



ANDRÉ DA SILVA COURA

EXPERIMENTOS COM PROBABILIDADE E  
ESTATÍSTICA: JANKENPON, MONTE CARLO,  
VARIÁVEIS ANTROPOMÉTRICAS

CAMPINAS  
2014





UNIVERSIDADE ESTADUAL DE CAMPINAS  
Instituto de Matemática, Estatística e Computação Científica

André da Silva Coura

**EXPERIMENTOS COM PROBABILIDADE E ESTATÍSTICA:  
JANKENPON, MONTE CARLO, VARIÁVEIS  
ANTROPOMÉTRICAS**

Dissertação apresentada ao Instituto de Matemática,  
Estatística e Computação Científica da Universidade  
Estadual de Campians como parte dos requisitos  
exigidos para a obtenção do título de Mestre.

**Orientadora: Profa. Dra. Laura Leticia Ramos Rifo**

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA  
DISSERTAÇÃO DEFENDIDA PELO ALUNO  
ANDRÊ DA SILVA COURA, ORIENTADO PELA  
PROFA. DRA. LAURA LETICIA RAMOS RIFO.

**Assinatura do(a) Orientador(a)**

**CAMPINAS  
2014**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Matemática, Estatística e Computação Científica  
Maria Fabiana Bezerra Muller - CRB 8/6162

C833e Coura, André da Silva, 1984-  
Experimentos com probabilidade e estatística : Jankenpon, Monte Carlo, variáveis antropométricas / André da Silva Coura. – Campinas, SP : [s.n.], 2014.

Orientador: Laura Leticia Ramos Rifo.  
Dissertação (mestrado profissional) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Estatística. 2. Pesquisa experimental. 3. Estatística matemática. 4. Probabilidades. 5. Organização da informação. I. Rifo, Laura Leticia Ramos, 1970-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Experiments with probability and statistics : Jankenpon, Monte Carlo, anthropometric variables

**Palavras-chave em inglês:**

Statistics

Experiential research

Mathematical statistics

Probabilities

Information organization

**Área de concentração:** Matemática em Rede Nacional

**Titulação:** Mestre

**Banca examinadora:**

Laura Leticia Ramos Rifo [Orientador]

Verónica Andrea González-López

Teresa Cristina Martins Dias

**Data de defesa:** 19-11-2014

**Programa de Pós-Graduação:** Matemática em Rede Nacional

**Dissertação de Mestrado Profissional defendida em 19 de novembro de 2014 e  
aprovada Pela Banca Examinadora composta pelos Profs. Drs.**



---

**Prof.(a). Dr(a). LAURA LETICIA RAMOS RIFO**



---

**Prof.(a). Dr(a). VERÓNICA ANDREA GONZÁLEZ-LÓPEZ**



---

**Prof.(a). Dr(a). TERESA CRISTINA MARTINS DIAS**



## Abstract

This dissertation presents a practical approach for teaching mathematics in the elementary and secondary levels. More specifically, presents concepts of Basic Statistics as information processing and the study of probabilities. These concepts are of great importance in scientific (experimental way, for example) and social (understanding of population characteristics), besides being inserted into the daily student's lives. Therefore, it was understood that is necessary to develop the skills and abilities to organize and understand information. Experiments were carried out for the application of the concepts presented in classroom. Also a search posing questions to analyze aspects of food and physical exercise. The realization of these experiments purpose, besides the application of classroom learnt concepts, develop in students, logic reasoning and critical look at issues related to the discipline of mathematics and daily situations by organizing and interpreting information with charts and graphs. The research aimed to show how it is used statistical theory for decision making and, if so, to improve their quality of life. Thus, we intend that presented methodology in this study may contribute to the dissemination of these mathematical knowledge tools for elementary and high school levels.

**Keywords:** Statistics, Experiments, Applied Mathematics, Probability, Information Processing.

## Resumo

A dissertação apresenta uma abordagem prática para o ensino da matemática nos níveis fundamental e médio. De forma mais específica, apresenta conceitos de estatística básica como tratamento de informações e estudo de probabilidades. Estes conceitos são de grande importância no âmbito científico (parte experimental, por exemplo) e social (compreensão de características populacionais), além de estarem inseridos na vida cotidiana dos alunos. Sendo assim, foi entendido que é primordial desenvolver as competências e habilidades para organizar e compreender informações. Foram realizados experimentos para a aplicação dos conceitos apresentados em sala de aula. Também uma pesquisa propondo questões para analisar aspectos sobre alimentação e prática de exercícios físicos. Estes experimentos, além da aplicação dos conceitos, pretendem desenvolver no público-alvo, raciocínio lógico e olhar crítico, para assuntos relacionados à disciplina de matemática, utilizando situações cotidianas. Para análise, organizamos e interpretamos as informações por meio de tabelas e gráficos. A pesquisa teve como objetivo principal mostrar como é usada

a teoria estatística para a tomada de decisão e, nesse caso, para melhorar a própria qualidade de vida. Desse modo, pretendemos que a metodologia apresentada neste trabalho possa contribuir para a disseminação do conhecimento destas ferramentas matemáticas para os níveis fundamental e médio do ensino escolar.

**Palavras-chave:** Estatística, Experimentos, Matemática aplicada, Probabilidade, Tratamento da informação.



# Sumário

Abstract . . . . .	vii
Resumo . . . . .	vii
<b>Agradecimentos</b>	<b>xiii</b>
<b>Prefácio</b>	<b>1</b>
<b>1 Dobra a língua e coça a orelha</b>	<b>5</b>
1.1 O experimento . . . . .	5
1.2 Introdução teórica . . . . .	6
1.3 Execução . . . . .	14
1.4 Análise das informações . . . . .	18
1.5 Conclusão . . . . .	24
1.6 Complementar . . . . .	26
1.6.1 Elementos de amostragem . . . . .	26
1.6.2 Teste de hipóteses . . . . .	36
1.6.3 Função indicadora . . . . .	45
<b>2 Jankenpon</b>	<b>47</b>
2.1 O Experimento . . . . .	47
2.2 Introdução teórica . . . . .	48
2.3 Aplicação . . . . .	50
2.3.1 Par ou ímpar? . . . . .	50
2.3.2 Jogadores usam uma das mãos . . . . .	51
2.3.3 Um jogador usa uma mão e o outro usa as duas . . . . .	52
2.3.4 Ambos os jogadores com as duas mãos . . . . .	53
2.4 Lei dos Grandes Números . . . . .	54
2.4.1 Jankenpon . . . . .	60
2.5 Execução . . . . .	64
2.6 Análise das informações e conclusão . . . . .	66

<b>3</b>	<b>Método de Monte Carlo</b>	<b>69</b>
3.1	O Experimento . . . . .	69
3.2	Introdução teórica . . . . .	69
3.2.1	Propriedades da esperança de uma variável aleatória . . . . .	72
3.2.2	Propriedades da variância de variáveis aleatórias . . . . .	73
3.2.3	Teorema Central do Limite . . . . .	75
3.3	Execução e análise dos resultados . . . . .	79
3.3.1	Etapa 1: Entendendo o método . . . . .	79
3.3.2	Etapa 2: Aplicando o método . . . . .	84
3.4	Conclusão . . . . .	88
3.5	Complementar . . . . .	89
3.5.1	Área com integral . . . . .	89
3.5.2	Teorema Fundamental do Cálculo . . . . .	89
3.5.3	Volume com integral . . . . .	92
<b>4</b>	<b>Variáveis antropométricas</b>	<b>95</b>
4.1	O experimento . . . . .	95
4.2	Introdução teórica . . . . .	95
4.3	Execução . . . . .	100
4.4	Análise das Informações . . . . .	101
4.5	Conclusão . . . . .	110
4.5.1	Importando um conjunto de dados com números decimais do <i>Excel</i> para o <i>R</i> . . . . .	110
4.6	Elementos de regressão linear . . . . .	112
<b>5</b>	<b>Vida com Qualidade: Alimentação saudável e atividades físicas</b>	<b>117</b>
5.1	Sujeitos e métodos . . . . .	118
5.2	Informações prévias . . . . .	119
5.2.1	Obtenção dos dados . . . . .	123
5.3	Análise das informações . . . . .	125
5.3.1	Análise bivariada . . . . .	135
5.4	Conclusão . . . . .	145
	<b>Considerações Finais</b>	<b>147</b>
	<b>Referências Bibliográficas</b>	<b>151</b>

<b>Apêndice</b>	<b>155</b>
Termo de Consentimento Livre e Esclarecido . . . . .	155
Questionário aos Alunos . . . . .	157
<b>Anexos</b>	<b>159</b>
Termo de Ciência e Concordância do Coordenador Pedagógico/Diretor . .	159
Autorização do Diretor do IMECC . . . . .	161
Parecer consubstanciado do CEP . . . . .	163



# Agradecimentos

Primeiramente agradeço a Deus.

Agradeço também aos meus pais, Pedro e Arlinda, pelo apoio. Aos meus irmãos, Douglas e Aline, que sempre foram fraternos.

Gratifico à minha companheira, Jaqueline Rodrigues Teixeira, pela compreensão e demonstração de amor. Por estar sempre ao meu lado, seja nas dificuldades ou na alegria de cada vitória. Ao meu filho, Vitor Rodrigues Coura, por servir de inspiração. Agradeço também aos familiares de minha esposa, Vilma, Wilson, Daniele, Ana Maria, Juracy, Vanilda, Gabriel e Guilherme, por colaborarem para a execução deste.

Agradeço aos amigos e colegas de trabalho pela presença e afetividade. Cito também, os amigos do Mestrado Profissional em Rede Nacional (Profmat), especialmente, a turma de 2012 da Universidade Estadual de Campinas (Unicamp). Também gratifico a todos que não se incluem nestes grupos, mas são igualmente importantes.

Gratulo aos funcionários do IMECC (Instituto de Matemática, Estatística e Computação Científica), que sempre foram solícitos e eficientes nos serviços prestados. Ao professor Dr. Sérgio Antônio Tozoni, coordenador acadêmico do Profmat pela Unicamp, pelos esclarecimentos. Ao diretor do IMECC, professor Dr. Caio Negreiros, por autorizar prontamente a realização da pesquisa e demais professores do IMECC e PROFMAT que direta ou indiretamente contribuíram para a construção deste trabalho.

Agradeço cordialmente à amiga e orientadora, professora Dra. Laura Leticia Ramos Rifo, pela colaboração na elaboração, execução e conclusão desta dissertação, pois acreditou e motivou do começo ao desfecho, solucionando todas as dúvidas, dando confiança e credibilidade.

Aos idealizadores do Profmat, pelo desafio e oportunidade, para professores e alunos, de explorar e abordar novas práticas sobre conceitos e conteúdos de matemática e desenvolvimento e aperfeiçoamento no âmbito profissional.

Gratulo aos alunos e à equipe gestora da escola estadual Elvira de Pardo Mêo Muraro. Por fim, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Campinas, 15 de dezembro de 2014.

André da Silva Coura.



# Lista de Figuras

1.1	Alelos dominantes $\times$ alelos recessivos. . . . .	13
1.2	Dobrar a língua em forma de U. . . . .	15
1.3	Lóbulo da orelha solto e colado. . . . .	15
1.4	Questionário aplicado aos alunos e familiares. . . . .	16
1.5	Frequência absoluta do gênero com tamanho da amostra $n = 135$ . . .	19
1.6	Frequência relativa aluno $\times$ familiares. . . . .	20
1.7	Frequência relativa aluno $\times$ familiares. . . . .	22
1.8	Intervalo de confiança $(1 - \alpha)$ para $\theta$ . . . . .	37
1.9	Intervalo de confiança considerando desvios inteiros em uma $N(0, 1)$ . .	38
1.10	Região crítica para teste de hipóteses para $\theta$ . . . . .	42
2.1	Proporção de bolas azuis $n = 30$ bolas extraídas. . . . .	57
2.2	Proporção de bolas azuis $n = 30$ bolas extraídas. . . . .	58
2.3	Proporção de bolas azuis $n = 30$ bolas extraídas. . . . .	59
2.4	Jankenpon com 15 sinais. . . . .	61
3.1	Região $B$ contida na região $A$ . . . . .	71
3.2	Distância do Monte Caburaí até o Chuí. . . . .	76
3.3	Método da escada para mudança de unidade de comprimento. . . . .	77
3.4	Mapas quadriculados em quadrados $8cm \times 8cm$ . . . . .	79
3.5	Cálculo de área com $f \geq g$ . . . . .	90
3.6	Cálculo de área com sinal $f - g$ variando . . . . .	91
3.7	Sólido de revolução . . . . .	93
4.1	Pescoço $\times$ braço na origem e escala $5cm$ , $n = 32$ . . . . .	103
4.2	Pescoço $\times$ braço no ponto $P$ e escala $2cm$ , $n = 32$ . . . . .	103
4.3	Altura $\times$ circunferência do quadril, em $cm$ , com $n = 135$ . . . . .	104
4.4	Altura $\times$ circunferência do pescoço, em $cm$ , com $n = 135$ . . . . .	105
4.5	Altura $\times$ comprimento do braço, em $cm$ , com $n = 135$ . . . . .	106
4.6	Altura $\times$ comprimento da perna, em $cm$ , com $n = 135$ . . . . .	107

4.7	Altura $\times$ comprimento do dedo indicador, em <i>cm</i> , com $n = 135$ . . . .	108
4.8	Altura $\times$ número do calçado considerando o gênero, com $n = 135$ . . .	109
5.1	Pirâmide alimentar. . . . .	120
5.2	Pirâmide alimentar adaptada às necessidades da população brasileira.	123
5.3	Frequência das alturas no intervalo, $n = 68$ . . . . .	126
5.4	Frequência do peso no intervalo, $n = 68$ . . . . .	127
5.5	Frequência da circunferência do quadril no intervalo, $n = 68$ . . . . .	128
5.6	Diagrama de Venn para refeições intermediárias, $n = 68$ . . . . .	131
5.7	Dispersão da altura $\times$ peso, $n = 68$ . . . . .	136
5.8	Dispersão da altura $\times$ quadril, $n = 68$ . . . . .	138
5.9	Frequência relativa de nutrientes por refeição. . . . .	139
5.10	Tempo que dorme $\times$ peso. . . . .	141
5.11	Frequência do <i>IAC</i> para o gênero masculino, $n = 32$ . . . . .	143
5.12	Frequência do <i>IAC</i> para o gênero feminino, $n = 36$ . . . . .	144



# Lista de Tabelas

1.1	Tabela de contingência $m \times n$ na forma geral. . . . .	8
1.2	Tabela de contingências $2 \times 2$ (elementos com sub-índices). . . . .	9
1.3	Tabulação do experimento dobra a língua e coça a orelha. . . . .	16
1.4	Exemplo de combinação genética de zigotos. . . . .	17
1.5	Tabela de contingência: Dobra a língua x lobo colado. . . . .	23
1.6	Tabela de valores esperados: Dobra a língua x lobo colado. . . . .	24
1.7	Valores críticos para $z_{\frac{\alpha}{2}}$ associados ao grau de confiança amostral. . .	33
1.8	Erro de decisão em um teste estatístico de hipóteses. . . . .	41
1.9	Valores críticos de $z$ com lateralidade do teste. . . . .	44
2.1	Combinações para o evento $A$ , jogadores usando uma das mãos. . . .	52
2.2	Combinações para $A$ , um jogador usa uma mão e o outro usa ambas. . .	53
2.3	Tabela para preenchimento do resultado das partidas. . . . .	65
2.4	Tabela para preenchimento do resultado das partidas com três dados. . .	65
2.5	Frequência de vitória de cada dado nas disputas. . . . .	66
2.6	Frequência de vitória dos dados na disputa simultânea. . . . .	67
3.1	Tabela com $n$ pontos gerados de forma aleatória. . . . .	80
3.2	Frequência de pontos dentro do mapa $2cm \times 2cm$ , $n = 10$ . . . . .	81
3.3	Frequência de pontos dentro do mapa $2cm \times 2cm$ , $n = 50$ . . . . .	82
3.4	Frequência de pontos dentro do mapa $2cm \times 2cm$ , $n = 100$ . . . . .	83
3.5	Frequência de pontos dentro do mapa $1cm \times 1cm$ , $n = 10$ . . . . .	84
3.6	Frequência de pontos dentro do mapa $1cm \times 1cm$ , $n = 50$ . . . . .	85
3.7	Frequência de pontos dentro do mapa $1cm \times 1cm$ , $n = 100$ . . . . .	85
3.8	Frequência de pontos dentro do mapa $0,5cm \times 0,5cm$ , $n = 10$ . . . . .	86
3.9	Frequência de pontos dentro do mapa $0,5cm \times 0,5cm$ , $n = 50$ . . . . .	87
3.10	Frequência de pontos dentro do mapa $0,5cm \times 0,5cm$ , $n = 100$ . . . .	87
4.1	Tabela com referidas mensurações dos alunos. . . . .	101
4.2	Estatísticas descritivas das variáveis mensuradas, $n = 135$ . . . . .	101

5.1	Tabela de Classificação de <i>IMC</i> e <i>IAC</i> . . . . .	125
5.2	Estatísticas descritivas da variável altura, $n = 68$ . . . . .	126
5.3	Estatísticas descritivas da variável peso, $n = 68$ . . . . .	127
5.4	Estatísticas descritivas da variável quadril, $n = 68$ . . . . .	128
5.5	O que geralmente os alunos comem no café da manhã, $n = 68$ . . . . .	129
5.6	O que geralmente os alunos comem no almoço, $n = 68$ . . . . .	129
5.7	O que geralmente os alunos comem no jantar, $n = 68$ . . . . .	130
5.8	Alimentação entre as refeições principais, $n = 68$ . . . . .	130
5.9	Prática de atividades físicas no ambiente escolar, $n = 68$ . . . . .	132
5.10	Prática de atividades físicas fora do ambiente escolar, $n = 68$ . . . . .	132
5.11	Atividades físicas mais praticadas fora da escola, $n = 68$ . . . . .	132
5.12	Tempo de utilização de eletrônicos, $n = 68$ . . . . .	133
5.13	Alimentação utilizando aparelhos eletrônicos, $n = 68$ . . . . .	134
5.14	Possui horários regulares para dormir e acordar, $n = 68$ . . . . .	134
5.15	Estatísticas do tempo que dorme os alunos com horários regulares. . .	135
5.16	Estatísticas descritivas, por gênero, $n = 68$ . . . . .	137
5.17	Estatísticas descritivas, por gênero, da circunferência do quadril. . .	137
5.18	Tabela de nutrientes por refeição. . . . .	139
5.19	Refeição mais provida de nutrientes. . . . .	140
5.20	Índice de Adiposidade Corporal dos alunos, $n = 68$ . . . . .	142

# Prefácio

Esta dissertação do Mestrado Profissional de Matemática em Rede Nacional (PROFMAT) tem como um dos objetivos ajudar o professor de matemática do ensino fundamental e médio a desenvolver ferramentas de aplicação no ensino de estatística em sala de aula. Dentro desta ideia, fazemos uso do material criado por professores, estudantes e pesquisadores da Unicamp para o Projeto  $M^3$  - Matemática Multimídia para o ensino fundamental e médio. Em particular, dado nosso interesse na área de probabilidade e estatística, o material escolhido foram experimentos e softwares relacionados o uso de ferramentas estatísticas de coleta, representação e análise de dados.

Este trabalho consiste na tentativa de uma abordagem diferente e prática no ensino de conceitos da teoria estatística e de probabilidade para ensino fundamental e médio. Foram realizados quatro experimentos, 1. Dobra a Língua e Coça a Orelha, 2. Jankenpon, 3. Método de Monte Carlo para aproximação de áreas, 4. Variáveis Antropométricas e 5. uma pesquisa referente aos hábitos alimentares e prática de atividades físicas por parte dos alunos. Os experimentos citados e escolhidos foram considerados, após avaliação, condizentes com as competências e habilidades pertinentes ao currículo escolar para alunos do ensino fundamental, além de serem de fácil e rápida execução e envolverem conceitos dentro da realidade do público alvo. A resposta aos alunos, atinge em grande parte o propósito da dissertação, que é mostrar como a estatística (e matemática) está diretamente ligada à vida das pessoas e suas várias formas de tratar informações e dados e auxiliar nas decisões. Tais experimentos permitem a coleta dos dados brutos, a tabulação dos mesmos de forma a facilitar a compreensão, a criação de diversos tipos de gráficos e suas interpretações, as conclusões que a análise dos dados nos permite e a decisão mais conveniente dada a situação. Por fim, os experimentos possuem uma multidisciplinaridade escolar e uma forma lúdica de abordagem, possibilitando um aprendizado mais consistente e maior interesse dos alunos. Cada capítulo, exceto o último, está dividido em sinopse, introdução, motivação, objetivos, experimento/pesquisa, análise dos resultados e conclusão.

O intuito das propostas deste material inclui a participação ativa dos alunos em

todos os aspectos deste processo: técnicas de amostragem e coleta dos dados, representação em tabelas e gráficos adequados, e uma análise estatística básica, compatível com o conteúdo do ensino fundamental. Pretendemos, neste estudo, aplicar tais técnicas na temática particular de coletar, organizar e analisar dados de uma amostra. A abordagem escolhida é que os alunos colem e analisem dados retratando seu próprio perfil. Acreditamos que esta abordagem permite: criar maior proximidade do aluno com o ambiente de estudo, ampliar o grau de informação deste sobre estatística e tratamento da informação, motivar o interesse pela pesquisa científica, coleta e análise de dados.

A tomada de decisão deve, sempre que possível, ser amparada por métodos e ferramentas que possam contribuir para sua melhor qualidade e resultado. Nesse sentido, de forma sucinta, o presente trabalho aborda, situações em que as decisões devem ser tomadas a partir da análise da informação disponível. Desta forma, as análises estatísticas orientam quanto à melhor decisão a ser tomada. Trata-se, dessa maneira, de um trabalho que pode despertar interesse em outros profissionais da educação que procurem instruir de uma maneira mais prática, porém com alicerce na teoria estatística e probabilística, como tomar decisões ótimas, após analisar informações.

O Capítulo 1 trata do primeiro dos quatro experimentos, denominado “Dobra a língua e coça a orelha” [1]. Este experimento multidisciplinar, explora a observação de características físicas, propondo um estudo sobre a condição de hereditariedade e uma possível relação entre a ocorrência dessas características em um indivíduo. As características estudadas são: a capacidade de “dobrar a língua em forma de U” e ter o lobo (ou lóbulo) da orelha descolado. Neste experimento, os alunos analisam estas características em amigos e familiares. Desta forma, pretendemos estimular que o aluno assuma o papel de pesquisador e avaliador de aspectos físicos de pessoas próximas a ele e que a partir destas informações, investigue possíveis correlações e frequências que podem ser observadas na amostra. Este experimento une informações e conhecimentos das disciplinas de matemática e ciências (ensino fundamental) ou biologia (ensino médio), permitindo assim a criação de atividades paralelas, enriquecendo as aulas de ambas as disciplinas com discussões como, por exemplo, quais os aspectos genéticos envolvidos? Qual a característica dominante (mais frequente)? Quais as relações genéticas das características que serão observadas?

O Capítulo 2 apresenta o segundo experimento, denominado “Jankenpon” [2], conhecido como “Pedra-Papel-Tesoura”. É baseado em um jogo simples de origem japonesa que usa uma das mãos para representar cada um de 3 objetos: (pedra = mão fechada, papel = mão aberta, tesoura = o número 2, feito com os dedos indicador e médio). O papel vence a pedra embrulhando-a; a pedra, por sua vez,

vence a tesoura amassando-a (ou quebrando-a); e a tesoura vence o papel, cortando-o. Observe que neste jogo, não existe um objeto melhor, já que para cada objeto, os resultados possíveis são: vitória, empate ou derrota. O experimento possui três dados com numerações específicas, sendo que o dado  $A$  possui quatro faces com o número 1 e duas faces com o número 10, o dado  $B$  possui todas as faces com o número 4 e o dado  $C$  tem duas faces com o número 0 e quatro faces numeradas com 6. A regra é que o dado vencedor de cada partida é aquele que obtiver a maior pontuação na face voltada para cima. A pergunta inicial é: Qual dos dados é melhor em um confronto dois a dois? Formando duplas, cada aluno escolherá um dos três dados:  $A$ ,  $B$  e  $C$  e terá em posse uma tabela para anotarem o dado vencedor em cada partida. Ganha quem vencer o maior número de partidas no final. À primeira vista, o dado  $B$  tem maior chance de vitória que o dado  $A$ , e o dado  $A$  tem maior chance de vencer que o dado  $C$ . A primeira pergunta é se o dado  $B$  tem, portanto, maior chance de vitória que o dado  $C$ ? Sabemos que no jogo Jankenpon não há uma opção que é a melhor de todas, sempre existe outra que pode vencer. Neste experimento, à semelhança do jankenpon, ao escolher dois dados, a vitória de um não é certa, pois dependerá do valor obtido na face voltada para cima. Pela descrição dos dados, temos combinações de resultados que ora resulta na vitória de um ora na vitória do outro. Uma segunda pergunta é se algum dado tem mais chances de ganhar que os outros dois? Em outras palavras, existe o melhor dado em um ensaio envolvendo os três simultaneamente? Para responder essas questões, os alunos realizam jogadas e registram os resultados obtidos. Fizemos tal procedimento analisando três exemplos de jogo, para cada par possível de dados, com o intuito de incumbir o aluno de observar e registrar os resultados, de modo a investigar as probabilidades que serão analisadas. Os alunos jogaram o tempo necessário para que cada grupo tivesse uma quantidade suficiente de resultados. Estes resultados foram dispostos em tabelas e comparados com as probabilidades teoricamente determinadas.

No Capítulo 3 é apresentado o “método de Monte Carlo” [3] para aproximação de área de figuras com qualquer contorno. Nesta aplicação, estimamos a área do território brasileiro. O método possui diversas formas de aproximação. Para realizar uma destas formas, criamos uma malha quadriculada com os eixos  $X$  e  $Y$  possuindo valores inteiros entre 1 e  $b$ . Após construir essa malha, foram geradas diversos pontos com coordenadas  $(x, y)$  através de uma distribuição Uniforme Discreta  $\{1, \dots, b\}$  para o eixo das abscissas e para o eixo das ordenadas. A seguir, foram contados quantos desses pontos aleatórios ficaram dentro ou sobre o contorno do mapa do Brasil e calculada a razão desses em relação ao total de pontos gerados. Feito isto diversas vezes e calculada a proporção média de pontos no interior do mapa e com o uso da Lei dos Grandes Números, supomos que essa proporção média é uma boa aproximação

para a área do mapa do Brasil em relação à área da malha retangular.

O Capítulo 4 aborda o experimento denominado “Variáveis antropométricas” [4], sendo observadas e registradas diversas medidas corporais dos alunos, tais como: circunferência do pescoço, cós da calça, comprimento da perna e do braço, altura e tamanho do dedo indicador, todas em centímetros, além de informações como tamanho do calçado e gênero. Após a coleta dos dados, estes foram tabulados e realizada uma análise gráfica para observar se há ou não relação entre as partes mensuradas, ou seja, se o tamanho do dedo indicador está relacionado com a altura ou se a circunferência do quadril tem alguma relação com a circunferência do pescoço ou são apenas medidas sem relação alguma.

O capítulo 5 apresenta uma pesquisa feita avaliando hábitos alimentares, prática de atividades físicas, tempo e forma de uso de aparelhos eletrônicos, como celulares, computadores, tablets, televisores e afins. A pesquisa foi baseada na aplicação de um questionário para alunos do ensino fundamental (especificamente às classes nas quais lecionei no ano de 2013) da E.E. Elvira de Pardo Mêo Muraro, no período da tarde. Participaram da pesquisa alunos que tiveram o Termo de Consentimento Livre e Esclarecido (TCLE) assinado pelos responsáveis legais e quiseram responder ao questionário de treze questões. O TCLE foi encaminhado pela escola e o questionário aplicado aos alunos para analisar os aspectos de uma vida saudável. O questionário foi analisado respeitando o anonimato dos respondentes. Não houveram riscos, custos nem remuneração para o participante da pesquisa. Os alunos foram informados da pesquisa e dos seus objetivos a fim de responder o questionário da forma mais precisa possível. Para evitar que as respostas fossem lidas por colegas, os alunos responderam tal questionário em casa sob a supervisão dos responsáveis legais. Esperamos que os resultados desta pesquisa possam contribuir para uma mudança nos hábitos alimentares dos alunos observados, também estimular a prática de atividades físicas. Um segundo objetivo do estudo foi mostrar a aplicação de ferramentas de estatística descritiva em dados reais, no caso, concernentes aos próprios alunos, de modo a propiciar maior envolvimento com o conteúdo visto em sala de aula.

Finalmente, no capítulo 6, fazemos as considerações finais sobre as atividades, salientando observações feitas na aplicação dos experimentos e da pesquisa.

# Capítulo 1

## Dobra a língua e coça a orelha

### 1.1 O experimento

#### Sinopse

O aluno deve coletar informações em sua própria família sobre duas características genéticas, a capacidade de dobrar a língua em “U” e formato do lobo da orelha. A mesma informação deve ser coletada na sala de aula entre os alunos para que seja possível estabelecer comparações. Em seguida, com tabelas e gráficos adequados, os alunos responderam uma série de perguntas, analisando os dados coletados com ferramentas de matemática e estatística.

#### Conteúdos

Estatística, elementos de amostragem, elementos de genética, interpretação de gráficos.

#### Objetivos do experimento

1. Desenvolver a capacidade de extrair informações a partir de dados brutos;
2. Estimular o uso de tabelas e gráficos para analisar dados sobre características físicas;
3. Mostrar aplicabilidade de ferramentas matemáticas e estatísticas em outras disciplinas.

## 1.2 Introdução teórica

Podemos dizer que realizar uma pesquisa científica no campo da saúde acarreta um processo que gera conhecimento para todos os envolvidos. O pesquisador poderá confirmar suas suposições e ideias iniciais sobre a linha de pesquisa e/ou adquirir uma série de informações novas, enquanto o pesquisado compreenderá de forma qualitativa suas características referente ao estudo. Nesse âmbito, para termos eficácia, utilizamos procedimentos estatísticos para obtenção, organização, análise e interpretação das informações que devemos adquirir. Sucintamente, podemos ressaltar as seguintes etapas para a realização de uma pesquisa científica [8]:

1. Desenvolvimento de conjectura que direcione ao(s) objetivo(s) da pesquisa;
2. Obtenção de informações por amostragem <sup>1</sup>;
3. Pareamento das informações adquiridas e esperadas pela conjectura pressuposta;
4. Comprovação ou negação da conjectura. No primeiro caso, manutenção da conjectura; no segundo, desenvolvimento de nova conjectura.

Estas etapas fazem parte de uma sequência atrelada à necessidade de coletar dados para comprovar teorias e compreender, cientificamente, o comportamento destes dados. De uma forma mais direta, coletamos dados por amostragem para que, através da análise, possamos comprovar ou negar suposições e teorias pré-estabelecidas. Em caso de inveracidade destas pressuposições, remodelamos as conjecturas e testamos com nova aquisição de dados até encontrar argumentos válidos.

Precisamos salientar que a Estatística (e suas ferramentas) abrange mais do que a construção e interpretação de tabelas, gráficos e cálculo de medidas centrais e de dispersão. Estas características numéricas possuem como finalidade a aquisição de informações para tomarmos decisões de acordo com critérios científicos. Desta forma, a Estatística possui uma gama de técnicas para planejarmos, executarmos e obtermos dados através de experimentos, depois, critérios para organização, síntese, análise e interpretação destes para tirarmos conclusões. [6]

As técnicas clássicas foram construídas para respostas ótimas sob conjecturas feitas previamente. No entanto, estudos mostraram que tais técnicas nem sempre representam o que sugerem as conjecturas quando temos candidatos a *outliers* (valores fora do padrão apresentado pelos dados amostrais). O desenvolvimento de métodos

---

<sup>1</sup>Veja mais sobre este tópico na seção “Complementar”, subseção “Elementos de amostragem”



exploratórios robustos (controlam a variação dos dados amostrais, mesmo com a presença de *outliers*) aumentou de forma significativa a eficiência da análise estatística descritiva.

Bons profissionais do ramo se atentam, de forma minuciosa, as evidências que os dados oferecem antes de formular e testar conjecturas. Esta análise exploratória é feita através da utilização de inúmeros métodos para extrairmos a maior informação possível dos dados, para, a partir desta análise, usarmos a inferência estatística para construirmos modelos e conjecturas condizentes.

Podemos dividir, de modo sucinto, a Estatística em duas áreas <sup>2</sup>:

1. *Estatística descritiva* é utilizada para organizar, apresentar e sintetizar dados por meio da utilização de tabelas e gráficos;
2. *Estatística inferencial ou indutiva* que possui um conjunto de postulados para generalizar dados amostrais fazendo uso de técnicas de amostragem, medidas e testes estatísticos.

Vale ressaltar que variável qualitativa possui dois grupos, *ordinal e nominal*, onde: **Variável qualitativa ordinal** é aquela que na qual se observa uma característica (qualidade) que pode ser colocada em ordem (ou hierarquia).

**Exemplo:** grau de satisfação; grau de gravidade de uma doença; grau de escolaridade; mês do ano.

**Variável qualitativa nominal** é aquela que “mede” uma característica que não pode ser ordenada ou hierarquizada.

**Exemplo:** Gênero; cor do cabelo, da pele e dos olhos; região onde mora uma pessoa.

As variáveis quantitativas também se dividem em dois subgrupos, *discretas e contínuas*, definidas como:

**Variável quantitativa discreta** é aquela que quantifica (enumera, via contagem, geralmente) um conjunto de forma finita (ou não) com uma ordem natural.

**Exemplo:** População municipal; número de acidentes por ano; número de gols marcados por um time.

**Variável quantitativa contínua** é aquela obtida por medição, geralmente, que assume qualquer valor em um intervalo de número reais dependendo da precisão com a qual é medida.

**Exemplo:** Peso e altura de uma pessoa; consumo de energia em kWh; tempo de uma volta no autódromo de Interlagos.

---

<sup>2</sup>**Observação:** Alguns autores dividem a Estatística em três áreas, sendo elas, Estatística Descritiva, Probabilidade e Inferência Estatística.

Na vida prática ou em pesquisas é habitual o interesse em obtermos a informação se duas variáveis estão, e o quanto estão, relacionadas. Por exemplo, se a altura de uma pessoa está relacionada com o peso; se o tamanho do pé tem relação com a circunferência do quadril. Frequentemente queremos verificar se as variáveis são dependentes (relação de causa e efeito), se é possível caracterizarmos uma das variáveis através da informação da outra e calcular a correlação, ou dependência, entre elas. [8]

Para estabelecermos uma resposta para as proposições do parágrafo anterior recorreremos à análise bidimensional, onde as variáveis abordadas podem ser qualitativas ou quantitativas e há técnicas apropriadas para cada tipo. Para analisar variáveis qualitativas usamos tabelas de contingência  $m \times n$ , estatística Qui-Quadrado ( $\chi^2$ ) ([11] ou [21]) e coeficiente de contingência modificado ( $C_{mod}$ ) [14]. No caso de duas variáveis qualitativas usamos tabela de contingência  $2 \times 2$ . Para a análise exploratória de variáveis quantitativas usamos diagramas de dispersão, análise de correlação, regressão linear simples, coeficiente de determinação e análise de resíduos [6].

Adiante estudaremos a análise de variáveis qualitativas, o estudo feito com variáveis quantitativas será explorado no capítulo 4.

Quando o estudo é feito com variáveis qualitativas, levamos em conta a frequência absoluta de cada categoria. Para duas variáveis são consideradas as frequências absolutas dos cruzamentos das categorias. Estes cruzamentos possibilitam responder, por exemplo, quantos indivíduos com determinada característica da categoria  $A_i$  possuem uma característica da categoria  $B_j$ , conforme ilustrada na Tabela 1.1 que apresenta uma tabela de contingência na forma geral.

Tabela 1.1: Tabela de contingência  $m \times n$  na forma geral.

