caso esse fio se rompesse com uma massa de 420 g, poderíamos garantir que outra amostra de fio vinda do mesmo rolo e com o mesmo tamanho suportaria uma massa de 410 g. Ou se necessariamente arrebentaria com 420 g. A figura 4.1 traz uma ilustração desse experimento imaginário. Após esses questionamentos, discutimos porque a medida única de 420 g fornece informação insuficiente sobre a massa limite.

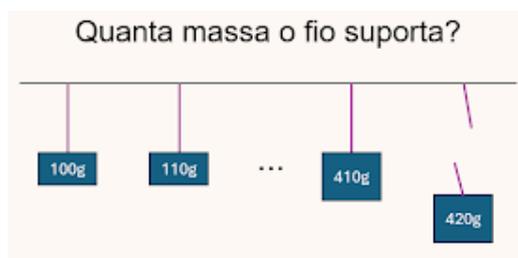


Figura 4.1: Ilustração do experimento da massa limite.

Para ver na prática uma medição em que uma única medida não fornece informação suficiente, propomos aos estudantes a medição do tempo de reação de uma pessoa. Explicamos o que é o tempo de reação a estímulos visuais e como isso está relacionado com segurança no trânsito, desempenho esportivo, habilidade em videogames, e perguntamos à turma qual deve ser o tempo de reação de um ser humano.

4.1.2 Medição do tempo de reação

Em seguida, convidamos um voluntário a medir seu próprio tempo de reação com o aplicativo. Após a medição, utilizando a imagem do aplicativo projetada para a turma, perguntamos se o valor daquela medida é o tempo de reação do voluntário. Anotamos o valor no quadro e pedimos que o voluntário faça outra medição para que possamos verificar se há alguma mudança em relação à primeira. Com o valor da segunda medida, e a projeção do aplicativo mostrando o início da construção de um histograma no eixo do tempo, perguntamos à turma qual o valor do tempo de reação do voluntário. Dada a diferença entre as duas medidas (em nenhuma aplicação os dois primeiros valores foram idênticos), escutamos algumas respostas dos estudantes: o valor mais baixo, o último valor obtido, a média, etc...

Com o propósito de entender melhor o comportamento das medidas, pedimos que o voluntário realize mais oito medições, completando um conjunto de dez. Todas as medidas foram anotadas no quadro, na forma de tabela. Com a tabela ao lado da projeção do aplicativo, comentamos como o histograma mostrado funciona e como ele facilita a visualização do conjunto de medidas.

Com a tabela e o histograma visíveis para toda a turma, perguntamos como representar o tempo de reação do voluntário. Escutamos várias respostas e suas justificativas. Geralmente, a média aritmética é uma das primeiras propostas. Após a sugestão da média, pedimos a justificativa para essa escolha e a resposta típica foi que toda a informação disponível é considerada pela média. Em seguida levantamos a questão se a média seria suficiente para representar o conjunto de dados.

4.1.3 Dispersão

Apontando para o histograma projetado, comentamos sobre a distribuição das medidas ao redor da média e salvamos a imagem do aplicativo no computador. Pedimos que outro voluntário realize dez medições, para que possamos comparar com os dados obtidos pelo primeiro. Tabelamos as novas dez medidas no quadro, salvamos a imagem do aplicativo contendo o novo histograma

e apresentamos as imagens com histogramas dos dois voluntários lado a lado (como na figura 3.7). Comparamos o formato dos dois histogramas, enfatizando a dispersão de cada um, e perguntamos qual dos dois voluntários apresenta maior consistência em suas medições. Com a resposta da turma, indicando o histograma mais estreito como o de medidas mais consistentes, apresentamos o desvio-padrão σ como forma de quantificar a dispersão das medidas, mostrando que ele indica o desvio típico de uma medida em relação a média.

Para isso, desenhamos as marcas que indicam a média \bar{t} e o intervalo de $\bar{t} - \sigma$ a $\bar{t} + \sigma$ sobre a projeção de cada histograma (produzindo algo como na figura 4.2). Comentamos como esse intervalo indica, no eixo, onde se concentram grande parte das medidas feitas e onde provavelmente uma nova medida deve ser encontrada.

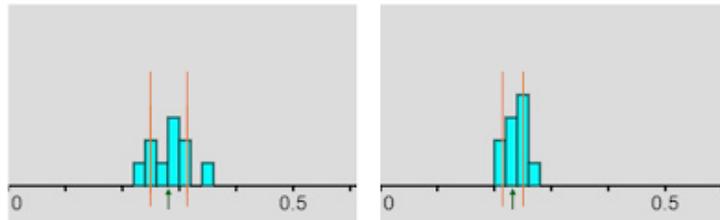


Figura 4.2: Ilustração das marcações de média e intervalos de $\bar{t} - \sigma$ a $\bar{t} + \sigma$ feitos no quadro em que a imagem era projetada durante a intervenção.

Em seguida, perguntamos aos estudantes: caso cada voluntário realizasse mais uma medição, a ordem dessas duas novas medidas poderia ser diferente da ordem das médias já encontradas. Apontamos que a resposta dependeria da possível sobreposição de intervalos correspondentes ao desvio-padrão.

4.1.4 Incerteza da média

Com as tabelas de medidas e as respectivas médias escritas no quadro, perguntamos aos alunos o que deve acontecer com o valor da média caso um voluntário refaça dez medições. As respostas dos estudantes geralmente apontam para a variação do valor da média, motivada pela provável diferença entre os valores das dez medições feitas e os valores das dez que seriam

feitas. Explicamos então que há uma forma de estimar o intervalo em que a média tende a ser encontrada em novas tomadas de medidas, partindo da tomada de um único conjunto.

O erro-padrão σ_M , dado por $\sigma_M = \sigma/\sqrt{n}$, é introduzido e apontamos que o intervalo que vai de $\bar{t} - \sigma_M$ a $\bar{t} + \sigma_M$ indica onde provavelmente será encontrada a média para um novo conjunto de medidas do tempo de reação, com mesmo número n de elementos. Enfatizamos que quanto maior for o número de medidas, menor será o erro-padrão. Justificamos qualitativamente esse comportamento argumentando que, quanto mais medidas fizermos, melhor conheceremos sua média. Ou seja, o erro-padrão fornece o grau de incerteza da média das medidas. Se, por resultado do processo de medição, queremos dizer a média do conjunto de medidas, o erro-padrão indicará a incerteza desse resultado.

Utilizando o valor do desvio-padrão das medidas dos dois voluntários, expostos nas imagens do aplicativo, calculamos o erro-padrão de cada voluntário e desenhamos um gráfico que mostra o tempo de reação médio e as barras de erro-padrão (intervalo de $\bar{t} - \sigma_M$ a $\bar{t} + \sigma_M$). Na figura 4.3 há uma foto feita no momento em que traçamos o gráfico em uma das aplicações.

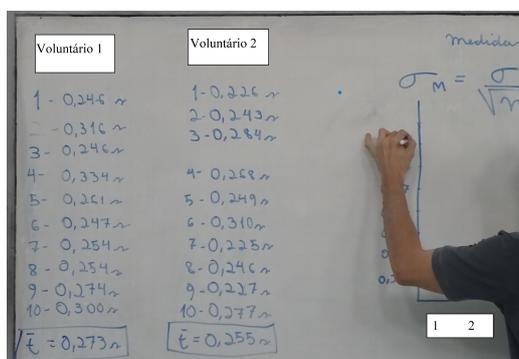


Figura 4.3: Tabela com as medidas de dois voluntários e construção do gráfico de média e incerteza.

4.1.5 Comparação entre grupos de dados

Por fim, com o gráfico das duas médias e respectivas incertezas desenhado no quadro (figura 4.4), perguntamos aos estudantes se é possível afirmar que um voluntário tem tempo de reação médio menor que o outro, sabendo que as médias podem variar em novas tomadas de medidas.

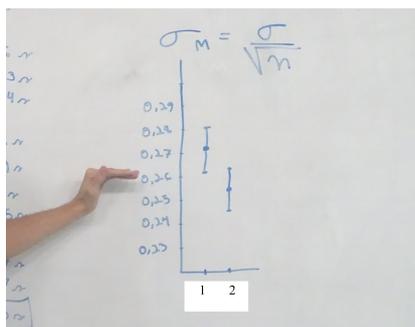


Figura 4.4: Gráfico de médias e incertezas do tempo de reação dos voluntários identificados como 1 e 2.

Generalizando a discussão já realizada sobre a comparação de medidas individuais, introduzimos nesse momento as regras de bolso apresentadas na seção 3.6, dando subsídios para que os estudantes encontrem uma resposta à questão. Na situação ilustrada na figura 4.4, por exemplo, os dados não permitiram afirmar que havia uma diferença entre os tempos de reação médios. A discussão termina com comentários sobre a generalidade desse processo de comparação e sua importância para tomadas de decisão (quem será o goleiro do seu time da escola?).

O texto de apoio distribuído aos estudantes apresenta vários dos pontos abordados acima em mais detalhes, principalmente no que diz respeito às formulações matemáticas. A leitura desse texto foi recomendada aos estudantes após a aula prática.

4.2 Aplicações e resultados

Em todas as três aplicações, o pré-teste *online* foi enviado aos estudantes dias antes do encontro, por intermédio do professor de física responsável pelas

turmas. O pós-teste, também *online*, foi fornecido logo após o encerramento da aula prática. Tanto o pré quanto o pós-teste consistiram em nossa versão do PMQ descrita na seção 2.3. Como já mencionamos, as questões 1 e 2 abordam a coleta de dados, a questão 3 versa sobre a sistematização dos resultados e as questões 4 e 5 tratam da comparação de conjuntos de medidas.

4.2.1 Primeira aplicação

Nossa primeira aplicação ocorreu com alunos do terceiro ano do ensino médio que participavam do clube de ciências do mesmo colégio em que realizamos as outras aplicações. O grupo tinha 16 estudantes.

