

**Pontifícia Universidade Católica de São Paulo
PUC-SP**

Anderson Anzai dos Santos

**A construção do letramento estatístico em estratégias com o uso
de tecnologias digitais em aulas de Estatística de cursos de
graduação**

Mestrado em Educação Matemática

**São Paulo
2019**

**Pontifícia Universidade Católica de São Paulo
PUC-SP**

Anderson Anzai dos Santos

**A construção do letramento estatístico em estratégias com o uso
de tecnologias digitais em aulas de Estatística de cursos de
graduação**

Mestrado em Educação Matemática

**Dissertação apresentada à Banca
Examinadora da Pontifícia
Universidade Católica de São Paulo,
como exigência parcial para
obtenção do título de MESTRE em
Educação Matemática, sob a
orientação da Prof.^a Dra. Cileda de
Queiroz e Silva Coutinho.**

São Paulo

2019

Banca Examinadora

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

Assinatura: _____ **Local e data:** _____

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - 88887.148921/2017-00.

AGRADECIMENTOS

Agradeço a todos que contribuíram, direta ou indiretamente, para a minha jornada nesta Universidade e à consequente conclusão deste trabalho.

À minha querida orientadora, a Prof.^a Dra. Cileda de Queiroz e Silva Coutinho, por todo o seu apoio durante esta pesquisa.

Aos demais professores do curso de Mestrado em Educação Matemática da Pontifícia Universidade Católica de São Paulo, pela excelente qualidade no ensino.

À CAPES, por todo auxílio estudantil, pela bolsa de estudo sob o processo de número 88887.148921/2017-00 e por ter contribuído de forma direta para a realização desta pesquisa.

Aos meus colegas de turma, Armando, Carolina, Eduardo, Jane, José Ronaldo e Viviane pelo excelente convívio e troca de conhecimento.

Ao grupo de estudos CEPEF (Combinatória, Estatística, Probabilidade e Educação Financeira) e, em especial a Cássio Giordano, pela contribuição direta e apoio na minha pesquisa.

Aos professores da banca Gerson Pastre de Oliveira e Fabiano Souza por todo o apoio na consecução desta pesquisa acadêmica.

Aos professores da Universidade Federal do Estado do Rio de Janeiro (UNIRIO), especialmente Alexandre Silva e Luciane Velasque pelo apoio na consecução das atividades didáticas aplicadas nos cursos da Universidade.

Aos meus grandes e melhores amigos(as) Camila Vitti, Keli Medeiros e Roberto Nogueira, por todo o incentivo e apoio durante o curso.

Aos meus pais, especialmente minha querida mãe, Helena, por estar sempre atenta a tudo que eu precisava, com todo apoio, amor, carinho e compreensão.

RESUMO

O letramento estatístico é caracterizado pela habilidade de ler, compreender, interpretar, analisar, avaliar e discutir corretamente conteúdos estatísticos. Nesse contexto, acreditamos que a construção desse letramento deve ser formada na vida do estudante desde o início da sua escolaridade. Entretanto, estudos indicam que essa formação não é suficiente e se agrava ainda mais na fase universitária. Assim, nesta pesquisa qualitativa, investigamos a situação atual das disciplinas introdutórias de estatística descritiva em cursos de graduação, analisando um da área de exatas (Licenciatura em Matemática) e dois de biológicas (Enfermagem e Medicina) na Universidade Federal do Estado do Rio de Janeiro (UNIRIO). A pesquisa foi realizada em Junho de 2018. Apoiamo-nos na fundamentação teórica da estruturação dos níveis de conceitualização dos conhecimentos pelos alunos para aplicar algumas sequências de atividades com o auxílio do *software* estatístico R e investigar se o aprendizado de conteúdos de estatística descritiva, utilizando o pacote *R Commander*, torna-se mais significativo e contribui para a construção do letramento estatístico dos estudantes universitários, bem como a contextualização da Estatística na formação profissional do estudante potencializa a construção do letramento estatístico, apoiando-o na formação de uma maior criticidade no uso dos dados e na possível articulação em seu cotidiano e formação de cidadania. Com relação aos procedimentos metodológicos, foram utilizados os pressupostos da Engenharia Didática, como estudos preliminares, análises a priori, experimentação, análises a posteriori e validação e durante a execução das atividades, identificamos que os alunos não tiveram dificuldade na manipulação do *software*, porém apresentaram dificuldade na abordagem das métricas estatísticas, na sua interpretação assim como em relacioná-las com os gráficos. Muitos demonstraram possuir apenas o nível técnico de seu conhecimento e tiveram dificuldade para ultrapassar os níveis mais elementares em seu letramento estatístico.

Palavras-chave: Educação Estatística, Letramento Estatístico, Software R, Atividades Didáticas, Engenharia Didática

ABSTRACT

Statistical Literacy is characterized by the ability of reading, understanding, interpreting, analyzing, evaluating and discussing statistical content correctly. Therefore, we believe that the construction of this literacy must be formed in the student's life from the beginning of his schooling. However, studies indicate that this training is not enough and is even worse during university stage. In this qualitative research, we expect to investigate the current situation of Intro to Descriptive Statistics major in undergraduate courses, being one of them in Exact Sciences field (Mathematics Degree) and two in Biology field (Nursing and Medicine) at the Federal University of Rio de Janeiro State (UNIRIO). The search was conducted in June 2018. We were based on theoretical foundation of levels of conceptualization of knowledge structuring by students in order to apply activities with support of the statistical software R and investigate whether the learning of Descriptive Statistics contents, using *R Commander* package, becomes more significant and contributes to the construction of statistical literacy in university students as well as whether the student's vocational training contextualization of statistics strengthens the construction of statistical literacy, basing it on the construction of superior criticality in the use of data and on the possible articulation in their daily life and constitution as an individual. Regarding the methodological procedures, Didactic Engineering assumptions were used as methodological procedures, such as preliminary studies, a priori analysis, experimentation, a posteriori analysis and validation, and, during the accomplishment of the activities, we identified that the students did not have problems in manipulating the software, however they had difficulty in approaching the statistical metrics, in its interpretation, as well as in relating them to the graphics. Many students demonstrated to have only the technical level of their knowledge and difficulties to exceed the most elementary levels in their Statistical literacy.

Keywords: Statistical Education, Statistical Literacy, Software R, Didactic Activities, Didactic Engineering.

LISTA DE ILUSTRAÇÕES

Figura 1 – Classificação de variáveis	41
Figura 2 – Gráfico de barras - distribuição dos alunos por curso e sexo.....	47
Figura 3 – Gráfico de setores - performance de 7 vendedores	48
Figura 4 – Histograma da altura de alunos de um curso de graduação	49
Figura 5 – Exemplo de um <i>boxplot</i>	50
Figura 6 – Etapas da pesquisa.....	60
Figura 7 – Habilitando o pacote <i>R Commander</i>	69
Figura 8 – Importando um banco de dados.....	69
Figura 9 – Importando um banco de dados e selecionando a planilha	70
Figura 10 – Solicitando uma tabela geral com as medidas descritivas	70
Figura 11 – Selecionando as variáveis de interesse	71
Figura 12 – Tabela de medidas descritivas no R para as variáveis peso e altura.....	71
Figura 13 – Solicitando uma tabela de multientrada no <i>R Commander</i>	72
Figura 14 – Selecionando as variáveis para a tabela de multientrada	72
Figura 15 – Saída da tabela de multientrada	73
Figura 16 – Saída da tabela com as medidas descritivas	73
Figura 17 – Solicitando uma matriz de correlação no <i>R Commander</i>	74
Figura 18 – Selecionando variáveis para uma matriz de correlação no <i>R Commander</i>	74
Figura 19 – Matriz de Correlação no <i>R Commander</i>	75
Figura 20 – Solicitando um gráfico de dispersão no <i>R Commander</i>	76
Figura 21 – Gráfico de dispersão de Glicose e Colesterol por Sexo	76
Figura 22 – Solicitando um gráfico <i>boxplot</i> no <i>R Commander</i>	77
Figura 23 – Gráfico <i>boxplot</i> no <i>R Commander</i>	77
Figura 24 – Tabela de medidas descritivas para Região e Aprovação de 2005.....	79
Figura 25 – Tabela de medidas descritivas para Região e Aprovação de 2007.....	80
Figura 26 – Tabela de medidas descritivas para Região e Aprovação de 2009.....	80
Figura 27 – Tabela de medidas descritivas para Região e Aprovação de 2005 a 2009	81
Figura 28 – Solicitando uma gráfico Histograma no <i>R Commander</i>	82
Figura 29 – Histograma de região por indicador de rendimento no <i>R Commander</i> ..	82
Figura 30 – Gráfico <i>boxplot</i> de Região por nota de Português 2005 no <i>R Commander</i>	83
Figura 31 – Resposta do grupo Med1 para o problema (1) da Atividade 1	91

Figura 32 – Resposta do grupo Med2 para o problema (1) da Atividade 1	91
Figura 33 – Resposta do grupo Med3 para o problema (1) da Atividade 1	92
Figura 34 – Resposta do grupo Enf1 para o problema (1) da Atividade 1	92
Figura 35 – Resposta do grupo Med1 para o problema (2) da Atividade 1	93
Figura 36 – Resposta do grupo Enf2 para o problema (2) da Atividade 1	94
Figura 37 – Resposta do grupo Enf3 para o problema (2) da Atividade 1	94
Figura 38 – Resposta do grupo Enf4 para o problema (2) da Atividade 1	95
Figura 39 – Resposta do grupo Med1 para o problema (3) da Atividade 1	96
Figura 40 – Resposta do grupo Med2 para o problema (3) da Atividade 1	96
Figura 41 – Resposta do grupo Med3 para o problema (3) da Atividade 1	97
Figura 42 – Resposta do grupo Enf2 para o problema (3) da Atividade 1	97
Figura 43 – Resposta do grupo Enf4 para o problema (3) da Atividade 1	98
Figura 44 – Resposta do grupo Med2 para o problema (4) da Atividade 1	99
Figura 45 – Resposta do grupo Enf2 para o problema (4) da Atividade 1	99
Figura 46 – Resposta do grupo Med1 para o problema (1) da Atividade 2	100
Figura 47 – Resposta do grupo Med2 para o problema (1) da Atividade 2	101
Figura 48 – Resposta do grupo Med3 para o problema (1) da Atividade 2	101
Figura 49 – Resposta do grupo Enf2 para o problema (1) da Atividade 2	102
Figura 50 – Resposta do grupo Enf3 para o problema (1) da Atividade 2	102
Figura 51 – Resposta do grupo Med2 para o problema (2) da Atividade 2	103
Figura 52 – Resposta do grupo Enf4 para o problema (2) da Atividade 2	104
Figura 53 – Resposta do grupo Enf5 para o problema (2) da Atividade 2	104
Figura 54 – Resposta do grupo Enf6 para o problema (2) da Atividade 2	105
Figura 55 – Resposta do grupo Mat1 para o problema (1) da Atividade 1	106
Figura 56 – Resposta do grupo Mat2 para o problema (1) da Atividade 1	106
Figura 57 – Resposta do grupo Mat1 para o problema (2) da Atividade 1	107
Figura 58 – Resposta do grupo Mat2 para o problema (2) da Atividade 1	107
Figura 59 – Resposta do grupo Mat3 para o problema (2) da Atividade 1	108
Figura 60 – Resposta do grupo Mat1 para o problema (3) da Atividade 1	108
Figura 61 – Resposta do grupo Mat1 para o problema (4) da Atividade 1	109
Figura 62 – Resposta do grupo Mat2 para o problema (4) da Atividade 1	110
Figura 63 – Resposta do grupo Mat1 para o problema (1) da Atividade 2	111
Figura 64 – Resposta do grupo Mat2 para o problema (1) da Atividade 2	111
Figura 65 – Resposta do grupo Mat1 para o problema (2) da Atividade 2	112
Figura 66 – Resposta do grupo Mat2 para o problema (2) da Atividade 2	112

LISTA DE TABELAS

Tabela 1 – Notas de alunos de quatro turmas de uma universidade	37
Tabela 2 – Idade dos alunos do curso de Enfermagem	46
Tabela 3 – Perfil dos alunos da UNIRIO por curso (por sexo)	47
Tabela 4 – Vendas de 7 funcionários em uma loja de brinquedo.....	48
Tabela 5 – Distribuição das alturas de 55 alunos de um curso de graduação	49
Tabela 6 – Caracterização dos alunos (por curso).....	60
Tabela 7 – Tabela de dupla entrada de Tabagismo e Hipertensão por Sexo Feminino	84
Tabela 8 – Tabela de dupla entrada de Tabagismo e Hipertensão por Sexo Masculino	85

LISTA DE QUADROS

Quadro 1 – Um modelo de letramento estatístico	30
--	----

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Objetivo geral	17
1.2 Objetivos específicos	17
1.3 Organização da pesquisa	18
2. REVISÃO BIBLIOGRÁFICA E APORTE TEÓRICO	19
2.1 Pesquisas na área	19
2.2 Letramento estatístico	29
2.3 Níveis de funcionamento dos conhecimentos mobilizados pelos alunos	34
2.3.1 O nível de conhecimentos técnicos	35
2.3.2 O nível de conhecimentos mobilizáveis	36
2.3.3 O nível de conhecimentos disponíveis	37
3 ESTATÍSTICA DESCRITIVA	39
3.1 O que é estatística descritiva?	39
3.2 Tipos de Variáveis	40
3.3 Medidas Descritivas	41
3.3.1 Medidas de posição	41
3.3.1.1. Média	42
3.3.1.2. Mediana	42
3.3.1.3. Moda	43
3.3.1.4. Quartil	43
3.3.2 Medidas de Dispersão	43
3.3.2.1. Variância	44
3.3.2.2. Desvio padrão	44
3.3.2.3. Amplitude	45
3.4 Distribuição de frequência - representação tabular	45
3.5 Representações gráficas	46
4 PROCEDIMENTOS METODOLÓGICOS	53
4.1 A tecnologia em contexto	53
4.2 O <i>software</i> R e o pacote <i>R Commander</i>	57
4.3 A pesquisa	58
4.3.1 Local da pesquisa	59
4.3.2 Participantes da pesquisa	59
4.3.3 Encontros e horários da coleta dos dados	60

4.4 Método: Pressupostos da Engenharia Didática	61
4.4.1 Análises Prévias	61
4.4.2 Análises <i>a priori</i>	63
4.4.3 Experimentação	64
4.4.4 Análises <i>a posteriori</i> e validação.....	65
5 ATIVIDADES DIDÁTICAS	67
5.1 Uma proposta de sequência didática	67
5.2 Atividades	67
5.2.1 Atividade 1 – Enfermagem e Medicina	68
5.2.2 Atividade 2 – Enfermagem e Medicina	75
5.2.3 Atividade 1 – Licenciatura em Matemática	78
5.2.4 Atividade 2 – Licenciatura em Matemática	81
5.2.5 Análise <i>a priori</i> da Atividade 1 – Enfermagem e Medicina.....	83
5.2.6 Análise <i>a priori</i> da Atividade 2 – Enfermagem e Medicina.....	86
5.2.7 Análise <i>a priori</i> da Atividade 1 – Licenciatura em Matemática.....	87
5.2.8 Análise <i>a priori</i> da Atividade 2 – Licenciatura em Matemática.....	88
5.2.9 Experimentação e análise <i>a posteriori</i> da Atividade 1 – Enfermagem e Medicina	90
5.2.10 Experimentação e análise <i>a posteriori</i> da Atividade 2 – Enfermagem e Medicina	100
5.2.11 Experimentação e análise <i>a posteriori</i> da Atividade 1 – Licenciatura em Matemática	105
5.2.12 Experimentação e análise <i>a posteriori</i> da Atividade 2 – Licenciatura em Matemática	110
6 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS	115
REFERÊNCIAS.....	121
ANEXOS	127
APÊNDICES	131

1 INTRODUÇÃO

As ferramentas fornecidas pela Estatística têm ganhado importância dada a sua relevância nas diversas áreas do conhecimento, sendo bastante difundida no mercado de trabalho. No entanto, pesquisas apontam para as dificuldades de aprendizagem, sejam essas inerentes à própria ciência ou consequências de escolhas didáticas. Temos interesse em aprofundar essas questões, analisando o processo de aprendizagem da Estatística introdutória básica em cursos de graduação.

Para isso, investigamos as pesquisas na área e propomos uma sequência de atividades didáticas para a aprendizagem da Estatística, utilizando estratégias com o uso de tecnologias digitais.

Para a realização deste trabalho, nos fundamentamos na ideia de autores, como Coutinho (2013), que comenta sobre o aumento das pesquisas na área da Educação Estatística, que apontam que, por proporcionar um método de análise e tratamento dos dados, essa ciência, atualmente, tem aplicações em outras ciências.

Nesse contexto, apontamos a importância de sua pertinência nos currículos dos mais diversos níveis de escolaridade, desde a escola básica até o curso superior.

Dada a minha jornada no mercado de trabalho como Estatístico, com doze anos no mercado financeiro, esses anos dedicados à modelagem estatística dos dados, à percepção do crescimento do uso da Estatística em diversos setores do mercado de trabalho e o nível de conhecimento técnico estatístico que os recém formados chegam ao mercado de trabalho, resolvi dedicar um momento de reflexão aos temas apresentados e decidi aprofundar uma pesquisa, investigando como a Estatística é abordada nas universidades.

Assim, neste estudo pretendemos discutir aspectos didáticos apresentados em algumas pesquisas na área, investigando as dificuldades de estudantes universitários na aprendizagem de conceitos da estatística descritiva básica e pensar em métodos e estratégias que possam ser utilizadas para potencializar a construção do conhecimento estatístico desses alunos.

Nas salas de aula, analisando o contexto de ensino e de aprendizagem da Estatística, percebemos que muitos alunos e professores não conseguem observar as particularidades dessa ciência e muitas vezes ela é caracterizada como

Matemática. Poderíamos pensar em uma relação entre a Matemática e a Estatística, em que entendemos que esta é um conteúdo que faz parte daquela. Porém deve ser observado que a Estatística tem suas particularidades, funcionalidades e características próprias e a trataremos assim em nossa pesquisa.

Em nossa pesquisa, trabalhamos com a ideia de Campos (2007), que utilizou a definição de Moore (2001) de que a Estatística é assumida como ciência do número em contexto, que possui técnicas próprias que permitem organizar, descrever, analisar e interpretar dados.

Nos apoiamos também em Batanero (2001), que apontou pontos de diferenciação entre a Matemática e a Estatística, comentando que a natureza da Estatística é muito diferente da cultura tradicional da Matemática, por exemplo, ainda hoje em dia prosseguem as controvérsias filosóficas sobre a interpretação e a aplicação de conceitos básicos como os de probabilidade, aleatoriedade, independência ou teste de hipóteses, enquanto estas controvérsias não existem em álgebra ou geometria.

Neste trabalho, assumimos a concepção adotada por Batanero, e seguimos a hipótese da autora de que a Matemática e a Estatística devem ser analisadas separadamente, em razão de suas particularidades.

Dessa forma, aplicamos algumas sequências didáticas, com atividades de estatística descritiva, utilizando o *software R* com o uso do pacote *R Commander*, em um curso superior da área de exatas (Licenciatura em Matemática) e dois da área de biológicas (Medicina e Enfermagem) na Universidade Federal do Estado do Rio de Janeiro (UNIRIO). A pesquisa foi realizada em Junho de 2018 e abordou 36 estudantes.

O objetivo foi analisar essas abordagens e investigar se tais estratégias contribuem para a construção de um letramento estatístico. Para isso, nos apoiamos no proposto por Gal (2002) que entende que o letramento é construído a partir de uma postura crítica e investigativa, de conhecimentos prévios de Estatística e Matemática, com habilidades de leitura e análise, crenças, atitudes e conhecimento sobre o homem e o mundo a seu redor. Segundo o autor tal letramento pode ser representado por um modelo que contempla dois tipos de elementos: (1) elementos de conhecimento, compostos por habilidades de conhecimento, conhecimento estatístico, conhecimento matemático, conhecimento do contexto e questionamento

crítico e (2) elementos de disposição que envolvem crenças e atitudes e postura crítica.

Ao desenvolver nos alunos tais elementos, partimos da premissa da melhoria de condições para a aprendizagem da Estatística, bem como a construção da criticidade e da cidadania, com habilidades de tomar decisões, econômicas, políticas ou sociais, com base em um raciocínio analítico, de modo a otimizá-las e torná-las mais assertivas.

Como estratégia para o aprendizado da Estatística e a construção do letramento, nos apoiamos na tecnologia para um processo didático, utilizando o *software* R como componente do processo.

Frente ao cenário exposto, as dificuldades dos alunos em cursos introdutórios de Estatística na graduação e a presença da tecnologia em diversos âmbitos profissionais, introduzimos a nossa questão de pesquisa: **Que contribuições podem ser evidenciadas pelo uso de estratégias didáticas com tecnologias digitais, mais especificamente o *software* estatístico R, na construção do letramento estatístico por alunos de cursos de graduação que possuem a disciplina Estatística em seu currículo?**

1.1 Objetivo geral

Investigar as possíveis contribuições do uso da tecnologia, por meio de estratégias didáticas com o *software* R, na construção do letramento estatístico por alunos de cursos de graduação da Universidade Federal do Estado do Rio de Janeiro (UNIRIO).

1.2 Objetivos específicos

A partir do objetivo geral, abordamos em nossa pesquisa dois objetivos específicos:

- Analisar as possíveis contribuições do uso da tecnologia para o aprendizado específico de conteúdos da estatística descritiva em diferentes cursos de graduação, comparando-os; e
- Investigar se estratégias didáticas que envolvem a utilização do *software* R permitem ao aluno a mobilização conceitual para a resolução de

problemas, ficando as mobilizações técnicas a cargo de estratégias específicas construídas pelo uso do *software*.

Assim, pretendemos contribuir para a área de educação estatística, investigando o contexto atual da aprendizagem da Estatística e novas estratégias que auxiliem o estudante a adquirir determinadas competências e habilidades de letramento estatístico que, ao fim, construam assertivamente tal letramento.

1.3 Organização da pesquisa

Esta pesquisa está organizada da seguinte forma:

Neste Capítulo 1, temos uma introdução de nossa pesquisa, com seus aspectos motivadores e como a Estatística está inserida nos cursos de graduação e a identificação da questão de pesquisa, seus objetivos e justificativas.

No Capítulo 2, apresentamos uma revisão bibliográfica e a fundamentação teórica utilizada na pesquisa, com a abordagem sobre o que é letramento estatístico e quais são os níveis de funcionamento dos conhecimentos mobilizados pelos alunos, teorias de Aline Robert (1998).

No Capítulo 3, realizamos uma revisão dos conceitos principais de estatística descritiva e dos conceitos que serão abordados nas atividades didáticas.

No Capítulo 4, apresentamos os procedimentos metodológicos da pesquisa bem como a utilização dos pressupostos da Engenharia Didática e suas etapas: análise preliminar, análise *a priori*, experimentação, análise *a posteriori* e validação.

No Capítulo 5, analisamos os resultados das atividades didáticas, seguidos da exposição de suas devidas conclusões.

Por fim, no Capítulo 6 apresentamos as considerações finais e as perspectivas de pesquisas que se abrem a partir de nossos resultados observados.

2. REVISÃO BIBLIOGRÁFICA E APORTE TEÓRICO

2.1 Pesquisas na área

Neste capítulo fizemos a revisão de algumas pesquisas da área. Para isso, abordamos primeiramente as discussões sobre o ensino da Estatística nos cursos de graduação, trazendo algumas considerações dos pesquisadores: Cordani (2001), Novaes (2004), Bifi (2006) e Costa (2012).

Na visão de Cordani (2001),

O que se assiste, em grande parte dos cursos universitários, é uma aversão generalizada dos alunos por parte da disciplina Estatística, geralmente oferecida nos primeiros semestres de graduação, e que apresenta um alto índice de reprovação. (CORDANI, 2001, p.19).

Estamos de acordo com a autora e identificamos que esse cenário permanece nos cursos de graduação. Nos últimos anos, tem-se discutido bastante sobre a educação estatística e qual a formação estatística, além das competências e habilidades que o aluno deve adquirir frente ao cenário atual da nossa sociedade, na qual há um mercado de trabalho cada vez mais exigente e competitivo. Assim, identificamos um papel importante da universidade, a qual deve impulsionar o estudante na construção de sua criticidade e contribuir para que saia do curso com conhecimento adequado ao mercado de trabalho e a vida cidadã.

Nesse contexto, decidimos investigar qual a reação e percepção dos alunos acerca do uso de conceitos estatísticos em sua formação pessoal e profissional. Frente a essa averiguação, realizamos também uma reflexão de como esses indivíduos chegam às universidades, ou seja, avaliamos se o conhecimento adquirido anteriormente na escola básica é suficiente para levar adiante um curso introdutório de Estatística. Cordani (2001) discute sobre esse tema:

Os alunos ingressantes no curso superior geralmente vêm com conhecimentos insuficientes (ou até mesmo inexistentes) de temas como, por exemplo, Probabilidade, variabilidade, raciocínio aleatório, fato este agravado por problemas de entendimento de linguagem (o qual afeta não só o desempenho em Estatística, como também em todas as outras áreas). Além disso, boa parte dos alunos tem uma visão determinística do mundo, culpa em parte da visão difundida pela maioria das disciplinas, nos ensinamentos fundamental e médio. Todos esses componentes, incluindo a perspectiva muitas vezes equivocada do professor, aliados a um assunto que tem complexidade conceitual, servem para dificultar o trajeto discente, tomando

essa disciplina uma das responsáveis pelo alto índice de retenção de alunos, principalmente nas universidades públicas. (CORDANI, 2001, p.11).

Concordamos com a autora e defendemos a hipótese de que a grade curricular e a abordagem da Estatística nos ensinos fundamental e médio não são adequadamente tratadas. Muitas vezes, as aulas são preparadas apenas com intuito formal e técnico da Estatística, com aplicação de regras e fórmulas, na qual o aluno não é incentivado a interpretar seus conceitos básicos e a relacioná-los com seu cotidiano sendo esse um dos fatores que geram as dificuldades dos alunos na graduação.

O objetivo geral de Cordani foi investigar como é ministrada a disciplina introdutória de Estatística nos cursos universitários, especificamente como a inferência é apresentada aos alunos em uma disciplina básica de Estatística. De acordo com a autora, o trabalho foi pensando para avaliar dois pontos: em primeiro lugar, o professor de Estatística de um curso superior, que está ministrando a disciplina introdutória de Probabilidade e Estatística e em segundo lugar, o usuário da Estatística, advindo de outro campo de estudo e interessado em conhecer alguns aspectos importantes que permeiam as suas análises ou as apresentadas em artigos de sua área de concentração.

Assim, sua questão de pesquisa foi: “Qual inferência que deve ser ensinada numa disciplina básica de um curso introdutório? ”.

Em seu trabalho, comentou sobre a grade curricular dos cursos de graduação em que as disciplinas introdutórias de Estatística são oferecidas nos primeiros semestres e, muitas vezes, o aluno não consegue associar a importância em sua vida pessoal e profissional. Segundo a autora,

Os alunos de um curso superior, geralmente sem formação inicial prévia na área de probabilidade e estatística, quase sempre desconhecem a razão pela qual a disciplina de estatística está em seu currículo, uma vez que, não tendo preocupações científicas e/ou tecnológicas e por estarem no início de sua formação (a disciplina de estatística é geralmente oferecida no início da grade curricular), não vêem como essa disciplina possa vir algum dia a lhes ser de utilidade. Isso poderia ser mudando logo na apresentação da disciplina, desde que fosse prática usual integrá-la com outras disciplinas do curso. Entretanto, como geralmente os cursos são divididos por disciplinas “não comunicantes”, o aluno não tem a oportunidade dessa visão global da aprendizagem como um processo. Além disso, como o material apresentado está centrado grande parte em exemplos com jogadas de dados, moedas ou sorteios de bolas em urnas, mais simples para o professor e mais disponíveis nos textos, embora sem qualquer apelo para os alunos, a distância entre o aluno e a disciplina vai se tornando intransponível em

grande parte dos casos, e é responsável, em boa parte, pela repetência, desistência e evasão. (CORDANI, 2001, p.109-110).

Concordamos com Cordani e defendemos que no processo de aprendizagem da Estatística, os conceitos e técnicas estatísticas devem estar associados a um contexto, que irá proporcionar equilíbrio entre a técnica e a prática aos alunos, de forma a trazer exemplos reais e motivadores para que possam visualizar a problemática e relacioná-la com alguma situação do seu dia a dia, tirando sua abstração. As estratégias didáticas praticadas em sala de aula são insuficientes e, muitas vezes, os alunos não se motivam a participar, levando alguns a desistirem de frequentar a disciplina. Cordani complementa, discorrendo sobre o ensino da Estatística:

Não basta ensinar bem. É preciso que o aluno aprenda e isto acontecerá somente quando o objeto de aprendizagem fizer sentido para este aluno. A compreensão dos conceitos em uma disciplina de estatística seria muito facilitada se os professores também estivessem preocupados com a aprendizagem, a cada momento da ação didática, que seria certamente favorecida pela inclusão de projetos ligados à área de interesse profissional do curso. (CORDANI, 2001, p.113).

Cordani (2001) abordou o ensino da Estatística e da Inferência Estatística nos cursos introdutórios da disciplina. Como metodologia, foi proposto um conjunto de problemas apresentando dois tipos de concepções em análises inferenciais: as teorias Clássica e Bayesiana.

A primeira foi encarada como paradigma nas disciplinas básicas, em que elementos de inferência são apresentados por meio de um conjunto de regras e técnicas, que o aluno aprende de forma instrumental, sem nenhuma discussão dos fundamentos subjacentes. Uma das características dessa teoria é a objetividade, no sentido de neutralizar o observador, que não incorpora suas opiniões à análise, realizando uma análise frequentista. A segunda foi vista como um contraponto à perspectiva clássica, pois defende o uso da subjetividade, levando em conta a opinião do observador no estabelecimento das chamadas análises *a priori*.

Assim, a autora abordou a Estatística Inferencial, trazendo as teorias Clássica e Bayesiana, comparando-as: a abordagem da teoria clássica com Fisher, Neyman e Pearson e a teoria Bayesiana com Bayes. Em sua tese, Cordani comentou que Fisher propôs vários esquemas de experimentação, baseados em estruturas lógicas, que produziram caminhos para a obtenção de inferências completamente válidas.

Assim como Fisher, Neyman e Pearson estabeleceram suas ferramentas inferenciais sem fazer uso das probabilidades *a priori*.

Com relação às ideias de Bayes, a autora comentou que o subjetivismo é umas de suas características, em que, na probabilidade de um evento, a análise *a priori* é sempre levada em consideração.

Outra pesquisa importante na área de educação estatística que abordou a aprendizagem em cursos de graduação foi a de Novaes (2004).

Em seu trabalho, Novaes (2004) defendeu que a Estatística, na formação universitária:

É ferramenta fundamental na interpretação e análise de dados, fornece elementos para controle, gestão e melhoria constante de processos e serviços. Consideramos ainda o fato de que essa área do saber é reconhecida mundialmente por seu papel na formação da cidadania crítica, por capacitar o sujeito para interpretar, avaliar criticamente e discutir a informação estatística nos diversos meios. (NOVAES, 2004, p.6).

Estamos de acordo com a autora e percebemos que na maioria dos cursos o aluno não consegue contextualizar o problema e associar as informações de uma tabela a um gráfico assim como não tem a visibilidade da importância da Estatística para a sua formação profissional, pessoal e crítica.