Variável A	Variável B				
Categorias	$B_1$	$B_2$	$\dots$	$B_n$	Total
$A_1$	$a_1b_1$	$a_1b_2$	$\dots$	$a_1b_n$	$a_1$
$A_2$	$a_2b_1$	$a_2b_2$	$\dots$	$a_2b_n$	$a_2$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$A_m$	$a_mb_1$	$a_mb_2$	$\dots$	$a_mb_n$	$a_m$
Total	$b_1$	$b_2$	$\dots$	$b_n$	$N$

Neste experimento, temos duas variáveis dicotômicas (com duas categorias: sim; não; ou valores: 0; 1), sendo uma delas ter ou não a capacidade de dobrar a língua em U e a outra possuir ou não o lóbulo da orelha colado. Portanto, para análise dos cruzamentos do par de características, os resultados são dispostos em uma tabela de

contingência, como exemplo abaixo (nesse caso,  $2 \times 2$ , pois temos duas variáveis com duas categorias cada).

Tabela 1.2: Tabela de contingências  $2 \times 2$  (elementos com sub-índices).

	Variável 2		
Variável 1	Sim (1)	Não (0)	Total
Sim (1)	$a_{11}$	$a_{12}$	$l_1$
Não (0)	$a_{21}$	$a_{22}$	$l_2$
Total	$c_1$	$c_2$	$N$

A Tabela 1.2 (de contingência) possibilita descrever o grau de associação entre as variáveis, avaliando, caso exista, a força desta associação assim como uma função indicadora (veremos sua construção e propriedades na subseção “Função indicadora” da seção “Complementar”). Além disto, se esta associação for forte é possível fazer predições dos valores de uma variável através da outra. Caso haja independência entre as variáveis ou associação fraca, as frequências registradas na tabela de contingências seguem o padrão apresentado nos totais marginais normalizados. Caso contrário, se as variáveis forem dependentes, as frequências seguem um padrão diferente.

Para avaliar se as frequências observadas na tabela de contingências seguem o padrão dos totais marginais normalizados, ou seja, das frequências esperadas, no caso de independência entre as variáveis, precisamos de uma estatística que leve em consideração a diferença entre tais frequências para assim compará-las. Daí, a primeira questão é como calcular os valores das frequências esperadas?

As frequências esperadas são calculadas, usando os dados da Tabela 1.2 (ou Tabela 1.1, para  $i \in \{1, \dots, m\}$  e  $j \in \{1, \dots, n\}$ ), pela seguinte expressão:

$$e_{ij} = \frac{l_i c_j}{N} \text{ para } i, j \in \{1, 2\}.$$

As frequências esperadas  $e_{ij}$  correspondem aos valores esperados para os respectivos  $a_{ij}$ , com a suposição de independência entre as variáveis (subseção “Teste de hipóteses” na seção “Complementar”). Com os valores de  $e_{ij}$ , usamos a estatística amostral Qui-quadrado ( $\chi_{ob}^2$ ), dada pela seguinte fórmula [11]:

$$\chi_{ob}^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(a_{ij} - e_{ij})^2}{e_{ij}}$$

sendo  $e_{ij}$  o valor esperado (médio) para o respectivo valor de  $a_{ij}$ .

Com o cálculo desta estatística, fazemos o teste  $\chi^2$  para avaliarmos, de forma quantitativa, a relação entre as frequências observadas e as frequências esperadas com a suposição de independência entre as variáveis. De maneira clara e objetiva, o valor  $\chi_{ob}^2$  revela com que certeza temos evidências para dizer que existe ou não uma relação de dependência entre as variáveis.

O teste  $\chi^2$  consiste em comparar o valor de  $\chi_{ob}^2$  amostral com o  $\chi_k^2$  tabelado para  $k = (m - 1)(n - 1)$  graus de liberdade <sup>3</sup> [11], sendo  $m$  o número de linhas e  $n$  o número de colunas da tabela de contingência. Para realizar o teste  $\chi^2$ , precisamos nos atentar as restrições de que,  $e_{ij} < 5$  não exceda 20% do total de caselas e nenhuma  $e_{ij} < 1$ . O critério de decisão sobre a estatística amostral  $\chi_{ob}^2$  independência das variáveis da tabela de contingência é o seguinte:

Se  $\chi_{ob}^2 < \chi_k^2$ , existe evidências na amostra de independência entre as variáveis.

Para uma tabela de contingência  $2 \times 2$ , temos 1 grau de liberdade. O valor tabelado da estatística  $\chi^2$  para 1 grau de liberdade é de 3,841 a um nível de significância  $\alpha = 0,05(5\%)$  (probabilidade de rejeitar a independência das variáveis sendo que estas são dependentes). Na subseção “Teste de hipóteses” da seção “Complementar” estes termos serão esclarecidos ao leitor. Isto significa que se  $\chi_{ob}^2 \geq 3,841$ , temos evidências para rejeitar a independência das variáveis.

O teste  $\chi^2$  permite avaliar se as variáveis possuem, ou não, uma relação de dependência, no entanto não mede, caso exista, o quão forte é esta relação.

Para determinar qual o grau da correlação entre as variáveis usamos a estatística  $C_c$ , denominada coeficiente de contingência e definida como [44]:

$$C_c = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

A estatística  $C_c$  só pode ser comparada a outro coeficiente de contingência calculado em tabelas de contingências de iguais dimensões, deve satisfazer as mesmas restrições para a realização do teste  $\chi^2$  e está sempre  $-1 < C_c < 1$  [44]. O valor máximo desta estatística  $C_c$  para uma tabela de contingências de dimensões  $m \times n$  é dado por [11]:

---

<sup>3</sup>**NOTA:** O grau de liberdade é o produto entre o número de linhas subtraído de 1 e o número de colunas subtraído de 1, isto porque a última (ou uma das outras) frequência esperada de cada linha (ou coluna), pode ser calculada usando a frequência observada menos a soma de todas as outras frequências esperadas (já calculadas). No caso de uma tabela de contingência  $2 \times 2$ , precisamos calcular apenas uma frequência esperada.

$$C_{max} = \sqrt{\frac{k-1}{k}}$$

sendo  $k$  o  $\min\{m, n\}$ .

O coeficiente de contingência modificado,  $C_{mod}$ , consiste na multiplicação de  $C_c$  pelo inverso do seu valor máximo, obtendo  $C_{mod}$  pela expressão abaixo [14]:

$$C_{mod} = \sqrt{\frac{\chi^2}{n + \chi^2}} \cdot \sqrt{\frac{k}{k-1}}$$

em que  $\chi^2$  é o valor da estatística  $\chi^2_{ob}$ ;

$n$  é o número total de observações;

$k$  é o mínimo entre o número de linhas e colunas.

O coeficiente de contingência modificado varia de,  $-1 \leq C_{mod} \leq 1$ , sendo zero (independência total) e 1 (dependência total ou associação perfeita) possuindo a classificação <sup>4</sup>, conforme o índice de relação entre as variáveis, mostrada abaixo [15].

- Se  $C_{mod} \in (0; 0,3)$ , a amostra revela evidências de uma correlação fraca entre as variáveis;
- Se  $C_{mod} \in [0,3; 0,7)$ , a amostra entrega evidências de uma correlação moderada entre as variáveis;
- Se  $C_{mod} \in [0,7; 1)$ , a amostra mostra evidências de uma correlação forte entre as variáveis.

Do ponto de vista de elementos de Genética, temos que os conceitos de dominância e recessividade estão entre os mais antigos da história desta [22]. Estes termos aparecem desde os trabalhos de Gregor Johann Mendel (entitulado “Pai da Genética”) e são utilizados até os dias atuais. No entanto, como nem todas as situações e suposições da análise genética contemplam estes conceitos (dominância e recessividade), ao longo do tempo foram feitos ajustes, surgindo os conceitos de *codominância* e *dominância incompleta*.

Na **genética clássica**, uma característica é dita *dominante* quando se manifesta fenotipicamente (característica observável, seja ela morfológica, fisiológica ou

---

<sup>4</sup>Existem outras classificações para coeficientes, por exemplo, Carla Santos [12] (2007), considera uma associação forte, se o coeficiente está entre 0,8 e 1; moderada se está entre 0,5 e 0,8; fraca entre 0,1 e 0,5; ínfima entre 0 e 0,1.

comportamental) no heterozigoto anulando a característica *recessiva*. O fenótipo *dominante* é aquele resultante do cruzamento de duas linhagens puras com fenótipos contrastantes (dominante e recessivo). Ou seja, neste caso, o fenótipo do heterozigoto é igual ao do indivíduo com genótipo homozigoto com dois alelos dominantes.

Esclarecendo, de forma sucinta, os termos *genes*, *alelos*, *homozigoto*, *homozigótico*, *heterozigoto*, *heterozigótico* e as *Leis de Gregor J. Mendel*, temos [16]:

1. **Genes** são seguimentos da molécula de DNA (Ácido Desoxirribonucleico) que contêm informações gênicas codificadas para a síntese de proteínas.
2. **Alelos** são formas alternativas da sequência nucleotídica de um gene, contêm as características genéticas. Em organismos diploides, como por exemplo, seres humanos, cada locus gênico (posição ocupada no cromossomo) possui um par de alelos;
3. **Herança monogênica** é a característica determinada por um par de genes, dependendo da relação entre os alelos;
4. **Herança oligogênica** é a característica determinada por poucos pares de genes;
5. **Herança poligênica** é a característica determinada por diversos pares de genes;
6. **Herança multifatorial** é a característica determinada por diversos pares de genes mais influência de fatores ambientais;
7. **Homozigoto** possui o par de alelos iguais;
8. **Homozigótico** é o indivíduo com o par de alelos iguais;
9. **Heterozigoto** possui o par de alelos diferentes;
10. **Heterozigótico** é o indivíduo com o par de alelos diferentes;
11. **Primeira Lei de Mendel (Monoibridismo ou Segregação dos fatores):**  
Cada característica é determinada pela união de um fator do gameta paterno com um fator do gameta materno, ou seja, o par de genes do filho será formado por uma informação do pai e outra da mãe;

12. **Segunda Lei de Mendel (Diíbridismo ou Segregação independente dos genes):** uma característica é herdada independentemente da existência de outra característica. Ou seja, a presença (ou ausência) de uma característica não está ligada a presença (ou ausência) de outra.

Portanto, na genética clássica, um heterozigoto com característica dominante resulta da combinação de dois homozigotos, um com característica dominante e outro com característica recessiva. Porém, existem inúmeras situações em que o cruzamento de homozigóticos (indivíduos com alelos iguais) resultam em indivíduos heterozigóticos com característica (fenótipo) intermediária (*dominância incompleta*) das apresentadas pelos dois indivíduos progenitores. Por fim, nos casos de codominância, observamos um fenótipo que contém simultaneamente, ambas as características dos pares de alelos. É importante compreender que alelos dominantes “não impedem nem inibem” os alelos recessivos de se manifestarem [22].

A Figura 1.1, retirada de <http://goo.gl/NwtmAu>, apresenta características de alelos dominantes e recessivos.

	ALELO DOMINANTE		ALELO RECESSIVO	
Nariz		Aquilino		Reto
Lobo da orelha		Destacado		Colado
Queixo		Com covinha		Sem covinha
Queixo		Prognato		Reto
Lábios		Grossos		Finos
Olhos		Escuros		Azuis
Cabelos		Escuros		Loiros
Língua		Capacidade de enrolar		Sem capacidade de enrolar
Calvície (nos homens)		Sim		Não
Dedo mindinho		Curvado		Reto
Curvatura do polegar		Curvado		Reto

Figura 1.1: Alelos dominantes × alelos recessivos.

## 1.3 Execução

Seria agradável mostrar a língua para as pessoas, analisar a orelha de alguém e formular hipóteses sobre a genética através de dados e cálculos estatísticos? Pensando nisso, a proposta deste experimento foi investigar certas características em amigos e familiares, coletando dados e testando hipóteses. Para isso, os alunos observaram duas características, a capacidade de dobrar a língua em forma de U e o formato do lobo (ou lóbulo) de orelha. A partir destas informações foram investigadas correlações e construídas tabelas de frequências observadas.

Este experimento aborda uma série de elementos que o tornam interdisciplinar. A relação com ciências (ensino fundamental) ou biologia (ensino médio) é direta, permitindo a criação de atividades paralelas com os professores destas disciplinas, possibilitando discussões, como por exemplo, sobre os aspectos genéticos envolvidos.

As duas variáveis (dobrar a língua em forma de U e lobo colado) que serão observadas são qualitativas e dicotômicas, o que justifica, de acordo com a introdução teórica, a análise de independência usando tabela de contingência  $2 \times 2$ , estatística Qui-Quadrado ( $\chi^2_1$ ) e coeficiente de contingência modificado ( $C_{mod}$ ).

Para sistematizar as informações produzidas foram utilizados gráficos e tabelas de frequências dos dados coletados pelos alunos e na sala de aula. Estes dados descrevem características genéticas que quase sempre passam despercebidas. Os materiais utilizados foram um questionário para que os alunos fizessem a coleta de dados de seus familiares e uma tabela para sistematização dos mesmos.

A realização do experimento consistiu em uma entrevista prévia feita em sala de aula com os alunos, determinando, após discussão crítica e consenso, quais características que os mesmos possuíam e as possíveis variações que estas poderiam ter, enfatizando a importância desta atenta observação para que os alunos aplicassem corretamente o questionário junto aos seus familiares. Foi discutido e definido um modelo padrão, após consenso, para ambas as características, da seguinte forma: dobrar a língua em forma de U seria caracterizada somente se a pessoa efetivamente conseguisse obter o formato em U sem a ajuda dos lábios (algo tentado por alguns alunos em sala de aula), salientando que a característica é em relação a mobilidade própria da língua. Já a definição da característica referente ao lóbulo da orelha foi da seguinte maneira, se houvesse uma concavidade (curva) onde a junção do lóbulo com a face tivesse um vinco seria considerado solto, caso contrário, seria considerado colado.



O levantamento estatístico foi feito sobre as seguintes características genéticas:

**(a) Capacidade de dobrar a língua em formato de U:** como visto na Figura 1.1, temos que esta característica é dominante em relação à incapacidade de fazê-lo. Ou seja, ao fazermos um cruzamento de um homozigótico que possui como fenótipo a capacidade de dobrar a língua em forma de U e um homozigótico com ausência desta característica, observaremos um heterozigótico com o fenótipo do primeiro homozigótico, assim como observado na Figura 1.2.

**(b) Formato do lobo da orelha:** São dois os fenótipos possíveis para o lobo auricular: ele pode ser solto ou colado, porém, nem sempre é fácil identificar essa diferença. Temos, pela Figura 1.1, que a característica lobo descolado (solto) é dominante, ou seja, ao fazermos um cruzamento de um homozigótico que possui o lobo solto e um homozigótico com o lobo colado, observaremos um heterozigótico com o lobo solto, assim como os observados, respectivamente, na Figura 1.3.



Figura 1.2: Dobrar a língua em forma de U.

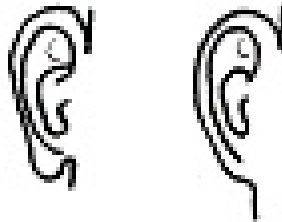


Figura 1.3: Lóbulo da orelha solto e colado.

Entregamos em sala de aula para aplicação em casa, o questionário abaixo, apresentado na Figura 1.4, (*retirado de <http://goo.gl/wDwt3c> - experimento “Dobra a língua e coça a orelha”*) para registrar as informações sobre o grau de parentesco do entrevistado com o aluno, o gênero, capacidade de dobrar a língua em U e característica do lóbulo da orelha.

**Questionário**

1. Grau de parentesco com o aluno (eu, mãe, pai, avô, avô, tio(a), primo(a) etc):

2. Sexo  
☐ *Masculino*  
☐ *Feminino*

3. Dobra a língua?  
☐ *Sim*  
☐ *Não*

4. Tem o lobo da orelha esquerda colado?  
☐ *Sim*  
☐ *Não*

Figura 1.4: Questionário aplicado aos alunos e familiares.

Os familiares entrevistados são consanguíneos com o aluno: mãe, pai, avós, irmãos, irmãos dos pais, irmãos dos avós, primo-irmãos. Cada aluno entrevistou 2 parentes, sendo um de cada gênero. Depois foi apresentada a Tabela 1.3 aos alunos, como forma de organização das informações coletadas.

Tabela 1.3: Tabulação do experimento dobra a língua e coça a orelha.

			Aluno				Parente masculino				Parente feminino			
Sexo			Língua em U?		O. C.?		Língua em U?		O. C.?		Língua em U?		O. C.?	
N	F	M	S	N	S	N	S	N	S	N	S	N	S	N
1														
2														
3														
...														
40														

**Legenda:** M = Masculino; F = Feminino; O.C. = Orelha colada; S = Sim; N = Não.

Previamente conversamos com o professor da disciplina de Ciências para que este, paralelamente, trabalhasse introdução à genética, esclarecendo conceitos como gene, alelo, herança monogênica, homozigótico, heterozigótico, as Leis de Mendel,

necessários para que entendessem do que se tratava o experimento e o porquê de determinadas características serem dominantes ou recessivas. Dessa forma, a referência foi o monge agostiniano Gregor Johann Mendel (1822-1884), considerado "o pai da Genética" por seus trabalhos de análise de características hereditárias.

Para ilustrar e esclarecer as proporções esperadas de cada característica e entender a formação do fenótipo fizemos tabelas considerando alelos homozigóticos e alelos heterozigóticos, denotando por uma letra maiúscula a característica dominante (mais forte) e por uma letra minúscula a característica recessiva, como apresentado na Tabela 1.4 (a) e (b).

Tabela 1.4: Exemplo de combinação genética de zigotos.

Genótipo	$a$	$a$
$A$	$Aa$	$Aa$
$A$	$Aa$	$Aa$

(a)

Genótipo	$A$	$a$
$A$	$Aa$	$Aa$
$a$	$Aa$	$aa$

(b)

Considerando a Tabela 1.4, sendo  $A$  a característica dominante e  $a$  a característica recessiva. Na tabela (a), teríamos 100% dos indivíduos dobrando a língua em U, na tabela (b), teríamos 75% dos indivíduos com a característica dominante.

Como já mencionado, cada aluno fez uma pesquisa individual determinando, conforme pré-estabelecido, as características de dois familiares consanguíneos, um de cada gênero. De posse dos dados, preenchemos a Tabela 1.3 para cada sala. Previamente ministramos aulas sobre organização de informação em tabelas, tabelas de frequências absoluta e relativa, com exemplos como tabela de campeonato de futebol com informações como pontos ganhos, jogos, vitórias, empates, derrotas, gols pró, gols contra, saldo de gols e aproveitamento de pontos (em %). Este exemplo, devido à maior familiaridade, foi mais apreciado, de maneira geral, pelos garotos, porém foi instrutivo às garotas por inserir conceitos e ideias sobre futebol.

Outro exemplo foi a tabulação de uma receita culinária, no qual as garotas possuíam muitas informações prévias, ajudando consideravelmente na construção da mesma. Além disto, acharam interessante a nova forma de organizar receitas, mencionando que iriam ensinar os familiares. Neste exemplo foi trabalhado conceitos de razão, proporção e regra de três simples, explorando as relações com questionamentos da quantidade necessária de ingredientes para fazer, por exemplo, um bolo para **sete** pessoas a partir de uma receita tabulada para **quatro** pessoas.

## 1.4 Análise das informações

Durante a aquisição das informações foram ministradas aulas sobre análise de dados por meio de construção de tabelas e gráficos, mostrando formas de organização para melhor visualização e entendimento de dados utilizando tabelas e gráficos. Foi ensinado a construção de tabelas usando frequência absoluta, relativa e acumulada, exemplificando como e quando é mais conveniente o uso destas. Também foi ensinado a construção de gráficos, suas características e como interpretá-los.

Com os alunos organizados na sala de informática, foi entregue, a cada um, uma folha com instruções sobre as atividades da aula e a Tabela 1.3 em branco, para eles perceberem que é necessário atenção para preencher uma tabela, observando quais informações serão depositadas nas linhas e colunas. Com uma planilha do programa *Excel* contendo as mesmas informações da Tabela 1.3, preenchemos, juntos, os dados dos cinco primeiros alunos, escalrecendo as dúvidas que, eventualmente, surgiram e foram compartilhadas. Cada aluno queria que fosse preenchido seu número, sem respeitar a ordem, por isso deixamos três alunos preencherem os campos de seu número. Eles demoraram mais tempo que os cinco primeiros alunos, assim perceberam que preencher na ordem numérica é mais eficiente do que de forma aleatória.

Após o preenchimento da Tabela 1.3 fizemos uma discussão, usando o *Excel* (um software do qual poucos alunos conheciam as funcionalidades), sobre quais tabelas e gráficos deveriam ser mais significativos para analisarmos os dados. Visando estimular discussões e ampliar a capacidade dos alunos de lidar com informações, fizemos perguntas como: Alguém observou algum dado curioso? É mais frequente menina ou menino com o lobo da orelha solto? Usar brincos influencia o fenótipo lobo solto?

Para as análises finais aqui colocadas utilizamos o software *R* [17], uma importante ferramenta para cálculos, análise e manipulação de dados, testes paramétricos e não paramétricos, modelagem linear e não linear, análises de séries temporais, de resíduos e de sobrevivência, fácil elaboração de gráficos, dentre outras funcionalidades. Pode ser instalado em sistemas operacionais como Windows, Linux e MacOS, além de ser gratuito e de uso livre. O leitor pode fazer o download do programa *R* em <http://cran.r-project.org>, que também oferece ao usuário informações, orientações e como baixar os pacotes. Os comandos e pacotes utilizados para a construção dos gráficos estarão disponibilizados ao final deste capítulo.

Participaram do experimento 135 alunos, sendo 62 meninas e 73 meninos. Para ilustrar esses dados, foi feito um gráfico de barras verticais, representado na Figura 1.5, levando em conta as frequências absolutas observadas.

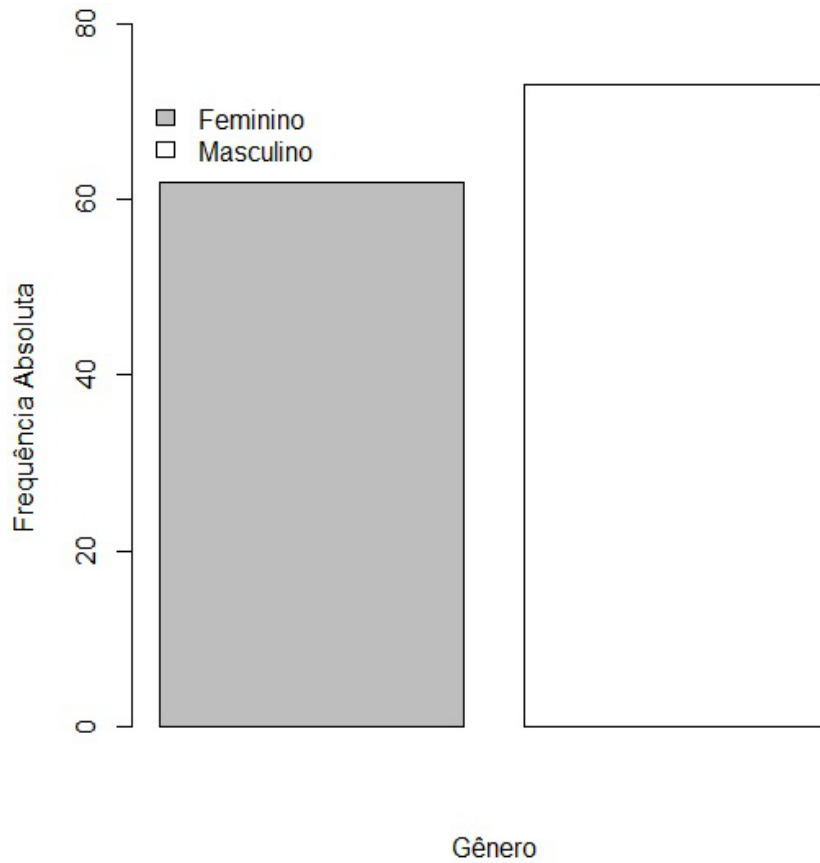


Figura 1.5: Frequência absoluta do gênero com tamanho da amostra  $n = 135$ .

Obtemos que 47 meninos e 41 meninas conseguem dobrar a língua em U. Para frisar a utilidade, calculamos a frequência relativa para cada grupo, obtendo que 64,38% dos meninos e 66,13% das meninas que participaram do estudo conseguem dobrar a língua em U.

Depois fizemos a Figura 1.6, que leva em consideração as frequências relativas da característica dobrar a língua em U considerando-a presente (ou ausente) no aluno e seus familiares, sendo que o par  $(C_M, C_F)$  diz respeito, respectivamente, a característica do familiar do gênero masculino e feminino. Ou seja, o par (Sim, Não) representa que o parente masculino é capaz de dobrar a língua e o parente feminino não. Vale ressaltar que observamos 64 pares (Sim, Sim); 19 pares (Sim, Não); 18 pares (Não, Sim) e 34 pares (Não, Não). A Figura 1.6 mostra a frequência relativa, em cada par, dos alunos que dobram (cinza) ou não a língua em U.

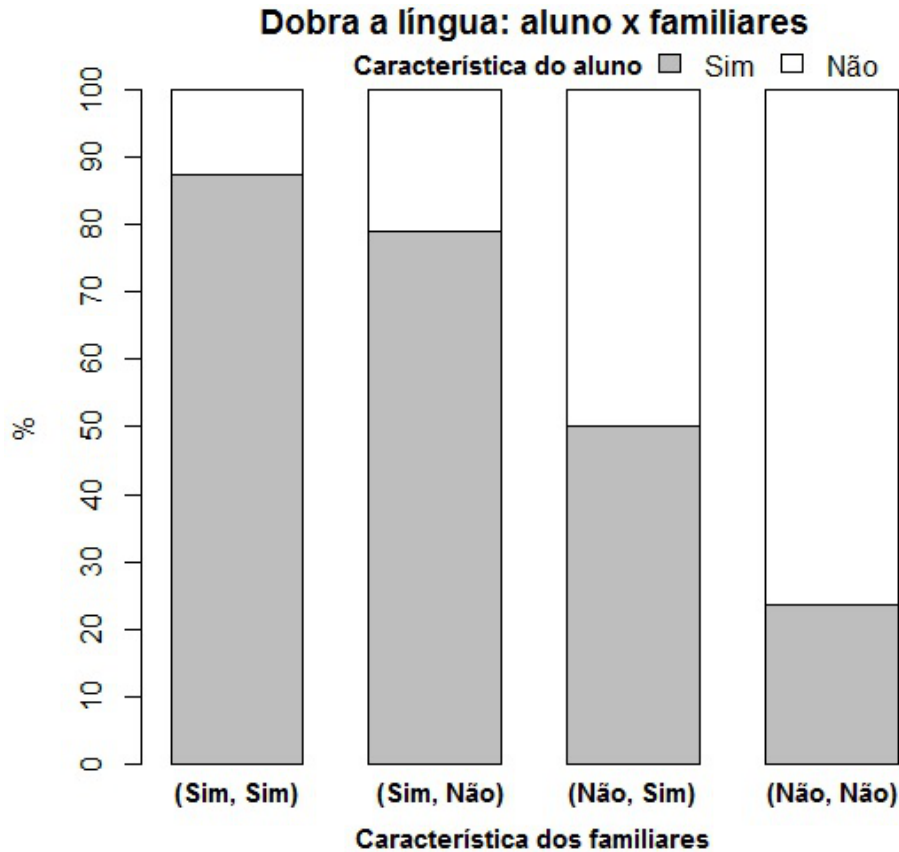


Figura 1.6: Frequência relativa aluno  $\times$  familiares.

Compreendendo o que significa os pares  $(C_M, C_F)$ , temos que a Figura 1.6 é repleta de informações. Por exemplo, temos que 83 (64+19) parentes do gênero masculino e 82 (64+18) parentes do gênero feminino são capazes de dobrar a língua em U, além disso, em 101 (64+19+18 ou 135-34) dos casos *pelo menos um* dos parentes é capaz de dobrar a língua.

Com estes dados, temos que 130 (47+83) dos 208 (73+135) homens, 62,5%, e 123 (41+82) das 197 (62+135) mulheres, 62,44%, são capazes de dobrar a língua. Sendo assim, a amostra entrega evidências de que a característica estudada não está associada ao gênero sexual. Além disso, como as frequências relativas foram superiores a 60%, **observação feita por aluno**, então podemos dizer que, *pelo menos*, 6 em cada 10 pessoas que participaram do estudo são capazes de dobrar a língua em forma de U (ou 60 em cada 100; 600 em 1000). Mais ainda, a probabilidade

de uma pessoa, selecionada ao acaso, conseguir dobrar a língua em U no conjunto amostrado é superior a 0,6 <sup>5</sup>.

Ainda na Figura 1.6, podemos observar que, a proporção de alunos que conseguem dobrar a língua em U diminui gradativamente conforme o par  $(C_M, C_F)$  de características de seus familiares. Para o par (Sim, Sim) temos que 87,5% dos alunos herdaram a característica; para o par (Sim, Não), a frequência relativa dos alunos que herdaram a característica é 78,95%; para o par (Não, Sim), em 50% os alunos herdaram a característica e para o par (Não, Não), menos de 25% dos alunos dobraram a língua. Observamos uma maior tendência do aluno herdar a característica do consanguíneo masculino nos casos onde apenas um dos familiares consegue dobrar a língua. Temos ainda que, dos 88 alunos que apresentaram a característica, 80 possui pelo menos um dos parentes com tal característica, o que significa que em 90,9% dos casos a característica dominante prevaleceu. A hereditariedade da característica dobrar a língua em U é destacada pelo gráfico.

Vale ressaltar que é possível que algum aluno consiga dobrar a língua, mas seus pais não, algo observado em 8 casos. Biólogos não conhecem com exatidão os genes que influenciam no surgimento dessa característica, sabendo apenas que é hereditária. Além disso, nem todos os alunos realizaram a pesquisa com os pais biológicos. Pedimos que, de preferência a entrevista fosse com os progenitores, mas na impossibilidade valeria qualquer parente consanguíneo, desde que fosse um homem e uma mulher.

A análise da característica lóbulo colado seguiu o mesmo padrão da característica dobrar a língua em forma de U. Dessa forma, fizemos a Figura 1.6, que leva em consideração as frequências relativas da característica lobo da orelha considerando-o solto (ou colado) no aluno e seus familiares, sendo que o par  $(C_M, C_F)$  diz respeito, respectivamente, a característica do familiar do gênero masculino e feminino. Ou seja, o par (Sim, Não) representa que o parente masculino possui o lobo colado e o parente feminino solto. Vale ressaltar que observamos 29 pares (Sim, Sim); 25 pares (Sim, Não); 11 pares (Não, Sim) e 70 pares (Não, Não). A Figura 1.7 mostra a frequência relativa, em cada par, dos alunos que possuem (cinza) ou não o lobo colado.

---

<sup>5</sup>Veremos Probabilidade no capítulo 2, porém esta observação se permite porque o experimento “Dobra a língua e coça a orelha” ocorreu após o experimento “Jankenpon”.

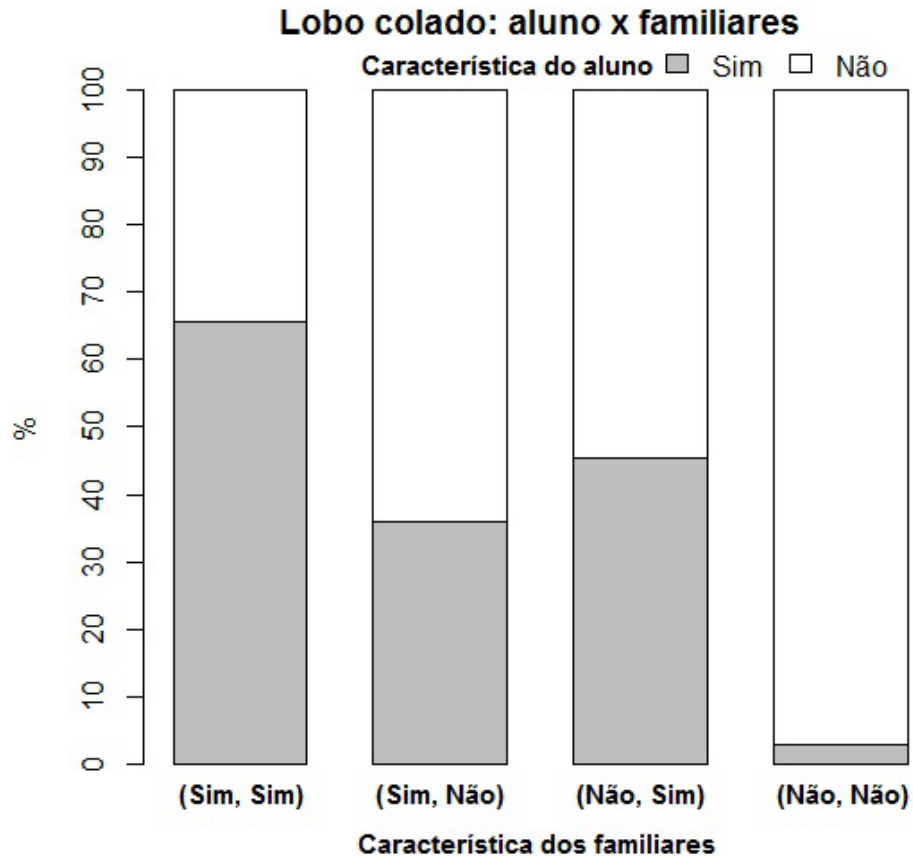


Figura 1.7: Frequência relativa aluno  $\times$  familiares.

É preciso ressaltar que a característica dominante neste caso é o lobo solto. Sendo assim, temos 81 (70+11) parentes do gênero masculino e 95 (70+25) parentes do gênero feminino possuem esta característica, além disso, em 106 (70+11+25 ou 135-29) dos casos *pelo menos* um dos parentes possui o lobo solto. Pelas informações dos familiares, poderíamos pensar que a característica lobo solto é mais comum em mulheres (*justamente o que pensaram os alunos, "justificando" que o uso de brincos ajuda na criação desse fenótipo*). No entanto, temos que 62 meninos e 38 meninas que possuem o lobo solto. Portanto, temos que 143 (62+81) dos 208 (73+135) homens, 68,75%, e 133 (38+95) das 197 (62+135) mulheres, 67,5%, possuem o lobo solto. Sendo assim, a amostra entrega evidências de que a característica estudada não está associada ao gênero sexual. Além disso, como as frequências relativas foram



superiores a 65%, então podemos dizer que, *pelo menos*, 65 em cada 100 pessoas que participaram do estudo possuem o lobo solto (ou 650 em cada 1000; 6500 em 10000). Mais ainda, a probabilidade de uma pessoa, selecionada ao acaso, possuir o lobo solto no conjunto amostrado é superior a 0,65. Pelos dados observados, temos evidências que o lobo solto é uma característica suscetível tanto na mulher quanto no homem. Mais uma vez, mostramos aos alunos, que a análise de dados deve ser prudente e minuciosa, sem conclusões preconcebidas, e provavelmente erradas, ao analisarmos apenas casos particulares.

Ainda na Figura 1.7, podemos observar que, a proporção de alunos que possuem o lobo solto aumenta gradativamente conforme o par  $(C_M, C_F)$  de características de seus familiares. Para o par (Sim, Sim), que significa (Colado, Colado) menos de 35% dos alunos possuem lobo solto; para o par (Sim, Não), 64% dos alunos herdam a característica do consanguíneo feminino; para o par (Não, Sim), temos que 54,5% dos alunos herdam a característica do consanguíneo masculino; para o par (Não, Não), em mais de 97% dos casos o aluno herda a característica. Observamos nos dados uma maior tendência do aluno herdar a característica do familiar feminino nos casos onde apenas um dos parentes possui o lobo solto. O gráfico evidencia a hereditariedade da característica.