Nessa aplicação, a motivação para a medida do tempo de reação foi o experimento histórico de Galileu sobre a velocidade da luz [26, p. 42-44]. Na descrição simplificada que utilizamos desse experimento, um feixe de luz é emitido pelo experimentador posicionado no alto de um morro, no mesmo instante em que o tempo começa a ser medido. O feixe é refletido por um espelho posicionado no topo de um segundo morro, situado a uma distância conhecida do primeiro. Ao ver a luz refletida, o experimentador interrompe a contagem de tempo. Para uma distância pequena entre o experimentador e o espelho, é impossível diferenciar o tempo de reação do tempo de percurso da luz¹. Foi colocada para os alunos a questão de se o tempo de reação limitaria o valor da velocidade da luz que Galileu seria capaz de medir (o resultado citado por Galileu foi que a propagação da luz, “se não é instantânea, é extraordinariamente rápida”).

Nessa primeira aplicação, realizamos a aula prática de forma semelhante ao procedimento exposto no início desse capítulo. Entretanto, por falta de tempo, não abordamos o erro-padrão e a comparação de conjuntos de dados.

Resultados do pré e pós-teste dos 16 estudantes estão mostrados na figura 4.5. O gráfico apresenta a fração dos estudantes que utilizou o raciocínio de conjunto para responder cada questão. Essa fração foi calculada como discutido na seção 2.3.

¹O experimento original e a influência do tempo de reação no seu resultado são discutidos em detalhe no artigo de Foschi e Leone [27].

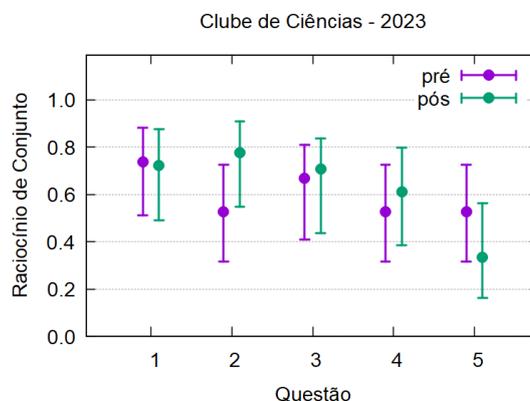


Figura 4.5: Resultados da primeira aplicação em 2023. As barras de erro representam intervalos de confiança de 95%.

Os intervalos de confiança encontrados na figura 4.5 não permitem diferenciar a proporção de raciocínios de conjunto dos estudantes do pré para o pós-teste. É interessante notar a semelhança com os resultados encontrados por Volkwyn et al. [15], mostrados na figura 2.5, principalmente no que diz respeito à maior dificuldade nas questões envolvendo comparação. Esse resultado nos levou a retirar a discussão sobre o experimento de Galileu das intervenções seguintes, de modo a ter tempo disponível para abordar a comparação de medidas.

4.2.2 Segunda aplicação

Nossa segunda aplicação foi realizada em 2023 em três turmas regulares de segundo ano do ensino médio do mesmo colégio. As atividade transcorreram em aulas de 50 minutos cedidas pelo professor de física da instituição. Alterando o procedimento da primeira aplicação, removemos a referência ao experimento de Galileu, que servia apenas como um motivador para a realização das medições do tempo de reação. Em seu lugar introduzimos uma discussão sobre a existência ou não de diferenças significativas entre os tempos de reação dos voluntários a partir da comparação dos histogramas.

Os resultados do pré e pós-teste dessa intervenção estão apresentados na figura 4.6, semelhante à mostrada para a primeira aplicação. Responderam

ao pré e pós-teste 48 e 28 estudantes, respectivamente.

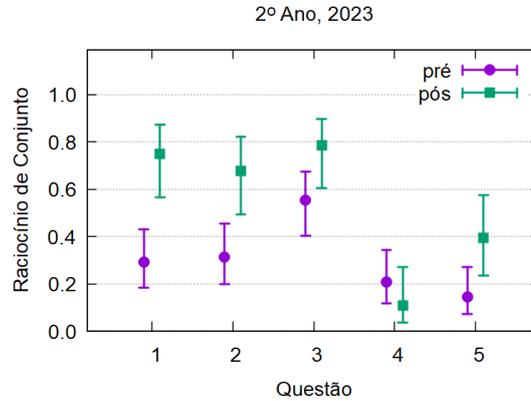


Figura 4.6: Resultados da segunda aplicação. As barras de erro representam intervalos de confiança de 95%.

Ao contrário da primeira aplicação, agora é possível notar uma diferença entre os resultados do pré e pós-teste. A tabela 4.1 mostra o valor-p e o coeficiente- ϕ [28, p. 232] da comparação do pré-teste com o pós-teste de cada questão.

Questão	Valor-p	coeficiente- ϕ
1	0,0001	0,43
2	0,002	0,34
3	0,0498	0,19
4	0,2584	0,06
5	0,0145	0,26

Tabela 4.1: Valor-p e ϕ da comparação do pré com o pós-teste cada questão segunda aplicação.

O valor-p indica que a intervenção teve um efeito estatisticamente significativo ($p \leq 0,05$) nas respostas a quase todas as questões, com exceção da de número 4. O coeficiente- ϕ mede o *tamanho do efeito* da intervenção. Costumemente atribui-se $0 \leq \phi < 0,2$ a um efeito pequeno, $0,2 \leq \phi < 0,6$ a um efeito moderado e $0,6 \leq \phi \leq 1$ a um efeito grande. O tamanho do efeito da

intervenção pode ser qualificado como moderado nas questões 1, 2 e 5, e pequeno na questão 3. A questão 4 sequer tem uma diferença estatisticamente significativa.

É possível perceber que a turma do clube de ciências teve resultados melhores em algumas questões durante o pré-teste (figura 4.5), quando comparada às turmas de segundo ano do ensino médio de 2023 (figura 4.6). Atribuímos essa diferença ao contato anterior do primeiro grupo com atividades experimentais.

4.2.3 Terceira aplicação

A terceira aplicação foi realizada em 2024 em três turmas regulares de segundo ano do ensino médio do mesmo colégio. Essa aplicação, assim como a segunda aplicação, transcorreu em aulas de 50 minutos cedidos pelo professor de física. O procedimento adotado na terceira aplicação é o descrito na seção 4.1 deste capítulo.

Devido a persistência de uma baixa proporção de raciocínios de conjunto nas duas primeiras aplicações, na resolução das questões que lidam com comparação de grupos de dados (5 e 6), ampliamos a abordagem desse tema. Introduzimos a discussão sobre flutuação da média e o conceito de erro-padrão. Também propusemos regras de bolso para simplificar a comparação de conjuntos de medidas.

Como forma de reforçar o conteúdo abordado durante a atividade prática, elaboramos um texto de apoio e, ao final da aplicação, recomendamos que os estudantes o leiam antes de responderem ao pós-teste. Esse texto constitui o material instrucional desta dissertação e está reproduzido no apêndice A.

Resultados do pré e pós-teste da nossa terceira aplicação podem ser vistos na figura 4.7, semelhante às mostradas nas aplicações anteriores. Responderam ao pré e pós-teste 45 e 28 estudantes, respectivamente.

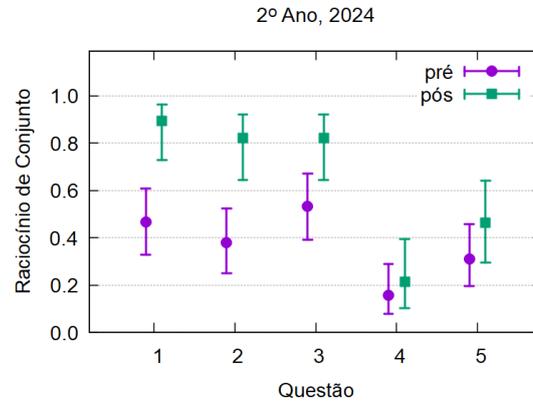


Figura 4.7: Resultados da terceira aplicação. As barras de erro representam intervalos de confiança de 95%.

Novamente é possível notar uma diferença entre os resultados do pré e pós-teste. A tabela 4.2 mostra o valor-p e o ϕ da comparação do pré-teste com o pós-teste de cada questão da avaliação feita em 2024.

Questão	Valor-p	coeficiente- ϕ
1	0,0002	0,42
2	0,0002	0,42
3	0,0124	0,27
4	0,5236	0
5	0,1874	0,1

Tabela 4.2: Valor-p e ϕ da comparação do pré com o pós-teste cada questão da oportunidade de intervenção de 2024.

O pós-teste apresenta diferenças estatisticamente significativas ($p < 0,05$) em relação ao pré-teste nas questões 1, 2 e 3. O tamanho do efeito da intervenção pode ser qualificado como moderado nas questões 1, 2 e 3. As questões 4 e 5 sequer possuem uma diferença estatisticamente significativa entre pré e pós-teste.

Comparando com as turmas de segundo ano de 2023, os resultados de 2024 não apresentaram um progresso significativo, apesar da reformulação realizada. Uma possível explicação seria que a introdução de flutuações da

média e comparações de conjuntos demandam mais tempo que o disponível numa aula de 50 minutos. A isso se soma o fato dos estudantes aparentemente não terem lido o texto de apoio.

Por fim, o resultado dos pré-testes indicam que as turmas de segundo ano de 2023 e de 2024 são praticamente equivalentes, como se vê na figura 4.8. Portanto, faz sentido comparar os resultados das diferentes intervenções nessas turmas.

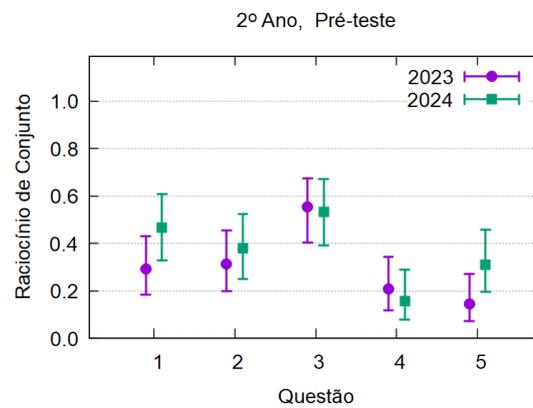


Figura 4.8: Comparação do pré-testes de 2023 com os de 2024, para as turmas de segundo ano. As barras de erro representam intervalos de confiança de 95%.

