



**SOCIEDADE BRASILEIRA DE MATEMÁTICA  
FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA  
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL**

**LEONARDO MOTA DE ANDRADE**

**ESTATÍSTICA NO ENSINO MÉDIO: UMA PROPOSTA DE ENSINO  
USANDO O *SOFTWARE R*.**

**PORTO VELHO  
2016**



**SOCIEDADE BRASILEIRA DE MATEMÁTICA  
FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA  
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL**

**LEONARDO MOTA DE ANDRADE**

**ESTATÍSTICA NO ENSINO MÉDIO: UMA PROPOSTA DE ENSINO  
USANDO O *SOFTWARE R*.**

Trabalho de Conclusão apresentado ao Mestrado Profissional em Matemática em Rede Nacional – PROFMAT no Polo da Universidade Federal de Rondônia – UNIR, como requisito parcial para obtenção do título de Mestre em Matemática Profissional.  
Orientador: Prof. Dr. Flávio Batista Simão

**PORTO VELHO  
2016**

## Ficha Catalográfica

A553e Andrade, Leonardo Mota

Estatística no Ensino Médio: uma proposta de ensino usando o *software R* / Leonardo Mota Andrade – Porto Velho : UNIR, 2016.  
96 f. : il.

Orientador: Prof. Flávio Batista Simão.

Dissertação (mestrado) – Universidade Federal de Rondônia,  
Área de Matemática Profissional, 2016.

1. Matemática Aplicada. 2. Estatística – Estudo e ensino. 3.  
Estatística – *Software*. I. Simão, Flávio Batista. II. Título.

CDD 519.5

CDU 519.2

Elaborada pela bibliotecária Cleuza Diogo Antunes CRB 11/864

**Leonardo Mota de Andrade**

**ESTATÍSTICA NO ENSINO MÉDIO: UMA PROPOSTA DE ENSINO USANDO O  
*SOFTWARE R.***

Este trabalho foi julgado Aprovado para obtenção do título de Mestre em Matemática do Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, do Departamento de Matemática da Fundação Universidade Federal de Rondônia, *Campus* de Porto Velho.

**Banca Examinadora:**

---

Prof. Dr. Flávio Batista Simão  
Orientador

---

Prof. Dr. Tomás Daniel Menéndez Rodríguez

---

Prof<sup>a</sup>. Dra. Maria das Graças Viana de Souza

## AGRADECIMENTOS

Primeiramente agradeço a **Deus**, Criador e Mantenedor de todas as coisas. Que de forma maravilhosa me permitiu concluir esta importante etapa da minha vida.

Aos meus pais **Samuel** e **Virgínia** pela educação, exemplo e apoio em todos os momentos da minha vida, amo vocês.

A minha esposa **Rosana**, pelo apoio, carinho, incentivo nas horas difíceis, compreensão na hora da ausência o meu muito obrigado, te amo.

As minhas duas princesas, minhas filhas **Juliana** e **Ana Júlia** que ao longo destes 2 anos de estudos em alguns momentos tiveram que conviver com a minha ausência ou não assistir TV pois o papai tinha que estudar, papai ama vocês.

Ao meu irmão **Maxwel** e sua esposa **Juliana** pelo apoio, incentivo e estadia em Porto Velho.

Ao meu irmão **Kenedy** e sua esposa **Tacide** pelo apoio, incentivo e o empréstimo do carro para algumas viagens a Porto Velho.

Ao amigo e advogado **Miguel Ângelo** pela dedicação em buscar de forma judicial o direito de eu fazer as provas após o pôr-do-sol de sábado, sem a sua contribuição isto não seria possível.

Ao meu orientador, **Prof. Dr. Flávio Simão**, pela disposição em me acompanhar neste processo, pelas orientações e pelos almoços sou muito grato.

Aos professores **Abel Ahbid Ahmed Delgado Ortiz**, **Adeilton Fernandes**, **Marinaldo Felipe da Silva**, **Marizete Nink de Carvalho**, **Ronaldo Cavalcanti** e **Tomás Daniel Menéndez Rodríguez**, por terem proporcionados momentos de aprendizado e crescimento profissional.

Aos amigos de mestrado **Emerson**, **Elihebert**, **Evandro** e **Marcelo** pelos grupos de estudos realizados em diversas ocasiões, pelo companheirismo e pelos ensinamentos.

Aos **amigos** e **familiares** que me apoiaram nos momentos de dificuldade com palavras de incentivo e orações.

## RESUMO

Vive-se um momento na educação em que não é mais possível se eximir do uso das tecnologias e suas aplicações em sala de aula. Esta dissertação propõe mostrar que o *software R* pode ser usado com efetividade como ferramenta auxiliar de ensino de estatística para o ensino médio com uso de oficinas e aulas teóricas no laboratório de informática. A metodologia adotada é a engenharia didática, que se caracteriza como uma sequência de aulas concebidas, organizadas e articuladas no tempo, de forma constante, por um professor para realizar um projeto de aprendizagem para certa população de alunos. Os participantes realizaram duas avaliações uma a priori e outra a posteriori. Os alunos participaram de cinco oficinas onde aprenderam a utilizar o *software R* para resolver situações em que envolvia estatística básica. Durante a realização das oficinas os alunos registraram os pontos positivos, pontos negativos e as dificuldades encontradas. Os resultados obtidos mostram que o uso do *software R* contribuiu para o aprendizado dos alunos, pois a média do teste a priori foi de 52,5 enquanto que a média do teste a posteriori foi de 61,25 numa escala de 0 a 100. A comparação das médias foi realizada com o Teste t de *Student* para amostras pareadas e o resultado foi estatisticamente significativo, ou seja, a média do teste a posteriori foi significativamente maior do que a média do teste a priori. Com os resultados obtidos concluímos que o uso deste *software* pode contribuir de forma significativa para o ensino e aprendizagem de estatística no ensino médio.

**Palavras-chave:** *Software R*. Estatística. Ensino Médio.

## ABSTRACT

We live in a time where it is no longer possible for education to escape the use of different technologies or their applications in the classroom. This dissertation proposes the use of *software R* can be used effectively as an auxiliary tool of statistical teaching for high school, with the use of workshops and theoretical classes in the computer lab. The methodology adopted is didactic engineering; which is characterized as a sequence of classroom(s) designed, organized and articulated in time, in a constant way, by an teacher to carry out a learning project for a certain student population. The participants performed two evaluations, *a priori* and *a posteriori*. Students participated in five workshops, where they learned to use the *software R* to solve situations involving basic statistics. During the workshops, students recorded the positive and negative results and the difficulties they encountered. The results obtained show that the use of the *software R* contributed to the students' learning, since the *a priori* test average was 52.5 while the *a posteriori* test had an average of 61.25 on a scale of 0 to 100. The comparison of the averages was performed with the *Student* t-test for paired samples and the result was statistically significant, that is, the average of the *a posteriori* test was higher than the *a priori* test average. With the results obtained, we can conclude that the use of this *software* could contribute significantly to the teaching and learning of statistics in high schools.

**Keywords:** *Software R*. Statistic. High school.

## LISTA DE FIGURAS

<b>Figura 01:</b> Janela da Interface do Usuário do R no Windows .....	<b>23</b>
<b>Figura 02:</b> Janela da Interface do Usuário do RStudio no Windows .....	<b>24</b>



## LISTA DE GRÁFICOS

<b>Gráfico 01:</b> Opinião de 16 alunos quanto ao time que será campeão brasileiro da série A em 2016.....	38
<b>Gráfico 02:</b> Opinião de 16 alunos quanto ao time que será campeão brasileiro da série A em 2016.....	38
<b>Gráfico 03:</b> Opinião de 16 alunos quanto ao time que será campeão brasileiro da série A em 2016.....	40
<b>Gráfico 04:</b> Histograma das idades de 40 turistas que visitaram Ji-Paraná em 7 de setembro de 2016. ....	41
<b>Gráfico 5:</b> Gráfico de linha da porcentagem de votos para os três candidatos mais bem posicionados na campanha para prefeito de Ji-Paraná/RO em 2016.....	42

## LISTA DE TABELAS

<b>Tabela 1:</b> Lista dos 10 modelos de automóveis mais vendidos no Brasil no mês de agosto de 2016. ....	31
<b>Tabela 2:</b> Tabela de frequência de ovos danificados em um carregamento para o mercado municipal de Lavras.....	32
<b>Tabela 3:</b> Tabela de frequência que mostra a opinião de 16 alunos em setembro de 2016, sobre qual time vai ganhar o campeonato brasileiro série A. ....	33
<b>Tabela 4:</b> Idade de 40 turistas que visitaram Ji-Paraná em 7 de setembro de 2016. ....	36
<b>Tabela 5:</b> Porcentagem de votos para os três candidatos mais bem posicionados na campanha para prefeito de Ji-Paraná/RO em 2016.....	41
<b>Tabela 6:</b> Tabela de frequência de ovos danificados em um carregamento para o mercado municipal de Lavras, com coluna auxiliar para o cálculo da média.....	44
<b>Tabela 7:</b> Idade de 40 turistas que visitaram Ji-Paraná em 7 de setembro de 2016 com coluna auxiliar para cálculo de média. ....	45
<b>Tabela 8:</b> Tabela com o número de gols por partida registrado na 24ª e 25ª rodadas do campeonato série A 2016. ....	48
<b>Tabela 9:</b> Tabela com o resultado de 40 observações relativas ao índice pluviométrico em Ji-Paraná/RO nos dias chuvosos de janeiro a março de 2016. ....	48
<b>Tabela 10:</b> Tabela de resultados de exame de colesterol de 200 funcionários que foram submetidos a exames clínicos.....	50
<b>Tabela 11:</b> Tabela com o resultado do número de peças vendidas pela rede de lojas Z, durante a promoção de natal de 2015. ....	51
<b>Tabela 12:</b> Tabela de valores de uma amostra do tempo de vida útil da bateria de uma marca de celular.....	53

<b>Tabela 13:</b> Tabela de valores de uma amostra do tempo de vida útil da bateria de uma marca de celular com colunas auxiliares para o cálculo do desvio médio. ....	55
<b>Tabela 14:</b> Tabela de valores de diárias praticada no litoral norte de São Paulo na período de baixa temporada de 2016.....	57
<b>Tabela 15:</b> Notas dos alunos na análise a priori sobre conhecimentos de estatística básica realizada no IFRO/Ji-Paraná-RO no segundo semestre de 2016.....	61
<b>Tabela 16:</b> Notas dos alunos na análise a posteriori sobre conhecimentos de estatística básica realizada no IFRO/Ji-Paraná-RO no segundo semestre de 2016.....	62

# SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>13</b>
<b>1-REFERÊNCIAL TEÓRICO.....</b>	<b>15</b>
1.1-As tecnologias digitais no processo ensino aprendizagem. ....	15
1.2-O uso de software na prática escolar .....	20
<b>2-O PROGRAMA R.....</b>	<b>22</b>
2.1-Instalando o R.....	22
2.3-Estrutura de Dados .....	24
<b>3-PRINCIPAIS CONCEITOS DE ESTATÍSTICA .....</b>	<b>27</b>
<b>3.1- Divisões da Estatística.....</b>	<b>27</b>
3.1.1- Estatística descritiva .....	27
3.1.2- Estatística Inferencial .....	27
3.2- População e Amostra.....	27
<b>3.3- Amostragem.....</b>	<b>28</b>
<b>3.4- Variável.....</b>	<b>28</b>
3.4.1- Variável Quantitativa .....	28
3.4.2- Variável Qualitativa .....	29
<b>3.5- Dados Brutos e Rol.....</b>	<b>29</b>
<b>3.6- Tabelas .....</b>	<b>30</b>
<b>3.7- Tabelas de frequências.....</b>	<b>31</b>
<b>3.8- Representação Gráfica.....</b>	<b>37</b>
3.8.1- Gráfico de barras ou colunas .....	37
3.8.2- Gráfico de setores (pizza).....	39
3.8.3- Histograma .....	40
3.8.4- Gráfico de linhas .....	41
<b>3.9- Medidas de Tendência Central .....</b>	<b>42</b>
3.9.1- <i>Média aritmética</i> .....	42
3.9.1.1- Média aritmética para dados não agrupados.....	43
3.9.1.2- Média aritmética ponderada (dados agrupados) .....	44
3.9.2- <i>Mediana</i> .....	45
3.9.2.1- Mediana para dados não agrupados.....	46
3.9.2.2- Mediana para dados agrupados sem intervalo de classes .....	47
3.9.2.3- Mediana para dados agrupados com intervalo de classes .....	49
3.9.3- <i>Moda</i> .....	50
3.9.3.1- Moda para dados não agrupados .....	50
3.9.3.2- Moda para dados agrupados sem intervalo de classe .....	51
3.9.3.3- Moda para dados agrupados com intervalo de classe .....	52
<b>3.10- Medidas de variabilidade.....</b>	<b>53</b>

3.10.1- Desvio médio.....	54
3.10.2- Variância e desvio padrão .....	56
<b>4-METODOLOGIA E PROCEDIMENTOS.....</b>	<b>59</b>
4.1- Engenharia Didática .....	59
4.2- Descrição das fases do projeto.....	60
<b>5-ANÁLISE E DISCUSSÃO DOS RESULTADOS .....</b>	<b>64</b>
<b>6-CONSIDERAÇÕES FINAIS .....</b>	<b>71</b>
<b>7-REFERÊNCIAS .....</b>	<b>72</b>
<b>APÊNDICE A .....</b>	<b>74</b>
<b>LISTA DE INTERESSADOS.....</b>	<b>74</b>
<b>APÊNDICE B.....</b>	<b>75</b>
<b>APÊNDICE C .....</b>	<b>76</b>
<b>APÊNDICE D .....</b>	<b>77</b>
<b>APÊNDICE E.....</b>	<b>78</b>
<b>APÊNDICE F .....</b>	<b>83</b>
<b>APÊNDICE G .....</b>	<b>84</b>
<b>APÊNDICE H.....</b>	<b>86</b>
<b>APÊNDICE I.....</b>	<b>88</b>
<b>APÊNDICE J .....</b>	<b>89</b>
<b>APÊNDICE K .....</b>	<b>90</b>
<b>APÊNDICE L.....</b>	<b>91</b>

## INTRODUÇÃO

A tecnologia caminha a passos largos. A sociedade mudou. O que acontece aqui hoje e agora o mundo inteiro sabe em milésimos de segundos. Não dá mais para viver sem interagir com a tecnologia.

A forma de compreender o mundo e tudo o que se vive diariamente já não é tão simples dadas às complexidades da modernidade.

Isso mostra que a sociedade não mais suporta uma prática escolar estática que deixe de fazer uso das tecnologias que estão presentes na sociedade.

No entanto não basta introduzir aparelhos tecnológicos pois a escola tem a função de formar indivíduos pensantes e que se apoderem da tecnologia em benefício da cidadania. Sabe-se que a educação é direito de todos e dever do Estado e da família, sendo uma atividade que deve preparar a pessoa para o exercício da cidadania e para sua qualificação profissional.

Dentre os demais saberes o saber matemático é que possibilitará ao aluno/cidadão ser capaz de decidir sobre as vantagens e/ou desvantagens de uma compra à vista ou à prazo; avaliar o custo de um produto em função da quantidade; conferir se estão corretas informações em embalagens de produtos quanto ao volume; calcular impostos e contribuições previdenciárias; avaliar modalidades de juros bancários, interpretar gráficos, tabelas e dados numéricos veiculados nas diferentes mídias; ler faturas de contas de consumo de água, luz e telefone, ou seja, exercer de forma consciente e autônoma a sua cidadania.

O mundo pós-moderno se caracteriza pela velocidade com que as informações circulam e pela quantidade destas informações. Cada pessoa deve ser capaz de analisar e interpretar dados, para a tomada de decisões. Desta forma, a estatística desempenha um papel muito importante, pois auxilia no desenvolvimento de competências sócias cognitivas nos indivíduos e a apropriação dos seus conhecimentos é essencial para o exercício de uma cidadania plena.

Embora os alunos tenham contato com estes conceitos no ensino médio geralmente não aprendem a dominar nenhuma ferramenta que possibilitem a eles uma análise de dados de forma rápida e prática que mesmo poderia ser usada num ambiente profissional posteriormente.

Diante destas reflexões elaboramos este trabalho cujo objetivo principal é mostrar que o *software R* pode ser usado com efetividade como ferramenta auxiliar de ensino de estatística para o ensino médio.

Os objetivos específicos são:

- Ensinar os alunos a utilizarem o *software R* no campo da estatística descritiva na vida acadêmica, social e profissional;
- Capacitar os alunos a apliquem corretamente os conceitos de estatística básica em diversas situações;
- Comparar o rendimento dos alunos em estatística básica com uma avaliação a priori e outra avaliação a posteriori ao uso do *software R*.

Este trabalho se torna relevante para a formação social, acadêmica e profissional do aluno, pois possibilita ao aluno dominar uma nova ferramenta que o auxilie na análise e interpretação de dados para a tomada de decisões.

Adotamos como metodologia na execução desta pesquisa a Engenharia Didática que é uma metodologia de validação de resultados de uma pesquisa educacional, proposta pela pesquisadora francesa Michelle Artigue.

## 1- REFERÊNCIAL TEÓRICO

### 1.1-As tecnologias digitais no processo ensino aprendizagem.

É uma característica do ser humano buscar facilitar sua vida em sociedade. Segundo ORWELL, escritor britânico do século XX, citado no site aprendendo e brincando com tecnologias, 2013: “O homem é tão bom quanto seu desenvolvimento tecnológico o permite ser.”

Desde os primeiros registros da existência humana, percebeu-se uma evolução nas técnicas de sobrevivência para assim se adaptar no meio em que viviam e conseqüentemente vencer os desafios de cada época. Trilhou-se milênios e o mundo transformou-se e estamos na era digital.

Está inerente a esse trabalho essa essência da busca humana. As adaptações ou descobertas para melhorar a sociedade, competir no mundo moderno e pós moderno a técnica ou tecnologia é que será foco dessa temática porém delimitando-se no papel da escola no processo de construção do saber institucionalizado, o saber escolar, uma vez que o foco da pesquisa é a utilização de um *software* na aprendizagem de um conteúdo, ou seja o uso da tecnologia digital de forma a beneficiar o fazer educativo no componente curricular de matemática no estudo do conteúdo de estatística.

Sabendo que o conhecimento é dinâmico e somos “seres inacabados” faz-se necessário ter em mente que daqui alguns anos os instrumentos tecnológicos de hoje talvez seja uma bobagem aos olhos do amanhã.

Não é raro ouvir-se comentários até com certa razão de que a escola pouco mudou, pouco caminhou é arcaica e que ainda tenta-se ensinar com “giz e saliva”. Antes dessas indagações quase num flash vem uma alerta: será que os estudantes dessa década chegaram a conhecer o giz? Porém para alguns de nós parece que ele, o giz, esse magnífico instrumento de um determinado tempo, também fruto da tecnologia humana, não parece tão distante.

Há o entendimento, em que se tratando de meios tecnológicos, a escola brasileira caminhou a passos lentos, no entanto não se pode negar que nas últimas décadas iniciou-se um processo de inserção tecnológica muito importante. É claro que se pensarmos na nossa história e no contexto educacional brasileiro de forma reflexiva veremos que isso ainda é pouco porém temos um processo histórico que ao longo dos anos aplicou poucos insumos na educação. Mas é necessário caminhar e entender que hoje:



No contexto escolar atual, é impensável fazermos algumas tarefas sem a ajuda de um computador. Pilhas de cadernos, agendas e planilhas de papel foram substituídas por arquivos no computador, que facilitam o fechamento de notas, o controle de presenças, a emissão do histórico dos alunos, etc. Provas são ricamente elaboradas com o uso de *softwares*, internet e editores de texto. Chega um momento, porém, em que a presença de alguns recursos tecnológicos deve deixar de ser imprescindível apenas no espaço administrativo e ocupar seu lugar onde será mais útil e mais ricamente aproveitada: a sala de aula. (SOUZA, 2016, REVISTA PÁTIO)

É certo que ainda tem-se muito caminho a percorrer e histórias a serem reescritas no entanto é importante ressaltar que por longas décadas o giz, o quadro negro, o mimiógrafo reinavam como uma proeza tecnológica existente na escola. E com o tempo incluiu-se o retroprojetor simples com lâminas, depois, mais sofisticados, depois de longos períodos seguiu-se com a TV no formato do programa TV escola com a tecnologia da parabólica, e por fim a era dos computadores, isso na escola pois em outros setores os computadores já haviam chegado.

Na atualidade uma boa parte das escolas estão equipadas com PCs com seus aplicativos e demais programas, porém a maioria dos professores não se adequaram a esse tempo a utilização para fins de aprendizagem e pesquisa que ainda é pouca. Para o professor Moran, nessa nova era educacional já não cabem à acomodação e os roteiros prontos. “Temos muitos professores que preferem receber tudo pré-formatado e não ter que planejar nem preparar detalhadamente cada aula e atividade”. Isso explica o sucesso dos sistemas de ensino que oferecem aulas prontas sobre todas as matérias da Educação Básica.

Evidencia-se nas escolas, alunos super conectados tecnologicamente transitando de forma fluente entre *software*, aplicativos e programas todavia com desinteresse pelas aulas. O maior desafio está no ensino médio. Por esse motivo a pesquisa científica acadêmica precisa pensar em desenvolver caminhos para uso das tecnologias de forma sedutora e que facilitem o desenvolvimento do currículo de forma que os estudantes não passem apenas pela escola e a eles sejam negados novas possibilidades para acesso ao conhecimento.

Vive-se um momento na educação em que não é mais possível se eximir do mundo do uso das tecnologias e suas aplicações em sala de aula. A escola precisa se munir de instrumentos tecnológicos como pressupostos para preparar indivíduos que nesse meio tecnológico se identifique como protagonista na construção de saberes significativos.

O fato é que a era digital está desencadeando uma nova forma de organização humana, e a educação não pode cometer o erro de pensar que possa estar imune a essa revolução, pois este será o novo caminho dos processos diversos da sociedade. Preparando o indivíduo para pensar digitalmente, ele poderá usufruir em conjunto das informações que fluem na rede eletrônica, e transformar isto em valor, seja em produto ou serviço, e

para seu crescimento humano, quando o acesso à informação for bem mais amplo, as barreiras hierárquicas, diminuirão na organização social aumentando as transformações sociais, culturais e até mesmo o sentido de nacionalidade ganhara novos valores, a comunicação vai fluir em todas as direções. (TAVARES, 2007)

A escola precisa desenvolver a capacidade de inserir a tecnologia atual, na perspectiva de um pensamento capaz de utilizar essa tecnologia no mundo presente e também no futuro e sem sombra de dúvidas que ser digital é fundamental para o mundo moderno e a escola precisa caminhar a passos largos no sentido de criar indivíduos pensantes que tendo acesso a tecnologias desenvolvam novas técnicas que façam a diferença num mundo em que a competitividade cada vez mais bate a porta.