Visando analisar se as características dobrar a língua em forma de U e lóbulo da orelha são dependentes, ou seja, se o fato dobrar a língua em U tem alguma relação com a característica do lóbulo da orelha, construímos a Tabela 1.5, considerando as características estudadas nos alunos e seus familiares.

Tabela 1.5: Tabela de contingência: Dobra a língua x lobo colado.

Dobra língua	Lobo colado		Total
	Sim	Não	
Sim	88	165	253
Não	41	111	152
Total	129	276	405

Pela análise da Tabela 1.5, notamos que 129 das 405 pessoas amostradas, quase 32%, possuem o lobo colado, enquanto 88 de 253 pessoas que dobram a língua em U, quase 35%, possuem o lobo colado. Ou seja, a frequência relativa das pessoas que possuem o lobo colado no total é próxima da frequência relativa das pessoas que possuem o lobo colado e conseguem dobrar a língua. Essa proximidade nas frequências relativas aponta para uma independência entre as características estudadas, porém para confirmar este pressuposto calculamos o valor da estatística  $\chi^2$  para os dados observados. Os valores esperados são mostrados na Tabela 1.6.

Tabela 1.6: Tabela de valores esperados: Dobra a língua x lobo colado.

	Lobo colado		
Dobra língua	Sim	Não	Total
Sim	80,58	172,42	253
Não	48,42	103,58	152
Total	129	276	405

Desta forma, temos que, usando a fórmula para calcular  $\chi_{ob}^2$ , página 9, obtemos

$$\chi_{ob}^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(a_{ij} - e_{ij})^2}{e_{ij}} = \frac{(88-80,58)^2}{80,58} + \frac{(165-172,42)^2}{172,42} + \frac{(41-48,42)^2}{48,42} + \frac{(111-103,58)^2}{103,58} = 2,695$$

Como  $\chi_{ob}^2 < 3,841$  ( $\chi_1^2$  tabelado), ao nível de significância  $\alpha = 5\%$ , a amostra não entrega evidências para rejeitarmos a independência entre as variáveis dobrar a língua em forma de U e lobo da orelha colado. Em outras palavras, temos evidências para concluir que estas variáveis não possuem dependência. Desta forma, a ocorrência da característica dobrar a língua em U está pouco relacionada com a ocorrência do lobo colado. Calculamos o coeficiente de contingência modificado,  $C_{mod}$ , para quantificar o grau dessa relação.

$$C_{modificado} = \sqrt{\frac{\chi^2}{n + \chi^2}} \cdot \sqrt{\frac{k}{k-1}} = \sqrt{\frac{2,695}{405 + 2,695}} \cdot \sqrt{2} \cong 0,115$$

Com esse valor para a estatística  $C_{mod}$  podemos dizer que, caso exista, a relação entre dobrar a língua em forma de U e possuir o lobo da orelha colado é **fraca**, usando a classificação da página 11, comprovando o fato de termos recusado que haja dependência entre tais características.

## 1.5 Conclusão

Para finalizar o experimento, apresentamos e discutimos as conclusões obtidas sobre as características estudadas em todas as salas que participaram da amostragem, evidenciando que ambas as características dobrar a língua em forma de U e lobo da orelha descolado são hereditárias e independentes.

O foco deste experimento foi o uso de tabelas e gráficos na sistematização dos dados, utilizando estes recursos para extrair as informações de interesse com mais

facilidade. Além disso, contribuir para o desenvolvimento dos alunos ao lidar com informações, mostrando que é preciso atenção e responsabilidade, desde a obtenção dos dados até se obter as conclusões. Também mostrou a aplicabilidade dos conceitos matemáticos e estatísticos no cotidiano e a associação entre as disciplinas escolares ministradas ao longo do ano letivo e da vida estudantil, nesse caso, matemática e ciências, no ensino fundamental (ou matemática e biologia, no ensino médio). Foram feitos diferentes tipos de gráficos na lousa e na sala de informática de forma que os alunos percebam a utilidade de cada um.

Para esta análise final foi utilizado, como já mencionado, o software gratuito *R* [17]. A seguir os comandos utilizados para criar os gráficos deste capítulo. Por questão de estética os comandos que não couberam em apenas uma linha serão apresentados de forma vertical (como o `barplot`).

**Gráfico** *Frequência absoluta dos alunos - Figura 1.5.*

```
Genero <- c("Feminino", "Masculino")
FA <- c(62, 73)
barplot(
  FA, main="Frequência absoluta dos alunos por gênero",
  xlab="Gênero", ylab="Frequência Absoluta", yaxp= c(0, 80, 4),
  space=0.2, width=1, col=c("gray", "white")
)
legend(
  "topleft", legend=c("Feminino", "Masculino"),
  fill=c("gray", "white"), bty="n"
)
```

**Gráfico** *Frequência relativa aluno  $\times$  familiares (Dobra a língua) - Figura 1.6.*

```
Dobra <- c("Sim", "Não")
Sim.Sim <- c(87.5, 12.5)
Sim.Não <- c(78.95, 21.05)
Não.Sim <- c(50.0, 50.0)
Não.Não <- c(23.53, 76.47)
graf<- data.frame(Sim.Sim, Sim.Não, Não.Sim, Não.Não)
barplot(
  as.matrix(graf), main="Dobra a língua: aluno x familiares",
  xlab="Característica dos familiares", ylab="%", yaxp=c(0, 100, 10),
  space=c(0.5, 0.5), col=c("gray", "white")
)
```

```

legend(
x=4, y=109, xpd=TRUE, ncol=2, legend=c("Sim", "Não"),
fill=c("gray", "white"), bty="n"
)

```

**Gráfico** *Frequência relativa aluno  $\times$  familiares (Lobo colado)* - Figura 1.7.

A Figura 1.7 foi construída de maneira similar à Figura 1.6, variando apenas as frequências relativas dos grupos e o nome do gráfico.

## 1.6 Complementar

### 1.6.1 Elementos de amostragem

Quando se trata de realizar uma pesquisa científica, grande parte das vezes, mesmo com todos os recursos computacionais, não é possível ou viável estudar todo o conjunto de dados no qual se tem interesse, por diversos motivos, por exemplo, escassez de tempo e custo elevado. Nestes casos, faz-se necessário o uso de técnicas e processos de amostragem aplicáveis nas mais diversas áreas do conhecimento no âmbito científico.

A teoria de amostragem é um dos instrumentos que possibilita a aquisição do conhecimento científico pois é um estudo das relações existentes entre uma população e a(s) amostra(s) dela extraídas [11]. Através da amostragem podemos avaliar parâmetros populacionais através de estatísticas amostrais, determinar diferenças entre duas (ou mais) amostras e se estas amostras são cientificamente significativas.

Os termos mais usuais em amostragem são [7]:

- **População** é o conjunto completo de todos os indivíduos ou objetos que possuam a(s) características que serão estudadas;
- **Unidade** é todo e qualquer indivíduo ou objeto que faça parte da população;
- **Amostra** é parte ou subconjunto da população que será estudado para obtermos informações sobre a população;
- **Variável** é a característica que será medida ou observada em cada unidade amostral.

É chamado de **CENSO** o estudo de uma (ou mais) característica(s) em todos os elementos da população. Outros fatores, além dos citados acima, que inviabilizam a realização de um censo são população muito grande (por exemplo, formigas)

ou infinitas, além da dificuldade para manter os dados atualizados. No entanto, o censo gera informações precisas e completas e é altamente indicado para populações pequenas.

A **Inferência Estatística** quando faz estudo a respeito de uma população utiliza técnicas de amostragem e teoria de probabilidade para inferir com determinada precisão a representatividade populacional adquirida com os dados amostrais.

Para que a amostra seja representativa é necessário **planejamento** da pesquisa (ou estudo) antes da obtenção propriamente dita dos dados. Primeiro, precisamos identificar e determinar quais as informações deverão ser recolhidas e o(s) instrumento(s) que viabilizará(ão) a coleta (questionário, teste, observação por determinado período, entre outros) e sua análise dos dados de acordo com o(s) instrumento(s) escolhido.

A coleta de dados deve ser sistematizada para assegurar a fidedignidade das informações que esta amostra produzirá. As etapas primordiais de uma amostragem são [7]:

1. **Identificar a população alvo** - É imprescindível determinar e identificar de forma clara e objetiva quais os indivíduos e características que serão estudados ou pesquisados; É preciso diferenciar **população alvo** de **população inquirida**; *População alvo* é composta por todos os elementos dos quais se deseja obter informações, já *população inquirida* é aquela composta pelos elementos efetivamente estudados ou pesquisados.

Por exemplo, se desejarmos fazer uma pesquisa sobre intenção de voto em uma eleição presidencial, teremos como **população alvo** pessoas entre 18 e 70 anos (que devem obrigatoriamente votar), porém o voto é facultativo para os analfabetos, maiores de 70 anos e jovens entre 16 e 18 anos (que podem ou não ser inclusos na população alvo, dependendo dos critérios do pesquisador); já a **população inquirida** será composta pelas pessoas que efetivamente opinarão, nesta pesquisa, sobre suas intenções de voto.

2. **Método de seleção do elemento que fará parte da amostra** - O principal objetivo da amostragem é uma representação significativa e fidedigna para que as estimativas e inferências sobre as características amostrais estejam relativamente próximas das características populacionais. Em outras palavras, que a amostra seja parte representativa da população. Existem dois métodos para realizar amostras de uma população [7]:

- **Amostragem não aleatória** - são métodos que usam caráter pragmático ou intuito, tem como pontos positivos promover um estudo rápido e

de baixo custo; os pontos negativos é que a inclusão de um elemento como parte da amostra parte de critérios subjetivos, geralmente opinião do pesquisador, isso faz com que exista elementos (que podem ser significativos) sem possibilidade de compor a amostra, além disso, estas amostras podem ser enviesadas (tortas, viciadas) e não representativas.

Os principais tipos de amostras não aleatórias são [7]:

- a **Intencional** possui elementos com características típicas ou representativas da população de acordo com o entendimento do pesquisador;
  - b **“Snowball”** é uma espécie de amostra intencional, na qual o pesquisador escolhe um grupo inicial, depois pede que estes indivíduos indiquem outros e assim sucessivamente. Desta forma, a amostra cresce similar a uma “bola de neve” (daí o nome), porém pode gerar resultados enviesados, já que a amostra, geralmente, terá pessoas com comportamentos e características específicas e similares;
  - c **Por quotas:** a população é dividida em grupos, conjuntos, categorias ou estratos e seleciona de modo não aleatório uma quantidade de elementos de cada categoria;
  - d **Por conveniência:** os elementos da amostra são determinados por conveniência ou facilidade. Por exemplo, fazer perguntas aos participantes de um programa de auditório ou torcedores de uma mesma equipe.
- **Amostragem aleatória** - são métodos onde todos os elementos da população possuem uma probabilidade diferente de zero de compor uma das possíveis amostras da população alvo, dessa forma, pode-se avaliar, de forma direta, as características populacionais a partir da amostra, pois podemos controlar e estimar os erros e evitar o viés mais provável em amostras não aleatórias. Os principais pontos contra esta amostragem consiste em obter a relação de todos os elementos pertencentes à população alvo e casos de não resposta (por exemplo, não encontrar uma pessoa em sua residência por motivos como mudança de endereço, viagem, horário de trabalho, entre outros). A seguir alguns métodos de amostragem aleatória [7]:
- a **Simples:** são selecionados  $n$  elementos de uma população de tamanho  $N$ ,  $n < N$ , onde todos os elementos da população possuem, inicialmente, a mesma probabilidade de fazer parte da amostra. A amostragem aleatória simples pode ser feita de duas maneiras.

*Com reposição:* Neste caso, a probabilidade de um elemento ser selecionado é sempre dada por  $1/N$ , pois o elemento retirado é recolocado novamente na população e, portanto, pode ser novamente escolhido. O número de amostras que podem retiradas neste método é dado por  $N^n$ , portanto uma determinada amostra tem probabilidade  $N^{-n}$  de ser realizada.

*Sem reposição:* Cada amostra aleatória simples de tamanho  $n$  tem probabilidade igual a  $1/\binom{N}{n}$  de ser selecionada, já que temos  $\binom{N}{n} = C_{N,n}$  amostras possíveis. Geralmente  $n < N/2$ , ou seja,  $n$  menor que a metade da população alvo  $N$ . Então, temos que

$$C_{N,n} = \binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N \times (N-1) \times \cdots \times (N-n+1)}{n!}$$

Neste método, a probabilidade de um elemento compor a amostra varia conforme as retiradas acontecem, se comportando da seguinte maneira:

Considere  $I_k$  o indivíduo escolhido na retirada  $k \in \{1, 2, \dots, n\}$ , então

$$\mathbb{P}(I_1) = \frac{1}{N}; \quad \mathbb{P}(I_2) = \frac{1}{N-1}; \quad \mathbb{P}(I_3) = \frac{1}{N-2};$$

Generalizando, temos  $\mathbb{P}(I_k) = \frac{1}{(N+1)-k}$ , para  $k \in \{1, 2, \dots, n\}$ .

Note que a probabilidade de um indivíduo não selecionado pertencer a amostra aumenta conforme as sucessivas retiradas.

Para populações infinitas ou suficientemente grandes ambos os métodos são equivalentes.

- b Casual sistemática** - é um método também conhecido por *quase-aleatório* pois as amostras não possuem a mesma probabilidade de ocorrer. Neste método determinamos o **rácio K** (relação, proporção, taxa) que é dado por  $K = N/n$ , depois escolhemos aleatoriamente um número no intervalo  $[1, K]$  como primeiro elemento da amostra. Adicionamos o rácio  $K$  a este primeiro elemento e encontramos o segundo, por adições sucessivas do rácio  $K$  obtemos os outros elementos que compõem a amostra.

**Exemplo:** Suponha que premiaremos 20 pessoas em um grupo de 1000 clientes cadastrados.

Temos  $n = 20$  e  $N = 1000$ , portanto  $K = N/n = 50$ .

Selecionamos aleatoriamente um número entre 1 e 50.

Suponhamos 27.

então

o primeiro premiado será a 27ª pessoa cadastrada;  
o segundo premiado será a  $(27 + 50)$  77ª pessoa cadastrada;  
o terceiro premiado será a  $(77 + 50)$  127ª pessoa cadastrada; e assim sucessivamente.

Generalizando, para este exemplo, temos um Progressão Aritmética (P.A.), sendo seu primeiro termo  $a_1$  um número aleatório entre 1 e 50, denotaremos por  $a_{1,50}$  (27, neste caso) com razão igual ao rácio  $K(= 50)$  e  $n(= 20)$ .

Portanto, o premiado  $P_n$  para  $n \in \{1, 2, \dots, 20\}$  será a pessoa com o cadastro numerado por:

$$P_n = 27 + (n - 1) \times 50;$$

na forma geral:  $P_n = a_{1,50} + (n - 1) \times K$ .

Observe que se, por exemplo,

$K = 2$ , então  $n = N/2$ , metade da população seria amostrada (50%);

$K = 10$ , então  $n = N/10$ , a décima parte da população seria amostrada (10%);

Ou seja, se  $K$  for divisor de 100, usamos  $(100/k)\%$  da população na amostra.

Caso  $K$  não seja um número inteiro, usamos o maior inteiro menor que  $K$ .

Matematicamente denotamos por  $K = \lfloor K \rfloor$ .

### **Exemplos:**

Se  $K = 16, 25$ , usamos  $K = 16$ ;

Se  $K = 19, 89$ , usamos  $K = 19$ .

- c Estratificada** - Neste método dividimos a população em grupos (chamados de estratos) mutuamente exclusivos (sem elementos em comum) e o mais homogêneo possível. Seleccionamos amostras aleatórias simples e independentes em cada estrato. Podemos retirar estas amostras nos estratos de duas maneiras:

*Proporcional:* O número de elementos  $n_{ek}$  amostrados em cada estrato deve ser proporcional ao número de elementos  $N_{ek}$  deste estrato na população. Por exemplo, se desejamos extrair uma amostra de tamanho 100 de dois estratos,  $e_1$  e  $e_2$ , onde na população total de 10000 elementos,  $N_{e1} = 6700$  e  $N_{e2} = 3300$ , então devemos ter  $n_{e1} = 67$  e  $n_{e2} = 33$ . Nesse caso, os resultados amostrais podem ser “diretamente” inferidos à população.

*Ponderada:* Se o número  $n$  de elementos amostrados em cada estrato



for o mesmo, então devemos ponderar os resultados de acordo com a proporção que estes estratos representam em relação a população alvo total.

É vantajoso estratificar uma população quando o comportamento dos indivíduos em relação a característica (ou variável) nos estratos são semelhantes, porém distintos entre os estratos. geralmente a estratificação é feita considerando outras características (ou variáveis) que estão relacionadas a característica (ou variável) de interesse. Por exemplo, suponha que trabalhe em uma seguradora. O valor final da apólice veicular depende do gênero do motorista, da idade, da localização geográfica de sua residência ou local onde o veículo ficará estacionado, da forma de uso (lazer, trabalho, entre outros), do modelo e ano do veículo, se possui equipamentos de segurança, quais equipamentos o veículo possui (vidro e trava elétrica, direção hidráulica, ar condicionado, computador de bordo, air bag, dentre outros), do tempo de habilitação, se possui multas no histórico, abrangência do seguro, estado civil, se a garagem é coberta, dentre outros fatores. Cada uma dessas variáveis foram estratificadas e estudadas para a criação de um modelo que determina o valor final a ser pago pela apólice.

- d **por Clusters** - Neste método de amostragem, a população geralmente está dividido por grupos (chamado de *clusters*, nesse caso). Assim selecionamos um dos grupos aleatoriamente e incluímos todos os elementos deste *cluster* na amostra.

**Exemplo:** Suponha que desejamos fazer uma pesquisa de opinião com os alunos de uma determinada escola (população alvo). Sabemos que esta possui uma listagem por série/ano e turma. Teremos uma amostragem por *clusters*, se selecionarmos, ao acaso, uma (ou mais) turma por série/ano e entrevistar todos os alunos desta(s) turma(s).

- e **Multietapas** - é similar a amostragem por *clusters*, porém selecionamos  $k$  elementos do grupo escolhido em vez de avaliar todos ( $n$ ),  $k < n$ . Este método pode ter mais de duas etapas se, de forma sucessiva, um grupo estiver dividido em subgrupos.

**Exemplo:** Desejamos avaliar de forma específica a opinião de estudantes de uma escola qualquer. Precisamos definir continente, país, região, estado, município, diretoria de ensino, escola, depois, número de alunos por turma (ou série/ano).

- f **Multifásicas** - Não podemos confundir este método com o anterior. Neste escolhemos, em uma primeira etapa,  $n$  unidades e avaliamos

de forma superficial estas unidades, depois, dependendo da resposta da primeira fase, selecionamos  $k$  unidades dentre as  $n$  escolhidas inicialmente para um estudo mais aprofundado. Este método pode ter mais de duas etapas dependendo do nível de informação que se deseja obter.

**Exemplo:** Na primeira fase de uma pesquisa, obtemos informações como nome, idade e estado civil;

Para uma segunda etapa, suponhamos que interessa apenas pessoas casadas com idade entre 25 e 50 anos, neste caso, continuamos o estudo com as pessoas entrevistadas na primeira fase que possuem tais características;

Assim segue o estudo, conforme a quantidade de informações que desejamos obter.

3. **Tamanho da amostra:** Vimos nos parágrafos anteriores que é preciso determinarmos qual a população que será estudada e quais os métodos para extrair e obter informações desta população. Outro fator importante em amostragem é definir o tamanho  $n$  desta amostra. O cálculo do tamanho da amostra depende de uma série de fatores como:

- a **Características da população alvo** - Uma população com comportamento homogêneo precisa de uma amostra menor do que uma população heterogênea, pois esta última possui maior variabilidade. Para determinar a variabilidade de uma população, podemos fazer uma amostra-piloto, buscar em referências bibliográficas ou utilizar a maior variabilidade possível.
- b **Características da pesquisa** - As mais comuns são estimar parâmetros populacionais como média e proporção ou comparar estes parâmetros em populações distintas.
- c **Grau de precisão** - Precisamos determinar qual o erro amostral que será tolerado. Usualmente os erros relativos são 1%, 5% e 10%.
- d **Tamanho da população** - Uma população pequena precisa de uma amostra menor do que uma população muito grande ou infinita.
- e **Tipo de amostragem** - Uma amostra aleatória simples **com reposição** tem tamanho diferente de uma **sem reposição** para uma população não muito grande, por exemplo.

- f Possíveis perdas ou não resposta** - Perder um elemento amostral (falecimento, por exemplo) ou ausência de resposta (mudança de endereço) deste ocasiona uma necessidade de repor ou compensar esta perda.

Veremos como calcular um tamanho de amostra mínimo para obtermos estimativas confiáveis para a média,  $\mu$ , e proporção,  $p$ , populacionais para populações infinitas ou muito grande e para populações finitas com  $N$  conhecido [18].

### **Tamanho da amostra para estimação da média populacional $\mu$ :**

Para determinarmos o tamanho da amostra, população infinita (ou muito grande), a ponto de conseguir uma boa estimativa amostral para a média populacional  $\mu$ , precisaremos conhecer  $\sigma$  (ou  $\sigma^2$ ), desvio padrão (ou variância) populacional, o valor crítico de  $z_{\frac{\alpha}{2}}$  referente a uma  $\sim N(0, 1)$  - *Normal padrão*, sendo  $\alpha$  o grau de confiança desejado, e o erro máximo  $\varepsilon$  que será tolerado. Os valores mais usuais de  $z_{\frac{\alpha}{2}}$  associados ao grau de confiança estão na Tabela 1.7.

Tabela 1.7: Valores críticos para  $z_{\frac{\alpha}{2}}$  associados ao grau de confiança amostral.

Confiança (%)	$\alpha$	$z_{\frac{\alpha}{2}}$
90	0,10	1,645
95	0,05	1,960
99	0,01	2,575

O tamanho da amostra é calculado pela seguinte expressão:

$$n = \left( \frac{\sigma \times z_{\alpha/2}}{\varepsilon} \right)^2 = \frac{\sigma^2 \times (z_{\alpha/2})^2}{\varepsilon^2}$$

### **Tamanho da amostra para estimação da proporção populacional $p$ :**

O tamanho da amostra, população infinita (ou muito grande), para conseguirmos uma boa estimativa amostral para a proporção populacional é similar a da média  $\mu$ , devemos apenas trocar  $\sigma^2$  por  $p \times q$ , sendo  $q = (1 - p)$ . Portanto, temos a seguinte expressão:

$$n = \frac{p \times q \times (z_{\alpha/2})^2}{\varepsilon^2}$$

Note que para determinar o tamanho da amostra precisamos conhecer  $\sigma$  (ou  $\sigma^2$ ) e  $p$ . Caso estes valores sejam desconhecidos, podemos fazer um estudo piloto e usarmos a estatística amostral  $s$  para estimar  $\sigma$  e  $\hat{p}$  para estimar  $p$ .

Podemos usar  $\sigma \approx \frac{A}{4}$ , sendo  $A = X_{(n)} - X_{(1)}$ , a diferença entre o máximo e o mínimo populacional.

Podemos usar  $p = q = 0,5$ , assim teremos:

$$n = \frac{0,25 \times (z_{\alpha/2})^2}{\varepsilon^2}$$

Vale salientar que, para calcular o desvio padrão amostral  $s$ , geralmente, calculamos primeiro a média amostral  $\bar{x}$  dada por [6]:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Depois, a variância amostral  $s^2$  é dada por:

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

Por fim, extraímos a raiz quadrada de  $s^2$  obtendo  $s$ .

Quando os dados estão agrupados (frequência absoluta  $f_i$  ou frequência relativa  $r_i$ ), para  $i \in \{1, 2, \dots, k\}$  e  $k$  o número de observações com frequências distintas, temos  $s^2$  dado por:

$$s^2 = \frac{1}{\sum_{i=1}^k f_i - 1} \sum_{i=1}^k (x_i - \bar{x})^2 \times f_i = \frac{1}{\sum_{i=1}^k r_i - 1} \sum_{i=1}^k (x_i - \bar{x})^2 \times r_i$$

sendo

$$\bar{x} = \frac{\sum_{i=1}^k x_i \times f_i}{\sum_{i=1}^k f_i}$$

Quando os dados estão agrupados em  $k$  intervalos de classe, temos  $\bar{x}$  e  $s^2$  dados por [10]:

$$\bar{x} = \frac{\sum_{i=1}^k x_{mi} \times f_i}{\sum_{i=1}^k f_i}; \quad \text{sendo } x_{mi} = \frac{l_s + l_i}{2} = \text{valor médio da classe.}$$

$$s^2 = \frac{1}{\sum_{i=1}^k f_i - 1} \sum_{i=1}^k (x_{mi} - \bar{x})^2 \times f_i = \frac{1}{\sum_{i=1}^k r_i - 1} \sum_{i=1}^k (x_{mi} - \bar{x})^2 \times r_i$$

Por fim, vale ressaltar que  $s^2$  pode ser calculado **sem** utilizar o valor da média pela seguinte fórmula [10]:

$$s^2 = \frac{\sum_{i=1}^n (x_i)^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2}{n - 1}$$

#### Tamanho da amostra para população finita com $N$ elementos:

Para uma população finita  $N$ , o tamanho da amostra para encontrarmos boas estimativas para média  $\mu$  e proporção  $p$  é dado por [18]:

$$\text{Média: } n = \frac{N \times \sigma^2 \times (z_{\alpha/2})^2}{(N - 1) \times \varepsilon^2 + \sigma^2 \times (z_{\alpha/2})^2}$$

$$\text{Proporção: } n = \frac{N \times p \times q \times (z_{\alpha/2})^2}{(N - 1) \times \varepsilon^2 + p \times q \times (z_{\alpha/2})^2}$$

Para finalizar, indicamos o site <http://goo.gl/D8rKl6> [19], elaborado por Glauber Eduardo de Oliveira Santos, que calcula o tamanho da amostra para população finita utilizando amostra aleatória simples para variáveis categóricas. A página possui todas as informações sobre como inserir os dados para obter o tamanho da amostra.

## 1.6.2 Teste de hipóteses

Testar uma hipótese significa empregar métodos e ferramentas de inferência estatística, construídos e embasados probabilisticamente, analisando conjecturas sobre parâmetros (ou características) desconhecidos (as) de uma (ou mais) população (ções). A teoria desenvolvida para construir e realizar por testes de hipóteses em inúmeros livros, por exemplo [6], passa primeiramente por teoria de amostragem (visto na subseção acima) e construção de intervalos de confiança. Portanto é importante compreender o que significa um intervalo de confiança para entender como é feito um teste de hipóteses.

### Intervalo de confiança

Geralmente construímos um intervalo de confiança  $I.C.(1 - \alpha) \times 100\%$  para uma estimativa amostral pontual  $\hat{\theta}$  de um parâmetro populacional  $\theta$  para que possamos garantir que  $\theta$  estará neste intervalo com o nível de confiança desejado.

Este intervalo de confiança é construído levando em consideração  $\hat{\theta}$ ,  $\sigma_{\hat{\theta}}$  e  $z_{\alpha/2}$ , para um nível de confiança  $(1 - \alpha)$  da seguinte forma.

É ideal, para que  $I.C.$  tenha maior confiabilidade, que  $\hat{\Theta}$  seja um **ENVUMV** (estimador não viciado uniformemente de mínima variância) [6] (ou pelo menos seja um estimador consistente), ou seja, que tenha as seguintes propriedades:

$$\mathbb{E}(\hat{\Theta}_n) = \theta; \quad \text{Var}(\hat{\Theta}_n) \rightarrow 0 \text{ para } n \text{ grande}, n \rightarrow \infty.$$

Estas propriedades garantem que  $\hat{\theta}$  é um bom estimador para  $\theta$ . Se  $\hat{\Theta}$  for um **ENVUMV**, a variância será a menor.

Com estas propriedades, construímos  $I.C.(1 - \alpha) \times 100\%$  da seguinte forma:

$$\theta \in (\hat{\theta} - z_{\alpha/2} \times \sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2} \times \sigma_{\hat{\theta}}) \text{ com probabilidade } (1 - \alpha).$$

ou podemos representar também como

$$\mathbb{P}(\hat{\theta} - z_{\alpha/2} \times \sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{\alpha/2} \times \sigma_{\hat{\theta}}) = (1 - \alpha).$$

Precisamos compreender o porquê aparece  $z_{\alpha/2}$  nesta construção.

Temos que, pelo *Teorema Central do Limite (TCL)* (Veja seção 3.2.3), a distribuição da variável aleatória  $\hat{\Theta} \sim N(\theta, \sigma_{\hat{\Theta}})$  para  $n$  grande ( $n > 30$ ).

Portanto, a variável aleatória  $Z$ , também por TCL, definida como:

$$Z = \frac{\hat{\Theta} - \theta}{\sigma_{\hat{\Theta}}} \sim N(0, 1).$$

O que nem sempre fica claro é porque  $Z$  tem esta distribuição.

Primeiro, precisamos compreender que, na verdade, escrevemos a variável aleatória  $\hat{\theta}$  em função de  $Z$  da seguinte forma:

$$\hat{\theta} = \theta + Z \times \sigma_{\hat{\theta}} \sim N(\theta, \sigma_{\hat{\theta}}); \theta \text{ constante}; \sigma_{\hat{\theta}} > 0 (\text{constante e conhecido})$$

Temos que <sup>6</sup>,

$$\mathbb{E}(\hat{\theta}) = \mathbb{E}(\theta + Z \times \sigma_{\hat{\theta}}) = \mathbb{E}(\theta) + \mathbb{E}(Z \times \sigma_{\hat{\theta}}) = \mathbb{E}(\theta) + \sigma_{\hat{\theta}} \times \mathbb{E}(Z) \Rightarrow$$

$$\mathbb{E}(\hat{\theta}) = \theta + \sigma_{\hat{\theta}} \times \mathbb{E}(Z) \xrightarrow{\mathbb{E}(\hat{\theta})=\theta} \theta = \theta + \sigma_{\hat{\theta}} \times \mathbb{E}(Z) \xrightarrow{\text{como } \sigma_{\hat{\theta}} > 0} \mathbf{E(Z)} = \mathbf{0} \blacksquare$$

$$\mathbb{V}ar(\hat{\theta}) = \mathbb{V}ar(\theta + Z \times \sigma_{\hat{\theta}}) = \mathbb{V}ar(Z \times \sigma_{\hat{\theta}}) \Rightarrow \mathbb{V}ar(\hat{\theta}) = \sigma_{\hat{\theta}}^2 \times \mathbb{V}ar(Z)$$

$$\xrightarrow{\mathbb{V}ar(\hat{\theta})=\sigma_{\hat{\theta}}^2} \sigma_{\hat{\theta}}^2 = \sigma_{\hat{\theta}}^2 \times \mathbb{V}ar(Z) \xrightarrow{\sigma_{\hat{\theta}} > 0} \mathbf{Var(Z)} = \mathbf{1} \blacksquare$$

Por estes resultados,  $Z \sim N(0, 1)$ . Desta forma, o valor  $z_{\alpha/2}$  determina quantos desvios padrões usaremos para criar um intervalo  $I.C.(1 - \alpha) \times 100\%$  para o parâmetro populacional.

A Figura 1.8, adaptada de <http://goo.gl/I5aj2N>, ilustra um  $I.C.(1 - \alpha) \times 100\%$  para o estimador  $\hat{\theta}$  com nível  $(1 - \alpha) \times 100\%$  de conter  $\theta$ .

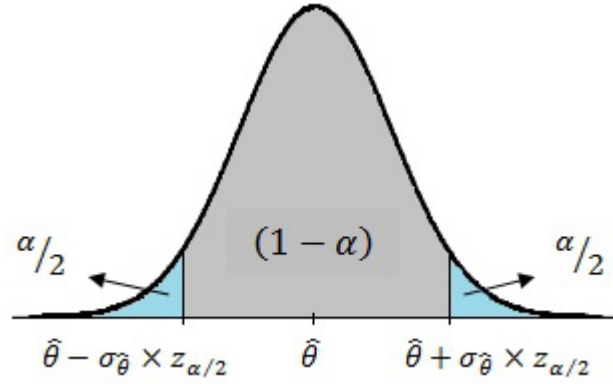


Figura 1.8: Intervalo de confiança  $(1 - \alpha)$  para  $\theta$ .

Já a figura 1.9, adaptada de <http://goo.gl/klva6M>, ajuda a compreender quais seriam os intervalos de confiança que teríamos se considerarmos  $\pm 1, \pm 2, \pm 3$  desvios padrão em uma  $\sim N(0, 1)$ .

---

<sup>6</sup>se necessário consulte as seções 3.2.1 (Propriedades de média) e 3.2.2 (Propriedades de variância)

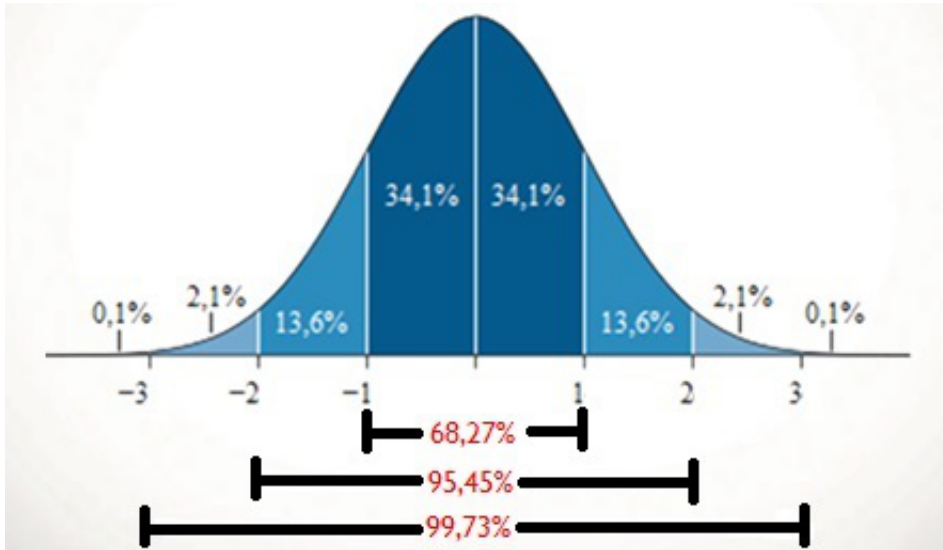


Figura 1.9: Intervalo de confiança considerando desvios inteiros em uma  $N(0, 1)$ .

Os intervalos de confiança mais comuns são para média  $\mu$ , para proporção  $p$  e soma ou diferença entre médias ou proporções [6] e os valores de confiança mais usuais estão dispostos na tabela 1.7.

**Intervalo de confiança para média populacional  $\mu$ :**

Estimador de  $\mu$  é  $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ ;  $\mathbb{E}(\bar{X}) = \mu$ ;  $\mathbb{V}ar(\bar{X}) = \frac{\sigma^2}{n}$



*População infinita ou finita com reposição*

$$\bar{x} \pm \frac{\sigma \times_{\alpha/2}}{\sqrt{n}}$$

*População finita  $N$  sem reposição*

$$\bar{x} \pm \frac{\sigma \times_{\alpha/2}}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

**Intervalo de confiança para proporção populacional  $p$ :**

Estimador de  $p$  é  $\bar{P} = \frac{\sum_{i=1}^n s_i}{n}$ ;  $\mathbb{E}(\bar{P}) = p$ ;  $\text{Var}(\bar{P}) = \frac{p \times q}{n}$ ;  $s_i = \# \text{sucessos}$

*População infinita ou finita com reposição*

$$\bar{p} \pm z_{\alpha/2} \times \sqrt{\frac{p \times (1-p)}{n}}$$

*População finita  $N$  sem reposição*

$$\bar{p} \pm z_{\alpha/2} \times \sqrt{\frac{p \times (1-p)}{n}} \times \sqrt{\frac{N-n}{N-1}}$$

**Intervalo de confiança para soma ou diferença de médias  $\mu_1$  e  $\mu_2$**

*Soma*

$$\bar{x}_1 + \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

*Diferença*

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

## Intervalo de confiança para soma ou diferença de proporções $p_1$ e $p_2$

*Soma*

$$\bar{p}_1 + \bar{p}_2 \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

*Diferença*

$$\bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

## Construindo testes de hipóteses

Um dos procedimentos mais utilizados pela *Inferência estatística* é o *teste de hipóteses* usando dados amostrais para decidir sobre o valor de um parâmetro ou distribuição probabilística de uma população. Os teste de hipóteses são chamados de **paramétricos** se forem usados para estimar um parâmetro de uma distribuição de probabilidade conhecida ou pré-estabelecida, já os testes **não paramétricos** excluem características prévias sobre a forma probabilística que a população está distribuída. Para saber mais consulte [6], [13] e [18].