Capítulo 5

Conclusões

Neste trabalho, desenvolvemos uma sequência didática que se mostrou eficaz em aproximar os estudantes de uma perspectiva mais adequada do processo de medição. Nela, é posta em evidência a necessidade de combinar medidas para obter uma melhor representação do valor de uma grandeza, bem como a importância de estimar uma incerteza para essa representação. Além disso, propomos uma forma simplificada para a comparação de dados (regras de bolso), que pode ser utilizada para fazer inferências e avaliar decisões tomadas baseadas em dados.

Investigamos por meio de pré e pós-testes, a eficácia dessa proposta em três aplicações da sequência didática no ensino médio. Na primeira aplicação, feita em um clube de ciências, a intervenção não parece ter tido efeito. Uma possível explicação para isso seria a familiaridade desses alunos com o tema, devido à realização de experimentos no clube. Essa explicação é reforçada quando notamos que esse resultado é semelhante com o de alunos universitários da disciplina de Física Experimental I (veja as figuras 4.5 e 2.6).

Na segunda e terceira aplicações, em alunos regulares do segundo ano do ensino médio, a intervenção produziu efeitos significativos nos temas relativos à coleta e sistematização de dados. No entanto, não parece ter havido um progresso claro no que diz respeito à comparação de conjuntos de dados (veja figuras 4.6 e 4.7). Os resultados dessas duas aplicações são semelhantes aos

encontrados por Volkwyn et al. em um curso introdutório de laboratório de física da universidade de Cape Town (ver figura 2.5). Assim como os nossos alunos, os estudantes universitários apresentaram resultados positivos ao responderem questões relacionadas à coleta e sistematização de dados e apresentam maiores dificuldades nas questões sobre comparação de conjuntos de dados. Aparentemente, o tema de comparações é mais complexo que os demais e, por conta disso, é de difícil aprendizagem mesmo em diferentes níveis de ensino.

É importante ressaltar que em uma intervenção de apenas 50 minutos foi possível melhorar a perspectiva dos estudantes em aspectos fundamentais como a necessidade da repetição de medidas e a sistematização dos dados. Entretanto, parece claro que temas como a comparação de conjuntos de dados necessitam de muito mais tempo para sua compreensão. Com mais aulas disponíveis, a discussão sobre inferências e comparações poderia ser estendida, acompanhada por experimentos realizados por todos os estudantes. Além disso outros conteúdos de interesse poderiam ser abordados, como a estimativa da incerteza para o caso de uma medida única.

Nossa proposta teve o objetivo de fornecer aos estudantes uma perspectiva adequada do processo de medição e de como utilizar medidas para fazer inferências e tomar decisões. A ideia mais geral por trás desse objetivo é prover meios para que, em seu dia a dia, esses estudantes elaborem raciocínios e cheguem a conclusões consistentes com evidências empíricas. Seria muito interessante, em futuro próximo, fazer uma extensão dessa proposta voltada para a compreensão de dados do nosso cotidiano.

Apêndice A

Medidas e Incertezas



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Instituto de Física
Programa de Pós-Graduação em Ensino de Física
Mestrado Profissional em Ensino de Física

Medidas e Incertezas

Caio M. Ferreira
Carlos Eduardo Aguiar
Roberto A. Pimentel Jr
Gustavo M. Rubini

Material instrucional apresentado ao Programa de Pós-Graduação em Ensino de Física, do Instituto de Física da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Ensino de Física.

Rio de Janeiro
Dezembro de 2024

Cara professora, caro professor:

O texto apresentado a seguir, *Medidas e Incertezas*, introduz noções básicas sobre a obtenção e análise de dados experimentais por meio de uma sequência de atividades desenvolvidas para estudantes do ensino médio. As atividades envolvem medidas do tempo de reação dos alunos a estímulos visuais, realizadas com auxílio de um aplicativo criado especialmente para essa finalidade, que pode ser acessado de qualquer *smartphone* ou computador. A importância de se realizar várias medições é discutida durante as atividades, assim como maneiras de sistematizar e apresentar os resultados encontrados. Nesse processo, construtos como o histograma são apresentados e os conceitos de valor médio e incerteza das medidas são expostos sem recorrer a formulações matemáticas elaboradas, de modo a enfatizar seu significado intuitivo. O papel da incerteza em qualquer processo de medida é ressaltado, mostrando como é essencial avaliar quantitativamente a confiabilidade de um resultado experimental para que se possa tomar decisões baseadas nele.

Na escola, as atividades e discussões propostas podem ser realizadas em dois tempos de aula de 50 minutos cada. O texto disponível aqui pode ser utilizado tanto pelo professor, como um roteiro para as aulas, quanto pelos alunos, como material de estudo e aprofundamento.

Mais informações sobre o ensino de mensuração e incerteza podem ser encontrados na dissertação de mestrado de Caio Ferreira, *Medição, Incerteza e Evidência Empírica*, em http://pef.if.ufrj.br/producao_academica/dissertacoes/2024_Caio_Ferreira/dissertacao_Caio_Ferreira.pdf .

Medidas e Incertezas

1 Introdução

Imagine a situação em que precise conhecer a maior massa que pode ser suportada por um barbante preso ao teto, sem que ele se rompa. Para medir esse valor penduramos no fio massas cada vez maiores, até que ele arrebente. Vamos supor que o barbante suportou todas as massas até 410 g, mas rompeu-se quando a massa suspensa foi aumentada para 420 g. Com o rompimento, temos uma medida da massa limite suportada pelo fio, mas será que conhecemos bem esse limite? O que aconteceria se pendurássemos, por exemplo, uma carga de 414 g? E o que é mais importante, se utilizarmos outra amostra do mesmo tipo de barbante (afinal, a primeira foi arrebentada), ela também seria rompida entre 410 g e 420 g? Podemos ter certeza que ela aguentaria 400 g, por exemplo? Ou que seria rompida por 430 g?

Em situações como a descrita acima, devemos avaliar com cuidado a confiança que temos no resultado da medição, ou seja, é necessário estimar a incerteza associada à medida. É impossível medir exatamente a massa limite para o rompimento de uma dada amostra de um determinado tipo de fio. Entretanto, é possível elaborar o quanto de confiança podemos depositar na medida desse limite e buscar formas de ter mais segurança em nossas medições.

Esta questão ocupa todas as ciências: o tamanho das peças do motor de um carro, a eficácia de um medicamento, o produto interno bruto de um país, todas essas medidas possuem o que chamamos de **incerteza de medição**. Contudo, apesar de possuírem incertezas, elas desempenham papéis fundamentais em nossa sociedade. Isso acontece porque a incerteza de cada medida é algo calculável, e nos fornece o próprio grau de confiança que podemos ter na medida.

Neste texto vamos discutir a realização de medições, como representar seus resultados, como calcular incertezas e como utilizar tudo isso para fazer comparações e tomar decisões. Faremos isso usando como exemplo a medida do tempo de reação de uma pessoa. Essa medição pode ser realizada em sala de aula ou em casa com um aplicativo desenvolvido para este fim. O mesmo aplicativo permite a análise da incerteza associada ao resultado obtido.

2 Tempo de reação

Ao dirigir um veículo, é importante que o tempo de reação do condutor seja pequeno, para que, caso ocorra algum imprevisto, ele possa reagir o mais rápido possível. O tempo de reação também é relevante para situações esportivas, desde a largada em uma prova de corrida, até a defesa de um pênalti em uma partida de futebol. Existem ao menos três categorias de tempo de reação, baseados em diferentes tipos de estímulo: auditivo, tátil e visual. As medições que iremos considerar são do tempo de reação a um estímulo visual.

3 Coleta e apresentação de dados

3.1 Medição do tempo de reação

Existem várias formas de fazer a medição do tempo de reação. Nós utilizaremos um aplicativo que simplifica bastante essa medição e organiza os dados em uma forma gráfica que nos permite entender melhor os resultados encontrados.

O aplicativo pode ser usado a partir de qualquer celular ou computador conectado à internet. Ele é executado em http://pef.if.ufrj.br/producao_academica/dissertacoes/2024_Caio_Ferreira/tempo-reacao/tempo.html. Ao acessar essa página, pode-se ver uma área cinza contendo o botão “Fazer medida”, como mostrado na figura 1.



Figura 1: Aspecto inicial do aplicativo.

Quando esse botão é acionado, surge uma mensagem “Clique na tela quando a cor mudar”. Após um intervalo escolhido ao acaso pelo aplicativo, a área cinza muda de cor. O aspecto do aplicativo antes e logo depois desse intervalo está na figura 2.



Figura 2: Aplicativo iniciado.

No instante da mudança de cor da tela, inicia-se uma contagem de tempo, que será interrompida por um toque na tela do celular ou clique do mouse na área laranja. O tempo entre a mudança de cor e o toque ou clique é o tempo de reação da pessoa, que será escrito na tela como se vê na medida mostrada na figura 3. A unidade de tempo utilizada é o segundo (abreviado como s).

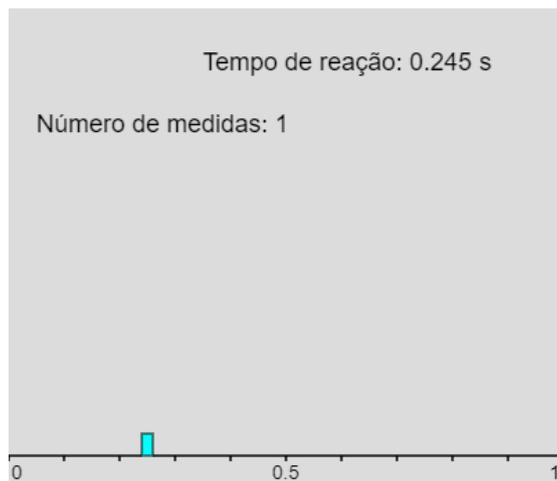


Figura 3: Medida do tempo de reação realizada com o aplicativo. O resultado é indicado numericamente e também pela posição de uma caixa sobre o eixo de tempo, que vai de 0 a 1 s em intervalos de 0,1 s.