Mas para ser uma escola que utiliza esses instrumentos os professores devem buscar cada dia mais caminhos que estimulem os jovens a gostar e compreender esse mundo digital o uso de *software* no ensino pode tornar-se um elemento facilitador na ação docente e também na aprendizagem dos estudantes.

É evidente a insatisfação dos alunos em relação a aulas ditas "tradicionais", ou seja, aulas expositivas nas quais são utilizados apenas o quadro e o giz. O aprender por aprender já não existe: hoje, os alunos precisam saber para que e por que precisam saber determinado assunto. Essa é a típica aprendizagem utilitária, isto é, só aprendo se for útil, necessário para entrar no mercado de trabalho, visando ao retorno financeiro. A internet invade nossos lares com todas as suas cores, seus movimentos e sua velocidade, fazendo o impossível tornar-se palpável, como navegar pelo corpo humano e visualizar a Terra do espaço sem sair do lugar. É difícil, portanto, prender a atenção do aluno em aulas feitas do conjunto lousa + professor. Então, por que fazer o mesmo quando se pode fazer diferente? Uma vez que os alunos gostam tanto de aulas que utilizam a tecnologia, por que não aproveitar essa oportunidade e usá-la a seu favor? A aula pode entusiasmar os alunos de maneira ao menos parecida com que são excitados pelos jogos e filmes de alta qualidade em efeitos especiais. A escola precisa modernizar-se a fim de acompanhar o ritmo da sociedade e não se tornar uma instituição fora de moda, ultrapassada e desinteressante. Embora lentamente, ela está fazendo isso. Saber que o aluno aprende com o que lhe prende a atenção todos sabem. (SOUZA, 2016, REVISTA PÁTIO)

Por ser um grande caminho que produz saberes a educação, como instituição, terá que ir além dos paradigmas do giz e saliva ou do canetão e quadro branco, precisa entrar no universo dos chips e assim avançar e preparar indivíduos com novas competências para inserir-se na sociedade competitiva desenvolvendo a criticidade e autonomia de forma rigorosa, a tecnologia não deve ser usada para uma simples navegação de lazer na escola, o professor precisa preparar-se para a rigorosidade metódica tanto defendida por Paulo Freire que lembra que ensinar exige rigorosidade metódica.

Existem vários instrumentos tecnológicos digitais para utilizar-se na aprendizagem e há de se utilizá-los de forma crítica. Alcançar a criticidade exige comprometimento com o que se manipula. Não é mera aceitação. É preciso embate. Concordância e discordância, questionamentos sendo levantados sempre que são necessários dúvidas quanto as ideias defendidas, contrapontos e contestações, sempre baseado em argumentos. Rigorosidade metódica nos falta nestes tempos acelerados da internet. O tempo é um recurso escasso que nos escapa pelos dedos toda vez que gastamos segundos e minutos preciosos a deriva nas ondas da internet, sem saber ao certo por onde andamos o que fazemos. Método é processo. É a forma que estabelecemos para atingir uma meta, um objetivo. Nos estudos significa ler, realizar exercícios, traçar paralelos, buscar novas fontes, escrever sobre o que está sendo estudado, rever anotações anteriores, debater com os pares, concordar ou contestar... Significa também compromisso e comprometimento. Estudar sabendo o porquê desta ação, tornando-a consequente, associando-a a seu viver, transformando os elementos deste estudo em algo que seja útil a você e a sociedade. Rigor é o nível que se exige neste método que torna-se então vigoroso, forte, capaz de legar frutos se associado a disciplina, comprometimento, organização e planejamento claro de suas ações. Onde quero chegar? Como isso vai me ajudar? Há muito a se aprender quando a rigorosidade metódica passa a fazer parte das ferramentas tecnológicas que temos a mão. Para isso a escola precisa entender que apesar de que é importante que na escola se pense digitalmente também a partir de uma rigorosidade metódica.

Desse modo, é de se esperar que a escola, tenha que “se reinventar”, se desejar sobreviver como instituição educacional. É essencial que o professor se aproprie de gama de saberes advindos com a presença das tecnologias digitais da informação e da comunicação para que estes possam ser sistematizadas em sua prática rigorosa e pedagógica. A aplicação e mediação que o docente faz em sua prática pedagógica do computador e das ferramentas multimídia em sala de aula, depende, em parte, de como ele entende esse processo de transformação e de como ele se sente em relação a isso, se ele vê todo esse processo como algo benéfico, que pode ser favorável ao seu trabalho, ou se ele se sente ameaçado e acuado por essas mudanças. (SERAFIM & SOUSA, 2011, p. 20)

É preponderante salientar que:

Independentemente do recurso tecnológico em questão, o professor é o sujeito capaz de mediar o aprendizado e torná-lo mais atrativo, divertido e interessante para os alunos. Os recursos tecnológicos, bem mais do que aguçam a curiosidade do aluno em relação ao que está sendo ensinado, ajudam a prepará-lo para um mundo em que se espera que ele conheça, além dos conteúdos escolares, todos os recursos por meio dos quais esses conteúdos foram trabalhados.[...]

Assim, a escola cumpre duplamente seu papel: ensina e educa, educando para um mundo no qual a tecnologia é não só necessária, mas também essencial.

São muitos os benefícios trazidos pelos recursos tecnológicos à educação. Contudo, é preciso que o professor conheça as ferramentas que tem à sua disposição se quiser que o aprendizado aconteça de fato. O uso das tecnologias na escola está além de disponibilizar tais recursos; ele implica aliar método e metodologia na busca de um ensino mais interativo. (SOUZA, 2016, REVISTA PÁTIO)

Nesse contexto vale entender que a escola precisa disseminar um pensar tecnológico digital significativo, para assim imbuir uma constante postura de renovação impactando o futuro com melhorias na sociedade facilitando a busca por uma cidadania plena, possibilitando uma sensível transformação das camadas mais necessitadas de nossa população. É importante lembrarem-se do direito ao acesso a tudo o que for proposto como possibilidades no currículo da escola. E a tecnologia já é uma conquista de cidadania preconizada nas ações legais no planejamento educacional do Ministério da educação.

O aluno tem direito ao acesso às tecnologias na escola: "A sólida base teórica sobre informática educativa no Brasil existente em 1989 possibilitou ao MEC instituir, através da Portaria Ministerial nº 549/89, o Programa Nacional de Informática na Educação - PRONINFE, com o objetivo de desenvolver a informática educativa no Brasil, através de atividades e projetos articulados e convergentes, apoiados em fundamentação pedagógica sólida e atualizada, de modo a assegurar a unidade política, técnica e científica imprescindível ao êxito dos esforços e investimentos envolvidos" (História da informática educativa no Brasil, site do MEC/SEED/PROINFO, nome do autor não indicado).

Neste sentido pressupõe-se que a tecnologia digital é a sustentação econômica da nova sociedade, temos que ser também bons competidores digitais para obter sucesso no desenvolvimento de todos os setores sociais e a escola não pode se eximir do seu papel.

Um novo relacionamento vai surgir entre mercado, poder e educação, pois ter informação não será o problema, mas sim como transformar isto em conhecimento, dinheiro e evolução social.

Do jeito como esta à nova organização mundial, tão competitiva, não introduzir tecnologia na educação e ensinar a utilizá-la poder significar a paralisação dos processos de evolução econômica e social.

Introduzir computadores na escola, não é o suficiente para colocar a educação na era digital, uma educação digital precisa usar a tecnologia, principalmente para promover o entendimento das forças que organizam o meio em que participa, elevando o grau de interação, transformação e evolução.

A era digital reinventa as oportunidades de evolução social, pois cria uma demanda por profissionais capazes de pensar e operar digitalmente, o processo educacional tem que avançar rumo à era digital, para que os seus participantes sejam capazes de dotar o seu meio de infra-estrutura tecnológica, aplicando seus conhecimentos, para criar e redesenhar novas estratégias sociais.

A totalidade da sociedade não precisa dominar os pormenores da tecnologia digital, mas é imprescindível que saibam de suas existência e de como utilizá-la se não investirmos, ou investirmos errado vai estar fadado ao

sistema social vigente tão ingrato, podemos ainda piorá-lo mais com perdas de competitividade, não podemos demorar muito a mudar, e investir em tecnologia digital na educação, pois sem dúvida o futuro será digital.(TAVARES, 2007)

## 1.2-O uso de software na prática escolar

O uso dos *softwares* não significa introduzir tecnologia na escola, porém significa sim usar de forma pontual a tecnologia digital na sala de aula, ora já existente, pois esses são instrumentos facilitadores para solucionar problemas da era tecnológica e que podem também serem facilitadores no processo de ensino aprendizagem se utilizados de forma correta e metódica. Ao utilizar *software* o professor está utilizando a tecnologia já existente de forma a atingir um objetivo de aprendizagem que de forma tradicional não motivam os estudantes e tampouco possibilita entender sua utilidade para a vida na sociedade ou em futuras profissões desses estudantes.

Os *softwares* educativos possibilitam a interação dos aprendentes, encorajando-os a cooperação entre si, tornando possível a consolidação de uma aprendizagem colaborativa e a realização de atividades extra-sala de aula. As falas dos sujeitos e os posicionamentos dos aprendentes permitiram concluir, com base na teoria das Inteligências Múltiplas e na abordagem metodológica construtivista, que o uso de *softwares* educativos podem ativar habilidades, acentuar mais o desenvolvimento de certas inteligências, variando sua combinação de indivíduo para indivíduo. (DANTAS & AQUINO, 2007)

As referidas pesquisadoras (2007) lembram que a virada do século XXI coincide com a revolução das tecnologias da informação e comunicação que alteraram os antigos paradigmas de compreensão do mundo e na escola, as tecnologias trazem consigo a possibilidade de ter que reinventar a prática deixando de lado aquilo que é tradicional, geralmente deve-se repensar os currículos, conteúdos, metodologias para que o ambiente escolar seja de fato centrado nos estudantes.

Assim sendo, o reconhecimento de uma sociedade da aprendizagem exige a conscientização do docente diante da necessidade de familiarizar-se com a aquisição de recursos tecnológicos cada vez mais sofisticados que ajudem seus aprendentes a colocar-se diante de uma “nova organização do trabalho educativo”. Nesse sentido, as universidades precisariam de uma reforma no seu modo de pensar a educação, dotando os ambientes de aprendizagem de uma infra-estrutura capaz de dar conta, de forma satisfatória, das novas exigências tecnológicas, tais como a introdução de equipamentos sofisticados e *softwares* educativos, a fim de que professores e alunos possam acompanhar a evolução tecnológica. Exige-se um investimento em meios tecnológicos (*softwares* educativos) devidamente disponibilizados não

simplesmente como uma distribuição vertical de recursos didáticos, mas voltados para um contínuo processo de aprendizagem e que funcionem como estímulos à construção do conhecimento. (DANTAS & AQUINO, 2007)

Ainda ressaltam citando alguns teóricos enfatizam que:

A informação digital tornou-se a matéria-prima para a comunicação do conhecimento, a constituição de saberes e a interlocução com as diversas áreas e disciplinas. Os *softwares educativos*, como produto de uma sociedade digital, em que a técnica constitui um dos fatos dominantes da nossa época (ESCÓSSIA, 1999), são um dos meios tecnológicos mais utilizados na educação e a finalidade de sua inserção nos ambientes de aprendizagem (laboratórios, salas de aula, entre outros) visa proporcionar uma dinamização das atividades informacionais, servindo tanto para o estímulo de uma série de competências e desenvolvimento de habilidades, quanto para apoiar as disciplinas dos cursos. (DANTAS & AQUINO, 2007)

Vale referendar teóricos mais consistentes onde enfatizam que os *softwares educativos* tem grande preponderância nos processos de construção da autonomia nos estudantes pois podem “[...] ativar o acesso à informação e à pesquisa individual e coletiva, favorecendo processos para aumentar a interação entre eles” (MORAN; MASETTO; BEHRENS, 2001, p. 97). Sem dúvida, os benefícios das tecnologias da inteligência (LÈVY, 1993) vão além de uma navegação no mundo da Internet, contudo ela “[...] é a mídia mais importante e a que resume todo o potencial de mudança que está acontecendo” (MORAN; MASETTO; BEHRENS, 2001, p. 54).

Nesse contexto vale refletir sobre a tarefa da escola ainda citando MORAN, 2001.

Uma das tarefas principais da educação é ajudar a desenvolver tanto o conhecimento de resposta imediata como o de longo prazo; tanto o que está ligado a múltiplos estímulos sensoriais como o que caminha em ritmos mais lentos, que exige pesquisa mais detalhada, e tem de passar por **decantação, revisão, reformulação**. Muitos dados, muita informação não significam necessariamente mais e melhor conhecimento. O conhecimento torna-se produtivo se o integramos em uma visão ética pessoal, transformando-o em sabedoria, em saber pensar para agir melhor. Aprendemos mais quando estabelecemos pontes entre a reflexão e a ação, entre experiência e a conceituação, entre a teoria e a prática; quando ambas se alimentam mutuamente. (MORAN, 2001, p. 22, 23, grifo nosso).

Nesse contexto de mudanças na educação provenientes da tecnologia digital estão intrínsecas a prática inclusiva do acesso a saberes que possibilitarão a plena cidadania. Não é mais possível falar em educação de forma estática e ignorar os meios tecnológicos que estão em todos os ambientes da sociedade, pesquisas e instrumentos que reinventem o protagonismo dos jovens numa sociedade que cada dia se torna mais competitivo e técnica.

## 2- O PROGRAMA R

Com o avanço tecnológico o uso de *softwares* e pacotes para a análise e interpretação de dados tornaram-se rotineiros. Contudo alguns *softwares* apresentam custos elevados para os usuários, assim é grande a procura por *softwares* livres.

Dentre os *softwares* de domínio público, um que tem ganhado destaque e cada vez mais adeptos e colaboradores é o Ambiente R, ou simplesmente R, como é usualmente conhecido pelos usuários (QUICK apoud MELLO; PETERNELLI, 2013).

Segundo Mello e Peternelli (2013, p.16) o *software R* “não é simplismente um programa estatístico, uma vez que permite operações matemáticas, manipulação de vetores e matrizes, confecção de gráficos e manipulação de banco de dados, entre outros”.

Ao contrário dos *softwares* pagos, ao invés de contar com o suporte técnico oferecido pela empresa que mantém o *software*, o R conta com a colaboração de dezenas (talvez centenas) de milhares de usuários ao redor do mundo. Alguns deles são renomados professores e pesquisadores em diversas áreas do conhecimento. Esse suporte é conseguido, principalmente, através de listas de emails (*mailing lists*) contendo dúvidas e respostas sobre diversos assuntos relacionados ao R. (MELLO; PETERNELLI, 2013, p. 16)

### 2.1- Instalando o R

O R está disponível para as plataformas UNIX, *Windows* e MacOS. Para baixar o R para o *Windows* entre no site do R <https://www.r-project.org/>.

Segundo Mello e Peternelli para a instalação do R basta conectar-se ao *site* e, em CRAN (*Comprehensive R Archive Network*), escolher o local de disponibilização (espelho) do programa mais próximo de onde você se encontra. No Brasil, existem atualmente quatro opções: Universidade Federal do Paraná (PR), Fundação Osvaldo Cruz (RJ), Universidade de São Paulo (USP-SP) e USP Piracicaba (SP).

Após escolher o espelho, em *Download and Install R*, clique no *link* que corresponde ao sistema operacional do seu computador, depois é só executar o arquivo.

Durante o processo de instalação no *Windows*, atalhos para o R são criados e podem ser acessados para rodar o programa.

Ao rodar o R no *Windows*, você deverá ver a interface do usuário, como mostrado na Figura 1. Na janela R GUI existem uma barra de menus, uma barra de ferramentas ou comandos e o *console* do R.

Com o R iniciado, você verá na tela o símbolo > em vermelho, que é o *prompt* do R, numa janela conhecida como *R Console*. Este sinal que é o *prompt* indica que o sistema está pronto para receber os comandos. Algumas vezes, o sinal de *prompt* do R muda de > para + indicando que o comando digitado na linha anterior ainda não foi finalizado.

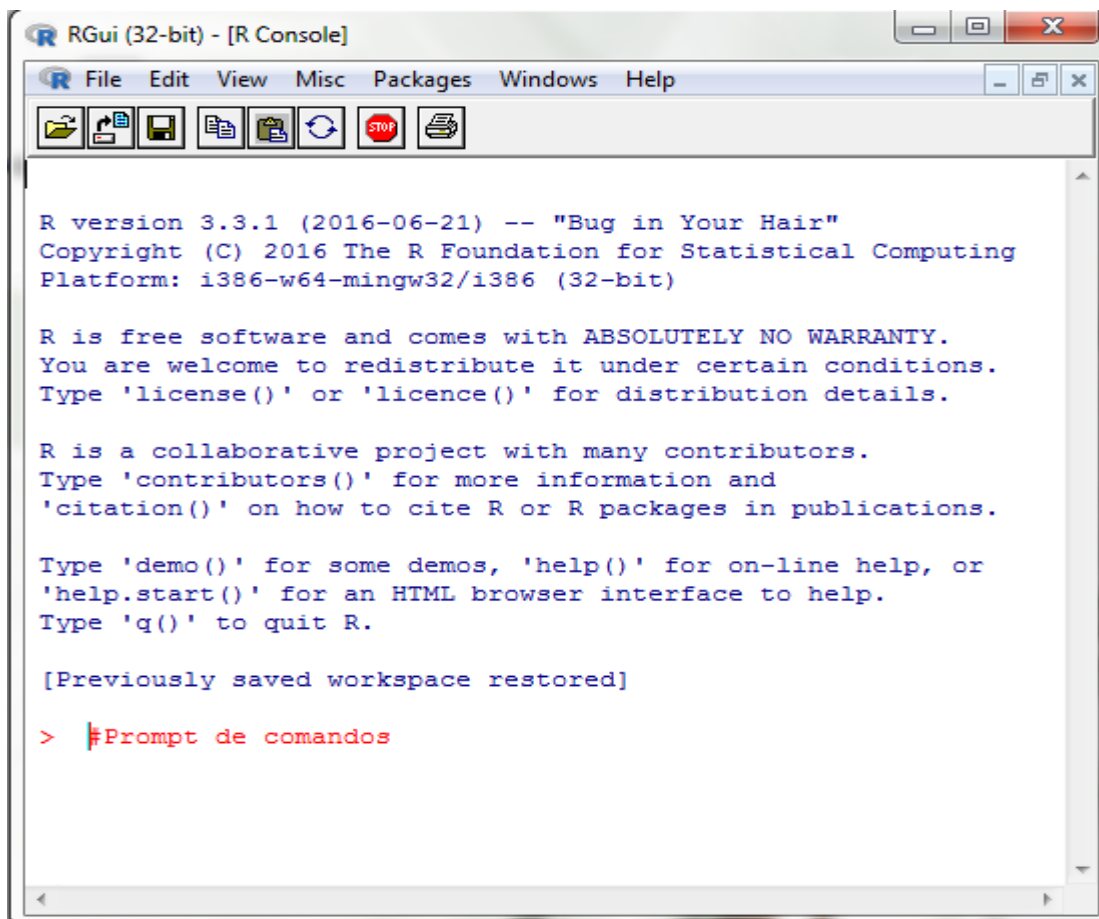


Figura 1- Janela da interface do usuário do R no Windows

## 2.2- Instalando o RStudio

Para Mello e Peternelli o RStudio trata-se de um *software* livre que oferece uma interface de usuário mais completa para o R. É o que os especialistas em computação chamam de IDE (*Integrated Development Environment* ou Ambiente Integrado de Desenvolvimento). Trata-se de uma integração entre o R e um avançado editor de textos voltados para a edição de comandos.

O RStudio é gratuito, tem versões para Linux, Mac OS X e *Windows* e pode ser obtido em <http://www.rstudio.com/>. Para baixar e instalar, acesse a página citada e escolha fazer o *download* da versão *RStudio Desktop*. O *site* irá recomendar a melhor versão para o seu



computador. Baixe e instale o RStudio. O RStudio deve ser instalado em computadores em que o R já esteja instalado.

No momento de instalação o Rstudio integra-se à última versão do R instalada em seu computador.

A janela do RStudio Figura 2 é dividida em quatro partes. A parte inferior esquerda é o *console* do R. Acessando o menu “File -> New -> R Script” o Rstudio abre, na parte superior esquerda, uma nova aba em branco onde os comandos serão digitados pelo usuário. O programa irá executar o comando quando este for enviado para o *console*, que pode ser feito com o atalho “CTRL+ENTER” ou clicando no ícone *Run* . Na parte superior direita, o RStudio mostra uma janela com os objetos criados na sessão atual do R e uma aba onde o histórico dos comandos executados no *console* podem ser visualizados. Na parte inferior direita há quatro abas importantes: *Files*, que mostra o conteúdo do diretório atual; *Plots*, que exhibe os gráficos criados pelo R; *Packages*, que apresenta os pacotes instalados e carregados; e *Help*, uma aba bastante útil pois exhibe uma ajuda do R.