Novaes observou que a Estatística está presente em diversos cursos de graduação. Segundo ela:

A Estatística está em praticamente todas as modalidades de cursos do Nível Superior, como bacharelado, licenciatura, tecnologia, devido à sua importância como ferramenta sintetizadora, na interpretação, análise de dados, tomada de decisões em situações com componentes de imprevisibilidade ou multiplicidade de variáveis pertinentes, fornecendo elementos para controle, gestão e melhoria constante de processos. (NOVAES, 2004, p.10).

A autora ainda comentou sobre o papel da Estatística na formação do cidadão:

Podemos considerar que os conhecimentos de Estatística sempre desenvolvem habilidades em qualquer cidadão. Entendemos como formação estatística adequada aquela que, segundo os preceitos da Educação, permite ao aluno construir seus conhecimentos de forma autônoma, por meio da mediação do professor. O aluno pode assim construir os conceitos de base da Estatística, de forma a poder mobilizá-los adequadamente em situações concretas de sua vida. (NOVAES, 2004, p.9).

Nesse contexto, Novaes, em sua pesquisa, teve por objetivo investigar o ensino e aprendizado da Estatística em um curso universitário tecnológico de

Turismo. Sua questão de pesquisa foi verificar se os alunos egressos em tal curso, no qual ocorre o contato inicial com a Estatística, estavam capacitados a utilizar/mobilizar de forma eficaz as noções estatísticas básicas para resolver problemas práticos de sua área de atuação.

Para isso, os alunos foram convidados a resolver diversos problemas em sua área de atuação, com interfaces nas áreas de Gestão, Marketing, Planejamento e Meio Ambiente. Como referencial teórico, Novaes (2004) utilizou os níveis de dimensões de análise dos conteúdos a ensinar, de Robert (1998), no qual a autora abordou quatro dimensões.

A primeira e a segunda dimensões propostas por Robert estavam relacionadas aos trabalhos de Douady (1986) e Duval (1996) e não foram objetos de sua pesquisa. A terceira dimensão, “Níveis de conceitualização”, estava associada à Teoria dos Campos Conceituais de Vergnaud e, também, não foi abordada. Sua pesquisa foi focada na quarta dimensão, que também foi utilizada aqui em nossa pesquisa, em que foram considerados os “Níveis de mobilização dos conhecimentos pelos alunos”, no qual Robert (1998) descreve três níveis para os conhecimentos adquiridos: técnico, mobilizável e disponível.

Como procedimentos metodológicos, Novaes em sua pesquisa propôs um problema prático para a área de Turismo. Para verificar o nível de mobilização dos conceitos, foi criada uma situação típica de estudo de demanda turística, em que 12 pessoas participaram, formando 6 duplas. Cada dupla foi observada individualmente, garantindo a não comunicação entre uma dupla e outra, e as interações entre os sujeitos foram audio-gravadas.

Sugeriu-se uma atividade onde foi explorada a demanda turística de uma região. Dessa forma, foi solicitado aos alunos que analisassem o perfil do visitante do local e baseando-se nos dados coletados justificassem a existência ou não de demanda por hospedagem na região determinada.

No desenvolvimento das atividades, Novaes esperava que o alunos mobilizassem os seguintes conhecimentos estatísticos: tamanho da amostra, amplitude amostral, construção de gráficos e tabelas; e cálculo de medidas descritivas, como média, mediana, moda, separatrizes e desvio-padrão.

Como resultados, a autora identificou cinco categorias de erros durante a consecução das atividades: (1) tratamento da amostra como se fosse à população; (2) incorreção no conceito média/moda/mediana; (3) dificuldade no trabalho com

unidades de medidas; (4) análise inadequada da variabilidade dos dados e (5) inexatidão no cálculo de porcentagens.

Novaes (2004) comentou que nenhuma das seis duplas pesquisadas conseguiu resolver a situação problema proposta completamente, ou seja, nenhuma atingiu o nível disponível de conhecimentos nos termos de Robert (1998) para a resolução da situação como um todo. Algumas duplas evoluíram mais do que outras no desenvolvimento da estratégia de resolução do problema e atingiram o nível mobilizável de conhecimento. Outras sequer foram capazes de começar a resolução. Os alunos participantes não tiveram êxito ao realizar a relação entre o real e a teoria que explica esse real. A maioria das duplas demonstraram habilidade em construir adequadamente gráficos e tabelas, bem como calcular as medidas de posição. Isso indicou que esses alunos mobilizam esses conhecimentos no nível técnico, que é o mais rudimentar para Robert, referencial utilizado na pesquisa.

Outra pesquisa importante na área foi a de Bifi (2006), que abordou a mobilização dos conhecimentos estatísticos em um curso de Administração de Empresas. Seu objetivo foi investigar se os alunos egressos do componente curricular Estatística do curso estavam capacitados a utilizar e/ou mobilizar, de forma eficaz, as noções estatísticas ligadas à variabilidade (média e desvio-padrão) na análise exploratória dos dados.

Nesse estudo, sua questão de pesquisa foi: “Qual o nível de funcionamento dos conceitos, segundo os preceitos de Robert, especificamente aqueles ligados ao estudo da variabilidade, dos alunos do Ensino Superior? ”.

Para o referencial teórico, Bifi também usou as propostas das quatro dimensões de Robert (1998) e os níveis já citados de conhecimentos mobilizados pelos alunos.

Quanto aos aspectos metodológicos, Bifi apontou que foi proposta uma situação problema na forma de uma atividade diagnóstica qualitativa, dividida em três etapas. Esperava-se que essa atividade permitisse diagnosticar o nível de funcionamento dos conceitos, segundo os preceitos de Robert (1998) especificamente aqueles ligados ao estudo da variabilidade, por parte dos alunos do ensino superior, e, também, identificar o(s) possível(is) erro(s) cometido(s) por eles.

Foi aplicada a situação problema para 3 duplas de alunos que já haviam cursado a disciplina de Estatística, oferecida no curso de Administração de Empresas, em um tempo estimado de 200 minutos (4 horas/aula). As atividades

envolveram aspectos da estatística descritiva: média, mediana, quartis, tabelas de frequência e medidas de variabilidade.

Como resultado, em resumo, muitos alunos tiveram dificuldades em interpretar ou justificar os cálculos realizados, apesar de estarem todos corretos. Assim, foi comentado que os alunos se classificaram em um nível técnico de mobilização de conhecimentos. Outros pontos identificados foram o trabalho do conjunto de dados na representação tabular e, em alguns momentos, algumas análises equivocadas ocasionadas pelas dificuldades em entender as diferenças conceituais entre média e mediana. Por fim, o autor comentou que os alunos se mostraram mais interessados em trabalhar com os dados analíticos e não com representações tabulares ou gráficas.

Agora, abordamos a pesquisa de Costa (2012) na qual a autora comentou sobre a importância da informação estatística nos cursos de graduação e sua formação profissional. Segundo a autora:

Por ser um instrumento que ajuda a compreender e a intervir nos fenômenos de ordem econômica, política e social, a estatística constitui uma linguagem que traduz e aproxima o conhecimento de fenômenos presentes no cotidiano, permitindo e possibilitando ao aluno a oportunidade de vivenciar situações problemas que o ajudem a compreender melhor como funciona a realidade do mundo profissional. (COSTA, 2012, p.16).

Concordamos com os apontamentos da autora e entendemos que o mercado de trabalho, cada vez mais acelerado e competitivo, exige desse estudante, futuro profissional desse mercado de trabalho, um preparo maior, que o capacite a discutir técnicas estatísticas em constante avanço e que saiba utilizar a grande quantidade de dados disponíveis, resumindo-os, interpretando e inserindo os mesmos em suas tomadas de decisão. Esta perspectiva apoiou nossa intenção de abordar a contextualização na utilização desses dados e as possíveis contribuições do uso da tecnologia em sala de aula.

Para tanto, Costa (2012) complementou que:

O avanço da tecnologia, aliado às exigências de sofisticadas competências e habilidades para o mundo do trabalho, bem como a facilidade e praticidade oferecida pela informática, levaram os meios de comunicação a recorrerem à linguagem estatística para implementar, organizar e sintetizar informações. Dentro de uma perspectiva de mundo, que busca a otimização de tempo, recursos e espaço, e da frequência da utilização de modelos estatísticos para interpretar acontecimentos, incrementar a comunicação ou explicar movimentos sociais. (COSTA, 2012, p. 11).

Em sua pesquisa, a autora teve por objetivo investigar a metodologia de projetos como uma alternativa para ensinar Estatística no ensino superior de Nutrição, em uma instituição particular de ensino da cidade de Campinas-SP.

Com relação ao aporte teórico, Costa (2012) foi buscar em referencial que pudesse articular afeto, aprendizagem e cognição para fundamentar as análises. Assim, apoiou-se em Vygotsky que, segundo a autora diz não ser possível negar que o afeto assume um papel essencial no processo de ensino e aprendizagem, uma vez sem ele não haveria interesse, nem necessidade e nem motivação. Consequentemente, perguntas ou problemas nunca seriam colocados; e, radicalizando, não haveria inteligência.

Segundo a autora, é importante entender que, no decorrer de todo o processo de desenvolvimento do ensino-aprendizagem, a afetividade é como uma energia que impulsiona as ações. Dessa forma, foi destacado a importância da relação entre professor e aluno também no ensino superior, de modo que ambos convivam em um ambiente de harmonia, e que a aprendizagem, assim, possa fluir com mais facilidade, havendo maior rendimento e interação entre ambos.

Em seu processo metodológico, Costa optou por fazer uma pesquisa de campo que, consistiu em uma modalidade de investigação na qual a coleta de dados foi realizada diretamente no local em que o problema ou fenômeno acontece e pôde ser dar por amostragem, entrevista, observação participante, pesquisa-ação, aplicação de questionário, testes, entre outros.

A coleta de dados foi realizada em uma universidade particular de Campinas-SP, contando com alunas, monitoras e professoras de um curso de Bioestatística. A metodologia de pesquisa contou com a consecução de projetos que segundo a autora criou uma articulação entre os interesses e motivações pessoais, profissionais, sociais, comunitários ou acadêmicos e ofereceu ao aluno a oportunidade de atuar de forma crítica, ponderada e reflexiva, no mercado de trabalho e no grupo social em que vive. Assim, foi adotada uma metodologia de trabalho com projetos como alternativa para ensinar Estatística no ensino superior.

Na pesquisa, foi proposto um questionário previamente testado, utilizado para coleta de dados relativos ao projeto interdisciplinar e sua aplicação teve por finalidade a obtenção de dados para caracterizar o perfil socioeconômico, de saúde e nutricional dos funcionários da instituição.

Para explorar o perfil socioeconômico dos funcionários da faculdade, foram coletadas as seguintes informações: nome, bairro onde mora, sexo, idade, escolaridade, faixa de renda, cargo e função desempenhada, tempo de exercício profissional na instituição. Para a determinação do perfil de saúde e nutricional, foram coletados os seguintes fatores: morbidade referida, tempo de diagnóstico da morbidade, antecedentes familiares de morbidade, uso de medicamentos, tempo de uso da medicação, prática de exercício físico, uso de dieta, uso de cigarro de tabaco e uso de bebida alcoólica. Já para o levantamento do consumo alimentar, foram utilizadas questões que o expressavam por grupos de alimentos.

Após a coleta dos dados, as alunas do curso os organizaram em gráficos e tabelas, utilizando os conceitos aprendidos na disciplina de Bioestatística para descobrir as possíveis relações entre as variáveis investigadas. Os resultados foram formalizados em um pôster e a divulgação dos resultados da pesquisa ocorreram na Semana da Nutrição no semestre seguinte.

Em sua pesquisa, Costa (2012) apresentou os principais resultados, atestando que as alunas gostaram de realizar projetos, principalmente por conta do incentivo de trabalhar em grupo e do relacionamento com as pessoas entrevistadas, que possuíam problemáticas próximas ao contexto da profissão. Segundo a autora, as observações realizadas em sala de aula e os conteúdos das menções relatadas levaram a crer que o trabalho com projetos ofereceu às alunas a oportunidade de participarem da busca de seu conhecimento, retomando e revendo valores e posicionamentos frente ao processo educacional.

Analisando o ponto de vista afetivo, de acordo com a autora, a interação entre os docentes e discentes envolvidos em seu trabalho, enquanto foi desenvolvida uma investigação estatística intermediada pelo projeto interdisciplinar no curso de Nutrição, bem como as dificuldades e imprevistos associados à experiência única em mediar as atividades relacionadas ao projeto, contribuíram para ressignificar crenças e saberes profissionais e acadêmicos.

Sob o ponto de vista da interação com a Estatística, a autora comentou que muitas alunas sentiram dificuldade no tratamento e análise dos dados. Entretanto, muitas se sentiram protagonistas do seu conhecimento, identificando a importância da Estatística em sua formação profissional e, comentaram que o trabalho por meio de projetos, assim como a visualização do contexto, motivaram e potencializaram o aprendizado.

As pesquisas de Cordani (2001), Novaes (2004), Bifi (2006) e Costa (2012) foram utilizadas como apoio em nossa pesquisa, dado o enfoque da presença da Estatística nos cursos de graduação. Em nosso trabalho, também investigamos quais as dificuldades dos alunos em cursos introdutórios de Estatística e identificamos se estratégias vinculadas ao uso de tecnologias digitais contribuíram para a sua aprendizagem. Para identificar os níveis de conhecimentos mobilizados dos alunos em cursos de graduação, nos baseamos na teoria de Robert (1998) e no modelo de letramento estatístico proposto por Gal (2002).

Ainda sobre a investigação da Estatística em cursos universitários de acordo com Wada (1996), foi possível observar que a disciplina de Estatística no ensino superior assume um caráter básico e introdutório, visando à formação estatística do aluno para o exercício de sua futura profissão. Trata-se de uma disciplina de serviço, que almeja preparar o estudante universitário a atuar ativa e efetivamente nas diversas esferas sociais, tais como escola, família e comunidade; e no exercício de sua profissão, lendo, interpretando e produzindo informações.

Porém, Jacobini (1999) apontou que é comum encontrar no ensino superior uma abordagem da Estatística que privilegia apenas aspectos como:

Construir gráficos, mas não interpretá-los. Valoriza-se a construção de uma distribuição de frequência e de um histograma, mas não a interpretação dos desvios ou padrões existentes nos gráficos. Durante todo esse processo, o aluno ouve, aprende a fazer cálculos e a construir gráficos (usando o computador ou manualmente) e trabalha com dados que pouco tem a ver com ele. O curso ministrado dessa forma não induz o aluno a pensar no por que do estudo da Estatística, mas apenas a como aplicar as fórmulas relacionadas com as técnicas estatísticas. (JACOBINI, 1999, p.13).

Tanto concordamos com o autor quanto levantamos a hipótese de que no ensino da Estatística, o professor deve contextualizar os problemas e capacitar o aluno a ir além de memorizar fórmulas ou apenas seguir regras e padrões. O professor deve incentivar o aluno a adquirir habilidades que o permita articular o problema, apropriando-se de conceitos estatísticos como um meio para atingir a sua resolução.

Com relação a essa construção de conhecimentos estatísticos, observamos que muitos pesquisadores concordam não apenas acerca do papel da Estatística como ferramenta nas diversas ciências, mas também quanto à importância da contextualização na consideração de seus dados para formação crítica do cidadão e na aquisição de habilidades em ler, interpretar, avaliar e tomar decisões mais

assertivas em seu cotidiano (WADA, 1996; NOVAES, 2004; BIFI, 2006; COSTA, 2012; GIORDANO, 2016).

Quando analisamos os cursos universitários, identificamos que a Estatística não está presente apenas na grade curricular de ciências exatas, mas, sim, em diferentes, compreendendo os cursos de ciências humanas (Pedagogia, Direito, Administração, etc.) e ciências biológicas (Biologia, Odontologia, Fisioterapia, etc.). Com isso, nos questionamos se a abordagem da Estatística nos diferentes cursos é contextualizada e direcionada para a formação específica do referido curso.

Diante desse cenário, percebemos que, ao abordamos a Estatística, o uso da tecnologia está cada vez mais presente na sociedade e no mercado de trabalho. Por outro lado, identificamos que os estudantes não são incentivados a se apropriar desses conhecimentos e relacionar com a Estatística. Visando potencializar tal apropriação, discutimos neste trabalho as possíveis relações e propomos algumas atividades didáticas utilizando o *software* estatístico *R*, trabalhadas com alunos de um curso da área de exatas (Licenciatura em Matemática) e dois da área de biológicas (Enfermagem e Medicina).

As teorias abordadas em nossa pesquisa nos auxiliaram a investigar a problemática e os nossos questionamentos apresentados anteriormente.

Assim, realizamos uma reflexão sobre o letramento estatístico e as propostas de Robert (1998), em que abordaram os níveis de funcionamento dos conhecimentos pelos alunos.

2.2 Letramento estatístico

Não podemos discutir aspectos da educação estatística, sem entendermos o que é o letramento estatístico. Alguns autores mencionam o letramento como literacia. Nesta pesquisa, consideramos ambos como sinônimos e adotamos a expressão “letramento estatístico”.

Gal (2002) entendeu que letramento estatístico é a habilidade para interpretar e avaliar criticamente as informações estatísticas e os argumentos baseados em dados que aparecem nas diversas mídias, bem como para discutir opiniões referentes a esse tipo de informação.

Já WATSON (1997 apud GIORDANO, 2012, p. 37) definiu letramento estatístico como capacidade de compreensão textual e das eventuais implicações das informações estatísticas contextualizadas, envolvendo entendimento básico de sua terminologia, de sua linguagem e de conceitos inseridos em um contexto social, bem como o desenvolvimento de atitudes investigativas críticas.

Segundo Gal (2002), o comportamento de um conhecedor de questões estatísticas é baseado na ativação conjunta de cinco bases de conhecimento inter-relacionadas (conhecimento geral, estatístico, matemático, contextual e crítico), junto a um grupo de disposições que servem como suporte e de crenças capacitadoras.

As implicações educacionais e de pesquisa são examinadas, e as responsabilidades enfrentadas pelos educadores, estatísticos e outros participantes. Dessa forma, Gal criou um modelo para o conhecimento estatístico, entendido também por letramento estatístico, que pode ser dividido em dois blocos: (1) elementos do conhecimento e (2) elementos de disposição, como apresentado na tabela a seguir.

Quadro 1 – Um modelo de letramento estatístico¹

Elementos de conhecimento	Elementos de disposição
Habilidades de letramento Conhecimento estatístico Conhecimento matemático Conhecimento do contexto Questionamento crítico	Crenças e atitudes Postura crítica
Letramento estatístico	

Fonte: Gal (2002, p.4, tradução nossa)

Com relação às habilidades de conhecimento, Gal (2002) aponta que é a etapa inicial nos exames das bases de conhecimento necessárias para o conhecimento estatístico, uma vez que praticamente todas as mensagens estatísticas são transmitidas através de textos escritos ou orais. Isto é, os leitores são obrigados a navegar por informações fornecidas em tabelas ou gráficos que requerem a ativação de habilidades específicas de conhecimento. Os leitores precisam entender o problema como um todo, ou seja, a parte em que a Estatística

¹Tabela traduzida do artigo Conocimientos básicos de estadística en adultos: significados, componentes, responsabilidades.

é discutida ou a que explica o gráfico ou a tabela apresentada, objetivando incluí-la no contexto apropriado.

Os leitores precisam se comunicar, ter opiniões claras, oralmente ou por escrito, momentos em que a sua resposta deve conter informação suficiente sobre a lógica ou a evidência na qual se baseia, de forma que o ouvinte ou o leitor entendam o que foi feito e julgam se é razoável. Portanto, o conhecimento estatístico e o conhecimento geral estão inter-relacionados.

Gal (2002) ainda explica que, um pré-requisito para entender e interpretar conceitos estatísticos é ter o conhecimento desses conceitos e procedimentos básicos bem como dos conceitos e temas matemáticos relacionados. Porém, quase todos os autores que estão preocupados com a capacitação dos adultos ou dos graduandos nas universidades não discutem que conhecimento é necessário para atingir o letramento estatístico e, geralmente limitam-se ao que deve ser ensinado nas escolas, afirmando que, em todos os graus das escolas secundárias (ou de universidades), os estudantes deveriam dominar um conjunto de temas estatísticos e assumindo que isso garantirá o conhecimento estatístico na fase adulta.

Além de concordarmos com o autor, defendemos a hipótese de que, durante o processo de construção do letramento estatístico, outras questões didáticas devem ser abordadas, as quais vão além de decorar regras e fórmulas.

O teórico Gal (2002) apontou as cinco características que devem servir de base para o conhecimento estatístico:

- (1) Conhecimento do porquê os dados são necessários e da forma como podem ser produzidos. Em geral, os adultos devem ter algum conhecimento sobre as origens dos dados que são apresentados em resultados ou representações gráficas, entender que é necessário saber como os dados foram extraídos e estar cientes de que um bom delineamento do projeto contribui para responder perguntas específicas;
- (2) Familiaridade com os termos e as ideias relacionadas com a estatística descritiva. Os adultos devem estar familiarizados com os conceitos básicos e apresentações de dados utilizados para transmitir os resultados ao público para os quais esses relatórios são abordados.
- (3) Familiaridade com os termos e ideias básicas relacionados às representações gráficas e tabulares. Espera-se que os adultos saibam que os

dados podem ser apresentados em gráficos e tabelas, os quais servem para organizar as múltiplas informações e permitir a detecção ou a comparação dos dados de tendência (TUFTE, 1997 apud GAL, 2002);

(4) Compreensão das noções básicas de probabilidade. Em muitos relatórios estatísticos, há a presença de informações probabilísticas em contexto, assim, espera-se que o adulto saiba compreender conceitos de eventos que têm a ver com ideias de acaso e aleatoriedade;

(5) Conhecimento de como obter as conclusões estatísticas ou inferenciais. Por conta da diversidade de modelos para coletar dados, os adultos devem ter uma ideia de como analisar os dados, chegar a um consenso e estar cientes dos problemas pertinentes da área. Devem ter domínio para entender possíveis erros em amostras, medições e inferências, percebendo que esses erros podem ser controlados pelo delineamento adequado do experimento.

Para o conhecimento matemático, MOORE, 1998 apud GAL, 2002, comentou que a determinação dos tipos de conhecimentos matemáticos esperados de adultos deve ser feita com cautela. Por um lado, os adultos devem estar claramente cientes de alguns dos procedimentos matemáticos subjacentes à produção de indicadores estatísticos comuns, como porcentagem ou média. Ao mesmo tempo, as expectativas em relação à quantidade e o nível de matemática formal necessários para compreender as ideias básicas de estatística que são ensinados em nível introdutório em universidades (ou secundário) apresentaram uma tendência de mudança nos últimos anos. Associado a isso, nos últimos anos, os estatísticos tem esclarecido a natureza de algumas diferenças fundamentais entre Matemática e Estatística e fizeram algumas hipóteses de trabalho sobre o nível geral de Matemática necessário para aprender Estatística, pelo menos no nível introdutório nas universidades.

Com relação ao conhecimento de contexto, Gal (2002) comentou que a interpretação adequada das mensagens estatísticas pelos adultos depende de sua capacidade de colocar as mensagens no contexto e de ter acesso a seu conhecimento do mundo. Esse tipo de conhecimento também apoia os processos de conhecimento geral e é fundamental para permitir dar sentido a qualquer tipo mensagem.

Moore (1997) argumentou que, na Estatística, o contexto motiva os procedimentos; os dados devem ser vistos como números com um contexto; e, portanto, o contexto é a fonte do significado e a base da interpretação dos resultados obtidos. No entanto, em contextos de leitura, as pessoas não se envolvem na geração de dados ou no desempenho de cálculos e análises. Sua familiaridade com os processos de geração de dados (como o design de estudos, o plano de amostragem e os questionários usados) ou com os procedimentos usados por pesquisadores ou estatísticos para analisá-los depende dos detalhes e da clareza, das informações fornecidas nas mensagens apresentadas.

O conhecimento do contexto é o principal determinante da familiaridade que o leitor tem com as fontes de variação e erro. Se ele ou um ouvinte não está familiarizado com o contexto no qual os dados são coletados, torna-se mais difícil imaginar por que uma diferença entre grupos pode ocorrer, quais interpretações alternativas podem existir para as conclusões apresentadas sobre a associação detectada entre determinadas variáveis ou como um estudo pode produzir resultados equivocados.

De acordo com Gal (2002) algumas perguntas devem ser consideradas fundamentais no acesso às informações estatísticas. O autor menciona que, com essas perguntas bem definidas, os adultos são capazes de adquirir habilidades para a validação das mensagens e a credibilidade das evidências implícitas nas informações e nas conclusões apresentadas, e a reflexão sobre possíveis interpretações alternativas das conclusões que elas lhes transmitem. Assim, os adultos devem ter em mente uma lista de “questões principais” acerca das informações a serem comunicadas ou mostradas. Fazemos a hipótese de que o mesmo ocorre com os jovens.

Bem definidos os elementos de conhecimento, podemos entender os elementos de disposição, que são divididos em dois componentes: “crenças e atitudes” e “postura crítica”.

Certas crenças e atitudes destacam a posição crítica das pessoas e sua disposição em fazer um esforço mental ou assumir riscos ocasionalmente como parte dos atos do conhecimento estatístico. Há um desafio em termos de definições ao se discutir “crenças” e “atitudes”, já que a distinção entre elas é um pouco vaga. Pesquisadores, por exemplo, geralmente definem implicitamente atitudes ou crenças estatísticas como algo que pode ser medido com seu instrumento de avaliação

favorito no contexto de uma população-alvo específica, como estudantes universitários ou adultos em geral. As habilidades metacognitivas mais amplas que são consideradas parte do funcionamento intelectual geral de indivíduos podem fornecer um apoio maior ao comportamento “*cult*” estatisticamente, como ter a propensão ao raciocínio lógico, curiosidade e pensamento de uma mente ampla.

Por fim, Gal (2002) apontou que, em condições de incerteza, a disposição dos adultos para recorrer a ações quando se deparam com informações ou mensagens estatísticas pode ser necessária. Exige-se, então, uma postura crítica bem clara e objetiva. Os adultos devem sentir-se seguros para explorar, supor e criticar, ou seja, criar um mecanismo em que fiquem confortáveis para estar na posição de um leitor ou para ouvir críticas, e acreditar na sua capacidade de argumentar como uma condição para desenvolver e sustentar sua motivação para a ação crítica.

Finalmente, chegamos ao ponto em que a “posição crítica” e as “crenças e atitudes” se encaixam. Para manter uma posição crítica, os adultos precisam desenvolver uma crença na legitimidade da ação crítica. É necessário que os leitores mantenham a ideia de que é legítimo ser crítico em mensagens ou argumentos estatísticos, de fontes oficiais ou outras fontes, por mais respeitáveis que sejam. Os adultos deveriam estar de acordo que é legítimo preocupar-se com qualquer aspecto de um estudo apresentado ou da interpretação de resultados e trazer questões relevantes, mesmo que não tenham aprendido muito sobre matemática ou estatística formal, ou que não tenham acesso a todos os detalhes necessários do contexto.

Para esta pesquisa, adotamos esse modelo proposto por Gal e concordamos em adotar os níveis de letramento propostos por ele a partir dos estudos de Shamos (1995) e investigamos tais níveis durante a aplicação das atividades didáticas.

2.3 Níveis de funcionamento dos conhecimentos mobilizados pelos alunos

Em nossas atividades, analisamos os níveis dos conhecimentos mobilizados pelos alunos. Robert (1998) propõe ferramentas de análise de mobilização de noções matemáticas por alunos de escolas secundárias, levando em conta sua especificidade e complexidade, os programas de ensino, as expectativas institucionais, e as hipóteses sobre o ensino e aprendizagem que admitimos e/ou que queremos colocar em jogo. A autora organizou a sua proposta em quatro

dimensões. As três primeiras mostram as características diretamente ligadas às noções que se pretendem ensinar, tal e como aparecem nos programas. São elas: ferramenta/objeto, quadros, registros; status de noções a ensinar quanto a sua inserção na paisagem matemática dos alunos; e os níveis de conceituação. A dimensão que será utilizada nesta pesquisa refere-se a identificar diferentes maneiras possíveis de os alunos colocarem em funcionamento tais noções para a resolução de atividades, abordando os níveis de mobilização dos conhecimentos pelos alunos.

Em nossa pesquisa, utilizamos a quarta dimensão da proposta de Robert, em que a autora considerou três níveis que podem ser utilizados em um nível escolar específico. Em todas as nossas atividades, estudamos a mobilização desses níveis de conhecimento dos alunos.

2.3.1 O nível de conhecimentos técnicos

O nível de conhecimentos técnicos corresponde, segundo Robert (1998), às mobilizações indicadas, isoladas, que explicitam aplicações imediatas de teoremas, propriedades, definições, fórmulas etc. Tratam-se, então, de contextualizações simples, locais e sem etapas, trabalhos preliminares de reconhecimento e adaptações, referindo-se ao funcionamento de ferramentas conceituais ou procedimentais que compreendem definições.

Nesse nível, a resolução de um problema está vinculada a uma técnica, como a aplicação de uma fórmula. Em uma aula de Estatística o professor apenas expõe o conteúdo, com as regras e fórmulas, sem se preocupar com os conceitos e contexto envolvidos. Nessa fase, em uma aula sobre métricas estatísticas, o docente apresenta uma série de fórmulas e regras e, por exemplo, pede para o aluno calcular a média, moda e mediana de um conjunto de dados. O aluno consegue aplicar tais fórmulas e resolver corretamente o problema, porém não tem habilidades para interpretar os resultados, saber diferenciar os conceitos de média e mediana ou mesmo reconhecer como etapa de outra atividade.

2.3.2 O nível de conhecimentos mobilizáveis

Este nível corresponde, baseado no conceito de Robert (1998), a funcionamentos mais amplos: ainda indicados, mas que passam da simples aplicação de uma propriedade. Isso pode ser, por exemplo, porque é necessário adaptar seus conhecimentos para aplicar o teorema adequado ou mudar de ponto de vista ou de quadro² (com indicações), uma vez que é necessário aplicar várias vezes seguidas a mesma noção ou noções diferentes, em etapas sucessivas, ou articular duas informações de naturezas diferentes.

Em todo caso, o nível de conhecimentos mobilizáveis testa um funcionamento no qual existe um princípio de justaposição de saberes em um dado domínio, e inclusive de organização. Nele, não há somente aplicação simples, pois o que está em jogo é explícito, ou seja: um saber é dito mobilizável se, quando está bem identificado, é bem utilizado pelo aluno, mesmo se houve lugar para se adaptar ao contexto particular.

Percebemos que temos neste nível uma atuação mais abrangente do sujeito, devendo haver um processo maior de adaptação e mudanças. Nas atividades aplicadas, serão exigidos tais conhecimentos mobilizados por parte dos estudantes. É importante ressaltar que nessa etapa, segundo a autora, tal adaptação por parte do aluno deve ocorrer mediante a interação com o professor ou estar explícita no próprio enunciado do problema. Essa indicação é o que diferencia o conhecimento mobilizado do conhecimento disponível.

Como exemplo prático, podemos expor em uma aula de Estatística um conjunto de dados e, no enunciado solicitar ao aluno que os represente por meio de gráficos. Assim, esse estudante irá mobilizar seus conhecimentos e refletir sobre qual gráfico é o que melhor representa o seu conjunto de dados e houver interesse em trabalhar com a visualização de medidas de posição, como mediana e quartis e identificar dados *outliers*, o aluno deve propor a construção de um gráfico *boxplot*.

²Um quadro é constituído dos objetos de um ramo da Matemática, das relações entre esses objetos, suas formulações eventualmente distintas e das imagens mentais associadas a esses objetos e essas relações. [...] Dois quadros podem conter os mesmos objetos e diferirem pelas imagens mentais e pela problemática desenvolvida. (DOUADY, 1986. p. 11).