Os termos de essencial entendimento para a realização de um teste de hipóteses são:

- a Hipótese estatística** é uma afirmação ou conjectura sobre um parâmetro populacional  $\theta$  ou de uma variável aleatória  $X$  com uma distribuição de probabilidades conhecida ou não.
- b Hipótese Nula  $H_0$**  é a conjectura com uma regra de decisão pré-estabelecida que será testada através de uma amostra.
- c Hipótese Alternativa  $H_A$  ou  $H_1$**  é a conjectura, geralmente, contrária a  $H_0$ .
- d Estatística do teste** é a expressão cujo valor, calculado a partir da amostra, usamos para tomar a decisão sob  $H_0$ .
- e Nível de significância do teste** é a probabilidade  $\alpha$  de cometer o erro tipo I, em outras palavras, significa com qual probabilidade o teste de hipóteses nos fará não aceitar  $H_0$  quando esta for verdade.
- f Nível de confiança do teste** é a probabilidade  $(1 - \alpha) \times 100\%$ , ou seja, é a confiança de  $(1 - \alpha) \times 100\%$  que a decisão acerca de  $H_0$  será correta.

### g Falsos positivos - Erros tipo I e tipo II

Tabela 1.8: Erro de decisão em um teste estatístico de hipóteses.

Realidade $H_0$	Decisão	
	Aceitar $H_0$	Rejeitar $H_0$
Verdadeira	Correta [ $\mathbb{P} = (1 - \alpha)$ ]	Erro tipo I [ $\mathbb{P} = \alpha$ ]
Falsa	Erro tipo II [ $\mathbb{P} = \beta$ ]	Correta [ $\mathbb{P} = (1 - \beta)$ ]

A Tabela 1.8 apresenta o tipo de decisão e o erro associado, no caso de uma decisão equivocada.

Em termos de probabilidade definimos os erros com a seguinte lógica:

$$\mathbb{P}(\text{Erro tipo I}) = \mathbb{P}(\text{Rejeitar } H_0 \mid H_0 \text{ verdadeira}) = \alpha$$

$$\mathbb{P}(\text{Erro tipo II}) = \mathbb{P}(\text{Aceitar } H_0 \mid H_0 \text{ falsa}) = \beta$$

Geralmente, modelamos um teste de hipóteses de maneira que cometer o **erro do tipo I** seja mais grave do que o **erro tipo II**. Ou seja, rejeitarmos  $H_0$  sendo esta verdadeira seria pior do que aceitá-la sendo falsa.

**h Teste unilaterais e bilaterais** - Existem três tipos de testes paramétricos que podem ser realizados, denominados como unilateral à esquerda, *unilateral à direita* e *bilateral*.

1. *Unilateral à esquerda*

$$\begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \xrightarrow{\text{Testada como}} \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

2. *Unilateral à direita*

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \xrightarrow{\text{Testada como}} \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

3. *Bilateral*

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \xrightarrow{\text{Testada como}} \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \text{ ou } \theta > \theta_0 \end{cases}$$

- i **Região crítica** ( $R_C$ ) é determinada pelos valores da variável aleatória ou estatística do teste para os quais  $H_0$  não será aceita. A Figura 1.10, adaptada de <http://goo.gl/wzuzQe>, nos auxilia a compreender qual a região crítica de cada teste de hipóteses, respectivamente, da esquerda à direita, apresentados no item h.

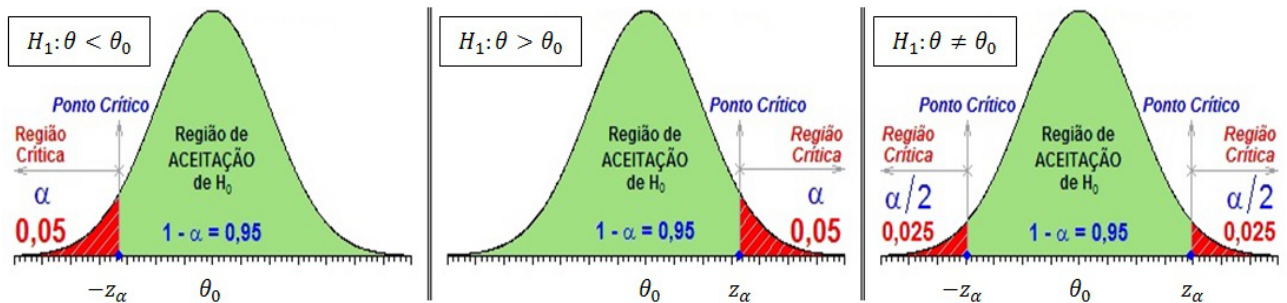


Figura 1.10: Região crítica para teste de hipóteses para  $\theta$ .

Note que, um teste de hipóteses bilateral coincide com um intervalo de confiança  $(1 - \alpha) \times 100\%$  para  $\theta$ .

As etapas para construirmos um teste de hipóteses paramétrico são:

1. Determinar o parâmetro populacional que será estimado;
2. Estabelecer  $H_0$  e  $H_1$  de maneira apropriada;
3. Escolher o nível de significância desejado para o teste;
4. Estabelecer a estatística de teste;
5. Determinar a região crítica  $R_C$ ;
6. Obter a(s) estimativas, a partir da amostra, para o(s) parâmetros populacionais;
7. Calcular o valor da estatística do teste;
8. Decidir acerca da aceitação ou rejeição de  $H_0$  no contexto estudado;

A seguir apresentaremos os testes de hipóteses paramétricos mais comuns e as respectivas estatística de teste mais apropriada [6].

**Média populacional**  $H_0 : \mu = \mu_0$

*Estatística de teste:*

$$z = \frac{\sqrt{n} \times (\bar{x} - \mu_0)}{\sigma}$$

Em todos os casos, comparamos o valor de  $z$  com o valor de  $z_\alpha$  ou  $z_{\alpha/2}$ .

**Proporção populacional**  $H_0 : p = p_0$

*Estatística de teste:*

$$z = \frac{\sqrt{n} \times (\bar{p} - p_0)}{\sqrt{p_0 \times (1 - p_0)}}$$

**Diferença entre médias**  $H_0 : \mu_1 = \mu_2$

*Estatística de teste:*

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{(\bar{x}_1 - \bar{x}_2)}}$$

**Diferença entre proporções**  $H_0 : p_1 = p_2$

*Estatística de teste:*

$$z = \frac{p_1 - p_2}{\sigma_{(p_1 - p_2)}}$$

É preciso salientar que o valor de  $z_\alpha < 0$  para um teste unilateral à esquerda e  $z_\alpha > 0$  para um teste unilateral à direita e para um teste bilateral, precisamos comparar  $z$  com  $\pm z_{\alpha/2}$

Se o desvio padrão populacional for desconhecido usamos o desvio padrão amostral como estimativa.

Vale ressaltar que, em todos os testes acima, estamos supondo  $n$  suficientemente grande ( $n > 30$ ) e, usando o TCL, supondo  $Z \sim N(0, 1)$ . Caso  $n < 30$ , usamos a distribuição “t-Student” com  $(n - 1)$  graus de liberdade [11], sendo que a estatística  $t$ , que usa o desvio padrão amostral  $s$ , é dada por:

$$t = \frac{\sqrt{n} \times (\hat{\theta} - \theta)}{s_{\hat{\theta}}}$$

Nestes casos, comparamos o valor de  $t$  com o valor de  $t_\alpha$  (*unilateral*) ou  $t_{\alpha/2}$  (*bilateral*) de uma distribuição t-Student com  $(n - 1)$  graus de liberdade<sup>7</sup>. Diferentemente de  $z_\alpha$  ou  $z_{\alpha/2}$ , o valor de  $t_\alpha$  ou  $t_{\alpha/2}$  depende do número de graus de liberdade.

---

<sup>7</sup>Temos  $(n - 1)$  graus de liberdade porque estimamos o desvio padrão populacional usando  $s$

A Tabela 1.9 mostra com quais os valores críticos de  $z_\alpha$  ou  $z_{\alpha/2}$  comparamos  $z$  conforme a lateralidade do teste de hipóteses.

Tabela 1.9: Valores críticos de  $z$  com lateralidade do teste.

	Nível de significância $\alpha$			
Valor crítico de $z$ dado teste	0,10	0,05	0,01	0,005
Unilateral à esquerda	-1,28	-1,645	-2,33	-2,575
Unilateral à direita	1,28	1,645	2,33	2,575
Teste bilateral	$\pm 1,645$	$\pm 1,96$	$\pm 2,575$	$\pm 3,08$

### Teste $\chi^2$ (Qui quadrado)

O teste  $\chi^2$  é usado para determinar (ou avaliar) a discrepância entre frequências observadas e esperadas ou para avaliar o quão próximo estão as distribuições teórica e empírica (obtida pela amostra), este último é chamado de teste de aderência. Um exemplo de teste de aderência é avaliar se a variável aleatória  $X$  igual a proporção de itens defeituosos em uma determinada linha de produção é descrito por uma distribuição Binomial com parâmetros  $n$  e  $p$ , ou seja, se  $X \sim b(n, p)$ .

No caso do experimento apresentado no capítulo 1, usamos a estatística de teste  $\chi^2$ , definida na página 9, para realizar um teste de hipóteses com as seguintes considerações:

$H_0 : a_{ij} = e_{ij}$ , ou seja, dobrar a língua em U e lobo da orelha colado são independentes;

$H_1 : a_{ij} \neq e_{ij}$ , ou seja, dobrar a língua em U e lobo da orelha colado são dependentes.

A estatística de teste tem distribuição assintótica  $\chi_\alpha^2$  com 1 grau de liberdade, pois precisamos encontrar apenas uma das frequências esperadas, por exemplo  $e_{11}$ , e calcular as outras por diferença da seguinte maneira:

$$e_{11} = \frac{l_1 \times c_1}{n}, \text{ então } e_{12} = l_1 - e_{11}; e_{21} = c_1 - e_{11}; e_{22} = l_2 - e_{21} = c_2 - e_{12}$$

O nível de significância  $\alpha$  foi 0,05, o valor crítico de  $\chi_\alpha^2 = 3,841$ . Um detalhe que devemos comentar é que, por  $H_1$  o teste é bilateral, porém a distribuição  $\chi_1^2$  não está definida à esquerda (veja mais em [6]), por isso comparamos apenas com um valor.

O número de graus de liberdade  $k$  de uma tabela de contingência  $m \times n$  é dado por  $k = (m - 1) \times (n - 1)$  pelo mesmo motivo [11], ou seja, precisamos calcular  $(m - 1)$  frequências esperadas em cada linha e  $(n - 1)$  em cada coluna e as restantes podem ser calculadas por diferenças similares às apresentadas acima. Neste caso, comparamos o valor da estatística de teste  $\chi^2$  com o valor de uma  $\chi_{(m-1) \times (n-1)}^2$  com nível de significância  $\alpha$ . O valor para  $\chi^2$  varia conforme o número de graus de

liberdade.

### 1.6.3 Função indicadora

Nos dias atuais, ouvimos muito sobre indicadores econômicos, educacionais, taxas de crescimento, inflação, índices, coeficientes, entre outros termos. Estes termos são muito importantes, pois através do estudo e interpretação destes é possível fazer previsões sobre características econômicas, analisar tendências, determinar o desenvolvimento, enfim, ajudam a compreender como se comportam as características que estudam <sup>8</sup>. Índices, taxas e indicadores são gerados a partir da chamada função indicadora. No caso de duas variáveis, por exemplo, indica se estas estão ou não relacionados e quantificam esta relação. Veremos como é definida uma função indicadora e suas propriedades.

Uma função indicadora de uma variável (ou característica)  $X$  em  $\Omega$  é denotado por  $\mathcal{I}_X : \Omega \rightarrow \{0, 1\}$ , sendo  $\Omega$  o conjunto universo (domínio) e  $\{0, 1\}$  os possíveis valores da função  $\mathcal{I}_X$ .

Formalmente [20],

$$\mathcal{I}_X = \begin{cases} 1 & \text{se } \omega \in X \\ 0 & \text{se } \omega \notin X \end{cases} = \begin{cases} 1 & \text{se } x \in X \\ 0 & \text{se } x \in X^c = \Omega - X \end{cases}$$

Perceba que podemos descrever  $X$  através de  $\mathcal{I}_X$  e denotamos por  $X = \{\omega : \mathcal{I}_X(\omega) = 1\}$ . Vejamos alguns exemplos particulares.

se  $\mathcal{I}_X(\omega) = 1, \forall \omega \in \Omega \Rightarrow X = \Omega$ ;

se  $\mathcal{I}_X(\omega) = 0, \forall \omega \in \Omega \Rightarrow X = \emptyset$ ;

se  $\mathcal{I}_X(\omega_0) = 1 \Rightarrow X = \{\omega_0\}$ ;

Generalizando temos que,

$$\mathcal{I}_X(\omega_i) = 1, i \in \{1, 2, \dots, n\} \Rightarrow X = \{\omega_1, \omega_2, \dots, \omega_n\}$$

Propriedades da função indicadora Se necessário revise operações com conjuntos  $\mathcal{I}_X$  e  $\mathcal{I}_Y$  para variáveis (conjuntos)  $X, Y$  em  $\Omega$  [9] :

1. Se  $Y \subseteq X$ , então  $\mathcal{I}_Y \leq \mathcal{I}_X$ ;
2.  $\mathcal{I}_\Omega = 1$  e  $\mathcal{I}_\emptyset = 0$ ;
3.  $\mathcal{I}_{X^c} = 1 - \mathcal{I}_X$ ;

---

<sup>8</sup>Veja mais sobre a construção de coeficientes em [44]

4.  $\mathcal{I}_{X \cap Y} = \min(\mathcal{I}_X, \mathcal{I}_Y) = \mathcal{I}_X \mathcal{I}_Y$ ;
5.  $\mathcal{I}_{X \cup Y} = \max(\mathcal{I}_X, \mathcal{I}_Y) = \mathcal{I}_X + \mathcal{I}_Y - \mathcal{I}_{X \cap Y}$
6.  $\mathcal{I}_{X - Y} = \max(\mathcal{I}_X - \mathcal{I}_Y, 0) = \mathcal{I}_X \mathcal{I}_{Y^c}$

A compreensão destas propriedades possibilita a construção de inúmeros argumentos sobre a relação entre  $X$  e  $Y$ . Vejamos alguns casos particulares.

Se  $\mathcal{I}_{X \cap Y} = 0$ , então  $X \cap Y = \emptyset$  e  $\mathcal{I}_{X \cup Y} = \max(\mathcal{I}_X, \mathcal{I}_Y) = \mathcal{I}_X + \mathcal{I}_Y$ .

Se  $\mathcal{I}_{X \cap Y} = 1$ , então  $X = Y$  e  $\mathcal{I}_{X \cup Y} = \mathcal{I}_X = \mathcal{I}_Y$ .

Se  $\mathcal{I}_X - \mathcal{I}_Y = 1$ , então  $X = \Omega$  e  $Y = \emptyset$ .

Se  $\mathcal{I}_{X \cap Y} = \mathcal{I}_{X \cup Y}$ , então  $X = Y$ .

A distribuição de Bernoulli, por exemplo, é definida em um espaço amostral binário  $\{0, 1\}$ , sendo que 0 é atribuído a uma resposta negativa (fracasso) e 1 se a resposta é positiva (sucesso). A distribuição de Bernoulli é associada, para efeito de compreensão, ao resultado do lançamento de uma moeda sendo que 1 é atribuído se o resultado for cara e 0 se for coroa ou vice e versa.

Cada ensaio de uma variável aleatória  $X$  com distribuição de Bernoulli é definido como

$$\mathcal{I}_X(S) \begin{cases} x = 1 & \text{se } \textit{Sucesso} \\ x = 0 & \text{se } \textit{Fracasso} \end{cases}$$

Seja  $X$  uma variável aleatória com distribuição de Bernoulli, sua função de probabilidade  $\mathbb{P}(X = x)$ , sendo  $p$  a probabilidade de sucesso é dada por

$$\mathbb{P}(X = x) = p^x \times (1 - p)^{1-x}$$

Outro exemplo de função indicadora é o coeficiente de correlação entre duas variáveis, valor que determina o grau de associação ou relação destas variáveis.

Para entendermos o que isto significa, a interpretação geométrica é valorosa. O coeficiente de correlação entre duas variáveis  $X$  e  $Y$ , com uma série de  $n$  valores,  $X = (x_1, x_2, \dots, x_n)$ ;  $Y = (y_1, y_2, \dots, y_n)$ ;  $X \neq Y$ ; é o cosseno do ângulo  $\alpha$  ( $\cos(\alpha)$ ) entre estes dois vetores (produto escalar normalizado entre  $X$  e  $Y$ ) no espaço com  $n$  dimensões [6]<sup>9</sup>. Portanto temos que se o valor do coeficiente de correlação é -1, ou seja,  $\cos \alpha = -1$ ; ( $\alpha = 180^\circ$ ) estas variáveis são colineares, porém seguem sentidos opostos; se o coeficiente de correlação é 1, ( $\alpha = 0^\circ$ ), estas variáveis são colineares e seguem no mesmo; se  $\cos \alpha = 0$ ; ( $\alpha = 90^\circ$ ), e estes vetores são ortogonais, ou seja, uma variável não sofre variação alguma se alterarmos a outra, em simples palavras, estas variáveis não possuem relação alguma (são totalmente independentes).

---

<sup>9</sup>Veremos com mais detalhes na seção 4.2



# Capítulo 2

## Jankenpon

### 2.1 O Experimento

#### Sinopse

Neste jogo, cada adversário escolhe e lança um dos três dados  $A$ ,  $B$  e  $C$ , numerados conforme segue.

- Dado  $A$  : 1, 1, 1, 1, 10, 10;
- Dado  $B$  : 4, 4, 4, 4, 4, 4;
- Dado  $C$  : 0, 0, 6, 6, 6, 6.

Ganha quem obtiver a maior pontuação no lançamento, considerando a face voltada para cima.

A pergunta inicial é qual dado tem maior chance de ganhar o jogo?

#### Conteúdos

Probabilidade: Independência, representação gráfica, eventos equiprováveis.

#### Objetivos do experimento

1. Aplicar conceitos de probabilidade;
2. Desenvolver habilidades para tratamento de informações através de gráficos;
3. Desenvolver raciocínio lógico, olhar crítico e formular estratégias a partir do conhecimento matemático.

## 2.2 Introdução teórica

A **probabilidade** é um tópico importante na Matemática, que lida com o conceito de incerteza. Utilizamos a probabilidade para modelar experimentos ou observações cujo resultado não conhecemos com precisão antes da realização do experimento. Para sua formalização, utilizamos o conceito de *experimento aleatório*, que é qualquer experimento ou observação cujo resultado não é conhecido exatamente antes de realizar o experimento [24].

Exemplos de *experimento aleatório* do nosso cotidiano:

1. A média bimestral de um aluno;
2. A soma dos dedos visíveis em uma disputa de “Par ou Ímpar”;
3. A quantidade de pacotes de salgadinhos premiados com brinquedo, brinde ou figurinha, em 100 pacotes;

É chamado de *espaço amostral* e denotado por  $\Omega$ , o conjunto de todos os resultados possíveis de um *experimento aleatório*. No primeiro exemplo, o *espaço amostral* poderia ser  $\Omega_1 = \{0, 1, 2, \dots, 10\}$ , no segundo exemplo,  $\Omega_2 = \{0, 1, 2, 3, \dots, 18, 19, 20\}$  e no último, podemos considerar a quantidade de pacotes premiados, como  $\Omega_3 = \{0, 1, 2, 3, 4, \dots, 100\}$ .

É definido como *evento* qualquer subconjunto do *espaço amostral* e denotado por letras maiúsculas do alfabeto. Dessa forma, seguindo os exemplos anteriores, podemos dizer que os eventos  $A, B$  e  $C$  fazem parte dos respectivos  $\Omega_1, \Omega_2, \Omega_3$ , sendo

$A = \text{“A média bimestral do aluno é maior que 5”} = \{6, 7, 8, 9, 10\}$ ;  $B = \text{“A soma dos dedos é par”} = \{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$ ;  $C = \text{“mais de 1 pacote é premiado”} = \{2, 3, 4, \dots, 99, 100\} = \{0, 1\}^C$ .

Dois eventos de um experimento aleatório são chamados *mutuamente exclusivos* se não podem ocorrer simultaneamente. Para o exemplo 1, dois eventos *mutuamente exclusivos*, são  $A = \text{“Média do aluno é maior do que 8”}$  e  $B = \text{“Nota do aluno é vermelha”}$ , ou seja, se o aluno tem média maior do que 8, sua nota é azul, portanto não pode ser vermelha. Para o exemplo 2, seja o evento  $C = \text{“Soma é par e maior do que 2”}$  e  $D = \text{“Soma é um número primo”}$ , então  $C$  e  $D$  são eventos *mutuamente exclusivos*. Para o exemplo 3, dois eventos *mutuamente exclusivos* são  $E = \text{“Pelo menos cinco pacotes contêm um brinquedo, brinde ou figurinha”}$  e  $F = \text{“Nenhum dos 100 pacotes possui prêmio”}$ .

O conceito de **probabilidade** está diretamente relacionado com o grau de informação <sup>1</sup>. Neste sentido, temos que a probabilidade de um evento representa a chance deste ocorrer baseado em uma informação conhecida pelo observador. Definimos probabilidade de maneira que satisfaça as seguintes condições [9].

**Definição 1.** Dizemos que  $\mathbb{P}$  define uma função de probabilidade em um conjunto de eventos associados ao espaço amostral  $\Omega$  se:

1.  $0 \leq \mathbb{P}(A) \leq 1$ , para todo evento  $A \in \Omega$ ;
2.  $\mathbb{P}(\Omega) = 1$ ;
3.  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ , se  $A$  e  $B$  são eventos mutuamente exclusivos.

A probabilidade de um evento pode ser alterada ao obtermos uma nova informação a seu respeito, este é o conceito de probabilidade condicional apresentado a seguir.

**Definição 2.** Sejam  $A$  e  $B$  eventos de um espaço amostral  $\Omega$ , com  $\mathbb{P}(B) > 0$ , então a probabilidade condicional de  $A$  dado  $B$  é:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Apresentamos um exemplo para ilustrar probabilidade condicional.

Suponha que precisamos descobrir uma senha de 3 dígitos distintos. Sem informação prévia, a probabilidade de acertar na primeira tentativa é 1 em 720 ( $10 \times 9 \times 8$ ). Se soubermos que todos os dígitos são *números primos*, as chances mudam para 1 em 24 ( $4 \times 3 \times 2$ ), pois os números primos entre 0 e 9 são 2, 3, 5 e 7. Se uma segunda informação for que os números estão em *ordem crescente*, então a probabilidade se torna 1/4, porque o primeiro dígito (da esquerda à direita) só poderia ser 2 (com 3 possíveis respostas: 235; 237 ou 257) ou 3 (apenas 357). De posse de uma terceira informação, na qual todos os *dígitos são ímpares*, descobrimos a senha com probabilidade 1 (357). Esse exemplo serviu, essencialmente, para os alunos entenderem o conceito de probabilidade condicional, porém mostrou que informações compreendidas ajudam a resolver problemas de matemática. **Observação:** Uma terceira informação que solucionaria o problema é que os números formam uma Progressão Aritmética (*para alunos do ensino médio*).

Se uma informação nova não altera a probabilidade de outro evento acontecer dizemos que os eventos são ***independentes***.

---

<sup>1</sup>Se pensarmos em probabilidade como função indicadora.

**Definição 3.** Sejam  $A$  e  $B$  dois eventos definidos em um espaço amostral  $\Omega$ , com  $\mathbb{P}(B) > 0$ , então  $A$  e  $B$  são independentes se:

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Ou ainda, usando a Definição 2, em que  $\mathbb{P}(B) \geq 0$ ,  $A$  e  $B$  são independentes se:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

Se os eventos não obedecem estas igualdades da Definição 3, dizemos que são eventos dependentes.

## 2.3 Aplicação

### 2.3.1 Par ou ímpar?

Um exemplo trabalhado em sala de aula foi qual a melhor escolha em uma disputa? Par ou ímpar? Discussão que se iniciou quando um dos alunos disse que sempre escolhe **par** porque a chance de ganhar é maior. Alguns alunos disseram que as chances são iguais, então a proposta foi averiguar se a afirmação do aluno era verdadeira. A primeira discussão veio sobre qual a característica apresentada pelo número 0. Muitos alunos diziam que o **zero** (0) é **neutro**. Dada essa afirmação, a primeira pergunta foi: “O que significa *neutro*?” Sem resposta, foi explicado que “*neutro*” (na verdade, **elemento neutro**) depende do contexto.

Por exemplo, o **zero** é o **elemento neutro** da **adição**, pois  $a+0 = a = 0+a$ , para qualquer valor de  $a$ . Já na **subtração**, o **zero** não tem o mesmo comportamento, pois  $a-0 = a \neq 0-a = -a$ , sendo  $-a$  o **simétrico** de  $a$ .

Por sua vez, a **multiplicação** também possui **elemento neutro**, sendo ele, o número 1 pois  $1.a = a = a.1$ . Na **divisão** o número 1 não é *elemento neutro*, pois  $a/1 = a \neq 1/a = a^{-1}$ , sendo  $a^{-1}$  o **inverso** de  $a$ .

Para sanar eventuais dúvidas, foram ministradas aulas sobre as propriedades das quatro operações básicas e as retas numéricas (como a reta  $\mathbb{R}$ ), exemplificando todas, inclusive a confusão que existia entre os termos: **simétrico** e **inverso**. Foi esclarecido que o primeiro, usamos na operação da *adição* e tem o significado de **oposto** em relação ao **zero**, na reta dos números Inteiros  $\mathbb{Z}$  (ou dos números Reais  $\mathbb{R}$ ), o segundo termo está relacionado a *multiplicação* e temos que “o inverso de  $a$  é o número 1 dividido por  $a$ , se  $a \neq 0$  e denotado por  $a^{-1}$ ”. Geralmente, é apresentado como propriedade de *potenciação*.

Para explicar e eliminar as dúvidas sobre a paridade do **zero**, discutimos que nos números Naturais ( $\mathbb{N}$ ) e Inteiros ( $\mathbb{Z}$ ), temos que um número é **par** ou **ímpar**.

Portanto, haveria sempre um número **par** com sucessor (ou antecessor) **ímpar** e vice e versa. Também foi discutido critérios de paridade de um número, ensinando que um número é **par**, se o resto da divisão por 2 for igual a *zero*, e **ímpar** se o resto for igual a 1. Mais ainda, para saber se um número é par ou ímpar, basta olhar o algarismo da unidade. Se for igual a  $\{0, 2, 4, 6, 8\}$  é **par**, se for  $\{1, 3, 5, 7, 9\}$  é **ímpar**. Por fim, foi desenhada a reta  $\mathbb{Z}$  e preenchida usando o conhecimento adquirido. Primeiro preenchemos os números inteiros, como os números -1 e 1 eram **ímpares**, então o *zero* só poderia ser **par**.

O próximo passo, foi definir as condições em que seria disputado o par ou ímpar? A maioria dos alunos disputam com apenas uma das mãos, eventualmente, um ou outro, apresenta as duas mãos. Desta forma, foi perguntado de quantas formas este poderia ser jogado, ou melhor, quais as situações que poderiam ou não influenciar os resultados.

Foram caracterizadas três maneiras de disputas:

1. Jogadores usam uma das mãos;
2. Um jogador usa uma mão e outro usa as duas;
3. Jogadores usam as duas mãos.

### 2.3.2 Jogadores usam uma das mãos

Neste momento falamos de eventos *equiprováveis*, considerando que a probabilidade de indicar qualquer quantidade com os dedos era a mesma, eliminando possibilidades citadas como “Sempre coloco o 2 porque é o meu número da sorte”. Primeiro foi descrito todo possível evento que cada jogador poderia realizar, sendo colocar a quantidade de dedos pertencente ao conjunto  $\{0, 1, 2, 3, 4, 5\}$  com 6 valores. Desta forma,  $\Omega = \text{Resultado da soma dos dedos destacados} = \{0, 1, 2, 3, \dots, 10\}$ , com  $6 \times 6 = 36$  resultados possíveis para as combinações dos pares da quantidade de dedos (*jogador 1, jogador 2*). Contando o número de elementos de  $\Omega$  temos 11 possíveis resultados para a soma dos dedos, sendo 6 pares. Mostrando isso, veio o comentário: “Não disse que **par** tem mais chances de ganhar”. Explicamos que ter mais números pares no conjunto dos possíveis resultados não prova que par é mais provável. Precisamos descobrir as frequências absolutas de cada resultado. Sejam os eventos:  $A = \{\text{Resultado da soma é par}\}$  e  $B = \{\text{Resultado da soma é ímpar}\}$ , como a soma é um número **par** ou **ímpar**, precisamos encontrar a probabilidade de apenas um dos eventos. Analisemos o evento  $A$ .

Resultados **pares** =  $A = \{0, 2, 4, 6, 8, 10\}$ .

Seja o par  $(d_1, d_2)$ , em que  $d_i$  = a quantidade de dedos destacados pelo jogador  $i$ , para  $i = 1, 2$ . Então as possíveis combinações para o evento  $A$ , estão registradas na Tabela 2.1.

Tabela 2.1: Combinações para o evento  $A$ , jogadores usando uma das mãos.

Soma	$d_1$	$d_2$	Soma	$d_1$	$d_2$
0	0	0	6	1	5
2	0	2		2	4
	1	1		3	3
	2	0		4	2
4	0	4	8	5	1
	1	3		3	5
	2	2		4	4
	3	1		5	3
	4	0	10	5	5

Usando a definição clássica de probabilidade [9], que leva em conta o quociente entre o número de casos favoráveis e o número de casos possíveis, a probabilidade do evento  $A$  é dada por:

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{1 + 3 + 5 + 5 + 3 + 1}{36} = \frac{18}{36} = \frac{1}{2} = 0,5 \text{ (1:2, metade)}$$

**Conclusão:** Se ambos os jogadores usarem apenas uma das mãos, a probabilidade da soma ser **par** é igual à probabilidade de ser **ímpar**.

### 2.3.3 Um jogador usa uma mão e o outro usa as duas

Com os alunos mais críticos, descrevemos todos os possíveis dos eventos que cada jogador poderia realizar nesta situação. Sem perda de generalidade, definimos que o jogador 1 usaria uma das mãos, portanto poderia colocar a quantidade de dedos pertencente ao conjunto  $\{0, 1, 2, 3, 4, 5\}$  e o jogador 2 usaria as duas mãos, tendo o conjunto  $\{0, 1, 2, 3, \dots, 10\}$ . Dessa forma, ao todo teríamos  $6 \times 11 = 66$  resultados possíveis,  $\Omega = \{0, 1, 2, 3, \dots, 15\}$ , sendo 8 números pares e 8 números ímpares. Sejam os mesmos eventos  $A$  e  $B$  definidos em 2.3.2.

Resultados **pares** =  $A = \{0, 2, 4, 6, 8, 10, 12, 14\}$ .

A Tabela 2.2 registra a frequência dos eventos de  $A$ .

Tabela 2.2: Combinações para  $A$ , um jogador usa uma mão e o outro usa ambas.

Soma	$d_1$	$d_2$	Soma	$d_1$	$d_2$
0	0	0	8	2	6
2	0	2		3	5
	1	1		4	4
	2	0	10	5	3
4	0	4		0	10
	1	3		1	9
	2	2		2	8
	3	1		3	7
	4	0		4	6
6	0	6		5	5
	1	5	12	2	10
	2	4		3	9
	3	3		4	8
	4	2		5	7
	5	1	14	4	10
8	0	10		5	9
	1	9			

Sendo assim, a probabilidade do evento  $A$  é:

$$\mathbb{P}(A) = \frac{1 + 3 + 5 + 6 + 6 + 6 + 4 + 2}{66} = \frac{33}{66} = \frac{1}{2} = 0,5.$$

**Conclusão:** Se um jogador usar uma das mãos e o outro jogador usar ambas, a probabilidade da soma ser PAR é igual a probabilidade de ser ÍMPAR.

### 2.3.4 Ambos os jogadores com as duas mãos

Como os dois jogadores usariam as duas mãos, então cada jogador poderia colocar a quantidade de dedos do conjunto  $\{0, 1, 2, 3, \dots, 10\}$  tendo 11 possibilidades. Ao todo teríamos,  $11 \times 11 = 121$  resultados possíveis para as combinações dos pares da quantidade de dedos. Com  $\Omega = \{0, 1, 2, 3, \dots, 18, 19, 20\}$ , tendo 21 possíveis resultados, com 11 números pares, com os definidos eventos  $A$  e  $B$ .

Resultados **pares** =  $A = \{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$

Temos que a frequência absoluta de cada evento de  $A$  é:

$A = 0 \Rightarrow 1$  possibilidade.

$A = 2 \Rightarrow 3$  possibilidades.

$A = 4 \Rightarrow 5$  possibilidades.

$A = 6 \Rightarrow 7$  possibilidades.

$A = 8 \Rightarrow 9$  possibilidades.

$A = 10 \Rightarrow 11$  possibilidades.

$A = 12 \Rightarrow 9$  possibilidades.

$A = 14 \Rightarrow 7$  possibilidades.

$A = 16 \Rightarrow 5$  possibilidades.

$A = 18 \Rightarrow 3$  possibilidades.

$A = 20 \Rightarrow 1$  possibilidade.

Sendo assim, a probabilidade do evento  $A$  é:

$$\mathbb{P}(A) = \frac{1+3+5+7+9+11+9+7+5+3+1}{121} = \frac{61}{121} \cong 0,5041$$

**Conclusão:** A probabilidade da soma ser **par** é ligeiramente maior que a probabilidade de ser **ímpar**. Para melhor entendimento, foi dito o seguinte, se dois jogadores disputassem 10000 partidas de **par ou ímpar** e um deles escolhesse sempre **par**, esse jogador ganharia mais de 5000 partidas e se esses jogadores fizessem muitas vezes 10000 disputas, pela Lei dos grandes números (LGN - seção 2.4), o que escolhesse sempre **par**, em média, ganharia 5041 partidas e perderia 4959.

Neste experimento “Jankenpon”, é esperado que, se realizarmos um número suficientemente grande de partidas, as frequências observadas estejam próximas da proporção de faces com determinado resultado no respectivo dado. Este resultado é devido a Lei dos Grandes Números (LGN) [23], conceito fundamental no ensino e estudo de probabilidade. Vale ressaltar que se o número de partidas for pequeno a variabilidade da frequência observada em torno do valor teórico pode ser grande.