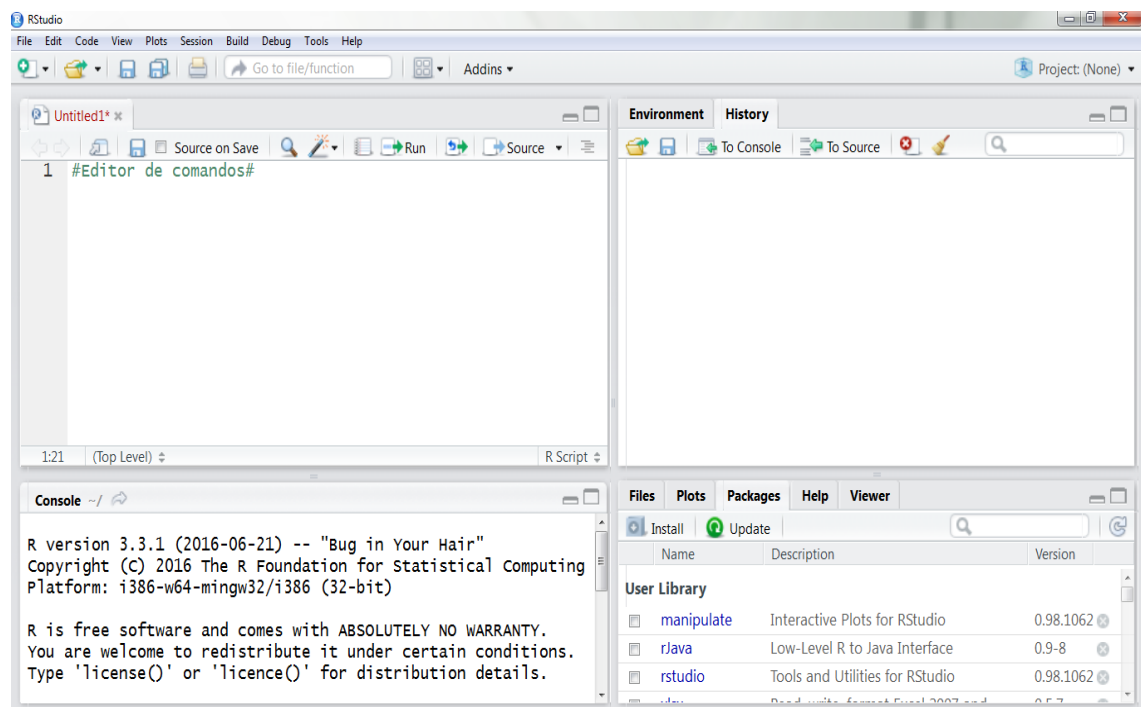


Figura 2 - Janela da interface do RStudio para Windows

### 2.3- Estrutura de Dados

Os dados no R podem estar organizados em diferentes estruturas, as estruturas mais usadas são:

- *vector*: vetor com um ou mais elementos com uma dimensão;
- *matrix*: matriz com duas dimensões;
- *factor*: vetor que representa dados categóricos;
- *data.frame*: parecido com a estrutura de matriz, mas permite colunas de diferentes tipos de um mesmo objeto.

Segundo Mello e Peternelli o vetor é uma das principais estruturas do R, pois podem armazenar elementos do tipo numérico, caractere, lógico ou complexo.

A maneira mais simples de criar um vetor é usando o comando `c( )`, que concatena elementos num mesmo objeto.

*Exemplo*

Criar o vetor  $X = \{1, 2, 3, 4\}$ .

Resolvendo com o Rstudio:

```
#Sintaxe:
X<-c(1,2,3,4)#Comando para criar o vetor x#

#Resultado:
> X #escrevo X e executo (CTRL + ENTER) ele mostra o vetor criado#
[1] 1 2 3 4
```

Para criar um novo vetor repetindo os valores do primeiro vetor podemos usar o comando `rep( )`.

*Exemplo*

Usar o vetor  $X$  para criar o vetor  $Y = \{1, 1, 2, 2, 2, 3, 4, 4\}$ .

Resolvendo com o RStudio:

```
#Sintaxe:
X<-c(1,2,3,4)
fi<-c(2,3,1,2)#cria o vetor da frequência de cada valor no vetor x#
Y<-c(rep(X,fi))#cria o vetor dos valores de X repetidos fi vezes#

#Resultado:
> X
[1] 1 2 3 4
> fi
[1] 2 3 1 2
> Y
[1] 1 1 2 2 2 3 4 4
```

Podemos criar vetores com caracteres que são os fatores, para isto pode-se usar o comando `factor(c( ))`.

*Exemplo*

Criar o fator  $E = \{\text{primavera}, \text{verão}, \text{outono}, \text{inverno}\}$ .

Resolvendo com o RStudio:

```
#Sintaxe:
E<-factor(c("primavera","verão","outono","inverno"))#cria um vetor co
```

```
m os nomes desejados( sempre separados por aspas duplas)#
#Resultado:
> E
[1] primavera verão outono inverno
Levels: inverno outono primavera verão
```

Podemos transformar um vetor ou um fator em uma matriz pois o R é capaz de operar com matrizes. O comando para criação de matrizes é `matrix( )`.

*Exemplo*

*Transforme o vetor Y e o fator E em matrizes.*

Resolvendo com o RStudio:

```
#Sintaxe:
Matriz1<-matrix(Y,nrow=4, ncol=2,byrow=T)#cria a matriz com os valores de Y, com 4 linhas e 2 colunas (4x2) onde os valores estão organizados por linhas#
Matriz2<-matrix(E,nrow=2, ncol=2,byrow=F) #cria a matriz com os dados de E, com 2 linhas e 2 colunas (2x2) onde os valores estão organizados por colunas#

#Resultado:
Matriz1
  [,1] [,2]
[1,]  1   1
[2,]  2   2
[3,]  2   3
[4,]  4   4

> Matriz2
  [,1] [,2]
[1,] "primavera" "outono"
[2,] "verão" "inverno"
```

Os `data.frame` são semelhantes as matrizes pois possuem duas dimensões entretanto podem armazenar em colunas diferentes tipos de dados diferentes podendo a primeira coluna ser numérica enquanto a segunda ser constituída de caracteres.

Mello e Peternelli comentam que os `data.frames` são uma das melhores formas de armazenar dados pois cada linha representa uma unidade e cada coluna representa uma variável observada em cada unidade. Entretanto cada coluna deve ter o mesmo tamanho. O comando para criar um `data.frame` é `data.frame( )`.

*Exemplo*

*Criar um data.frame com o fator E e o vetor fi.*

Resolvendo com o RStudio:

```
#Sintaxe:
data.frame<-data.frame(E,fi)#cria o data.frame com a colunas E e fi#

  E      fi
1 primavera 2
2 verão     3
3 outono    1
4 inverno   2
```

### 3- PRINCIPAIS CONCEITOS DE ESTATÍSTICA

As definições apresentadas neste capítulo são fundamentadas nas obras de “Estatística Básica” (FERREIRA, 2009), “Estatística Vol. I” (PINTO & SILVA, 2013), “Estatística Para a Educação Profissional e Tecnológica” (NOVAES & COUTINHO, 2013) e “Fundamentos da matemática elementar, 11: matemática comercial, matemática financeira e estatística descritiva” (IEZZI; HAZZAN & DEGENSZAJN, 2004).

#### 3.1- Divisões da Estatística

A Estatística se divide em Descritiva e Inferencial, passando pela Teoria de Probabilidade como elemento de ligação entre as duas.

##### 3.1.1- Estatística descritiva

É aquela que se preocupa com a coleta, organização, classificação, apresentação, interpretação e análise de dados referentes ao fenômeno estudados através de gráficos e tabelas além de calcular medidas que permita descrever o fenômeno.

##### 3.1.2- Estatística Inferencial

É a aquela que partindo de uma amostra, estabelece hipóteses, tira conclusões sobre a população de origem e que formula previsões.

#### 3.2- População e Amostra

O conjunto total de levantamentos, sobre o qual se deseja retirar conclusões, é denominado de **população** ou universo.

Se a população sob estudo for muito pequena pode ser prático obter todos os levantamentos. Quando se tem uma população muito grande é razoável obter um subconjunto desta população. Esse subconjunto é conhecido como **amostra**.

Se uma empresa está estudando a proporção de peças defeituosas em sua produção total, a população será constituída por todas as peças produzidas no período considerado e uma amostra pode ser constituída por qualquer subconjunto de peças.

### 3.3- Amostragem

É o processo de fixar critérios para a composição de uma amostra que tenha representatividade necessária no estudo em questão. A amostragem pode ser probabilística e não probabilística.

A amostragem é probabilística quando todo elemento da população tem a mesma probabilidade de sair na amostra, caso contrário, teremos uma amostragem não probabilística.

### 3.4- Variável

É uma grandeza observável e sujeita a variações. Trata-se de um conjunto de dados específicos que pode assumir valores e/ou aspectos distintos, segundo os casos particulares ou circunstâncias. As variáveis podem ser classificadas em quantitativas e qualitativas.

#### 3.4.1- Variável Quantitativa

É uma variável que pode ser mensurada numericamente de forma *discreta* ou *continua*.

Uma variável quantitativa é discreta quando entre dois valores consecutivos da variável não podemos inserir outro valor.

*Exemplos*

*Número de filhos, número de turistas em uma região, quantidades de gols em uma partida de futebol, etc.*

Uma variável quantitativa é contínua quando seus valores podem assumir qualquer valor dentro de um intervalo real.

*Exemplo*

*A altura dos estudantes, temperatura ambiente, peso das bagagens de um avião.*

### 3.4.2- Variável Qualitativa

É uma variável que não pode ser mensurada numericamente e seus valores são expressos por atributos *nominais* ou *ordinais*.

Uma variável qualitativa é nominal quando não se pode estabelecer uma relação de ordem ou hierarquia entre os possíveis valores a serem assumidos pela variável.

*Exemplos*

*Cor dos olhos, sexo, estado civil.*

Uma variável qualitativa é ordinal quando permite que se estabeleça uma hierarquia coerente ou uma relação de ordem entre os valores assumidos.

*Exemplos*

*Grau de satisfação (muito satisfeito, satisfeitos, nada satisfeitos), grau de escolaridade, classe econômica.*

### 3.5- Dados Brutos e Rol

Os dados são coletados numa forma sem ordenação e sem nenhum tipo de arranjo sistemático. Nesse caso, eles são denominados **dados brutos**.

$X = \{a_1, a_2, a_3, \dots, a_i, \dots, a_n\}$ , com  $i \in \{1, 2, 3, \dots, n\}$ , onde **n** é número total de observações.

*Exemplo*

*Dados brutos referentes ao número de ovos danificados em uma inspeção feita em 20 embalagens, de uma dúzia cada, em um carregamento para o mercado municipal de Lavras proveniente de uma cidade distante.*

$X = \{3, 2, 1, 2, 3, 5, 1, 1, 3, 4, 1, 1, 1, 1, 1, 1, 0, 2, 0, 0\}$ ,  $n = 20$

Resolvendo no RStudio:

```
#sintaxe:
Dados.bruto<-c(3, 2, 1, 2, 3, 5, 1, 1, 3, 4, 1, 1, 1, 1, 1, 1, 0, 2,
0, 0)

#Resultado:
>Dados.bruto
[1] 3 2 1 2 3 5 1 1 3 4 1 1 1 1 1 1 0 2 0 0
```

Para que se possa analisar este conjunto, fica mais simples se estes forem organizados de alguma forma, quando possível, em ordem crescente ou decrescente de grandeza numérica,

seja em categorias semelhantes ou ordem alfabética, no caso de variáveis qualitativas. Quando estes dados estão ordenados passam a ser considerados um **rol**. Quando queremos indicar um valor qualquer da variável que esteja no rol denotaremos por  $(x_i)$ .

$X = \{x_1, x_2, x_3, \dots, x_i, \dots, x_n\}$ , com  $i \in \{1, 2, 3, \dots, n\}$ , onde **n** é número total de observações.

*Exemplo*

*Rol referente ao número de ovos danificados em uma inspeção feita em 20 embalagens, de uma dúzia cada, em um carregamento para o mercado municipal de Lavras proveniente de uma cidade distante.*

$X = \{0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 5\}$ ,  $n = 20$

Resolvendo no RStudio:

```
#sintaxe:
Rol<-sort(Dados.bruto)

Resultado:
> Rol
[1] 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 3 3 3 4 5
```

### 3.6- Tabelas

Um dos objetivos da Estatística é sintetizar os valores que uma ou mais variáveis podem assumir, para que tenhamos uma visão global da variação dessa ou dessas variáveis. E isso ela consegue, inicialmente, apresentando esses valores em tabelas e gráficos.

Tabela é um quadro que resume um conjunto de observações. Uma tabela compõe-se de:

*Título:* conjunto de informações, as mais completas possíveis, respondendo às perguntas. (*O que? Onde? E quando?*)

*Cabeçalho:* parte superior da tabela que especifica o conteúdo das colunas;

*Coluna (ou Linha) indicadora:* parte da tabela que especifica o conteúdo das linhas (ou colunas);

*Linhas:* retas imaginárias que facilitam a leitura, no sentido horizontal, de dados que se inscrevem nos seus cruzamentos com as colunas;

*Célula:* espaço destinado a uma só informação;

*Corpo:* conjunto de linhas e colunas que contém informações sobre a variável em

estudo.

### Exemplo

**Tabela 1** - Lista dos 10 modelos de automóveis mais vendidos no Brasil no mês de agosto de 2016.

Ranking	Carro	Marca	Quantidade
1	Onix	Chevrolet	12.422
2	HB20	Hyundai	10.263
3	Ka	Ford	7.043
4	Palio	Fiat	6.750
5	Sandero	Renault	6.657
6	Corolla	Toyota	6.030
7	Prisma	Chevrolet	5.848
8	HR-V	Honda	4.906
9	Gol	Volkswagen	4.525
10	Renegade	Jeep	4.483

Fonte: Federação dos Distribuidores de Veículos.

Resolvendo no RStudio:

```
#sintaxe:
Ranking<-c(1:10)

Carro<-factor(c("Onix","HB20","Ka","Palio","Sandero","Corolla","Prisma",
,"HR-V","Gol","Renegade"))

Marca<-factor(c("Chevrolet","Hyundai","Ford","Fiat","Renault","Toyota",
,"Chevrolet","Honda","Volkswagen","Jeep"))

Quantidade<-c(12422,10263,7043,6750,6657,6030,5848,4906,4525,4483)

TabelaI<-data.frame(Ranking,Carro,Marca,Quantidade)
Retorno:
> TabelaI
  Ranking  Carro      Marca  Quantidade
1       1  Onix    Chevrolet    12422
2       2 HB20    Hyundai     10263
3       3   Ka     Ford        7043
4       4 Palio    Fiat        6750
5       5 Sandero Renault     6657
6       6 Corolla Toyota     6030
7       7 Prisma   Chevrolet    5848
8       8  HR-V    Honda      4906
9       9   Gol    Volkswagen  4525
10      10 Renegade Jeep       4483
```

### 3.7- Tabelas de frequências

O primeiro procedimento que possibilita uma leitura mais resumida dos dados após o rol é a construção de tabelas de frequências.

Para cada variável estudada, contamos o número de vezes que ocorre cada um de seus



valores (ou realizações). O número obtido é a *frequência absoluta* e pode ser indicada por  $f_i$ , onde  $i$  corresponde à classe (linha) da tabela.

Temos ainda a *frequência relativa* ( $f_r$ ), *frequência percentual* ( $f_p$ ) e a *frequência acumulada* ( $F_{ac\downarrow}$ ) onde:

- $f_r = \frac{f_i}{n}$ , sendo  $n$  o número total de observações;
- $f_p = f_r \times 100$ ;
- $F_{ac\downarrow}$ , é obtida acumulando-se (somando) as frequências absolutas das linhas.

#### Exemplo 1

Dados referentes ao número de ovos danificados em uma inspeção feita em 20 embalagens, de uma dúzia cada, em um carregamento para o mercado municipal de Lavras proveniente de uma cidade distante.

$$X = \{0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 5\}, n = 20$$

**Tabela 2** - Tabela de frequência de ovos danificados em um carregamento para o mercado municipal de Lavras.

Classe (k)	Nº de ovos danificados ( $x_i$ )	$f_i$	$f_r$	$f_p\%$	$F_{ac\downarrow}$
1	0	3	0,15	15	3
2	1	9	0,45	45	12
3	2	3	0,15	15	15
4	3	3	0,15	15	18
5	4	1	0,05	5	19
6	5	1	0,05	5	20
	Total	20	1,00	100	

Fonte: Adaptado de Daniel Furtado

Resolvendo no RStudio:

```
#sintaxe:
X<-c(0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 5)
table(X) #conta os valores de cada xi#
xi<-c(0,1,2,3,4,5)
fi<-c(3,9,3,3,1,1)
fr<-c(fi/20)
fp<-c(fr*100)
Fac<-c(3,3+9,3+9+3,3+9+3+3,3+9+3+3+1,3+9+3+3+1+1)
Tab.freq<-data.frame(xi,fi,fr,fp,Fac)
#Resultado:
X
 [1] 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 3 3 3 4 5
> table(X)
X
0 1 2 3 4 5
3 9 3 3 1 1
> xi
 [1] 0 1 2 3 4 5
> fi
```

```
[1] 3 9 3 3 1 1
> fr
[1] 0.15 0.45 0.15 0.15 0.05 0.05
> fp
[1] 15 45 15 15 5 5
> Fac
[1] 3 12 15 18 19 20
> Tab.freq
  xi fi  fr  fp  Fac
1  0  3  0.15 15   3
2  1  9  0.45 45  12
3  2  3  0.15 15  15
4  3  3  0.15 15  18
5  4  1  0.05  5  19
6  5  1  0.05  5  20
```

### Exemplo 2

Em uma turma de 16 alunos perguntou-se no mês de setembro de 2016, qual time vai ganhar o campeonato brasileiro série A de futebol em 2016?

Respostas: Palmeiras, Flamengo, Corinthians, Grêmio, Atlético Mineiro, Fluminense, Flamengo, Corinthians, Santos, Flamengo, Palmeiras, Flamengo, Atlético Mineiro, Palmeiras, Corinthians, Palmeiras.

**Tabela 3** - Tabela de frequência que mostra a opinião de 16 alunos em setembro de 2016, sobre qual time vai ganhar o campeonato brasileiro série A.

Times	$f_i$	$f_r$	$f_p$	$F_{ac\downarrow}$
Atlético Mineiro	2	0,1250	12,50	2
Corinthians	3	0,1875	18,75	5
Flamengo	4	0,2500	25,00	9
Fluminense	1	0,0625	6,25	10
Grêmio	1	0,0625	6,25	11
Palmeiras	4	0,2500	25,00	15
Santos	1	0,0625	6,25	16

Fonte: Dados fictícios

Resolvendo no RStudio:

```
#Sintaxe:
Respostas<-factor(c("Palmeiras","Flamengo","Corinthians","Grêmio",
"Atlético Mineiro","Fluminense","Flamengo","Corinthians","Santos",
"Flamengo","Palmeiras","Flamengo","Atlético Mineiro","Palmeiras",
"Corinthians","Palmeiras"))

table(Respostas)

xi<-factor(c("Atlético Mineiro","Corinthians","Flamengo","Fluminense",
"Grêmio","Palmeiras","Santos" ))

fi<-c(2,3,4,1,1,1,4,1)

fr<-c(fi/16)
```

```

fp<-c(fr*100)
Fac<-c(2,2+3,2+3+4,2+3+4+1,2+3+4+1+1,2+3+4+1+1+1,2+3+4+1+1+1+3,2+3+4+1
+1+1+3+1)

#Resultado:
> Respostas
 [1] Palmeiras      Flamengo      Corinthians   Grêmio
Atletico Mineiro Fluminense
 [7] Flamengo      Corinthians   Santos        Flamengo
Palmeiras        Flamengo
[13] Atlético Mineiro Palmeiras      Corinthians   Palmeiras
> xi
 [1] Atlético Mineiro Corinthians   Flamengo      Fluminense
Grêmio        Palmeiras
 [7] Santos
> fi
 [1] 2 3 4 1 1 1 4 1
> fr
 [1] 0.1250 0.1875 0.2500 0.0625 0.0625 0.0625 0.2500 0.0625
> fp
 [1] 12.50 18.75 25.00 6.25 6.25 6.25 25.00 6.25
> Fac
 [1] 2 5 9 10 11 12 15 16

```

Quando estamos trabalhando com dados quantitativos contínuos em algumas situações é conveniente trabalhar com tabelas de frequência com dados agrupados em intervalo de classes.

Para determinar o número  $k$  de classes, podemos utilizar a regra da raiz quadrada ou a regra de Sturges, uma prática comum é utilizar um mínimo de 5 e um máximo de 10 classes, observando se o valor escolhido não interfere na clareza da análise e se é possível evitar classes vazias.

- $k \cong \begin{cases} 5, & \text{para } n \leq 25 \\ \sqrt{n}, & \text{para } n > 25 \end{cases}$ , regra do quadrado
- $k \cong 1 + 3.32 \log n$ , regra de Sturges

Para montar a tabela de frequência precisamos determinar a amplitude das classes ( $h$ ) em função da amplitude total ( $A_t$ ) dos dados:

- $A_t = X_{\text{máximo}} - X_{\text{mínimo}}$
- $h = \frac{A_t}{k}$

### Exemplo 3

*Idade de 40 turistas que visitaram Ji-Paraná/RO no último feriado de 7 de setembro de 2016.*

53, 23, 30, 48, 42, 40, 14, 32, 52, 17, 51, 48, 45, 49, 38, 62, 51, 47, 50, 52, 55, 63, 66, 54, 62, 60, 27, 42, 51, 37, 43, 15, 42, 28, 48, 49, 40, 39, 48, 62

*Rol*

14, 15, 17, 23, 27, 28, 30, 32, 37, 38, 39, 40, 40, 42, 42, 42, 43, 45, 47, 48, 48, 48, 48, 49, 49, 50, 51, 51, 51, 52, 52, 53, 54, 55, 60, 62, 62, 62, 63, 66

$$A_t = 66 - 14 = 52$$

$$k \cong 1 + 3.32 \log 40 = 6,32, \text{ adotaremos } k = 6$$

$h = \frac{52}{6} = 8,67$ , sempre que possível podemos arredondar os dados para o inteiro mais próximo neste caso usaremos  $h = 9$ .

A primeira classe pode-se iniciar com o valor mínimo observado ou um valor menor. Para estes dados começaremos com o valor 13, logo teremos como limites superiores das classes os valores:

- Classe1:  $13 + 9 = 22$ ,
- Classe2:  $22 + 9 = 31$ ,
- Classe3:  $31 + 9 = 40$ ,
- Classe4:  $40 + 9 = 49$ ,
- Classe5:  $49 + 9 = 58$ ,
- Classe6:  $58 + 9 = 67$ .

*Pontos médios das classes ( $x_i$ ):*

- $x_1 = \frac{13+22}{2} = 17,5$
- $x_2 = \frac{22+31}{2} = 26,5$
- $x_3 = \frac{31+40}{2} = 35,5$
- $x_4 = \frac{40+49}{2} = 44,5$
- $x_5 = \frac{49+58}{2} = 53,5$
- $x_6 = \frac{58+67}{2} = 62,5$

Para representar os intervalos a simbologia mais utilizada é o intervalo fechado à esquerda e aberto à direita com a seguinte notação  $a \vdash b$ , fechado em  $a$  e aberto em  $b$ , caso necessário podemos fechar o intervalo da última classe à direita.

Tabela 4 - Idade de 40 turistas que visitaram Ji-Paraná em 7 de setembro de 2016.