Além de registrar o valor do tempo de reação, o aplicativo desenha uma caixa azul que representa a medida em um eixo de tempo que vai de 0,0 a 1,0 s. Note que essa caixa tem uma certa largura, dentro da qual está o valor medido. No caso a medida foi 0,245 s e a caixa vai de 0,24 s a 0,26 s, ou seja, sua largura é 0,02 s.

3.2 Medidas e histograma

Com o valor que obtivemos (0,245 s), **podemos afirmar que conhecemos o tempo de reação da pessoa que realizou a medição?** Para verificar se a afirmativa é correta, a mesma pessoa faz outra medição com o aplicativo. O resultado está mostrado na figura 4.

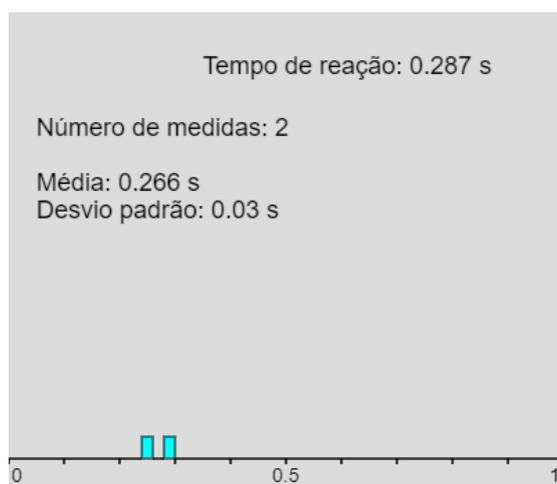


Figura 4: Duas medidas coletadas.

Na segunda medição, há um valor diferente para o tempo de reação, 0,287 s, em vez da repetição do primeiro valor 0,245 s. O aplicativo mostra em sua tela o tempo que acaba de ser medido, o número de medidas já realizadas e outras informações – a média e o desvio padrão – cujo significado discutiremos mais à frente. Além disso, mais uma caixa azul é colocada sobre o eixo, representando a nova medida. Vemos que há uma diferença perceptível entre as duas medidas, mas alguma delas é melhor que a outra?

As duas foram realizadas corretamente pelo mesmo processo, portanto não há motivos para escolhermos apenas uma como o tempo de reação, excluindo a outra. Para compreender melhor o que está acontecendo, vamos realizar mais oito medidas, completando um total de dez. Os resultados estão mostrados na tabela 1.

Medida	Valor (s)
1	0,245
2	0,287
3	0,266
4	0,264
5	0,257
6	0,274
7	0,236
8	0,284
9	0,276
10	0,260

Tabela 1: Tabulação das medidas.

Não é fácil perceber a partir da tabela como os resultados estão distribuídos. Uma forma de facilitar essa análise está na figura 5, que mostra a tela do aplicativo ao final das dez medidas. Na parte de baixo dessa figura está o que chamamos de **histograma** dos resultados. Esse histograma é construído a partir da divisão do eixo horizontal em intervalos de 0,02 s, assim temos: 0,00 a 0,02, de 0,02 a 0,04, de 0,04 a 0,06... Cada medida do tempo de reação é marcada no intervalo que a contém, no gráfico, colocando sobre esse intervalo uma caixa, como vimos na figura 3. Caso uma nova medida seja encontrada em um intervalo já ocupado por uma medida prévia, a caixa da nova medida será empilhada sobre a anterior. Com mais medidas, surgem torres de caixas empilhadas, indicando a ocorrência de várias medidas no intervalo. Esse tipo de gráfico é útil para visualizar e comparar a frequência com que eventos acontecem em cada faixa.

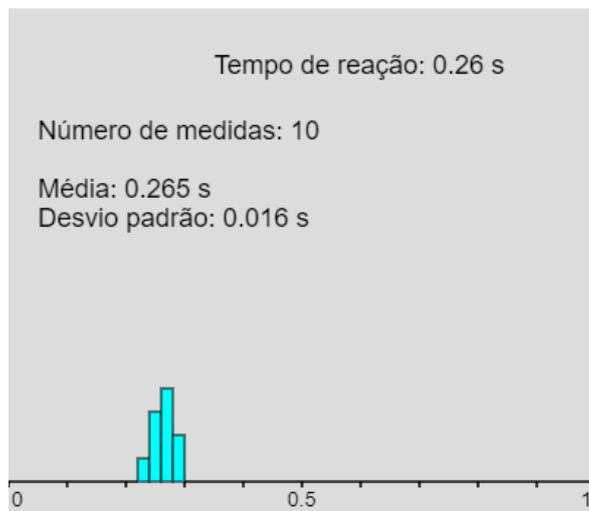


Figura 5: Histograma das dez medidas. Note como o histograma é mais útil para visualizar a distribuição das medidas do que a tabela 1.

Em virtude da variação do valor do tempo de reação nas diferentes medições, torna-se difícil escolher uma determinada medida para descrevê-lo. Todavia, podemos utilizar a **média** de todas as medidas realizadas (0,265 segundos) para obter uma representação do tempo de reação que leve em consideração a informação que está disponível.

4 A representação pela média

A média de um conjunto de n números, x_1, x_2, x_3, \dots , é definida como:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

No caso de n medidas de tempo de reação, como as discutidas acima, aplicando essa definição encontraríamos o tempo de reação médio.

$$\bar{t} = \frac{t_1 + t_2 + \dots + t_n}{n}$$

Por exemplo, a média das duas medidas ilustradas na figura 4 é:

$$\bar{t} = \frac{0,245 + 0,287}{2} = 0,266.$$

Essa média foi calculada pelo aplicativo e escrita na tela, como também se vê na figura 4. Note que a média ficou bem no meio entre as duas medidas. Para dez medidas, como na figura 5, temos o seguinte:

$$\bar{t} = \frac{0,245 + 0,287 + 0,266 + \dots + 0,284 + 0,276 + 0,260}{10} = 0,2649$$

Da mesma forma que em $n = 2$, a média das dez medidas é calculada pelo aplicativo, conforme mostrado na figura 5. Ela é escrita com o mesmo número de casas decimais das medidas; por isso o programa arredonda 0,2649 para 0,265. A média representa razoavelmente bem o valor típico das medidas, mesmo que nenhuma medição tenha encontrado esse valor. A figura 6 mostra o histograma das dez medidas com a média indicada no eixo horizontal. Vemos que o valor médio dá uma boa indicação sobre os tempos de reação encontrados.

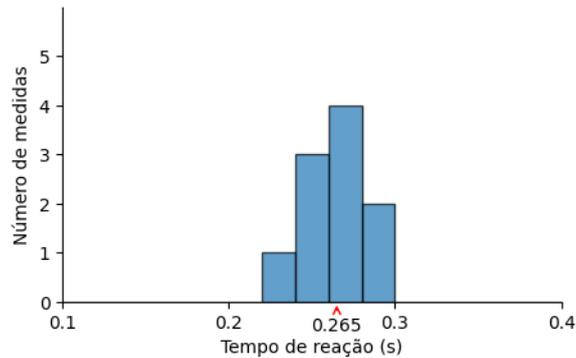


Figura 6: A seta indica a localização da média no histograma das 10 medidas.

5 Dispersão das medidas e desvio-padrão

Embora a média dê uma boa descrição do conjunto de medidas, ela não diz nada sobre a dispersão dos valores encontrados em cada medição. A dispersão das medidas do tempo de reação indica como a resposta da pessoa a um estímulo varia de uma medição para outra. Medidas próximas umas das outras indicam que as reações são sistematicamente semelhantes, enquanto medidas muito espalhadas apontam para respostas pouco consistentes entre si. Na figura 7, há dois histogramas de medidas do tempo de reação obtidos por pessoas diferentes. As medidas do histograma da esquerda (obtidos pela pessoa *A*) estão mais dispersas que as medidas do histograma à direita (obtidos pela pessoa *B*).

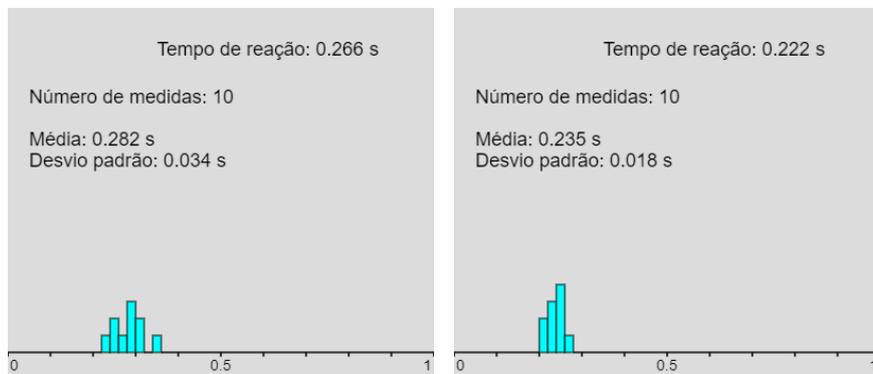


Figura 7: Histograma da pessoa *A* à esquerda e da pessoa *B* à direita.

A largura do intervalo ocupado pelo histograma dá uma ideia de como as medidas estão dispersas em torno de sua média. Quantitativamente, o grau dessa dispersão costuma ser representado pelo **desvio-padrão**.