## 2.4 Lei dos Grandes Números

1. *Se um evento de probabilidade  $p$  é observado repetidamente de forma independente, a proporção da frequência observada deste evento em relação ao número total de repetições (frequência relativa) converge para  $p$  à medida que o número de repetições se torna arbitrariamente grande.*



A definição matematicamente formal da *LGN* é dada por:

**Lei dos Grandes Números.** *Sejam  $X_1, X_2, X_3, \dots, X_n$ , variáveis aleatórias independentes e identicamente distribuídas (v.a.i.i.d.), com  $\mathbb{E}(\bar{X}_n) = \mu$  e  $\text{Var}(\bar{X}_n) = \sigma^2 < \infty$  (Média e Variância finitas, respectivamente), e definamos  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Então, para cada  $\epsilon > 0$  temos:*

**Lei FRACA dos Grandes Números:**  $\bar{X}_n$  converge em **probabilidade** para  $\mu$ , ou seja,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| < \epsilon) = 1$$

**Lei FORTE dos Grandes Números:**  $\bar{X}_n$  converge **quase certamente** para  $\mu$  (valor esperado = esperança), ou seja,

$$\mathbb{P}(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| < \epsilon) = 1$$

Em síntese, a *LGN* diz que é mais provável uma amostra de tamanho suficientemente grande ter a(s) característica(s) de um todo do que uma amostra pequena. As formas apresentadas como Lei Fraca e Lei Forte dos Grandes Números não definem leis distintas, e sim, maneiras diferentes de representar a aproximação da probabilidade experimentalmente observada em relação à probabilidade teórica.

A Lei Fraca dos Grandes Números nos diz que, para  $n^*$  grande, é provável que a média amostral,  $\bar{X} = (X_1 + \dots + X_{n^*})/n^*$ , esteja próxima de  $\mu$  (média populacional) [24]. Para uma proporção  $\bar{p}$  (ou frequência relativa) temos que, se  $n^*$  suficientemente grande, então é provável que esta seja próxima de  $p$  (proporção populacional). No entanto, esta Lei não diz que  $\bar{X}_n$  (ou  $\bar{p}_n$ ) permanecerá próxima de  $\mu$  (ou  $p$ ) para todo valor  $n > n^*$ . Assim, fica aberta a possibilidade de que  $|\bar{X} - \mu| > \epsilon$  (ou  $|\bar{p} - p| > \epsilon$ ) inúmeras vezes em intervalos não determináveis, mas este evento tem probabilidade de ocorrer muito próxima de zero, ou seja,  $\mathbb{P}(\bar{X} = \mu)$  (ou  $\mathbb{P}(\bar{p} = p)$ ) tende a 1.

Já a Lei Forte dos grandes números mostra que, para qualquer  $\epsilon > 0$ , com probabilidade 1,  $|\bar{X} - \mu| > \epsilon$  (ou  $|\bar{p} - p| > \epsilon$ ) ocorre apenas um número finito de vezes. Em outras palavras, dizemos que  $\bar{X} \rightarrow \mu$  (ou  $\bar{p} \rightarrow p$ ) quase certamente, onde  $\bar{X} \nrightarrow \mu$  (ou  $\bar{p} \nrightarrow p$ ) possui probabilidade zero. Desta forma, dizemos que a Lei Forte possui uma convergência pontual, na qual a estatística  $\bar{X}$ , para  $n$  suficientemente grande é idêntica a  $\mu$  (ou  $\bar{p} \equiv p$ ).

Para ajudar os alunos a compreenderem o significado da *LGN*, utilizamos o software de probabilidade com urnas (Atividade 1), de <http://goo.gl/nWD9FI>, que trata da retirada, com reposição, de bolas de uma urna. Esta urna conta com a quantidade de 20 bolas e determinamos a mesma quantidade de bolas de cada cor (10 azul e 10 vermelha). Nesta atividade, o software gera um gráfico interativo com os valores da proporção de bolas extraídas da cor azul em relação ao total de bolas retiradas, em cada extração. Computa também a quantidade de bolinhas retiradas de cada cor e possui um histórico com a cor selecionada em cada extração. Na atividade podem ser feitas até 30 extrações. O software tem as opções para extrair aleatoriamente uma bola por vez ou extrair cinco vezes uma bola. No canto inferior esquerdo temos 4 ícones que são, respectivamente, referências da teoria abordada, registro com a quantidade de bolas de cada cor determinada inicialmente, calculadora e acessibilidade. Além de 7 questões de resposta direta e 2 dissertativas sobre a atividade que podem ser resolvidos através da análise gráfica (ou usando o histórico de extrações) e são corrigidos pelo software (as opções de correção são item a item ou todas as questões da aba). A atividade possibilita ao aluno uma aprendizagem de maneira prática e lúdica. A atividade junto ao software, de certa forma, ilustrou o significado de probabilidade condicional, pois a proporção de bolas extraídas da cor azul é alterada a cada extração.

Visamos com a aplicação desta atividade, garantir a compreensão de forma leiga e ilustrada do significado de limite usando o recurso gráfico disponibilizado pelo software. O intuito é fazer com que os alunos “enxerguem” que, de maneira geral, a proporção de bolas azuis extraídas se aproxima de 0,5 (neste caso) e quanto mais extrações fizermos mais a proporção amostral se aproxima da probabilidade teórica.

Cada aluno fez a atividade uma vez respondendo as 7 questões (maioria numérica, porém observacional sem a necessidade expressa do cálculo) e registraram (no caderno) as 2 questões discursivas e a quantidade de bolas azuis extraídas em 30 retiradas para avaliarmos em sala de aula. Denotaremos

$$p_f = \frac{\#b_A}{30}; \quad p_{30} = \frac{\sum_{i=1}^{30} p_i}{30}; \quad \delta = \sum_{i=1}^{30} (p_i - 0,5)$$

sendo  $\#b_A$  o número de bolas azuis em 30 extrações;

$p_i$  a proporção de bolas azuis extraídas até a retirada  $i$ ;

$\delta$  soma das diferença entre  $p_i$  e 0,5.

As Figuras 2.1, 2.2 e 2.3 ilustram três dessas atividades, sendo que o eixo  $Y$  determina a proporção <sup>2</sup> e o eixo  $X$  o número da extração.



Figura 2.1: Proporção de bolas azuis  $n = 30$  bolas extraídas.

Na Figura 2.1, temos  $p_f = 0,4\bar{3}$ ,  $p_i \geq 0,5$ ,  $i \in \{1, 2, \dots, 18\}$  e  $p_i < 0,5$ ,  $i \in \{19, 20, \dots, 30\}$ . Um fato curioso é que da extração 16 até a 25 foram retiradas 10 bolas vermelhas. Apesar de  $p_f = 0,4\bar{3}$ , temos  $\delta > 0$ , mais exatamente,

$$\delta = \sum_{i=1}^{30} (p_i - 0,5) = 1,465$$

A média das proporções  $p_i$ ,  $p_{30}^-$ , é dada por

$$p_{30}^- = \frac{\sum_{i=1}^{30} p_i}{30} = 0,549$$

Em duas extrações (2 e 18) <sup>3</sup> o valor de  $p = 0,5$ .

<sup>2</sup>Podemos associar as extrações a uma função indicadora  $\mathcal{I}_p(b_i)$ , sendo  $b_i$  a cor da bola na extração  $i$  e  $\mathcal{I}_p(b_i) = 1$ , se a bola extraída é azul e  $\mathcal{I}_p(b_i) = 0$ , se a bola extraída é vermelha.

<sup>3</sup>Note que só podemos ter  $p_i = 0,5$ , se  $i = 2k$ ,  $i \in \{1, 2, \dots, 10\}$ , ou seja, se o valor de  $i$  for par.



Figura 2.2: Proporção de bolas azuis  $n = 30$  bolas extraídas.

Observamos na Figura 2.2 um total de 17 bolas azuis em 30 extrações com  $p_f = 0,567$ . Temos  $p_i \leq 0,5$ ,  $i \in \{1, 2, \dots, 15\}$  e  $p_i \geq 0,5$ ,  $i \in \{16, 17, \dots, 30\}$ . Da extração 11 até a 19 obtemos 9 bolas azuis. Apesar da proporção final  $p_f = 0,567$ , temos  $\delta < 0$ , sendo especificamente,

$$\delta = \sum_{i=1}^{30} (p_i - 0,5) = -2,5$$

A média das proporções  $p_{30}$  é

$$\bar{p} = \frac{\sum_{i=1}^{30} p_i}{30} = 0,417$$

Em três extrações (2, 16 e 24) o valor de  $p = 0,5$ .

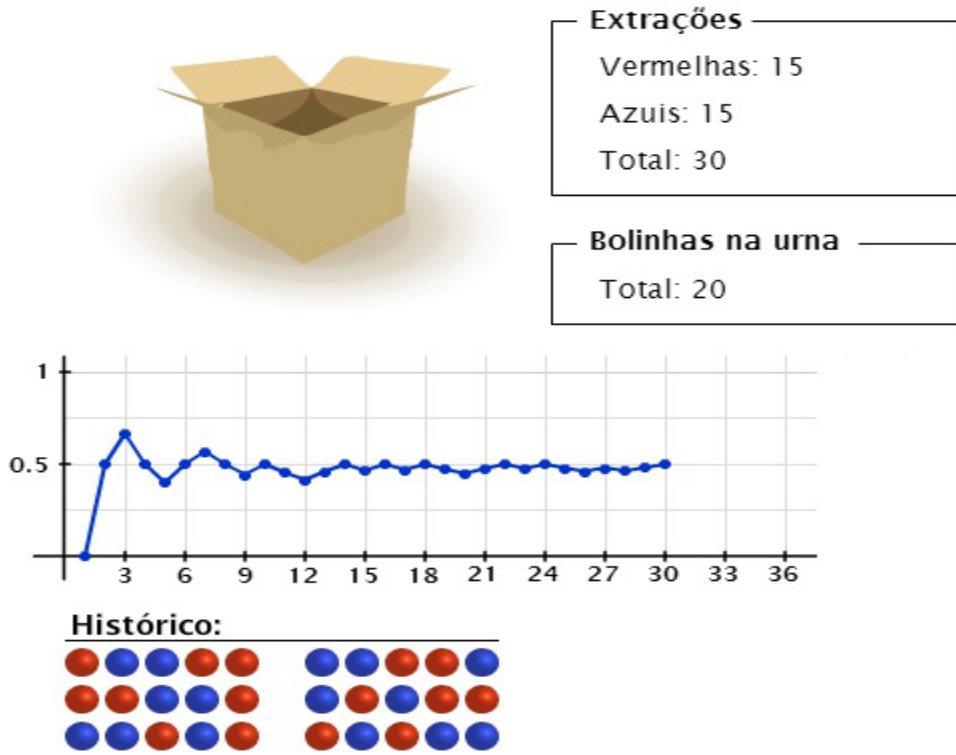


Figura 2.3: Proporção de bolas azuis  $n = 30$  bolas extraídas.

Observamos na Figura 2.3 um total de 15 bolas azuis com  $p_f = 0,5$ . Neste figura  $p_i$  oscila menos em torno linha horizontal em  $p = 0,5$ , tendo um comportamento mais próximo da probabilidade teórica. Este comportamento menos oscilatório e comprovado pelo valor encontrado de  $\delta$ , sendo este

$$\delta = \sum_{i=1}^{30} (p - 0,5) = -0,899$$

Se considerarmos a proporção média  $\bar{p}$  nas 30 extrações teremos

$$\bar{p} = \frac{\sum_{i=1}^{30} p_i}{30} = 0,47$$

Temos ainda que em onze extrações (2, 4, 6, 8, 10, 14, 16, 18, 22, 24 e 30) o valor de  $p = 0,5$ .

Se fossemos considerar estas três figuras como uma atividade com 90 extrações, teríamos  $p_f = 0,5$  e  $p_{90} = 0,479$ <sup>4</sup>. O intuito da atividade foi justamente mostrar que

<sup>4</sup>Em geral não usamos  $\bar{p}_n$  e sim a proporção  $p_f = b_a/n$ .

quanto maior o número de extrações, teríamos  $p_f \approx 0,5$  (probabilidade teórica de  $p$ ). Ou seja, se a probabilidade de retirarmos uma bola azul é de 0,5, esperamos que em 10000 extrações tenhamos um número de 5000 bolas azuis (ou próximo disto).

No total participaram 133 alunos desta atividade, realizando 3990 extrações das quais 1999 foram bolas azuis, desta forma, o valor de  $p_f = 0,501$  (erro de 1 milésimo).

Esta atividade serve de exemplo para ilustrar as Leis Fraca e Forte dos Grandes Números. Pelo resultado temos que a proporção  $p_f$  se aproxima cada vez mais de  $p = 0,5$  (Lei Forte), porém nada impede que, após algumas extrações com bola da mesma cor (por exemplo, Figura 2.1, da extração 16 até 25 ou Figura 2.2, da extração 11 até a 19), a diferença entre proporção  $p_i$  e 0,5 seja maior que um  $\epsilon > 0$  (Lei Fraca).

### 2.4.1 Jankenpon

Visando uma motivação extra foi perguntamos: “Quem já brincou de Jankenpon (ou pedra-papel-tesoura)?”. A resposta positiva foi geral e todos entenderam do que se trataria o jogo, pois é comum usarem no lugar do *par ou ímpar* para decidir quem começa a bater cartinhas (ou figurinhas) e até para outras atividades que envolvem disputas. No Jankenpon, os jogadores devem simultaneamente esticar a mão, após pronuniarem o nome do jogo, cada um forma um símbolo que representa pedra, papel ou tesoura, sendo a pedra simbolizada pelo punho fechado; a tesoura por dois dedos esticados e o papel pela mão aberta. Caso dois jogadores façam o mesmo gesto, ocorre um empate e, geralmente, joga-se de novo até desempatar. Os jogadores comparam os símbolos para decidir quem ganhou, da seguinte forma:

- Pedra ganha da tesoura (amassando-a ou quebrando-a).
- Tesoura ganha do papel (cortando-o).
- Papel ganha da pedra (embrulhando-a).

Existem variações mais complexas adotadas em campeonatos que usam até 15 sinais visuais distintos, cada qual com o seu significado conforme a Figura 2.4, onde cada um ganha de 7 outros sinais e perde para outros 7, sendo assim um jogo equilibrado cada sinal possui a mesma chance de ganhar ou perder. A interpretação da figura é simples: se a seta chega ao símbolo, significa que tal símbolo é derrotado pelo outro de onde parte a seta, poderíamos interpretar a seta como **VENCE**, por exemplo, Água vence Fogo. Resumindo a interpretação da Figura 2.4, temos que o símbolo vence os 7 símbolos dispostos imediatamente à sua esquerda (sentido anti-horário) e é derrotado pelos 7 sinais imediatamente à sua direita (sentido horário).

Os 15 sinais da Figura 2.4 (retirada de <http://goo.gl/pS2hhM>) significam, em português, air = ar, water = água, fire = fogo, rock = pedra, scissors = tesoura, gun = arma (revólver), snake = cobra, human = humano, tree = árvore, wolf = lobo, sponge = esponja, paper = papel, dragon = dragão, devil = demônio (diabo), lightning = raio. Existem outros sinais não incluídos na figura, como, por exemplo, um parecido com o raio, porém com o dedo indicador e que simboliza agulha (ou linha) e tem características similares a tesoura, ou seja, fura o papel e é amassado pela pedra.

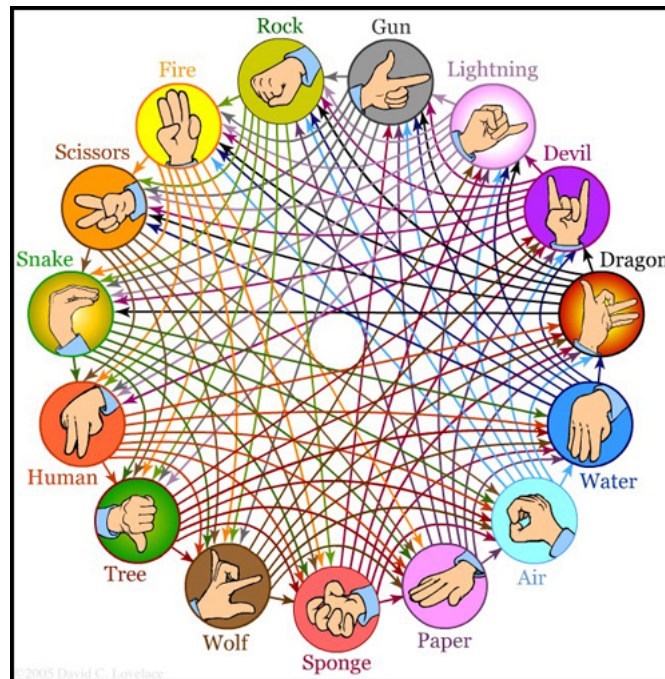


Figura 2.4: Jankenpon com 15 sinais.

Sabemos que no Jankenpon não há uma opção que é a melhor de todas, sempre existe outra que pode vencer. No jogo apresentado neste experimento, ao invés de pedra, papel e tesoura, transferimos essa noção para três dados:  $A$ ,  $B$  e  $C$ . Ao escolher dois destes, a vitória de um não é certa como no Jankenpon, uma vez que depende das faces obtidas.

As perguntas a serem respondidas são se algum dado tem mais chances de ganhar em uma disputa direta. Em outras palavras, existe o melhor dado entre  $A$ ,  $B$  e  $C$ ? Jogando os três dados simultaneamente, algum tem maior probabilidade de vitória?

O objetivo inicial era analisar os casos dois a dois,  $A \times B$ ,  $A \times C$  e  $B \times C$ , similar ao Jankenpon, porém um aluno indagou: “... e se jogarmos os três dados simultaneamente, existe algum melhor?”. Outro aluno observou que o dado  $B$  resultaria sempre no valor 4. Visando responder também esta pergunta foram mantidas as duplas e explicado que, em outra aula, os alunos deveriam escolher entre os dados  $A$  e  $C$  e considerar a cada partida que o resultado do dado  $B$  é sempre igual a 4.

Com os conceitos da Lei dos Grandes Números, determinamos a probabilidade teórica de vitória para cada par e trio na disputa de dados para fazer uma comparação com os resultados obtidos no experimento. Supondo que os dados são balanceados e que os lançamentos são equilibrados, podemos assumir que cada uma das 6 faces tem a mesma probabilidade de ser obtida em um lançamento,  $1/6$ . Sendo assim, para o dado  $A$ , o resultado 1 tem probabilidade  $4/6$  e o resultado 10 tem probabilidade  $2/6$ , para o dado  $B$ , o resultado 4 tem probabilidade 1 e para o dado  $C$ , o resultado 0 tem probabilidade  $2/6$  e o resultado 6 tem probabilidade  $4/6$ .

Podemos assumir que os resultados obtidos em cada dado são independentes, ou seja, saber o resultado no dado  $A$  não altera a probabilidade de qualquer resultado no dado  $B$  ou  $C$ . Para cada par de dados, denotemos por  $(i, j)$ , sem perda de generalidade e mantendo a ordem alfabética como referência,  $i$  é o resultado obtido no primeiro dado e  $j$  o resultado obtido no segundo dado. Com estas suposições, a probabilidade de obter o resultado  $(i, j)$  é o produto das probabilidades de obter o resultado  $i$  e o resultado  $j$ , nos respectivos dados.

Portanto, para uma disputa entre os dados  $A$  e  $B$ , os possíveis resultados são  $(i, j) = \{(1, 4), (10, 4)\}$ , com probabilidades

$$\mathbb{P}(1, 4) = \mathbb{P}_A(1)\mathbb{P}_B(4) = \frac{4}{6} \times 1 = \frac{4}{6}$$

$$\mathbb{P}(10, 4) = \mathbb{P}_A(10)\mathbb{P}_B(4) = \frac{2}{6} \times 1 = \frac{2}{6}$$

Poderíamos usar probabilidade complementar, considerando  $\mathbb{P}_B = \mathbb{P}_{A^C}$ :

$$\mathbb{P}(10, 4) = 1 - \mathbb{P}(1, 4) = 1 - \frac{4}{6} = \frac{2}{6}$$

Como o par  $(1, 4)$  é mais provável, o dado  $B$  é melhor que o dado  $A$  em um sistema de disputas com  $n$  suficientemente grande.



Para a disputa entre  $A$  e  $C$ , os pares  $(i, j)$  são:  $\{(1, 0), (1, 6), (10, 0), (10, 6)\}$ , temos que os pares  $(1, 0), (10, 0), (10, 6)$  resultam na vitória do dado  $A$ , portanto a probabilidade de  $C$  vencer é

$$\mathbb{P}(1, 6) = \mathbb{P}_A(1)\mathbb{P}_C(6) \Rightarrow \frac{4}{6} \times \frac{4}{6} = \frac{16}{36} \Rightarrow \mathbb{P}(1, 6) = \frac{4}{9}$$

Dessa forma, a probabilidade de  $A$  vencer é

$$\mathbb{P}(A \text{ vencer } C) = 1 - \frac{4}{9} = \frac{5}{9}$$

Logo, o dado  $A$  é melhor que o dado  $C$  em  $n$ , suficientemente grande, disputas.

Para uma disputa entre os dados  $B$  e  $C$ , temos  $(i, j) = \{(4, 0), (4, 6)\}$ , com a probabilidade do dado  $B$  vencer dada por

$$\mathbb{P}(4, 0) = \mathbb{P}_B(4)\mathbb{P}_C(0) = 1 \times \frac{2}{6} = \frac{2}{6}$$

e a probabilidade do dado  $C$  sair vencedor é:

$$\mathbb{P}(4, 6) = 1 - \frac{2}{6} = \frac{4}{6}$$

Sendo assim, o dado  $C$  é melhor que o  $B$  em  $n$  disputas ( $n$  suficientemente grande).

Analisando os resultados teóricos dos trios  $(a; b; c)$ , sendo que  $a, b$  e  $c$  representam o resultado do lançamento dos dados  $A, B$  e  $C$ , nessa ordem. Temos o conjunto de possíveis trios  $\{(1, 4, 0)(1, 4, 6)(10, 4, 0)(10, 4, 6)\}$ , e a probabilidade de cada dado vencer é:

$$\mathbb{P}(A \text{ vencer}) = \mathbb{P}(10, 4, 0) + \mathbb{P}(10, 4, 6) = \frac{2}{6} \times 1 \times \frac{2}{6} + \frac{2}{6} \times 1 \times \frac{4}{6} = \frac{1}{3} = \frac{3}{9}$$

$$\mathbb{P}(B \text{ vencer}) = \mathbb{P}(1, 4, 0) = \frac{4}{6} \times 1 \times \frac{2}{6} = \frac{8}{36} = \frac{2}{9}$$

$$\mathbb{P}(C \text{ vencer}) = \mathbb{P}(1, 4, 6) = \frac{4}{6} \times 1 \times \frac{4}{6} = \frac{16}{36} = \frac{4}{9}$$

Portanto, em  $n$  disputas ( $n$  grande) envolvendo os três dados, o dado  $C$  é teoricamente melhor.

## 2.5 Execução

### Jogar e coletar informações

Foram previamente preparados três dados com fita crepe colada nas faces com a seguinte numeração:

- Dado  $A$  : 1, 1, 1, 1, 10, 10;
- Dado  $B$  : 4, 4, 4, 4, 4, 4;
- Dado  $C$  : 0, 0, 6, 6, 6, 6.

Os alunos foram divididos em duplas. Desta forma, cada aluno escolheu um dos dados  $A$ ,  $B$  ou  $C$ . Foi registrado quais duplas fariam as disputas  $A \times B$ ,  $A \times C$  e  $B \times C$ . Cada aluno deveria colocar na tabela qual dado tinha escolhido, esta informação é importante pois reflete a opinião do aluno sobre qual dado tem mais chance de sair vencedor. Foram ressaltadas as regras do jogo, descritas abaixo:

- O dado escolhido é usado até o final da disputa;
- Os jogadores devem lançar os dados simultaneamente;
- Marca ponto o jogador que obtiver o maior valor na face que ficar voltada para cima;
- Não há limite para o número de partidas;
- O vencedor será aquele com o maior número de pontos ao final de todas as partidas.
- Em caso de empate, realize uma última disputa para determinar o vencedor.

Foram preenchidas tabelas similares às encontradas na tabela 2.3 de acordo com a respectiva disputa. Os alunos anotaram qual o dado vencedor da disputa e a quantidade de partidas que cada dado ganhou.

Tabela 2.3: Tabela para preenchimento do resultado das partidas.

	Dado Vencedor	
Partida	$A$	$B$
1		
2		
3		
4		
...		
$n$		
Total		

	Dado Vencedor	
Partida	$A$	$C$
1		
2		
3		
4		
...		
$n$		
Total		

	Dado Vencedor	
Partida	$B$	$C$
1		
2		
3		
4		
...		
$n$		
Total		

Como já mencionado, o objetivo inicial era analisar os casos dois a dois, assim como no Jankenpon, porém um aluno sugeriu que jogássemos os três dados simultaneamente para verificar se existe algum melhor. Visando responder tal pergunta foram mantidas as duplas, os alunos escolheram entre os dados  $A$  e  $C$ , consideraram, a cada partida, o resultado do dado  $B$  sempre igual a 4 e preencheram uma tabela similar à Tabela 2.4.

Tabela 2.4: Tabela para preenchimento do resultado das partidas com três dados.

	Dado Vencedor		
Partida	$A$	$B$	$C$
1		4	
2		4	
3		4	
4		4	
...		4	
30		4	
Total		4	

## 2.6 Análise das informações e conclusão

Ao todo foram realizadas 2247 partidas, sendo 19 disputas  $A \times B$ , 28 disputas  $A \times C$  e 16 disputas  $B \times C$ . Os resultados gerais das disputas para cada par de dados estão apresentados na Tabela 2.5.

Tabela 2.5: Frequência de vitória de cada dado nas disputas.

$A \times B$	Vitórias	$F_{\text{Observada}}$	Probabilidade
$A$	219	32,59	0,3333
$B$	453	67,41	0,6667
Total	672	100	1

$A \times C$	Vitórias	$F_{\text{Observada}}$	Probabilidade
$A$	521	56,2	0,5556
$C$	406	43,8	0,4444
Total	927	100	1

$B \times C$	Vitórias	$F_{\text{Observada}}$	Probabilidade
$B$	211	32,56	0,3333
$C$	437	67,44	0,6667
Total	648	100	1

Com o resultado do experimento e o resultado teórico, apresentados na Tabela 2.5, podemos comparar o quão próximos eles são. Vale ressaltar que essa comparação é válida pela Lei Grandes Números, sendo a média  $\mu = 4$  em cada dado e cada variância é finita. Para verificar a proximidade dos resultados, fizemos a diferença entre a frequência relativa observada e a probabilidade esperada.

Na disputa  $A \times B$

$$\mathbb{P}(A) - F_{\text{Observada}}(A) = 0,3333 - 0,3259 = 0,0074$$

Um diferença observada de pouco mais de 7 milésimos, muito próxima do resultado teórico. Como as probabilidades são complementares, a diferença entre a  $F_{\text{Observada}}(B)$  e  $\mathbb{P}(B)$ , em módulo, é a mesma.

Na disputa  $A \times C$

$$F_{R_{observada}}(A) - \mathbb{P}(A) = 0,5620 - 0,5556 = 0,0064$$

A diferença é pouco superior a 6 milésimos.

Na disputa  $B \times C$

$$F_{R_{observada}}(C) - \mathbb{P}(C) = 0,6744 - 0,6667 = 0,0077$$

A diferença é próxima de 8 milésimos.

Por fim, verificamos se na disputa simultânea entre  $A$ ,  $B$  e  $C$ , algum dado tem mais chance de ganhar. Para essa forma de disputa, foram realizadas 2187 partidas e obtivemos as frequências absolutas relatadas na Tabela 2.6.

Tabela 2.6: Frequência de vitória dos dados na disputa simultânea.

$A \times B \times C$	Vitórias	$F_{R_{observada}}$	Probabilidade
$A$	731	33,42%	0,3333
$B$	479	21,9%	0,2222
$C$	977	44,68%	0,4445
Total	2187	100%	1

Pelos resultados da tabela 2.6, temos evidências que o dado  $C$  é melhor em um sistema de  $n$  lançamentos simultâneos dos três dados. Comparando as frequências relativas observadas e os resultados teóricos para a disputa  $A \times B \times C$ , temos

$$F_{R_{observada}}(A) - \mathbb{P}(A) = 0,3342 - 0,3323 = 0,0009$$

$$\mathbb{P}(B) - F_{R_{observada}}(B) = 0,2222 - 0,219 = 0,0032$$

$$F_{R_{observada}}(C) - \mathbb{P}(C) = 0,4468 - 0,4445 = 0,0023$$

O principal objetivo deste experimento foi trabalhar conceitos de probabilidade mostrando a aplicação da teoria abordada na forma de uma atividade lúdica e informativa para que os alunos compreendam o significado dos resultados teóricos e tenham mais proximidade com o tema.



# Capítulo 3

## Método de Monte Carlo

### 3.1 O Experimento

#### Sinopse

Neste experimento será apresentado um processo, conhecido como método de Monte Carlo, utilizado para estimar a área de qualquer figura. Calcularemos uma estimativa para a área do território brasileiro a partir de um mapa. Conceitos prévios serão necessários para o entendimento e realização do experimento.

#### Conteúdos

Probabilidade: Probabilidade geométrica;  
Razão e proporção: Proporcionalidade direta.

#### Objetivos do experimento

1. Apresentar um método que permite estimar a área de qualquer figura;
2. Abordar probabilidade geométrica.

### 3.2 Introdução teórica

Há vários métodos com o nome *Monte Carlo* e que seguem o mesmo padrão. Designa-se por **método de Monte Carlo** [9] qualquer método que se baseie em amostragens aleatórias massivas para obter resultados numéricos. Repetindo sucessivas simulações um grande número de vezes, podemos calcular, por exemplo, a média dos valores das proporções encontradas se aproxima da probabilidade térica. Este método é utilizado

em simulações estocásticas (não determinísticas) com diversas aplicações em áreas como física, matemática e biologia, em variados campos, como desenho de reator nuclear [25], cromodinâmica quântica [26], terapia de câncer por radiação [27], fluxo de tráfego [28], evolução estelar [29], sistemas de muitos corpos interagentes [30], econometria [31], entre outros.

Neste experimento vamos trabalhar com uma das aplicações do método estimando a área de uma figura qualquer. Esta estimativa será explicada na introdução teórica com o uso do conceito de probabilidade geométrica. Há muitos cálculos em Física Nuclear e de Altas Energias que precisam determinar quantidades equivalentes a áreas (integrais) que são impossíveis de se calcular de maneira exata ou muito demoradas para o computador calcular com precisão. Nestes casos, as estimativas fornecidas pelos métodos de Monte Carlo são importantes para obter respostas em tempos aceitáveis.

A ideia da aplicabilidade surgiu das aulas de cálculo de áreas <sup>1</sup> de figuras bem definidas como triângulos, quadriláteros e outras figuras que pudessem ser compostas (ou decompostas) com a junção dessas figuras. Foi proposto o desafio de calcular a área da superfície do Brasil, pelo método de Monte Carlo, usando um mapa com o contorno do território brasileiro. Esta atividade prática lida com estatística e probabilidade, relacionando-as a conceitos geométricos. O principal objetivo do experimento não é a obtenção da área, e sim, a percepção de como responder perguntas de cálculo utilizando ferramentas probabilísticas. Este experimento usa diversos conhecimentos matemáticos para desenvolvê-lo, tais como razão e proporcionalidade direta, mudança de unidades de comprimento e área, notação científica, localização de pontos no plano cartesiano, conhecimento geográfico, dentre outros.

Previamente foram esclarecidos e retomados conceitos matemáticos como razão, proporcionalidade direta, mudança de unidade de comprimento e área (método da escada), escrita de números em notação científica (potência de 10), localização de pontos no primeiro quadrante do plano cartesiano. Também utilizamos conceitos da disciplina de geografia para esclarecer, por exemplo, que a antiga expressão “Do Oiapoque (Amapá) ao Chuí (Rio Grande do Sul)”, usada para referenciar a extensão territorial do Brasil do extremo norte ao extremo sul, deveria ser trocada por “Do Monte Caburaí (Roraima) ao Chuí”, pois o Monte Caburaí encontra-se mais de 80 quilômetros (Km) acima do Oiapoque, sendo o extremo norte. Outra utilidade, em termos de conhecimentos geográficos (e matemáticos) foi fixar a ideia de escala, amplamente usada na disciplina que estuda a terra e que aparece com frequência em exercícios matemáticos que envolvem planta de residência e desenhos de figuras.

---

<sup>1</sup>Na seção “Complementar” deste capítulo veremos outras formas para o cálculo de área e volume



Vamos justificar em termos de probabilidade geométrica por que o método de Monte Carlo funciona.

Considere que um único ponto, gerado de maneira aleatória, estará situado em uma malha quadriculada, totalmente ordenada, que possui uma figura plana em seu interior. Usando o conceito de probabilidade geométrica [32], se tivermos uma região  $B$  com área  $A_B$  contida em uma região  $A$  com área  $A_A$ , a probabilidade de um ponto de  $A$  também pertencer a  $B$  é proporcional à área de  $A_B$  em relação a área de  $A_A$  e independe da posição que  $B$  ocupa em  $A$ . Portanto, selecionado ao acaso um ponto  $p$  de  $A$ , a probabilidade de que ele pertença a  $B$  será:

$$\mathbb{P}(p \in B) = \frac{A_B}{A_A}$$

Neste experimento,  $A_B$  é a área do contorno da superfície do Brasil e  $A_A$  é a área da malha quadriculada. A razão entre as áreas é igual à probabilidade de que um ponto da malha quadriculada também pertença ao mapa com o contorno do território brasileiro. Experimentalmente, essa probabilidade é a razão entre o número de pontos dentro do contorno e o número total de pontos na malha. Uma ilustração da região  $B$  contida na região  $A$  na figura 3.1 (retirada de [3]).

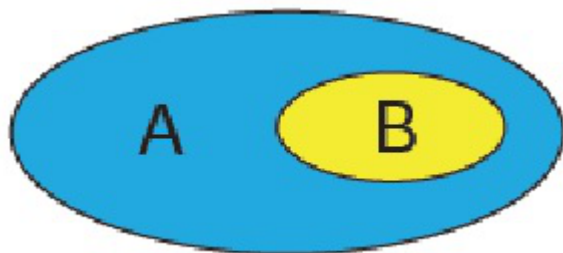


Figura 3.1: Região  $B$  contida na região  $A$ .

Para o experimento “sortear um ponto da região  $A$ ”, podemos definir o evento “ponto sorteado está na região  $B$ ”. Denotemos por  $p$  a probabilidade de que este evento ocorra. Desta forma, a variável aleatória  $X$ , definida como  $X = 1$ , se o evento ocorrer, e  $X = 0$ , se não, tem distribuição de Bernoulli de parâmetro  $p$ , ( $X \sim \text{Bern}(p)$ ), onde  $p$  é a probabilidade de um ponto aleatório estar dentro da região  $B$ , nesse caso, dentro do contorno do mapa.

Definamos a variável aleatória  $P$  sendo a proporção de pontos dentro da figura em  $n$  repetições independentes de um experimento com  $X_i \sim \text{Bern}(p)$ , então veremos que

$$P = \sum_{i=1}^n X_i \sim b(n, p), \text{ ou seja, tem distribuição Binomial de parâmetros } n \text{ e } p.$$

### 3.2.1 Propriedades da esperança de uma variável aleatória

A esperança de uma variável aleatória  $X$  é definida como o valor médio de  $X$  em seu espaço amostral. Matematicamente,

- Se  $X$  é uma variável aleatória discreta, então  $\mathbb{E}(X) = \sum_{i=1}^n x_i P(X = x_i)$
- Se  $X$  é uma variável aleatória contínua, então  $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$

Sejam  $X$  e  $Y$  variáveis aleatórias,  $a$ ,  $b$  e  $c$ , constantes, então a esperança tem as seguintes propriedades [9], para quaisquer constantes  $a$ ,  $b$  e  $c$  e variáveis  $X, Y$  com  $\mathbb{E}(X) < \infty$  e  $\mathbb{E}(Y) < \infty$ :

- E1.**  $\mathbb{E}(a) = a$ ;
- E2.**  $\mathbb{E}(a + X) = a + \mathbb{E}(X)$ ;
- E3.**  $\mathbb{E}(bX) = b\mathbb{E}(X)$ ;
- E4.**  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ ;
- E5.**  $\mathbb{E}(a + bX + cY) = a + b\mathbb{E}(X) + c\mathbb{E}(Y)$ ;
- E6.** Se  $X$  e  $Y$  são independentes, então  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .

**Esperança da soma de variáveis aleatórias independentes e identicamente distribuídas.** *Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas (v.a.i.i.d.), com esperança  $\mathbb{E}(X_i) = \mu$  para  $i = 1, 2, \dots, n$ , então:*

$$Y = \sum_{i=1}^n X_i, \text{ tem } \mathbb{E}(Y) = n\mu$$

**Demonstração.**  $\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \mathbb{E}(X_1 + X_2 + \dots + X_n)$  (por E4)

$$\Rightarrow \mathbb{E}(Y) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n) = \underbrace{\mu + \mu + \dots + \mu}_{n \text{ vezes}} = n\mu \quad \blacksquare$$

Desse modo, se  $X_i \sim \text{Bern}(p)$ ,  $\mathbb{E}(X_i) = p$ , então  
 $\mathbb{E}(P) = \mathbb{E}(X_1 + X_2 + \cdots + X_n) = np$  ( $P \sim b(n, p)$ ).

### 3.2.2 Propriedades da variância de variáveis aleatórias

Seja  $X$  uma variável aleatória, sua variância é definida como

$$\mathbb{V}ar(X) = \mathbb{E}(X - \mathbb{E}(X))^2.$$

Para  $c$  constante, a variância tem as seguintes propriedades:

- V1.**  $\mathbb{V}ar(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- V2.**  $\mathbb{V}ar(X + c) = \mathbb{V}ar(X) \Rightarrow \mathbb{V}ar(c) = 0$
- V3.**  $\mathbb{V}ar(cX) = c^2 \mathbb{V}ar(X)$
- V4.** Se  $X$  e  $Y$  são variáveis aleatórias independentes, então

$$\mathbb{V}ar(X + Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y)$$

- V5.** Se  $X_1, X_2, \dots, X_n$  são variáveis aleatórias independentes, então

$$\mathbb{V}ar(X_1 + X_2 + \cdots + X_n) = \mathbb{V}ar(X_1) + \mathbb{V}ar(X_2) + \cdots + \mathbb{V}ar(X_n)$$

**Variância da Soma de Variáveis Aleatórias Independentes e Identicamente Distribuídas.** *Sejam  $X_1, X_2, \dots, X_n$  são variáveis aleatórias independentes e identicamente distribuídas (v.a.i.i.d.), com  $\mathbb{V}ar(X_i) = \sigma^2 < \infty$ , então*

$$P = \sum_{i=1}^n X_i, \text{ tem } \mathbb{V}ar(P) = n\sigma^2$$

**Demonstração.** *Primeiro vamos verificar se vale para  $n = 2$  (V4).*

$$\begin{aligned} \mathbb{V}ar(X_1 + X_2) &= \mathbb{E}[(X_1 + X_2)^2] - [\mathbb{E}(X_1 + X_2)]^2 = \\ &= \mathbb{E}[(X_1 + X_2)(X_1 + X_2)] - [\mathbb{E}(X_1 + X_2)\mathbb{E}(X_1 + X_2)] = \\ &= \mathbb{E}(X_1^2 + 2X_1X_2 + X_2^2) - (\mathbb{E}(X_1)\mathbb{E}(X_1) + 2\mathbb{E}(X_1)\mathbb{E}(X_2) + \mathbb{E}(X_2)\mathbb{E}(X_2)) = \\ &= \mathbb{E}(X_1^2) + 2\mathbb{E}(X_1X_2) + \mathbb{E}(X_2^2) - \mathbb{E}(X_1)^2 - 2\mathbb{E}(X_1)\mathbb{E}(X_2) - \mathbb{E}(X_2)^2 \end{aligned}$$

*Como  $X_1$  e  $X_2$  são independentes temos que  $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$ , portanto podemos reescrever a expressão como*

$$\mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 + \mathbb{E}(X_2^2) - \mathbb{E}(X_2)^2 = \mathbb{V}ar(X_1) + \mathbb{V}ar(X_2) = \sigma^2 + \sigma^2 = 2\sigma^2 \quad \blacksquare$$

Agora, por indução, vamos supor que vale para  $n \in \mathbb{N}$  e verificar que também vale para  $(n + 1)$ .

**Hipótese:** Se temos v.a.i.i.d., então  $\mathbb{V}ar(X_1 + X_2 + \cdots + X_n) = n\sigma^2$

**Tese:** Se temos v.a.i.i.d., então  $\mathbb{V}ar(X_1 + X_2 + \cdots + X_n + X_{n+1}) = (n + 1)\sigma^2$

**Prova:**

$$\mathbb{V}ar(Y) = \mathbb{V}ar(X_1 + X_2 + \cdots + X_n + X_{n+1})$$

$$\begin{aligned} \mathbb{V}ar(Y) &= \mathbb{E}[(X_1 + X_2 + \cdots + X_n + X_{n+1})^2] - [\mathbb{E}(X_1 + X_2 + \cdots + X_n + X_{n+1})]^2 = \\ &= \mathbb{E}[(X_1 + \cdots + X_{n+1})(X_1 + \cdots + X_{n+1})] - [\mathbb{E}(X_1 + \cdots + X_{n+1})\mathbb{E}(X_1 + \cdots + X_{n+1})] = \\ &= \mathbb{E}\left(\sum_{i=1}^{n+1} X_i^2 + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n+1} X_i X_j\right) - \left(\sum_{i=1}^{n+1} (\mathbb{E}X_i)^2 + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n+1} \mathbb{E}(X_i)\mathbb{E}(X_j)\right) = \\ &= \left(\sum_{i=1}^{n+1} \mathbb{E}(X_i^2)\right) + 2 \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n+1} \mathbb{E}(X_i X_j)\right) - \left(\sum_{i=1}^{n+1} (\mathbb{E}X_i)^2\right) - 2 \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n+1} \mathbb{E}(X_i)\mathbb{E}(X_j)\right) \end{aligned}$$

Como as variáveis são independentes, então:

$$2 \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n+1} \mathbb{E}(X_i X_j)\right) = 2 \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n+1} \mathbb{E}(X_i)\mathbb{E}(X_j)\right)$$

Portanto, podemos reescrever e temos que

$$\begin{aligned} \mathbb{V}ar(Y) &= \left(\sum_{i=1}^{n+1} \mathbb{E}(X_i^2)\right) - \left(\sum_{i=1}^{n+1} (\mathbb{E}X_i)^2\right) \Rightarrow \\ &= \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 + \cdots + \mathbb{E}(X_n^2) - \mathbb{E}(X_n)^2 + \mathbb{E}(X_{n+1}^2) - \mathbb{E}(X_{n+1})^2 = \\ &= \underbrace{\mathbb{V}ar(X_1) + \cdots + \mathbb{V}ar(X_n)} + \mathbb{V}ar(X_{n+1}) = n\sigma^2 + \sigma^2 = (n + 1)\sigma^2 \quad \blacksquare \end{aligned}$$

Desse modo, se  $\mathbb{V}ar(X_i) = p(1 - p)$  ( $X_i \sim \text{Bern}(p)$ ) então

$$\mathbb{V}ar(P) = \mathbb{V}ar\left(\sum_{i=1}^n X_i\right) = n\mathbb{V}ar(X_i) = np(1 - p) \quad (P \sim b(n, p)).$$

Nesse experimento, não sabemos precisamente qual a proporção que a área do contorno da superfície do território brasileiro representa em relação malha quadriculada, então precisamos estimar essa proporção através de amostragem. A estatística amostral  $\hat{p}$ , definida como

$$\hat{p} = \frac{P}{n} = \frac{\sum_{i=1}^n X_i}{n}.$$

possui todas propriedades apresentadas para a esperança e variância, podemos justificar o uso dessa estatística amostral como estimativa para a proporção populacional  $p$ , pelo Teorema Central do Limite (TCL).

### 3.2.3 Teorema Central do Limite

**Teorema** (Teorema Central do Limite). *Sejam  $X_1, X_2, X_3, \dots$ , variáveis aleatórias independentes e identicamente distribuídas (v.a.i.i.d.) com qualquer distribuição de probabilidade, tal que  $\mathbb{E}(X_i) = \mu, \text{Var}(X_i) = \sigma^2 < \infty$  (Média e Variância finitas) e*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

*Então, se  $n \rightarrow \infty$  ( $n$  é suficientemente grande), temos que:*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

*ou seja, o TCL garante que qualquer que seja a distribuição da variável aleatória  $X$  (discreta ou contínua), para uma amostra suficientemente grande, a distribuição amostral do quociente da diferença entre a média amostral e a populacional pelo desvio padrão em torno da média é  $\sim N(0; 1)$ .*

Para que a proporção de pontos,  $\hat{p}$ , dentro da figura com o contorno do território brasileiro seja uma boa estimativa para a área, é importante construir uma malha com um grande número de pontos e a proporção da figura em relação a esta não ser extremamente pequena (nem extremamente grande). Pelo Teorema Central do Limite, temos que o erro de previsão é igual a:

$$\varepsilon = \frac{\sigma}{\sqrt{n}}$$

Por causa desta relação, entendemos que, quanto maior for a quantidade de pontos gerados, melhor será a estimativa da área, uma vez que este erro é inversamente proporcional a  $n$ . Isso porque, pela LGN, a proporção se aproxima da probabilidade para  $n$  suficientemente grande e, pelo TCL, o erro amostral entre a proporção estimada e a probabilidade se aproxima de zero.

Como a estatística  $\hat{p}$  é o produto da soma de v.a.i.i.d. por uma constante, são válidas todas as propriedades para esperança e variância e pelo  $TCL$ <sup>2</sup>, temos

$$\frac{\hat{p} - p}{\sigma_{\hat{p}}/n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

Amparados nestes conceitos, temos que a relação (1) vale para qualquer figura que seja desenhada na malha quadriculada:

$$(1) \quad \frac{\text{Número médio de pontos dentro da figura}}{\text{Número de pontos gerados}} \approx \frac{\text{Área da figura}}{\text{Área do quadrado}}$$

Para a relação (2), usamos a área territorial,  $8514877Km^2$  e da distância do Monte Caburaí ao Chuí de  $4395Km$ . Depois, no mapa, construimos a distância do Monte Caburaí até o Chuí, extremo norte até extremo sul, Figura 3.2.



Figura 3.2: Distância do Monte Caburaí até o Chuí.

---

<sup>2</sup>Outros exemplos do uso do  $TCL$  é o movimento de um pêndulo (ou mola) que tende a estabilizar [37]; o erro em experimentos físicos e o bombardeamento molecular que uma partícula sofre suspensa em um líquido [9].

Usaremos a relação (2) (deixando área em  $cm^2$  e distância em  $cm$  conforme o “método da escada”) para calcular uma estimativa para a Área do território brasileiro:

$$(2) \frac{\text{Área do mapa}}{\text{Área do Brasil}} = \frac{(\text{Distância do Caburaí ao Chuí no mapa})^2}{(\text{Distância real do Caburaí ao Chuí})^2}$$

O método da escada para mudança de unidade de comprimento foi ensinado também para área e volume. Tal método consiste em multiplicar (ou dividir) o número requerido por 10 elevado a potência do número de degraus que se desce (ou sobe) para ir da unidade inicial até a unidade final.

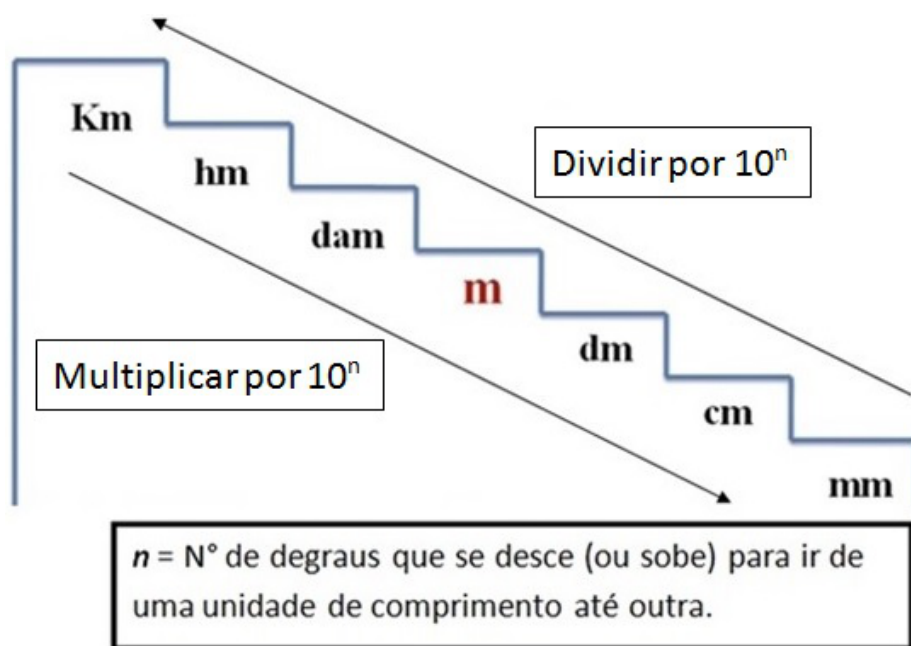


Figura 3.3: Método da escada para mudança de unidade de comprimento.

Por exemplo, suponha que queremos transformar  $5,375hm$  (hectômetros) em  $dm$  (decímetros). Para ir de  $hm$  até  $dm$  descemos 3 degraus, então

$$5,375hm = 5,375 \cdot 10^3 dm = 5375dm.$$

Se quisermos transformar  $4,5mm$  (milímetros) em  $dam$  (decâmetros). Temos que de  $mm$  para  $dam$  subimos 4 degraus então

$$4,5mm = \frac{4,5}{10^4} dam = 0,00045dam = 4,5 \cdot 10^{-4} dam.$$

Para área e volume, basta elevarmos a relação linear, em ambos os lados, pela respectiva dimensão. Ou seja, para sabermos quantos  $mm^2$  temos em  $18cm^2$ , fazemos a relação  $1cm = 10mm$  ao quadrado, obtemos que  $1cm^2 = 100mm^2$ , então

$$18cm^2 = 18 \times 100mm^2 = 1800mm^2 = 1,8 \times 10^3mm^2.$$

Para calcularmos quantos  $dam^3$  tem em  $5dm^3$ , fazemos  $1dm = 10^{-2}dam$  ao cubo obtendo que  $1dm^3 = 10^{-6}dam$ , portanto

$$5dm^3 = \frac{5}{10^6}dam^3 = 5 \times 10^{-6}dam^3.$$

Podemos dizer que a característica aleatória desse procedimento está no fato de que a geração é feita para que os pontos, em média, fiquem distribuídos de forma uniforme.

Vale salientar que os pontos foram gerados, para cada eixo  $X$  e  $Y$ , a partir de uma distribuição Uniforme Discreta  $\{a, a + 1, \dots, b - 1, b\}$  ( $\sim Unif\{a, b\}$ ), para  $a = 1$ , usando o comando  $(=ALEATÓRIOENTRE(a, b))$ , no programa *Excel* (ou no programa *R*,  $x < -sample(a : b, n)$ , que gera  $n$  valores aleatórios). Os pontos gerados são denominados *pseudoaleatórios*, pois “simulam” o comportamento de números aleatórios.

A distribuição Uniforme Discreta é característica de uma variável aleatória  $X$  cujos resultados são valores inteiros, equiprováveis, entre  $a$  e  $b$ ,  $a < b$ . Esta distribuição possui as seguintes expressões para o valor da esperança e variância.

$$\mathbb{E}(X) = \frac{b + a}{2}; \quad \mathbb{V}ar(X) = \frac{(b + a)(b - a)}{12} = \frac{b^2 - a^2}{12}.$$

Vale lembrar que o objetivo principal é aplicar os conceitos de probabilidade e estatística aprendidos em aula e não propriamente a obtenção da área. Além disso, quando trabalhamos com estimativas existem fatores que podem ocasionar erros, dentre eles, destacamos:

- **Erros nas medidas:** A figura copiada apresenta distorções, tanto na planificação de parte do globo terrestre quanto na cópia em uma impressora. As medidas obtidas das distâncias entre as cidades mencionadas e a área real do Brasil são as melhores possíveis no globo e não necessariamente mantêm a proporcionalidade no mapa utilizado. A medida tomada com o uso de uma régua milimetrada tem a limitação da precisão da ordem de 1 milímetro tendo dois algarismos significativos;
- **Erros estatísticos:** É de se esperar que a distribuição dos pontos gerados aleatoriamente seja aproximadamente uniforme em relação à malha quadriculada



criada, sendo que as possíveis flutuações da distribuição uniforme implicam erros nas estimativas feitas. Os pontos foram arredondados, conforme critérios computacionais, porém, em especial, na malha com a quadrícula de 2 centímetros para a qual foram gerados 10 pontos aleatórios de 16 possíveis, houve casos com reincidência de pontos. Por exemplo, aparecer duas vezes em 10 pontos, o par ordenado  $(x, y) = (3, 3)$ , que estaria 'dentro' do mapa, segundo os critérios adotados, sendo assim, foi considerado dois pontos dentro do mapa, resultado que poderia ser alterado se fosse gerado outro ponto. Além disso, o critério é, de certa forma, problemático, pois quadrados com partes da área do contorno teriam o mesmo “peso” de quadrados totalmente dentro do contorno;

### 3.3 Execução e análise dos resultados

#### 3.3.1 Etapa 1: Entendendo o método

Esta etapa pretende dar alguma habilidade de manejo para esclarecer o funcionamento do método no contexto de medida e de experimento, isto é, trata-se de estimativas estatísticas com erros devido a flutuação na amostra, além dos erros de limitações das medidas diretas. Pretendemos também avaliar as duas variantes envolvidas que são o tamanho da quadrícula e a quantidade de pontos gerados.

Os alunos presentes na sala de aula foram divididos em 5 grupos que receberam uma folha com 3 mapas com quadrículas diferentes, como mostra a Figura 3.4.

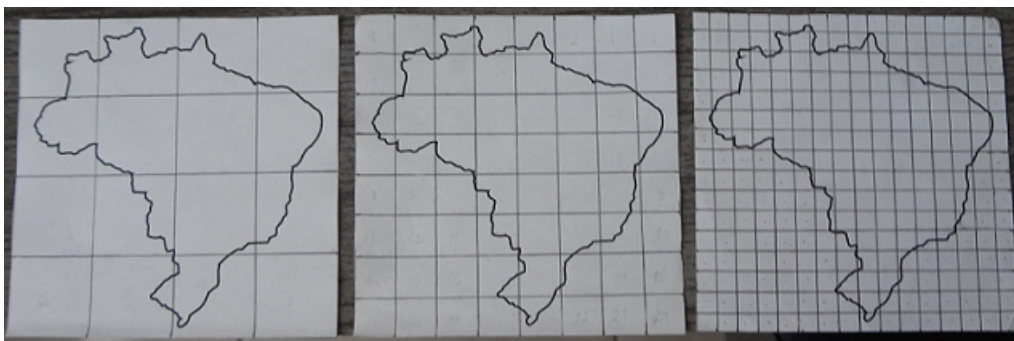


Figura 3.4: Mapas quadriculados em quadrados  $8cm \times 8cm$ .

O mapa à esquerda foi quadriculado de modo que seus lados medissem 2 centímetros; o mapa que se encontra ao centro tem quadrados com 1 centímetro de

lado e o que se encontra à direita possui lados com 0,5 centímetro. Diminuímos o tamanho do lado para que a proporção da área do contorno do mapa se aproxime da probabilidade de, um ponto, gerado ao acaso, estar dentro do mapa.

Foram dadas as seguintes orientações:

- Numerar, no mapa com quadrícula de 2 cm, os quadrados, na parte de baixo da folha, de 1 a 4, da esquerda à direita;
- Numerar o lado esquerdo do mapa de 1 a 4, começando de baixo para cima;

A seguir foram entregues, a cada aluno, 10 pares de pontos gerados aleatoriamente (dispostos como na Tabela 3.1) a partir de uma distribuição Uniforme Discreta de 1 até 4 ( $\sim Unif\{1, 4\}$ ), definida sobre o conjunto  $\{1, 2, 3, 4\}$  para os eixos  $X$  e  $Y$ .

Tabela 3.1: Tabela com  $n$  pontos gerados de forma aleatória.

Ponto	$X$	$Y$
1	$x_1$	$y_1$
2	$x_2$	$y_2$
...	...	...
$n$	$x_n$	$y_n$

Após a entrega dos pares de pontos, foram dadas as seguintes orientações:

- Localize todos os pontos, lembrando que à esquerda temos o valor de  $X$  e à direita, o valor de  $Y$ ;
- Pontos cujo quadrado houver uma parte do contorno do mapa, por menor que seja, serão considerados dentro do mapa; caso contrário, o ponto será considerado fora (se o quadrado for totalmente branco);
- Contar quantos pontos ficaram nas delimitações do contorno do mapa da superfície do território brasileiro.

Na primeira parte do experimento, fizemos a distribuição uniforme de  $n$  pontos em uma área de  $64cm^2$  com 16 quadrados de área  $4cm^2$ , ou seja, cada ponto cobre uma área de  $4cm^2$ . Dados os critérios de decisão e a malha quadriculada, foram obtidos 13 de 16 pontos com alguma parte do contorno, assim, a proporção de pontos no mapa foi de 13/16 (81,25%). Outro critério seria considerar o centro do ponto gerado contendo o contorno, o que diminuiria para 6 de 16 (37,5%), a proporção de pontos dentro do contorno da superfície. Reforçamos que o intuito dessa etapa era fazer com que os alunos entendessem quais procedimentos seriam seguidos, os critérios adotados e as relações estabelecidas.

A Tabela 3.2 mostra a frequência observada de pontos dentro do mapa.

Tabela 3.2: Frequência de pontos dentro do mapa  $2cm \times 2cm$ ,  $n = 10$ .

Pts Dentro	frequência	# pts $\times$ frequência
7	2	14
8	21	168
9	39	351
10	67	670
Total	129	1203

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

Esta Tabela 3.2 fornece o número médio de  $1203/129=9,326$  pontos dentro da figura. Usando a relação (1), a estimativa para área do mapa foi

$$\frac{9,326}{10} \approx \frac{A_{mapa}}{64} \Rightarrow A_{mapa} = 64 \times 0,9326 \Rightarrow A_{mapa} \cong 59,69cm^2$$

Usando a relação (2), com as medidas para área em  $cm^2$  e distância em  $cm$ , calculamos uma estimativa para a área do território brasileiro.

$$\frac{59,69cm^2}{A_{BRASIL}} = \frac{(7,5cm)^2}{(4,395 \times 10^8cm)^2} \Rightarrow A_{BRASIL} = \frac{59,69cm^2 \times 19,316025 \times 10^{16}cm^2}{56,25cm^2}$$

$$A_{BRASIL} \cong 20,32561 \times 10^{16}cm^2 = 20325610km^2.$$

Essa parte do experimento gerou uma estimativa de área 2,387 vezes a área territorial brasileira, mas foi útil para destacarmos os possíveis erros (nas medidas e estatísticos) e de grande valia para o entendimento e aplicação do método de Monte Carlo. A execução dessa parte contribuiu para que os alunos entendessem como localizar pontos na malha quadriculada, usar as relações estabelecidas, aplicar conceitos como mudança na unidade de medida, notação científica, razão e proporção, arredondamento de números, semelhança de figuras planas, desenvolver o raciocínio lógico e crítico.

Depois entregamos, a cada aluno, 50 pares de pontos gerados (aleatoriamente) de  $(\sim Unif\{1,4\})$  e pedimos para que contassem o número de pontos para que fazermos os mesmos procedimentos acima e obter uma nova estimativa para a área do território brasileiro.

Os valores observados estão apresentados na Tabela 3.3.

Tabela 3.3: Frequência de pontos dentro do mapa  $2cm \times 2cm$ ,  $n = 50$ .

Pts Dentro	frequência	#pts $\times$ frequência
36	3	108
37	6	222
38	8	304
39	13	507
40	17	680
41	22	902
42	20	840
43	18	774
44	14	616
45	8	360
Total	129	5313

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

O número médio de  $5313/129=41,186$  pontos dentro da figura. Usando a relação (1), a estimativa para área do mapa foi

$$\frac{41,186}{50} \approx \frac{A_{mapa}}{64} \Rightarrow A_{mapa} = 64 \times 0,82372 \Rightarrow A_{mapa} \cong 52,718cm^2$$

Usando a relação (2), a estimativa para a área do território brasileiro foi

$$\frac{52,718cm^2}{A_{BRASIL}} = \frac{(7,5cm)^2}{(4,395 \times 10^8cm)^2} \Rightarrow A_{BRASIL} = \frac{52,718cm^2 \times 19,316025 \times 10^{16}cm^2}{56,25cm^2}$$

$$A_{BRASIL} \cong 18,1032 \times 10^{16}cm^2 = 18103200km^2.$$

A estimativa da área é 2,126 vezes a área real.

Posteriormente entregamos, a cada aluno, 100 pares de pontos gerados (aleatoriamente) de  $(\sim Unif\{1,4\})$  e pedimos para que contassem o número de pontos para que fazemos os mesmos procedimentos acima e obter uma nova estimativa para a área do território brasileiro.

Os valores observados estão apresentados na Tabela 3.4.

Tabela 3.4: Frequência de pontos dentro do mapa  $2cm \times 2cm$ ,  $n = 100$ .

Pts Dentro	frequência	#pts $\times$ frequência
76	2	152
77	5	385
78	9	702
79	13	1027
80	16	1280
81	21	1701
82	20	1640
83	18	1494
84	16	1344
85	9	765
Total	129	10490

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

O número médio de  $10490/129=81,318$  pontos dentro da figura. Usando a relação (1), a estimativa para área do mapa foi

$$\frac{81,318}{100} \approx \frac{A_{mapa}}{64} \Rightarrow A_{mapa} = 64 \times 8,1318 \Rightarrow A_{mapa} \cong 52,043cm^2$$

Usando a relação (2), a estimativa para a área do território brasileiro foi

$$\frac{52,043cm^2}{A_{BRASIL}} = \frac{(7,5cm)^2}{(4,395 \times 10^8cm)^2} \Rightarrow A_{BRASIL} = \frac{52,043cm^2 \times 19,316025 \times 10^{16}cm^2}{56,25cm^2}$$

$$A_{BRASIL} \cong 17,872 \times 10^{16}cm^2 = 17872000km^2.$$

A estimativa da área é 2,099 vezes a área real.

### 3.3.2 Etapa 2: Aplicando o método

Nesta segunda etapa de realização do método de Monte Carlo, usamos o mapa com quadrícula de  $1cm$  e foram repetidos os processos anteriores. Desta vez, cada aluno recebeu 10, 50 e 100 pares de pontos gerados aleatoriamente a partir de  $\sim Unif\{1, 8\}$  para cada eixo.

Temos a distribuição uniforme de  $n$  pontos em uma área de  $64cm^2$  com 64 quadrados de 1 cm de lado, o que implica que cada ponto cobre a área de  $1cm^2$ . Pelos critérios adotados, a proporção de pontos no mapa é 40 quadrados (62,5%). A Tabela 3.5 registra a frequência observada de pontos dentro do mapa para  $n = 10$ .

Tabela 3.5: Frequência de pontos dentro do mapa  $1cm \times 1cm$ ,  $n = 10$ .

Pts Dentro	frequência	# pts $\times$ frequência
3	4	12
4	9	36
5	17	85
6	29	174
7	22	154
8	19	152
9	16	144
10	13	130
Total	129	887

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

O número médio de  $887/129=6,876$  pontos dentro da figura. Usando a relação (1), a estimativa para área do mapa foi

$$\frac{6,876}{10} \approx \frac{A_{mapa}}{64} \Rightarrow A_{mapa} = 64 \times 0,6876 \Rightarrow A_{mapa} \cong 44,006cm^2$$

Usando a relação (2), a estimativa para a área do território brasileiro foi

$$\frac{44,006cm^2}{A_{BRASIL}} = \frac{(7,5cm)^2}{(4,395 \times 10^8cm)^2} \Rightarrow A_{BRASIL} = \frac{44,006cm^2 \times 19,316025 \times 10^{16}cm^2}{56,25cm^2}$$

$$A_{BRASIL} \cong 15,11155 \times 10^{16}cm^2 = 15111550km^2.$$

A estimativa da área é 1,775 vezes a área real.

A Tabela 3.6 registra a frequência para  $n = 50$ .

Tabela 3.6: Frequência de pontos dentro do mapa  $1cm \times 1cm$ ,  $n = 50$ .

Pts Dentro	frequência	#pts $\times$ frequência
25	2	50
26	7	156
28	12	336
29	16	464
30	19	570
31	34	1054
32	17	544
33	13	429
35	6	210
37	3	111
Total	129	3924

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

A proporção média é 30,42 pontos, a área estimada do mapa é  $38,94cm^2$  e a área estimada do território brasileiro,  $13370400Km^2$ , sendo 1,57 vezes a área real.

A Tabela 3.7 registra a frequência para  $n = 100$ .

Tabela 3.7: Frequência de pontos dentro do mapa  $1cm \times 1cm$ ,  $n = 100$ .

Pts Dentro	frequência	#pts $\times$ frequência
56	6	336
57	8	456
58	11	638
59	14	826
60	15	900
61	22	1342
62	27	1674
63	14	882
64	7	448
65	4	260
66	1	66
Total	129	7828

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

A proporção média é 60,682 pontos dentro do mapa, a área estimada do mapa é  $38,83cm^2$  e a área estimada do território brasileiro de  $13336330Km^2$ , sendo 1,566 vezes a área real.

Por fim, usamos o mapa com quadrícula de  $0,5cm$  de lado. Foi entregue, a cada aluno, uma folha com 10, 50 e 100 pares de pontos gerados aleatoriamente de  $\sim Unif\{1, 16\}$  e proposto como atividade de casa, localizar e contar os pontos dentro do contorno do mapa. A folha com o mapa continha 256 quadradinhos de área igual a  $0,25cm^2$ , com 132 quadradinhos dentro do contorno do mapa com proporção de 0,5156. Dos 129 alunos que receberam a folha para realizar a atividade em casa, 93 calcularam o número de pontos dentro do mapa. Fato que pode ter contribuído (ou não) para o resultado, pois quase 20% dos alunos não realizaram a atividade.

A Tabela 3.8 registra a frequência para  $n = 10$ .

Tabela 3.8: Frequência de pontos dentro do mapa  $0,5cm \times 0,5cm$ ,  $n = 10$ .

Pts Dentro	frequência	#pts $\times$ frequência
2	2	4
3	4	12
4	11	44
5	17	85
6	23	138
7	18	126
8	13	104
9	5	45
Total	93	558

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

A proporção média é 6 pontos, a área estimada do mapa é  $38,4cm^2$  e a área estimada do território brasileiro,  $13186410Km^2$ , sendo 1,549 vezes a área real.



A Tabela 3.9 registra a frequência para  $n = 50$ .

Tabela 3.9: Frequência de pontos dentro do mapa  $0,5cm \times 0,5cm$ ,  $n = 50$ .

Pts Dentro	frequência	#pts $\times$ frequência
21	2	42
22	3	66
23	5	115
24	9	216
25	14	350
26	25	650
27	13	351
28	10	280
29	7	203
30	4	120
31	1	31
Total	93	2424

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

A proporção média é 26,065 pontos, a área estimada do mapa é  $33,363cm^2$  e a área estimada de  $11456580Km^2$ , sendo 1,345 vezes a área real.

A Tabela 3.10 registra a frequência para  $n = 100$ .

Tabela 3.10: Frequência de pontos dentro do mapa  $0,5cm \times 0,5cm$ ,  $n = 100$ .

Pts Dentro	frequência	# pts $\times$ frequência
46	4	184
47	6	282
48	9	432
49	12	539
50	17	900
51	23	1173
52	12	728
53	6	424
54	3	216
55	1	55
Total	93	4668

**Legenda:** Pts Dentro = Pontos Dentro; #pts  $\times$  frequência = Número de pontos vezes sua frequência.

A proporção média é 50,19 pontos dentro do mapa, a área estimada do mapa é  $32,12\text{cm}^2$  e a área estimada do território brasileiro de  $11031210\text{Km}^2$ , sendo 1,296 vezes a área real.

Uma estimativa “final” com um erro relativo (diferença entre área estimada e a área real dividida pela área real) de 0,296. Este erro era inicialmente de 1,387 e diminuiu gradativamente conforme a execução do experimento. Lembramos que o objetivo não era obter a área exata, e sim, mostrar a aplicabilidade dos conceitos ministrados em sala de aula e desenvolver a capacidade de usar relações matemáticas.

### 3.4 Conclusão

Os resultados obtidos eram esperados no seguinte sentido, conforme diminuimos o tamanho da quadrícula, também diminuimos a quantidade de quadrados que possuíam apenas uma parte do contorno, e dessa maneira, o erro relativo médio também diminui, pois estamos nos aproximando da proporção real da área da figura (mapa com o contorno do Brasil) em relação à área do quadrado, além disso, quando aumentamos o número de pontos gerados aleatoriamente  $n$  aproximamos a proporção de pontos dentro do mapa da proporção do contorno do mapa em relação ao quadrado (probabilidade geométrica). Ou seja, dado o critério de decisão, era esperado que, quanto menor a área do quadrado da malha mais próximos estaremos da representatividade real do contorno do mapa em relação à área do quadrado no qual a figura está inserida e quanto maior for  $n$ , mais próxima é a razão entre o número de pontos dentro do mapa e a quantidade de pontos gerados e a probabilidade da figura em relação ao todo.

Uma quarta malha feita com a quadrícula de  $0,25\text{cm}$  com 1024 quadrados de área  $0,0625\text{cm}^2$  e um aumento no número de pontos para 200 ou 500, por exemplo, resultaria em uma área estimada mais próxima da área real da superfície do território brasileiro.

Foi muito edificante quanto profissional, a realização deste experimento. Primeiro, pela quantidade de conteúdos envolvidos, promovendo um crescimento intelectual e crítico nos alunos, além de mostrar aplicabilidade dos conceitos aprendidos em aula. Também despertou curiosidade em outros professores. Um exemplo é que em meio aos comentários sobre o experimento, um professor indagou se o método seria capaz de determinar de forma aproximada a área de uma ameba. Respondi que provavelmente sim, porém deveríamos adotar critérios e relações microscópicas para tal estimação e fazer medições com instrumentos apropriados para não cometer grandes erros.

A conclusão final é que ambos os fatores, tamanho da quadrícula e número de pontos gerados influenciam diretamente para a obtenção de uma estimativa melhor.

## 3.5 Complementar

### 3.5.1 Área com integral

Neste capítulo apresentamos um método estatístico para encontrar uma estimativa para a área de qualquer figura inserida em um plano. Nesta seção veremos uma aplicação do cálculo integral definida para a obtenção de área de uma região limitada e volume de sólidos de revolução <sup>3</sup>. O caso mais trivial que podemos ter é uma função contínua  $f : [a, b] \rightarrow \mathbb{R}$  tal que  $f(x) \geq 0, \forall x \in [a, b]$ , então a área dessa região  $\mathcal{R}$  delimitada pelo gráfico e o eixo  $X$  é dada por

$$\mathcal{R} = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, 0 \leq y \leq f(x)\} = \int_a^b f(x)dx$$

O resultado que determina o valor desta área e conhecido como **Teorema Fundamental do Cálculo (TFC)**.

### 3.5.2 Teorema Fundamental do Cálculo

**Teorema Fundamental do Cálculo.** *Seja uma função contínua  $f : [a, b] \rightarrow \mathbb{R}$ , então*

$$\int_a^b f(x)dx = F(b) - F(a) = F(x)|_a^b$$

*Sendo  $F(x)$  uma primitiva qualquer de  $f(x)$ .*

Porém, na grande maioria dos problemas que envolvem o cálculo de área a aplicação direta do *TFC* não resolve diretamente o problema, pois geralmente temos uma região limitada por duas funções  $f$  e  $g$  contínuas em um intervalo  $[a, b]$ , nestes casos temos outro Teorema nos ajuda a calcular a área compreendida entre estas funções.