Idades	$x_i$	$f_i$	$f_r$	$f_p$	$F_{ac}$
13 † 22	17,5	3	0,075	7,5	3
22 † 31	26,5	4	0,100	10,0	7
31 † 40	35,5	4	0,100	10,0	11
40 † 49	44,5	12	0,300	30,0	23
49 † 58	53,5	11	0,275	27,5	34
58 † 67	62,5	6	0,150	15,0	40
<b>Totais</b>		40	1,000	100,0	

Fonte: Dados fictícios

Resolvendo no RStudio:

```
#Sintaxe
Idade<-c(53, 23, 30, 48, 42, 40, 14, 32, 52, 17, 51, 48, 45, 49, 38,
62, 51, 47, 50, 52, 55, 63, 66, 54, 62, 60, 27, 42, 51, 37, 43, 15, 4
2, 28, 48, 49, 40, 39, 48, 62)
RoI<-sort(Idade)
n<-length(Idade)
AT<-(max(Idade)-min(Idade))
n.classe<-n.class.Sturges(Idade)
Tabela.IV<-table(cut(Idade,seq(13,67,l=7),include.lowest = T,right =
F)) #construção dos intervalos de classe, onde 13 e 67 são os limites
, l=7 é o n.classes, com intervalo fechado à esquerda e aberto à dir
eita#
Idades<-factor(c("[13,22)","[22,31)","[31,40)","[40,49)","[49,58)","[
58,67]"))
xi<-c(17.5,26.5,35.5,44.5,53.5,62.5)
fi<-c(3,4,4,12,11,6)
fr<-c(fi/n)
fp<-c(fr*100)
Fac<-c(3,3+4,3+4+4,3+4+4+12,3+4+4+12+11,3+4+4+12+11+6)

#Resultado
> Idade
[1] 53 23 30 48 42 40 14 32 52 17 51 48 45 49 38 62 51 47 50 52 55 6
3 66 54 62 60
[27] 27 42 51 37 43 15 42 28 48 49 40 39 48 62
> RoI
[1] 14 15 17 23 27 28 30 32 37 38 39 40 40 42 42 42 43 45 47 48 48 4
8 48 49 49 50
[27] 51 51 51 52 52 53 54 55 60 62 62 62 63 66
> n
[1] 40
> AT
[1] 52
> n.classe
[1] 7
> Tabela
[13,22) [22,31) [31,40) [40,49) [49,58) [58,67]
      3      4      4      12      11      6
> Idades
[1] [13,22) [22,31) [31,40) [40,49) [49,58) [58,67]
Levels: [13,22) [22,31) [31,40) [40,49) [49,58) [58,67]
> xi
[1] 17.5 26.5 35.5 44.5 53.5 62.5
> fi
[1] 3 4 4 12 11 6
> fr
[1] 0.075 0.100 0.100 0.300 0.275 0.150
```

```

> fp
[1] 7.5 10.0 10.0 30.0 27.5 15.0
> Fac
[1] 3 7 11 23 34 40
> Tabela.IV
  Idades  xi  fi  fr  fp  Fac
1 [13,22) 17.5 3 0.075 7.5 3
2 [22,31) 26.5 4 0.100 10.0 7
3 [31,40) 35.5 4 0.100 10.0 11
4 [40,49) 44.5 12 0.300 30.0 23
5 [49,58) 53.5 11 0.275 27.5 34
6 [58,67] 62.5 6 0.150 15.0 40

```

### 3.8- Representação Gráfica

A representação gráfica compreende o conjunto de ferramentas utilizadas em Estatística para condensar dados de maneira simples, clara e verdadeira.

Para a escolha da representação gráfica mais adequada, faz-se necessário considerar a natureza dos dados. Se a variável for qualitativa ou quantitativa com dados discretos, podemos utilizar um gráfico de barras, colunas ou setor.

#### 3.8.1- Gráfico de barras ou colunas

Os gráficos em barra ou colunas são utilizados principalmente para comparar as frequências dos valores observados de uma mesma variável.

*Exemplo*

*Em uma turma de 16 alunos perguntou-se no mês de setembro de 2016, qual time vai ganhar o campeonato brasileiro série A de futebol em 2016?*

*Respostas:*

*Palmeiras, Flamengo, Corinthians, Grêmio, Atlético Mineiro, Fluminense, Flamengo, Corinthians, Santos, Flamengo, Palmeiras, Flamengo, Atlético Mineiro, Palmeiras, Corinthians, Palmeiras.*

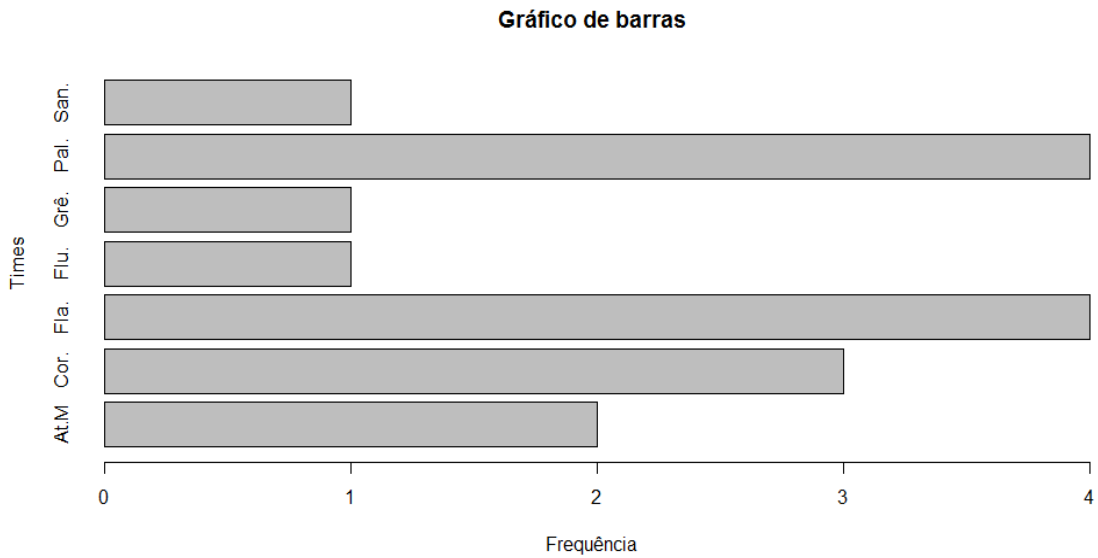
Resolvendo no RStudio:

```

#Sintaxe:
dados<-c("Pal.", "Fla.", "Cor.", "Grê.", "At.M", "Flu.", "Fla.", "Cor.", "San
.", "Fla.", "Pal.", "Fla.", "At.M", "Pal.", "Cor.", "Pal.")

barplot(table(dados), xlab="Times", ylab="Frequência", main="Gráfico de
barras", hor=T)#hor=T, barras horizontais#

```

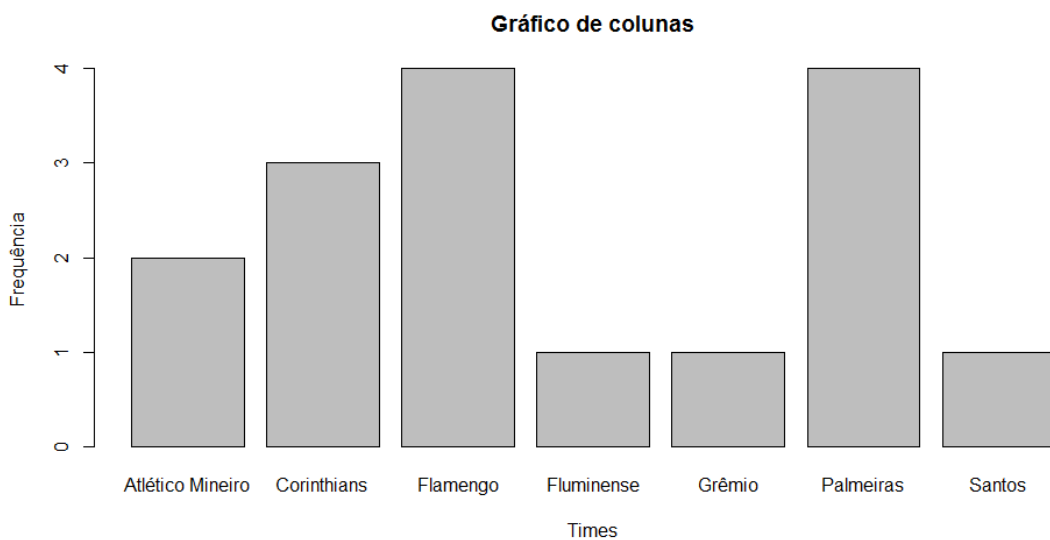


**Gráfico 01:** Opinião de 16 alunos quanto ao time que será campeão brasileiro da série A em 2016.

Resolvendo no RStudio:

```
#Sintaxe:
dados<-c("Palmeiras","Flamengo","Corinthians","Grêmio","Atlético Mineiro",
"Fluminense","Flamengo","Corinthians","Santos","Flamengo","Palmeiras",
"Flamengo","Atlético Mineiro","Palmeiras","Corinthians","Palmeiras")

barplot(table(dados),xlab="Times",ylab="Frequência",main="Gráfico de
colunas",hor=F)
```



**Gráfico 02:** Opinião de 16 alunos quanto ao time que será campeão brasileiro da série A em 2016.

### 3.8.2- Gráfico de setores (pizza)

Usamos esse tipo de gráfico para representar distribuições de uma mesma variável em relação ao total de informações ele é construído com base num círculo e é útil para representar frações em relação ao total.

#### Exemplo

*Em uma turma de 16 alunos perguntou-se no mês de setembro de 2016, qual time vai ganhar o campeonato brasileiro série A de futebol em 2016?*

*Respostas: Palmeiras, Flamengo, Corinthians, Grêmio, Atlético Mineiro, Fluminense, Flamengo, Corinthians, Santos, Flamengo, Palmeiras, Flamengo, Atlético Mineiro, Palmeiras, Corinthians, Palmeiras.*

Resolvendo com o RStudio:

```
#Sintaxe
dados<-c("Palmeiras","Flamengo","Corinthians","Grêmio","Atlético Mineiro",
"Fluminense","Flamengo","Corinthians","Santos","Flamengo","Palmeiras",
"Flamengo","Atlético Mineiro","Palmeiras","Corinthians","Palmeiras")

table(dados) )
#contar a frequência de cada time#

Quant.<-c(2,3,4,1,1,4,1)
#vetor de frequência dos times em ordem alfabética#

names(Quant.)<-c("Atlético Mineiro", "Corinthians", "Flamengo","Fluminense",
"Grêmio","Palmeiras","Santos")
#dando nomes a cada quantidade#

porc<-round(Quant.*100/sum(Quant.),2)
# calcula a porcentagem de cada time usando duas casas decimais#

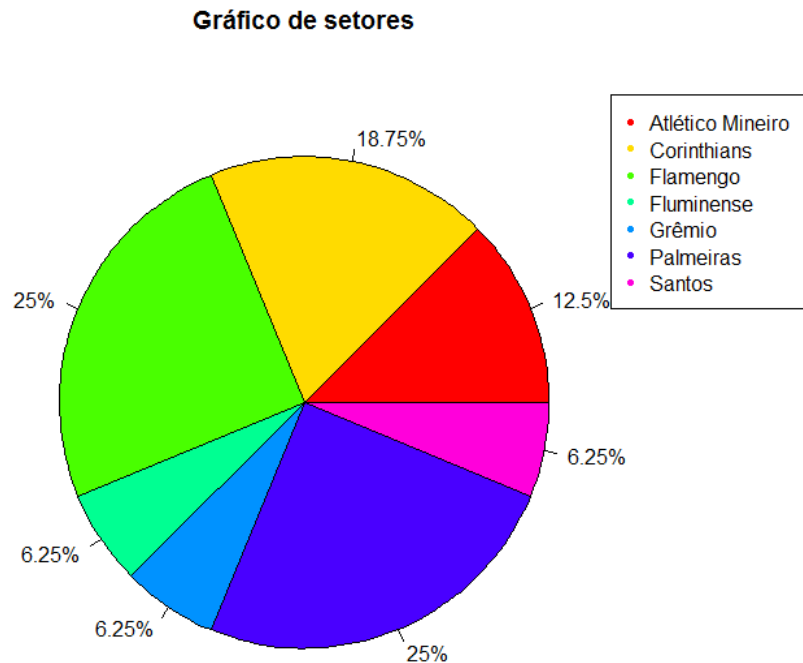
rotulos<-paste("",porc,"%",sep="")#coloca as porcentagem como rótulo
no gráfico de setores#

x11()
# abre uma janela em branco para exibição do gráfico#

pie(Quant.,main="Gráfico de setores",col = rainbow(7),labels=rotulos)
# cria o gráfico de setores com nome, sete cores, exibindo a porcentagem
de cada time#

legend(1,1,names(Quant.),col = rainbow(7),pch=rep(20,6))
#cria a legenda do gráfico#
```





**Gráfico 03:** Opinião de 16 alunos quanto ao time que será campeão brasileiro da série A em 2016.

### 3.8.3- Histograma

Para dados de variáveis quantitativas contínuas a representação gráfica mais indicada é o histograma que constitui de colunas verticais justapostas, onde estas colunas representam os intervalos de cada classe.

*Exemplo*

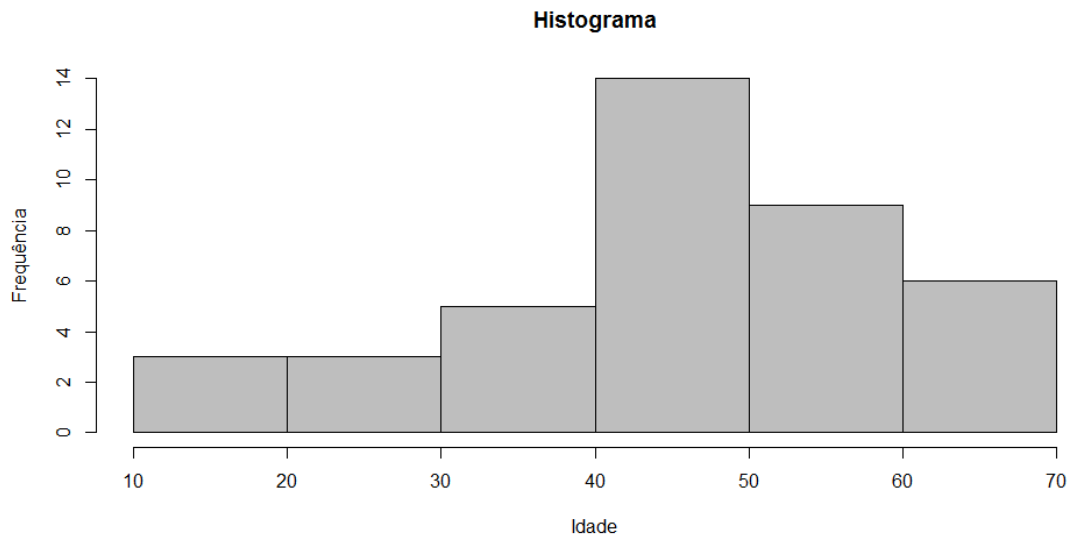
*Idade de 40 turistas que visitaram Ji-Paraná/RO no último feriado de 7 de setembro de 2016.*

53, 23, 30, 48, 42, 40, 14, 32, 52, 17, 51, 48, 45, 49, 38, 62, 51, 47, 50, 52, 55, 63, 66, 54, 62, 60, 27, 42, 51, 37, 43, 15, 42, 28, 48, 49, 40, 39, 48, 62

Resolvendo com o RStudio:

```
#Sintaxe:
idade<-c(53, 23, 30, 48, 42, 40, 14, 32, 52, 17, 51, 48, 45, 49, 38,
62, 51, 47, 50, 52, 55, 63, 66, 54, 62, 60, 27, 42, 51, 37, 43, 15, 4
2, 28, 48, 49, 40, 39, 48, 62)

hist(idade, nc=7, right=F, main="Histograma",xlab="Idade",ylab="Frequ
ência",col="Grey")
#cria o histograma com seis classes (7-1), aberto à direita, com títu
lo Histograma e nomes nos eixos#
```



**Gráfico 04:** Histograma das idades de 40 turistas que visitaram Ji-Paraná em 7 de setembro de 2016.

### 3.8.4- Gráfico de linhas

O gráfico de linhas é muito usado quando se quer representar o comportamento de uma variável cujos valores diminuem ou aumentam ao longo do tempo de maneira contínua.

#### *Exemplo*

*Uma pesquisa eleitoral mostra a evolução de três candidatos mais bem posicionados na campanha eleitoral para prefeito em 2016 na cidade Ji-Paraná/RO.*

**Tabela 5:** Porcentagem de votos para os três candidatos mais bem posicionados na campanha para prefeito de Ji-Paraná/RO em 2016.

Mês/2016	Candiato 1	Candidato 2	Candidato 3
3	9	17	26
5	13	16	24
7	17	15	20
9	25	12	21

Fonte: Dados fictícios.

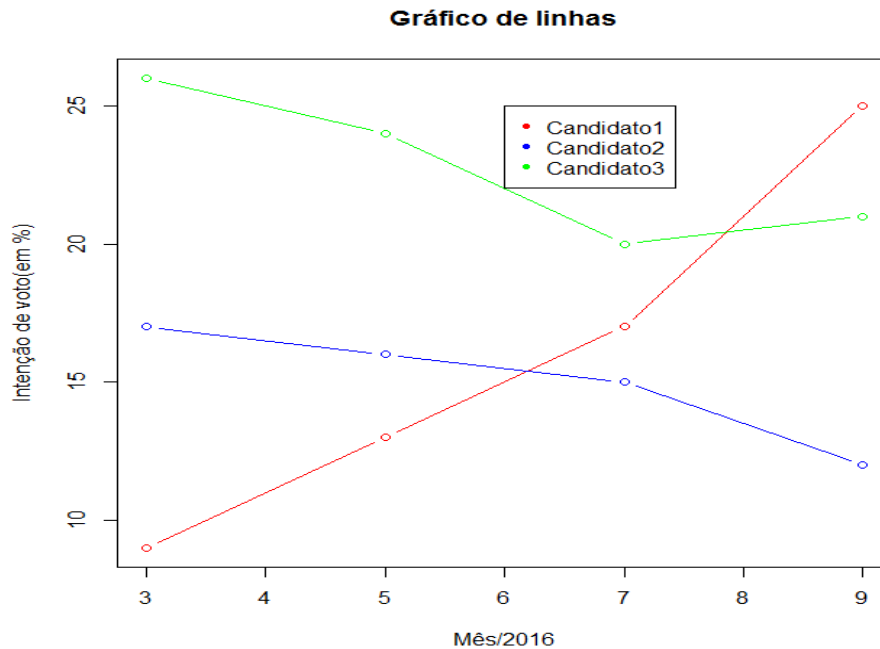
Resolvendo com o RStudio:

```
#Sintaxe:
mes<-c(3,5,7,9)
Candidato1<-c(9,13,17,25)
Candidato2<-c(17,16,15,12)
Candidato3<-c(26,24,20,21)
x11()#cria uma janela vazia onde o gráfico será construído#
plot(mes,Candidato1, main = "Gráfico de linhas",xlim = c(3,9),ylim=c(
9,26), ylab = "Intenção de voto(em %)",xlab = "Mês/2016",col="red",ty
pe="b")#cria um gráfico entre mês e candidato 1 contendo: nome, eixos
delimitados, eixos com nome, linha de cor vermelha e ponto em formato
de círculo.#
lines(mes,Candidato2,col="blue",type="b")#cria linha entre mes e cand
```

```

idato2 na cor azul#
lines(mes,Candidato3,col="green",type="b")"#cria linha entre mes e c
andidato3 na cor verde#
legend(6,25,c("Candidato1","Candidato2","Candidato3"),col =c("red","b
lue","green"), pch=rep(20,3))#cria a legenda do gráfico na coordenada
(6,25) com as cores de cada candidato#

```



**Gráfico 5:** Gráfico de linha da porcentagem de votos para os três candidatos mais bem posicionados na campanha para prefeito de Ji-Paraná/RO em 2016.

### 3.9- Medidas de Tendência Central

Um conjunto de dados pode ser representado por tabelas e gráficos, no entanto pode-se utilizar outras formas de representação, algumas delas são as medidas de tendência central: média, mediana e moda. Podemos dizer que a análise das medidas de tendência central permite a representação dos dados por um único valor. Consideremos, por exemplo, o caso das notas dos alunos que os professores passam para a secretaria escolar. O professor não passa todas as notas das avaliações que o aluno faz, mas uma média das notas das avaliações em certo período de tempo.

#### 3.9.1-Média aritmética

A média aritmética é a medida de tendência central mais utilizada para representar um conjunto de dados. Podemos interpretar a média como sendo o ponto de equilíbrio dos desvios dos valores da distribuição. A média é altamente influenciada pelos extremos. Por

isso seu valor analisado sem o valor de outras medidas que mostram a variabilidade nos dados pode ficar comprometido.

Usualmente, em Estatística usamos letras gregas minúsculas para representar medidas populacionais e letras do alfabeto latino para medidas amostrais. No caso da média usaremos a seguinte notação:

- $\bar{x}$ : média amostral
- $\mu$ : média populacional

A média do conjunto de dados será dada por  $\mu = \frac{\sum_{i=1}^N x_i}{N}$ , onde  $x_i$  representa cada um dos valores observados e  $N$  é o número da população.

Para dados amostrais, teremos  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ , onde  $n$  é o número de elementos da amostra.

A média aritmética admite algumas propriedades que podem ser úteis na resolução de problemas.

Seja  $x$  uma variável quantitativa que assume valores  $x_1, x_2, \dots, x_n$ .

P1) Se a cada  $x_i$  ( $i = 1, 2, 3, \dots, n$ ) adicionarmos uma constante real  $c$ , a média aritmética fica adicionada de  $c$  unidades.

P2) Se a cada  $x_i$  ( $i = 1, 2, 3, \dots, n$ ) multiplicarmos uma constante real  $c$ , a média aritmética fica multiplicada por  $c$ .

### 3.9.1.1- Media aritmética para dados não agrupados

Para calcular a média aritmética de dois ou mais números, adicionamos esses números e dividimos o resultado obtido pela quantidade de números adicionados.