2.3.3 O nível de conhecimentos disponíveis

Robert (1998) explica que o nível de conhecimentos disponíveis corresponde ao fato de saber resolver o que está proposto sem indicações, de procurar em seus próprios conhecimentos o que pode intervir na solução. Fornecer contraexemplos (encontrar ou inventar), mudar de quadros sem sugestão (relacionar) e aplicar métodos não previstos são comportamentos.

Espera-se que o aluno utilize seus conhecimentos adquiridos em outras tarefas e que seja autônomo na realização de uma nova tarefa. Ou seja, nessa etapa, o estudante faz tudo o que tem no nível de conhecimentos mobilizados, mas de forma autônoma e sem indicações. Por exemplo: em uma coleta de dados, deve ter competência para selecionar uma amostra representativa de uma população e, baseando-se nessa amostra, é necessário que saiba construir uma tabela de frequência que o auxilie a responder a uma determinada questão ou um gráfico adequado para o seu conjunto de dados.

Também podemos utilizar de exemplo uma aula de variabilidade em que é proposto aos alunos um conjunto de dados, os quais possuem as notas de dez estudantes em quatro turmas distintas.

Tabela 1 – Notas de alunos de quatro turmas de uma universidade

Aluno	Turma A	Turma B	Turma C	Turma D
1	8,5	7,0	5,0	7,0
2	10,0	6,0	5,5	6,5
3	7,0	9,5	8,0	8,0
4	7,0	6,0	9,0	10,0
5	9,0	5,0	8,0	9,5
6	7,5	8,0	9,5	9,0
7	6,5	8,5	9,0	4,0
8	9,0	8,0	4,0	5,5
9	10,0	7,5	3,5	6,0
10	8,5	5,0	9,5	7,0
Média	8,3	7,1	7,1	7,3

Fonte: o autor (2019).

Nesse problema, pede-se que (1) calcule as medidas resumo para as notas das quatro turmas e (2) explique como seria identificado o comportamento da variação das notas dos alunos dessas quatro turmas e como argumentaria qual turma obteve o melhor desempenho. Dado o enunciado, o aluno calcularia as medidas resumo: média, mediana e moda facilmente e, ao falar em variabilidade, a variância e o desvio padrão, também sem grandes dificuldades. Isso caracteriza o nível de conhecimento técnico.

Porém, uma vez que as relações entre as medidas e a interpretação dos resultados não está explícito no enunciado, o aluno teria de atuar de forma autônoma e construir o seu raciocínio para chegar a uma conclusão. Se o aluno percorrer esse caminho na sua completude, dizemos que ele atingiu o nível de conhecimento disponível.

3 ESTATÍSTICA DESCRITIVA

3.1 O que é estatística descritiva?

Neste capítulo discutimos aspectos da estatística descritiva, temas relevantes para a aplicabilidade das atividades propostas neste trabalho.

De acordo Lima e Magalhães (2004, p.2), a estatística descritiva é, em geral, utilizada na etapa inicial da análise, quando tem-se o contato com os dados pela primeira vez. Objetivando tirar conclusões de modo informal e direto, a maneira mais simples seria a observação dos valores colhidos. Entretanto, ao depararmos com uma grande massa de dados, percebemos, imediatamente, que a tarefa pode não ser simples. Assim, para tentar depreender dos dados a respeito do fenômeno sob estudo, é necessário aplicar alguma técnica que permita resumir a informação daquele conjunto de valores. Em outras palavras, a estatística descritiva pode ser definida como um conjunto de técnicas destinadas a descrever e resumir os dados, a fim de que possamos tirar conclusões a respeito de características de interesse.

Batanero (2001, p. 10) mencionou que esta estatística tem por objetivo apresentar um resumo de um conjunto de dados e expor suas características, mediante as representações gráficas. O interesse central é descrever o conjunto de dados e não estender as conclusões para outro conjunto ou outra população.

Entendemos que uma etapa extremamente importante para a análise descritiva é a organização dos dados. Lima e Magalhães (2004) apontaram que a questão central é como “tratar” os valores numéricos ou não, a fim de extrair informações a respeito de uma ou mais características de interesse, lembrando-se de que a natureza dos dados deve ser considerada para todo esse procedimento. Para isso, é necessário fazer uso de tabelas de frequências, métricas descritivas e gráficos e, por fim, entender o conceito de variável e os tipos existentes.

Cordani (2001) comentou sobre a importância da organização dos dados, com conceitos de padrão e regularidades e seus objetivos serem pautados em classificar variáveis e descobrir relações entre elas. Além disso, a autora aborda as disciplinas introdutórias de Estatística:

A organização dos dados exige uma padronização que facilita a operação à tomada de decisão mediante potentes recursos computacionais. A disciplina introdutória de Estatística, da forma como é desenvolvida atualmente, não

está preparada para colocar em discussão este conceito. Uma reforma nesta disciplina, que atenda às expectativas do novo século, dependerá fundamentalmente do preparo dos professores. (CORDANI, 2001, p.5).

Acreditamos que essa etapa de organização dos dados é fundamental bem como defendemos a hipótese de que um aluno letrado estatisticamente será capaz de reconhecer tal etapa e naturalmente trabalhar corretamente com o banco de dados disposto.

3.2 Tipos de Variáveis

Conforme apontam Lima e Magalhães (2004), variável é qualquer tipo de característica associada a um tipo grupo ou população, podendo ser classificada como quantitativa ou qualitativa.

Para ilustrarmos o entendimento, em uma sala de aula, temos as características dos alunos, tais como o peso, a altura e a idade. A variável altura, por exemplo, pode assumir os valores em metros de 1,60 a 1,85. Deve-se considerar esses fatos nas análises assim como dois tipos de variáveis: numéricas e não numéricas. As numéricas serão denominadas quantitativas, ao passo que as não numéricas, qualitativas.

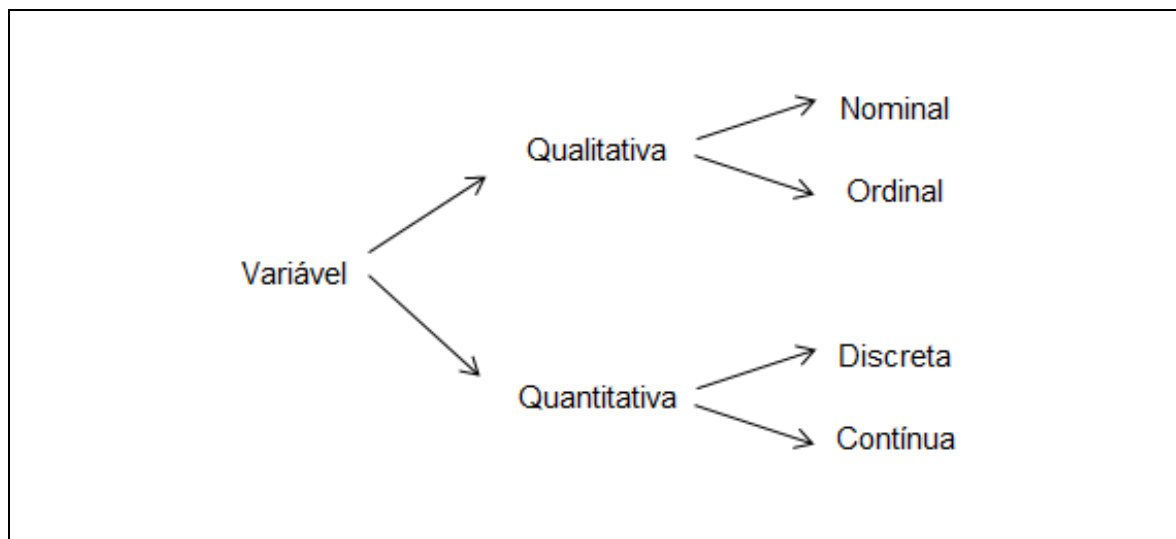
A variável é qualitativa quando os possíveis valores que assume representam atributos e/ou qualidades. Se as variáveis têm uma ordenação natural, indicando intensidades crescentes de realização, serão classificadas como qualitativas ordinais. Quando não é possível estabelecer uma ordem natural entre seus valores, são classificadas como qualitativas nominais. Por exemplo, variáveis como tipo da turma (A ou B), sexo (feminino ou masculino) e fumante (sim ou não) são qualitativas nominais; já variáveis como tamanho (pequeno, médio ou grande) e classe social (baixa, média ou alta), são qualitativas ordinais.

A variável quantitativa tem natureza numérica e está subdividida em discretas e contínuas. Variáveis quantitativas discretas podem ser vistas como resultantes de contagens, assumindo geralmente valores inteiros. De uma maneira mais formal, o conjunto dos valores assumidos é finito ou enumerável. Já as variáveis quantitativas contínuas, assumem valores em intervalos dos números reais e, na maioria das vezes, são provenientes de uma mensuração. Por exemplo, número de irmãos (0, 1,

2, ...) e número de defeitos (0, 1, 2, ...) são variáveis quantitativas discretas, enquanto peso e altura são contínuas.

Abaixo apresentamos uma figura que demonstra o esquema dos tipos de variáveis:

Figura 1 – Classificação de variáveis



Fonte: Lima e Magalhães (2006, p.8).

3.3 Medidas Descritivas

Nesta seção, iremos entender quais são as principais medidas descritivas que buscam resumir as informações de um determinado conjunto de dados. Para estudá-las, iremos dividi-las em dois blocos: as medidas de posição e as medidas de dispersão.

3.3.1 Medidas de posição

As medidas de posição mais conhecidas são a média, a moda, a mediana e os quartis. A média é, de modo geral, o ponto de equilíbrio dos desvios da distribuição; a mediana é uma medida posicional; a moda, o valor que mais se repete em uma distribuição de um conjunto de dados; e os quartis são medidas que dividem um conjunto de dados em quatro partes.

3.3.1.1. Média

A média é uma das principais medidas de tendência central e na prática, consideramos que ela é o resultado da soma de todas as informações do conjunto de dados dividida pelo número de informações que foram consideradas.

Por ser robusta sob o ponto de vista matemático, uma vez que considera em seu cálculo todos os valores da distribuição e, para a interpretação, as discrepâncias entre eles, a média é a medida mais utilizada (NOVAES, 2004, p. 38). Embora não seja adequada para os casos de grande variabilidade nos dados, é utilizada nos cálculos estatísticos avançados.

Assim, vale ressaltar que, essa medida deve ser utilizada e analisada com cuidado, dado que ela é mais adequada para situações em que os dados estejam distribuídos de forma uniforme, ou seja, sem grandes discrepâncias. A principal aplicação da média é identificar a representatividade de um elemento em um conjunto de dados, ou seja, um elemento que resuma adequadamente o conjunto de dados analisado.

No nível de mobilização do conceito desejado em um processo de aprendizagem eficaz, pode-se esperar que o aluno faça mais do que a simples aplicação da fórmula para o cálculo de média simples ou ponderada, identificando suas propriedades e aplicabilidade na situação analisada. Assim, a média para um conjunto de valores de dados de uma observação é feita pela soma dos valores dividida pelo número total de observações, conforme vemos na fórmula a seguir:

$$X_{obs} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

3.3.1.2. Mediana

A mediana é uma medida de localização que identifica o centro da distribuição dos dados, dividindo o conjunto em dois subconjuntos com exatamente o mesmo número de elementos. Para encontrar o valor da mediana é necessário colocar os valores em ordem crescente ou decrescente.

Quando o número elementos de um conjunto é ímpar, o número central é o valor da mediana, já quando o número elementos de um conjunto é par, a mediana é

encontrada pela média dos dois valores centrais. Assim, esses valores são somados e divididos por dois.

Por exemplo, se em uma sala de aula, tivermos 9 alunos e as alturas desses alunos, forem: 1,54 m, 1,67 m, 1,50 m, 1,65 m, 1,75 m, 1,69 m, 1,60 m, 1,55 m e 1,78 m, onde de m é a medida em metros, para determinarmos o valor da mediana, primeiramente, devemos colocar os valores em ordem crescente, assim, temos: 1,50 m; 1,54 m; 1,55 m; 1,60 m; 1,65 m; 1,67 m; 1,69 m; 1,75 m; 1,78 m. Como nesse conjunto de dados o número de elementos é ímpar (9 elementos), a mediana é o elemento central, ou seja, o quinto elemento, assim a mediana é 1,56 m.

3.3.1.3. Moda

Como já foi citado, a moda é o valor que mais se repete em uma distribuição de um conjunto de dados, sendo assim, para identificá-la basta observar a frequência com que os valores aparecem.

Por exemplo, em uma sala de um curso universitário de Medicina, temos 20 alunos e a distribuição da idade destes alunos é dada por: 18, 18, 19, 19, 21, 23, 21, 19, 19, 19, 19, 21, 22, 19, 18, 19, 18, 19, 20 e 19. Podemos observar que 10 dos 20 alunos tem 19 anos de idade, assim consideramos que a moda da idade dos alunos deste curso é 19 anos, ou seja, essa é a idade do maior número de alunos.

3.3.1.4. Quartil

Por fim, o quartil é a medida que dividem um conjunto de dados em quatro partes com o mesmo número de elementos, como mencionado anteriormente.

Para isso, os dados precisam estar em ordem crescente. O primeiro quartil é o número que deixa 25% das observações abaixo e 75% acima, enquanto que o terceiro quartil deixa 75% das observações abaixo e 25% acima. Já segundo quartil é a mediana e deixa 50% das observações abaixo e 50% das observações acima.

3.3.2 Medidas de Dispersão

Para identificarmos a variabilidade de dados, utilizamos as medidas de dispersão, sendo que as principais são: o desvio padrão, a variância e a amplitude.

Tanto a variância quanto o desvio padrão nos auxiliam a identificar a dispersão do conjunto de dados em relação à média. Por uma questão interpretativa, ao trabalharmos com tais medidas, recomendamos o uso do desvio padrão, uma vez que está na mesma escala que os dados.

3.3.2.1. Variância

A variância é uma medida de dispersão que, em um conjunto de dados, mostra o quanto cada elemento está distante do valor central (da média). Assim, quanto menor é o seu valor, mais próximos os valores estão da média. A variância populacional é dada pela seguinte expressão:

$$\text{Variância } (X) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

Quando desejamos determinar a variância para uma amostra da população, ou seja, quando temos uma parte de uma população, devemos considerar esta expressão:

$$\text{Variância } (X) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$$

3.3.2.2. Desvio padrão

O desvio padrão, nada mais é do que a padronização da variância, ou seja, é determinada pela raiz quadrada da variância.

$$\text{Desvio Padrão } (X) = \sqrt{\text{Var}(X)}$$

Assim, tanto a variância quanto o desvio padrão nos auxiliam a identificar a dispersão do conjunto de dados em relação à média. Por uma questão interpretativa, ao trabalharmos com tais medidas, recomendamos o uso do desvio padrão dado que está na mesma escala que os dados.

3.3.2.3. Amplitude

A amplitude referente a certa variável é definida como a diferença entre o maior e o menor valor do conjunto de dados (LIMA; MAGALHÃES, 2004). A amplitude só leva em conta dois valores de todo o conjunto e, assim, seria mais conveniente considerarmos uma medida que utilizasse todas as observações.

Uma ideia inicial é considerar o desvio de cada observação em relação a um ponto de referência e, então, tomar sua média. Caso a observação seja menor do que a referência, o desvio será negativo; caso seja maior, positivo. A soma de tais desvios fará com que termos de sinais diferentes se compensem, podendo ocultar o efeito da variabilidade.

3.4 Distribuição de frequência - representação tabular

Quando desejamos descrever um determinado conjunto de dados, é necessário verificar a frequência dos diversos valores existentes da variável. Assim, definimos o que é frequência simples (ou absoluta): o número de vezes que o valor de uma variável (qualitativa ou quantitativa) foi observado (denotaremos por “fi”). Dessa forma, determinando por “n” o total de elementos, podemos considerar que a soma de todas as observações é dada pela expressão a seguir, em que “k” é o número dos valores diferentes existentes da variável.

$$\sum_{i=1}^k f_i = n, \text{ onde } k \text{ é o número dos valores diferentes existentes da variável.}$$

A distribuição de frequência é a função empírica que associa, a cada valor da variável, o fator observado de sua frequência absoluta.

Com isso, definimos que a frequência relativa (denotada por “fr”) é a proporção de um dado valor de uma variável (qualitativa ou quantitativa), como exemplificado abaixo:

$$f_r = \frac{f_i}{n}$$

Por fim, definimos que a frequência absoluta acumulada (“Fac”) é a frequência acumulada até o intervalo que temos interesse em observar.

Para ilustrar o que foi exposto anteriormente, vamos considerar o seguinte conjunto de dados: idade dos alunos do curso de Enfermagem da Universidade Federal do Estado do Rio de Janeiro (UNIRIO).

Tabela 2 – Idade dos alunos do curso de Enfermagem

Idade	Frequência (fi)	Frequência Relativa (fr)	Frequência Acumulada (Fac)
19	5	28%	28%
20	6	33%	61%
21	4	22%	83%
23	1	6%	89%
24	1	6%	94%
25	1	6%	100%
TOTAL	18	100%	100%

Fonte: o autor (2019).

Na tabela de frequência acima, identificamos 18 alunos no curso de Enfermagem da UNIRIO. Além disso, a faixa etária (amplitude) desses é de 19 a 25 anos, sendo que a maior frequência observada (moda) é de alunos com 20 anos (6 alunos, representando 33% das observações). Observamos que a frequência relativa desses alunos com 20 anos é de 33% e a frequência acumulada até esse valor é de 61% (acumulado de 19 e 20 anos).

3.5 Representações gráficas

Nesta seção, abordamos os gráficos mais usuais para as variáveis qualitativas e quantitativas.

A fim de representar as qualitativas, os gráficos mais usuais são de barras e decomposição em setores (conhecidos como “gráficos de pizza”). Utilizando o conjunto de dados composto por perfil de alunos dos cursos da UNIRIO e observando a quantidade de alunos por sexo, elaboramos a tabela 3.

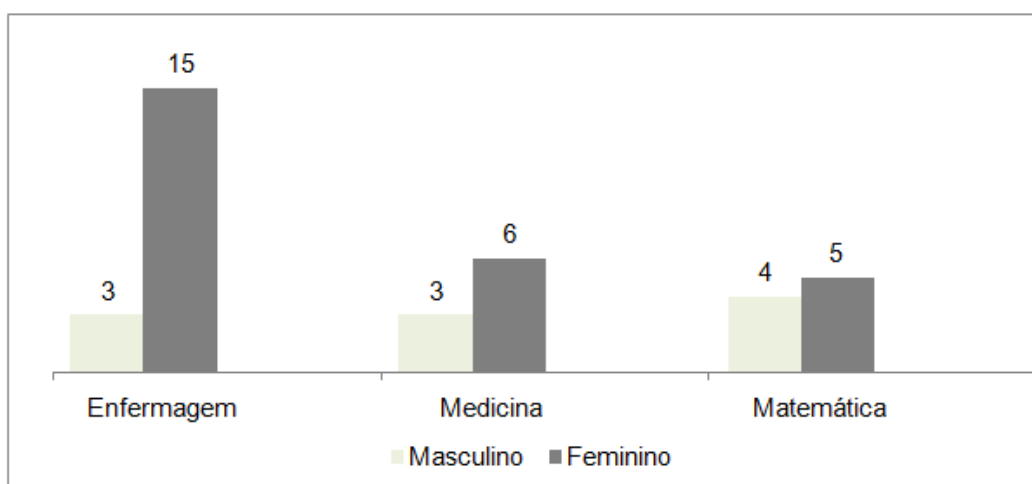
Tabela 3 – Perfil dos alunos da UNIRIO por curso (por sexo)

Curso	Quantidade de Alunos		
	Masculino	Feminino	Total
Enfermagem	3	15	18
Medicina	3	6	9
Matemática	4	5	9
Total	10	26	36

Fonte: o autor (2019).

Para uma visualização gráfica, observamos a quantidade de alunos (por sexo) e comparamos os cursos.

Figura 2 – Gráfico de barras - distribuição dos alunos por curso e sexo



Fonte: o autor (2019).

Outros dois gráficos importantes, que também serão trabalhados nas atividades didáticas deste estudo, são o histograma e o gráfico setorial (ou “gráfico de pizza”).

Supondo que queremos analisar a performance de sete vendedores em uma loja de brinquedo, sabendo que o faturamento total do estabelecimento foi de R\$ 55.000 no mês, montamos uma tabela com a venda de cada funcionário:

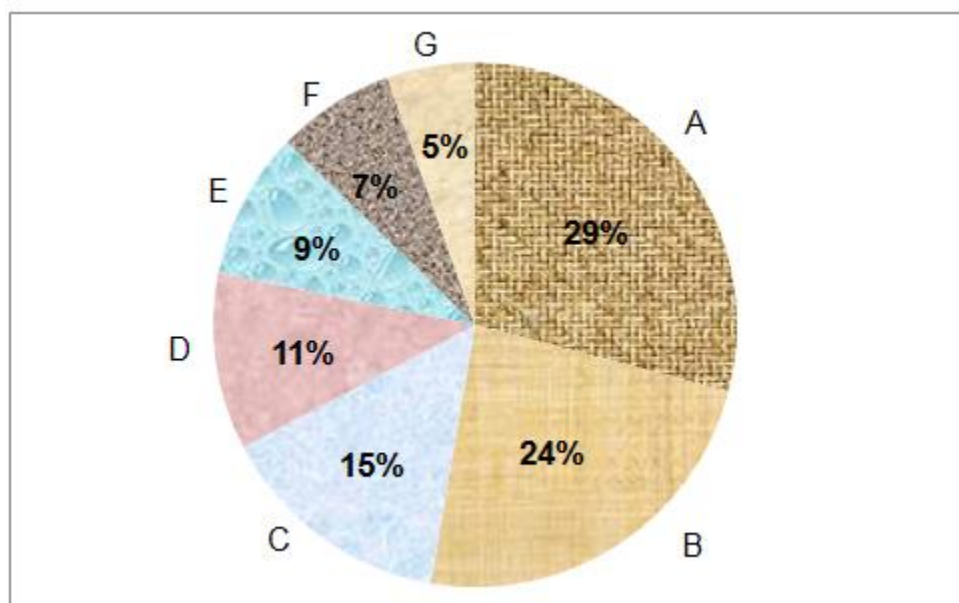
Tabela 4 – Vendas de 7 funcionários em uma loja de brinquedo

Funcionário	Valor - R\$	% Valor
A	16.000	29%
B	13.000	24%
C	8.000	15%
D	6.000	11%
E	5.000	9%
F	4.000	7%
G	3.000	5%
TOTAL	55.000	100%

Fonte: autor (2019).

Baseando-se na tabela 4, podemos construir um gráfico de setor, conforme ilustrado abaixo:

Figura 3 – Gráfico de setores - performance de 7 vendedores



Fonte: o autor (2019).

Assim, analisando o gráfico, identificamos que o vendedor A foi o que apresentou melhor performance na venda dos brinquedos da loja.

Agora, demonstramos a construção de um histograma, um gráfico de frequência que tem o objetivo de caracterizar um conjunto de dados organizados em intervalos de classes. Para isso, vamos considerar a altura de 55 alunos de um curso de graduação.

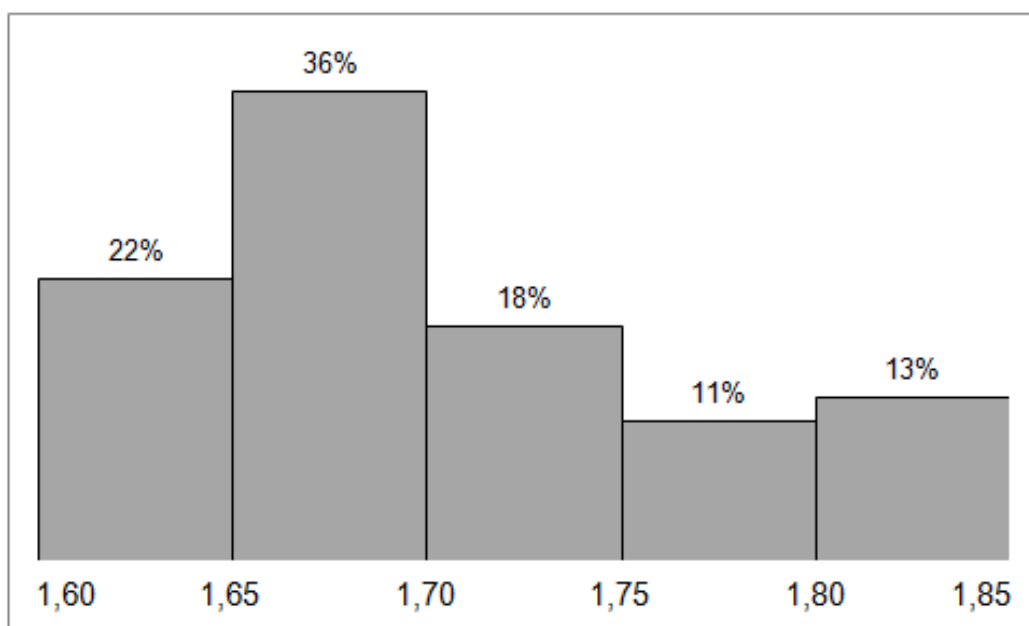
Tabela 5 – Distribuição das alturas de 55 alunos de um curso de graduação

Altura (metros)	Frequência (fi)	Frequência Relativa (fr)
[1,60 ; 1,65]	12	22%
[1,66 ; 1,70]	20	36%
[1,71 ; 1,75]	10	18%
[1,76 ; 1,80]	6	11%
[1,81 ; 1,85]	7	13%
Total	55	100%

Fonte: o autor(2019).

Com base nesses dados, é possível construir um histograma conforme ilustrado abaixo:

Figura 4 – Histograma da altura de alunos de um curso de graduação



Fonte: o autor (2019).

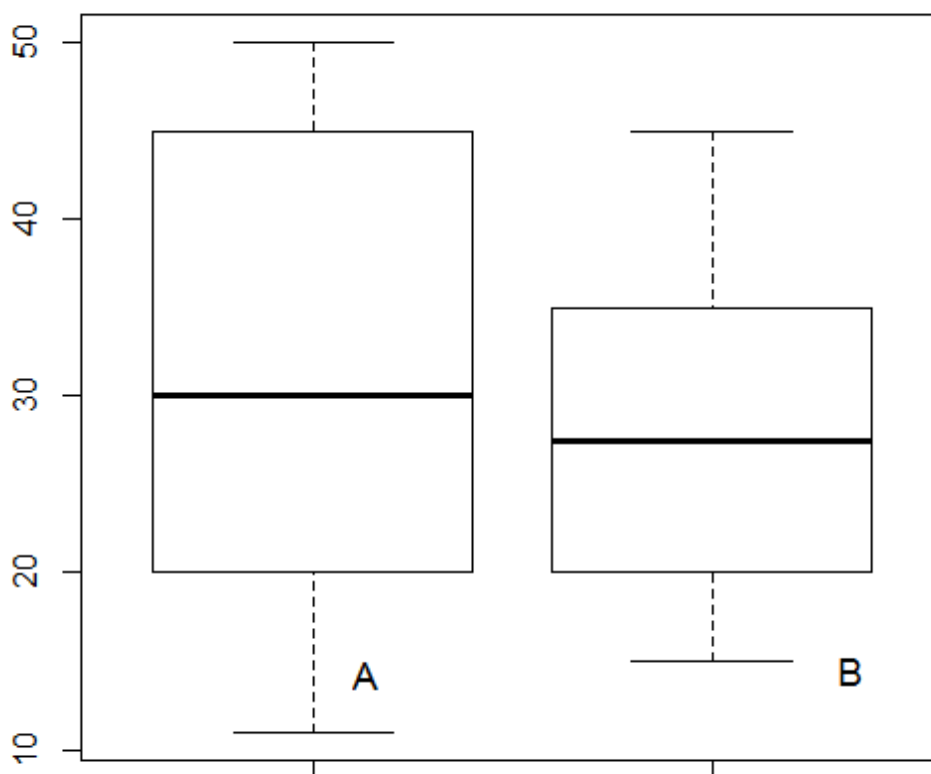
Por fim, analisamos o gráfico *boxplot* (ou “gráfico de caixas”), utilizado para avaliar a distribuição empírica dos dados, formado pelo primeiro e terceiro quartil e pela mediana. As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior ao limite inferior, e do quartil superior até o maior valor não superior ao limite superior.

Agora, supondo que temos dois conjuntos de dados, decidimos construir um *boxplot* para comparar os quartis, mediana e amplitude. Em dois consultórios médicos (A e B), foram avaliados dez pacientes, em cada, com as idades abaixo. Neste caso, é solicitado que seja realizada a construção de um *boxplot* para cada um deles.

A = {11, 17, 20, 21, 25, 35, 40, 45, 48, 50}

B = {15, 18, 20, 22, 25, 30, 32, 35, 38, 45}

Figura 5 – Exemplo de um *boxplot*



Fonte: o autor (2019).

Nesse *boxplot*, observamos as idades dos dez pacientes avaliados nos consultórios A e B e identificamos que a mediana da idade dos pacientes do consultório A é maior do que a do B, bem como sua variabilidade. Também notamos que no consultório A os pacientes são mais velhos do que no consultório B.

Nesta pesquisa, abordamos os gráficos histograma e *boxplot* e esperamos que os alunos mobilizem seus conhecimentos para interpretá-los de forma assertiva.

No primeiro, temos o interesse em investigar se os alunos conseguem relacioná-lo às tabelas de frequências e às medidas de posição: média, mediana e moda. Para o segundo, que saibam relacioná-lo às medidas de quartis e tenham a habilidade de visualizar os quartis, a mediana e a amplitude dos dados, sempre associando a um contexto e construindo, adequadamente, seu letramento estatístico.

4 PROCEDIMENTOS METODOLÓGICOS

Abordamos neste capítulo os procedimentos metodológicos e a metodologia utilizados em nossa pesquisa. Durante a sequência de atividades aplicadas, empregamos como estratégia o uso de tecnologias digitais, em especial o *software* R. Essas atividades foram aplicadas na Universidade Federal do Estado do Rio de Janeiro (UNIRIO), em três cursos de graduação: Medicina, Enfermagem e Licenciatura em Matemática. Em relação à coleta de dados, nos valem das respostas na resolução das atividades, de filmagem e de gravação de áudio.

4.1 A tecnologia em contexto

A maioria das tecnologias é utilizada como auxiliar no processo educativo, não sendo nem objeto, nem sua substância ou finalidade (KENSKI, 2007, p. 45). Elas estão presentes em todos os momentos do processo pedagógico, desde o planejamento das disciplinas e elaboração da proposta curricular até a certificação dos alunos concluintes.

De acordo com o Relatório GAISE - *Guidelines for Assessment and Instruction in Statistics Education* (2016), os alunos devem ser capazes de interpretar e tirar conclusões de todas as saídas de um *software* estatístico. A análise moderna de dados envolve o uso de *softwares* estatísticos que sejam capazes de armazenar e analisar potencialmente grandes bancos de dados.

Na prática, devemos orientar os nossos alunos no desenvolvimento do letramento estatístico. São objetivos que se referem a como o aluno deve entender e analisar dados e conceitos estatísticos frente a um primeiro ano de um curso de Estatística básica. Assim, listamos os seguintes objetivos, que serão revisitados indiretamente em nossa pesquisa.