Para calcular o desvio padrão, adotamos o seguinte procedimento. Começamos calculando os desvios de cada medida em relação à média,

$$\text{desvios: } (t_1 - \bar{t}), (t_2 - \bar{t}), \dots (t_n - \bar{t}).$$

Como alguns desvios serão positivos e outros negativos, elevamos cada um ao quadrado, para obter sempre números positivos, e os somamos:

$$(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \dots + (t_n - \bar{t})^2.$$

De maneira semelhante à média, dividimos o resultado da soma pelo número de elementos envolvidos, obtendo

$$\frac{(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \dots + (t_n - \bar{t})^2}{n}.$$

Observe que se tivéssemos tomado a média dos desvios o resultado seria zero, pois os desvios negativos cancelariam os positivos. Como estamos interessados no tamanho dos desvios, e não para qual lado da média eles foram, tomar seu quadrado elimina essa distinção. Como o resultado é a média do quadrado dos desvios, o tamanho típico do desvio será a raiz quadrada dessa média, ou seja,

$$\sqrt{\frac{(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \dots + (t_n - \bar{t})^2}{n}}.$$

É costumeiro substituir o n no denominador por $n - 1$, definindo o desvio-padrão, que chamaremos de σ , por

$$\sigma = \sqrt{\frac{(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + \dots + (t_n - \bar{t})^2}{n - 1}}.$$

Existem justificativas para colocar esse $n - 1$ no denominador. De maneira simplista, podemos dizer que ele indica a impossibilidade de calcular o desvio-padrão de apenas uma medida, pois isso resultaria em uma divisão de zero por zero na fórmula acima.

Como exemplo, abaixo está o cálculo do desvio-padrão das duas medidas presentes na figura 4:

$$\sigma = \sqrt{\frac{(0,287 - 0,266)^2 + (0,245 - 0,266)^2}{2 - 1}} = 0,0296\dots$$

Note que, na figura 4, o aplicativo arredondou o valor do desvio-padrão para 0,030 s. Pode ser trabalhoso obter o desvio-padrão de muitas medidas sem o uso de uma ferramenta computacional. Por isso o aplicativo já faz esse cálculo.

Como já mencionamos, o desvio-padrão fornece a distância típica entre as medidas e a média. Mais explicitamente, caso muitas medidas sejam feitas a maioria delas provavelmente estará entre $\bar{t} - \sigma$ e $\bar{t} + \sigma$. A figura 8 abaixo

mostra o histograma dos dados presentes na tabela 1 com indicação da média \bar{t} e da faixa entre $\bar{t} - \sigma$ e $\bar{t} + \sigma$. Como $\bar{t} = 0,265$ s e $\sigma = 0,016$ s (ver figura 5), essa faixa vai de 0,249 s até 0,281 s. Uma inspeção na tabela 1 mostra que seis das dez medidas estão nesse intervalo.

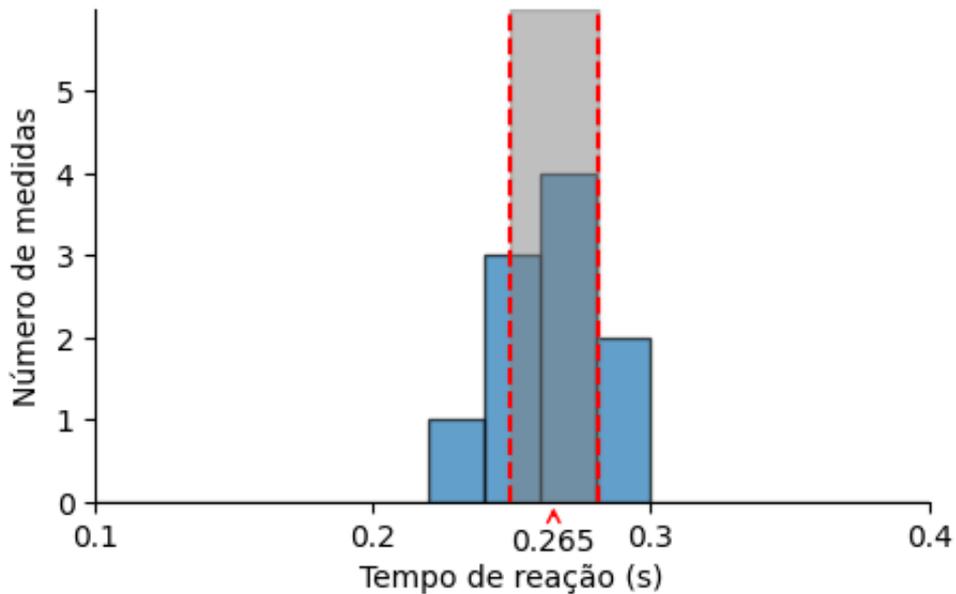


Figura 8: A seta indica a posição e o valor da média no gráfico e as retas tracejadas delimitam o intervalo ocupado pelo desvio-padrão.

Na figura 9 abaixo, vemos o gráfico com as médias e desvios-padrão dos dois grupos de medidas mostrados anteriormente na figura 7. O grupo de medidas obtidas pela “pessoa A” possui um desvio-padrão maior que o grupo da “pessoa B”. Ou seja, as medidas de B estão menos dispersas (foram mais consistentes) que as medidas de A.

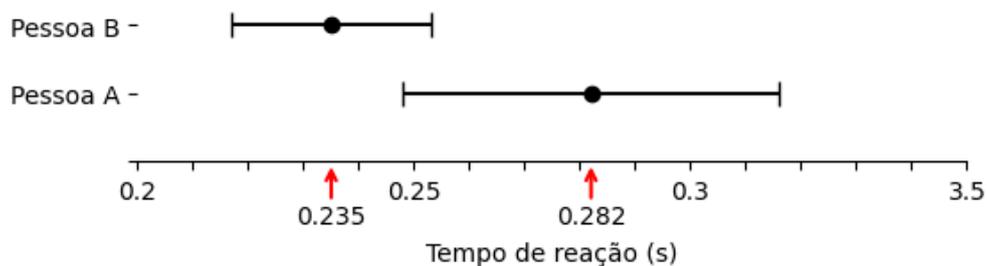


Figura 9: Representação gráfica da média e desvio-padrão do tempo de reação de duas pessoas A e B.

Os desvios-padrão permitem responder a perguntas interessantes. Por exemplo, vemos que o tempo de reação médio de B é menor do que o de A . Entretanto, se essas duas pessoas medirem mais uma vez seus tempos de reação, é possível que nesse caso o tempo de A seja menor que o de B ? A figura 9 sugere que a resposta é sim, é possível. Como há uma interseção entre os intervalos dados pelo desvio-padrão das duas pessoas, é possível que uma particular reação de B seja mais lenta que uma de A . Contudo, considerando a pequena interseção entre os intervalos, essa inversão será pouco provável.

6 Flutuação da média e desvio-padrão da média

Já sabemos como representar um conjunto de medidas, pela sua média, e como medir sua dispersão, pelo desvio-padrão. Entretanto, para a realização de comparações, temos que saber o quão precisa é a nossa representação pela média do objeto em questão (no caso, o tempo de reação). Note que se repetíssemos a realização de dez medições, como as discutidas na seção 4, os resultados individuais não seriam exatamente os mesmos da tabela 1. Por isso, é provável que a média das dez novas medidas, feitas pela mesma pessoa, seja diferente da anterior. A figura 10 mostra o histograma e a média das dez medidas que apresentamos na seção 4. Ao lado, está o histograma de outras dez medidas feitas pela mesma pessoa. Note que os dois histogramas são diferentes e que, conseqüentemente, as médias não são idênticas.

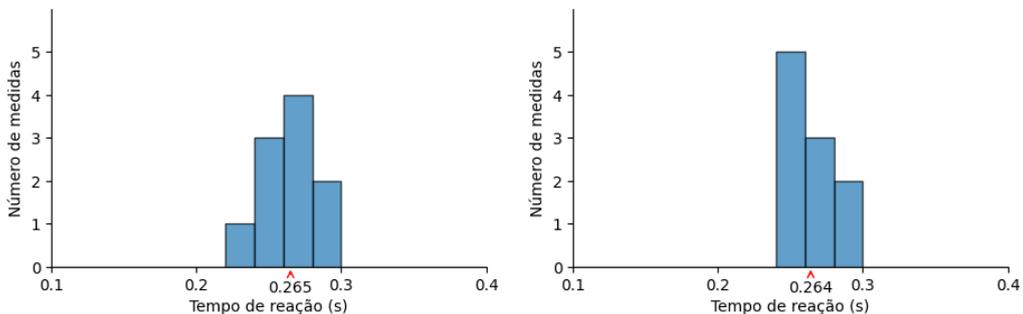


Figura 10: Dois grupos de dados obtidos pela mesma pessoa. As médias de cada grupo estão indicadas pelas setas

Como diferentes grupos de medidas da mesma quantidade podem ter médias diferentes, isso significa que há uma incerteza associada à média. Para avaliar a variação da média, a pessoa que obteve as medidas da figura 10 realizou mais oito conjuntos de dez medições e anotou o valor das médias obtidas em cada conjunto na tabela 2.

Conjunto	Média (s)
1	0,265
2	0,264
3	0,262
4	0,261
5	0,275
6	0,267
7	0,258
8	0,260
9	0,264
10	0,265

Tabela 2: Médias dos conjuntos de dez medidas. Os histogramas dos conjuntos 1 e 2 são aqueles mostrados na figura 10.

A distribuição dessas médias está mostrada na figura 11. Vemos que o histograma das médias é bem mais estreito do que os histogramas de medidas individuais apresentados na figura 10.