Sejam  $x_1, x_2, \dots, x_n$ , portanto,  $n$  valores da variável  $X$ . A média aritmética de  $X$  será representada por:

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

*Exemplo*

*Dados brutos referentes ao número de ovos danificados em uma inspeção feita em 20 embalagens, de uma dúzia cada, em um carregamento para o mercado municipal de Lavras proveniente de uma cidade distante. Qual é a média de ovos danificados por embalagem?*

$$X = \{3, 2, 1, 2, 3, 5, 1, 1, 3, 4, 1, 1, 1, 1, 1, 1, 0, 2, 0, 0\}, n = 20$$

$$\bar{x} = \frac{3 + 2 + 1 + 2 + 3 + 5 + 1 + 1 + 3 + 4 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 0 + 2 + 0 + 0}{20} = \frac{33}{20}$$

$$\bar{x} = 1,65$$

Resolvendo com o RStudio:

```
#Sintaxe:
x<-c(3, 2, 1, 2, 3, 5, 1, 1, 3, 4, 1, 1, 1, 1, 1, 1, 1, 0, 2, 0, 0)
media<-mean(x)#comando que calcula a média#

#Retorno:
> x
[1] 3 2 1 2 3 5 1 1 3 4 1 1 1 1 1 1 0 2 0 0

> media
[1] 1.65
```

### 3.9.1.2- Média aritmética ponderada (dados agrupados)

Se os dados dos quais queremos determinar a média estão organizados em tabela de frequência usamos a seguinte notação:

- $\mu = \frac{\sum_{i=1}^N x_i \cdot f_i}{\sum_{i=1}^N f_i}$  ou  $\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i}$

*Exemplo 1*

*Determinar a média para os dados que estão apresentados na tabela 2.*

**Tabela 6:** Tabela de frequência de ovos danificados em um carregamento para o mercado municipal de Lavras, com coluna auxiliar para o cálculo da média.

Classe (k)	Nº de ovos danificados ( $x_i$ )	$f_i$	$f_r$	$f_p\%$	$x_i * f_i$
1	0	3	0,15	15	0
2	1	9	0,45	45	9
3	2	3	0,15	15	6
4	3	3	0,15	15	9
5	4	1	0,05	5	4
6	5	1	0,05	5	5
Total		20	1,00	100	33

Fonte: Adaptado de Daniel Furtado

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{33}{20} = 1,65$$

Resolvendo no RStudio:

```
#Sintaxe:
xi<-c(0,1,2,3,4,5)
fi<-c(3,9,3,3,1,1)
media<-mean(rep(xi,fi))#o comando rep(xi,fi) cria um vetor com valores de xi repetidos fi vezes#

#Retorno:
> xi
[1] 0 1 2 3 4 5
> fi
[1] 3 9 3 3 1 1
> media
[1] 1.65
```

### Exemplo2

Determinar a média para os dados apresentados na tabela 4.

**Tabela 7:** Idade de 40 turistas que visitaram Ji-Paraná em 7 de setembro de 2016 com coluna auxiliar para cálculo de média.

Idades	$x_i$	$f_i$	$f_r$	$f_p$	$x_i * f_i$
13 † 22	17,5	3	0,075	7,5	52,5
22 † 31	26,5	4	0,100	10,0	106
31 † 40	35,5	4	0,100	10,0	142
40 † 49	44,5	12	0,300	30,0	534
49 † 58	53,5	11	0,275	27,5	588,5
58 † 67	62,5	6	0,150	15,0	375
<b>Totais</b>		40	1,000	100,0	1798

Fonte: Dados fictícios

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i} = \frac{1798}{40} = 44,95$$

Resolvendo no RStudio:

```
#Sintaxe:
> xi<-c(17.5,26.5,35.5,44.5,53.5,62.5)
> fi<-c(3,4,4,12,11,6)
> mean(rep(p,F))

#Retorno:
> xi
[1] 17.5 26.5 35.5 44.5 53.5 62.5
> fi
[1] 3 4 4 12 11 6
> mean(rep(p,F))
[1] 44.95
```

### 3.9.2- Mediana

A média pode ser muito afetada quando encontramos valores discrepantes em um conjunto de dados, podendo se tornar uma medida de centralidade pouco representativa do resumo dos dados. Para contornar questões desta natureza, definiremos, a seguir, uma medida de centralidade mais resistente aos valores discrepantes denominada mediana (Me).

### 3.9.2.1- Mediana para dados não agrupados

Sejam  $x_1 \leq x_2 \leq \dots \leq x_n$  os  $n$  valores ordenados de uma variável  $X$ . A mediana é o elemento que ocupa a posição central dos dados, ou seja:

$$\bullet \quad Me = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par} \end{cases}$$

#### Exemplo1

*Dados brutos referentes ao número de ovos danificados em uma inspeção feita em 20 embalagens, de uma dúzia cada, em um carregamento para o mercado municipal de Lavras proveniente de uma cidade distante. Qual é a mediana de ovos danificados por embalagem?*

$X = \{3, 2, 1, 2, 3, 5, 1, 1, 3, 4, 1, 1, 1, 1, 1, 1, 0, 2, 0, 0\}$ ,  $n = 20$

*Rol*

$X = \{0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 5\}$ ,  $n = 20$

*Temos  $n$  par, logo:*

$$Me = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} = \frac{x_{10} + x_{11}}{2} = \frac{1 + 1}{2} = 1$$

Resolvendo com o RStudio:

```
#Sintaxe:
X<-c(3, 2, 1, 2, 3, 5, 1, 1, 3, 4, 1, 1, 1, 1, 1, 1, 0, 2, 0, 0)
rol<-sort(X)
mediana<-median(X)#comando que determina a mediana#
> X
[1] 3 2 1 2 3 5 1 1 3 4 1 1 1 1 1 1 0 2 0 0
> rol
[1] 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2 3 3 3 4 5
> mediana
[1] 1
```

#### Exemplo2

Considere a seguinte distribuição  $Y$  das notas de 15 alunos em Estatística. Qual é a nota que divide o conjunto dos 15 valores apresentados em duas partes com exatamente o mesmo número de elementos em cada uma delas?

$$Y = \{2, 2, 7, 7, 7, 8, 3, 4, 4, 6, 9, 9, 4, 4, 7\}$$

Rol

$$Y = \{2, 2, 3, 4, 4, 4, 4, 6, 7, 7, 7, 7, 8, 9, 9\}$$

Temos  $n$  ímpar, logo:

$$Me = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_8 = 6$$

Resolvendo no RStudio:

```
#Sintaxe:
Y<-c(2,2,7,7,7,8,3,4,4,6,9,9,4,4,7)
rol<-sort(Y)
mediana<-median(Y)

#Resultado:
> Y
[1] 2 2 7 7 7 8 3 4 4 6 9 9 4 4 7
> rol
[1] 2 2 3 4 4 4 4 6 7 7 7 7 8 9 9
> mediana
[1] 6
```

### 3.9.2.2- Mediana para dados agrupados sem intervalo de classes

Para dados agrupados sem intervalo de classes temos que:

- $Me = \begin{cases} \frac{x_i + x_{(i+1)}}{2}, & \text{se } F_{(acl)Me} = \frac{\sum f_i}{2} \\ x_i, & \text{se } F_{(acl)Me} > \frac{\sum f_i}{2} \end{cases}$ ,
- com  $i = \{1, 2, 3, \dots, k\}$
- onde  $F_{(acl)Me}$  é a primeira frequência acumulada que é maior ou igual  $\frac{\sum f_i}{2}$ .

*Exemplo 1*

Determine a mediana para os dados apresentados na tabela 8.



**Tabela 8:** Tabela com o número de gols por partida registrado na 24ª e 25ª rodadas do campeonato série A 2016.

Número de gols	$f_i$	$F_{ac\downarrow}$
0	1	1
1	6	7
2	4	11
3	5	16
4	2	18
6	1	19
8	1	20
<b>Total</b>	20	

Fonte: Globo Esporte

*Resolução:*

$\frac{\sum f_i}{2} = \frac{20}{2} = 10$ , não temos nenhuma  $F_{ac\downarrow}$  igual a 10 então,  $F_{(ac\downarrow)Me} = 11$ , determinará a classe mediana, logo:

$$Me = x_3 = 2$$

Resolvendo no RStudio:

```
#Sintaxe:
xi<-c(0,1,2,3,4,6,8)
fi<-c(1,6,4,5,2,1,1)
mediana<-median(rep(xi,fi))#comando que determina a mediana para valores de xi repetidos fi vezes#

#Resultado:
> xi
[1] 0 1 2 3 4 6 8
> fi
[1] 1 6 4 5 2 1 1
> mediana
[1] 2
```

*Exemplo2*

Determine a media para os dados apresentados na tabela 9.

**Tabela 9:** Tabela com o resultado de 40 observações relativas ao índice pluviométrico em Ji-Paraná/RO nos dias chuvosos de janeiro a março de 2016.

Quantidade de chuva (mm)	$f_i$	$F_{ac\downarrow}$
135	7	7
144	8	15
148	5	20
157	10	30
162	6	36
175	4	40
<b>Total</b>	40	

Fonte: Dados Fictícios

*Resolução:*

$\frac{\sum f_i}{2} = \frac{40}{2} = 20$ , como temos a  $F_{ac\downarrow} = 20$  na terceira classe a mediana será:

$$Me = \frac{x_i + x_{(i+1)}}{2} = \frac{x_3 + x_{(3+1)}}{2} = \frac{148 + 157}{2} = 152,5$$

Resolvendo com o RStudio:

```
#Sintaxe:
chuva<-c(135,144,148,157,162,175)
fi<-c(7,8,5,10,6,4)
mediana<-median(rep(chuva,fi))

#Retorno:
> chuva
[1] 135 144 148 157 162 175
> fi
[1] 7 8 5 10 6 4
> mediana
[1] 152.5
```

### 3.9.2.3- Mediana para dados agrupados com intervalo de classes

Para calcular a mediana com dados agrupados em intervalos de classes podemos proceder da seguinte maneira:

- 1º) Determinamos as frequências acumuladas;
- 2º) Calculamos  $\frac{\sum f_i}{2}$ ;
- 3º) Marcamos a classe correspondente à frequência acumulada imediatamente superior ou igual à  $\frac{\sum f_i}{2}$ . Tal classe será a classe mediana;
- 4º) Calculamos a Mediana pela seguinte fórmula:

- $Me = li + \frac{(\frac{\sum f_i}{2} - F_{ac\downarrow(ant.)}) \cdot h}{f_{i(Me)}}$ , onde:
- $li$  é o limite inferior da classe mediana;
- $F_{ac\downarrow(ant.)}$  é a frequência acumulada anterior a classe mediana;
- $f_{i(Me)}$  é a frequência absoluta da classe mediana;
- $h$  é a amplitude da classe mediana.

*Exemplo*

*Calcule a mediana para os dados da tabela 10.*

**Tabela 10:** Tabela de resultados de exame de colesterol de 200 funcionários que foram submetidos a exames clínicos.

Colesterol (em mg/dl de sangue)	$x_i$	$f_i$	$F_{ac\downarrow}$
140 † 180	160	21	21
180 † 220	200	45	66
220 † 260	240	73	139
260 † 300	280	34	183
300 † 340	320	27	200
<b>Total</b>		200	

Fonte: Adaptado de Gelson Iezzi

*Resolução:*

1º) Temos a  $F_{ac\downarrow}$  na tabela;

$$2^\circ) \frac{\sum f_i}{2} = \frac{200}{2} = 100$$

3º) A classe mediana será a classe onde  $F_{ac\downarrow} = 139 > 100$ ;

4º)  $li = 220$ ,  $F_{ac\downarrow}(ant.) = 66$ ,  $f_{i(Me)} = 73$ ,  $h = 40$ ,

*Calculando:*

$$Me = 220 + \frac{(100-66) \cdot 40}{73} = 238,63$$

Resolvendo com o RStudio:

```
#Sintaxe:
Me<-(220 +(100-66)*40/73)#como os dados estão tabelados podemos usar
a função calculadora do R#

#Resultado:
> Me
[1] 238.6301
```

### 3.9.3- Moda

A moda de um conjunto de valores corresponde ao valor que ocorre mais vezes, portanto uma série de dados pode ser amodal, unimodal, bimodal, trimodal ou polimodal.

#### 3.9.3.1- Moda para dados não agrupados

Seja  $X$  uma variável quantitativa que assume os valores  $x_1, x_2, \dots, x_k$ , com frequências absolutas iguais a  $f_1, f_2, \dots, f_k$ , respectivamente. Se o máximo entre  $f_1, f_2, \dots, f_k$  é igual a  $f_j, j \in \{1, 2, \dots, k\}$ , dizemos que a moda (Mo) é igual ao valor  $x_j$ .

*Exemplo*

Os seguintes dados representam o rol dos diferentes preços (em R\$) de um determinado produto pesquisado em 20 lojas.

$Rol = \{50, 50, 51, 51, 51, 51, 51, 51, 52, 52, 52, 52, 52, 52, 52, 53, 53, 53, 53, 53, 53, 54\}$

$Mo_1 = 52$  e  $Mo_2 = 53$ , temos uma série bimodal

Resolvendo com o RStudio:

```
#Sintaxe:
Rol<-c(50,50,51,51,51,51,51,51,52,52,52,52,52,52,52,53,53,53,53,53,53,54)
tabela<-table(Rol)
moda <- names(tabela)[tabela == max(tabela)] #determina o valor da moda da tabela#

#Resultado:
> Rol
 [1] 50 50 51 51 51 51 51 51 52 52 52 52 52 52 53 53 53 53 53 53
[20] 54
> tabela
Rol
50 51 52 53 54
 2  5  6  6  1
> moda
 [1] "52" "53"
```

**3.9.3.2- Moda para dados agrupados sem intervalo de classe**

Seja X uma variável quantitativa que assume os valores  $x_1, x_2, \dots, x_k$ , com frequências absolutas iguais a  $f_1, f_2, \dots, f_k$ , respectivamente. Se o máximo entre  $f_1, f_2, \dots, f_k$  é igual a  $f_j, j \in \{1, 2, \dots, k\}$ , dizemos que a moda (Mo) é igual ao valor  $x_j$ .

*Exemplo*

Uma rede de lojas Z de eletrodomésticos colocou um produto em oferta durante o natal de 2015 e prometeu um prêmio para os vendedores com o maior número de unidades vendidas. O número de peças vendidas por esses vendedores está na tabela 11. Qual o número modal de peças vendidas?

**Tabela 11:** Tabela com o resultado do número de peças vendidas pela rede de lojas Z, durante a promoção de natal de 2015.

Quantidade peças vendidas	$f_i$	$f_p$
10	1	2,3
18	3	6,8
22	2	4,5
23	8	18,2
28	15	34,1
35	13	29,6

<b>48</b>	<b>2</b>	<b>4,5</b>
<b>Total</b>	<b>44</b>	<b>100</b>

Fonte: Dados Fictícios

*Resolução:*

Como o maior  $f_i$  é o da quinta classe o valor modal de peças vendidas será:

$$Mo = 28$$

Resolvendo com o RStudio:

```
#Sintaxe:
Pecas<-c(10,18,22,23,28,35,48)
fi<-c(1,3,2,8,15,13,2)
tabela<-table(rep(Pecas,fi))#tabela os valores de Pecas repetidos fi
vezes#
moda<-names(tabela)[tabela == max(tabela)]#determina o valor da moda
da tabela#

#Resultado:
> Pecas
[1] 10 18 22 23 28 35 48
> fi
[1] 1 3 2 8 15 13 2
> tabela
10 18 22 23 28 35 48
1 3 2 8 15 13 2
> moda
[1] "28"
```

### 3.9.3.3- Moda para dados agrupados com intervalo de classe

A classe que apresenta a maior frequência é denominada classe modal. Pela definição, podemos afirmar que a moda, neste caso, é o valor dominante que está compreendido entre os limites da classe modal.

A forma mais simples de encontrar a moda é calcular o ponto médio da classe modal:

- $Mo = \frac{LiMo + LsMo}{2}$ , o valor calculado por esse processo é chamado de moda bruta.

Uma outra maneira de calcular a moda para dados agrupados em intervalo de classe é utilizando a fórmula de Czuber.

- $Mo = li + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$ , onde:
- $li$  é o limite inferior da classe modal;

- $\Delta_1$  é a diferença entre a frequência da classe modal e a da classe imediatamente anterior;
- $\Delta_2$  é a diferença entre a frequência da classe modal e a da classe imediatamente posterior.

*Exemplo*

*Determine a moda bruta e a moda pelo método de Czuber dos dados apresentados na tabela 12.*

**Tabela 12:** Tabela de valores de uma amostra do tempo de vida útil da bateria de uma marca de celular.

Nº de meses (vida útil)	$x_i$	$f_i$
0 – 06	03	3
06 – 12	09	42
12 – 18	15	86
18 – 24	21	127
24 – 30	27	64
30 – 36	33	8
<b>Total</b>		<b>330</b>

Fonte: Adaptado de Novaes e Coutinho

*Resolução:*

*A classe modal é a quarta classe, logo temos que a moda bruta é 21.*

*Para a moda usando o método de Czuber, temos:*

$$li = 18,$$

$$\Delta_1 = 127 - 86 = 41$$

$$\Delta_2 = 127 - 64 = 63$$

$$h = 6$$

$$Mo = li + \frac{\Delta_1 \cdot h}{\Delta_1 + \Delta_2} \rightarrow Mo = 18 + \frac{41 \cdot 6}{41 + 63} = 20,37$$

Resolvendo com o RStudio:

```
#Sintaxe:
> Mo=18+41*6/(41+63)

#Resultado:
> Mo
[1] 20.36538
```

### 3.10- Medidas de variabilidade

A análise isolada das medidas de tendência central pode-se levar a conclusões incorretas pois estas medidas não levam em consideração a variabilidade de dados, uma vez que conjunto de dados diferentes podem apresentar mesma medida de tendência central. Para analisar o grau de variabilidade de um conjunto de dados há necessidade de medidas específicas de variabilidade. Desvio médio, variância e desvio padrão são as medidas de variabilidade mais utilizadas em Estatística.

### 3.10.1- Desvio médio

É a média do valor absoluto dos desvios em relação a média aritmética da distribuição. O desvio médio pode ser obtido da seguinte maneira:

a) Dados não agrupados:

- $dm = \sum_{i=1}^n \frac{|x_i - \bar{x}|}{n}$ , para amostra;
- $dm = \sum_{i=1}^N \frac{|x_i - \mu|}{N}$ , para população.

*Exemplo*

*Considere a seguinte distribuição amostral Y das notas de 15 alunos em Estatística.*

*Qual é o desvio médio destas notas?*

$$Y = \{2, 2, 7, 7, 7, 8, 3, 4, 4, 6, 9, 9, 4, 4, 7\}$$

*Resolução:*

*Rol*

$$Y = \{2, 2, 3, 4, 4, 4, 4, 6, 7, 7, 7, 7, 8, 9, 9\}$$

$$\bar{x} = \frac{2 + 2 + 3 + 4 + 4 + 4 + 4 + 6 + 7 + 7 + 7 + 7 + 8 + 9 + 9}{15} \cong 5,5$$

$$dm = \frac{|2 - 5,5|}{15} + \frac{|2 - 5,5|}{15} + \dots + \frac{|9 - 5,5|}{15} = \frac{31,5}{15} = 2,1$$

Resolvendo com o RStudio:

```
#Sintaxe:
Y<-c(2,2,7,7,7,8,3,4,4,6,9,9,4,4,7)
dm<-sum(c(abs(Y - mean(Y))))/length(Y)#retorna a soma dos desvios div
idido pelo número de valores do vetor Y#

#Resultado:
> Y
```

```
[1] 2 2 7 7 7 8 3 4 4 6 9 9 4 4 7
> dm
[1] 2.097778
```

b) Dados agrupados:

- $dm = \sum_{i=1}^n \frac{|x_i - \bar{x}| \cdot f_i}{\sum_{i=1}^n f_i}$ , para amostra;
- $dm = \sum_{i=1}^N \frac{|x_i - \mu| \cdot f_i}{\sum_{i=1}^N f_i}$ , para população.

*Exemplo*

Determine o desvio médio do tempo de vida útil para uma amostra de dados da tabela

13.

**Tabela 13:** Tabela de valores de uma amostra do tempo de vida útil da bateria de uma marca de celular com colunas auxiliares para o cálculo do desvio médio.

Nº de meses (vida útil)	$x_i$	$f_i$	$x_i \cdot f_i$	$ x_i - \bar{x} $	$ x_i - \bar{x}  \cdot f_i$
0 ─ 06	03	3	9	16,2	48,6
06 ─ 12	09	42	378	10,2	428,4
12 ─ 18	15	86	1290	4,2	361,2
18 ─ 24	21	127	2667	1,8	228,6
24 ─ 30	27	64	1728	7,8	499,2
30 ─ 36	33	8	264	13,8	110,4
<b>Total</b>		<b>330</b>	<b>6336</b>		<b>1676,4</b>

Fonte: Adaptado de Novaes e Coutinho

*Resolução:*

Os valores necessários estão na tabela:

$$\bar{x} = \frac{6336}{330} = 19,2$$

$$dm = \sum_{i=1}^n \frac{|x_i - \bar{x}| \cdot f_i}{n} = \frac{1676,4}{330} = 5,08$$

Resolvendo com o RStudio:

```
#Sintaxe:
xi<-c(3,9,15,21,27,33)
fi<-c(3,42,86,127,64,8)
media<-mean(rep(xi,fi))
dm<-sum(c(abs(xi-media))*fi)/sum(fi)#retorna a soma dos desvios em va
lores absolutos dividido pela soma dos valores de fi#

#Resultado:
> xi
[1] 3 9 15 21 27 33
> fi
[1] 3 42 86 127 64 8
> media
[1] 19.2
```



```
> dm
[1] 5.08
```

### 3.10.2- Variância e desvio padrão

A variância populacional é definida como a soma dos quadrados dos desvios dividida pelo tamanho da população, enquanto que a variância amostral é definida como a soma dos quadrados dos desvios dividida pelo tamanho da amostra, menos 1.

A variância assim como a média aritmética admite algumas propriedades importantes.

Seja  $x$  uma variável quantitativa que assume valores  $x_1, x_2, \dots, x_n$ . Considere  $\bar{x}$  a média aritmética e  $\sigma^2$  a variância correspondente.

P1) Se a cada  $x_i (i = 1, 2, 3, \dots, n)$  for adicionado uma constante real  $c$ , a variância não se altera.

P2) Se cada  $x_i (i = 1, 2, 3, \dots, n)$  for multiplicado por uma constante real  $c$ , a variância fica multiplicada por  $c^2$ .

Sejam  $x_1, x_2, \dots, x_n$  os valores assumidos por uma variável  $x$ . Chamamos desvio padrão de  $x$  a raiz quadrada da variância de  $x$ .

a) Dados não agrupados:

- $\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$ , variância populacional;
- $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ , variância amostral;
- $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$ , desvio padrão populacional;
- $S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ , desvio padrão amostral.