- (1) Os alunos devem se tornar críticos frente a resultados estatísticos reportados em mídia popular, identificando se os resultados apresentados estão condizentes após o estudo e análises feitas;
- (2) Os alunos devem ser capazes de reconhecer as questões para as quais o processo investigativo em Estatística seria útil e devem ser capazes de responder as perguntas usando o processo investigativo;
- (3) Os alunos devem aprender a trabalhar com gráficos e resumos numéricos e interpretar o que os gráficos representam ou não;
- (4) Os alunos devem reconhecer e saber explicar o papel da variabilidade no campo da Estatística;

- (5) Os alunos devem reconhecer e saber explicar o papel da aleatoriedade em estudos e tirar conclusões;
- (6) Os alunos devem ganhar experiência na construção de modelos estatísticos, incluindo modelos multivariados;
- (7) Os alunos devem demonstrar uma compreensão e a capacidade de usar ideias básicas de inferência estatística, testes de hipóteses e intervalos de confiança em diversas situações;
- (8) Os alunos devem ser capazes de interpretar e tirar conclusões de saídas dos pacotes estatísticos;
- (9) Os alunos devem demonstrar uma consciência em questões éticas associadas à prática Estatística.(GAISE, 2016, p.8) ³

Além disso, buscamos algumas pesquisas que abordam o tema da expansão do uso da tecnologia na educação estatística. Casini, Gonzales e Padro (2011) abordam o crescente número de estudantes nas universidades e a dificuldade do alcance da relação professor-aluno que permita o acompanhamento individual do processo de aprendizagem.

Associada a essa expansão, também houve a adaptação às mudanças introduzidas pelo novo sistema de tecnologias da informação e comunicação, as quais impactaram diretamente nos métodos tradicionais de ensino. Os autores mencionam que essas novas tecnologias aumentaram consideravelmente as informações disponíveis, alterando a metodologia para seu tratamento estatístico. Ademais, apontam que essas transformações estão, progressivamente, sendo incorporadas ao ensino universitário e devem assumir o desafio de treinar pessoas que estejam cada vez mais preparadas para desenvolver atividades no domínio da nova sociedade da informação.

Por isso, é necessário otimizar o grau de inovação, eficácia e avaliação correspondente, proporcionando aos alunos ferramentas intelectuais apropriadas para entender e operar junto às mudanças rapidamente produzidas.

Casini, Gonzales e Padro (2011) avaliaram as experiências na aplicação de tecnologias digitais para a aprendizagem da Estatística, descrevendo o papel da utilização de um *software* para a realização de atividades práticas e analisando as possibilidades da internet em uma dupla perspectiva: como complemento ao ensino presencial e como suporte de uma sala de aula virtual. Com base no exposto, gera-

³ Texto traduzido do GAISE College Report ASA Revision Committee, "Guidelines for Assessment and Instruction in Statistics Education College Report 2016", disponível em <http://www.amstat.org/education/gaise>.

se uma proposta metodológica que procura adequar o uso da tecnologia a diversas propostas curriculares pertencentes a universidades privadas e públicas, em vários estabelecimentos de ensino e em situações de marcada diferença no número de alunos a cargo do professor.

Também entendemos que, nesta pesquisa, devemos discutir aspectos da fluência em tecnologias digitais. Para isso, recorreremos ao artigo de Oliveira e Marcelino (2015), que comentam que existem quatro etapas principais para o professor empregar determinadas tecnologias no âmbito de suas estratégias didáticas: adquirir fluência nas tecnologias; pensar com elas; elaborar temas com base nas tecnologias; e criar estratégias didáticas a partir do conhecimento da lógica das mídias eleitas para seu trabalho.

Os autores ainda comentam sobre a fluência em relação à tecnologia e que, para o professor, a primeira fase de um processo de uso das mídias em sala de aula é dividida em duas etapas relacionadas.

Na primeira, ocorre a exploração dos elementos da interface, para adquirir desenvoltura nos instrumentos existentes, os quais podem permitir encaminhar a construção do que se deseja (gráficos, expressões etc.). Dominar as interfaces sempre foi importante, tanto em relação à oralidade quanto à escrita e às mídias digitais.

A segunda etapa, também bastante relevante, refere-se à apropriação da lógica da interface em uso, destinada a uma compreensão, que vai além daquela adquirida na etapa anterior sobre como a tecnologia utilizada trabalha com o aspecto matemático, “ou seja, como se dá a integração entre o conhecimento matemático, fundamental para a resolução de um problema, e a expressão desta resolução sob o ponto de vista da forma como a interface opera”.

Oliveira e Marcelino (2015) apontam, também, que a fluência tecnológica remete ao uso intensivo do computador nos processos de ensino ou de aprendizagem em Matemática, mas não como substituto do trabalho intelectual das pessoas (professores e alunos) nem como suplementar em relação à atividade mental humana.

De acordo com o trabalho de GOOS (2003) apud OLIVEIRA (2018) a criação de uma autêntica etimologia da mediação pessoas/tecnologias pode ser de quatro ordens distintas, sendo elas:

(1) A tecnologia como mestre: ocorre quando a falta de competência, técnica ou matemática, limita os sujeitos a uma gama bastante estreita de possibilidades, por um lado, a falta de conhecimentos matemáticos atrapalha a avaliação de resultados de atividades usando softwares bem como restringe, significativamente, os avanços a partir dos mesmos, por outro, as limitações relativas ao uso dos programas podem tornar muito difícil explorar temas matemáticos significativos;

(2) A tecnologia como serva: os recursos tecnológicos servem, predominantemente, como uma forma de mudar os suportes midiáticos de exposição, além de serem empregados como um meio confiável e rápido para substituir cálculos feitos com instrumentos, como lápis ou papel, ou mentalmente. Um símbolo desse tipo de mediação é o caráter de troca de lousa tradicional por projeções/lousas digitais sem qualquer proposta didática distinta das práticas reprodutivas tenha lugar - neste sentido, o professor continua propondo a reprodução de seus métodos de cálculo, por exemplo, valendo-se de um meio mais rápido e confiável;

(3) A tecnologia como parceira: a ideia predominante neste tipo de mediação é aumentar as possibilidades abertas, para que os estudantes tenham um papel ativo na própria aprendizagem, por meio de atividades que promovam experimentação e visualização. Outra possibilidade é de que as tecnologias sejam empregadas como elementos que permitam discussões em sala de aula;

(4) A tecnologia como extensão em si: consiste no desenvolvimento de competência tecnológica como parte integrante dos repertórios didáticos e/ou matemático dos sujeitos bem como inclui o desenvolvimento de expertise no uso de diversos recursos tecnológicos e formas de "pensar diferente", a partir do desenvolvimento da autonomia na construção do conhecimento matemático. (GOOS apud OLIVEIRA, 2018, p. 63).

Em concordância com os autores, aprofundamos o emprego da tecnologia com base nos itens (3) e (4), ou seja, utilizando-a como uma estratégia didática para potencializar a aprendizagem da Estatística com o uso do *software* R.

Ainda sobre a fluência, OLIVEIRA (2018) comentou que no âmbito de processos educacionais, ela pressupõe dois tipos de construção de conhecimento, os quais, em sua configuração ideal, ocorrem simultaneamente- ou em momentos muito próximos/conectados: a exploração dos elementos da interface e a apropriação da lógica de uso dos recursos disponíveis. São processos ligados aos cenários didáticos e sistemas de ensino, como aqueles presentes nas salas de aulas e ambientes virtuais, por exemplo, ou podem ocorrer a partir de uma proposta de autoaprendizagem.

O autor comentou que outra possibilidade aberta no âmbito da fluência consiste em elaborar estratégias didáticas e/ou de autoaprendizagem com tecnologias, como aprimoramento da possibilidade que se abre em relação à exploração de trajetórias investigativas ou mesmo como movimento advindo da reorganização do pensamento a partir do desenvolvimento de fluência no uso dos recursos.

Concordamos com o autor e levantamos a hipótese de que a fluência em um *software* computacional, especificamente nesta pesquisa o R, auxiliará o aluno em sua autoaprendizagem na busca da construção do letramento estatístico e entendemos que o aluno deva apropriar-se da sua utilização e relacioná-lo com os conceitos estatísticos de base.

4.2 O *software* R e o pacote *R Commander*

O R é uma ferramenta para tratamento estatístico de dados, que oferece gráficos bem desenhados e com ótima qualidade para publicação, podendo incluir símbolos matemáticos e fórmulas quando necessário (SOUZA, 2013, p. 87). O autor destaca, também, as vantagens do uso do pacote Rcmdr mais conhecido com *R Commander*, desenvolvido no ano 2000, visando contornar algumas dificuldades encontradas pelos usuários, principalmente no que se referia à sintaxe dos comandos. A vantagem da utilização desse pacote é a existência de vários menus que conduzem a procedimentos e comandos estatísticos, sem a necessidade de programá-los com a sintaxe usual do R, bem como de uma interface para a saída dos resultados, sendo que os gráficos também são gerados em janelas separadas. Além disso, o *software* está disponível de forma livre e código aberto.

Para nossa pesquisa, fizemos um levantamento dos cursos de graduação que possuísssem um curso introdutório de Estatística na grade curricular e utilizassem algum *software* estatístico nas aulas. Uma vez que optamos pelo *software* estatístico R por ser gratuito e para desmistificar o alto grau de complexidade da ferramenta que os usuários comentam, identificamos que os cursos de graduação da UNIRIO utilizavam a ferramenta, assim propusemos a aplicação de uma sequência de atividade aos coordenadores dos cursos de ciências exatas e biológicas.

Definidos o local e o público da nossa pesquisa, propomos as atividades, que incluirão cálculos de medidas descritivas, como por exemplo, a média, moda e a mediana, análises de tabelas de frequências e construção de gráficos, como por exemplo, histogramas, de dispersão e *boxplot* (gráficos de caixa).

4.3 A pesquisa

Nossa pesquisa é qualitativa do tipo estudo de caso. Araújo e Borba (2012), enfatizam que pesquisa qualitativa deve ter por trás uma visão de conhecimento que esteja em sintonia com procedimentos como entrevistas, análises de vídeos, etc. e interpretações. O que se convencionou chamar de pesquisa qualitativa, prioriza procedimentos descritivos à medida que sua visão de conhecimento explicitamente admite a interferência subjetiva, o conhecimento como compreensão que é sempre contingente, negociada e não é verdade rígida.

Referente a pesquisa de caso, Severino (2007), comenta que é uma pesquisa que se concentra no estudo de um caso particular, considerado representativo de um conjunto de casos análogos, por ele significativamente representativo. A coleta de dados e sua análise se dão da mesma forma que nas pesquisas de campo, no geral. O caso escolhido para a pesquisa deve ser significativo e bem representativo, de modo a ser apto a fundamentar uma generalização para situações análogas, autorizando inferências. Os dados devem ser coletados e registrados com o necessário rigor e seguindo todos os procedimentos da pesquisa de campo. Devem ser trabalhados, mediante análise rigorosa, e apresentados em relatórios qualificados.

A partir disso, identificamos que nossa entidade e público estão bem definidos: estudantes de cursos de graduação, frequentando disciplinas introdutórias de Estatística, que mobilizam seus conhecimentos com estratégias ao uso de tecnologias digitais e à visualização da construção do letramento estatístico.

Para a coleta de dados, utilizamos durante a consecução das atividades, filmagem e gravações, aplicação de questionários, além de nos embasarmos nas respostas dos estudantes, devidamente documentadas. A coleta de dados foi fundamental para percorrermos todas as etapas na investigação dos níveis de conhecimentos mobilizados pelo aluno e analisarmos a construção do letramento estatístico.

4.3.1 Local da pesquisa

Com relação ao local da pesquisa, selecionamos a Universidade Federal do Estado do Rio de Janeiro (UNIRIO), uma instituição federal brasileira de ensino superior, sediada no bairro da Urca, Rio de Janeiro - RJ.

Para a escolha, realizamos algumas pesquisas com a intenção de buscar uma instituição que possuísse, em seus cursos de graduação, uma abordagem da Estatística com o uso de estratégias de tecnologia digitais e, se possível, com enfoque no *software* R, buscando desmistificar a dificuldade do seu uso, comumente comentada pelos alunos. Após algumas pesquisas, identificamos que a UNIRIO trabalha especificamente com o R e aborda a Estatística de uma forma contextualizada e por meio de projetos. Assim, ficamos três dias na Universidade, abordando três cursos distintos de graduação, graças à parceria estabelecida com seus professores.

4.3.2 Participantes da pesquisa

Com relação aos participantes da pesquisa, aplicamos as atividades para 36 alunos de três cursos de graduação da UNIRIO: nove do curso de Licenciatura em Matemática, nove do curso de Medicina e dezoito da Enfermagem.

Esses estudantes estavam regularmente matriculados nos cursos e em contato com a primeira disciplina de Estatística na graduação. Além disso, todos já possuíam conhecimentos prévios do *software* R, dado que a pesquisa foi realizada no final do semestre em um curso introdutório de Estatística, em junho de 2018.

O ensino e aprendizagem da Estatística nesta universidade se dão por meio de projetos, com temas específicos e contextualizados por curso e as ementas permeiam os conceitos básicos de estatística, como por exemplo, medidas estatísticas (de posição e dispersão), tipos de variáveis (qualitativas e quantitativas), apresentação dos dados em gráficos e tabelas, noções de amostragem e testes de hipóteses. Ao final do semestre, os alunos devem apresentar os seus projetos desenvolvidos em formato de pôster.

Para a análise dos dados, os alunos foram organizados em grupos de três indivíduos. Abaixo, apresentamos a caracterização das turmas, conforme tabela 6:

Tabela 6 – Caracterização dos alunos (por curso)

Curso	Quantidade de Alunos			Idade	
	Masculino	Feminino	Total	Mínima	Máxima
Enfermagem	3	15	18	19	25
Medicina	3	6	9	19	21
Matemática	4	5	9	20	59
TOTAL	10	26	36		

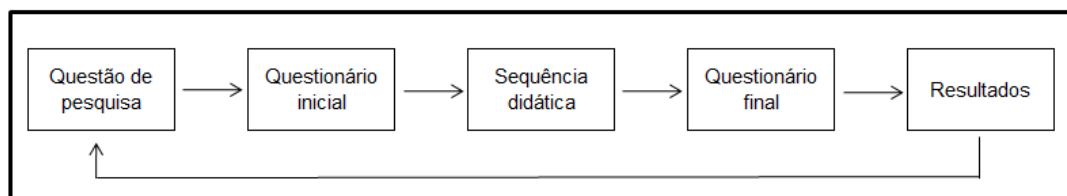
Fonte: o autor (2019)

4.3.3 Encontros e horários da coleta dos dados

Em cada turma, a sequência didática teve duração de duas horas e as atividades foram desenvolvidas em três etapas, conforme demonstrado na figura 6:

- (1) Primeira etapa: aplicação de um questionário sobre a percepção dos estudantes acerca da Estatística em seu âmbito pessoal e profissional, antes e depois de entrar na graduação.
- (2) Segunda etapa: aplicação de uma sequência didática, na qual foi explorado um banco de dados com uma problemática contextualizada, abordando um tema vinculado à sua formação profissional.
- (3) Terceira etapa: aplicação de um questionário, disponível no apêndice deste trabalho sobre as atividades da sequência didática, identificando os conceitos estatísticos mobilizados e a utilização do R pelos alunos.

Figura 6 – Etapas da pesquisa



Fonte: o autor (2019).

A figura 6 demonstra todas as etapas de pesquisa, sendo que ao final, os resultados foram confrontados com as questões de pesquisa, da forma como seus pontos questionados foram respondidos.

As atividades foram aplicadas e estruturadas de forma que abrangessem os principais conteúdos da estatística descritiva: exploração de banco de dados; cálculos e interpretação das principais medidas de posição e dispersão; e construção e interpretação de tabelas de frequências e gráficos (histograma, dispersão e *boxplot*).

Os dados observados foram extraídos das interações com os alunos e capturados por áudios, vídeos e fotos. Durante as atividades, os alunos utilizaram o *software* estatístico R, com o pacote *R Commander*.

4.4 Método: Pressupostos da Engenharia Didática

Como método da pesquisa, utilizamos os pressupostos da Engenharia Didática, de Michèle Artigue (1995). A autora expõe que, como método de pesquisa, a Engenharia Didática é caracterizada, em primeiro lugar, por um esquema experimental baseado nas “realizações didáticas” em aula, isto é, na construção, realização, observação e análise de sequências de ensino. Esse método também é caracterizado em comparação com outros tipos de pesquisa baseados em experimentação em classe, pelo registro em que se encontra e pelas formas de validação à qual está associado.

Segundo Almouloud (2007),

A engenharia didática pode ser utilizada em pesquisas que estudam os processos de ensino e aprendizagem de um dado objeto matemático e, em particular, a elaboração de gêneses artificiais para um dado conceito. Esse tipo de pesquisa difere daqueles que são transversais aos conteúdos, mesmo que o suporte seja o ensino de um certo objeto matemático (um saber ou um saber-fazer). (ALMOULOU, 2007, p. 171).

Segundo Artigue (1995), a Engenharia Didática apresenta quatro fases: (1) as análises prévias, (2) as análises *a priori*, (3) a experimentação e (4) as análises *a posteriori* e validação. Estudaremos cada uma das etapas a seguir.

4.4.1 Análises Prévias

Para a construção de uma pesquisa com Engenharia Didática, não se faz necessário apenas o quadro didático e o conhecimento didático anteriormente adquirido no campo de estudo mas também um conjunto de análises preliminares,

explica Artigue (1995). Segundo a autora, os pontos mais importantes que devem ser observados nessa etapa são: a análise epistemológica do conteúdo contemplado no ensino; a análise da educação tradicional e seus efeitos; a análise das concepções dos estudantes e das dificuldades e obstáculos que determinam sua evolução; e a análise das restrições em que será realizada a sequência didática.

Acerca do assunto, Almouloud (2007) expõe que:

Um dos objetivos das análises preliminares é identificar os problemas de ensino e aprendizado do objeto de estudo e delinear de modelo fundamentado a(s) questão(ões), as hipóteses, os fundamentos teóricos e metodológicos da pesquisa. (ALMOULOU, 2007, p,172).

O autor ainda complementa que essa primeira fase pode comportar as seguintes vertentes:

- (1) Deve-se analisar a organização matemática, em que está presente o estudo da gênese histórica do saber e suas manifestações, funcionalidades e obstáculos epistemológicos envolvidos; analisar a estrutura matemática do conceito investigado: analisar o ensino atual e seus efeitos; evidenciar os saberes (matemáticos) e os conhecimentos (matemáticos e/ou culturais ou pessoais relacionados; analisar as condições e fatores dos quais depende a construção didática efetiva das situações de ensino; e considerar os objetivos específicos da pesquisa.
- (2) Deve-se analisar a organização didática do objeto matemático escolhido: analisar as diferentes instituições de ensino em que o saber deve ser ensinado/aprendido bem como estudar a evolução do tratamento do conceito; fazer uma análise das propostas curriculares e PCN; analisar os livros didáticos; estudar as concepções de alunos e/ou professores a propósito dos saberes em jogo, aspecto que necessita da elaboração e da aplicação de um instrumento de coleta de dados; e levantar referências bibliográficas sobre os fatores que interferem nos processos de ensino e aprendizagem do objeto em questão.
- (3) Definição da questão de pesquisa: verificar se os estudos anteriores, abordados em (1) e (2) auxiliam o pesquisador na definição da(s) questão(ões) e hipóteses da pesquisa.

Como nosso objetivo era identificar se o uso do *software* potencializa a construção do letramento estatístico nos estudantes universitários, decidimos comparar a abordagem dos principais conceitos estatísticos em diferentes cursos de graduação. Dessa forma, escolhemos a abordagem de dois cursos da área de ciências biológicas (Enfermagem e Medicina) e de um da área de ciências exatas (Licenciatura em Matemática), identificamos a ementa das disciplinas e fizemos uma investigação prévia do perfil dos alunos.

Para o desenvolvimento das atividades, procuramos trazer temas vinculados às formações profissionais dos indivíduos, buscando trabalhar problemáticas que despertassem interesse nesses alunos.

4.4.2 Análises *a priori*

Esta etapa tem a finalidade de responder à questão levantada na fase anterior, como é apontado por Almouloud (2007). Para isso, o pesquisador deve elaborar e analisar uma sequência de situações problema. De acordo com o autor, essas situações devem levar em conta as seguintes características:

- Os alunos entendem facilmente os dados do problema e podem se engajar na resolução, usando seus conhecimentos disponíveis.
- Estas situações devem colocar em jogo um campo conceitual que se deseja efetivamente explorar e no qual o conhecimento está inserido.
- Os conhecimentos antigos dos alunos são insuficientes para a resolução completa do problema.
- Os conhecimentos, o objeto de aprendizagem, são as ferramentas que devem ser mobilizadas, em última instância, para obter a solução final.
- O problema pode envolver vários domínios de conhecimentos: álgebra, geometria, domínio numérico, entre outros. (ALMOULOUD, 2007, p. 174).

Segundo o autor, as atividades devem ser concebidas levando-se em consideração os resultados dos estudos preliminares e permitir aos alunos desenvolverem certas competências e habilidades, tendo por objetivo:

- Auxiliar o aluno na construção de conhecimentos e saberes de uma maneira construtiva e significativa.
- Desenvolver certas habilidades como, por exemplo, saber ler, interpretar e utilizar as diferentes representações matemáticas, bem como desenvolver o raciocínio dedutivo. (ALMOULOUD, 2007, p. 174).

O autor ainda afirma que a aplicação de cada atividade deve levar em consideração as seguintes condições:

- Os alunos devem mobilizar os objetos matemáticos de saber disponíveis como ferramenta explícita para resolver, pelo menos parcialmente, o problema.
- O professor/aplicador deve provocar um debate de confrontação dos resultados dos alunos. Nessa fase, diversas formas de saber podem aparecer. O objetivo visado é homogeneizar e construir o saber da classe, assim como promover o progresso, na aquisição individual os conhecimentos.
- É importante que o professor/aplicador, após o debate, selecione e organize as descobertas dos alunos e sistematize os novos conhecimentos e saberes, a fim de promover, para o aluno, uma melhor compreensão dos novos objetos matemáticos, Além disso, é preciso fazer a institucionalização dos saberes novos estudados.
- É imprescindível haver uma fase de familiarização, na qual o professor deve propor outras situações cujo objetivo é consolidar os novos conhecimentos. (ALMOULOU, 2007, p. 175).

Em nossa pesquisa, consideramos a universidade e os alunos de cursos diferentes de graduação como contexto. A UNIRIO foi o local para o desenvolvimento da pesquisa, com uso da tecnologia, como o *software* R e o pacote *R Commander*, para auxiliar na consecução das atividades didáticas, onde tivemos a participação dos alunos de diferentes cursos e investigamos a mobilização dos seus saberes matemáticos a fim de lidar com os conceitos estatísticos envolvidos.

4.4.3 Experimentação

Almouloud (2007) explica que a fase da experimentação é clássica, pois é o momento de colocar em funcionamento todo o dispositivo construído, corrigindo-o quando as análises locais do desenvolvimento experimental identificam essa necessidade, o que implica em um retorno à análise *a priori*, um processo de complementação.

É necessário que, na fase de experimentação, haja a apresentação do dispositivo experimental, o qual deve, de acordo com Almouloud:

- Discutir os objetivos e o quadro teórico que sustentam o dispositivo experimental para seu conhecimento.
- Justificar as escolhas feitas.
- Descrever as condições e o contexto da experimentação.
- Justificar o dispositivo experimental em relação à(s) questão(ões) e às hipóteses da pesquisa.
- Apresentar as situações experimentais (cronograma da experimentação e organograma do trabalho em fases).
- Justificar o encadeamento das fases da experimentação. (ALMOULOU, 2007, p.178).

Nesta etapa, aplicamos de fato a atividade didática, abordando temas vinculados à formação profissional e ao curso do indivíduo, contextualizado com seu cotidiano e abordando os principais conceitos da estatística descritiva com o uso do *R Commander*. Além disso, durante todas as atividades, houve uma etapa de observação e interação do pesquisador com os alunos, realizando uma retrospectiva às análises *a priori*, conforme comentado por Almouloud.

4.4.4 Análises *a posteriori* e validação

A análise *a posteriori* de uma sessão é, de acordo com Almouloud (2007), o conjunto de resultados que pode ser retirado da exploração de dados recolhidos e que contribui para a melhoria dos conhecimentos didáticos que possuímos sobre as condições da transmissão do saber em jogo. Segundo o autor, ela é feita à luz da análise *a priori*, dos fundamentos teóricos, das hipóteses e da problemática da pesquisa. Assim, afirma que:

A análise *a posteriori* depende das ferramentas técnicas (material didático, vídeo) ou teóricas (teoria das situações, contrato didático, etc.) utilizadas com as quais se coletam os dados que permitirão a construção dos protocolos de pesquisa. Esses protocolos serão analisados profundamente pelo pesquisador e as informações daí resultantes serão confrontadas com a análise *a priori* realizada. O objetivo é relacionar as observações com os objetivos definidos *a priori* e estimar a reprodutibilidade e a regularidade dos fenômenos didáticos identificados. (ALMOULOU, 2007, p.177).

Nessa etapa de aplicação e análise da sequência, como apontado por Almouloud (2007), deve-se pelo menos:

- Prever os instrumentos de coletas de dados.
- Organizar e analisar as produções dos alunos, levando em consideração as atividades propostas e as informações coletadas no decorrer da experimentação. Essa análise deve ser feita considerando as diferentes interações dos alunos (aluno-situação, aluno-aluno, aluno-professor) com o milieu (meio) adidático e didático.
- Estudar as modificações possíveis no estudo proposto (questão de pesquisa, os fundamentos teóricos e metodológicos, as situações propostas, as variáveis escolhidas, o esquema experimental).
- Análise dos principais resultados, em relação à questão de pesquisa, às hipóteses, à metodologia adotada e aos resultados de outras pesquisas sobre o mesmo tema.
- Retomada do problema, com síntese das conclusões e avaliação das limitações da pesquisa. Em última instância, o pesquisador deve discutir os principais resultados e as questões levantadas pela pesquisa, que podem ser objeto de outras pesquisas, e destacar a importância dos resultados alcançados para a área da didática da matemática, a instituição escolar, os

alunos, a formação inicial e continuada de professores. (ALMOULOU, 2007, p.178).

Em nossa pesquisa, seguimos os pontos observados pelo autor e, durante a consecução das atividades, criamos uma articulação entre todas as etapas da Engenharia Didática.

5 ATIVIDADES DIDÁTICAS

5.1 Uma proposta de sequência didática

Neste trabalho, seguimos com uma proposta de sequência de atividades utilizando como estratégia o uso de tecnologias digitais, especificamente o *software* R e seu pacote *R Commander*. Ao elaborarmos as atividades, fizemos uma reflexão, que norteou a nossa pesquisa e, de forma assertiva, contribuiu para as respostas referentes aos nossos objetivos iniciais.

Iniciamos as análises buscando as principais pesquisas na área, que abordassem temas vinculados ao aprendizado da Estatística em cursos de graduação, e realizamos inúmeras leituras que guiassem a nossa investigação. Posteriormente, nos apropriamos dos conceitos de letramento estatístico e dos níveis de conhecimentos mobilizados nos alunos, conforme mencionado nos subcapítulos 2.2 e 2.3.

Para as atividades, procuramos atingir todos os níveis de conhecimento, nos termos de Robert (1998) nos alunos e observar elementos que indicassem como ocorre a construção do letramento estatístico. Além disso, consideramos abordar os principais conhecimentos da estatística descritiva, como análises das principais métricas estatísticas, como média, moda, mediana, desvio padrão; análises de tabelas; e análises de gráficos, especialmente o comportamento e a abordagem de histogramas e *boxplots*.

5.2 Atividades

Para a realização das atividades com os três cursos, agrupamos os alunos em trios para análise, dispostos desta forma:

- (1) Do curso de Enfermagem, 18 alunos participaram da pesquisa e, por isso, tivemos 6 grupos, sendo eles Enf1, Enf2, Enf3, Enf4, Enf5 e Enf6;
- (2) Do curso de Medicina, 9 alunos participaram da pesquisa e, por isso, tivemos 3 grupos, sendo eles Med1, Med2 e Med3.;
- (3) Do curso de Licenciatura em Matemática, 9 alunos participaram da pesquisa e, por isso, tivemos 3 grupos, sendo eles Mat1, Mat2 e Mat3.

Procuramos contextualizar os temas e trabalhar com dados que estivessem próximos à formação profissional desses estudante. Assim, nos cursos de Enfermagem e Medicina, utilizamos um banco de dados com o tema “Insuficiência Cardíaca” e, no de Licenciatura em Matemática, um banco de dados do IDEB.

Demonstramos, a seguir, as atividades propostas.

5.2.1 Atividade 1 – Enfermagem e Medicina

Uma vez apresentado o banco de dados e expostos os dados em uma tabela em formato Excel, foi solicitado que os alunos dos cursos de Enfermagem e Medicina realizassem estes problemas.

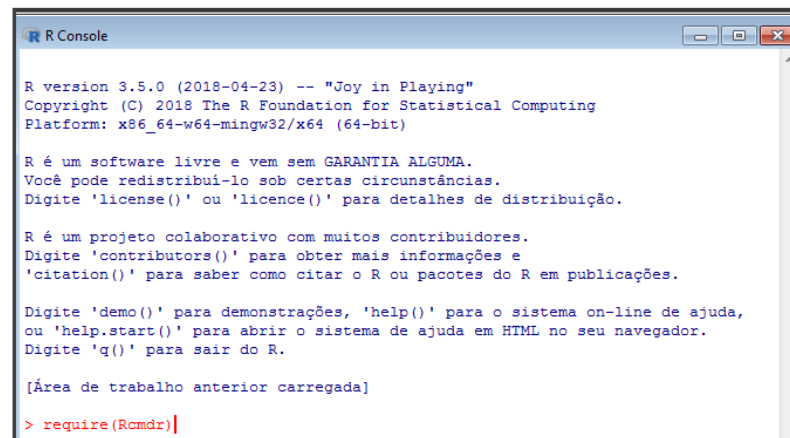
Atividade

1. Com o uso do R, determine a média e o desvio padrão de peso e altura por sexo. Qual(is) característica(s) você percebeu?
2. Com o uso do R, construa uma tabela de frequência de múltipla entrada e determine a proporção de tabagismo e hipertensão por sexo. Qual(is) característica(s) você percebeu?
3. Com o uso do R, determine a média de colesterol total, colesterol HDL e colesterol LDL por sexo. Qual(is) característica(s) você percebeu?
4. Com o uso do R, determine a correlação de Pearson entre colesterol total, glicose, número de medicamentos, triglicerídeos e ácido úrico. Qual(is) característica(s) você percebeu?

Resolução esperada do aluno

1. Abrir o *software* R.
2. Habilitar o pacote *R Commander*, digitando o comando “require (Rcmdr)”, conforme demonstrado na figura 7.

Figura 7 – Habilitando o pacote *R Commander*



```
R version 3.5.0 (2018-04-23) -- "Joy in Playing"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

[Área de trabalho anterior carregada]

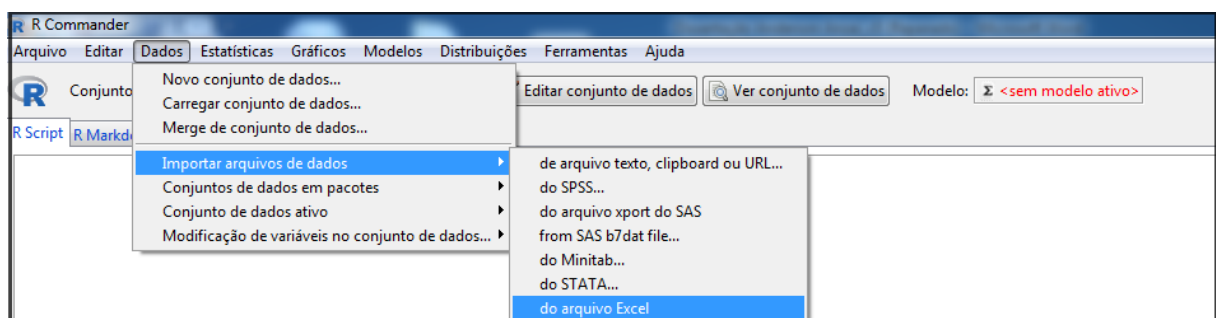
> require(Rcmdr)
```

Fonte: Dados da pesquisa.