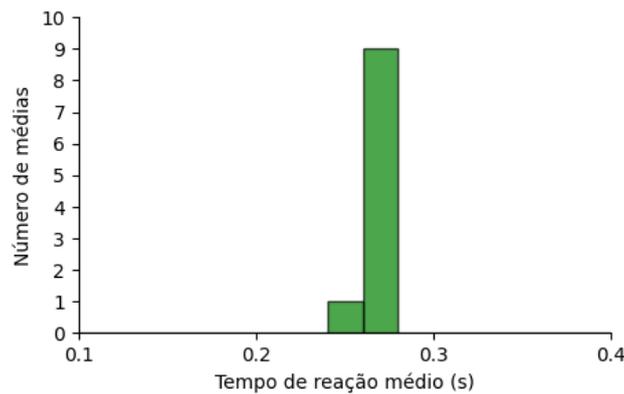


Figura 11: Histograma com os dados presentes na tabela 2. Oito das dez médias estão entre 0,26 s e 0,27 s

Quanto maior o número n de medidas, mais precisa (menos incerta) será a média. Caso tomemos as médias de dez conjuntos com cem medidas do tempo de reação da mesma pessoa em cada, essas dez médias serão ainda mais próximas entre si do que as médias de conjuntos de dez medidas mostradas na figura 11. A estabilização do valor médio com o aumento do número de medidas pode ser atribuída ao ganho de informação sobre o objeto medido. Para um conjunto de n medidas, a incerteza da média é dada pelo **desvio-padrão da média** σ_M , também chamado de **erro-padrão**.

Pode-se demonstrar que o erro-padrão de um conjunto com n medidas é dado por

$$\sigma_M = \frac{\sigma}{\sqrt{n}},$$

onde σ é o desvio-padrão das medidas, discutido na seção 5. O intervalo entre $\bar{t} - \sigma_M$ e $\bar{t} + \sigma_M$ nos fornece onde a média de um novo grupo de n medidas provavelmente será encontrada. Utilizando os dados fornecidos na tabela 1, temos o seguinte erro-padrão:

$$\sigma_M = \frac{0,016}{\sqrt{10}} = 0,005.$$

Com isso o intervalo de $\bar{t} - \sigma_M$ até $\bar{t} + \sigma_M$ vai de 0,260 a 0,270 s. É fácil verificar que esse intervalo contém 8 das 10 médias presentes na tabela 2.

Portanto, mais medidas levam à diminuição do erro-padrão e do intervalo de incerteza na média. Isso está ilustrado na figura 12, que mostra o efeito do aumento do número n de medidas sobre a representação final dos dados. Nela estão marcados a média e o erro-padrão para as nossas medidas originais ($n = 10$) e os para os valores obtidos com $n = 50$ e $n = 100$.

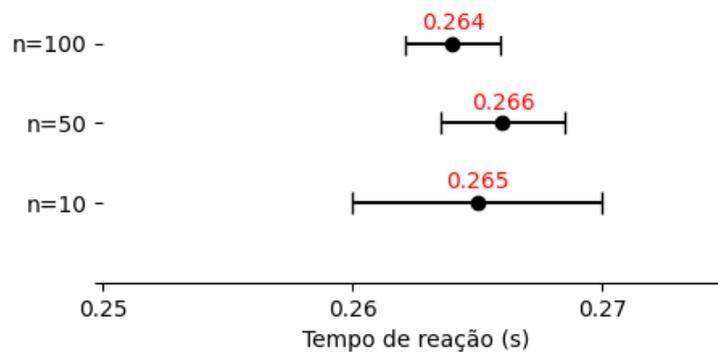


Figura 12: O tempo de reação médio de uma pessoa obtido com diferentes números de medidas. As barras representam o erro-padrão.

Um ponto a ser ressaltado é que, ao contrário do erro-padrão σ_M , o desvio-padrão σ tende a se estabilizar quando o número de medidas aumenta. O desvio-padrão indica a dispersão das medidas individuais, enquanto o erro-padrão indica a nossa confiança sobre a média.

7 Comparações entre grupos de dados

Para lidar com comparações entre grupos de dados, vamos olhar para a figura 13. Nela, posicionamos verticalmente o eixo que informa o tempo de reação, pois esse estilo gráfico é mais comum do que aquele que vínhamos utilizando. Temos as médias dos tempos de reação das pessoas A e B , já apresentadas anteriormente na seção 5 (figura 7), desta vez com as barras de erro-padrão. Olhando para as informações no gráfico, é possível afirmar que uma das duas pessoas tem, em média, reação mais rápida?

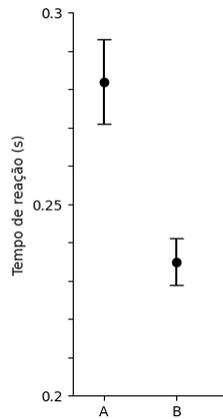


Figura 13: Médias e barras de erro-padrão de 10 medidas de tempo de reação das pessoas A e B .

Pela figura, parece claro que A tem um tempo de reação médio maior que B . Um detalhe importante para essa conclusão é a observação das barras de erro-padrão. Elas mostram a incerteza das médias e vemos que há um grande espaço (vão) entre o limite inferior da barra de A e o limite superior da barra de B . É por conta disso que podemos afirmar que os tempos médios de A e B são diferentes: a probabilidade de encontrar uma diferença tão grande por mero acidente é muito pequena. Note que sem a informação dada pelas barras de erro não teríamos noção de quão imprecisas são as médias e quais valores poderiam ser encontrados caso fizéssemos outras medidas e obtivéssemos novas médias de A e B .

Vamos agora comparar duas outras pessoas, C e D . Na figura 14 vemos um gráfico parecido com o anterior. Temos as médias das medidas do tempo de reação de C e D , junto das suas barras de erro-padrão. Olhando para o gráfico, é possível afirmar que uma dessas pessoas tem, em média, reação mais rápida?

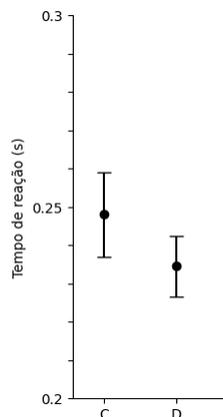


Figura 14: Médias e barras de erro-padrão de 10 medidas de tempo de reação das pessoas C e D .

Como na figura anterior, as médias possuem valores diferentes. Contudo, temos uma situação em que não há vão entre as barras, há inclusive uma sobreposição entre elas. Os dados do gráfico não nos permitem **garantir** que os tempos de reação médios de C e D são diferentes. A sobreposição entre as barras indica que se fizéssemos outras medidas e obtivéssemos novas médias, essas novas médias poderiam resultar até em uma inversão de ordem entre C e D .

Podemos generalizar o raciocínio que nos levou às conclusões acima com algumas “regras de bolso”:

- Quando o vão entre as barras de erro-padrão de duas médias a serem comparadas tem o comprimento de aproximadamente um erro-padrão, podemos dizer que há uma **probabilidade razoável de existir uma diferença** entre as quantidades representadas pelas médias. A figura 15 mostra a situação em que o vão tem comprimento de um erro-padrão.

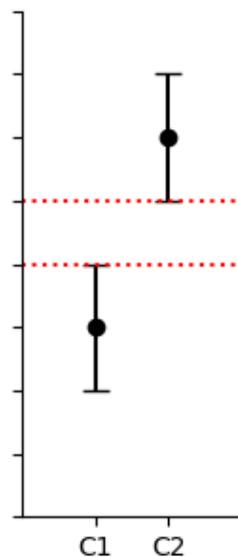


Figura 15: Comparação entre as médias de dois conjuntos de medidas, $C1$ e $C2$. A distância entre cada traço no eixo vertical equivale a um erro-padrão.

- Se há interseção entre as barras dos dois conjuntos de medidas ou se o vão tem comprimento bem menor que um erro-padrão, então **os dados não permitem afirmar que há uma diferença** entre as quantidades que comparamos. A figura 16 traz gráficos que ilustram essas situações.

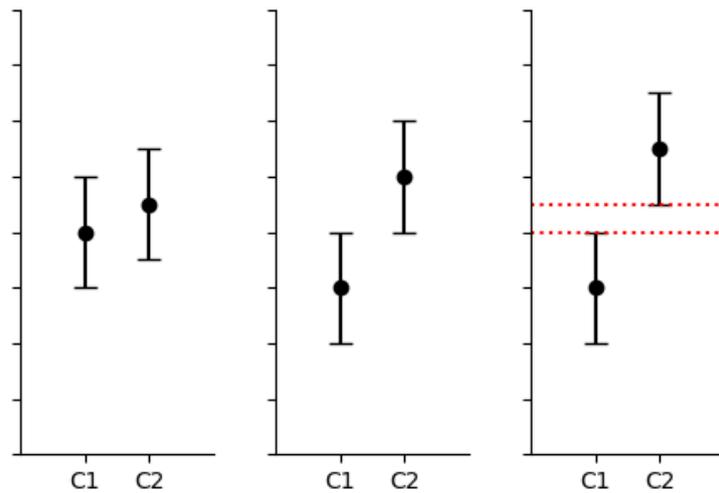


Figura 16: Situações em que os dados não permitem afirmar que há uma diferença entre as quantidades representadas pelas médias.

- Caso o vão seja bem maior que um erro-padrão, podemos afirmar que **é praticamente certo que haja uma diferença** entre as quantidades que estamos comparando. Essa situação está ilustrada na figura 17.

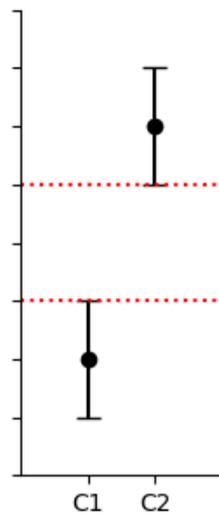


Figura 17: Vão com o comprimento de dois erros-padrão.

- Na situação em que lidamos com médias que possuem **erros-padrão diferentes**, aplicamos as regras de bolso comparando o vão entre as barras ao maior dos erros-padrão.

A regra correspondente à figura 17 se aplica às medidas de tempo de reação mostradas na figura 13, e foi assim que concluímos que o tempo de reação da pessoa *A* deve ser maior que o da pessoa *B*, estando praticamente certos disso.

É importante ressaltar que esse tipo de comparação não nos permite afirmar que duas quantidades são iguais. Só podemos dizer que elas são (muito provavelmente) diferentes, ou que não podem ser diferenciadas a partir dos dados disponíveis. Quando duas grandezas não podem ser diferenciadas, como nas situações da figura 16, isso não significa que elas sejam iguais.