*Exemplo*

*Os dados seguintes referem-se aos gastos mensais com ônibus e metrô (expressos em reais) que um estudante universitário tem durante um semestre. Determine a variância e o desvio padrão.*

$$X = \{42, 50, 54, 48, 56, 59\}, N = 6$$

*Resolução:*

$$\bar{x} = \frac{42 + 50 + 54 + 48 + 56 + 59}{6} = 51,5$$

$$\sigma^2 = \frac{\sum_{i=1}^6 (x_i - \bar{x})^2}{6} =$$

$$\frac{(42 - 51,5)^2 + (50 - 51,5)^2 + (54 - 51,5)^2 + (48 - 51,5)^2 + (56 - 51,5)^2 + (59 - 51,5)^2}{6}$$

$$\sigma^2 = 31,25 \text{ (reais)}^2, \text{ variância}$$

$$\sigma = \sqrt{31,25} = 5,59 \text{ reais, desvio padrão}$$

Resolvendo com o RStudio:

```
#Sintaxe:
X<-c(42,50, 54, 48, 56, 59)
variância<-var(X) #retorna a variância do vetor X#
dp<-sqrt(variância)

#Resultado:
> X
[1] 42 50 54 48 56 59
> variância
[1] 31.25
> dp
[1] 5.590169
```

b) Dados agrupados

- $\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot f_i}{N}$ , variância populacional;
- $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{n-1}$ , variância amostral;
- $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot f_i}{N}}$ , desvio padrão populacional;
- $S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{n-1}}$ , desvio padrão amostral.

*Exemplo*

Uma pesquisa realizada em alguns hotéis do litoral norte de São Paulo teve o objetivo de identificar os preços praticados nas diárias da região na baixa temporada de 2016. Determine a variância e o desvio padrão para os dados coletados.

**Tabela 14:** Tabela de valores de diárias praticada no litoral norte de São Paulo na período de baixa temporada de 2016

Valor da diária(R\$)	$x_i$	$f_i$	$x_i \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
40 ┆ 60	50	6	300	1936	11616
60 ┆ 80	70	18	1260	576	10368
80 ┆ 100	90	12	1080	16	192
100 ┆ 120	110	12	1320	256	3072
120 ┆ 140	130	6	780	1296	7776
140 ┆ 160	150	6	900	3136	18816
<b>Total</b>		60	5640		51840

Fonte: Adaptado de Novaes e Coutinho

*Resolução:*

$$\bar{x} = \frac{5640}{60} = 94$$

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{n - 1} = \frac{51840}{60 - 1} = 878,64(\text{reais})^2$$

$$S = \sqrt{878,64} = 29,64 \text{ reais}$$

Resolvendo com o RStudio:

```
#Sintaxe:
diarias<-c(50,70,90,110,130,150)
fi<-c(6,18,12,12,6,6)
variancia<-var(rep(diarias,fi))#calcula a variância para uma amostra#
desvio.padrao<-sqrt(variancia)

#Resultados:
> diarias
[1] 50 70 90 110 130 150
> fi
[1] 6 18 12 12 6 6
> variancia
[1] 878.6441
> desvio.padrao
[1] 29.64193
```

## 4- METODOLOGIA E PROCEDIMENTOS

A metodologia representa um método, um caminho ou um meio adequado para se alcançar determinado objetivo.

### 4.1- Engenharia Didática

A metodologia utilizada para realização deste trabalho foi a Engenharia Didática que surgiu na França no final da década de 1960.

A Engenharia Didática surgiu no transcorrer das discussões desenvolvidas no IREM (Instituto de Investigação do Ensino de Matemática), na França, ao final da década de 1960. Em seus primórdios, o IREM desenvolvia uma complementação na formação de professores de matemática e na produção de meios materiais de apoio para a sala de aula, destacando-se o desenvolvimento de jogos, brinquedos, problemas, exercícios e experimentos. (POMMER, 2013, p.10).

Para Brousseau (1996, apud POMMER, 2013, p.13), o modelo de pesquisa da Engenharia Didática requer do pesquisador/professor a participação e análise das situações didáticas. Um elemento essencial da situação didática é sua intencionalidade de ser construída para a aprendizagem do aluno.

Segundo Artigue apud Pommer, a Engenharia Didática é um processo empírico que objetiva conceber, realizar, observar e analisar as situações didáticas que podem assumir dupla função, a qual pode ser compreendida como uma produção para o ensino tanto como uma metodologia de pesquisa qualitativa.

Deste modo podemos dizer que a Engenharia Didática se caracteriza por propor:

[...] uma seqüência de aula(s) concebida(s), organizada(s) e articulada(s) no tempo, de forma constante, por um professor-engenheiro para realizar um projeto de aprendizagem para certa população de alunos. No decurso das trocas entre professor e alunos, o projeto evolui sob as reações dos alunos e em função das escolhas e decisões do professor (MACHADO, 2002, p. 198, apud DOUADY, 1993, p. 2, apud POMMER, 2013, p.20).

A Engenharia Didática, como metodologia descrita por Artigue (1996) apud Pommer (2013), compreende quatro fases: a 1ª fase, das *análises preliminares*, a 2ª fase, da *concepção e da análise a priori*, a 3ª fase, da *experimentação* e a 4ª e última fase, da *análise a posteriori e validação*.

- *Análises preliminares:* Nesta fase, são coletados dados e informações a fim de que se possa refletir sobre eles e com isso estruturar uma maneira positiva de interferência no ensino.
- *Concepção e da análise a priori:* Nesta fase, Machado (2002) apud Pommer (2013) ressalta que a pesquisa delimita as variáveis de comando, as variáveis são aquelas para as quais as escolhas de valores provocam modificações nas estratégias de resolução de problemas, de modo a fazer evoluir o desempenho dos alunos.
- *Experimentação:* De acordo com Machado (2002) apud Pommer (2013), consiste basicamente no desenvolvimento da aplicação da Engenharia Didática, concebida a um grupo de alunos, objetivando verificar as ponderações levantadas na análise *a priori*.
- *Análise a posteriore e validação:* De acordo com Artigue (1996) apud Pommer (2013), se apóia sobre o conjunto de dados obtidos ao longo da experimentação pelas observações do pesquisador. Segundo a autora, esta fase se caracteriza pelo tratamento dos dados colhidos e a confrontação com a análise *a priori*, permitindo a interpretação dos resultados. Assim, é possível analisar se ocorrem e quais são as contribuições para a superação do problema, o que permitirá a validação interna do objetivo da pesquisa.

#### **4.2- Descrição das fases do projeto**

Esta pesquisa foi desenvolvida no Instituto Federal de Rondônia – IFRO, no *campus* de Ji-Paraná situado na rua rio Amazonas nº 151, bairro Jardim dos Migrantes na cidade de Ji-Paraná/RO, no segundo semestre de 2016. A escolha deste local se deve ao fato do pesquisador ser servidor desta instituição.

Segue a descrição das fases desta pesquisa.

1º) *Análises preliminares:* nesta fase foi feito uma revisão bibliográfica, leitura de Projetos Pedagógicos de Curso, leitura de Planos de Ensinos, diálogo com professores de matemática da instituição.

2º) *Concepção e da análise a priori:* Após a leitura dos documentos e diálogo com os professores foi verificado que os alunos estudam estatística, no entanto não aprendem a dominar nenhum *software* que os auxilie no campo da Estatística.

A turma do 1ºA química do curso Técnico Integrado em Química tem uma disciplina de Estatística com carga horária de 80 horas em sua grade. O professor da disciplina sugeriu que fosse realizada a pesquisa com 20 alunos desta turma uma vez que os mesmos já haviam estudado os conteúdos de estatística básica no primeiro semestre de 2016. Os alunos foram convidados a participarem das atividades e espontaneamente colocaram seus nomes em uma lista de interessados, a escolha dos 20 foi feita por indicação do professor da disciplina de Estatística. Os pais ou responsáveis receberam os devidos esclarecimentos ao lerem e assinaram o “*Termo de Consentimento Livre e Esclarecido- TCLE*”.

Os alunos foram devidamente esclarecidos dos procedimentos da pesquisa ao lerem e assinarem o “*Termo de Assentimento Livre e Esclarecido- TALE*”.

Os alunos foram divididos em 5 grupos: A, B, C, D e E, cada aluno do grupo ficou identificado com um número de 1 a 4. Toda e qualquer referência ao aluno participante será pelo seu grupo e número. A divisão dos alunos em grupos foi feita para facilitar o desenvolvimento das atividades de cada oficina, pois os participantes desenvolveram as atividades em grupo.

A análise a priori foi realizada com a aplicação de uma avaliação contendo 10 questões de múltipla escolha de estatística descritiva: conceitos básicos, medidas de tendência central, medidas de dispersão, interpretação de tabelas e gráficos.

Na correção desta avaliação foi considerada uma escala de 0 a 100, escala usada pela instituição de ensino em que os participantes da pesquisa estudam. Para serem aprovados sem recuperação e exame segundo o “*Regulamento de Ordem Acadêmica – ROA*” o aluno deve alcançar média igual ou superior a 60 pontos. As notas dos alunos na análise a priori estão apresentadas na tabela 15.

**Tabela 15:** Notas dos alunos na análise a priori sobre conhecimentos de estatística básica realizada no IFRO/Ji-Paraná-RO no segundo semestre de 2016.

Estudantes	Notas
A1	40
A2	60
A3	40
A4	50
B1	40
B2	60
B3	50
B4	60
C1	70
C2	60

C3	50
C4	80
D1	30
D2	40
D3	40
D4	50
E1	40
E2	60
E3	50
E4	40
Média	50,5

Observa-se que a média dos alunos está abaixo de 60 pontos e que apenas 7 alunos alcançaram média igual ou superior a 60 pontos, resultando que apenas 35% dos participantes alcançariam a média estabelecida pela instituição para a aprovação.

3º) *Experimentação*: Nesta fase da pesquisa os alunos participaram de cinco oficinas divididas em 10 encontros semanais de duas aulas de 50 minutos cada ministrada em laboratório de informática.

- *Oficina I*: Instalação do *software R* e do *software RStudio*, apresentação da interface de cada *software*, criação de elementos no ambiente R: vetores, fatores, matriz e *data.frame*.
- *Oficina II*: Principais conceitos de Estatística, séries estatísticas e tabelas de frequências.
- *Oficina III*: Mediadas de tendência central: média, mediana e moda.
- *Oficina IV*: Medidas de variabilidade: desvio médio, variância e desvio padrão.
- *Oficina IV*: Medidas de variabilidade: desvio médio, variância e desvio padrão.

4º) *Análise a posteriori e validação*: para análise a posteriori foi aplicada uma avaliação contendo 10 questões de múltipla escolha de estatística descritiva: conceitos básicos, medidas de tendência central, medidas de dispersão, interpretação de tabelas e gráficos.

Na correção desta avaliação foi considerada a mesma escala da avaliação na análise a priori. As notas dos alunos na análise a posteriori estão apresentadas na tabela 16.

**Tabela 16:** Notas dos alunos na análise a posteriori sobre conhecimentos de estatística básica realizada no IFRO/Ji-Paraná-RO no segundo semestre de 2016.

Estudante	Notas
A1	60

A2	70
A3	50
A4	70
B1	40
B2	70
B3	70
B4	40
C1	90
C2	90
C3	70
C4	70
D1	60
D2	40
D4	50
E2	40
Média	61,25

Na segunda avaliação para análise a posteriori tivemos 16 participantes realizando a avaliação. Os participantes D3 e E1 saíram da escola e os participantes E3 e E4 não compareceram.

Observa-se que a média geral foi de 61,25 e que 10 alunos atingiram nota igual ou superior a 60 pontos, ou seja, 62,5 % dos alunos conseguiriam aprovação na instituição em que estudam. A validação dos dados segue no próximo capítulo.



## 5- ANÁLISE E DISCUSSÃO DOS RESULTADOS

No desenvolvimento das atividades os alunos foram questionados sobre os pontos positivos, pontos negativos e principais dificuldades que encontraram. Segue abaixo comentários dos grupos participantes para cada oficina.

- *Oficina I:* Instalação do *software R* e do *software RStudio*, apresentação da interface de cada *software*, criação de elementos no ambiente R: vetores, fatores, matriz e `data.frame`.

### **Pontos Positivos:**

Grupo A: *“Aprendizado de utilização do software R, que será utilizado tanto na vida escolar quanto profissional”*.

Grupo B: *“Facilitou e deixou mais fácil de realizar”*.

Grupo C: *“Acesso gratuito e instalação rápida”*.

Grupo D: *“A explicação, na hora das dúvidas ele nos atendeu”*.

Grupo E: *“Aprendemos a instalar o R e o RStudio, criar vetores, entender como funciona”*.

### **Pontos Negativos:**

Grupo A: Não relatou.

Grupo B: *“É complicado criar”*.

Grupo C: *“Outro Idioma”*.

Grupo D: *“Nenhuma”*.

Grupo E: *“Nenhum ponto negativo”*.

### **Dificuldades:**

Grupo A: *“Identificar os erros nos comandos”*.

Grupo B: *“Um pouco de dificuldade para acompanhar”*.

Grupo C: *“Nenhuma”*.

Grupo D: *“Em alguns momentos o professor não pode nos atender pois estava ocupado com outro grupo. É complicado olhar os slides do lugar que ficamos”*.

Grupo E: *“Por estarmos iniciando, tivemos uma certa dificuldade ao usar o RStudio”*.

- *Oficina II:* Principais conceitos de Estatística, séries estatísticas e tabelas de frequências.

### **Pontos Positivos:**

Grupo A: *“Melhor aprendizagem de um conteúdo já visto mas esquecido. Aprendizagem da estatística com o software, que facilita os cálculos de um futuro projeto”*.

Grupo B: *“Deixou mais fácil fazer”*.

Grupo C: *“Desperta mais interesse nos alunos”*.

Grupo D: *“Aprender a fazer tabela no software R. Fácil entendimento, ou seja, professor ensinou o conteúdo de forma clara”*.

Grupo E: *“Revisão do conteúdo já aplicado ajudando na compreensão do restante da oficina”*.

**Pontos Negativos:**

Grupo A: *“Nenhum”*.

Grupo B: *“É complicado”*.

Grupo C: *“Palavras em inglês”*.

Grupo D: Não relatou.

Grupo E: *“Não tem”*.

**Dificuldades:**

Grupo A: *“Nenhuma”*.

Grupo B: *“Dificuldade de acompanhar”*.

Grupo C: *“Nenhuma”*.

Grupo D: *“Apesar de explicar de forma clara é ruim para esclarecer os alunos de forma individual às vezes demorando para responder as questões individuais”*.

Grupo E: *“Nenhuma”*.

- *Oficina III: Medidas de tendência central: média, mediana e moda.*

**Pontos Positivos:**

Grupo A: *“O conteúdo foi absorvido facilmente pelas alunas, tanto na parte teórica quanto na parte prática do RStudio”*.

Grupo B: *“Fácil aprendizado, utilização da ferramenta R”*.

Grupo C: *“Aprendizado rápido, fácil, dinâmica, simplificado.”*

Grupo D: *“Apesar de já conhecermos o conteúdo, houveram algumas dúvidas e o professor soube explicar com clareza, tornando o conteúdo mais fácil”*.

Grupo E: *“A explicação muito nos ajuda na compreensão. Dúvidas forma sanadas”*.

**Pontos Negativos:**

Grupo A: *“Nenhum”*.

Grupo B: *“Não temos pontos negativos”*.

Grupo C: *“Nenhum”*.

Grupo D: *“Não”*.

Grupo E: *“Não tem”*.

**Dificuldades:**

Grupo A: *“Nenhuma dificuldade”*.

Grupo B: *“Algumas dificuldades para entender a ferramenta R”*.

Grupo C: *“Nenhuma”*.

Grupo D: *“Como só vimos o conteúdo na 1º etapa, não nos lembrávamos do mesmo o professor soube relembrar”*.

Grupo E: *“Não consigo calcular corretamente por causa da quantidade de valores”*.

- *Oficina IV: Medidas de variabilidade: desvio médio, variância e desvio padrão.*

**Pontos Positivos:**

Grupo A: *“Relembrou as fórmulas, ajudou no entendimento da teoria e ajudou no ENEM. Além de uma maior praticidade nos cálculos”*.

Grupo B: *“Fácil compreensão, aumento de conhecimento, melhora na digitação”*.

Grupo C: *“A facilidade e a velocidade que se executa”*

Grupo D: *“Aprendemos a utilizar o RStudio para encontrar variância, desvio médio e desvio padrão”*.

Grupo E: *“Conseguimos completar todas as atividades. Compreendemos melhor os gráficos”*.

**Pontos Negativos:**

Grupo A: *“Ter que ter conhecimentos do computador, como manusear ”*.

Grupo B: *“Não há”*.

Grupo C: *“Nenhum”*.

Grupo D: *“Nenhum”*.

Grupo E: *“Estrapolou o horário da aula”*.

**Dificuldades:**

Grupo A: *“Entender o significado de alguns códigos, por mais que haja explicação”*.

Grupo B: *“Não há”*.

Grupo C: “Decorar as fórmulas”.

Grupo D: “Nenhuma”.

Grupo E: “Nem uma”.

- *Oficina V*: Representação gráfica: gráfico de barras, gráfico de colunas, gráfico de setores, histogramas e gráficos de linhas.

**Pontos Positivos:**

Grupo A: “Maior entendimento dos gráficos, o que significa, como é feito, etc.”.

Grupo B: “Ajudará no futuro, facilitará a criação de novos relatórios.”.

Grupo C: Não Relatou.

Grupo D: Não Relatou.

Grupo E: Não Relatou.

**Pontos Negativos:**

Grupo A: “Ter que ter um conhecimento de computador”.

Grupo B: “Não há”.

Grupo C: Não Relatou.

Grupo D: Não Relatou.

Grupo E: Não Relatou.

**Dificuldades:**

Grupo A: “Entender alguns códigos mesmo com explicação”.

Grupo B: “Lembrar os comandos, porém os slides auxiliaram”.

Grupo C: Não Relatou.

Grupo D: Não Relatou.

Grupo E: Não Relatou.

Entre os pontos positivos os participantes destacaram a gratuidade do *software*, a facilidade de instalação, despertou o interesse do aluno, a facilidade e a velocidade com que se executam os cálculos e que ajudará no futuro para desenvolvimento de futuros projetos entre outros.

Quando analisamos os pontos negativos o fato do programa usar a linguagem em inglês foi o ponto negativo mais citado entre os participantes.

As principais dificuldades apresentadas pelos participantes foram a linguagem em inglês, entender os comandos e identificar os erros dos comandos.

Fazendo uma análise entre pontos positivos e pontos negativos verifica-se que os pontos positivos vão ao encontro dos objetivos propostos enquanto que os pontos negativos podem ser sanados com uma versão do *software* em português ou uma melhor qualidade no ensino da língua inglesa.

Para validação dos resultados comparamos as notas a priori com as notas a posteriori vamos considerar apenas os 16 participantes que concluíram todas as etapas da pesquisa. Segue a estatística descritiva do teste a priori e do teste a posteriori.

Resolvendo no R:

```
#Sintaxe:
x<-c(40,60,40,50,40,60,50,60,70,60,50,80,30,40,50,60)#conjunto de notas a priori#

var(X)
[1] 166.6667

summary(X)#apresenta um resumo estatístico dos dados#
Min.    1st Qu.  Median    Mean   3rd Qu.    Max.
30.0    40.0    50.0    52.5   60.0    80.0

Y<-c(60,70,50,70,40,70,70,40,90,90,70,70,60,40,50,40) #conjunto de notas a posteriori#

var(Y)
[1] 278.3333

summary(Y)
Min.    1st Qu.  Median    Mean   3rd Qu.    Max.
40.00   47.50   65.00   61.25   70.00   90.00
```

Pelo resumo observamos que todas as estatísticas: nota mínima, 1º quartil, mediana, média, 3º quartil, nota máxima e variância apresentada pelas notas da avaliação a posteriori são superiores as da avaliação a priori. Para validação desta observação aplicaremos um teste de hipótese para comparação de médias para dados pareados (duas observações de mesmo indivíduo).

Segundo Bussab e Morettin um Teste de Hipótese pode ser construído usando os cinco passos abaixo:

1º) Passo: Fixe qual a hipótese nula  $H_0$  a ser testada e qual a hipótese alternativa  $H_1$ .

Hipóteses:  $\begin{cases} H_0: A \text{ média a posteriori não difere da média a priori.} \\ H_1: A \text{ média a posteriori é maior do que média a priori.} \end{cases}$

2º) Passo: Use a teoria estatística e as informações disponíveis para decidir qual estatística será usada para testar  $H_0$ .

$$\text{Hipóteses: } \begin{cases} H_0: \mu_y = \mu_x \\ H_1: \mu_y > \mu_x \end{cases}$$

3º) Passo: Fixe a probabilidade  $\alpha$  de cometer o erro de tipo I, pois qualquer que seja a decisão tomada estamos sujeitos a cometer erros.

*Erro de tipo I:* rejeitar  $H_0$  quando essa é verdadeira. Chamamos de  $\alpha$  a probabilidade de cometer este erro.

*Erro do tipo II:* não rejeitar  $H_0$  quando  $H_0$  é falsa. A probabilidade de cometer esse erro é denotado por  $\beta$ .

Adotaremos  $\alpha = 0,05$ .

4º) Use as observações da amostra para calcular o valor da estatística do teste.

5º) Para a tomada de decisão de rejeitar ou não  $H_0$  pode-se usar o valor-p ou (p-value), para valor-p  $\leq 0,05$  indica que devemos rejeitar a hipótese nula.

Dados duas amostras  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$ , de observações pareadas, isto é, podemos considerar que temos uma amostra de pares  $(X_1, Y_1), \dots, (X_n, Y_n)$ . Se definirmos a variável aleatória  $D = Y - X$ , teremos a amostra  $D_1, \dots, D_n$  resultante das diferenças entre os valores de cada par.

Esta amostra deve possuir distribuição normal para aplicarmos o teste de comparação de médias para dados pareados.

Para Mello e Peternelli (2013, p.162), podemos usar o teste de Shapiro-Wilk para verificar a normalidade dos dados que pode ser realizado no R com o uso do comando `shapiro.test ( )`.

As hipóteses do teste de Shapiro-Wilk:

$$\text{Hipóteses: } \begin{cases} H_0: \text{Os dados seguem uma distribuição normal} \\ H_1: \text{Os dados não seguem uma distribuição normal} \end{cases}$$

Resolvendo com o R:

```
#Sintaxe:
D<-c(Y-X)#cria o vetor da diferença entre as duas médias a serem test
adas#
D
[1] 20 10 10 20 0 10 20 -20 20 30 20 -10 30 0 0 -20
shapiro.test(D)
      shapiro-wilk normality test
data:  D
w = 0.91207, p-value = 0.1257
```

Como o valor- $p > 0,05$  o conjunto de dados D segue uma distribuição normal de probabilidade.