3. Importar o banco de dados no pacote *R Commander*, seguindo estes passos:

- (a) Clicar em “Dados” no menu;
- (b) Selecionar “Importar arquivos de dados”;
- (c) Selecionar “do arquivo Excel”, conforme apresentado na figura 8;

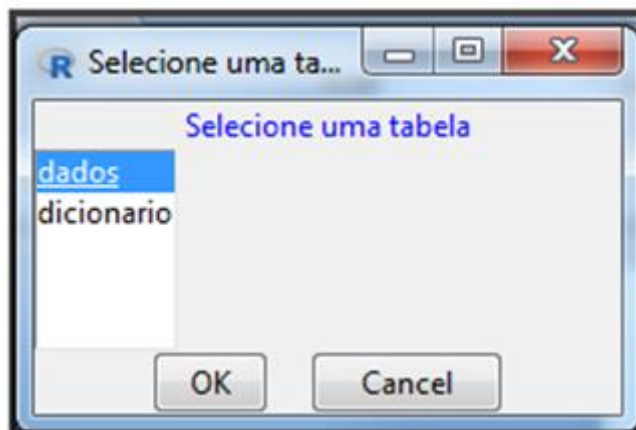
Figura 8 – Importando um banco de dados



Fonte: Dados da pesquisa.

- (d) Nomear o arquivo e clicar em “OK”;
- (e) Localizar o caminho onde se encontra o arquivo;
- (f) Identificar a planilha com os dados e clicar em “OK”, conforme apresentado na figura 9.

Figura 9 – Importando um banco de dados e selecionando a planilha



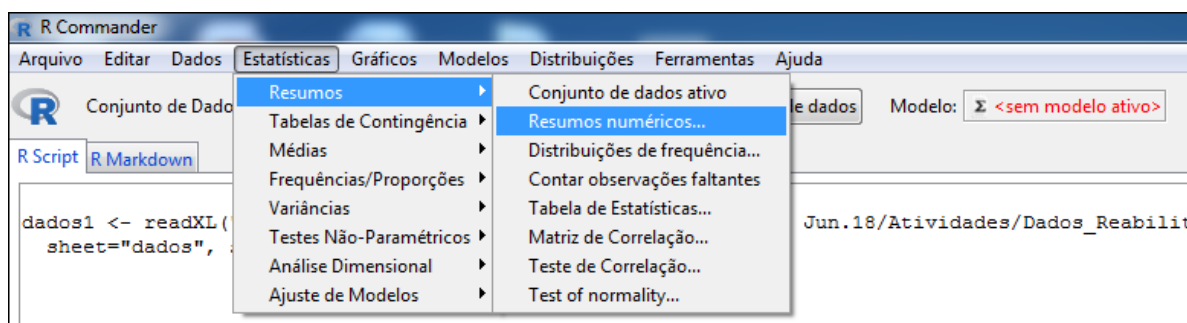
Fonte: Dados da pesquisa.

Com o banco de dados importado, seguir com a resolução de cada item da atividade 1.

Para o problema 1, o aluno deveria construir no *R Commander* uma tabela geral com todas as medidas descritivas para as variáveis peso e altura por sexo, seguindo o passo a passo com os comandos no R.

4. Clicar em “Estatísticas”>“Resumos”>“Resumos numéricos”, conforme apresentado na figura 10:

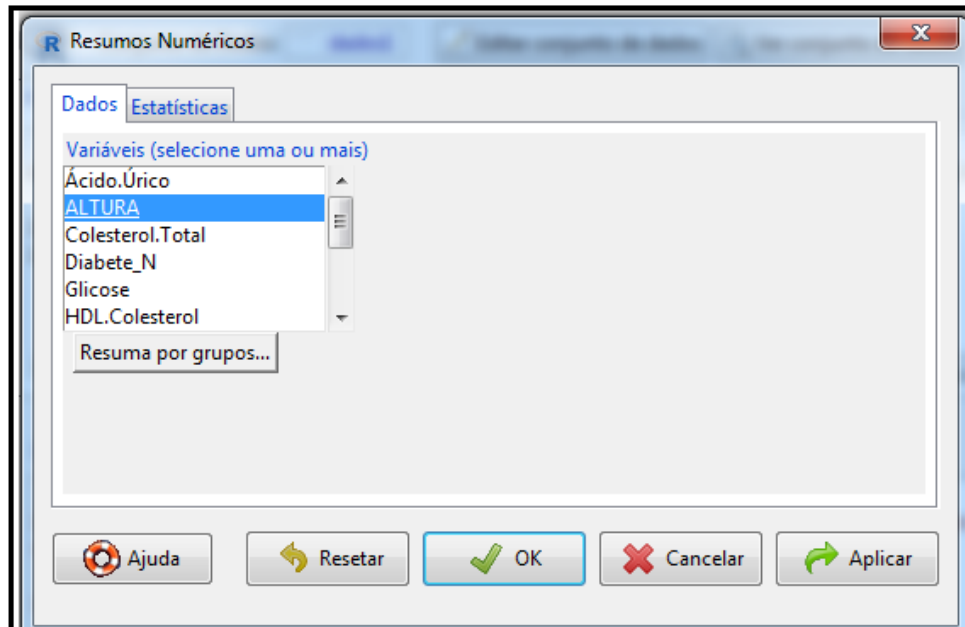
Figura 10 – Solicitando uma tabela geral com as medidas descritivas



Fonte: Dados da pesquisa.

5. Selecionar as variáveis de interesse, neste caso, peso e altura e clicar em “Resuma por Grupos”, selecionando a variável “sexo”, conforme exposto na figura 11.

Figura 11 – Selecionando as variáveis de interesse



Fonte: Dados da pesquisa.

Com isso, temos a saída da tabela no *R Commander*, que apresenta as médias descritivas para as variáveis altura e peso por sexo, conforme a figura 12:

Figura 12 – Tabela de medidas descritivas no R para as variáveis peso e altura

Variable: ALTURA										
	mean	sd	IQR	cv	0%	25%	50%	75%	100%	n
Feminino	1.569677	0.07481015	0.085	0.04765957	1.44	1.515	1.56	1.60	1.81	31
Masculino	1.692185	0.06955256	0.090	0.04110222	1.46	1.650	1.69	1.74	1.91	119

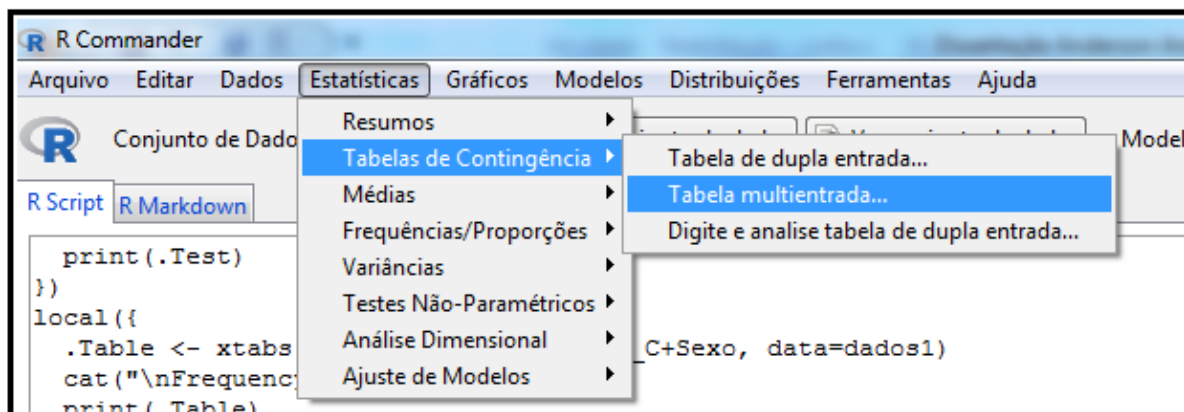
Variable: PESO										
	mean	sd	IQR	cv	0%	25%	50%	75%	100%	n
Feminino	65.57419	12.24835	10.50	0.1867861	47.0	60.00	62.5	70.5	93	31
Masculino	76.60420	12.07734	14.25	0.1576590	53.5	68.75	76.2	83.0	119	119

Fonte: Dados da pesquisa.

Para o problema 2, o aluno deveria construir no *R Commander* uma tabela de frequência de múltipla entrada para as variáveis tabagismo e hipertensão por sexo.

6. Clicar em “Estatísticas” > “Tabelas de Contingência” > “Tabela multientrada”, conforme a figura 13;

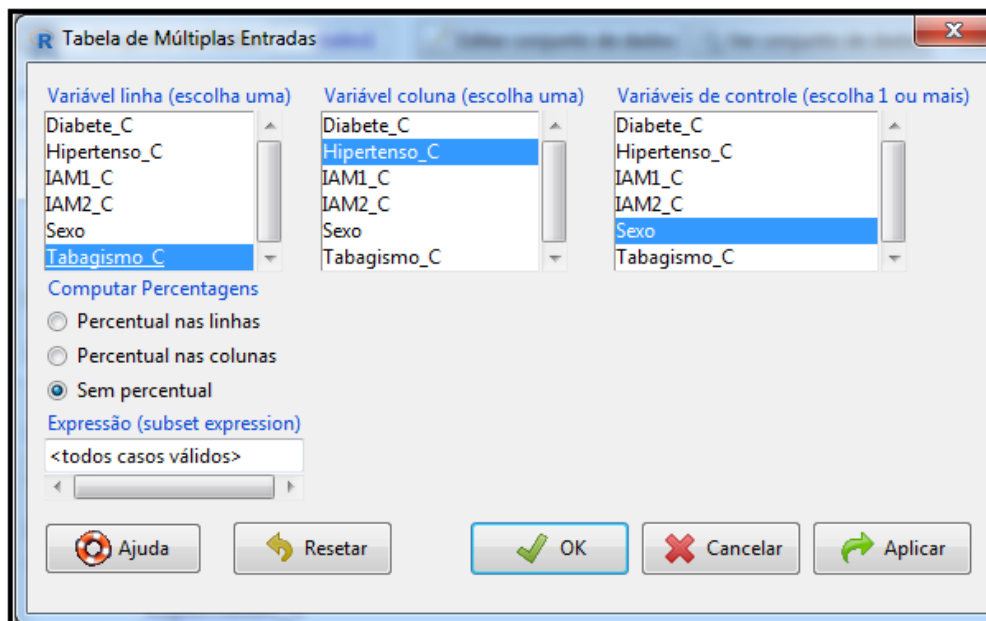
Figura 13 – Solicitando uma tabela de multientrada no *R Commander*



Fonte: Dados da pesquisa.

7. Escolher as variáveis de interesse, neste caso, tabagismo, hipertensão e sexo e clicar em “OK”, conforme apresentado na figura 14.

Figura 14 – Selecionando as variáveis para a tabela de multientrada



Fonte: Dados da pesquisa.

Com isso, temos a saída da tabela no *R Commander*, conforme a figura 15. Na figura temos a saída de uma tabela de dupla entrada, onde temos as quantidades de tabagistas e hipertensos por sexo.

Figura 15 – Saída da tabela de multientrada

```

Frequency table:
, , Sexo = Feminino

      Hipertenso_C
Tabagismo_C Não Sim
      Não    1  14
      Sim    3  13

, , Sexo = Masculino

      Hipertenso_C
Tabagismo_C Não Sim
      Não   17  14
      Sim   49  39

```

Fonte: Dados da pesquisa

Para o problema 3, o aluno deveria construir no *R Commander* uma tabela geral com todas as medidas descritivas para as variáveis colesterol total, colesterol HDL e colesterol LDL, seguindo o passo a passo com os comandos no R.

8. Clicar em “Estatísticas”>“Resumos”>“Resumos numéricos”. A figura 10 mostra o passo a passo. Escolher as variáveis de interesse: colesterol total, HDL colesterol e LDL colesterol, conforme segue na figura 16. Na figura, temos todas as métricas descritivas das variáveis colesterol Total, colesterol HDL e colesterol LDL por sexo.

Figura 16 – Saída da tabela com as medidas descritivas

```

Variable: Colesterol.Total
      mean      sd  IQR      cv  0%   25%  50%  75% 100%   n
Feminino 233.7097 45.14436 58.5 0.1931643 155 206.5 227 265 329 31
Masculino 215.9160 51.33647 65.0 0.2377614 118 179.0 206 244 415 119

Variable: HDL.Colesterol
      mean      sd  IQR      cv  0%  25%  50%  75% 100%   n
Feminino 43.87097 10.378638 11 0.2365719 28 37 42 48 70 31
Masculino 39.01681 9.084335 12 0.2328313 14 33 39 45 65 119

Variable: LDL.Colesterol
      mean      sd  IQR      cv  0%   25%  50%   75% 100%   n
Feminino 152.2581 36.74595 41.5 0.2413399 83 128.0 146 169.5 256 31
Masculino 140.5042 37.83413 50.0 0.2692740 62 113.5 135 163.5 264 119

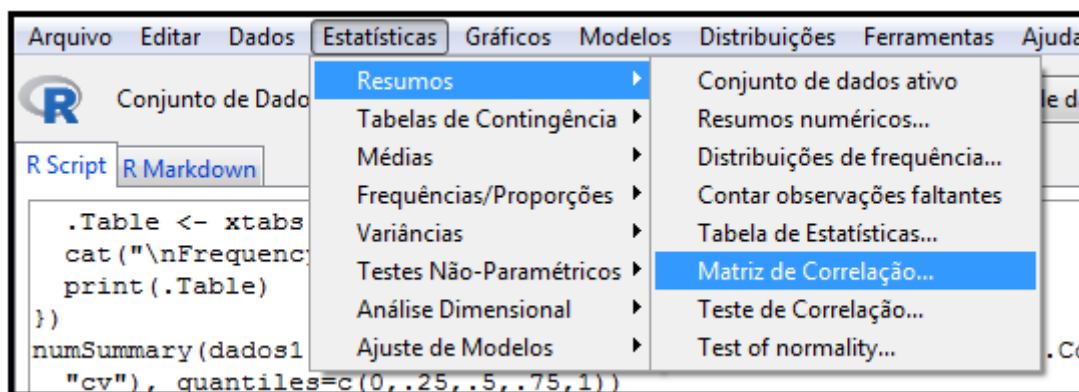
```

Fonte: Dados da pesquisa

Para o problema 4, o aluno deveria construir no *R Commander* uma tabela geral com a correlação de Pearson para as variáveis colesterol total, glicose, número de medicamentos, triglicerídeos e ácido úrico.

9. Clicar em “Estatísticas”>“Resumos”>“Matriz de Correlação”, conforme apresentado na figura 17.

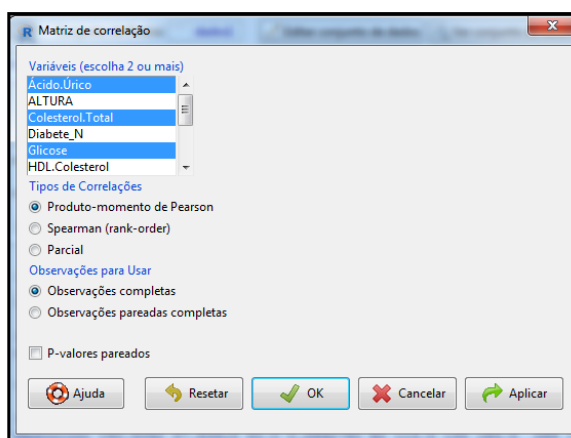
Figura 17 – Solicitando uma matriz de correlação no *R Commander*



Fonte: Dados da pesquisa.

10. Selecionar as variáveis de interesse, neste caso, colesterol total, glicose, número de medicamentos, triglicerídeos e ácido úrico e clicar em “OK”, conforme apresentado na figura 18.

Figura 18 – Selecionando variáveis para uma matriz de correlação no *R Commander*



Fonte: Dados da pesquisa.

Com isso, temos a saída com os valores de correlação para as variáveis ácido úrico, colesterol total, glicose, número de medicamentos e triglicerídeos, conforme apresentado na figura 19.

Figura 19 – Matriz de Correlação no *R Commander*

	Ácido.Úrico	Colesterol.Total	Glicose	Número.de.Medicamentos	Triglicerídeos
Ácido.Úrico	1.000000000	0.17611700	0.002611108	0.14618425	0.2133045
Colesterol.Total	0.176117002	1.00000000	0.010217463	0.07785279	0.3585576
Glicose	0.002611108	0.01021746	1.00000000	0.05433971	0.2720859
Número.de.Medicamentos	0.146184248	0.07785279	0.054339706	1.00000000	0.1203238
Triglicerídeos	0.213304519	0.35855760	0.272085935	0.12032380	1.0000000

Fonte: Dados da pesquisa.

5.2.2 Atividade 2 – Enfermagem e Medicina

Da mesma forma que a atividade anterior, depois de apresentado o banco de dados e expostos os dados em uma tabela em formato Excel, foi solicitado que os alunos dos cursos de Enfermagem e Medicina realizassem estes problemas.

Atividade

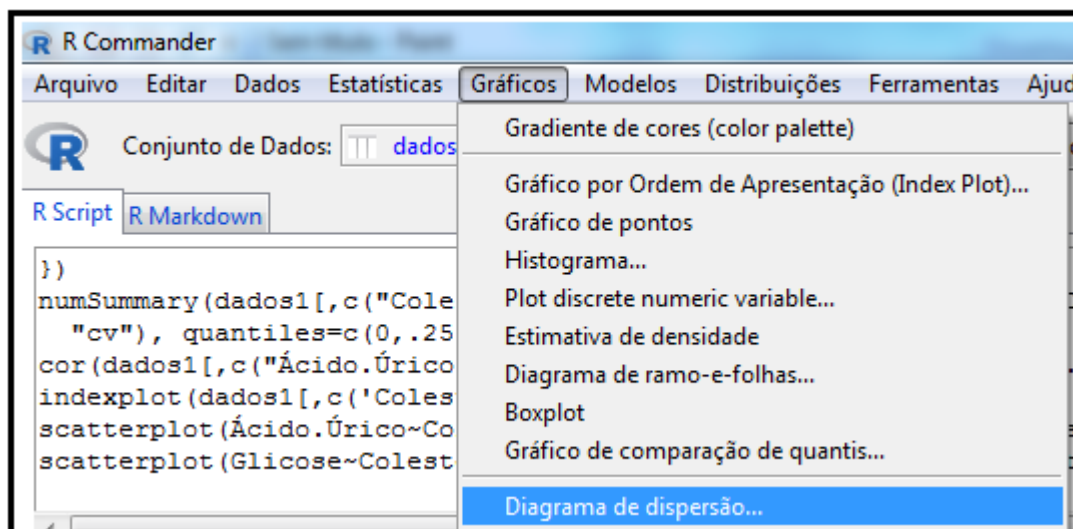
1. Com o uso do R, construa um gráfico de dispersão das variáveis colesterol e glicose por sexo. Qual(is) características você percebeu?
2. Com o uso do R, construa um gráfico *boxplot* para as variáveis sexo e ácido úrico. Qual(is) características você percebeu?

Resolução esperada do aluno

Para o problema 1, o aluno deveria construir no *R Commander* um gráfico de dispersão das variáveis colesterol e glicose por sexo.

1. Clicar em “Gráficos”>“Diagramas de dispersão”, conforme apresentado na figura 20.

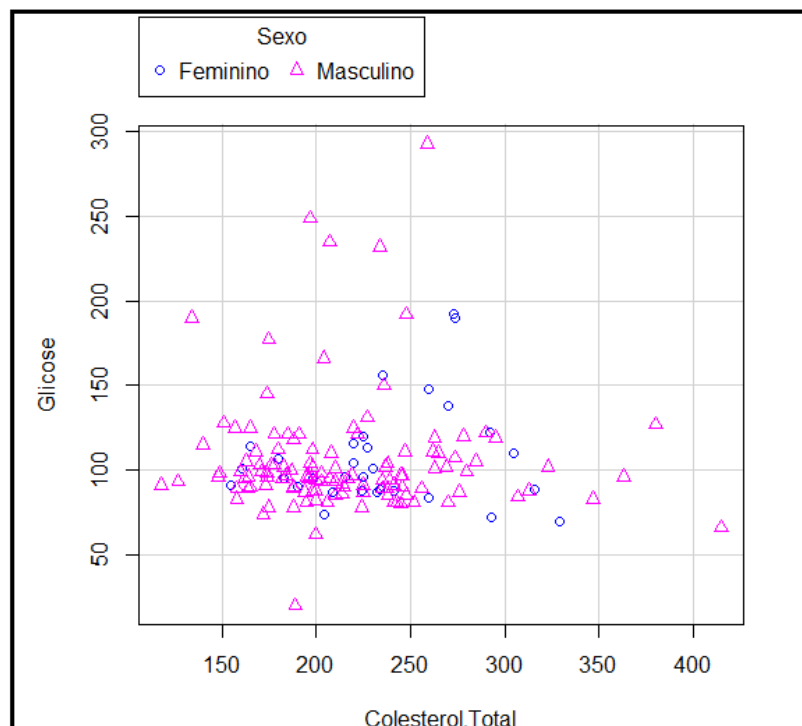
Figura 20 – Solicitando um gráfico de dispersão no *R Commander*



Fonte: Dados da pesquisa.

Com isso, temos a saída do gráfico de dispersão das variáveis Glicose e Colesterol por Sexo, conforme figura 21.

Figura 21 – Gráfico de dispersão de Glicose e Colesterol por Sexo

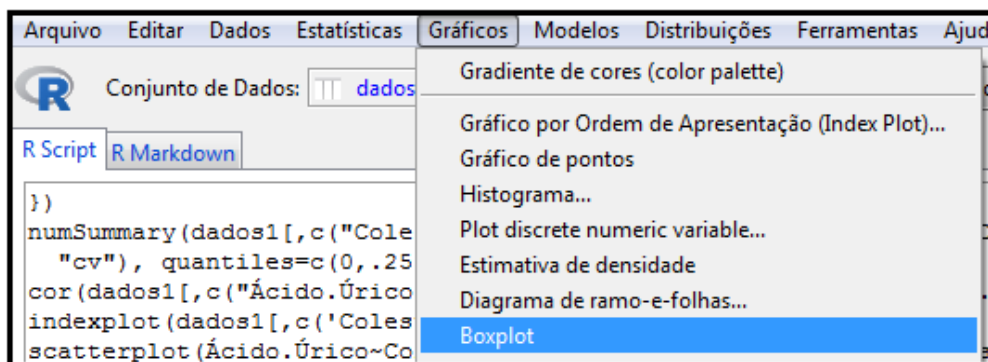


Fonte: Dados da pesquisa

Para o problema 2, o aluno deveria construir no *R Commander* um gráfico *boxplot* para as variáveis sexo e ácido úrico.

2. Clicar em “Gráficos”>“Boxplot”, conforme apresentado na figura 22.

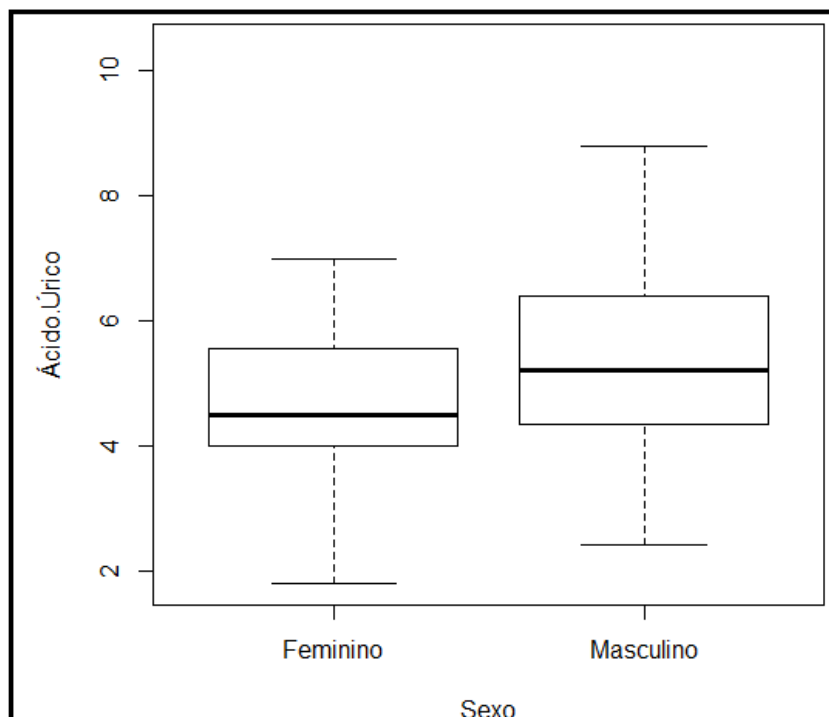
Figura 22 – Solicitando um gráfico *boxplot* no *R Commander*



Fonte: Dados da pesquisa.

Assim, temos a saída do gráfico *boxplot* da variável ácido úrico por sexo, conforme figura 23.

Figura 23 – Gráfico *boxplot* no *R Commander*



Fonte: Dados da pesquisa.

5.2.3 Atividade 1 – Licenciatura em Matemática

Assim como nas atividades anteriores, apresentamos o banco de dados e expostos os dados em uma tabela em formato Excel, solicitamos aos alunos do curso de Licenciatura em Matemática que realizassem estes problemas.

Atividade

1. Utilizando o R, construa uma tabela com as médias das três regiões pelas aprovações de 2005 para 5ª a 8ª série. Qual(is) característica(s) você percebeu?
2. Utilizando o R, construa uma tabela com as médias das três regiões pelas aprovações de 2007 para 5ª a 8ª série. Qual(is) característica(s) você percebeu?
3. Utilizando o R, construa uma tabela com as médias das três regiões pelas aprovações de 2009 para 5ª a 8ª série. Qual(is) característica(s) você percebeu?
4. Utilizando o R, construa uma tabela com as médias das três regiões pelas aprovações de 2005 a 2009 para 5ª a 8ª série. Podemos perceber algum tipo de evolução?
5. Utilizando o R, construa uma tabela com as médias e desvios padrões por região e aprovações de 2005 a 2009. Qual(is) característica(s) você percebeu?

Resolução esperada do aluno

1. Abrir o *software* R.
2. Habilitar o pacote *R Commander*, digitando o comando “require (Rcmdr)”, conforme a figura 7.
3. Importar o banco de dados no pacote *R Commander*, seguindo estes passos:
 - (a) Clicar em “Dados” no menu;
 - (b) Selecionar “Importar arquivos de dados”;
 - (c) Selecionar “do arquivo Excel”, conforme figura 8;
 - (d) Nomear o arquivo e clicar em “OK”;
 - (e) Localizar o caminho onde se encontra o arquivo;
 - (f) Identificar a planilha com os dados e clicar em “OK”, conforme figura 9.

Com o banco de dados importado, seguir com a resolução de cada item da atividade 1 para a turma de Licenciatura em Matemática.

Para o problema 1, o aluno deveria construir no *R Commander* uma tabela geral com todas as medidas descritivas para as variáveis região e aprovações de 2005 para 5ª a 8ª série, seguindo o passo a passo com os comandos no R.

4. Clicar em “Estatísticas” > “Resumos” > “Resumos numéricos”, conforme figura 10.
5. Selecionar as variáveis de interesse, neste caso, região e aprovações de 2005 para 5ª a 8ª série e clicar em OK.

Com isso, será gerada a saída da tabela no *R Commander*, conforme apresentado na figura 24.

Figura 24 – Tabela de medidas descritivas para Região e Aprovação de 2005

	mean	sd	IQR	0%	25%	50%	75%	100%	X2005_Aprov_5_8:n
Nordeste	64.29333	5.614709	9.50	54.7	60.05	63.8	69.55	72.7	15
Sudeste	71.06667	5.855726	6.80	61.9	67.85	69.6	74.65	81.9	15
Sul	82.52667	7.371031	12.05	68.1	76.25	85.1	88.30	93.4	15

Fonte: Dados da pesquisa.

Para o problema 2, o aluno deveria construir no *R Commander* uma tabela geral com todas as medidas descritivas para as variáveis região e aprovações de 2007 para 5ª a 8ª série, seguindo o passo a passo com os comandos no R.

6. Clicar em “Estatísticas” > “Resumos” > “Resumos numéricos”, conforme figura 10.
7. Selecionar as variáveis de interesse, neste caso, região e aprovações de 2007 para 5ª a 8ª série e clicar em OK.

Com isso, será gerada a saída da tabela no *R Commander*, conforme apresentado na figura 25.

Figura 25 – Tabela de medidas descritivas para Região e Aprovação de 2007

	mean	sd	IQR	0%	25%	50%	75%	100%	X2007_Aprov_5_8:n
Nordeste	68.42667	6.419553	10.55	58.7	64.35	67.1	74.90	80.8	15
Sudeste	71.95333	5.939681	7.15	58.3	68.80	73.3	75.95	81.7	15
Sul	85.08000	5.500545	8.45	76.1	79.90	88.2	88.35	93.0	15

Fonte: Dados da pesquisa.

Para o problema 3, o aluno deveria construir uma tabela geral com todas as medidas descritivas para as variáveis região e aprovações de 2009 para 5ª a 8ª série, seguindo o passo a passo com os comandos no R.

8. Clicar em “Estatísticas” > “Resumos” > “Resumos numéricos”, conforme a figura 10 e selecionar as variáveis de interesse — neste caso, região e aprovações de 2009 para 5ª a 8ª série.

Com isso, será gerada a saída da tabela no *R Commander*, conforme apresentado na figura 26.

Figura 26 – Tabela de medidas descritivas para Região e Aprovação de 2009

	mean	sd	IQR	0%	25%	50%	75%	100%	X2009_Aprov_5_8:n
Nordeste	70.40000	6.788856	6.45	61.4	66.20	70.5	72.65	88.2	15
Sudeste	72.34000	6.844268	9.20	62.1	67.25	73.0	76.45	83.0	15
Sul	82.46667	4.510728	4.55	72.8	80.10	82.7	84.65	89.0	15

Fonte: Dados da pesquisa.

Para o problema 4, a aluno deveria construir uma tabela geral com todas as medidas descritivas para as variáveis região e aprovações de 2005 a 2009 para 5ª a 8ª série, seguindo o passo a passo com os comandos no R.

9. Clicar em “Estatísticas” > “Resumos” > “Resumos numéricos”, conforme mostra a figura 10e selecionar as variáveis de interesse — neste caso, região e aprovações de 2005 a 2009 para 5ª a 8ª série.

Com isso, será gerada a saída da tabela no *R Commander*, conforme apresentado na figura 27.