Apêndice B

Questionário de Medidas Físicas

Este apêndice contém a versão do PMQ utilizada em nosso trabalho. São 5 questões traduzidas para o português e adaptadas para aplicação *online*, via Google Forms. Cada questão é respondida em duas etapas. Na primeira o estudante faz uma escolha sobre o procedimento experimental ou a análise do resultado. Na segunda etapa ele justifica essa escolha a partir de alternativas pré-determinadas. Essas alternativas foram selecionadas a partir de relatos de aplicação do PMQ disponíveis na literatura.

Questionário de Medidas Físicas

Tradução para o português de cinco questões do [Physics Measurement Questionnaire](#), adaptadas para uso online.

Instruções:

- Um experimento de física está descrito logo abaixo. Leia o texto atentamente, informe seu nome e curso e, em seguida, responda às questões que irão aparecer.
- Responda a todas as questões, na ordem em que forem surgindo. Em alguns casos você poderá marcar mais de uma opção, desde que concorde com elas.
- Após responder a uma pergunta, use o botão *Próxima* para seguir em frente.
- Ao final do questionário, use o botão *Enviar* para submeter suas respostas.

[Faça login no Google](#) para salvar o que você já preencheu. [Saiba mais](#)

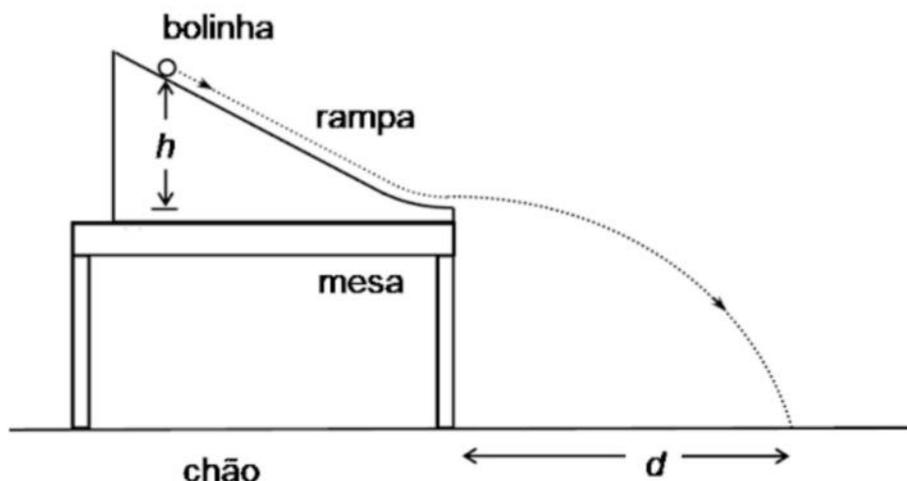
* Indica uma pergunta obrigatória

Um experimento de física

O seguinte experimento está sendo realizado por estudantes em um laboratório de física.

Uma rampa de madeira é fixada próxima à borda de uma mesa. Uma bolinha é solta da altura h acima da mesa, como mostrado no diagrama. A bolinha sai da rampa movendo-se horizontalmente e cai no chão a uma distância d da borda da mesa. O impacto produz uma pequena marca sobre uma folha de papel especial colocada no chão.

Os estudantes investigam como a distância d depende da altura h . Uma régua é usada para medir d e h .



Nome *

Sua resposta

Nome do seu professor

Sua resposta

Próxima

Limpar formulário

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários

Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

Questão 1

Repetir a medida? *

Os estudantes trabalham em grupos no experimento. A primeira tarefa deles é determinar d quando $h = 400$ mm. Um grupo lança a bolinha da altura $h = 400$ mm e, usando uma régua, eles medem que d é 436 mm.

A seguinte discussão ocorre então entre os estudantes.

Eu acho que deveríamos lançar a bolinha mais algumas vezes da mesma altura e medir d em cada vez.

Por quê? Já temos o resultado. Não precisamos fazer mais nenhum lançamento.

Eu acho que deveríamos lançar a bolinha apenas mais uma vez da mesma altura.



A



B



C

Com quem você concorda mais?

- A
- B
- C

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

1-A

Eu acho que a bolinha deveria ser lançada mais algumas vezes da mesma altura, *

porque...

- com a prática faremos medidas melhores.
- devemos repetir as medidas até que elas fiquem próximas umas das outras.
- as novas medidas podem mostrar qual é o resultado correto.
- precisamos de mais medidas para achar a média.
- precisamos de mais medidas para achar a média e avaliar o grau de incerteza dos resultados.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

1-B

Eu acho que não é preciso fazer mais nenhum lançamento, porque... *

- repetir vai levar ao mesmo resultado.
- repetir pode levar a resultados diferentes e não saberemos qual é o correto.
- repetir vai desperdiçar tempo e esforço.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

1-C

Eu acho que a bolinha deveria ser lançada apenas mais uma vez da mesma altura, porque... *

- a prática fará com que a segunda medida seja melhor que a primeira.
- precisamos confirmar o resultado da primeira medida.
- com duas medidas podemos achar a média.
- mais de dois lançamentos seriam um desperdício de tempo e esforço.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

Questão 2

Repetir a medida novamente? *

O grupo de estudantes decide lançar a bolinha novamente de $h = 400$ mm. Dessa vez eles medem $d = 426$ mm.

Primeiro lançamento: $h = 400$ mm $d = 436$ mm

Segundo lançamento: $h = 400$ mm $d = 426$ mm

A seguinte discussão acontece então entre os estudantes.



Com quem você concorda mais?

- A
- B
- C

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

2-A

Eu acho que não é preciso repetir a medida mais uma vez, porque... *

- continuaremos sem saber qual é a medida correta.
- a terceira medida seria um desperdício de tempo e esforço.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

2-B

Eu acho que é preciso repetir a medida apenas mais uma vez, porque... *

- a prática fará com que a terceira medida seja ainda melhor que as anteriores.
- a terceira medida pode confirmar um dos resultados anteriores.
- precisamos de três medidas para achar a média.
- precisamos de três medidas para achar a média e o grau de incerteza dos resultados.
- quatro ou mais medidas só aumentariam a confusão.
- quatro ou mais medidas são um desperdício de tempo e esforço.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

2-C

Eu acho que é preciso repetir a medida muitas vezes mais, porque... *

- a qualidade das medidas melhora com a prática.
- alguma das medidas pode confirmar um resultado anterior.
- devemos repetir as medidas até que elas fiquem próximas umas das outras.
- precisamos de mais medidas para avaliar melhor a média.
- precisamos de mais medidas para avaliar melhor a média e o grau de incerteza dos resultados.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

Questão 3

Como usar repetições? *

Os estudantes continuam a lançar a bolinha de uma altura $h = 400$ mm. Seus resultados após cinco lançamentos são:

<u>Lançamento</u>	<u>d (mm)</u>
1	436
2	426
3	438
4	426
5	434

Os estudantes então discutem o que escrever para d como seu resultado final.

Fico imaginando o que deveríamos escrever como nosso resultado final de d .



Escreva na linha abaixo o que você acha que os estudantes deveriam registrar como seu resultado final para d .

Sua resposta

Escrevi esse valor porque... *

- foi o resultado da última medida.
- apareceu duas vezes nas medidas.
- está entre o maior e o menor valor medido.
- é a média das medidas.
- é a medida que está mais próxima da média.
- Outro:

Voltar

Próxima

Limpar formulário

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários

Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

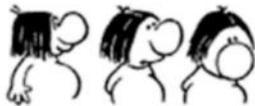
Questão 4

Qual o melhor resultado? *

Dois grupos de estudantes comparam seus resultados para d , obtidos lançando a bolinha de $h = 400$ mm. Os resultados de cinco lançamentos estão mostrados abaixo:

Lançamento	Grupo A	Grupo B
	d (mm)	d (mm)
1	444	441
2	432	460
3	424	410
4	440	424
5	435	440
Média:	435	435

Nossos resultados são melhores. Eles estão todos entre 424 mm e 444 mm. Os seus estão dispersos entre 410 mm e 460 mm.



A

Nossos resultados são tão bons quanto os seus. Nossa média é igual à sua. Ambos obtivemos 435 mm para d .



B

Eu acho que os resultados do grupo B são melhores que os do grupo A



C

Com quem você concorda mais?

- A
- B
- C

Voltar

Próxima

Limpar formulário

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários

Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

4-A

Os resultados do grupo A são melhores porque... *

- uma das medidas deles (435 mm) coincide com a média.
- a dispersão das medidas deles é menor.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

4-B

Os resultados de B são tão bons quanto os de A porque... *

- os dois grupos obtiveram a mesma média.
- a diferença na dispersão das medidas não é importante.
- os dois grupos realizaram mesmo número de medidas.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

4-C

Os resultados do grupo B são melhores porque... *

- a dispersão das medidas é menor que a de A.
- a dispersão das medidas é maior que a de A.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

Questão 5

Os resultados concordam? *

Dois outros grupos de estudantes comparam seus resultados para d , obtidos lançando a bolinha de $h = 400$ mm.

Seus resultados para cinco lançamentos estão mostrados abaixo:

<u>Lançamento</u>	<u>Grupo A</u> <u>d (mm)</u>	<u>Grupo B</u> <u>d (mm)</u>
1	440	432
2	438	444
3	433	426
4	422	433
5	<u>432</u>	<u>440</u>
Média:	433	435

Nosso resultado concorda com o seu.

Não. Seu resultado não concorda com o nosso.



A



B

Com qual grupo você concorda mais?

A

B

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

5-A

Os resultados concordam porque... *

- as médias são aproximadamente as mesmas.
- a separação entre as médias é pequena comparada à dispersão das medidas.
- há medidas coincidentes nos dois grupos.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Questionário de Medidas Físicas

Faça login no [Google](#) para salvar o que você já preencheu. [Saiba mais](#)

* Indica uma pergunta obrigatória

5-B

Os resultados não concordam porque... *

- as médias são diferentes.
- a separação entre as médias é grande comparada à dispersão das medidas.
- há medidas diferentes nos dois grupos.