Uma vez que o conjunto diferença entre as notas tem distribuição normal podemos fazer o teste de comparação de médias para dados pareados.

Um teste bastante usado é o Teste t (de Student) para fazer afirmação sobre a média ou fazer comparações entre duas médias.

As hipóteses do Teste t (de Student):

$$\text{Hipóteses: } \begin{cases} H_0: \mu_y - \mu_x = 0 \\ H_1: \mu_y - \mu_x > 0 \end{cases}$$

Para Mello e Peternelli (2013, p.158), para compararmos médias de duas amostras pareadas podemos usar comando *t.test()*.

Resolvendo com o R:

```
#Síntese:
t.test(Y,X,conf.level = 0.95,paired = T, alternative = "greater")
#aplica o teste t para a comparação de médias de dados pareados unila
teral à direita#

Paired t-test

data: Y and X
t = 2.2063, df = 15, p-value = 0.02169
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.7974      Inf
sample estimates:
mean of the differences
      8.75
```

Pelo resultado apresentado no Teste t de *Student* observa-se que o valor- $p = 0,02169$ , ou seja, menor do que 0,05 levando então a rejeitar a hipótese nula, logo o teste indica que a média a posteriori é maior do que a média a priori.

O teste realizado indica que a diferença média entre as notas dos alunos na avaliação a posteriori em relação às notas da avaliação a priori é de 8,75 pontos.

## 6- CONSIDERAÇÕES FINAIS

A linguagem dos comandos em inglês foi o principal aspecto negativo apontado pelos participantes quanto ao uso do *software R* durante as oficinas.

As principais dificuldades apontadas pelos participantes foram acompanhar o desenvolvimento das oficinas, entender os comandos e saber identificar os erros apontados pelo sistema.

Todavia ao considerarmos os depoimentos dos participantes ao realizarem as atividades, este projeto contribuiu no processo de ensino e aprendizagem de Estatística, pois segundo os participantes foi um aprendizado fácil, rápido, dinâmico que despertou interesse e facilitou os cálculos. Aprendizado que será utilizado tanto na vida escolar quanto na vida profissional.

Diante dos depoimentos apresentados pelos participantes e pelo teste t de *Student* realizado concluímos que os objetivos propostos foram alcançados, pois os alunos aprenderam a utilizar os *software R* e *RStudio* para situações em que se aplica a estatística básica e contribuiu para que os alunos apliquem corretamente os conceitos de estatística básica, fato este comprovado pela avaliação a posteriori.

Este projeto será apresentado para a comunidade escolar do IFRO – *Campus Ji-Paraná* e para a comunidade externa como ação de extensão visando contribuir para a melhoria do ensino de Estatística na região.

Este material servirá de incentivo para que outros professores adotem o laboratório de informática para trabalhar estatística básica utilizando o *software R* ou outro *software*.



## 7- REFERÊNCIAS

BRASIL. Ministério da Educação e do Desporto. *Lei de Diretrizes e Bases da Educação Nacional*. Brasília: MEC, 2005.

\_\_\_\_\_. Ministério da Educação e do Desporto. *Orientações Curriculares para o Ensino Médio: ciências da natureza, matemática e suas tecnologias*. Brasília: MEC; SEB, 2006.

\_\_\_\_\_. Ministério da Educação e do Desporto. *Parâmetros Curriculares Nacionais do Ensino Médio*. Vol. 1. Bases Legais. MEC; SEMTEC, 1999/2002.

\_\_\_\_\_. Ministério da Educação e do Desporto. *Parâmetros Curriculares Nacionais do Ensino Médio*. Vol. 3. Ciências da Natureza, Matemática e suas Tecnologias. MEC; SEMTEC, 1999/2002.

BUSSAB, Wilton de Oliveira; MORETTIN, Pedro Alberto. *Estatística Básica*. 8º ed. São Paulo, SP: Saraíva, 2013.

DANTAS, Geórgia Geogletti Cordeiro; AQUINO, Mirian de Albuquerque . *Aprendendo com o uso de softwares educativos para ativar inteligências múltiplas (IM)*. <http://seer.ufrgs.br/index.php/EmQuestao/article/view/55/1576> . Acessado em 15 de novembro de 2016.

D`AMBROSIO, Ubiratan. *Educação Matemática: da teoria à pratica*. 23º ed. Campinas, SP: Papyrus, 2012.

ESCÓSSIA, Liliana de. *Relação homem-técnica e processos de individuação*. Aracaju: Fundação Oviêdo Teixeira, 1999.

FERREIRA, Daniel Furtado. *Estatística Básica*. 2. ed. rev., Lavras: Ed. UFLA, 2009.

FREIRE, Paulo. *Pedagogia da autonomia*. São Paulo: Paz e Terra, 1996.

IEZZI, Gelson; DEGENSZAJN, David. *Fundamentos de matemática elementar: Matemática comercial, Financeira, Estatística*, 11. 1. ed. , São Paulo: Atual Editora, 2004.

LÈVY, P. *As Tecnologias da inteligência: o futuro do pensamento na era da informática*. Rio de Janeiro: Editora 34, 1993.

MELLO, Marcio Pupin; PETERNELLI, Luiz Alexandre. *Conhecendo o R: Uma visão mais que Estatística*. Viçosa: UFV, 2013.

MORAN, J. M.; MASETTO, M. T.; BEHRENS, M. A. *Novas tecnologias e mediação pedagógica*. 3. ed. Campinas: Papyrus, 2001.

NOVAES, Diva Valério; COUTINHO, Cileda Queiroz e Silva. *Estatística Para Educação Profissional e Tecnológica*. 2. ed., São Paulo: Atlas, 2013.

NUNES, C. *Informação Planejada*. Moderna online. Disponível em

<<http://www.moderna.com.br/pedagogia/infoplan> . Acessado em 20 de novembro de 2016.

PINTO, Suzi Samá; SILVA, Carla Silva da. *Estatística Vol. I*. Porto Alegre: A autora, 2013.

POMMER, Wagner Marcelo. *A Engenharia Didática em sala de aula: Elementos básicos e uma ilustração envolvendo as Equações Diofantinas Lineares*, 2013. Disponível em: <http://stoa.usp.br/wmpommer/files/3915/20692/Livro+Eng%C2%AA+Did%C3%A1tica+2013.pdf> , Acessado em 21 de julho de 2016.

SERAFIM, Maria Lúcia; SOUZA, Robson Pequeno de. Multimídia na educação: o vídeo digital integrado ao contexto escolar. In (org) *Tecnologias Digitais na Educação* 2011. Campina Grande.PB. <http://static.scielo.org/scielobooks/6pdyn/pdf/sousa-9788578791247.pdf> . Acessado em 17 de novembro de 2016.

SILVA, Mozart Linhares da. *A urgência do tempo: novas tecnologias e educação contemporânea*. In: \_\_\_\_ (org.) *Novas Tecnologias: educação e sociedade na era da informática*. Belo Horizonte: Autêntica, 2001.

SOUZA, Renata Beduschi. *O uso das tecnologias na educação*. <https://loja.grupoa.com.br/revista-patio/artigo/5945/o-uso-das-tecnologias-na-educacao.aspx> . Acessado em 20 de novembro de 2016.

TAVARES, Silvio Mauro. *Tecnologia em Educação*. 2007. Postado por Heloiza Piller em agosto de 2012. <http://helopiller.blogspot.com.br/2012/08/tecnologia-e-educacao.html> . Acessado em 16 de novembro de 2016.

## APÊNCIDE A

### LISTA DE INTERESSADOS

Prezado(a) aluno(a) se é de seu interesse participar do projeto do mestrando pesquisador *Leonardo Mota de Andrade*, aluno devidamente matriculado no Programa de Pós-Graduação Stricto Sensu em Matemática – PROFMAT da Universidade Federal de Rondônia, que realizará a pesquisa intitulada “**Estatística no ensino médio: uma proposta de ensino usando o software R**” assine o teu nome abaixo. A assinatura não garante a participação no projeto, pois o número de vagas é limitado em 20 vagas. Caso o número de interessados exceda o número de vagas a escolha dos participantes ficará a cargo do professor da disciplina de Estatística.

Nº	ASSINATURA
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	

## APÊNDICE B

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezados pais ou responsáveis,

O mestrando pesquisador *Leonardo Mota de Andrade*, aluno devidamente matriculado no Programa de Pós-Graduação Stricto Sensu em Matemática – PROFMAT da Universidade Federal de Rondônia, que realiza a pesquisa intitulada “**Estatística no ensino médio: uma proposta de ensino usando o software R.**”, para que possa contar com vossa autorização para que seu filho ou menor sob sua responsabilidade participem da pesquisa da seguinte forma:

- Respondam duas avaliações de estatística básica sendo uma no início e outra no final do projeto;
- Participem das oficinas ministradas em laboratórios de informática em horário oposto ao de estudos regular. (*obs: O dia e horário serão definidos com os participantes e será uma vez na semana num total de 8 a 10 encontros*)

Informamos que o caráter ético desta pesquisa assegura o sigilo das informações coletadas, mediante vossa autorização, garantindo, a preservação da identidade e da privacidade da instituição e dos sujeitos, bem como, o retorno dos resultados da pesquisa aos sujeitos envolvidos, na forma de um relatório de dissertação de pesquisa.

Esclarecemos que a autorização é uma pré-condição bioética para a execução de qualquer estudo envolvendo seres humanos, sob qualquer forma ou dimensão, em consonância com a Resolução 196/96, do Conselho Nacional de Saúde.

Agradecemos vossa compreensão e colaboração com o desenvolvimento de pesquisa científica em nossa Região.

Colocamo-nos à vossa disposição, conforme segue: Celular do Professor Pesquisador: (69) 99206-4635 ou (69) 3421-6870 E-mail: [leonardo.mota@ifro.edu.br](mailto:leonardo.mota@ifro.edu.br).

Sendo o que tínhamos para o momento, agradecemos antecipadamente.

Eu, \_\_\_\_\_, após a leitura deste documento e ter a oportunidade de conversar com o pesquisador responsável, para esclarecer todas as minhas dúvidas, acredito está suficientemente informado, ficando claro para mim. A minha assinatura a seguir representa o meu aceite em liberar o meu filho ou criança sob minha responsabilidade para a referida pesquisa, bem como fica assegurado minha privacidade de acordo com a Resolução CNS 466/2012 da CONEP, que regulamenta a realização de pesquisa envolvendo seres humanos.

Ji-Paraná, \_\_\_\_ de \_\_\_\_\_ de 2016.

---

**Leonardo Mota de Andrade**  
Mestrando Pesquisador

---

**Assinatura**  
Pais ou responsáveis.

## APÊNDICE C

### TERMO DE CONSENTIMENTO E ESCLARECIMENTO

Eu, \_\_\_\_\_ portador (a) do Registro de Identidade (ou outro documento), nº \_\_\_\_\_, concordo em participar da pesquisa Intitulada: **“Estatística no ensino médio: uma proposta de ensino usando o software R”** objeto de estudo do Mestrando *Leonardo Mota de Andrade*, da Universidade Federal de Rondônia, sob orientação da *Prof<sup>o</sup>. Dr<sup>o</sup>. Flávio Batista Simão*, do Programa de Pós-Graduação Stricto Sensu PROFMAT/UNIR. O presente trabalho tem por objetivo propor a utilização de uma ferramenta para o ensino de estatística no ensino médio o *software R*.

Assim, tenha conhecimento que:

1. A participação do senhor(a) na pesquisa será respondendo duas avaliações com questões de estatística básica sendo uma no início e outra no final do projeto, como também participando das oficinas em horário de aula a ser realizada no laboratório de informática;
2. Suas respostas serão tratadas de forma anônima e confidencial, isto é, em nenhum momento será divulgado o seu nome em qualquer fase do estudo;
3. O benefício relacionado à sua participação será de aumentar o conhecimento científico para a área da Educação;
4. Os resultados da pesquisa serão destinados exclusivamente à elaboração de trabalhos de pesquisas e/ou publicações científicas em revistas nacionais ou internacionais;
5. Fica assegurada ao senhor (a) a possibilidade de desistir de participar da pesquisa a qualquer momento, ou mesmo recusar-se a responder perguntas que lhe cause constrangimento, sem nenhum prejuízo à sua pessoa;
6. Fica assegurado uma cópia do Termo de Consentimento Livre e Esclarecido ao participante da pesquisa;
7. Fica esclarecido que o senhor (a) não receberá nenhuma remuneração por estar participando da pesquisa;

A assinatura a seguir representa o meu aceite em participar do estudo, bem como fica assegurado minha privacidade de acordo com a Resolução CNS 466/2012, que regulamenta a realização de pesquisa envolvendo seres humanos.

Ji-Paraná, \_\_\_\_\_ de \_\_\_\_\_ de 2016.

---

**Leonardo Mota de Andrade**  
Pesquisador

---

**Assinatura**  
Participante

## APÊNDICE D

### TERMO DE ANUÊNCIA

De conformidade com o Regimento Interno do Mestrado Profissional em Matemática em Rede Nacional, PROFMAT/UNIR - nível mestrado profissional, Eu, *Prof. Dr. Flávio Batista Simão*, aceito o compromisso de orientar e acompanhar, *Leonardo Mota de Andrade*, aprovado no exame nacional de qualificação, ENQ 2016/01, na condução dos trabalhos da pesquisa intitulada: **“Estatística no ensino médio: uma proposta de ensino usando o software R”** para elaboração de dissertação visando à obtenção do título de Mestre em Matemática.

Ji-Paraná, \_\_\_\_ de \_\_\_\_\_ de 2016.

---

Orientador

## APÊNDICE E

UNIVERSIDADE FEDERAL DE RONDÔNIA - UNIR  
PROGRAMA DE MESTRADO PROFISSIONAL EM MATEMÁTICA PROFMAT

### AVALIAÇÃO A PRIORI NOÇÕES DE ESTATÍSTICA BÁSICA

ALUNO: \_\_\_\_\_ GRUPO: \_\_\_\_\_

#### COMPETÊNCIAS E HABILIDADES

- Ler e interpretar dados e informações de caráter estatístico apresentados em diferentes linguagens e representações, na mídia ou em outros textos e meios de comunicação.
- Identificar formas adequadas para descrever e representar dados numéricos e informações de natureza social, econômica, política, científico-tecnológica ou abstrata.
- Obter médias e avaliar desvios de conjuntos de dados ou informações de diferentes naturezas.
- Compreender e emitir juízos sobre informações estatísticas de natureza social, econômica, política ou científica apresentadas em textos, notícias, propagandas, censos, pesquisas e outros meios.
- Quantificar e fazer previsões em situações aplicadas a diferentes áreas do conhecimento e da vida cotidiana que envolva o pensamento estatístico.
- Identificar em diferentes áreas científicas e outras atividades práticas modelos e problemas que fazem uso de estatísticas.

1) **Questão**

Qual a definição correta de amostra no campo da estatística?

- a) amostra é o levantamento que se faz com todos os indivíduos de uma população.
- b) amostra é uma tabela de dados obtida com uma pesquisa.
- c) amostra é o conjunto de dados obtidos em uma pesquisa.
- d) amostra é um subconjunto finito de uma população.
- e) amostra é um brinde que ganhamos em um evento.

2) **Questão**

Qual a definição de variável no campo da estatística.

- a) uma letra que representa um valor desconhecido
- b) uma letra que representa um valor variável
- c) é um subconjunto finito de uma população.
- d) é o levantamento que se faz com todos os indivíduos de uma população.
- e) é o conjunto de resultados possíveis de um fenômeno

3) **Questão**

A pesca não predatória pressupõe que cada peixe retirado de seu hábitat já tenha procriado, pelo menos uma vez. Para algumas espécies, isso ocorre depois dos peixes apresentarem a máxima variação anual de seu peso.

O controle de pesca no Pantanal é feito com base no peso de cada espécie.

A tabela fornece o peso do pacu, uma dessas espécies, em cada ano:

Idade	1	2	3	4	5	6	7	8	9	10	11	12	13
Peso(kg)	1,1	1,7	2,6	3,9	5,1	6,1	7	7,8	8,5	8,9	9,1	9,3	9,4

Considerando esses dados, a pesca do pacu deve ser autorizada para espécimes com peso de, no mínimo:

- a) 3
- b) 4
- c) 5
- d) 6
- e) 8

4) **Questão**

Os candidatos K, L, M, N e P estão disputando uma única vaga de emprego em uma empresa e fizeram provas de português, matemática, direito e informática. A tabela apresenta as notas obtidas pelos cinco candidatos.

Candidatos	Português	Matemática	Direito	Informática
K	33	33	33	34
L	32	39	33	34
M	35	35	36	34
N	24	37	40	35
P	36	16	26	41



Segundo o edital de seleção, o candidato aprovados será aquele para a qual a mediana das notas obtidas por ele nas quatros disciplinas for maior.

O candidato aprovado será:

- a) K
- b) L
- c) M
- d) N
- e) P

**5) Questão**

Em um grupo de dez amigos suas idades são 12 anos, 10 anos, 13 anos, 11 anos, 10 anos, 13 anos, 11 anos, 13 anos, 12 anos e 15 anos. Nessas condições podemos concluir que a média de idade desse grupo é:

- a) 10 anos;
- b) 10,5 anos
- c) 11 anos
- d) 11,5 anos
- e) 12 anos

**6) Questão**

A tabela abaixo se refere a uma pesquisa realizada com 20 alunos, a respeito da área da carreira universitária que cada um pretende seguir.

Área	Frequência simples	Frequência relativa	porcentagem
Humanas	8	0,4	40%
Biológicas	?	0,35	35%
Exatas	5	?	25%
Total	20	1,00	100%

A frequência simples e a frequência relativa que faltam (assinaladas por ?) são respectivamente:

- a) 7 e 0,25
- b) 7 e 0,35
- c) 5 e 0,35
- d) 7 e 0,4
- e) 5 e 0,25

**7) Questão**

Uma loja que vende sapatos recebeu diversas reclamações de seus clientes relacionadas à venda de sapatos de cor branca ou preta. Os donos da loja anotaram as numerações dos sapatos com defeito e fizeram um estudo estatístico com o intuito de reclamar com o fabricante.

A tabela contém a média, a mediana e a moda desses dados anotados pelos donos.

<b>Estatísticas sobre as numerações dos sapatos com defeito</b>			
	Média	Mediana	Moda
Numerações dos sapatos com defeito	36	37	38

Para quantificar os sapatos pela cor, os donos representaram a cor branca pelo número 0 e a cor preta pelo número 1. Sabe-se que a média da distribuição desses zeros e uns é igual a 0,45.

Os donos da loja decidiram que a numeração dos sapatos com maior número de reclamações e a cor com maior número de reclamações não serão mais vendidas.

A loja encaminhou um ofício ao fornecedor dos sapatos, explicando que não serão mais encomendados os sapatos de cor:

- a) branca e os de número 38.
- b) branca e os de número 37.
- c) branca e os de número 36.
- d) preta e os de número 38.
- e) preta e os de número 37.

### 8) *Questão*

Em uma corrida de regularidade, a equipe campeã é aquela em que o tempo dos participantes mais se aproxima do tempo fornecido pelos organizadores em cada etapa. Um campeonato foi organizado em 5 etapas, e o tempo médio de prova indicado pelos organizadores foi de 45 minutos por prova. No quadro, estão representados os dados estatísticos das cinco equipes mais bem classificadas.

**Dados estatísticos das equipes mais bem classificadas (em minutos)**

Equipes	Média	Moda	Desvio-padrão
Equipe I	45	40	5
Equipe II	45	41	4
Equipe III	45	44	1
Equipe IV	45	44	3
Equipe V	45	47	2

Utilizando os dados estatísticos do quadro, a campeã foi a equipe:

- a) I
- b) II
- c) III
- d) IV
- e) V

### 9) *Questão*

A tabela a seguir representa a distribuição de frequências dos salários de um grupo de 50 empregados de uma empresa, num certo mês.

Número de classe	Salário do mês em R\$	Número de empregados
1	1.500 + 2.500	22
2	2.500 + 3.500	18
3	3.500 + 4.500	8
4	4.500 + 5.500	2

O salário médio desses empregados, nesse mês, foi de:

- a) R\$ 2 800,00
- b) R\$ 2 720,00
- c) R\$ 2 500,00
- d) R\$ 2 420,00
- e) R\$ 2 200,00

**10) Questão**

As seguintes distribuições apresentam médias iguais. Qual delas apresenta maior dispersão?

GRUPO	DISTRIBUIÇÃO	MÉDIA
A	7,6 – 7,6 – 7,6 – 7,6 – 7,6 – 7,6 – 7,6 – 7,6 – 7,6 – 7,6	7,6
B	7 – 8 – 8 – 9 – 7 – 6 – 8 – 9 – 8 – 6	7,6
C	10 – 9 – 6 – 5 – 9 – 12 – 7 – 7 – 5 – 6	7,6
D	4 – 5 – 4 – 7 – 9 – 12 – 6 – 8 – 10 – 11	7,6

- a) A
- b) B
- c) C
- d) D
- e) Não é possível identificar

## APÊNDICE F

### Atividades-Oficinal

- 1) Para cada conjunto abaixo use o RStudio para criar um vetor ou um fator.  
 $A = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$   
 $B = \{a, e, i, o, u\}$   
 $C = \{4, 5, 1, 2, 3, 4, 5, 6, 7, 2, 2, 3, 3, 4, 4, 6, 6, 7, 8, 3\}$   
 $D = \{\text{primavera, verão, outono, inverno}\}$   
 $E = \{\text{Palmeiras, Flamengo, Atlético Mineiro, Corinthians}\}$
- 2) Crie uma tabela do conjunto E organizada por linhas.
- 3) Crie uma tabela do conjunto D organizada por colunas.
- 4) Crie uma tabela 4 x 5 do conjunto C organizada por linhas.
- 5) Confira abaixo a lista dos 10 modelos de automóveis mais vendidos no mês de agosto de 2016.

Ranking	Carro	Marca	Quantidade
1	Onix	Chevrolet	12.422
2	HB20	Hyundai	10.263
3	Ka	Ford	7.043
4	Palio	Fiat	6.750
5	Sandero	Renault	6.657
6	Corolla	Toyota	6.030
7	Prisma	Chevrolet	5.848
8	HR-V	Honda	4.906
9	Gol	<b>Volkswagen</b>	4.525
10	Renegade	<b>Jeep</b>	4.483

Fonte: Federação dos Distribuidores de Veículos.

Usando os comandos necessários crie esta tabela no RStudio.