Figura 27 – Tabela de medidas descritivas para Região e Aprovação de 2005 a 2009

Variable: X2005_Aprov_5_8										
	mean	sd	IQR	0%	25%	50%	75%	100%	n	
Nordeste	64.29333	5.614709	9.50	54.7	60.05	63.8	69.55	72.7	15	
Sudeste	71.06667	5.855726	6.80	61.9	67.85	69.6	74.65	81.9	15	
Sul	82.52667	7.371031	12.05	68.1	76.25	85.1	88.30	93.4	15	

Variable: X2007_Aprov_5_8										
	mean	sd	IQR	0%	25%	50%	75%	100%	n	
Nordeste	68.42667	6.419553	10.55	58.7	64.35	67.1	74.90	80.8	15	
Sudeste	71.95333	5.939681	7.15	58.3	68.80	73.3	75.95	81.7	15	
Sul	85.08000	5.500545	8.45	76.1	79.90	88.2	88.35	93.0	15	

Variable: X2009_Aprov_5_8										
	mean	sd	IQR	0%	25%	50%	75%	100%	n	
Nordeste	70.40000	6.788856	6.45	61.4	66.20	70.5	72.65	88.2	15	
Sudeste	72.34000	6.844268	9.20	62.1	67.25	73.0	76.45	83.0	15	
Sul	82.46667	4.510728	4.55	72.8	80.10	82.7	84.65	89.0	15	

Fonte: Dados da pesquisa.

Para o problema 5, o aluno deveria construir uma tabela geral com todas as medidas descritivas para as variáveis região e aprovações de 2005 a 2009, a mesma representada na figura 27.

5.2.4 Atividade 2 – Licenciatura em Matemática

No mesmo molde das outras atividades, após apresentar o banco de dados e expor os dados em uma tabela em Excel, foi solicitado aos alunos do curso de Licenciatura em Matemática que realizassem estes problemas.

Atividade

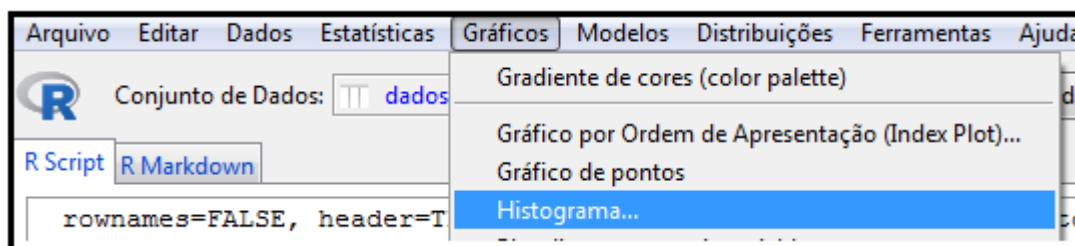
1. Com o uso do R, construa o histograma da região e do indicador de rendimento. Qual(is) característica(s) você percebeu?
2. Com o uso do R, construa um gráfico *boxplot* da região e da nota de Português em 2005. Qual(is) característica(s) você percebeu?

Resolução esperada dos alunos

Para a resolução do problema 1, deve construir no *R Commander* um gráfico histograma da região e do indicador de rendimento.

1. Clicar em “Gráficos” > “Histograma”, conforme apresentado na figura 28.

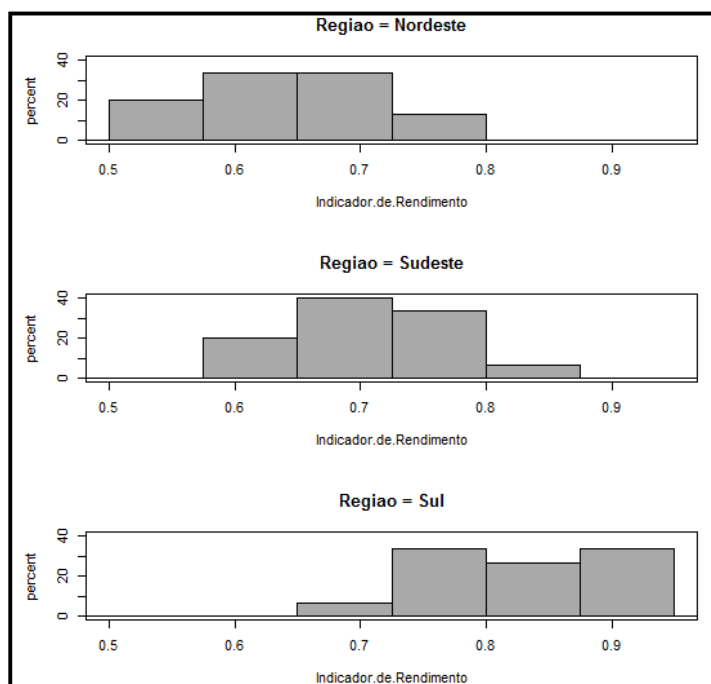
Figura 28 – Solicitando uma gráfico Histograma no *R Commander*



Fonte: Dados da pesquisa.

Com isso, será gerada a saída do gráfico histograma do indicador de rendimento por região.

Figura 29 – Histograma de região por indicador de rendimento no *R Commander*

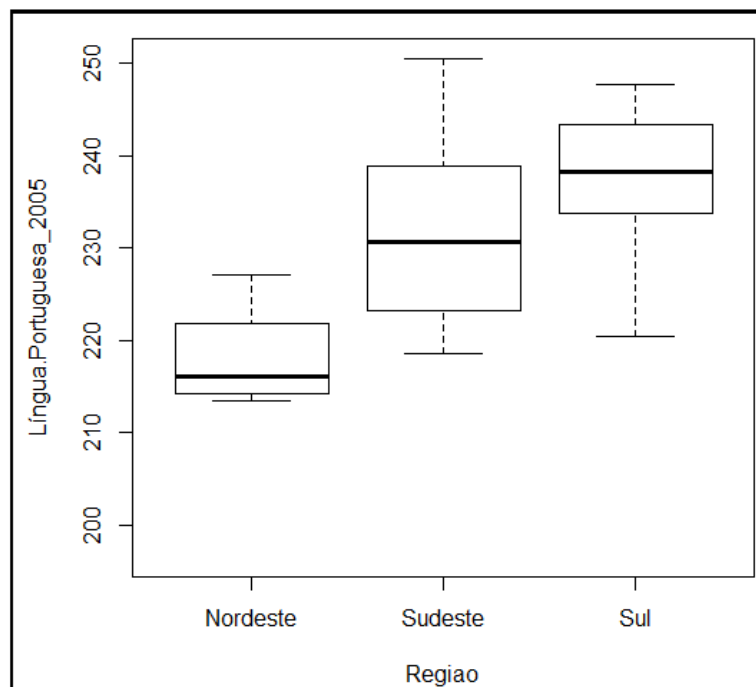


Fonte: Dados da pesquisa.

Para o problema 2, o aluno deveria construir no *R Commander* um gráfico *boxplot* da região e da nota de Português em 2005.

2. Clicar em “Gráficos” > “Boxplot”, e assim temos o gráfico, conforme apresentado na figura 30.

Figura 30 – Gráfico *boxplot* de Região por nota de Português 2005 no *R Commander*



Fonte: Dados da pesquisa.

5.2.5 Análise *a priori* da Atividade 1 – Enfermagem e Medicina

Em todas as análises, esperamos que o aluno mobilizasse uma série de conhecimentos matemáticos e estatísticos e que o uso da tecnologia, especificamente do R, contribuísse para a manipulação do banco de dados, por meio da visualização, dinamismo e experimentação, permitindo assim o desenvolvimento das atividades.

Para a resolução do problema 1, pedimos que os alunos construíssem no *R Commander* uma tabela com as métricas estatísticas para as variáveis altura e peso por sexo. Inicialmente, analisamos o comportamento dos alunos na manipulação do *software* e identificamos o quanto ele contribui para a manipulação dos dados. Construída a tabela, esperamos que os alunos observassem os valores das métricas e fizessem uma análise completa, passando pelos níveis de conhecimento técnico e atingindo, pelo menos, o nível mobilizado.

Nesse problema, os alunos tiveram de mobilizar os seus conhecimentos sobre métricas estatísticas e conceito de variabilidade dos dados, analisando a

saída do *R Commander*, e identificar se havia uma particularidade das variáveis de peso e altura por sexo. Para isso, se fez necessários alguns conhecimentos matemáticos prévios, como contagem e conceitos básicos de operações aritméticas.

Baseando-se no conjunto de dados, o aluno deveria identificar que a variabilidade dos dados tanto para o sexo masculino quanto para o feminino da variável peso é maior quando comparada com a altura. Com isso, o aluno poderia inferir que, para a variável peso, a média não é a melhor métrica para caracterizar o conjunto dos dados, atingindo assim o nível de conhecimento disponível.

Era esperado que, nesse problema, os alunos tivessem dificuldades em interpretar a variabilidade dos dados e significar as métricas, identificando quando usar a média ou quando usar a mediana para representar seus dados. Isso justificado pelo fato de que a abordagem usual da Estatística na escola básica (conhecimentos prévios) e nos livros didáticos não propicia tal tipo de análise, conforme resultados de pesquisas na área, como Coutinho (2013) e Yamauti (2013).

No problema 2, pedimos aos alunos que construíssem no *R Commander* uma tabela de múltipla entrada para as variáveis hipertensão e tabagismo por sexo. Assim como na atividade anterior, observamos a manipulação dos bancos de dados com o *software* e quais os níveis de conhecimento são mobilizados. Nesse momento, foi esperado que o aluno mobilizasse alguns conhecimentos estatísticos, como medidas descritivas e análise bivariada de dados, e conhecimentos matemáticos, como conceitos de proporção, porcentagem e contagem.

Caso o aluno fosse capaz de ler e interpretar a tabela de dupla entrada e realizar uma análise bivariada, calculando as proporções de pacientes hipertensos e tabagistas, teria atingido o nível disponível. Dessa forma, esperamos que construíssem as tabelas abaixo (tabelas 7 e 8) com as proporções e a interpretasse.

Tabela 7 – Tabela de dupla entrada de Tabagismo e Hipertensão por Sexo Feminino

Sexo: Feminino						
Tabagismo / Hipertenso	Não	Sim	Total	Não	Sim	Total
Não	1	14	15	3%	45%	48%
Sim	3	13	16	10%	42%	52%
Total	4	27	31	13%	87%	100%

Fonte: Dados da pesquisa.

Tabela 8 – Tabela de dupla entrada de Tabagismo e Hipertensão por Sexo

Masculino						
Sexo: Masculino						
Tabagismo / Hipertenso	Não	Sim	Total	Não	Sim	Total
Não	17	14	31	14%	12%	26%
Sim	49	39	88	41%	33%	74%
Total	66	53	119	55%	45%	100%

Fonte: Dados da pesquisa.

Baseando-se nas tabelas, o aluno teria condições de identificar que a proporção de tabagistas do sexo masculino é maior (74%) quando comparado com o sexo feminino (52%) e que a proporção de hipertensos é maior no sexo feminino (87%) do que no masculino (45%). Assim, poderia identificar, pela análise bivariada, que a proporção de tabagistas e hipertensos é mais representativa no sexo feminino (42%) do que no masculino (33%).

Neste problema, foi esperado que o aluno apresentasse dificuldades em fazer uma análise bivariada, relacionando duas variáveis ao mesmo tempo, e que apresentasse dúvidas na proporcionalidade das variáveis por sexo. Tais dificuldades podem ser justificadas pela abordagem usual na escola básica, com pouca discussão sobre as tabelas de dupla entrada e análises bivariadas de dados, além de pelos temas não estarem presentes nos livros didáticos.

Para a resolução do problema 3, solicitamos aos alunos que construíssem no *R Commander* uma tabela geral com todas as medidas descritivas para as variáveis colesterol total, colesterol HDL e colesterol LDL por sexo. Esperávamos que o aluno mobilizasse conhecimentos de métricas estatísticas, variabilidade dos dados e conhecimentos matemáticos de operações aritméticas. Foi ainda avaliado sua fluência na manipulação do *R Commander* para a construção da tabela com as medidas resumo.

Com o uso da tabela, esperamos que o aluno identificasse que a média de colesterol total, HDL e LDL são maiores para o sexo feminino. Atingido o nível disponível ao atuar de forma autônoma, o aluno deveria comentar sobre a variabilidade dos dados, sendo mais dispersos para o colesterol total, bem como relacionar e interpretar os valores do quartis.

No problema 4, solicitamos que os alunos construíssem no *R Commander* uma tabela geral com a correlação de *Pearson* para as variáveis colesterol total, glicose, número de medicamentos, triglicerídeos e ácido úrico. Sabemos que muitos alunos apresentam dificuldades na interpretação dos dados de correlação, questionando a partir de que número é considerado uma alta correlação e o que se deve fazer nesses casos, justificado pela falta de abordagem do tema na escola básica.

Era esperado que os alunos mobilizassem seus conhecimentos anteriores, atuassem de forma autônoma, atingindo assim o nível de conhecimento disponível e fazendo uma análise dos números de correlação de cada variável e identificassem as mais correlacionadas, tais quais colesterol total e triglicerídeos, que apresentam correlação de 35,85%. Além disso, que discutissem se esta correlação era alta ou baixa e refletissem sobre o que fazer com essas variáveis em uma possível análise inferencial.

5.2.6 Análise a priori da Atividade 2 – Enfermagem e Medicina

Para a resolução do problema 1, pedimos que os alunos construíssem no *R Commander* um gráfico de dispersão das variáveis colesterol e glicose por sexo. Esperamos que os alunos mobilizassem os três níveis de conhecimento: técnico, mobilizável e disponível, analisando a dispersão dos dados e procurando relacionar com os dados, as medidas descritivas e as tabelas.

Ao construir o gráfico de dispersão, era desejado que os alunos conseguissem visualizar a concentração dos valores para as variáveis. Nesta etapa, para colesterol a concentração está entre 150 e 250 ml e para glicose, entre 80 e 120 ml, atingindo assim o nível técnico.

Os alunos deveriam também comparar a dispersão por sexo, percebendo que os dados tanto de colesterol quanto glicose são mais dispersos para o sexo masculino, comentando sobre a variabilidade dos dados e atingindo assim o nível mobilizável. Estudantes que conseguiram fazer a análise de variabilidade dos dados e relacioná-los com a atividade 1, construíram uma tabela com os dados descritivos e analisaram, por exemplo, o desvio padrão dessas variáveis, atuando de forma autônoma, atingiram o nível disponível. Em relação ao item da fluência, entendemos

que muitos alunos poderiam confundir os tipos de gráficos, como o de pontos com o de dispersão.

Para o problema 2, foi solicitado que os alunos construíssem um *boxplot* para as variáveis sexo e ácido úrico. Era esperado que mobilizassem os três níveis de conhecimento: técnico, mobilizável e disponível. O estudante que foi capaz de construir o gráfico e interpretar a saída, relacionando com todas as medidas descritivas envolvidas, como os quartis, atingiu o nível de conhecimento técnico.

Na sequência, esperamos que os alunos identificassem que os valores de todos os quartis eram inferiores para o sexo feminino; comentassem, por exemplo, que o valor da mediana para a variável é menor no sexo feminino; fizessem uma reflexão do que isso representa no seu conjunto de dados, pensando na variabilidade dos dados; e identificassem que a dispersão dos dados no sexo masculino é maior quando comparada ao sexo feminino. Assim, atingiriam o nível mobilizável.

Por fim, os alunos atingiriam o nível de conhecimento disponível se, de forma autônoma, construíssem as tabelas com as medidas descritivas e relacionassem com o gráfico, comparando os valores da média e mediana por sexo e sua variabilidade, relacionando com as métricas de dispersão.

Considerando que o gráfico não é muito usual, podemos considerar que os alunos apresentem dificuldade na interpretação das medidas expostas e da variabilidade dos dados. Para a fluência, é esperado que tenham dificuldade em escolher o gráfico correto, podendo confundir qual o tipo que foi pedido na atividade.

5.2.7 Análise *a priori* da Atividade 1 – Licenciatura em Matemática

Para a resolução dos problemas 1, 2 e 3, pedimos aos alunos que construíssem no *R Commander* uma tabela geral com todas as medidas descritivas para as variáveis região e aprovações de 2005, 2007 e 2009 para 5ª a 8ª série. Analisamos o desenvolvimento dos alunos a partir da manipulação do *software* e de sua contribuição para a consecução da atividade.

Com relação à manipulação do *software*, avaliamos se o aluno apresentou dificuldades na execução das tarefas, construindo adequadamente as tabelas com as medidas descritivas. Construída a tabela, esperamos que os alunos observassem os valores das métricas e fizessem uma análise completa, passando pelos níveis de

conhecimento técnico, mobilizado e disponível, além de que comparassem as notas de aprovação das três regiões, analisando pelo quartis as notas máximas e mínimas e a dispersão dos dados, atingindo assim o nível de conhecimento técnico.

Caso consiga observar os valores médios das aprovações por região e identificar que a região Sul é a que apresenta as maiores taxas, seguida por Sudeste e Nordeste, o aluno atingirá o nível mobilizável. Também deve observar os valores do desvio padrão e identificar quais dados são mais dispersos, a fim de comparar a média com a mediana e fazer uma reflexão de qual é a mais indicada para representar o conjunto de dados. Ao fazer a reflexão sobre a variabilidade dos dados e construir de forma autônoma, por exemplo, um gráfico para visualizar essa dispersão, ele terá atingido o nível de conhecimento disponível.

Entendemos que os alunos podem sentir dificuldade em relacionar o desvio padrão com a variabilidade dos dados bem como em interpretar conceitos básicos de Estatística, como os valores dos quartis, e associá-los com a amplitude, máximos e mínimos e mediana. Isso ocorre devido aos conhecimentos prévios que são construídos na escola básica, insuficientes, ou até mesmo inexistentes, como menciona Cordani (2001).

Nos problemas 4 e 5, foi solicitado aos alunos que construíssem no *R Commander* uma única tabela geral com todas as medidas descritivas para as variáveis região e aprovações de 2005, 2007 e 2009 para 5ª a 8ª. Para esses itens, analisamos os mesmos quesitos dos itens anteriores, uma vez que os conhecimentos a serem mobilizados também são os mesmos.

Pedimos, propositalmente, que os alunos identificassem as possíveis evoluções e realizassem uma análise temporal dos dados. Eles deveriam perceber que, historicamente, a região Sul é a que apresenta a melhor performance de aprovações, porém as regiões Nordeste e Sudeste vêm apresentando aumento ao longo dos anos, sendo importante ressaltar que, embora a região Sul ainda tenha a melhor performance, no último ano, caiu 3 pontos na taxa.

5.2.8 Análise *a priori* da Atividade 2 – Licenciatura em Matemática

Para a resolução dos problemas 1 e 2, exploramos a parte gráfica do banco de dados bem como investigamos a fluência dos alunos na manipulação do *software*

e na interpretação dos gráficos, esperando que tenham conseguido relacioná-los com os itens da atividade anterior.

No primeiro, pedimos para que os alunos construíssem um histograma da região e do indicador de rendimento. Consideramos ser possível que os alunos mobilizassem os três níveis de conhecimento: técnico, mobilizável e disponível, interpretando a saída do gráfico e relacionando com os dados ao visualizar uma tabela de frequência, bem como alguns conhecimentos matemáticos, como porcentagem, proporção e contagem.

Observando os histogramas por região, primeiramente, o aluno deveria observar a frequência de cada indicador de rendimento por região, atingindo o nível de conhecimento técnico. Dessa forma, relacionar os dados com a atividade 1, visualizando o melhor rendimento da região Sul, concentrado em taxas de aprovação acima de 0,7 e associando as métricas, como média, mediana e amplitude, levaria o estudante a atingir o nível de conhecimento mobilizável.

Por fim, o aluno que, de forma autônoma, construísse também uma tabela com as métricas e uma de frequência de dados assim como discutisse a simetria dos histogramas, teria atingido o nível de conhecimento disponível.

Entendemos que alguns alunos poderiam ter dificuldade em relacionar o histograma com os valores das métricas, interpretar as porcentagens e identificar a concentração dos valores de rendimento. Isso pode ser justificado pelo fato de que na escola básica não são incentivados a interpretar os gráficos associando às métricas estatísticas.

No segundo problema, solicitamos aos alunos que construíssem um *boxplot* da nota de Português em 2005 por região. Esperamos que os alunos mobilizassem os três níveis de conhecimento: técnico, mobilizável e disponível, interpretando a saída do gráfico, sabendo interpretá-lo de forma assertiva e identificando todas as medidas expostas.

Assim, o aluno deveria identificar os quartis do gráfico e perceber que a região Sul é a que, de fato, também apresenta as melhores notas da disciplina Português, identificando os quartis, comentando sobre a mediana (segundo quartil) e, então, atingindo o nível de conhecimento técnico. Posteriormente, os alunos também poderiam comentar sobre a variabilidade dos dados, como notas máximas e mínimas e perceber que a região Sudeste é a que apresenta maior dispersão das notas, atingindo assim o nível de conhecimento mobilizável.

Por fim, se o aluno, de forma autônoma, transitar entre a representação gráfica e a tabular, ou seja, construir as tabelas com todas as medidas descritivas e conseguirem relacionar com a representação gráfica, terá atingido o nível de conhecimento disponível.

Como se trata de um gráfico que muitos alunos poderiam não conhecer, era esperado que tivessem dificuldade em avaliar os resultados e associá-los com os dados das métricas da atividade, temas pouco explorados na escola básica e nos livros didáticos, conforme menciona Coutinho (2013).

5.2.9 Experimentação e análise *a posteriori* da Atividade 1 – Enfermagem e Medicina

As atividades expostas neste capítulo fizeram parte da fase avaliativa da Engenharia Didática, e o desenvolvimento e institucionalização foram realizados pelos professores das turmas em aula regular.

Na resolução da Atividade 1, com relação à fluência na manipulação dos dados no R, tanto para Medicina como para Enfermagem identificamos que os alunos não apresentaram dificuldades. Alguns alunos comentaram que o fato do *software* apresentar essa versão por comandos facilita a manipulação.

Quanto às análises estatísticas, notamos que os alunos tiveram dificuldades em diferenciar a média da mediana e em quando usar uma ou outra para melhor representar o seu conjunto de dados, assim como aconteceu nas atividades aplicadas na pesquisa de Novaes (2004). Alguns estudantes não analisaram os quartis, associando, por exemplo, o segundo quartil à mediana e aos pontos de máximos e mínimos.

Vale ressaltar, também, que poucos alunos, associaram os valores de desvio padrão à variabilidade dos dados, conforme esperado na análise *a priori*. É importante frisar que são poucos os exercícios em livros didáticos que exploram a interpretação dessa medida e a associa à variabilidade.

Figura 31 – Resposta do grupo Med1 para o problema (1) da Atividade 1

1. Com o uso do R, determine a média e o desvio padrão de: Peso e Altura por sexo. Qual (is) característica (s) você percebeu?

Com relação a altura, ao se analisar a média para homens e mulheres, é possível ver que apresentam um baixo desvio padrão, demonstrando que na geral as alturas individualmente são próximas do valor encontrado na média. Entretanto, quanto ao peso, o desvio padrão foi mais elevado, demonstrando a persistência de valores disjuntos ou distantes da média.

Fonte: Dados da pesquisa.

O grupo Med1, conforme figura 31, explanou sobre os valores do desvio padrão, buscando relacionar o seu valor à média, mas não comparou as variáveis de peso e altura por sexo, conforme esperávamos que ocorresse nas análises *a priori*. Possivelmente, isso é explicado pelo fato de, conforme Cordani (2001), muitos alunos egressos no ensino superior possuírem conhecimento prévio insuficiente, ou até mesmo inexistente, de conceitos estatísticos básicos. Consideramos que esse grupo atingiu apenas o nível técnico de conhecimento.

Figura 32 – Resposta do grupo Med2 para o problema (1) da Atividade 1

1. Com o uso do R, determine a média e o desvio padrão de: Peso e Altura por sexo. Qual (is) característica (s) você percebeu?

O sexo feminino possui menor média de peso e um dp levemente maior. Já em relação a altura apresentam menor altura e um maior desvio padrão. Representando um grupo mais heterogêneo.

Fonte: Dados da pesquisa.

O grupo Med2, conforme apresentado na figura 32, comparou as variáveis de peso e altura por sexo e comentou sobre o desvio padrão, mas não explicou sobre a variabilidade dos dados, conforme esperávamos em nossa análise *a priori*. Dessa forma, podemos inferir que essa falta de mobilização de conhecimentos possa ser um efeito de contratos didáticos anteriores (escola básica), quando a Estatística é usualmente abordada apenas em seus aspectos computacionais, de forma

mecânica e pouco contextualizada. Assim, consideramos que o grupo apresentou também apenas o nível técnico do conhecimento.

Figura 33 – Resposta do grupo Med3 para o problema (1) da Atividade 1

1. Com o uso do R, determine a média e o desvio padrão de: Peso e Altura por sexo. Qual (is) característica (s) você percebeu?

Peso: média = 74,32467 e desvio = 12,8764

Altura: média = 1,666867 e sd = 0,08622693

As alturas são mais homogêneas que os pesos, as masculinas são um pouco maiores e mais pesadas

Fonte: Dados da pesquisa.

Já o grupo Med3, conforme exposto na figura 33, relatou os resultados de uma forma genérica, sem especificar se os valores apresentados eram do sexo masculino ou feminino, e comentou sobre dados homogêneos, acreditamos que inferidos pelos valores do desvio padrão. Aqui entendemos que o grupo atuou de forma mecânica e não buscou interpretar os dados, atingindo a fase inicial do nível técnico.

Figura 34 – Resposta do grupo Enf1 para o problema (1) da Atividade 1

1. Com o uso do R, determine a média e o desvio padrão de: Peso e Altura por sexo. Qual (is) característica (s) você percebeu?

A média de altura dos homens é de 1,69 (dp = 0,069), enquanto das mulheres é de 1,56 (dp = 0,074), demonstrando maior variância nos valores das mulheres

A respeito da variável peso, a média do sexo feminino é de 65,5 (dp = 12,2) enquanto que do sexo masculino a média é de 76,6 (dp = 12,07)

Fonte: Dados da pesquisa.

Analisando também a resposta do grupo Enf1, conforme apresentado na figura 34, percebemos que a maioria dos grupos expuseram os valores de médias e desvio padrão das variáveis peso e altura, sem interpretar os resultados e fazer

menção sobre a variabilidade dos dados. Isso pode ser justificado pela abordagem equivocada da Estatística na escola básica, mecanizada e pouco contextualizada.

Consolidada a análise *a posteriori* do problema 1 da Atividade 1, identificamos que os grupos mencionaram as medidas de dispersão e o desvio padrão, porém poucos comentaram sobre a variabilidade dos dados, identificando qual das duas variáveis apresentavam maior dispersão. Nenhum grupo escreveu sobre os quartis, e os valores máximos e mínimos. Já para o uso do R, não apresentaram dificuldade e conduziram de forma efetiva o que foi pedido nas atividades.

Assim, podemos dizer que, nesta atividade, em relação à manipulação do *software*, os alunos atingiram o nível de conhecimento disponível e, em relação às análises estatísticas, apenas o nível técnico.

Na resolução da Atividade 2, observamos que tanto os alunos de Medicina quanto de Enfermagem apresentaram dificuldade em calcular as proporções de hipertensos e tabagistas, demonstrando que não mobilizaram corretamente os conhecimentos matemáticos de porcentagem e proporcionalidade, assim como ocorreu na pesquisa de Novaes (2004).

Figura 35 – Resposta do grupo Med1 para o problema (2) da Atividade 1

2. Com o uso do R, construa uma tabela de frequência de múltipla entrada e determine a proporção de tabagismo e hipertensão por sexo. Qual (is) característica (s) você percebeu?

Das mulheres com diabetes hipertensas, mais de 60% não tem diabetes, por exemplo. Das que são hipertensas, houve uma boa correlação com a variável tabagismo, sendo possível quantificar as diferenças entre homens e mulheres.

Fonte: Dados da pesquisa.

Os alunos tiveram muita dificuldade para interpretar uma tabela de dupla entrada e conseguir fazer uma análise bivariada. Em sua resposta, o grupo Med1, conforme figura 35, comentou sobre a correlação entre os dados, sem nenhuma informação que desse respaldo para isso. Vale ressaltar que, em nossas análises prévias, não identificamos atividades que trabalhassem a análise bivariada de dados. Associado a isso, percebemos que na escola básica pouco se comenta sobre o assunto bem como há uma ausência do tema em livros didáticos.

Figura 36 – Resposta do grupo Enf2 para o problema (2) da Atividade 1

2. Com o uso do R, construa uma tabela de frequência de múltipla entrada e determine a proporção de tabagismo e hipertensão por sexo. Qual (is) característica (s) você percebeu?

O número de tabagistas e hipertensos é maior no sexo masculino, mais do que no sexo feminino, porém existem mais homens na pesquisa do que mulheres.

Fonte: Dados da pesquisa.

Para o grupo Enf2, conforme figura 36, identificamos que também houve complexidade em interpretar a tabela de dupla entrada e analisaram, erroneamente, que a proporção de tabagistas e hipertensos é maior no sexo masculino quando comparado ao feminino. A análise é semelhante à do grupo anterior, ratificando a dificuldade dos alunos em interpretar dados bivariados, assim, o grupo não atingiu nem ao menos o nível técnico de conhecimento.

Figura 37 – Resposta do grupo Enf3 para o problema (2) da Atividade 1

2. Com o uso do R, construa uma tabela de frequência de múltipla entrada e determine a proporção de tabagismo e hipertensão por sexo. Qual (is) característica (s) você percebeu?

A quantidade de homens e mulheres não tabagistas e hipertensas é a mesma, porém um número maior de homens apresentam ambas as morbidades.

Fonte: Dados da pesquisa.

O comentário feito pelo grupo Enf3, conforme exposto na figura 37, demonstra que os alunos confundiram os conceitos de proporcionalidade com valor absoluto em quantidade. Quando atestaram que a quantidade de indivíduos do sexo masculino e feminino é a mesma, consideraram apenas o valor em quantidade, não em proporção, e não observaram que a quantidade de homens era superior ao de mulheres. Nesse caso, o conhecimento mobilizado esperado não estava disponível

nesses estudantes e podemos considerar que o grupo também não atingiu o nível técnico do conhecimento.

Figura 38 – Resposta do grupo Enf4 para o problema (2) da Atividade 1

2. Com o uso do R, construa uma tabela de frequência de múltipla entrada e determine a proporção de tabagismo e hipertensão por sexo. Qual (is) característica (s) você percebeu? <i>Ao compararmos tais variáveis nota-se que há mais homens hipertensos que são fumantes do que em mulheres.</i>
--

Fonte: Dados da pesquisa.

O grupo Enf4, conforme exposto na figura 38, também confundiu conceitos de proporcionalidade com valor em quantidade. Ao afirmar que a quantidade de indivíduos do sexo masculino hipertensos que são fumantes (39 de 119 indivíduos, ou seja, 33%) é maior que a do sexo feminino (13 de 31 indivíduos, ou seja, 42%), foi considerando apenas o valor em quantidade, não em proporção. Assim como o grupo Enf3, este grupo não mobilizou corretamente seus conhecimentos matemáticos de porcentagem e proporcionalidade, fato que deve ser investigado com mais profundidade.

Consolidando as respostas dos grupos, constatamos que nas análises estatísticas muitos grupos tiveram dificuldades em interpretar os dados em uma tabela de dupla entrada, fazendo assim uma análise bivariada dos dados. Observamos que os alunos apresentaram dificuldade em relacionar conceitos de proporcionalidade a porcentagem, itens esperados na análise *a priori*.

Também verificamos que os alunos não identificaram que a quantidade de indivíduos do sexo masculino era superior ao do sexo feminino e que deveriam considerar os efeitos da proporcionalidade, calculando os percentuais para cada conjunto de situações de tabagistas e hipertensos por sexo.

Com relação à manipulação do *software*, apresentaram dificuldade apenas em diferenciar o tipo de tabela solicitada e discernir se deveriam clicar no item de Tabelas de Contingência ou de Frequência e Proporções, conforme a figura 12.