[Voltar](#)

[Próxima](#)

[Limpar formulário](#)

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. [Denunciar abuso](#) - [Termos de Serviço](#) - [Política de Privacidade](#)

Google Formulários



Apêndice C

Classificação das respostas ao Questionário de Medidas Físicas

A tabela apresenta as justificativas disponíveis para as respostas a nossa versão do PMQ. As justificativas estão classificadas segundo o paradigma de ponto (P) ou de conjunto (C). A coluna “Resposta” mostra a opção tomada na primeira parte da questão; por exemplo 1A refere-se à opção A da questão 1. A coluna “Código” indica a justificativa correspondente à resposta; por exemplo, 1A2 significa a segunda justificativa à resposta 1A. Em algumas respostas é possível dar mais de uma justificativa. Nesse caso a classificação P ou C é dada de acordo com o raciocínio mais utilizado. Em caso de empate a classificação é registrada como “mista”.

Questão	Resposta	Código	Paradigma	Justificativa
1	1A	1A1	P	com a prática faremos medidas melhores.
1	1A	1A2	P	devemos repetir as medidas até que elas fiquem próximas umas das outras.

Apêndice C. Classificação das respostas ao Questionário de Medidas Físicas

1	1A	1A3	P	as novas medidas podem mostrar qual é o resultado correto.
1	1A	1A4	C	precisamos de mais medidas para achar a média.
1	1A	1A5	C	precisamos de mais medidas para achar a média e avaliar o grau de incerteza dos resultados.
1	1B	1B1	P	repetir vai levar ao mesmo resultado.
1	1B	1B2	P	repetir pode levar a resultados diferentes e não saberemos qual é o correto.
1	1B	1B3	P	repetir vai desperdiçar tempo e esforço.
1	1C	1C1	P	a prática fará com que a segunda medida seja melhor que a primeira.
1	1C	1C2	P	precisamos confirmar o resultado da primeira medida.
1	1C	1C3	C	com duas medidas podemos achar a média.
1	1C	1C4	P	mais de dois lançamentos seriam um desperdício de tempo e esforço.
2	2A	2A1	P	continuaremos sem saber qual é a medida correta.

Apêndice C. Classificação das respostas ao Questionário de Medidas Físicas

2	2A	2A2	P	a terceira medida seria um desperdício de tempo e esforço.
2	2B	2B1	P	a prática fará com que a terceira medida seja ainda melhor que as anteriores.
2	2B	2B2	P	a terceira medida pode confirmar um dos resultados anteriores.
2	2B	2B3	P	precisamos de três medidas para achar a média.
2	2B	2B4	C	precisamos de três medidas para achar a média e o grau de incerteza dos resultados.
2	2B	2B5	P	quatro ou mais medidas só aumentariam a confusão.
2	2B	2B6	P	quatro ou mais medidas são um desperdício de tempo e esforço.
2	2C	2C1	P	a qualidade das medidas melhora com a prática.
2	2C	2C2	P	alguma das medidas pode confirmar um resultado anterior.
2	2C	2C3	P	devemos repetir as medidas até que elas fiquem próximas umas das outras.

Apêndice C. Classificação das respostas ao Questionário de Medidas Físicas

2	2C	2C4	C	precisamos de mais medidas para avaliar melhor a média.
2	2C	2C5	C	precisamos de mais medidas para avaliar melhor a média e o grau de incerteza dos resultados.
3	3X	3X1	P	foi o resultado da última medida.
3	3X	3X2	P	apareceu duas vezes nas medidas.
3	3X	3X3	C	está entre o maior e o menor valor medido.
3	3X	3X4	C	é a média das medidas.
3	3X	3X5	P	é a medida que está mais próxima da média.
4	4A	4A1	P	uma das medidas deles (435 mm) coincide com a média.
4	4A	4A2	C	a dispersão das medidas deles é menor.
4	4B	4B1	P	os dois grupos obtiveram a mesma média.
4	4B	4B2	P	a diferença na dispersão das medidas não é importante.
4	4B	4B3	P	os dois grupos realizaram mesmo número de medidas.
4	4C	4C1	C	a dispersão das medidas é menor que a de A.

Apêndice C. Classificação das respostas ao Questionário de Medidas Físicas

4	4C	4C2	C	a dispersão das medidas é maior que a de A.
5	5A	5A1	P	as médias são aproximadamente as mesmas.
5	5A	5A2	C	a separação entre as médias é pequena comparada à dispersão das medidas.
5	5A	5A3	P	há medidas coincidentes nos dois grupos.
5	5B	5B1	P	as médias são diferentes.
5	5B	5B2	C	a separação entre as médias é grande comparada à dispersão das medidas.
5	5B	5B3	P	há medidas diferentes nos dois grupos.

Referências Bibliográficas

- [1] Vocabulário internacional de metrologia: conceitos fundamentais e termos associados: VIM 2012. Inmetro; 2021.
- [2] Lubben F, Campbell B, Buffler A, Allie S. Point and set reasoning in practical science measurement by entering university freshmen. *International Journal of Science Education*. 2001;85(4):311–27.
- [3] Laburú CE, Barros MA. Problemas com a compreensão de estudantes em medição: razões para a formação do Paradigma Pontual. *Investigações em Ensino de Ciências*. 2016;14(2):151-62.
- [4] Buffler A, Allie S, Lubben F. The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education*. 2001;23(11):1137–56.
- [5] Guia para a expressão da incerteza de medição. 3a ed. Rio de Janeiro: ABNT, INMETRO; 2003.
- [6] Pigosso LT, Heidemann LA. Uma revisão da literatura sobre a abordagem do processo de medição científica no ensino de física na educação básica. *Investigações em Ensino de Ciências*. 2023;28(2):332–51.
- [7] Séré M, Journeaux R, Larcher C. Learning the statistical analysis of measurement errors. *International Journal of Science Education*. 1993;15(4):427–38.
- [8] Lubben F, Millar R. Children's ideas about the reliability of experimental data. *International Journal of Science Education*. 1996;18(8):955–68.
- [9] Allie S, Buffler A, Campbell B, Lubben F. First-year physics students' perceptions of the quality of experimental measurements. *International Journal of Science Education*. 1998;20(4):447–59.

- [10] Pollard B, Werth A, Hobbs R, Lewandowski HJ. Impact of a course transformation on students' reasoning about measurement uncertainty. *Physical Review Physics Education Research*. 2020;16(2):020160.
- [11] Gomes AD. Concepções de estudantes do Ensino Médio sobre os conceitos de média e dispersão de dados. *Caderno Brasileiro de Ensino Física*. 2016;33(1):51.
- [12] Stump EM, Hughes M, Passante G, Holmes NG. Comparing introductory and beyond-introductory students' reasoning about uncertainty. *Physical Review Physics Education Research*. 2023;19(2):020147.
- [13] Campbell B, Lubben F, Buffler A, Allie S. Teaching scientific measurement at university: understanding students' ideas and laboratory curriculum reform. Monograph, *African Journal of Research in Mathematics, Science and Mathematics Education*[Internet]. 2005.
- [14] Laburú CE, Silva OH, Sales DR. Superações conceituais de estudantes do ensino médio em medição a partir de questionamentos de uma situação experimental problemática. *Revista Brasileira de Ensino Física*. 2010;32(1):1402–15.
- [15] Volkwyn TS, Allie S, Buffler A, Lubben F. Impact of a conventional introductory laboratory course on the understanding of measurement. *Physical Review Special Topics - Physics Education Research*. 2008;4(1):010108.
- [16] Kung RL. Teaching the concepts of measurement: An example of a concept-based laboratory course. *American Journal of Physics*. 2005;73(8):771–7.
- [17] Buffler A, Allie S, Lubben F. Teaching Measurement and Uncertainty the GUM Way. *The Physics Teacher*. 2008;46(9):539–43.
- [18] Allie S, Buffler A, Campbell B, Lubben F, Evangelinos D, Psillos D, Valasiades O. Teaching Measurement in the Introductory Physics Laboratory. *The Physics Teacher*. 2003;41(7):394–401.
- [19] Pigosso LT, Heidemann LA, Veit EA. O processo de medição no ensino de Física a partir do enfoque no processo de modelagem científica: subsídios para o desenvolvimento de atividades no ensino de Física. *Caderno Brasileiro de Ensino de Física*. 2024;41(1):66–93.

- [20] Barlow R. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Chichester: John Wiley & Sons; 1993.
- [21] Wilcox RR. *Basic statistics: understanding conventional methods and modern insights*. New York Oxford: Oxford University Press; 2009.
- [22] Mendes A, Rosário PP, *Metrologia e Incerteza de Medição: Conceitos e Aplicações*. 1a ed. Rio de Janeiro: LTC; 2020.
- [23] Holmes NG, Wieman CE, Bonn DA. Teaching critical thinking. *Proceedings of the National Academy of Sciences*. 2015;112(36):11199-204.
- [24] Cumming G, Finch S. Inference by Eye: Confidence Intervals and How to Read Pictures of Data. *American Psychologist*. 2005;60(2):170-80.
- [25] Ferreira JC, Patino CM. O que realmente significa o valor-p?. *Jornal Brasileiro de Pneumologia*. 2015;41(5):485.
- [26] Galilei G, *Discorsi e dimostrazioni matematiche, intorno a due nuove scienze*. 1638, traduzido para o inglês por S. Drake em *Two New Sciences*. 1a ed. Madison: University of Wisconsin Press; 1974.
- [27] Foschi R, Leone M. Galileo, Measurement of the Velocity of Light, and the Reaction Times. *Perception*. 2009;38(8):1251–9.
- [28] Barbetta PA. *Estatística aplicada às ciências sociais*. 8a ed. Florianópolis: Editora UFSC; 2012.