**Teorema.** *Sejam  $f$  e  $g$  duas funções contínuas no intervalo  $[a, b]$ , tais que  $f(x) \geq g(x), \forall x \in [a, b]$ , então a área dessa região  $\mathcal{R}$  delimitada por*

$$\mathcal{R} = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, g(x) \leq y \leq f(x)\} = \int_a^b (f(x) - g(x))dx$$

*Usando as propriedades de integral [34], temos*

$$\mathcal{R} = \int_a^b f(x) - \int_a^b g(x)dx = F(x)|_a^b - G(x)|_a^b = F(b) + G(a) - [G(b) + F(a)]$$

---

<sup>3</sup>Veja mais em [34] e [35]

A Figura 3.5, extraída de <http://goo.gl/GXAWr1>, ilustra a situação do Teorema.

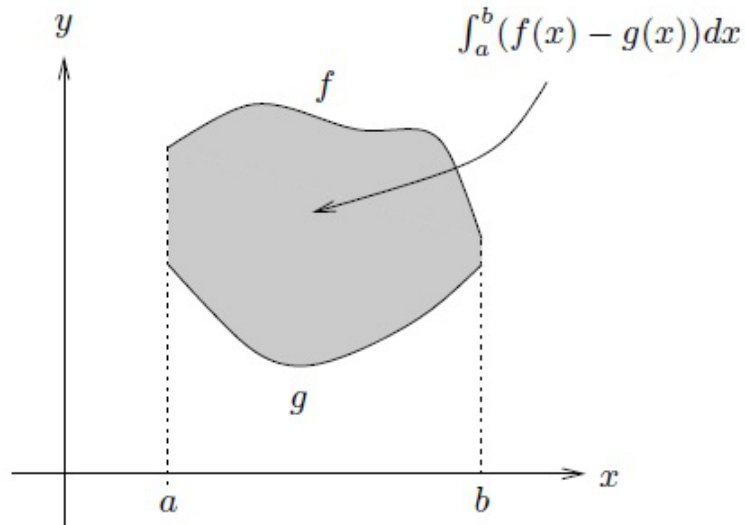


Figura 3.5: Cálculo de área com  $f \geq g$

Nos casos ilustrados pela Figura 3.5, os valores extremos da integral  $a$  e  $b$ , geralmente precisam ser determinados. Nestes casos fazemos  $f(x) = g(x)$  e encontramos os dois valores de  $x$  para os quais a igualdade é verdadeira. Caso não esteja claro qual a função  $f$  e qual a função  $g$ , ou seja, qual função está acima e qual está abaixo. Podemos “testar” substituindo em ambas qualquer valor de  $x \in (a, b)$ , como por exemplo  $(a + b)/2$ , a função com maior valor de  $y$  será  $f$ .

Existem casos onde  $(f - g)$  muda de sinal, ou seja, ora  $(f - g) \geq 0$  ora  $(f - g) < 0$ . Nesses casos precisamos determinar os valores onde ocorrem as mudanças de sinal. Para determinar estes valores com mais facilidade, um esboço confiável da região  $\mathcal{R}$  pode ajudar na determinação destes, geralmente quando fazemos  $f = g$  solucionamos. De posse destes valores, dividimos a região  $\mathcal{R}$  em subregiões  $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n$  e o intervalo  $[a, b]$  em subintervalos  $I_1, I_2, \dots, I_n$  e escrevemos em  $n$  integrais simples considerando em cada um a diferença entre as funções e seu intervalo de integração.

A Figura 3.6, adaptada de <http://goo.gl/GXAWr1>, exemplifica esta situação.

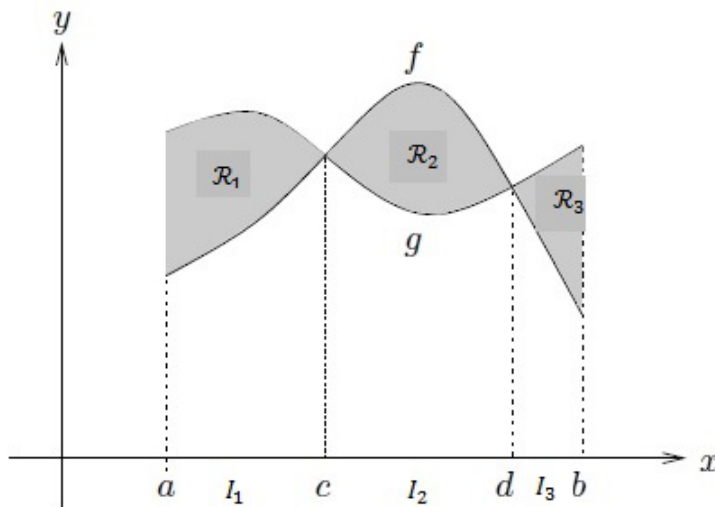


Figura 3.6: Cálculo de área com sinal  $f - g$  variando

Pela Figura 3.6, em  $\mathcal{R}_1$  temos  $g \geq f$  e  $I_1 = (a, c)$ ; em  $\mathcal{R}_2$  temos  $f \geq g$  e  $I_2 = (c, d)$ ; em  $\mathcal{R}_3$  temos  $g \geq f$  e  $I_3 = (d, b)$ . Logo,

$$\begin{aligned}\mathcal{R} &= \int_{I_1} \mathcal{R}_1 dx + \int_{I_2} \mathcal{R}_2 dx + \int_{I_3} \mathcal{R}_3 dx = \\ \mathcal{R} &= \int_a^c (g(x) - f(x)) dx + \int_c^d (f(x) - g(x)) dx + \int_d^b (g(x) - f(x)) dx = \\ \mathcal{R} &= [G(c) + F(d) + G(b)] - [F(a) + G(c) + F(d)]\end{aligned}$$

Compreender o comportamento das funções pode facilitar muitas integrais de áreas, principalmente quando o intervalo é simétrico em torno de 0. Mais especificamente, temos que:

Uma função  $f$  é dita função **PAR**, se  $f(x) = f(-x)$ ,  $\forall x \in D_f$ , (Domínio da função  $f$ ), mais claramente, se a função for simétrica em relação ao eixo  $Y$ .

**Exemplos:**  $f(x) = x^2$ ;  $f(x) = \cos x$ ;  $f(x) = |x|$ .

Uma função  $f$  é dita função **ÍMPAR**, se  $f(x) = -f(-x)$ ,  $\forall x \in D_f$ , mais claramente, se a função for simétrica em relação ao ponto de origem do plano  $XY$ .

**Exemplos:**  $f(x) = \pm x$ ;  $f(x) = x$ ;  $f(x) = x^3$ .

Para funções com estas características temos resultados importantes quando elas são integradas em um intervalo simétrico a zero.

- Seja  $f$  uma função par, então

$$\int_{-a}^a f(x)dx = 2 \times \int_0^a f(x)dx$$

- Seja  $f$  uma função ímpar, então

$$\int_{-a}^a f(x)dx = 0$$

### 3.5.3 Volume com integral

Podemos calcular volume com integral definida simples <sup>4</sup>, desde que seja de um sólido de revolução em torno do eixo  $X$  (ou do eixo  $Y$ ).

Seja uma função contínua  $f : [a, b] \rightarrow \mathbb{R}$  tal que  $f(x) \geq 0, \forall x \in [a, b]$ . Seja  $\mathcal{V} \subset \mathbb{R}^3$  o volume do sólido gerado a partir da revolução de  $f$  em torno do eixo  $X$ , então o volume deste sólido  $\mathcal{V}$  é delimitada por

$$\mathcal{V} = \{(x, y) : a \leq x \leq b, 0 \leq y \leq f(x)\}$$

**Teorema.** *O volume  $\mathcal{V}$  de um sólido de revolução em torno do eixo  $X$  gerado a partir de  $f(x) = y$  é dado por*

$$\mathcal{V} = \pi \times \int_a^b f^2(x)dx$$

---

<sup>4</sup>Em problemas que consideramos sólidos, que não são de revolução, no  $\mathbb{R}^3$ , o volume é calculado por integral dupla ou até tripla, veja mais em [36]

A Figura 3.7, retirada de <http://goo.gl/GXAWr1>, é um exemplo de sólido de revolução.

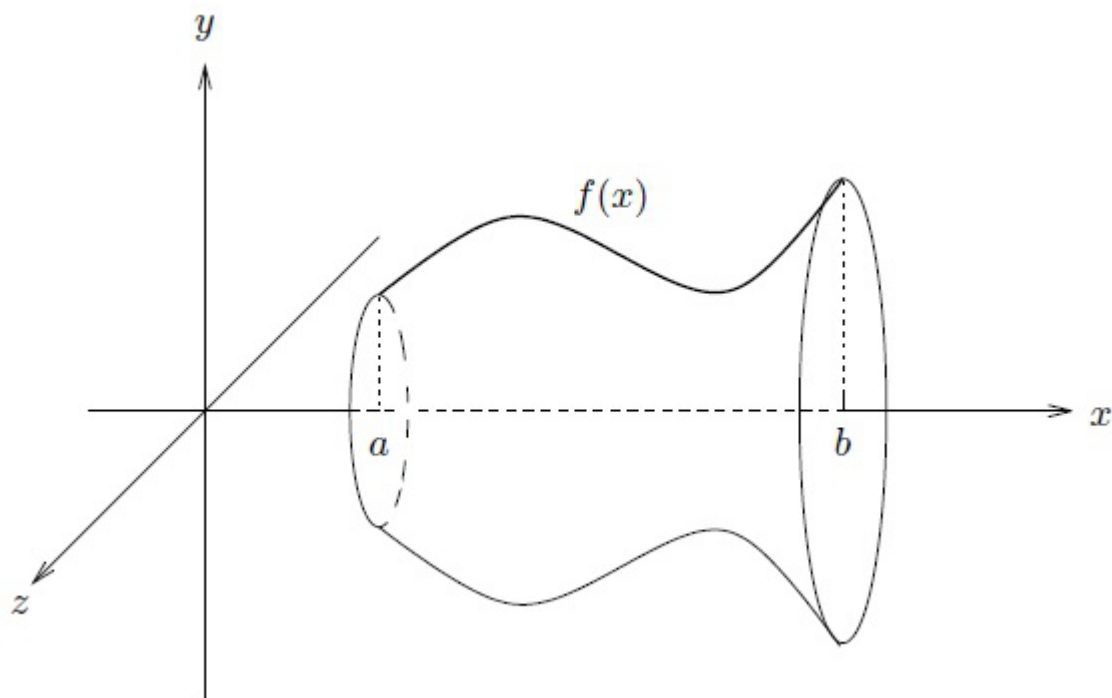


Figura 3.7: Sólido de revolução





# Capítulo 4

## Variáveis antropométricas

### 4.1 O experimento

#### Sinopse

Neste experimento, os alunos coletaram dados a partir de diversas medições de partes do corpo para fazerem análise sobre possíveis relações entre tais medidas através da construção de tabelas e gráficos.

#### Conteúdos

Representação gráfica e correlação de duas variáveis quantitativas.

#### Objetivos do experimento

1. Desenvolver no aluno a capacidade de extrair informações a partir de dados brutos;
2. Estimular o uso de tabelas e gráficos de frequências para analisar dados bivariados sobre características físicas.

### 4.2 Introdução teórica

Muito se diz sobre relações existentes entre as medidas do nosso corpo. Por exemplo, ouvimos que, quanto maior a altura de uma pessoa maior será sua envergadura e a medida do antebraço de uma pessoa é aproximadamente igual à medida de seu

pé. Neste experimento, os alunos fizeram, supervisionados pelo professor, uma investigação de algumas suposições sobre as medidas de seu corpo usando ferramentas estatísticas. Propomos a investigação do comprimento do dedo indicador, circunferência do pescoço e do quadril, tamanho do braço e da perna e o número do calçado e a altura do indivíduo.

Para estudar uma população é comum realizar uma amostragem escolhendo indivíduos desta população visando medir e estudar as características de interesse e, através de uma análise, inferir conclusões para a população inteira. A primeira pergunta que devemos fazer é se realmente conseguimos, a partir dessa amostra, informações precisas sobre a população, ou seja, se as características observadas na amostra podem ser generalizadas para toda população [8].

Neste experimento, o aluno realizará uma amostragem com o intuito de analisar o comportamento de variáveis antropométricas e as possíveis relações entre elas. Algumas perguntas que motivaram este experimento foram:

1. Pessoas mais altas têm maior número de calçado?
2. Existe alguma relação entre o comprimento do dedo indicador e a altura de uma pessoa?
3. Como verificar graficamente a existência ou a estrutura das relações sugeridas pelas perguntas anteriores?
4. Como quantificar as evidências obtidas pela amostra a respeito dessas relações?

Na primeira etapa deste experimento, os alunos fizeram a medição das partes do corpo citadas anteriormente. Na segunda etapa, analisaram os dados obtidos utilizando uma folha quadriculada, entregue a cada um. As análises aqui inseridas foram feitas no pacote estatístico  $R$  [17].

A análise gráfica dos valores da amostra permite verificar visualmente a possível existência de alguma relação entre as variáveis medidas. Por exemplo, nos dados coletados é possível observar que:

- Quanto maior for o valor da variável  $X$ , maior será o valor da variável  $Y$ ?

*ou*

- Quanto maior for o valor da variável  $X$ , menor será o valor da variável  $Y$ ?

Estas questões permitem identificar se as variáveis  $X$  e  $Y$  são ou não correlacionadas. Se a resposta da primeira questão for positiva, dizemos que as variáveis são positivamente correlacionadas, se a resposta da segunda questão for afirmativa, as variáveis são negativamente correlacionadas.

Intuitivamente, se houver correlação entre as variáveis, ao saber o valor de uma destas teríamos uma boa informação para prever o valor da outra. Uma correlação alta nos permite utilizar uma das variáveis para previsão ou diagnóstico da outra.

Outra característica que pode ser observada é se a variabilidade da variável  $Y$  depende (ou não) da região considerada de valores da variável  $X$ . Esta variabilidade permite identificar a qualidade da previsão que podemos fazer da variável  $Y$ , conhecendo o valor da variável  $X$ . Por conseguinte, se a variabilidade de  $Y$ , dado  $X$ , for pequena, a previsão pode ser bastante precisa.

Um modelo matemático para uma relação entre duas variáveis é o modelo linear, sendo que os valores de  $X$  e  $Y$  podem ser vistos como flutuações em torno de uma reta de regressão. O ajuste linear, ou seja, a qualidade do modelo, pode ser quantificado pelo coeficiente de correlação linear amostral  $r$ , pois este é calculado levando em conta as posições relativas de cada ponto amostral em relação ao ponto de médias, correspondente ao centro de massa da nuvem de pontos.

Como as ordens de grandeza e a unidade de medida das variáveis podem ser diferentes, é conveniente padronizá-las considerando seus valores em termos do desvio padrão. Dada uma variável  $X$ , definimos seu desvio padrão como [6]:

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu_X)^2}{n}}$$

sendo  $n$  o tamanho da amostra.

A transformação

$$Z_X = \frac{X - \mu_X}{\sigma_X}$$

sendo  $\mu_X$  e  $\sigma_X$ , respectivamente, a média e desvio padrão dos valores da variável  $X$ , é chamada escala padrão de  $X$ . Com esta transformação, as variáveis ficam na mesma escala e sem unidade. Dizemos que a variável  $X$  dessa forma está padronizada.

O coeficiente de correlação linear  $r_{X,Y}$  entre duas variáveis  $X$  e  $Y$  é o valor médio dos produtos entre as coordenadas dos pontos observados em escala padrão, ou seja,

$$r_{X,Y} = \frac{1}{n} \times \sum_{i=1}^n Z_X Z_Y = \frac{1}{n} \sum_{i=1}^n \left( \frac{(X_i - \mu_X)}{\sigma_X} \times \frac{(Y_i - \mu_Y)}{\sigma_Y} \right)$$

O valor de  $r$  para duas variáveis numéricas  $(X, Y)$  quaisquer varia de -1 a 1. Intuitivamente, quanto  $r$  for mais próximo de 1 (ou -1) melhor é a aproximação linear

e quanto mais próximo de zero, menos precisa é a relação linear entre as variáveis. No entanto, neste tipo de análise, o gráfico deve ser a principal ferramenta de decisão, deixando que o valor de  $r$  apenas corrobore o que percebemos visualmente.

Outras maneiras de escrevermos  $r_{X,Y}$  são

$$r_{X,Y} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sigma_X\sigma_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \times \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

A correlação é quantificada de acordo com valores numéricos usualmente utilizados [12] :

1. Se  $|r| \in (0; 0,3)$  as variáveis possuem correlação fraca;
2. Se  $|r| \in [0,3; 0,7)$  as variáveis possuem correlação moderada;
3. Se  $|r| \in [0,7; 1)$  as variáveis possuem correlação forte;

Temos o Resultado 1 acerca do coeficiente de correlação linear  $r$ .

**Resultado 1.** *O coeficiente de correlação linear  $r$  é adimensional, ou seja, não varia se mudarmos escala ou locação.*

**Demonstração.** *Primeiro vamos escrever a expressão do coeficiente de correlação linear  $r$  com variações  $a, b \neq 0$  em  $X$  e  $Y$ , respectivamente.*

$$r_{aX,bY} = \frac{1}{n} \sum_{i=1}^n \left( \frac{(aX_i - a\mu_X)}{a\sigma_X} \times \frac{(bY_i - b\mu_Y)}{b\sigma_Y} \right)$$

$$r_{aX,bY} = \frac{1}{n} \sum_{i=1}^n \left( \frac{a(X_i - \mu_X)}{a\sigma_X} \times \frac{b(Y_i - \mu_Y)}{b\sigma_Y} \right) \quad (3)$$

como  $a, b \neq 0$ , podemos eliminá-los e teremos que:

$$r_{aX,bY} = \frac{1}{n} \sum_{i=1}^n \left( \frac{(X_i - \mu_X)}{\sigma_X} \times \frac{(Y_i - \mu_Y)}{\sigma_Y} \right) = r_{X,Y}$$

Precisamos deixar claro que o **coeficiente de correlação  $r$  é limitado à linearidade**, ou seja, podemos ter duas variáveis com uma forte associação e  $r = 0$ .

Para compreendermos o coeficiente  $r$  é preciso associar a um resultado da geometria sobre o cosseno entre o ângulo entre dois vetores. Sejam  $X = (x_1, x_2, \dots, x_n)$  e  $Y = (y_1, y_2, \dots, y_n)$  dois vetores no  $\mathbb{R}^n$ , o cosseno do ângulo  $\alpha$  entre estes dois vetores é dado por [33]

$$\cos \alpha = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \times \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Este resultado para o cosseno do ângulo entre dois vetores é justamente o valor de  $r_{X,Y}$  apresentado em (3), então pensemos no que é a função cosseno.

Se  $\cos \alpha = 0$ , então estes vetores são perpendiculares em  $\mathbb{R}^2$ , e ortogonais no  $\mathbb{R}^n, n \geq 3$ , isto significa que estes vetores possuem um ângulo de  $90^\circ$ , ou seja, qualquer variação em um destes vetores não acarreta variação no outro. Pensando no plano cartesiano, por exemplo, temos os vetores  $v_1 = (x, 0)$  e  $v_2 = (0, y)$ , ou seja, alterar o valor da abcissa  $x$  em  $v_1$  nada produz em  $v_2$ .

Se  $\cos \alpha = 1$ , os dois vetores, então estes vetores são colineares em  $\mathbb{R}^2$ , e coplanares no  $\mathbb{R}^n, n \geq 3$ , isto significa que estes vetores possuem um ângulo de  $0^\circ$ , ou seja, qualquer variação em um destes vetores acarreta em uma variação de mesma magnitude no outro,  $\Delta_X = \Delta_Y$ . Pensando em variáveis,  $r = \cos \alpha$  determinada qual a taxa variação unidimensional (linear) ocorre em  $Y$  ao variarmos em uma unidade a variável  $X$ .

Um exemplo, dentre vários, é uma função (variável)  $Y = f(X) = X^2$ , totalmente explicada pela variável  $X$ , porém possui  $r = 0$ , porque  $r$  tem como finalidade determinar uma reta com ângulo  $\alpha$  em relação ao eixo  $X$  que associe a variação em  $Y$  dada a variação de  $X$ . Neste exemplo, esta reta seria o eixo  $Y$  é perpendicular ao eixo  $X$ , já que  $Y = X^2$  é uma função de  $X$  simétrica ao eixo  $Y$ , ou seja, o valor de  $r = 0$  diz que  $X$  e  $Y$  não possuem relação alguma e que variando  $X$  nada acontece com o valor de  $Y$  o que é contraditório, pois  $Y = X^2$ , e portanto, função de  $X$ .

Portanto, ressaltamos que o coeficiente de correlação linear  $r = 0 \nRightarrow Y \neq f(X)$ .

Usaremos a notação acadêmica para as estatísticas descritivas neste capítulo. Apresentaremos a notação acadêmica para estatísticas de ordem e que iremos adotar das páginas seguintes.

Considere  $x_1, x_2, \dots, x_n$  uma amostra de uma variável aleatória  $X$ . Denotamos essa amostra ordenada da seguinte forma:

$$\begin{aligned} x_{(1)} &= \min\{x_1, x_2, \dots, x_n\}; \\ x_{(2)} &= \text{segundo menor valor do conjunto } \{x_1, x_2, \dots, x_n\}; \\ &\vdots \\ x_{(n-1)} &= \text{segundo maior valor do conjunto } \{x_1, x_2, \dots, x_n\}; \\ x_{(n)} &= \max\{x_1, x_2, \dots, x_n\}. \end{aligned}$$

Para as estatísticas amostrais usaremos

$\bar{x}$  para a média;

$\tilde{x}$  para a mediana;

$s$  para o desvio padrão;

$q_1$  para o primeiro quartil;

$q_3$  para o terceiro quartil.

Usaremos o coeficiente de variação  $C_V$  que é uma medida de dispersão relativo à média, calculada pela razão entre o desvio padrão da variável e sua média. Este coeficiente é usado para comparar a variabilidade em torno da média para duas variáveis com média, desvio padrão ou unidade de medida distintos, e portanto não comparáveis [6]. O valor de  $C_V$  indica o comportamento da variabilidade (ou precisão experimental) e independe da unidade de medida. O  $C_V$  um importante indicador de dispersão, pois desvios só podemos comparar a e avaliar variabilidade de desvios se estes estão na mesma grandeza. Usamos a seguinte fórmula para obtermos  $C_V$

$$C_V = \frac{\sigma}{\mu} (Populacional); \quad C_V = \frac{\bar{s}}{S} (Amostral);$$

Multiplicando  $C_V$  por 100, obtemos a variabilidade em porcentagem (%).

## 4.3 Execução

Para a coleta dos dados neste experimento foram usadas fitas métricas. Visando maior rapidez, os alunos foram divididos em grupos menores. As análises foram feitas considerando todos os alunos que participaram do experimento.

As mensurações que os alunos registraram em uma tabela são gênero, altura, circunferências do pescoço e do quadril, comprimentos do dedo indicador direito, do braço e da perna e número do calçado. Nesta primeira parte, foi necessário cuidado para evitar erros nas medições. Uma precaução tomada foi pedir que todos os integrantes do grupo realizassem cada mensuração nos outros integrantes. Havendo discrepância entre as medições, o professor, tomando os cuidados necessários, realiza outra medição e considera valores centrais das medidas como sendo a mais próxima da realidade. É importante ressaltar, que as medidas limitam-se a centímetros e milímetros e feitas com instrumentos de medição que, eventualmente, possam ter alguma diferença de um para outro. A ideia foi supervisionar todas as medições e arredondamentos para que os alunos não usassem apenas centímetros, mas também milímetros.

Depois das medições, preenchemos a Tabela 4.1, para organizarmos as informações das respectivas variáveis a fim de produzir os gráficos e as tabelas para análise descritiva das mensurações.

Tabela 4.1: Tabela com referidas mensurações dos alunos.

	Gênero		medidas (em centímetros)						
N.º	F	M	Altura	Pescoço	Cós	Indicador	Perna	Braço	Calçado
1									
2									
3									
...									
40									

**Legenda:** F=Feminino; M=Masculino.

## 4.4 Análise das Informações

Para início e auxílio da análise, fizemos a Tabela 4.2 com estatísticas descritivas das variáveis mensuradas, considerando valor mínimo ( $x_{(1)}$ ), primeiro quartil ( $q_1$ ), mediana ( $\tilde{x}$ ), média ( $\bar{x}$ ), terceiro quartil ( $q_3$ ), valor máximo ( $x_{(n)}$ ) e desvio padrão ( $s$ ) de cada uma. Note que  $q_1$  concentra 25% dos valores amostrados, dispostos em ordem crescente, abaixo de seu valor, e  $q_3$  apresenta 75% dos valores nestas condições.

Tabela 4.2: Estatísticas descritivas das variáveis mensuradas,  $n = 135$ .

	Altura	Cós	Pescoço	Braço	Perna	Dedo	Calçado
$x_{(1)}$	148,7	54,2	26,2	37,4	58,8	6,1	34
$q_1$	156,8	67,6	30,7	46,8	70,9	7,2	36
$\tilde{x}$	158,9	74,3	33	48,6	75,6	7,6	37
$\bar{x}$	159,7	75,5	32,99	49,05	75,94	7,581	37
$q_3$	162,6	83,05	35,65	51,25	80,1	7,9	38
$x_{(135)}$	175,2	108,5	40,3	57,6	93,1	9,2	42
$s$	4,4842	10,8221	3,3312	3,7895	6,7609	0,6396	1,5835
$C_V$	0,0281	0,1433	0,1010	0,0773	0,0890	0,0844	0,0428

A Tabela 4.2 descreve as variáveis mensuradas de forma qualitativa. Notamos que a altura é a variável que possui menor variabilidade entre as mensuradas, pois

possui o menor valor para  $C_V$  enquanto a circunferência do quadril (cós) é que possui a maior variabilidade, superior a 14%. Os dados desta tabela permite a descrição e compreensão de diversas características em cada uma das variáveis. Podemos calcular também a **amplitude** de cada uma delas fazendo  $x_{(135)} - x_{(1)}$ .

Fizemos a análise gráfica dos valores amostrais para verificar visualmente a possível existência de alguma relação entre as variáveis mensuradas.

O primeiro passo para analisar os dados, foi pedir aos alunos para que construíssem um dos gráficos bidimensionais, usando o quadriculado entregue no início da aula. Para tal construção, foi ensinado como definir a escala de cada eixo e um melhor valor para o ponto no qual os eixos se cruzam.

Para ilustração foi feito um gráfico relacionando circunferência do pescoço e comprimento do braço com a origem do sistema de eixos no ponto  $(0, 0)$  e uma escala de 5 em 5 *cm* para os dois eixos, mostrando a difícil localização dos pontos.

Por fim, foi feito um segundo gráfico também relacionando circunferência do pescoço e comprimento do braço com escalas adequadas, facilitando a leitura, entendimento e análise dos dados. As representações gráficas estão nas Figuras 4.1 e 4.2.

Vale ressaltar que no segundo gráfico, a origem foi fixada no ponto  $P = (26, 42)$  com uma escala de 2*cm* para cada eixo, o que melhorou a visibilidade da dispersão dos pontos.

Feitos os gráficos, pedimos aos alunos que tentassem inferir se as variáveis possuem alguma relação, ressaltando que as relações podem ser:

- Se as grandezas são relacionadas positivamente, então quanto maior o valor de uma variável  $X$ , maior será o valor da variável  $Y$ ;
- Para grandezas relacionadas negativamente, teremos que quanto maior for o valor da variável  $X$ , menor será o valor da variável  $Y$ ;
- A variabilidade da variável  $X$  depende ou não da região considerada da variável  $Y$ ; ou seja, os valores da variável  $X$  estão mais dispersos ou concentrados em alguma região da variável  $Y$ .



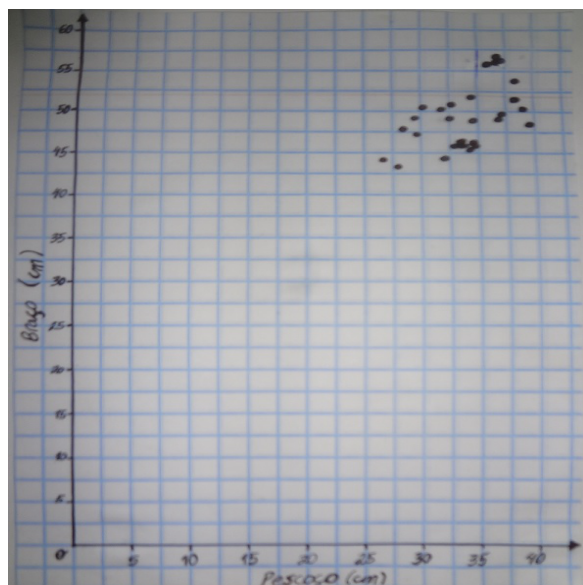


Figura 4.1: Pescoço  $\times$  braço na origem e escala  $5cm$ ,  $n = 32$ .

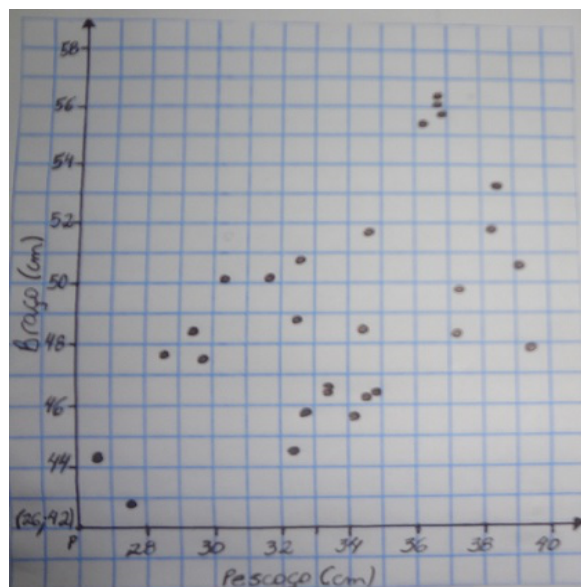


Figura 4.2: Pescoço  $\times$  braço no ponto  $P$  e escala  $2cm$ ,  $n = 32$ .

Com as informações da Tabela 4.1 foram feitos seis gráficos, com o eixo  $X$  sendo a altura dos alunos e o eixo  $Y$  representado por cada uma das demais variáveis. A grade do eixo  $X$  foi fixada em 5  $cm$ . As outras variáveis tiveram a escala conforme o *default* (padrão) do programa  $R$  que também determina o ponto de interseção dos eixos. Também calculamos o coeficiente de correlação  $r$  em cada caso.

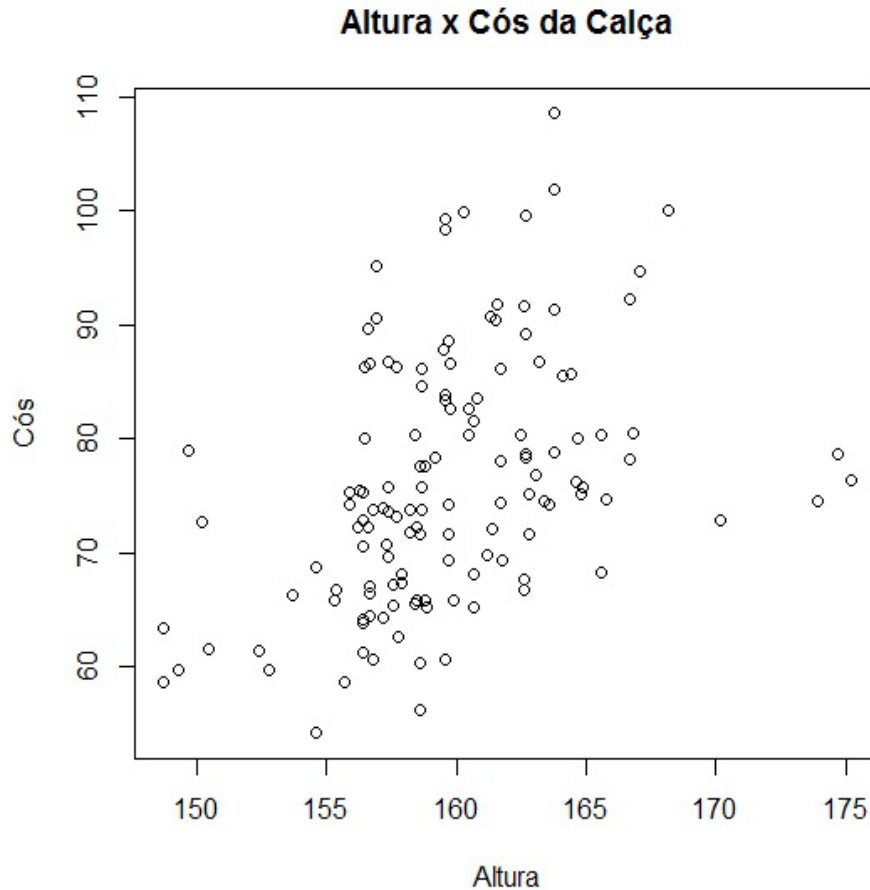


Figura 4.3: Altura  $\times$  circunferência do quadril, em  $cm$ , com  $n = 135$ .

O gráfico da Figura 4.3, altura por circunferência do quadril, visualmente apresenta uma tendência de correlação positiva entre as variáveis. Nota-se também, que existem alunos de baixa estatura e grande circunferência do quadril e alunos altos e com pequena circunferência do quadril. O valor do coeficiente de correlação  $r$  para altura e circunferência do quadril foi 0,3723 entregando evidências da existência correlação moderada positiva, conforme critérios de classificação da apresentados na página 96, entre as variáveis altura e circunferência do quadril.

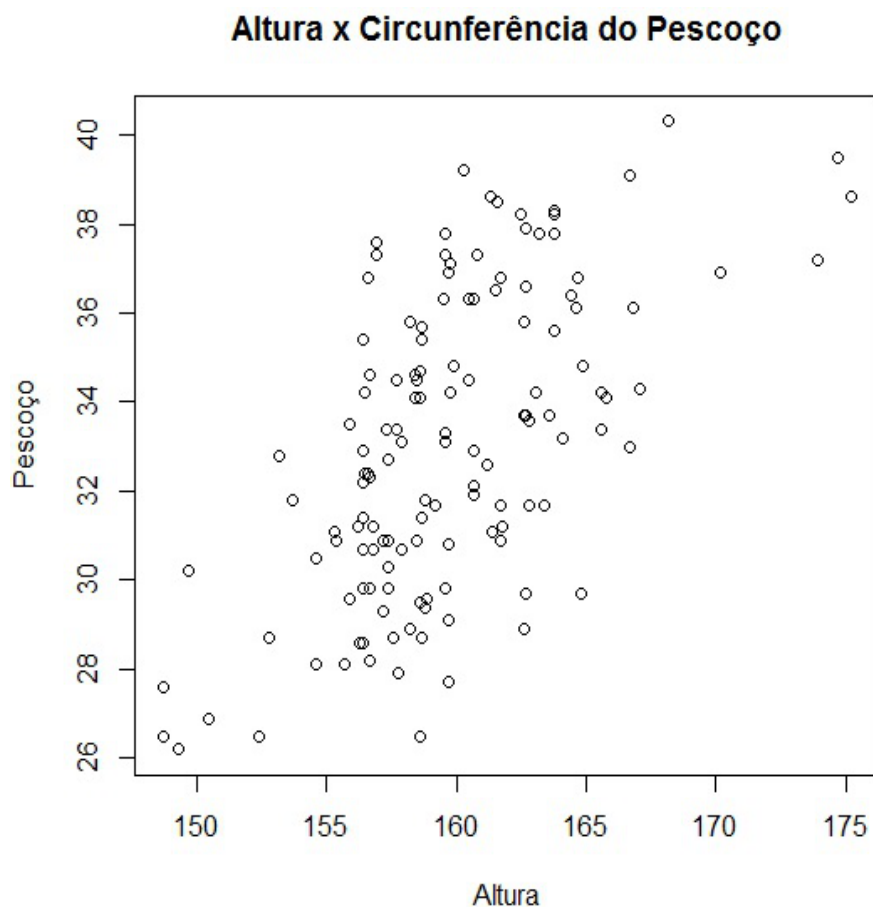


Figura 4.4: Altura  $\times$  circunferência do pescoço, em *cm*, com  $n = 135$ .