Na resolução do problema 3, observamos que tanto os alunos de Medicina quanto os de Enfermagem tiveram dificuldade em interpretar as medidas descritivas. Poucos comentaram sobre a variabilidade dos dados e quartis, o que pode ser explicado pelos conhecimentos prévios, dada a abordagem da Estatística na escola básica.

Figura 39 – Resposta do grupo Med1 para o problema (3) da Atividade 1

3. Com o uso do R, determine as médias de Colesterol Total, HDL-Colesterol e LDL-Colesterol por sexo. Qual (is) característica (s) você percebeu?

o distribuído foi bem heterogêneo.

Fonte: Dados da pesquisa.

O grupo Med1, conforme apresentado na figura 39, se equivocou nos termos estatísticos e utilizou o termo heterogêneo sem realizar a avaliação das variáveis isoladamente. Não comentaram sobre os dados descritivos, média, mediana, desvio padrão e quartis. Aqui percebemos que esse grupo não atingiu nem o nível técnico, identificamos a ausência de conhecimentos.

Figura 40 – Resposta do grupo Med2 para o problema (3) da Atividade 1

3. Com o uso do R, determine as médias de Colesterol Total, HDL-Colesterol e LDL-Colesterol por sexo. Qual (is) característica (s) você percebeu?

As mulheres apresentam um maior médio nos 3 fatores em relação homens

Fonte: Dados da pesquisa.

O grupo Med2, conforme figura 40, apresentou apenas um comparativo entre as médias das variáveis, sem comentar sobre a variabilidade dos dados e valores máximos e mínimos. Aqui podemos inferir que há um efeito do contrato assumido na

escola básica, em que o aluno não é incentivado a mobilizar os conceitos estatísticos adequadamente, contextualizando-os. Podemos assumir que o grupo atingiu apenas o nível técnico do conhecimento.

Figura 41 – Resposta do grupo Med3 para o problema (3) da Atividade 1

3. Com o uso do R, determine as médias de Colesterol Total, HDL-Colesterol e LDL-Colesterol por sexo. Qual (is) característica (s) você percebeu?

A média das mulheres é maior e desvio menor, mediana próxima a média (amostra $\oplus$ homogênea que a de homens).

LDL

Fonte: Dados da pesquisa.

O grupo Med3, conforme exposto na figura 41, apresentou em sua resposta apenas um comparativo entre as médias das variáveis, porém foi o único grupo a comentar de forma autônoma sobre os valores da mediana e comentou sobre a homogeneidade dos dados. Podemos considerar que esse grupo atingiu o nível mobilizável do conhecimento.

Figura 42 – Resposta do grupo Enf2 para o problema (3) da Atividade 1

3. Com o uso do R, determine as médias de Colesterol Total, HDL-Colesterol e LDL-Colesterol por sexo. Qual (is) característica (s) você percebeu?

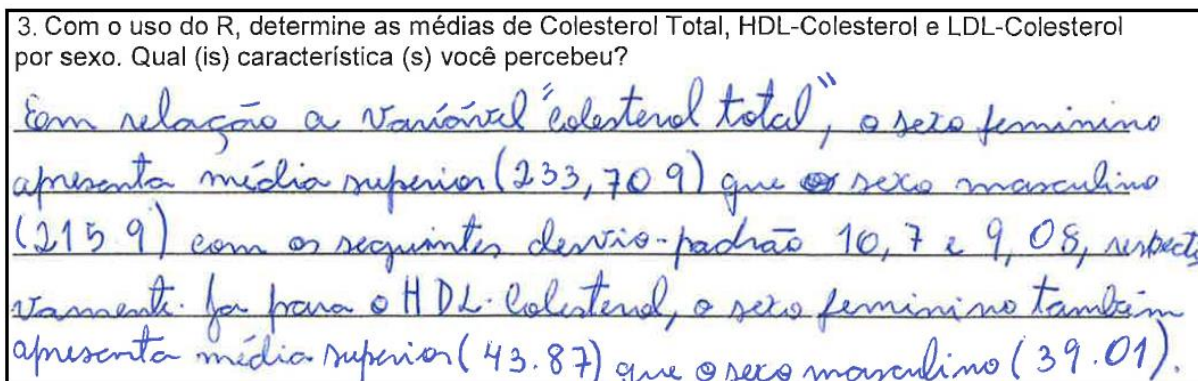
A média do colesterol total no sexo feminino é 233,7 e no sexo masculino é 215,9, a média do HDL-colesterol no sexo feminino é 43,87 e no sexo masculino é 39,01, a média do LDL-colesterol no sexo feminino é 152,2 e no sexo masculino é 140,5.

Fonte: Dados da pesquisa.

O grupo Enf2, conforme apresentado na figura 42, em suas análises descreveu apenas os valores das medidas descritivas sem elaborar uma análise comparativa entre as variáveis por sexo. Percebemos que essa forma mecânica de

apenas descrever os dados é reflexo dos conhecimentos prévios construídos na escola básica, atingindo assim apenas o nível técnico.

Figura 43 – Resposta do grupo Enf4 para o problema (3) da Atividade 1



Fonte: Dados da pesquisa.

O grupo Enf4, conforme exposto na figura 43, também se limitou a apenas expor os valores das médias descritas na tabela, inferimos que seja por conhecimentos prévios construídos na escola básica e também pela ausência de problemas contextualizados em livros didáticos, em que promovam ao aluno a capacidade de atuar como protagonista na resolução do problema, interpretando adequadamente os resultados. Assim, o grupo também atingiu apenas o nível técnico.

Consolidando as análises *a posteriori* desse problema, identificamos que a maioria dos alunos apenas informaram os valores da média, sem relacionar as medidas com o problema no geral, identificando uma postura pouco autônoma. Assim, muitos não conseguiram passar do nível técnico e não atingiram nem mesmo o nível mobilizável de conhecimento. Para a fluência, identificamos que os alunos mobilizaram de forma eficiente seus conhecimentos na manipulação do R.

Na resolução do problema (4), tanto para a Medicina quanto para a Enfermagem percebemos que os alunos sentiram dificuldades para interpretar a correlação e não conseguiram associar corretamente as variáveis. Tal dificuldade pode ser explicada pela abordagem do tema na escola básica e a forma que é abordada nos livros didáticos. Buscamos pesquisas que tratassem do tema e na área há poucas que exploram temas vinculadas a correlação dos dados.

Figura 44 – Resposta do grupo Med2 para o problema (4) da Atividade 1

4. Com o uso do R, determine a correlação (Pearson) entre Colesterol Total, Glicose, número de medicamentos, Triglicerídeos e Ácido Úrico. Qual (is) característica (s) você percebeu?

Colesterol com numero de medicamentos não possui relação. O que usario fezio a relação de colesterol total e triglicerídeos por apresentarem maior porcentagem

Fonte: Dados da pesquisa.

O grupo Med2, conforme apresentado na figura 44, comenta que o colesterol e número de medicamentos não tem correlação, mas na tabela apresenta um valor de 0,1. Foi comentado também sobre porcentagem, o que nos faz inferir que os alunos apresentam conhecimentos insuficientes do tema. Entendemos que esse grupo não atingiu nem o nível técnico de conhecimento.

Figura 45 – Resposta do grupo Enf2 para o problema (4) da Atividade 1

4. Com o uso do R, determine a correlação (Pearson) entre Colesterol Total, Glicose, número de medicamentos, Triglicerídeos e Ácido Úrico. Qual (is) característica (s) você percebeu?

Nenhuma das variáveis tem relação uma com as outras

Fonte: Dados da pesquisa.

Conforme apresentado na figura 45, o grupo Enf2 de forma equivocada que não há correlação entre nenhuma das variáveis apresentadas, o que demonstra uma ausência de conhecimento de conceitos básicos de estatística. Assim, identificamos que para essa atividade, o grupo também não atingiu nem o nível técnico de conhecimento.

Muitos grupos não apresentam resposta para essa questão comentando que não sabiam o que era correlação entre variáveis e qual seria sua utilização em uma

análise estatística. Assim, consolidando a análise desse item, chegamos à conclusão que o tema deve ser mais e melhor abordado nos cursos introdutórios de Estatística, abordando o que correlação entre as variáveis, a correlação de Pearson e a aplicabilidade em testes estatísticos.

5.2.10 Experimentação e análise *a posteriori* da Atividade 2 – Enfermagem e Medicina

Na resolução do problema (1) da Atividade 2, tanto para a Medicina quanto para a Enfermagem os alunos sentiram dificuldades para interpretar os gráficos de dispersão, não conseguindo visualizar a concentração e variabilidade dos dados. Sentimos que os alunos não conseguem relacionar o gráfico com as métricas estatísticas e as pesquisas que analisamos nos mostram que os alunos egressos em cursos superiores chegam com conhecimento insuficientes para trabalhar com os diversos tipos de representações estatísticas.

Figura 46 – Resposta do grupo Med1 para o problema (1) da Atividade 2

1. Com o uso do R, construa um gráfico de dispersão: Colesterol e Glicose por Sexo. Qual (is) característica (s) você percebeu?

A glicose dos homens apresenta-se maior do que nas mulheres, assim como colesterol.

Fonte: Dados da pesquisa.

Conforme apresentado na figura 46, o grupo Med1 comentou apenas sobre os valores máximos da variável glicose presente no sexo masculino. Não comentam sobre a variabilidade e concentração dos dados e não relacionaram com as medidas descritivas do item anterior, conforme esperávamos na análise *a priori*.

Figura 47 – Resposta do grupo Med2 para o problema (1) da Atividade 2

1. Com o uso do R, construa um gráfico de dispersão: Colesterol e Glicose por Sexo.
Qual (is) característica (s) você percebeu?

Apresento-se mais concentrado

Fonte: Dados da pesquisa.

O grupo Med2, conforme exposto na figura 47, comentou sobre a concentração dos dados, mas não especificou a qual a variável ele se referia, apresentando uma análise superficial, que não refletia o que o gráfico demonstrava. Percebemos que o grupo demonstrou ausência de conhecimento para desenvolvimento da atividade, não atingindo nem ao menos o nível de conhecimento técnico.

Figura 48 – Resposta do grupo Med3 para o problema (1) da Atividade 2

1. Com o uso do R, construa um gráfico de dispersão: Colesterol e Glicose por Sexo.
Qual (is) característica (s) você percebeu?

*Distribuição parecida entre homens e mulheres, porém
homens atingem taxas de glicose e colesterol maiores*

Fonte: Dados da pesquisa.

Já o grupo Med3, conforme apresentado na figura 48, comentou que as taxas de glicose e colesterol são maiores no sexo masculino quando comparada ao feminino, porém comentaram de forma equivocada que a distribuição é parecida. Assim, entendemos que o grupo confundiu alguns conceitos, não comentou sobre a variabilidade dos dados e não conseguiu relacionar os dados com as métricas estatísticas, conforme esperávamos na análise *a priori*. O grupo também não

mobilizou corretamente os seus conhecimentos, não atingindo também o nível técnico.

Figura 49 – Resposta do grupo Enf2 para o problema (1) da Atividade 2

1. Com o uso do R, construa um gráfico de dispersão: Colesterol e Glicose por Sexo. Qual (is) característica (s) você percebeu? O valor da glicose dos homens e das mulheres é parecido, já o colesterol dos homens é pior que o das mulheres.
--

Fonte: Dados da pesquisa.

Os grupos Enf2 e Enf3, conforme apresentado nas figuras 49 e 50, também apresentaram dificuldade na interpretação do gráfico de dispersão. Comentaram que os valores para as variáveis de glicose e colesterol são semelhantes. Foi comentado também que o colesterol é pior ou que apresentam maiores níveis no sexo masculino, inferimos que a conclusão esteja relacionada ao pontos de máximo para essa variável. Os grupos também não comentaram sobre a variabilidade dos dados e não relacionaram com as medidas descritivas, conforme esperávamos na análise *a priori*. Percebemos que os grupos apresentaram conhecimentos insuficientes da estatística básica, observação que está de acordo com o comentário de Cordani (2001). Assim, identificamos que os grupos também não atingiram o nível técnico de conhecimento.

Figura 50 – Resposta do grupo Enf3 para o problema (1) da Atividade 2

1. Com o uso do R, construa um gráfico de dispersão: Colesterol e Glicose por Sexo. Qual (is) característica (s) você percebeu? Os níveis de glicose entre mulheres e homens são similares na maioria e os níveis de colesterol são variados. Os homens apresentam maiores níveis de colesterol acima do indicado.
--

Fonte: Dados da pesquisa.

Na resolução desse item os grupos Enf4, enf5 e Enf6 não conseguiram responder à questão.

Para esse problema, no geral, identificamos que os alunos sentiram muita dificuldade para interpretar os dados. Muitos grupos não comentaram sobre a variabilidade dos dados, sobre sua concentração e dados máximos e mínimos.

Nenhum grupo conseguiu relacionar esse item com as medidas descritivas identificadas na atividade anterior. Durante a atividade muitos alunos questionaram sobre a utilização desse gráfico e questionaram se não poderiam construir um gráfico mais simples e usual. Para a fluência em tecnologia, os alunos sentiram dificuldade para encontrar no R qual o comando que deveria ser clicado para construir o gráfico.

Na resolução do problema (2) da Atividade 2, tanto para a Medicina quanto para a Enfermagem os alunos conseguiram interpretar parcialmente o gráfico *boxplot*. Os alunos apresentaram dificuldade para associar o gráfico com os quartis e apenas um grupo comentou sobre a variabilidade dos dados. Na análise *a priori*, comentamos que os alunos poderiam apresentar dificuldades em associar o gráfico com o quartis, pontos de máximo e mínimo e com a variabilidade dos dados, o que de fato ocorreu no decorrer da atividade.

Figura 51 – Resposta do grupo Med2 para o problema (2) da Atividade 2

2. Com o uso do R, construa um gráfico Boxplot do Sexo e Ácido Úrico. Qual (is) característica (s) você percebeu?

O sexo masculino apresenta maiores níveis de ácido úrico

Fonte: Dados da pesquisa.

Conforme apresentado na figura 51, o grupo Med2 apenas comentou sobre os níveis de ácido úrico, não conseguiram associar os quartis as medidas descritivas, não comentaram sobre os valores da mediana e pontos de máximo e mínimos do ácido úrico por sexo. Entendemos que esse grupo poderia fazer uma análise mais completa, conforme esperávamos na análise *a priori*, comentando ao

menos sobre as medidas apresentadas no gráfico, quartis, máximos e mínimos. O grupo não atingiu nem o nível técnico de conhecimento.

Figura 52 – Resposta do grupo Enf4 para o problema (2) da Atividade 2

2. Com o uso do R, construa um gráfico Boxplot do Sexo e Ácido Úrico.
Qual (is) característica (s) você percebeu?

Após analisarmos o boxplot, nota-se que o sexo feminino possui menor taxa para ácido úrico se compararmos ao sexo masculino. Essa afirmação, porém, só será considerada verdadeira ao realizarmos um teste de hipótese e comprovarmos tal relação entre as variáveis.

Fonte: Dados da pesquisa.

O Grupo Enf4, conforme exposto na figura 52, apenas comentou sobre os níveis de ácido úrico, que são menores nos sexo feminino, porém não comentou sobre os quartis e sobre a variabilidade dos dados, entretanto achamos interessante o comentário do grupo sobre a necessidade da realização de um teste de hipótese. Entendemos que para esse item, o grupo atingiu o nível técnico do conhecimento.

Figura 53 – Resposta do grupo Enf5 para o problema (2) da Atividade 2

2. Com o uso do R, construa um gráfico Boxplot do Sexo e Ácido Úrico.
Qual (is) característica (s) você percebeu?

Os valores de ácido úrico no geral para as mulheres é menor do que para os homens, sendo os valores máximo e mínimo feminino menores que o masculino, no entanto, os outliers são maiores.

Fonte: Dados da pesquisa.

O grupo Enf5, conforme apresentado na figura 53, também comentou sobre os níveis de ácido úrico, que são menores nos sexo feminino, foi um único grupo que apresentou em sua análise conceitos de dados máximos e mínimos e dados

outliers. Para esse grupo, podemos considerar que foi atingido o nível de conhecimento técnico.

Figura 54 – Resposta do grupo Enf6 para o problema (2) da Atividade 2

2. Com o uso do R, construa um gráfico Boxplot do Sexo e Ácido Úrico.
Qual (is) característica (s) você percebeu?

Há mais variabilidade dos níveis de ácido úrico entre os homens do que entre as mulheres.

Fonte: Dados da pesquisa.

O grupo Enf6, conforme figura 54, foi o único que comentou sobre variabilidade dos dados, porém não comentou sobre os quartis, pontos de mínimo e máximo e mediana. Podemos considerar que o grupo atingiu o início do nível técnico.

No geral, percebemos que muitos grupos interpretaram de forma parcial o gráfico, apresentaram dificuldades em associar a visualização do gráfico com as medidas descritivas. Durante a atividade dois grupos questionaram sobre a utilização do gráfico, que era a primeira vez que tinham contato com o gráfico e comentaram que preferem analisar os dados organizados em uma tabela. Considerando a nossa análise *a priori*, os alunos atingiram apenas o nível técnico, pouco foi comentado sobre a variabilidade dos dados, sobre a visualização dos quartis, pontos de máximo e mínimo. Podemos concluir, que o gráfico deve ser mais e melhor trabalhado na escola básica, bem como mais abordado nos livros didáticos. Para a fluência, os alunos apresentaram dificuldades para localizar o tipo do gráfico solicitado para pedir no R.

5.2.11 Experimentação e análise *a posteriori* da Atividade 1 – Licenciatura em Matemática

Para a fluência em tecnologia percebemos que os alunos da Licenciatura em Matemática não sentiram grandes dificuldades para manipular o banco de dados no

software estatístico R. Tivemos apenas alguns casos isolados na construção de gráficos, para diferenciar qual era o tipo de gráfico pedido na questão.

Com relação a mobilização dos conceitos estatísticos, nos problemas (1), (2) e (3), percebemos que os alunos conseguiram de forma assertiva relacionar as medidas com a performance de aprovação por região, mencionando que a região Sul é a que possui as melhores taxas de aprovação. Entretanto, de acordo com nossa análise *a priori*, esperávamos que os alunos comentassem sobre variabilidade dos dados, baseando-se no desvio padrão e sobre a menção do uso da mediana em relação à média. Também não tivemos nenhum comentário sobre a observação dos quartis.

Figura 55 – Resposta do grupo Mat1 para o problema (1) da Atividade 1

1. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2005 para 5 a 8 série). Qual (is) característica (s) você percebeu?

A região sul tem um índice de aprovação maior que as outras regiões. As regiões Nordeste e Sudeste apresentam índice de aprovação parecidos, exceto no 7º ano, que a região Sudeste apresentou uma média maior.

Fonte: Dados da pesquisa.

O grupo Mat1, conforme apresentado na figura 55, expôs em sua análise a diferenciação entre as regiões comentando que a Região Sul apresenta os melhores índices de aprovação, porém não comentou sobre a variabilidade dos dados e os quartis, conforme esperávamos na análise *a priori*. Assim identificamos que os alunos atingiram apenas o nível técnico de conhecimento.

Figura 56 – Resposta do grupo Mat2 para o problema (1) da Atividade 1

1. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2005 para 5 a 8 série). Qual (is) característica (s) você percebeu?

A média de aprovação na região sul é maior em todas as séries. Sua mediana também é a maior mas, em média, seu desvio padrão é um pouco maior. Com isso podemos dizer que a qualidade de ensino na região Sul é melhor.

Fonte: Dados da pesquisa

Baseando-se na figura 56, o grupo Mat2 apresentou uma análise mais completa que o grupo anterior, comentando também sobre os valores da mediana e desvio padrão, porém também não comenta sobre a variabilidade dos dados, o que pode estar associado aos conhecimentos insuficientes adquiridos na escola básica. Entendemos que o grupo também atingiu apenas o nível técnico.

Figura 57 – Resposta do grupo Mat1 para o problema (2) da Atividade 1

2. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2007 para 5 a 8 série). Qual (is) característica (s) você percebeu?

A região Sul tem um índice de aprovação bem maior que as demais novamente. As regiões Nordeste e Sudeste continuam com um índice de aprovação muito parecido.

Fonte: Dados da pesquisa.

Para o problema (2), conforme exposto na figura 57, o grupo Mat1 apresentou também em sua análise um comparativo entre as regiões, porém sem comentar sobre a variabilidade dos dados e as informações dos quartis, conforme esperávamos na análise *a priori*. Assim, o grupo não chegou a atingir o nível de conhecimento mobilizável, chegando apenas no nível técnico.

Figura 58 – Resposta do grupo Mat2 para o problema (2) da Atividade 1

2. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2007 para 5 a 8 série). Qual (is) característica (s) você percebeu?

Quase as mesmas características de 2005 porém o que difere é o desvio padrão da região sul diminuiu.

Fonte: Dados da pesquisa.

Já grupo Mat2, conforme figura 58, comentou sobre as características que são as mesmas do item anterior, comenta sobre o desvio padrão que diminuiu, mas

não interpreta essa queda, fazendo assim uma análise superficial, não comentando sobre a melhor performance da Região Sul, baseando-se nas métricas estatísticas, sobre a variabilidade dos dados e sobre os quartis. Observamos uma análise inferior ao que esperávamos na análise *a priori*. Assim, identificamos que o grupo não mobilizou adequadamente os conhecimentos básicos de Estatística e não atingiu nem o nível técnico de conhecimento.

Figura 59 – Resposta do grupo Mat3 para o problema (2) da Atividade 1

2. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2007 para 5 a 8 série). Qual (is) característica (s) você percebeu?

A região sul continua ~~maior~~ com maior taxa de aprovação, mas com pouca variabilidade menor que em 2005, o Sudeste continua com os mesmos médios com taxa semelhante, a variabilidade no Sudeste aumentou em relação a 2005.

Fonte: Dados da pesquisa.

O grupo Mat3, conforme exposto na figura 59, comparou corretamente os índices de aprovação por região e foi o único grupo que mencionou a variabilidade dos dados da região Sul, porém poderia se aprofundar um pouco mais comentando sobre os valores máximos e mínimos, quartis conforme esperávamos na análise *a priori*. Mesmo assim, podemos considerar com o grupo apresentou uma parte do nível mobilizável de conhecimento.

Figura 60 – Resposta do grupo Mat1 para o problema (3) da Atividade 1

3. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2009 para 5 a 8 série). Qual (is) característica (s) você percebeu?

A região sul teve novamente o maior índice de aprovação que as outras regiões. As regiões Sudeste e Nordeste apresentaram índices de aprovação muito próximos.

Fonte: Dados da pesquisa.

Para o problema (3), conforme apresentado na figura 60, o grupo Mat1 também comparou os índices de aprovação por região, porém não apresentou uma análise sobre a variabilidade dos dados e avaliação das taxas máximas e mínimas e dados de quartis, mobilizando assim apenas o nível técnico de conhecimento.

Nos itens (4) e (5) percebemos que os alunos conseguiram identificar as principais tendências de evolução na performance das regiões. Sentimos falta de comentários sobre noções de variabilidade dos dados. Assim, entendemos que a maior parte dos alunos apresentou apenas o nível técnico de conhecimento e que a construção do letramento ainda não está adequada ao esperado. Vale ressaltar que os conceitos estatísticos de base devem ser mais e melhor abordados na escola básica, sendo que os problemas devem ser contextualizados e os alunos incentivados a mobilizarem melhor seus conhecimentos.

Figura 61 – Resposta do grupo Mat1 para o problema (4) da Atividade 1

4. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2005 a 2009 para 5 a 8 série). Podemos perceber algum tipo de evolução?

A aprovação do Nordeste foi aumentando ao longo dos anos de avaliações. O Sul manteve o índice e o Sudeste apresentou pouquíssima melhora.

Fonte: Dados da pesquisa.

Conforme apresentado na figura 61, o grupo Mat1 apresentou em sua análise a evolução das taxas de aprovação por região, porém sentimos falta de uma análise mais completa comentando sobre a variabilidade dos dados, as notas de aprovação máxima e mínima conforme esperávamos na análise *a priori*. Entendemos que os alunos apresentam dificuldade em interpretar as métricas com maior profundidade, o que pode ser justificado pelos conhecimentos insuficientes e equivocados construídos na escola básica, conforme diagnosticado por Cordani (2001).

Figura 62 – Resposta do grupo Mat2 para o problema (4) da Atividade 1

4. Utilizando o R, construa uma tabela com as médias das três regiões (pelas aprovações de 2005 a 2009 para 5 a 8 série). Podemos perceber algum tipo de evolução?

Sim, na região do Nordeste a média melhorou, a mediana também e o desvio padrão deu uma leve aumentada. A média da região Sudeste não houve muitas alterações e seu desvio padrão variou pouco. Já no sul a média também não houve muitas alterações mas seu desvio padrão só diminuiu.

Fonte: Dados da pesquisa.

O grupo Mat2, conforme exposto na figura 62, apresentou uma análise mais completa, comentando sobre a mediana, média e desvio padrão, comentando que seu valor aumentou ao longo dos anos, mas em nenhum momento interpreta essa tendência associando com a variabilidade dos dados. Baseando-se na análise *a priori*, esperávamos também que o grupo comentasse sobre os quartis, notas de aprovação máxima e mínima. Assim, entendemos que o grupo atingiu apenas o nível técnico do conhecimento.

5.2.12 Experimentação e análise a posteriori da Atividade 2 – Licenciatura em Matemática

No problema (1) da Atividade 2, entendemos que os alunos não tiveram dificuldades para manipular adequadamente o *software*, apenas alguns grupos que tinham dúvidas para encontrar o campo adequado para gerar o gráfico que foi pedido no item na atividade.

No que diz respeito a interpretação dos gráficos, vimos que ocorreu de forma parcial. Percebemos que os alunos conseguiram interpretar os histogramas de forma isolada, ou seja, não conseguiram relacionar com as métricas estatísticas determinadas na atividade 1.

Figura 63 – Resposta do grupo Mat1 para o problema (1) da Atividade 2

1. Com o uso do R, construa o histograma da Região e indicador de Rendimento. Qual (is) característica (s) você percebeu?

A região sul tem o maior indicador de rendimento das 3 regiões. O sudeste tem o 2º maior índice de rendimento e o Nordeste está com o menor índice.

Fonte: Dados da pesquisa.

Esse grupo de alunos, conforme exposto na figura 63, apenas apresentou os resultados numéricos, não apresentando a interpretação dos dados. Na análise *a priori*, comentamos que esperávamos que os alunos conseguissem relacionar o histograma com uma tabela de frequência, identificando a distribuição dos dados, buscando identificar os valores médios, mediana e rendimento máximo e mínimo, o que não ocorreu. Assim, não identificamos uma postura autônoma dos alunos, identificando que o grupo atingiu apenas o nível técnico de conhecimento.

Figura 64 – Resposta do grupo Mat2 para o problema (1) da Atividade 2

1. Com o uso do R, construa o histograma da Região e indicador de Rendimento. Qual (is) característica (s) você percebeu?

Na região Nordeste, a frequência do indicador de rendimento é maior entre 0,6 e 0,7. Na região Sudeste, a maior frequência está em 0,7 e no sul entre 0,8 e 0,9.

Fonte: Dados da pesquisa.

Já o grupo Mat2, conforme exposto na figura 64, apresentou uma análise comentando sobre as frequências de cada região, o que era esperado em nossas análises *a priori*, chegando assim a performance do rendimento dos alunos por região, demonstrando que atingiram o nível técnico do conhecimento. Vale comentar que o grupo poderia também ter associado as medidas visualizadas no gráfico com a atividade anterior, comentando sobre as métricas trabalhadas na atividade 1.

Para o problema (2) da atividade 2, os alunos também conseguiram interpretar o gráfico *boxplot* parcialmente. Eles conseguiram identificar os quartis com as medidas descritivas, porém pouco comentaram sobre a variabilidade dos dados e relacionaram os gráficos com as estatísticas trabalhadas na atividade 1.

Figura 65 – Resposta do grupo Mat1 para o problema (2) da Atividade 2

2. Com o uso do R, construa um gráfico Boxplot da Região e nota Português de 2005 . Qual (is) característica (s) você percebeu?

A mediana da região Sudeste e Sul, estão representando bem as notas dessas regiões. As notas da região Nordeste são bem mais baixas que as outras regiões. No Nordeste há 3 municípios que apresentam notas bem mais baixas que a média na região.

Fonte: Dados da pesquisa.

O grupo Mat1, conforme figura 65, conseguiu visualizar nos gráficos a performance das notas de Português por Região, porém não comentaram sobre o intervalo interquartil dos gráficos, bem como a variabilidade dos dados. Esperávamos que os alunos comentassem sobre as notas máximas e mínimas e a dispersão dos dados, identificando que a região Sudeste é a que apresenta a maior variabilidade dos dados, o que não ocorreu. Com isso, o grupo atingiu apenas o nível técnico do conhecimento.

Figura 66 – Resposta do grupo Mat2 para o problema (2) da Atividade 2

2. Com o uso do R, construa um gráfico Boxplot da Região e nota Português de 2005 . Qual (is) característica (s) você percebeu?

Na região Nordeste, a mediana está entre 210 e 220, 3 notas abaixo de 210 e sua maior nota um pouco menor que 230. Na região Sudeste, a mediana é 230, a maior nota é um pouco maior que 250 e a menor, um pouco menor que 220. No sul, a mediana é um pouco menor que 240, sua maior nota está entre 240 e 250 e a sua menor é 220.

Fonte: Dados da pesquisa.

Conforme exposto na figura 66, o grupo Mat2 comentou apenas os valores da mediana, não comentando sobre os outros quartis e sobre a variabilidade dos dados. Também não conseguiram associar os dados apresentados no gráfico com as medidas descritivas do item anterior, o que esperávamos na análise *a priori*.

No geral, identificamos que a maioria dos grupos conseguiu interpretar os gráficos de forma parcial. Nenhum grupo comentou sobre a variabilidade dos dados, tampouco conseguiram associar os dados apresentados no gráfico com as métricas estatísticas. Assim, concluímos que a maior parte dos alunos se limitam ao nível de conhecimento técnico, produzem apenas o que é pedido no enunciado e pouco atuam de forma autônoma. Recomendamos que a parte gráfica seja melhor abordada na escola básica. Para a fluência em tecnologia, sentimos que os alunos não apresentaram grandes dificuldades e mobilizaram de forma correta seus conhecimentos tecnológicos durante a consecução das atividades.

6 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

Dada a importância da Estatística presente em inúmeras áreas do conhecimento, nesta pesquisa buscamos investigar como a disciplina de Estatística está sendo abordada em alguns cursos do ensino superior. Para isso, fizemos uma investigação de pesquisas na área de educação estatística e, de acordo com muitos pesquisadores, diversos alunos apresentam dificuldade em mobilizar os conceitos estatísticos adequados para a construção do seu letramento estatístico. Então, decidimos nos aprofundar no tema e investigar como ocorre o processo de aprendizagem da estatística na graduação, a fim de averiguar as possíveis dificuldades na absorção de conteúdo da estatística descritiva.

Como estratégia didática, utilizamos a abordagem de tecnologias digitais, mais especificamente o *software* estatístico R, com o objetivo de verificar as suas possíveis contribuições na construção do letramento estatístico por alunos de graduação. Além disso, para responder à questão desta pesquisa, aplicamos atividades em que foram adotados como metodologia os pressupostos da Engenharia Didática: análises prévias, *a priori*, experimentação e *a posteriori*.