- 6) Observe a tabela abaixo:

Nome	Sexo	Idade	Altura	Peso	Irmaos	Cabelo	Hobby	Sapato	Manequim
<b>Antonio</b>	M	15	156	49	2	Castanho	esporte	36	38
<b>Artur</b>	M	14	166	48	0	Castanho	musica	39	38
<b>Bruno</b>	M	15	165	66	1	loiro	patinacao	36	42
<b>Carla</b>	F	14	175	63	0	Ruivo	danca	36	42
<b>Claudia</b>	F	15	165	57	2	Preto	esporte	36	40

Usando os comandos necessários crie esta tabela no RStudio.

## APÊNDICE G

### Atividades-OficinaII

1) Relacione as colunas:

( a ) <i>População</i>	( ) convencionalmente, é o conjunto de resultados possíveis de um fenômeno.
( b ) <i>Amostra</i>	( ) identifica-se pelo caráter variável do fator cronológico. O local e a espécie (fenômeno) são elementos fixos.
( c ) <i>Variável</i>	( ) conjunto de elementos em estudo que tem pelo menos uma característica em comum. (pessoas, coisas, objetos etc.)
( d ) <i>Série temporal</i>	( ) subconjunto de elementos extraídos de uma população. Ou ainda, uma parte da população que é selecionada para análise.
( e ) <i>Série geográfica</i>	( ) o caráter variável é apenas o fato ou espécie. Também é chamada de série categórica.
( f ) <i>Série específica</i>	( ) quando existe a combinação entre as séries cronológicas, geográficas e específicas, visto que podem variar o fato, o lugar e o tempo.
( g ) <i>Série conjugada</i>	( ) apresenta como elemento variável o fator geográfico. A época e o fato (espécie) são elementos fixos. Também é chamada de espacial, territorial ou de localização.
( h ) <i>Tabela</i>	( ) é um quadro que resume um conjunto de dados dispostos segundo linhas e colunas de maneira sistemática.

2) Classifique as variáveis abaixo em qualitativas nominais (N), qualitativas ordinais (O), quantitativas contínuas (C) ou quantitativas discretas (D).

- a) Cor dos olhos das alunas do 2º C... ( )
- b) Índice de liquidez nas indústrias capixaba... ( )
- c) Número de defeitos em aparelhos de TV... ( )
- d) Comprimento dos pregos produzidos por uma empresa... ( )
- e) O ponto obtido em cada jogada de um dado... ( )
- f) Grau de escolaridade dos funcionários de uma empresa... ( )
- g) Sexo ( )
- h) Idade ( )
- i) Número de alunos de uma sala de aula ( )
- j) Salário ( )
- k) Grau de escolaridade ( )
- l) Temperatura ( )
- m) Religião ( )
- n) Raça ( )
- o) Estatura (metros) ( )
- p) Nível sócio econômico ( )

q) Número de faltas recebido por um aluno ( )

3) Complete os dados que faltam na distribuição de frequência:

$I$	$x_i$	$f_i$	$fr_i$	$fp$	$F_{ac\downarrow}$
1	0	1	0,05		
2	1		0,15		4
3	2	4			
4	3		0,25		13
5	4	3	0,15		
6	5	2			18
7	6				19
8	7				
	<b>Total</b>	<b>20</b>	<b>1,00</b>		

4) Dados referentes ao número de ovos danificados em uma inspeção feita em 20 embalagens, de uma dúzia cada, em um carregamento para o mercado municipal de Lavras proveniente de uma cidade distante.

$$X = \{0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 5\}, n = 20$$

Considerando os dados acima, monte uma tabela de distribuição de frequências sem intervalo de classe, contendo a frequência absoluta, a frequência absoluta acumulada, a frequência relativa e a frequência percentual.

5) Idade de 40 turistas que visitaram Ji-Paraná/RO no último feriado de 7 de setembro de 2016.

53, 23, 30, 48, 42, 40, 14, 32, 52, 17, 51, 48, 45, 49, 38, 62, 51, 47, 50, 52, 55, 63, 66, 54, 62, 60, 27, 42, 51, 37, 43, 15, 42, 28, 48, 49, 40, 39, 48, 62

Considere os dados monte uma tabela de distribuição de frequências, contendo intervalo de classes, o ponto médio, a frequência absoluta, a frequência absoluta acumulada, a frequência relativa e a frequência percentual.

## APÊNDICE H

### Atividades-OficinaIII

- 1) Dados brutos referentes ao número de ovos danificados em uma inspeção feita em 20 embalagens, de uma dúzia cada, em um carregamento para o mercado municipal de Lavras proveniente de uma cidade distante. Qual é a média, mediana e moda de ovos danificados por embalagem?

$$X = \{3, 2, 1, 2, 3, 5, 1, 1, 3, 4, 1, 1, 1, 1, 1, 1, 0, 2, 0, 0\}, n = 20$$

- 2) A tabela apresenta os resultados obtidos por três candidatos na primeira fase de um concurso de vestibular.

PROVA	PESO	PONTOS		
		Candidato A	Candidato B	Candidato C
<i>Conhecimentos Gerais</i>	1,5	40	30	52
<i>Redação</i>	4,5	6	5	5,5
<i>Língua Portuguesa</i>	2,5	18	15	11
<i>Língua Estrangeira</i>	1,5	9	10	9

Calcule a média aritmética ponderada de cada candidato e coloque-os em ordem de classificação.

- 3) Foram registradas as seguintes medidas para o tempo de secagem, em horas, de certa marca de tinta látex.

3,4   2,5   4,8   2,9   3,6  
2,8   3,3   5,6   3,7   2,8  
4,4   4,0   5,2   3,0   4,8

Determine a media, mediana e a moda dos valores registrados.

- 4) Tabela com o número de gols por partida registrada na 24ª e 25ª rodadas do campeonato brasileiro série A de 2016.

Número de gols	$f_i$
0	1
1	6
2	4
3	5
4	2
6	1
8	1
<b>Total</b>	<b>20</b>

*Fonte: Globo Esporte*

De acordo com as informações do quadro calcule:

- a) Média
- b) Mediana

c) Moda

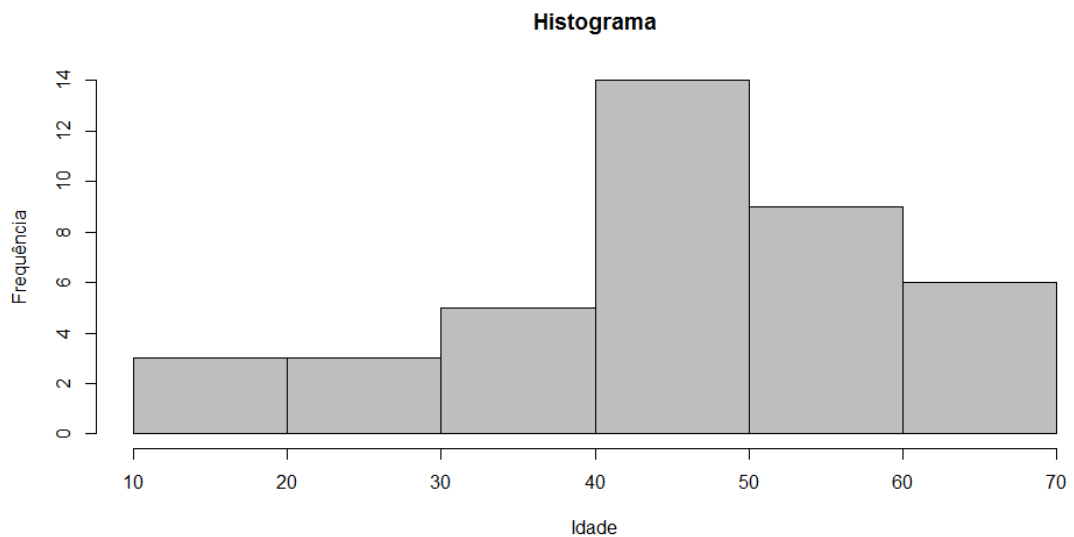
- 5) A tabela, a seguir, mostra os resultados de uma pesquisa sobre a faixa salarial dos funcionários de uma empresa que usam motocicletas para ir ao trabalho.

Faixa salarial (R\$)	$f_i$
650 ┆ 750	38
750 ┆ 850	26
850 ┆ 950	20
950 ┆ 1050	18
1050 ┆ 1150	12
1150 ┆ 1250	6
<b>Total</b>	

Calcule:

- Média.
- Mediana.
- Moda.

- 6) O histograma seguinte mostra as idades de 40 turistas que visitaram Ji-Paraná no feriado de 7 setembro de 2016.



Calcule:

- A idade média.
- A idade mediana.
- A idade modal.



## APÊNDICE I

### Atividades-OficinaIV

- 1) Considere a seguinte distribuição amostral  $Y$  das notas de 15 alunos em Estatística.

$$Y = \{2, 2, 7, 7, 7, 8, 3, 4, 4, 6, 9, 9, 4, 4, 7\}$$

Responda:

- a) Qual é o desvio médio destas notas?
  - b) Qual é a variância destas notas?
  - c) Qual é o desvio padrão?
- 2) A tabela abaixo apresenta os resultados das mensurações (em mm) da quantidade de chuva em Ji-Paraná no período de janeiro a março de 2016.

**Tabela I** – Tabela com o resultado de 40 observações relativas ao índice pluviométrico em Ji-Paraná/RO nos dias chuvosos de janeiro a março de 2016.

Quantidade de chuva (mm)	$f_i$
135	7
144	8
148	5
157	10
162	6
175	4
<b>Total</b>	<b>40</b>

Fonte: Dados Fictícios

Determine:

- a) O desvio médio.
  - b) A variância. Qual é a unidade da variância?
  - c) O desvio padrão. Qual é a unidade do desvio padrão?
- 3) Uma pesquisa realizada em alguns hotéis do litoral norte de São Paulo teve o objetivo de identificar os preços praticados nas diárias da região na baixa temporada de 2016. Determine a variância, desvio padrão e o desvio médio para os dados coletados.

**Tabela II**- Tabela de valores de diárias praticada no litoral norte de São Paulo na período de baixa temporada de 2016.

Valor da diária (R\$)	$f_i$
40 – 60	6
60 – 80	18
80 – 100	12
100 – 120	12
120 – 140	6
140 – 160	6
<b>Total</b>	<b>60</b>

Fonte: Adaptado de Novaes e Coutinho

## APÊNDICE J

### Atividades-OficinaV

- 1) Em uma turma de 16 alunos perguntou-se no mês de setembro de 2016, qual time vai ganhar o campeonato brasileiro série A de futebol em 2016?

*Respostas:*

*Palmeiras, Flamengo, Corinthians, Grêmio, Atlético Mineiro, Fluminense, Flamengo, Corinthians, Santos, Flamengo, Palmeiras, Flamengo, Atlético Mineiro, Palmeiras, Corinthians, Palmeiras.*

- a) Represente os dados utilizando um gráfico de colunas.  
 b) Represente os dados utilizando um gráfico de barras.  
 c) Represente os dados utilizando um gráfico de setores.
- 2) A tabela abaixo apresenta os 10 modelos de carro mais vendido no Brasil no mês de agosto de 2016.

**Tabela I** - Lista dos 10 modelos de automóveis mais vendidos no Brasil no mês de agosto de 2016.

Carro	Quantidade
Onix	12.422
HB20	10.263
Ka	7.043
Palio	6.750
Sandero	6.657
Corolla	6.030
Prisma	5.848
HR-V	4.906
Gol	4.525
Renegade	4.483

- a) Represente esta tabela usando um gráfico de colunas.  
 b) Represente esta tabela usando um gráfico de setores.  
 c) Qual é a melhor representação?
- 3) Idade de 40 turistas que visitaram Ji-Paraná/RO no último feriado de 7 de setembro de 2016.  
 53, 23, 30, 48, 42, 40, 14, 32, 52, 17, 51, 48, 45, 49, 38, 62, 51, 47, 50, 52, 55, 63, 66, 54, 62, 60, 27, 42, 51, 37, 43, 15, 42, 28, 48, 49, 40, 39, 48, 62
- a) Represente os dados em um histograma.  
 b) Represente os dados em gráfico de linhas.  
 c) Qual é a melhor representação?



## APÊNDICE L

### AVALIAÇÃO A POSTERIORI NOÇÕES DE ESTATÍSTICA BÁSICA

ALUNO: \_\_\_\_\_ GRUPO/Nº: \_\_\_\_\_

#### COMPETÊNCIAS E HABILIDADES

- Ler e interpretar dados e informações de caráter estatístico apresentados em diferentes linguagens e representações, na mídia ou em outros textos e meios de comunicação.
- Identificar formas adequadas para descrever e representar dados numéricos e informações de natureza social, econômica, política, científico-tecnológica ou abstrata.
- Obter médias e avaliar desvios de conjuntos de dados ou informações de diferentes naturezas.
- Compreender e emitir juízos sobre informações estatísticas de natureza social, econômica, política ou científica apresentadas em textos, notícias, propagandas, censos, pesquisas e outros meios.
- Quantificar e fazer previsões em situações aplicadas a diferentes áreas do conhecimento e da vida cotidiana que envolva o pensamento estatístico.
- Identificar em diferentes áreas científicas e outras atividades práticas modelos e problemas que fazem uso de estatísticas.

**1) Questão**

Qual a definição correta de população no campo da estatística?

- f) é o conjunto de elementos em estudo que tem pelo menos uma característica em comum. (pessoas, coisas, objetos etc.)
- g) convencionalmente, é o conjunto de resultados possíveis de um fenômeno.
- h) é o conjunto de dados obtidos em uma pesquisa.
- i) é o conjunto de pessoas que residem em um país.
- j) é um quadro que resume um conjunto de dados dispostos segundo linhas e colunas de maneira sistemática.

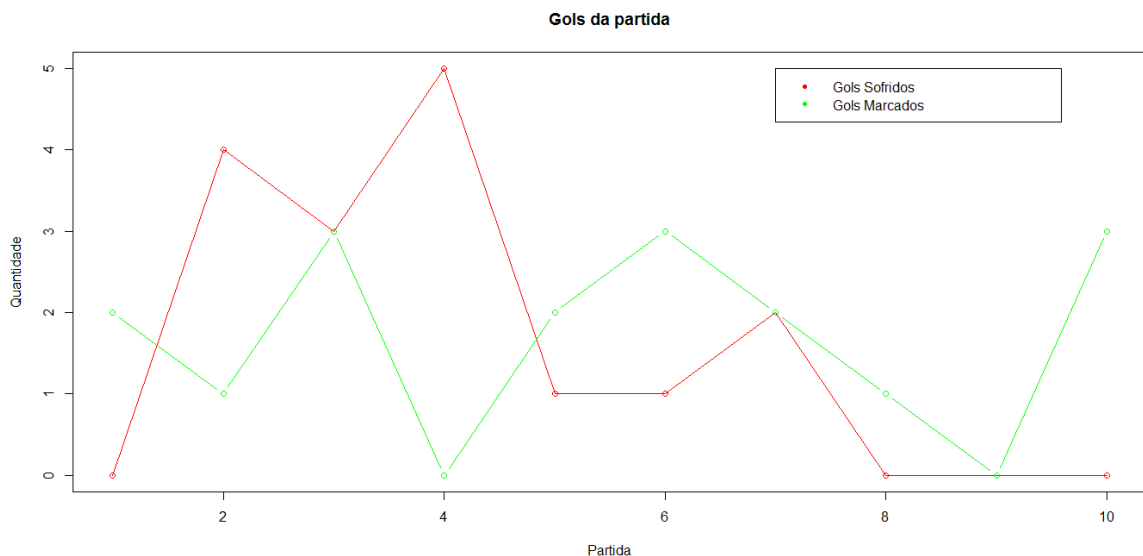
**2) Questão**

Qual a definição de *Série Geográfica* no campo da estatística?

- a) é um quadro que resume um conjunto de dados dispostos segundo linhas e colunas de maneira sistemática.
- b) é uma tabela que apresenta como elemento variável a localização. A época e o fato (espécie) são elementos fixos.
- c) é uma tabela onde o caráter variável é apenas o fato ou espécie. Também é chamada de série categórica.
- d) é uma tabela que identifica-se pelo caráter variável do fator cronológico. O local e a espécie (fenômeno) são elementos fixos.
- e) quando existe a combinação entre as séries cronológicas e específicas, visto que podem variar o fato e o tempo.

**3) Questão**

No gráfico estão apresentados os gols marcados e os gols sofridos nas 10 primeiras rodadas de um campeonato:



Considerando que, neste campeonato, as equipes somam 3 pontos para cada vitória, 1 ponto por empate e 0 ponto em caso de derrota, a equipe em questão, ao final da 10 partida, terá acumulado um número de pontos igual a:

- a) 15
- b) 17
- c) 18
- d) 20
- e) 24

4) **Questão**

Na tabela, são apresentados dados da cotação mensal do ovo extra branco vendido no atacado, em Brasília em reais, por caixa de 30 dúzias de ovos, em alguns meses de 2007 e 2008.

Mês	Cotação	Ano
Outubro	R\$ 83,00	2007
Novembro	R\$ 73,10	2007
Dezembro	R\$ 81,60	2007
Janeiro	R\$ 82,00	2008
Fevereiro	R\$ 85,30	2008
Março	R\$ 84,00	2008
Abril	R\$ 84,60	2008

De acordo com estes dados, o valor da mediana das cotações mensais do ovo extra branco nesse período era igual a:

- a) R\$ 73,10
- b) R\$ 81,50
- c) R\$ 82,00
- d) R\$ 83,00
- e) R\$ 85,30

5) **Questão**

Cinco amigos marcaram uma viagem à praia em dezembro. Para economizar, combinaram de ir num único carro. Cada amigo anotou quantos quilômetros seu carro fez, em média, por litro de gasolina, nos meses de setembro, outubro e novembro. Ao final desse trimestre, calcularam a média dos três valores obtidos para escolherem o carro mais econômico, ou seja, o que teve a maior média. Os dados estão representados na tabela:

Carro	Desempenho médio mensal (km/litro)		
	Setembro	Outubro	Novembro
I	6,2	9,0	9,3
II	6,7	6,8	9,5
III	8,3	8,7	9,0
IV	8,5	7,5	8,5
V	8,0	8,0	8,0

Qual carro os amigos deverão escolher para a viagem?

- a) I
- b) II
- c) III
- d) IV
- e) V

6) **Questão**

A tabela abaixo se refere a uma pesquisa realizada com 40 alunos a respeito da área da carreira universitária que cada um pretende seguir, porém a tabela está incompleta.

<i>Área</i>	<i>Frequência simples</i>	<i>Frequência relativa</i>	<i>Frequência acumulada</i>
<i>Humanas</i>	?		
<i>Biológicas</i>	12	?	
<i>Exatas</i>	8	0,20	?
<i>Outras</i>	10	0,25	40
<i>Total</i>	40	1,00	

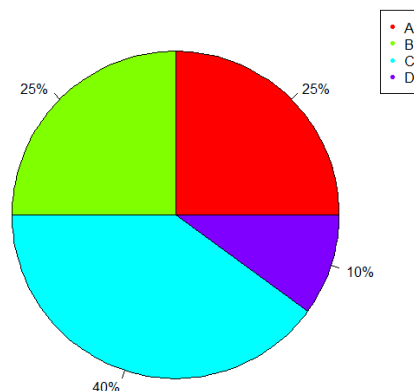
A frequência simples, frequência relativa e frequência acumulada que faltam (assinaladas por ?) são respectivamente:

- a) 8 – 0,25 – 28
- b) 8 – 0,20 – 30
- c) 10 – 0,25 – 38
- d) 12 – 0,30 – 32
- e) 10 – 0,30 – 30

7) **Questão**

Foi realizado um levantamento nos 200 hotéis de uma cidade, no qual foram anotados os valores, em reais, das diárias para um quarto padrão de casal e a quantidade de hotéis para cada valor da diária. Os valores das diárias foram: A = R\$ 200,00; B = R\$ 300,00; C = R\$ 400,00 e D = R\$ 600,00. No gráfico, as áreas representam as quantidades de hotéis pesquisados, em porcentagem, para cada valor da diária.

Porcentagem de Hotéis - Valor de Diária



O valor modal e o valor mediano da diária, em reais, para o quarto padrão de casal nessa cidade, são respectivamente:

- a) 200 e 300
- b) 300 e 400
- c) 300 e 350
- d) 400 e 300
- e) 400 e 350

**8) Questão**

A tabela a seguir mostra a evolução da receita bruta anual nos três últimos anos de cinco microempresas (ME) que se encontram à venda.

ME	2013 (em milhares de reais)	2014 (em milhares de reais)	2015 (em milhares de reais)
<b>Alfinetes V</b>	200	220	240
<b>Balas W</b>	200	230	200
<b>Chocolates x</b>	250	210	215
<b>Pizzaria Y</b>	230	210	230
<b>Tecelagem Z</b>	160	220	245

Um investidor deseja comprar duas empresas listadas na tabela. Para tal, ele calcula a variação máxima da receita bruta de cada empresa ao longo dos três anos (de 2013 até 2015) e escolhe as duas empresas com as menores variações.

As empresas que este investidor resolve comprar são:

- a) Balas W e Pizzaria Y
- b) Chocolates X e Tecelagem Z
- c) Pizzaria Y e Alfinetes V
- d) Pizzaria Y e Chocolates X
- e) Tecelagem Z e Alfinetes V

**9) Questão**

A tabela a seguir representa a distribuição de frequências das idades de um grupo de 40 pessoas.

Idade (em anos)	Frequência
15 + 25	8
25 + 35	9
35 + 45	11
45 + 55	7
55 + 65	5
<b>Total</b>	<b>40</b>

A idade média desde grupo, em anos, é de:

- a) 36,25



- b) 38
- c) 39
- d) 40,5
- e) 42

**10) Questão**

Em certo país, o governo financia um programa de assistências às famílias de baixa renda. Cada família recebe, de cinco em cinco semanas, a quantia de 100 UM (unidades monetárias) para comprar produtos de alimentação em estabelecimentos conveniados. O coordenador desse projeto selecionou em uma cidade quatro famílias e acompanhou a distribuição dos gastos semana a semana.

Observe a tabela:

Família	Gastos em UM				
	1° Semana	2° Semana	3° Semana	4° Semana	5° Semana
<b>I</b>	20	20	20	20	20
<b>II</b>	20	24	20	16	20
<b>III</b>	12	28	24	20	16
<b>IV</b>	36	32	20	8	4

Segundo os dados apresentados a família que apresenta a maior variância de gastos durante este período e a família IV com variância de  $160 \text{ UM}^2$ . Qual é aproximadamente o desvio padrão de gastos desta família?

- a) 12 UM
- b) 12,5 UM
- c) 12,6 UM
- d) 12,8 UM
- e) 13 UM