Para as análises *a priori*, buscamos o apoio em alguns referenciais teóricos, abordando o letramento estatístico e os níveis de conhecimentos mobilizados pelos alunos, de acordo com o conceito de Robert (1998), considerando a fluência em tecnologia. Na fase experimental, aplicamos uma sequência de atividades, na qual houve a participação de alunos de três cursos da Universidade Federal do Estado do Rio de Janeiro (UNIRIO): Enfermagem, Medicina e Licenciatura em Matemática. Por fim, para as análises *a posteriori*, avaliamos a consecução das atividades, em que avaliamos os níveis de conhecimentos mobilizados dos alunos, independentemente de serem matemáticos, estatísticos ou tecnológicos.

Apoiamo-nos em quatro pesquisas dos seguintes autores: Cordani (2001), Novaes (2004), Bifi (2006) e Costa (2012), que nos conduziram na investigação do cenário atual da abordagem da Estatística em cursos de graduação, apresentando as possíveis dificuldades dos alunos na aprendizagem da disciplina, tais como: interpretar as métricas estatísticas adequadamente, em que os alunos têm dúvidas sobre a média e mediana e sobre quando usar uma ou outra para representar o seu conjunto de dados; comentar sobre a variabilidade dos dados, especialmente em associar as medidas de desvio padrão à variabilidade; associar os diversos tipos de

representações estatísticas e mobilizar os conhecimentos adequados para interpretar os dados de uma tabela, como medida de frequência, e associá-la a um gráfico; compreender porcentagem e proporcionalidade, erros que podem estar associados à falta de atenção ou de conhecimento de base.

A partir dessa reflexão e identificadas as principais dificuldades nos alunos, decidimos investigar estratégias didáticas que contribuíssem para o aprendizado da Estatística, optando por utilizar o *software* estatístico R nas atividades.

Vale ressaltar, também, que as pesquisas de Novaes (2004) e Bifi (2006) nos incentivaram a dar continuidade à aplicação de atividades, investigando os níveis de conhecimentos mobilizados pelos alunos e o nível de conhecimento atingido, a fim de avaliar se o seu letramento estatístico foi construído de forma correta.

Em nossas atividades, avaliamos quais os níveis de conhecimentos estatísticos foram mobilizados pelos alunos e se a habilidade de interpretar e avaliar criticamente as informações estatísticas e os argumentos baseados em dados bem como de discutir opiniões referentes a esse tipo de informação estatística, mencionada por Gal (2002), esteve presente na consecução das atividades e foi construída adequadamente pelo aluno.

Durante a aplicação das atividades, na análise *a posteriori*, apoiamo-nos na teoria de Robert (1998) sobre os níveis de conhecimentos mobilizados pelos alunos: técnico, quando o aluno mobiliza de forma isolada mecanismos para aplicação de fórmulas, teoremas e definições, ou seja, se refere mais ou funcionamento de ferramentas conceituais ou procedimentais; mobilizável, que corresponde a funcionamentos mais amplos e o aluno deve adaptar seus conhecimentos, havendo uma atuação mais abrangente do sujeito; e disponível, caracterizado pela atuação autônoma do aluno, em que deve resolver um problema sem indicações, utilizando seus conhecimentos já adquiridos em outras tarefas.

Assim, propusemos duas atividades para cada turma dos cursos de Enfermagem, Medicina e Licenciatura em Matemática. A primeira abordou os conceitos de métricas estatísticas e a segunda, a parte gráfica.

Para a primeira atividade, nas análises *a priori*, esperávamos que os alunos mobilizassem alguns conhecimentos matemáticos: como contagem, conceitos básicos de operações aritméticas, porcentagem e proporcionalidade, bem como estatísticos de base, como análise de métricas estatísticas (média, moda, mediana) e de quartis, identificando pontos de máximos, mediana e mínimos e medidas de

dispersão, como o desvio padrão, associando à variabilidade dos dados. Por fim, analisamos a fluência em tecnologia com a manipulação do banco de dados no *software* estatístico R.

Na segunda atividade, nas análises *a priori*, foi esperado que os alunos mobilizassem os conceitos adquiridos na primeira atividade, para a visualização gráfica de todos os conteúdos estatísticos abordados. No gráfico de dispersão, os alunos deveriam abordar a variabilidade dos dados e identificar a concentração de dados, comentando sobre características das variáveis. No histograma, associar às métricas e à tabela de frequência dos dados. Já no *boxplot*, deveriam identificar no gráfico os pontos de mínimos e máximos bem como o valor da mediana (segundo quartil) e comentar sobre a variabilidade dos dados.

Na consecução das atividades, para a atividade 1, identificamos os seguintes resultados:

- Os alunos apresentaram dificuldade na interpretação das medidas descritivas, na visualização da variabilidade dos dados, em o que representa, de fato, o desvio padrão e em qual a relação com a média. Muitos alunos não conseguiram identificar as diferenças entre a média e a mediana, quando usar uma ou outra e relacioná-las com os quartis;
- Com relação a parte tabular, os alunos apresentaram dificuldade em associar as variáveis em uma análise bivariada bem como em compreender o tema de correlação entre variáveis;
- Na análise de tabelas de frequências, os alunos apresentaram dificuldade para calcular porcentagens e identificar a proporcionalidade entre as variáveis e por sexo;
- Com relação a fluência em tecnologia, identificamos que o uso de um *software*, especificamente o R, contribuiu de forma significativa para a consecução das atividades e os alunos se sentiram motivados a trabalhar com a manipulação dos dados nesse *software*, mostrando autonomia e protagonista durante as atividades.

Na consecução das atividades, para a atividade 2, identificamos os seguintes resultados:

- Para a parte gráfica, verificamos que os alunos conseguiram interpretar parcialmente os gráficos, muitos não estabeleceram uma relação com as medidas descritivas;
- Percebemos que para o gráfico histograma, os alunos sentiram dificuldade para visualizar as frequências dos dados, proporções e associar com as métricas estatísticas;
- Já para o gráfico de dispersão, notamos que os alunos sentiram dificuldade para visualizar as métricas envolvidas, dispersão e concentração dos dados;
- Para o gráfico *boxplot*, identificamos que os alunos sentiram dificuldade para visualizar as métricas estatísticas associadas, como por exemplo, quartis, mediana, valores mínimos e máximos e dispersão dos dados.

Com relação aos níveis de conhecimentos mobilizados, durante a execução das atividades, observamos que os alunos não trabalharam de forma autônoma nem construíram um pensamento e raciocínio fora do que foi solicitado no enunciado. Ao tratar as análises dos dados, percebemos que a maior parte dos alunos não ultrapassou o nível de conhecimento técnico e poucos alunos chegaram ao nível mobilizado.

Voltando a nossa questão de pesquisa, que foi: **Que contribuições podem ser evidenciadas pelo uso de estratégias didáticas com tecnologias digitais, mais especificamente o *software* estatístico R, na construção do letramento estatístico por alunos de cursos de graduação que possuem a disciplina Estatística em seu currículo?**, podemos identificar que em relação à estratégia com o uso do *software*, que é bem aceita pelos alunos e contribuiu significamente para o desenvolvimento da atividade, porém não conseguimos concluir se contribuiu para a construção do letramento estatístico dos alunos, dado que tivemos indícios que os alunos atingiram apenas o nível técnico de conhecimento nos termos de Aline Robert (1998).

Percebemos ao longo do desenvolvimento das atividades que os alunos ainda não possuem habilidades para interpretar e avaliar criticamente os dados estatísticos, entendimento primordial para Gal (2002). Também não identificamos nos alunos os elementos de conhecimento bem como os elementos de disposição, apresentado pelo autor. Podemos inferir que tais conhecimentos não foram mobilizados, dado a ausência da abordagem correta de conhecimentos estatísticos

na escola básica e pesquisas indicam a ausência de conteúdos relevantes em livros didáticos.

Para perspectivas futuras, devemos ressaltar que a reflexão acerca do letramento estatístico, discutindo estratégias para a melhoria da aprendizagem da Estatística, não se esgota por aqui. Cabe a nós, pesquisadores, incentivar e potencializar o conhecimento, utilizando as teorias estudadas, inserindo-as em nosso contexto e buscando apoio nas tecnologias e no uso de *softwares* disponíveis, a partir dos resultados de pesquisas na área de educação estatística.

Dada a particularidade e a relevância da Estatística, devemos tratar o tema com cuidado, sempre repensando novas formas de abordá-la, por exemplo, trabalhar com os alunos as diversas formas de sua representação (gráfica, tabular, numérica, etc.) a fim de potencializá-la e minimizar os possíveis conflitos encontrados na aprendizagem em sala de aula. Assim, podemos propor aulas dinâmicas, contextualizadas e aulas em laboratórios, incentivando o aluno a criar o seu próprio conhecimento, de maneira a mostrar a importância da disciplina não só para sua formação acadêmica mas também para sua formação profissional e, principalmente, como cidadão, expondo que, com a aquisição desses conhecimentos, poderá ter embasamentos e argumentos que o tornarão mais crítico, levando-o a tomar decisões mais assertivas.

REFERÊNCIAS

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

ARTEAGA, P.; BATANERO, C.; DÍAZ, C.; CONTRERAS, J. M. **El lenguaje de los gráficos estadísticos**. *Unión: Revista Iberoamericana de Educación Matemática*, v. 18, p. 93-104, 2009.

ARTIGUE, M. **Ingengería didáctica**. In: ARTIGUE, M.; DOUADY, R.; MORENO, L. *Ingeniería Didáctica en Educación Matemática Un esquema para la investigación y la innovación en la enseñanza y el aprendizaje de las matemáticas*. Bogotá: Una Empresa docente ® & Grupo Editorial Iberoamérica, S.A. de C.V., 1995. p. 33-60.

BATANERO, C. **Didáctica de la Estadística**. Granada: Universidad de Granada, Espanha, 2001.

BATANERO, C. **Significado y comprensión de las medidas de posición central**. Granada: Universidad de Granada, Espanha, 2000.

BATANERO, C.; DÍAZ, C. **El papel de los proyectos en la enseñanza y aprendizaje de la estadística**. In: ROYO, J.P. (ed.). *Aspectos didácticos de las matemáticas*. Zaragoza (España): ICE, 2004. p. 125-164.

BATANERO, C.; DÍAZ, C. **Estadística con proyectos**. Granada (España), Universidad de Granada, 2011.

BATANERO, C.; GODINO, J. D.; GREEN, D. R.; HOLMES, P.; VALLECILLOS, A. **Errores y dificultades en la comprensión de los conceptos estadísticos elementales**. *International Journal of Mathematics Education in Science and Technology*, v. 25, n. 4, p. 527-547, 1994.

BIFI, C. R. **Estatística em um curso de Administração de Empresas: mobilização dos conceitos estatísticos de base**. Mestrado – Pontifícia Universidade Católica de São Paulo, São Paulo, 2006.

BIFI, C. R. **Conhecimentos estatísticos no ciclo I do ensino fundamental: um estudo diagnóstico com professores em exercício**. Tese (doutorado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2014.

BORBA, M.C.; ARAUJO, J. L. (Orgs.). **Pesquisa qualitativa em Educação Matemática**. 4.ed. – Belo Horizonte: Autêntica Editora, 2012

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**, 8 ed. São Paulo. Atual, 2013.

CAMPOS, C. R. **A educação estatística: uma investigação acerca dos aspectos relevantes à didática da estatística em cursos de graduação**. Tese (doutorado) - Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, 2007.

CAMPOS, C. R.; WODEWOTZKI, M. L. L.; JACOBINI, O. R. **Educação estatística: teoria e prática em ambientes de modelagem matemática**. 2. ed. Belo Horizonte: Autêntica, 2013.

CARVER, R; EVERSON, M. **GAISE College Report ASA Revision Committee, “Guidelines for Assessment and Instruction in Statistics Education College Report 2016”**. Disponível em: <<http://www.amstat.org/education/gaise>>.

CASINI, R.; GONZALES, M.; PADRO, O. **Aplicación de Nuevas Tecnologías, en la enseñanza de Estadística, una experiencia ante diferentes propuestas educativas de universidades públicas y privadas de la Ciudad de Córdoba – República Argentina**, 2011.

CAZORLA, I. M. **A relação entre a habilidade viso-pictórica e o domínio dos conceitos estatísticos na leitura de gráficos**. Tese (doutorado) – Universidade Estadual de Campinas, Campinas, 2002.

CHAGAS, R. M. **Estatística para alunos do 6.º ano do ensino fundamental: um estudo sobre os conceitos mobilizados na resolução de problemas**. Dissertação (mestrado) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2010.

CONTI, K. C. **O papel da estatística na inclusão de alunos da educação de jovens e adultos em atividades letradas**. Dissertação (mestrado) - Faculdade de Educação, Universidade Estadual de Campinas, Campinas, 2009.

CONTI, K. C.; CARVALHO, D. L. **O letramento presente na construção de tabelas por alunos da educação de jovens e adultos**. Boletim de Educação Matemática, v. 24, n. 40, p.637-658, 2011.

CORDANI, L. K. **O Ensino de Estatística na Universidade e a controvérsia sobre os fundamentos da inferência**. Tese de doutorado. USP, São Paulo, 2001.

COSTA, G. D. F. **A metodologia de projetos como alternativa para ensinar estatística no ensino superior**. Tese (doutorado) - Faculdade de Educação, Universidade Estadual de Campinas, Campinas, 2012.

COUTINHO, C. Q. S. **Educação estatística e os livros didáticos para o ensino médio**. Revista Educação Matemática em Foco, Campina Grande, v. 2, n. 1, p. 68-86, 2013.

COUTINHO, C. Q. S.; SOUZA, F. S. **Potencial do uso do GeoGebra e do R na construção e interpretação de gráficos estatísticos**. In: SAMÁ, S.; PORCIÚNCULA, M. (Orgs.). *Educação estatística: ações e estratégias pedagógicas no ensino básico e superior*. Curitiba: CRV, 2015. p. 121-131.

DOUADY, R. **Jeux de Cadres et Dialectique outil-objet**. In: *RECHERCHES EM DIDACTIQUE DES MATHÉMATIQUES*, La Pensée Sauvage-Editions, v.7, nº 2, pp. 5-31. Grenoble, 1986.

DUVAL, R., **Quel cognitif retenir en didactique des mathématiques?** *In: Recherches en didactique des mathématiques*, 16 (3), 349-382. 1996.

FIORENTINI, D.; LORENZATO, S. **Investigação em educação matemática: percursos teóricos e metodológicos**. 2. ed. rev. Campinas: Autores Associados, 2007.

GAL, I. **Conocimientos básicos de estadística en adultos: significados, componentes, responsabilidades**. *Revista Internacional de Estadística*, Haifa (Israel), v. 70, n. 1, p. 1-25, 2002.

GARFIELD, J. **Teaching statistics using small-group cooperative learning**. *Journal of Statistics Education*, v. 1, n. 1, p. 1-9, 1993.

GARFIELD, J. **Cooperative learning revisited: from an instructional method to a way of life**. *Journal of Statistics Education*, v. 21, n. 2, p. 1-8, 2013.

GIORDANO, C. C. **O desenvolvimento do letramento estatístico por meio de projetos: um estudo com alunos do Ensino Médio**. Dissertação (Mestrado) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2016.

JACOBINI, O. R. **A Modelagem Matemática Aplicada no Ensino de Estatística em Cursos de Graduação**. 1999. Dissertação (Mestrado) - Instituto de Geociências e Ciências Exatas, UNESP, Rio Claro, 1999.

JACOBINI, O. R. **A modelagem matemática como instrumento de ação política na sala de aula**. Tese (doutorado) - Instituto de Geociências e Ciências Exatas, UNESP, Rio Claro, 2004.

KENSKI, V. M. **Educação e tecnologias: o novo ritmo da informação**. 3. ed. Campinas - São Paulo, 2007.

LIMA, A. C. P.; MAGALHÃES, M. N. **Noções de Probabilidade e Estatística**. 6. ed. São Paulo: Universidade de São Paulo, 2004.

LOPES, C. E. **A probabilidade e a estatística no ensino fundamental: uma análise curricular**. Dissertação (mestrado) – Universidade Estadual de Campinas, Campinas, 1998.

LOPES, J. A. **O livro didático, o autor e as tendências na educação matemática**. *In: NACARATO, A. M.; LOPES, C. E. (Orgs.). Escritas e leituras na educação matemática*. Belo Horizonte: Autêntica, 2009.

LOPES, C. E.; D'AMBROSIO, B. S. **Perspectivas para a educação estatística de futuros educadores matemáticos de infância**. *In: SAMÁ, S.; PORCIÚNCULA, M. (Orgs.). Educação estatística: ações e estratégias pedagógicas no ensino básico e superior*. Curitiba: CRV, 2015. p. 17-27.

LOPES, C. E.; COUTINHO, C. Q. S.; ALMOULOU, S. A. **Estudos e reflexões em educação estatística**. Mercado das Letras: Campinas, 2010.

MARCELINO, S. B.; OLIVEIRA, G. P. **Adquirir fluência e pensar com tecnologias em Educação Matemática: uma proposta com o software SUPERLOGO**, 2015.

MARQUES, M.; GUIMARÃES, G.; GITIRANA, V. **Compreensões de alunos e professores sobre média aritmética**. *Boletim de Educação Matemática*, v. 24, n. 40, p. 725-745, 2011.

MENDONÇA, L. O. **A educação estatística em um ambiente de modelagem matemática no ensino médio**. Dissertação (mestrado) – Universidade Cruzeiro do Sul, São Paulo, 2008.

MENDONÇA, L. O.; LOPES, C. E. **Modelagem matemática: um ambiente de aprendizagem para a implementação da educação estatística no ensino médio**. *Boletim de Educação Matemática*, v. 24, n. 40, p. 701-724, 2011.

MOORE, D. **New Pedagogy and new content: the case on Statistics**, *International Statistical Review*, v.65, 1997.

MORAIS, T. M. R. **Um estudo sobre o pensamento estatístico: componentes e habilidades**. Dissertação (mestrado) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2006.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. 8. ed. São Paulo: Saraiva, 2013.

NAGAMINE, C. M. L.; HENRIQUES, A.; UTSUMI, M. C.; CAZORLA, I. M. **Análise praxeológica dos passeios aleatórios da Mônica**. *Boletim de Educação Matemática*, v. 24, n. 39, p. 451-472, 2011.

NOVAES, D. V. **A mobilização de conceitos estatísticos: estudo exploratório com alunos de um curso de tecnologia em Turismo**. Dissertação (mestrado em educação matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2004.

NOVAES, D. V. **Concepções de professores da educação básica sobre variabilidade estatística**. Tese (doutorado em educação matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2011.

NOVAES, D. V.; COUTINHO, C. Q. S. **Estatística para a educação profissional**. São Paulo: Atlas, 2009.

OLIVEIRA, G. P. **Sobre tecnologias e educação matemática - fluência, convergência e o que isto tem a ver com aquilo** (páginas 21-80). Texto do livro *Educação Matemática - epistemologia, didática e tecnologia* (2018).

PFANNKUCH, M.; BEN-ZVI, D. **Developing teachers' statistical thinking**. In: **Teaching statistics in school mathematics: challenges for teaching and teacher education**. Netherlands: Springer, 2011. p. 323-333.

PONTE, J. P. **Estudos de caso em educação matemática.** *Boletim de Educação Matemática*, v. 25, p. 103-132, 2006.

PONTE, J. P.; BROCARD, J.; OLIVEIRA, H. **Investigações matemáticas na sala de aula.** Belo Horizonte: Authentic, 2005.

PORCIÚNCULA, M.; PINTO, S. S. **Teaching statistics through learning projects.** *Statistics Education Research Journal of Education (IASE/ISI)*, v. 13, n. 2, p. 177-186, 2014.

PORCIÚNCULA, M.; SAMÁ, S. **Projetos de aprendizagem: uma proposta pedagógica para a sala de aula de estatística.** In. SAMÁ, S.; PORCIÚNCULA, M. (Orgs.). *Educação estatística: ações e estratégias pedagógicas no ensino básico e superior.* Curitiba: CRV, 2015.

ROBERT, A. **Outils D'Analyse des Contenus Mathématiques à enseigner au lycée et à l'Université.** RECHERCHES EM DIDACTIQUE DES MATÉMATIQUES, Vol.18, nº2, pp.139-190. 1998.

RUMSEY, D. J. **Statistical literacy as a goal for introductory statistics courses.** *Journal of Statistics Education*, v. 10, n. 3, p. 6-13, 2002.

SÁ, D. L. **Elaboração e análise de um instrumento para verificar informações acerca do letramento estatístico de estudantes concluintes do ensino médio.** Dissertação (mestrado em educação em ciências) - Universidade Federal do Rio Grande, Rio Grande, RS, 2015.

SEVERINO, A. J. **Metodologia do trabalho científico.** São Paulo: Cortez, 2007.

SHAMOS, M. H. **The myth of scientific literacy.** New Brunswick: Rutgers University Press, 1995.

SILVA, C. B., **Pensamento estatístico e raciocínio sobre variação: um estudo com professores de matemática.** Tese (doutorado em educação matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2007.

SILVA, C. X.; BARRETO FILHO, B. **Matemática aula por aula.** São Paulo: FTD, 2005. 3 v.

SOARES, M. **Letramento e escolarização.** In: RIBEIRO, V. M. (Org.). *Letramento no Brasil.* São Paulo: Global, 2003. p. 89-113.

SOUZA, F.S. **A Importância das Atividades para o Ensino de Estatística na Educação Básica com o Uso do Software R e seu Pacote de Dados Rcmdr no III SIPEMAT.** *Perspectivas da Educação Matemática*, Campo Grande, MS, v.6, n. 11, p. 85-92, jan./jun. 2013.

STELLA, C. A. **Um estudo sobre o conceito de média, com alunos do ensino médio.** Dissertação (mestrado em educação matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2003.

TOMAZ, V. S.; DAVID, M. M. M. S. **Interdisciplinaridade e aprendizagem da matemática em sala de aula**. Belo Horizonte: Autêntica, 2012.

VENDRAMINI, C. M. M. **Implicações das atitudes e das habilidades Matemáticas na aprendizagem dos conceitos de Estatística**. Tese de doutorado. UNICAMP/SP, 2000.

VIEIRA, M. **Análise exploratória de dados: uma abordagem com alunos do Ensino Médio**. Dissertação (mestrado em educação matemática) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2008.

WATSON, J. **Assessing statistical thinking using the media**. In: GAL, I.; GARFIELD, J. (Orgs.) *The assessment challenge in statistics education*. Amsterdam: IOS, 1997.

WILD, C. J.; PFANNKUCH, M. **Statistical thinking in empirical enquiry**. *International Statistical Review*, v. 67, n. 3, p. 223-248, 1999.

WADA, R. S. **Estatística e Ensino: um estudo sobre representação de professores do 3º grau**. Tese de Doutorado. Universidade Estadual de Campinas (2006), Campinas, SP, 1996.

YAMAUTI, M. M. **Regressão linear simples nos livros de estatística para cursos de Administração: um estudo didático**. Dissertação (mestrado) - Pontifícia Universidade Católica, São Paulo, 2013.

ANEXOS

ANEXO 1 - CARTA DE ESCLARECIMENTO SOBRE O PROJETO E A PESQUISA

PROJETO: PROCESSOS DE ENSINO E APRENDIZAGEM DE MATEMÁTICA EM AMBIENTES TECNOLÓGICOS.

Pesquisador: Saddo Ag Almouloud

Informações sobre o projeto e sobre a pesquisa:

A pesquisa a ser realizada faz parte de projeto em desenvolvimento no Programa de Estudos Pós Graduated em Educação Matemática da Pontifícia Universidade Católica de São Paulo (PUC-SP).

A pesquisa tem como objetivos:

- Dar subsídios teóricos ao trabalho de professores e pesquisadores interessados na integração de pesquisas em Educação Matemática e de tecnologias em aulas de Matemática.
- Produzir conhecimento na área de formação de professores de matemática, utilizando tecnologias como recurso de apoio (como instrumento, no sentido que explicitaremos mais adiante neste texto).
- Elaborar em conjunto uma sequência utilizando o Geogebra e/ou Cabri 3D para ensinar conteúdos matemáticos para ser aplicada com grupo de professores de Ensino Básico e Superior.
- Contribuir para os debates a respeito do papel das tecnologias na formação de professores

Para este fim, serão levantados dados a partir de observações, questionários, documentos produzidos pelos participantes da pesquisa.

A pesquisa intitulada “A construção do letramento estatístico em estratégias com o uso de tecnologias digitais em aulas de Estatística de cursos de graduação ” é subprojeto de Processos de Ensino e Aprendizagem e Matemática em Ambientes Tecnológicos, conjugando assim dos objetivos e procedimentos metodológicos aqui apresentados.

Destacamos que, a qualquer momento, os participantes são livres para deixar de participar da pesquisa, sem qualquer ônus ou penalização de qualquer parte. Caracteriza-se firmemente a participação voluntária.

Os dados coletados serão tratados de forma a manter o sigilo dos participantes, sem possibilidade de identificação dos mesmos.

ANEXO 2 - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**Pesquisa: PROCESSOS DE ENSINO E APRENDIZAGEM DE MATEMÁTICA EM AMBIENTES TECNOLÓGICOS**

Eu, _____, com _____ anos de idade, portador (a) do RG _____, residente na _____, com número de telefone _____ e e-mail _____, abaixo assinado, dou meu consentimento livre e esclarecido para participar como voluntário da pesquisa supra citada, sob a responsabilidade do Prof. Dr Saddo Ag Almouloud, vinculada ao Programa de Estudos Pós-graduados em Educação Matemática da PUC-SP.

Assinando este Termo de Consentimento, estou ciente de que:

- 1) O objetivo da pesquisa.
- 2) Assim que for terminada a pesquisa, o relatório com seus resultados será divulgado para a comunidade acadêmica, mas primeiramente para o grupo de participantes.
- 3) Estou livre para interromper, a qualquer momento, minha participação nesta pesquisa.
- 4) Poderei entrar em contato com o pesquisador responsável, Saddo Ag Almouloud pelo e-mail saddoag@pucsp.br.
- 5) Obtive todas as informações necessárias para poder decidir conscientemente sobre a minha participação na referida pesquisa;

São Paulo, _____ de _____ de 2019.

Assinatura do participante

Assinatura do pesquisador

ANEXO 3 - SOLICITAÇÃO 1 DE AUTORIZAÇÃO PARA A PESQUISA

São Paulo, 22 de Junho de 2018

Ao **Alexandre Sousa da Silva**,

Eu, Anderson Anzai dos Santos, responsável pelo projeto de pesquisa do curso Mestrado Acadêmico em Educação Estatística, da Pontifícia Universidade Católica de São Paulo (PUC SP) , pertencente ao grupo CEPEF - Estudos em Probabilidade, Estatística e Educação Financeira, venho por meio deste documento, solicitar a autorização para realizar a pesquisa com alunos dos curso de Graduação em Enfermagem.

A pesquisa tem por objetivo investigar as contribuições do *software* Estatístico R para a construção do Letramento Estatístico em cursos de graduação.

O trabalho é orientado pela professora Dra. Cileda de Queiroz e Silva Coutinho, que também assina nesse documento.

Coloco-me à disposição para qualquer tipo de esclarecimento.

Atenciosamente,

Anderson Anzai dos Santos

Cileda de Queiroz e Silva Coutinho

Alexandre Sousa da Silva

ANEXO 4 - SOLICITAÇÃO 1 DE AUTORIZAÇÃO PARA A PESQUISA

São Paulo, 22 de Junho de 2018

À Luciane de Souza Velasque ,

Eu, Anderson Anzai dos Santos, responsável pelo projeto de pesquisa do curso Mestrado Acadêmico em Educação Matemática, da Pontifícia Universidade Católica de São Paulo (PUC SP) , pertencente ao grupo CEPEF - Estudos em Probabilidade, Estatística e Educação Financeira, venho por meio deste documento, solicitar a autorização para realizar a pesquisa com alunos dos curso de Graduação em Medicina e Licenciatura em Matemática.

A pesquisa tem por objetivo investigar as contribuições do software Estatístico R para a construção do Letramento Estatístico em cursos de graduação.

O trabalho é orientado pela professora Dra. Cileda de Queiroz e Silva Coutinho, que também assina nesse documento.

Coloco-me à disposição para qualquer tipo de esclarecimento.

Atenciosamente,

Anderson Anzai dos Santos

Cileda de Queiroz e Silva Coutinho

Luciane de Souza Velasque

APÊNDICES

APÊNDICE 1 - QUESTIONÁRIO APLICADO ANTES DA REALIZAÇÃO DAS ATIVIDADES

Curso: _____ Idade: _____

Sexo: () Masculino () Feminino

As notas de 1 a 4, representam o seu grau de concordância com a questão, ou seja, sendo 1 o que você menos concorda e o 4 o que você mais concorda.

1. Qual o seu o nível de interesse pela Estatística antes de entrar na graduação?

() 1 () 2 () 3 () 4

2. Classifique seus conhecimentos de Estatística antes de entrar na graduação.

() 1 () 2 () 3 () 4

3. Qual o seu nível de interesse pela Estatística após entrar na Graduação, com o contato com essa disciplina?

() 1 () 2 () 3 () 4

4. Na sua visão, qual a importância da Estatística no cotidiano das pessoas?

() 1 () 2 () 3 () 4

5. Na sua visão, qual a importância da Estatística na sua profissão?

() 1 () 2 () 3 () 4

6. Na sua visão, qual a importância em usar algum software (por exemplo, o R) no ensino da Estatística?

() 1 () 2 () 3 () 4

7. Na sua visão, qual a importância de trabalhar com projetos (estudos contextualizados) no ensino da Estatística?

() 1 () 2 () 3 () 4

8. Na sua visão, qual a importância da Estatística para a formação da criticidade no cidadão?

() 1 () 2 () 3 () 4

9. Comparando o conhecimento de Estatística antes e depois de entrar na Graduação, o quanto você acredita que colaborou para sua evolução pessoal?

() 1 () 2 () 3 () 4

10. Comparando o conhecimento de Estatística antes e depois de entrar na Graduação, o quanto você acredita que colaborou para sua evolução profissional?

() 1 () 2 () 3 () 4

APÊNDICE 2 - QUESTIONÁRIO APLICADO APÓS A REALIZAÇÃO DAS ATIVIDADES

Curso: _____ Idade: _____

Sexo: () Masculino () Feminino

As notas de 1 a 4, representam o seu grau de concordância com a questão, ou seja, sendo 1 o que você menos concorda e o 4 o que você mais concorda.

1. Qual era o seu conhecimento do R antes dessa atividade?

() 1 () 2 () 3 () 4

2. Após essa atividade, o quanto acredita que o uso de um software (especificamente o R), pode contribuir para o ensino da Estatística?

() 1 () 2 () 3 () 4

3. Quais as contribuições do R para o cálculo de métricas (média, mediana, desvio padrão, variância) e análises descritivas e seu apoio para sua formação pessoal?

() 1 () 2 () 3 () 4

4. Quais as contribuições do R para o cálculo de métricas (média, mediana, desvio padrão, variância) e análises descritivas e seu apoio para sua formação profissional?

() 1 () 2 () 3 () 4

5. Quais as contribuições do R para a construção de gráficos e seu apoio para sua formação pessoal?

() 1 () 2 () 3 () 4

6. Quais as contribuições do R para a construção de gráficos e seu apoio para sua formação profissional?

() 1 () 2 () 3 () 4

7. O quanto o R contribuiu para a sua formação crítica dos dados, pensando em sua articulação com seu cotidiano?

() 1 () 2 () 3 () 4

8. O quanto o R contribuiu para a sua formação crítica dos dados, pensando em sua articulação com seu futuro profissional?

() 1 () 2 () 3 () 4