Na Figura 4.4, referente à altura e circunferência do pescoço, o gráfico também apresenta uma tendência de correlação positiva. O valor do coeficiente de correlação  $r$  para altura e circunferência do pescoço foi 0,527; temos aqui que a amostra entrega evidências da existência de uma correlação moderada positiva entre a altura e a circunferência do pescoço. Este resultado mostra que quanto mais alta for a pessoa, maior tende a ser a circunferência de seu pescoço.

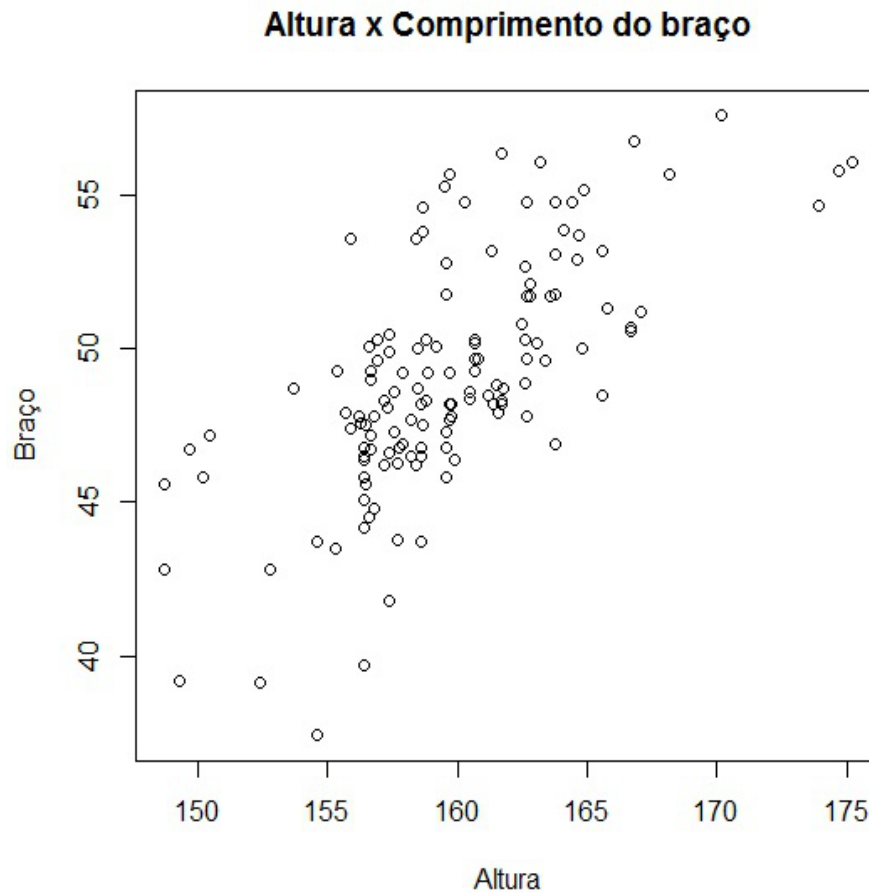


Figura 4.5: Altura  $\times$  comprimento do braço, em *cm*, com  $n = 135$ .

Na Figura 4.5, que relaciona altura e comprimento do braço, observamos novamente uma tendência de correlação positiva entre as variáveis. O valor do coeficiente de correlação linear  $r = 0,685$  para altura e comprimento do braço, novamente a amostra entrega evidências de uma correlação moderada positiva <sup>1</sup> positiva entre as variáveis. Ou seja, temos evidências que a altura de uma pessoa está relacionada ao comprimento do braço, em outras palavras, quanto mais alta a pessoa, maior tende a ser o comprimento do seu braço.

---

<sup>1</sup>Este valor de  $r$  pode caracterizar uma correlação forte em outros critérios de classificação.

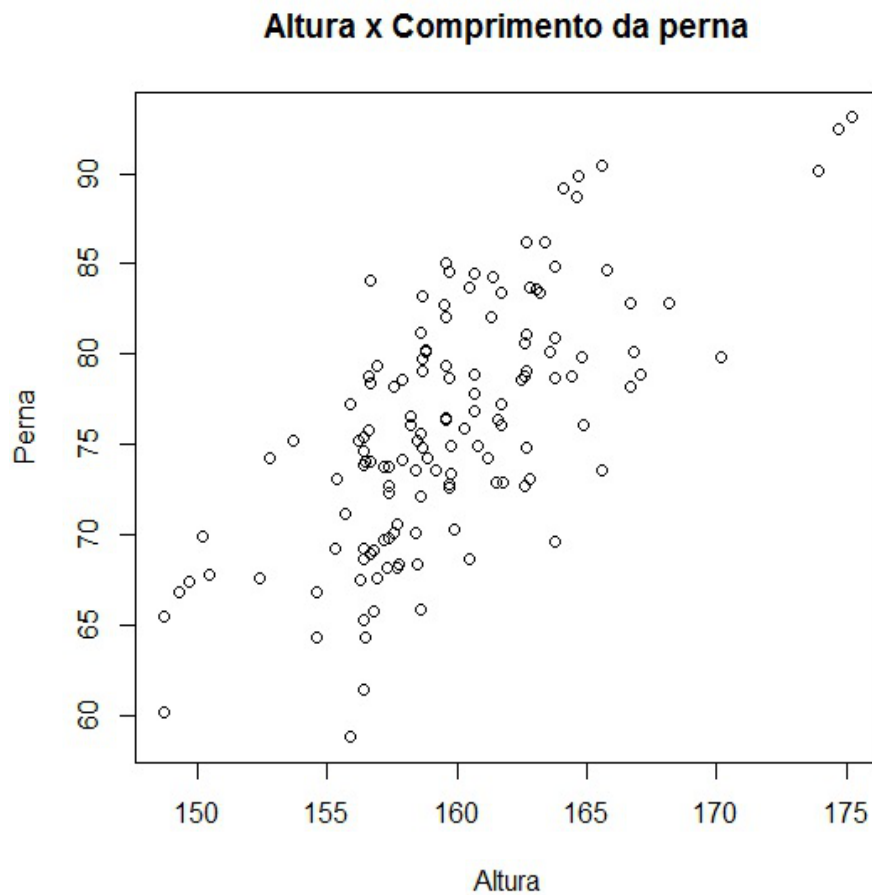


Figura 4.6: Altura  $\times$  comprimento da perna, em *cm*, com  $n = 135$ .

Na Figura 4.6, que relaciona as variáveis altura e comprimento da perna, podemos observar que as variáveis tendem a possuir uma correlação positiva. O valor do coeficiente de correlação linear  $r$  para altura e comprimento da perna foi 0,697. Desta forma, a amostra entrega evidências da existência de uma correlação moderada positiva entre as variáveis. Ou seja, temos evidências para acreditar que quanto mais alta for a pessoa, maior será o comprimento do sua perna.

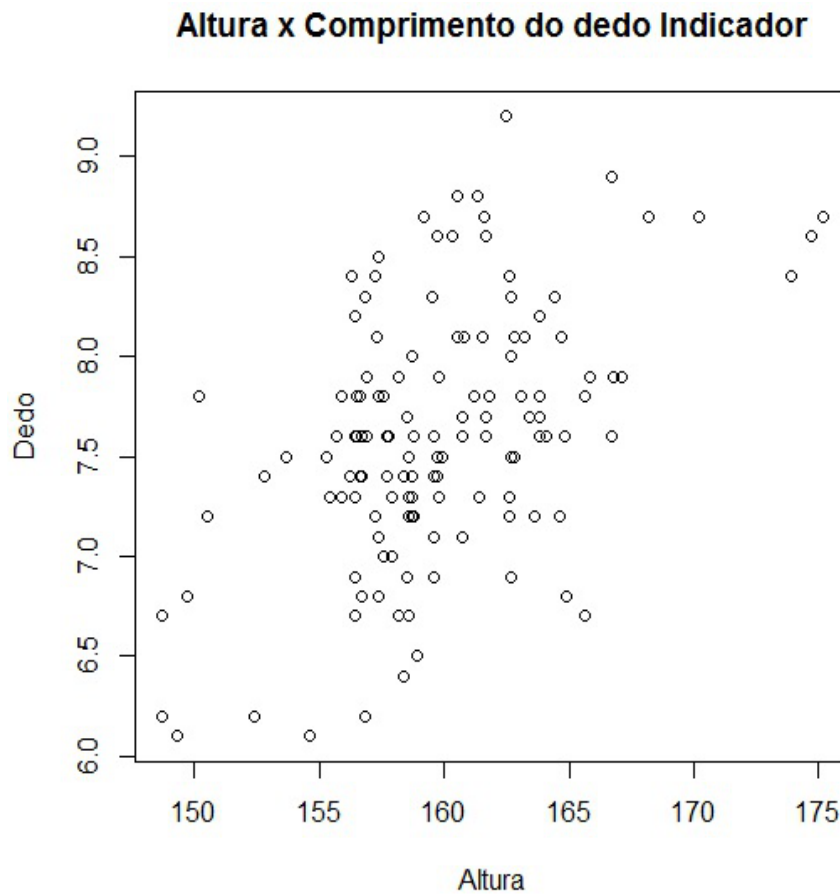


Figura 4.7: Altura  $\times$  comprimento do dedo indicador, em *cm*, com  $n = 135$ .

Pela análise da Figura 4.7, temos indícios de uma associação linear entre as variáveis altura e comprimento do dedo indicador. O valor do coeficiente de correlação linear  $r$  para estas variáveis foi 0,5002, sendo assim, há evidências de uma correlação linear moderada positiva entre altura e comprimento do dedo indicador de uma pessoa. Novamente, os resultados indicam que quanto mais alta for uma pessoa, maior será o seu dedo indicador.

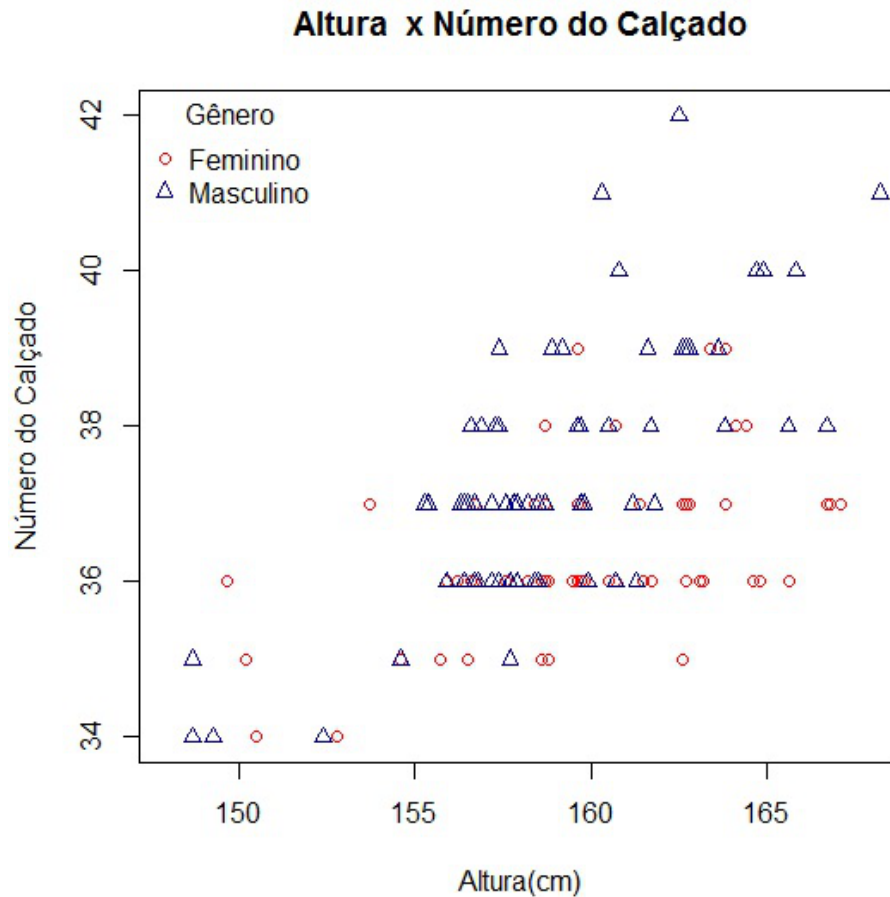


Figura 4.8: Altura  $\times$  número do calçado considerando o gênero, com  $n = 135$ .

A análise visual da Figura 4.8, altura por número do calçado, também mostra que as variáveis possuem associação de linear positiva. Notamos também que para uma mesma faixa de altura, de 155 a 165 cm, temos cinco numerações de calçados diferentes (do 35 ao 39), sendo os números 35, 36 e 37 são mais usados por garotas e 37, 38 e 39 usados por garotos. O valor observado de  $r$  para altura e número do calçado foi 0,6208, sendo assim, a amostra entrega evidências de correlação moderada positiva entre a altura e o número do calçado de uma pessoa e pelos resultados, temos que quanto mais alta for uma pessoa, maior será o número de seu calçado.

## 4.5 Conclusão

O foco deste experimento foi o uso de tabelas e gráficos para tratar as informações coletadas, discutindo com os alunos a utilidade destas ferramentas. Terminada a Etapa 2, foram feitas as conclusões, junto aos alunos que participaram do estudo, sobre as relações entre as variáveis mencionadas, incentivando a participação para que os mesmos entendessem como chegar a conclusões através da análise gráfica. A aplicação do experimento contribuiu para o trabalho de construção de gráficos, como tratar os eixos, escolher escala, determinar o ponto de cruzamento entre os eixos, mostrando a importância destes detalhes na construção de um gráfico.

Para análise final foi utilizado o software *R*. Primeiramente os dados foram inseridos no *Excel* e depois importados para o *R*. Porém o conjunto de dados possui números decimais que, no *Excel* são inseridos usando a **vírgula**, por exemplo,  $A, B$ , sendo  $A$  é a parte inteira e  $B$  a parte decimal. Já no *R* são inseridos com a utilização do **ponto**, sendo  $A.B$ .

Abaixo segue um “tutorial” de importar dados inseridos no *Excel* para o *R* e os comandos utilizados para criar os gráficos deste capítulo.

### 4.5.1 Importando um conjunto de dados com números decimais do *Excel* para o *R*

Com o conteúdo da planilha selecionado, clique na *Barra de Ferramentas*, **Salvar como** → **Outros Formatos** → **Tipo:** → **Texto (separado por tabulações)**; Nomeie o arquivo e selecione uma Pasta ou Local de Destino.

**Sugestão:** Crie uma Pasta na Área de Trabalho (Desktop).

Aparecerão duas mensagens, clique “OK” na primeira (para salvar a planilha ativa) e “SIM” na segunda (para manter o formato do arquivo); Feche o Arquivo no *Excel*. O programa perguntará se quer salvar as alterações no arquivo e se deseja substituir o arquivo já existente, clique em “SIM” para ambas;

Depois na Pasta ou Local de Destino, abra o arquivo (que estará no formato .txt) com o **Bloco de Notas**, clique em **Editar** → **Substituir** (ou o atalho  $CTRL+H$ ), clique em **Localizar:** coloque a VÍRGULA (,) e em **Substituir por:** coloque o ponto (.) → **Substituir Tudo**. Salve o arquivo e feche o Bloco de Notas.

Seu conjunto de dados estará pronto para ser “lido” pelo *R*, porém é necessário que o programa encontre este conjunto de dados. O software *R* possui um diretório onde são gravados os arquivos utilizados, portanto se o conjunto de dados não estiver nesse diretório, ele não será lido. É necessário, incluir o conjunto de dados no diretório do *R* ou mudar o diretório do programa para um local que contenha o banco de dados.



Veremos como resolver o problema inserindo o bando de dados em um local e depois mudar o diretório do programa *R*.

- Selecione e copie (CTRL+C)(ou recorte (CTRL+X)) o arquivo (.txt) e faça o seguinte:
- **Se o arquivo estiver em uma Pasta:** Clique em: *Computador* ► *(:C)* ► *Program Files (Arquivos de Programa)* ► *R* ► *Colar (CTRL+V)*.
- **Se o arquivo estiver na Área de Trabalho, fora de uma Pasta:** Clique em: *Iniciar* ► *Computador* ► *(:C)* ► *Program Files (Arquivos de Programa)* ► *R* ► *Colar (CTRL+V)*.

Seu conjunto de dados estará em um diretório do *R*, porém esse diretório não é o padrão, então é preciso alterar o diretório. Com o programa *R* aberto, é necessário os seguintes passos:

- *Arquivo* ► *Mudar Dir ...* ► *Computador* ► *(:C)* ► *Program Files (Arquivos de Programa)* ► *R* ► *OK*.

Agora o *R* usará como diretório o local onde está seu arquivo. Para finalizar é preciso fazer com que o programa “leia” seu conjunto de dados, usando o seguinte comando:

```
read.table("NOME DO ARQUIVO.txt", head=T)
```

A função *head=T* fará com que o programa leia e comece pelo cabeçalho do seu conjunto de dados, caso o mesmo tenha e conte os dados a partir deste. Se não colocar *head=T* (ou colocar *head=F*), o programa “criará” colunas  $V_1, V_2, \dots, V_N$  e contará a partir da primeira linha do banco de dados. Geralmente, cria-se um objeto para inserir o conjunto de dados, como por exemplo:

```
DADOS <- read.table("NOME DO ARQUIVO.txt", head=T)
```

Para este capítulo foram criados seis conjuntos de dados relacionando sempre a altura com: circunferências do quadril e pescoço; comprimentos do braço, perna e dedo indicador; e número do calçado. Foram nomeados, respectivamente, como *acos*, *ap*, *ab*, *aper*, *ad*, *acalc*. Portanto, os comandos para construir os gráficos foram:

**Figura 4.3** *Altura x Circunferência do Quadril*

```
acos <- read.table("acos.txt", head=T)  
plot(acos, main="Altura x Circunferência do Quadril")
```

**Figura 4.4** *Altura x Circunferência do Pescoço*

```
ap <- read.table("ap.txt", head=T)
plot(ap, main="Altura x Circunferência do Pescoço")
```

**Figura 4.5** *Altura x Comprimento do braço*

```
ab <- read.table("ab.txt", head=T)
plot(ab, main="Altura x Comprimento do braço")
```

**Figura 4.6** *Altura x Comprimento da perna*

```
aper <- read.table("aper.txt", head=T)
plot(aper, main="Altura x Comprimento do perna")
```

**Figura 4.7** *Altura x Comprimento do dedo indicador*

```
ad <- read.table("ad.txt", head=T)
plot(ad, main="Altura x Comprimento do dedo indicador")
```

**Gráfico** *Altura x Número do Calçado*

```
acalc <- read.table("acalc.txt", head=T)
plot(acalc, main="Altura x Número do Calçado")
```

Para construir uma tabela que contenha duas (ou mais) colunas de uma tabela ou banco de dados usamos o comando **data.frame**

```
alturaxquadril <- data.frame(DADOS$Altura,DADOS$Quadril)
```

O programa *R* criará uma tabela com as colunas **Altura** e **Quadril** do banco de dados nomeado *DADOS*.

Para obter estatísticas descritivas usamos o comando *summary(DADOS)*. Por exemplo, *summary(ap)* gera as estatísticas de  $x_{(1)}, q_1, \tilde{x}, \bar{x}, q_3, x_{(n)}$  para as variáveis altura e circunferência do pescoço. O comando *sd(DADOS\$VARIÁVEL)* calcula o desvio padrão amostral da variável. Por exemplo, *sd(ap\$Altura)* calcula o desvio padrão amostral da variável *Altura* no banco de dados *ap*.

## 4.6 Elementos de regressão linear

Modelos de regressão são de grande utilidade para estatísticos e economistas, dentre outros profissionais, pois permitem fazer previsões e compreender de maneira mais confiável possível o comportamento das variáveis em questão. O coeficiente de correlação tem limitação linear justamente por estar ligado aos modelos de regressão linear. Veremos com breve abordagem os elementos de um modelo linear e qual representatividade do coeficiente de correlação linear nestes modelos.

A equação geral de um modelo linear simples é dada por [38]:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

sendo  $Y_i$  a variável dependente (explicada por  $X$ ), e pelo modelo, escrita como função linear de  $X$ ;

$\alpha$  o coeficiente linear da reta, valor que intercepta o eixo  $Y$ ;

$\beta$  o coeficiente angular da reta, inclinação em relação ao eixo  $X$ ;

$X_i$  a variável independente (explicativa), fator explicativo da reta de regressão;

$\varepsilon_i$  é uma variável aleatória com variância desconhecida e independente de  $X_i$  que contém os resíduos e os erros do modelo de regressão.

$i \in \{1, 2, \dots, n\}$

Para encontrar a reta de regressão estimada, usa-se o Método dos Mínimos Quadrados <sup>2</sup>, fazendo com que a soma do quadrado dos erros seja a menor possível. Por fim, encontramos as estimativas para  $\alpha$  e  $\beta$  são dadas por:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = \frac{\sum_{i=1}^n y_i}{n} - \hat{\beta} \frac{\sum_{i=1}^n x_i}{n}$$

Calculados os valores de  $\hat{\alpha}$  e  $\hat{\beta}$  a reta estimada de regressão é representada por

$$Y_i = \hat{\alpha} + \hat{\beta} X_i$$

sendo que  $\sum_{i=1}^n \hat{\varepsilon}_i = 0$ .

O valor que determina se o modelo de regressão é adequado para descrever a relação dos dados é chamado de coeficiente de determinação  $R^2$  que indica qual a proporção (ou porcentagem) da variabilidade de  $Y$  o modelo consegue “explicar”. Para calcular o valor de  $R^2$ , precisamos definir as três somas de quadrados.

**Soma total dos quadrados** é a soma das diferenças entre cada valor observado e a média amostral.

$$SQ_T = \sum_{i=1}^n (y_i - \bar{y})^2$$

---

<sup>2</sup>Veja mais em [13].

**Soma dos quadrados dos preditos** é a soma das diferenças entre cada valor predito pelo modelo e a média amostral.

$$SQ_P = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

**Soma dos quadrados dos resíduos** é a soma das diferenças entre cada valor predito e observado.

$$SQ_R = \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Temos a seguinte igualdade

$$SQ_T = SQ_P + SQ_R$$

O valor de  $R^2$  é dado por

$$R^2 = \frac{SQ_P}{SQ_T} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

equivalente a

$$R^2 = 1 - \frac{SQ_R}{SQ_T} = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

No caso de regressão linear simples,  $R^2 = r^2$ , ou seja, o coeficiente de determinação do ajuste do modelo é o quadrado do coeficiente de correlação linear. O coeficiente  $0 \leq R^2 \leq 1$ , diferentemente de  $r$  que assume valores em  $-1 \leq r \leq 1$ , além disso  $R^2 < |r|$ .

Para modelos de regressão linear múltipla  $Y$  é descrito pela combinação linear de duas ou mais variáveis aleatórias independentes, também chamada de regressoras.

O modelo é dado por

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_k X_k + \varepsilon_i$$

Neste caso, teremos  $k$  variáveis regressoras  $X_i$  e precisaremos estimar  $(k + 1)$  coeficientes  $\alpha_i$ ,  $i \in \{1, 2, \dots, k\}$ . O método para estimação dos coeficientes é similar ao de regressão linear simples. Primeiro estimamos  $\alpha_k$ , depois  $\alpha_{k-1}$ , e assim sucessivamente até  $\alpha_0$ . Consulte [38].

Nos modelos de regressão não linear, procuramos um modelo preditor para  $Y$  através do teste de funções não lineares como polinômias e exponenciais, os métodos de estimação dos coeficientes, geralmente é o método dos quadrados mínimos. Veja mais em [39].



## Capítulo 5

# Vida com Qualidade: Alimentação saudável e atividades físicas

### Sinopse

Esta pesquisa foi baseada na aplicação de um questionário para alunos do ensino fundamental (especificamente nas classes em que lecionei no ano de 2013) da Escola Estadual Elvira de Pardo Mêo Muraro, no período da tarde. Participaram da pesquisa alunos que tiveram o Termo de Consentimento Livre e Esclarecido (TCLE) assinado pelos responsáveis legais e quiseram responder ao questionário de 13 questões (anexo no fim deste trabalho). O TCLE foi encaminhado e autorizado pela escola e o projeto aprovado pelo Comitê de Ética em Pesquisas (CEP) da Unicamp (vide anexos). O questionário aplicado aos alunos visa analisar os aspectos de uma vida saudável. Dessa forma, foi respondido diretamente pelos alunos e analisado respeitando o anonimato dos respondentes. Não houve riscos, custos nem remuneração que pudessem decorrer do presente estudo para o participante. Os alunos foram informados da pesquisa e dos seus objetivos a fim de responder o questionário de forma precisa e correta. Para evitar que as respostas fossem lidas por colegas, os alunos responderam o questionário em casa com supervisão e orientação dos responsáveis.

O intuito foi que os resultados desta pesquisa contribuíssem para uma conscientização dos hábitos alimentares dos alunos e estimular a prática de atividades físicas.

Um segundo objetivo do estudo foi ensinar aos alunos como aplicar ferramentas estatísticas em dados reais, nesse caso, concernentes a eles mesmos, propiciando maior envolvimento e motivação do aluno com o conteúdo visto em sala de aula.

Acreditamos que o tema do estudo e a abordagem escolhida permitiram criar uma maior proximidade do aluno com a teoria de estudo, ampliar a competência de extrair informações de dados, desenvolvendo um espírito crítico e despertar o interesse pela pesquisa científica com coleta, organização e análise de informações.

## Justificativa

Pretendemos contribuir para a sensibilização dos alunos sobre a importância de uma alimentação saudável, prática de atividades físicas e práticas conscientes de consumo, contribuindo para a formação de um cidadão participativo na sociedade. Também pretendemos oferecer condições para levar o aluno a ser capaz de compreender conceitos de estatística e aplicá-los no seu dia-a-dia para assimilar e divulgar informações de diferentes naturezas e abrangências, emitir juízos com informações de natureza social, econômica e política, apresentadas em textos, notícias, propagandas, censos, pesquisas e outros meios.

A necessidade de um direcionamento nesse tópico surgiu após a aplicação do experimento “variáveis antropométricas”. Percebemos, em uma breve análise, que diversos alunos de baixa estatura possuem uma cintura consideravelmente ampla e alunos com grande estatura possuem valores relativamente baixos para a circunferência do quadril.

## 5.1 Sujeitos e métodos

Este estudo foi baseado na aplicação de questionário para coleta, representação tabular, gráfica e análise de dados referentes a hábitos alimentares e desportivos de alunos de uma escola da rede de ensino público da cidade de Campinas, SP. Participaram deste estudo sessenta e oito alunos matriculados no ensino fundamental do período da tarde. O questionário foi aplicado pelo próprio pesquisador com ciência e concordância da coordenadora pedagógica do ensino fundamental e da diretora da referida escola (Anexo 1), ciência do diretor do Instituto de Matemática, Estatística e Computação Científica (IMECC) da Universidade estadual de Campinas (Unicamp) (Anexo 2). Também foi submetido e aprovado na Plataforma Brasil, base nacional e unificada de registros de pesquisas envolvendo seres humanos para todo o sistema Comitê de Ética em Pesquisa/ Conselho Nacional de Ética em Pesquisa (Anexo 3).

Para minimizar a chance de respostas viesadas ou pouco precisas dos estudantes, conversamos inicialmente sobre os objetivos do estudo e o compromisso de manter o sigilo das respostas, reforçando a não identificação no mesmo.

A teoria matemática necessária foi apresentada ao longo do terceiro bimestre e reforçada no quarto bimestre, através da realização dos experimentos dos capítulos anteriores, sendo ela composta por coleta, tabulação e análise de dados.

A autorização dos responsáveis foi fundamental para a realização da pesquisa, pois a mesma envolveu adolescentes. Entregamos o Termo de Consentimento Livre e Esclarecido (TCLE) para ser preenchido dentro de um prazo de quinze dias.



Participaram do estudo somente os alunos com o TCLE assinado e o questionário respondido.

## 5.2 Informações prévias

Paralelamente ao trabalho de coleta e tratamento de dados, os professores de Educação Física e Ciências ministraram aulas expositivas com abordagens sobre os benefícios de uma alimentação regrada e balanceada, incluindo exemplos de dietas de atletas de alta performance. Foi salientada a importância de uma alimentação nutritiva e a prática de atividades físicas para melhor aproveitamento dos nutrientes, tratando o assunto de forma a determinar os benefícios científicos destas medidas, exemplificando os porquês de não ser suficientemente ótimo, apenas se alimentar bem ou somente praticar esportes [42].

Nesse âmbito de esclarecimentos, explicamos aos alunos que se alimentar bem é mais do que ingerir uma grande quantidade de alimentos, é dar ao nosso organismo uma quantidade ideal de nutrientes essenciais ao seu bom funcionamento. Nenhum alimento possui todos os nutrientes essenciais ao corpo humano e alimentos do mesmo grupo podem oferecer nutrientes em excesso e até causar doenças. Dessa forma, o ideal é ter uma alimentação balanceada e atenta às características a seguir:

1. **Adequação:** A alimentação deve ser apropriada às diferentes fases e condições de vida, às atividades, às circunstâncias fisiológicas e patológicas.
2. **Qualidade:** Deve conter variedade de alimentos que satisfaça todas as necessidades do corpo. Os alimentos devem ser nutritivos e não apenas calóricos, devem contribuir para o bom funcionamento do organismo humano.
3. **Quantidade:** Deve ser suficiente para atender o organismo em todas as suas necessidades.
4. **Harmonia:** É o equilíbrio entre os nutrientes, em relação à quantidade e qualidade.
5. **Variedade:** Ingerir diariamente alimentos de uma ampla seleção para conseguir diferentes nutrientes.

Essas características precisam ser levadas em conta durante a alimentação e uma forma gráfica de orientar as pessoas é a Pirâmide Alimentar, que mostra como deve ser a alimentação diária para manter uma pessoa, acima de 2 anos de idade, saudável.

Cada parte da pirâmide representa um grupo de alimentos considerando funções, nutrientes e o número de porções que é recomendado ingerir diariamente. Devemos incluir todos os grupos recomendados para garantir que o nosso organismo tenha os nutrientes dos quais necessita, porém a quantidade que devemos ingerir varia conforme o grupo ou nível.

A pirâmide possui na base, alimentos essenciais e que devem ser consumidos em maior quantidade; à medida que subimos em direção ao topo da pirâmide, diminui a necessidade do consumo dos determinados alimentos. No topo, temos alimentos que devem ser consumidos em quantidades moderadas. De acordo com o Departamento de Nutrição da Universidade de Brasília, ([fs.unb.br/nutricao/](http://fs.unb.br/nutricao/)), a pirâmide alimentar da Figura 5.1 tem oito grupos divididos em quatro níveis. Nenhum grupo ou nível pode ser utilizado como única fonte de nutrientes.



Figura 5.1: Pirâmide alimentar.

A princípio, água e líquidos não fazem parte das tradicionais pirâmides alimentares, porém alguns nutricionistas acreditam que estes devem compor a base por serem essenciais à vida já que representam até 70% do corpo humano. A água e os líquidos hidratam o organismo, ajudam a dissolver alimentos e transportar nutrientes. É indicada a ingestão de *pelo menos* 2 litros de água por dia.

No grupo 1, localizado no Nível 1, estão alimentos ricos em carboidratos, chamados de *energéticos* são responsáveis pelo fornecimento de grande parte da energia que o organismo humano necessita. Composto por alimentos como cereais e derivados como pães, massas em geral, raízes e tubérculos, é indicado a ingestão de 8 porções de alimentos desse grupo.

No Nível 2, estão alimentos ricos em vitaminas, sais minerais e fibras denominados *reguladores*. Os alimentos do grupo 2 são hortaliças e verduras e é indicado o consumo de 3 porções diárias. No grupo 3, estão frutas e sucos naturais e é indicado o consumo de 3 porções.

No Nível 3, estão os alimentos *construtores*, com proteínas, cálcio, ferro e zinco. Também possui fibras, vitaminas, açúcar e gorduras. No grupo 4, estão alimentos como leite e seus derivados, com orientação para o consumo de 3 porções diárias. No grupo 5, estão carnes e ovos, é indicado 2 porções. O grupo 6 é composto por leguminosas como feijão, soja, ervilha, entre outros, tais alimentos oferecem boas quantidades de calorias (sem prejudicar a saúde, possuem HDL, o “colesterol bom”) e proteínas específicas que ajudam a combater várias doenças. É indicada a ingestão diária de 1 porção de alimentos deste grupo.

No Nível 4, temos alimentos *energéticos extras*, ricos em calorias e colesterol. No grupo 7, estão óleos e gorduras, que realizam o transporte de vitaminas. Devem ser consumidos em quantidades moderadas, sendo 2 porções ou 120 *kcal*. No grupo 8 temos alimentos com grande quantidade calórica e pouca qualidade nutricional, cujo consumo indicado é de 2 porções ou 80 *kcal*. Esses alimentos são açúcares, balas, chocolates e salgadinhos. Os valores acima relacionados são condizentes à uma dieta de 2500 calorias por dia e contêm no total 24 porções.

Alimentos como açúcar, gorduras e sal estão inseridos em diversos grupos, por estarem naturalmente nos alimentos. A ingestão destes três alimentos deve ser cautelosa porque o excesso pode acarretar vários problemas para a saúde. Excesso de açúcar pode provocar doenças cardiovasculares, hipertensão arterial (a longo prazo) e obesidade que facilita o surgimento de diabetes tipo 2, patologia crônica [43]. O excesso de sal (sódio) pode causar hipertensão arterial, pedra nos rins, insuficiência renal, também agrava a osteoporose, afeta o paladar e acelera o envelhecimento [45]. O excesso de gordura, segundo Halpern [46] (endocrinologista), pode causar problemas nas artérias, no cérebro e provocar doenças como diabetes e obesidade.

Para sabermos o número exato que devemos ingerir de cada grupo de alimento é preciso o auxílio do nutricionista, pois existe uma série de variáveis, tais como sexo, peso, idade, altura e necessidades individuais que causam variações. Grande parte das pessoas, precisam, em média, da quantidade mínima de porções de cada grupo de alimentos dentro das variações recomendadas.

Diversos estudos já mostraram que uma alimentação equilibrada e saudável juntamente com práticas regulares de atividades físicas é determinante para uma boa qualidade de vida. A última Pesquisa de Orçamento Familiar (POF) do IBGE (Instituto Brasileiro de Geografia e Estatística), disponível em <http://goo.gl/RAIJDh>, de 2008-9, aponta que o excesso de peso em adultos do gênero masculino (em 2006) era menos de 20%, e em 2010 já atingia 50,1%. Em relação às mulheres, no mesmo período, a proporção passou de 28,7% para 48%. Preocupada com o quadro e visando disseminar para a população a necessidade e relevância de uma alimentação balanceada e saudável para uma boa qualidade de vida, a pesquisadora Sônia Tucunduva Philippi, do departamento de nutrição da faculdade de saúde pública da USP (Universidade de São Paulo), colaborou para o desenvolvimento do "Guia Alimentar Brasileiro" (junto ao Ministério da Saúde), calculando o número de porções e o valor energético destas para os grupos alimentares reduzindo à dieta de 2000 calorias.

A recomendação atual é uma alimentação variando de quatro a seis refeições diárias, onde cada uma das três principais refeições devem conter de 15 a 35% da energia sugerida e lanches intermediários, cada um contendo de 5 a 15% da energia.

A Figura 5.2 mostra a pirâmide alimentar adaptada às necessidades atuais da população brasileira [47].



Figura 5.2: Pirâmide alimentar adaptada às necessidades da população brasileira.

### 5.2.1 Obtenção dos dados

O registro dos dados foi feito no *Excel* pelo responsável da pesquisa, garantindo o anonimato. A representação gráfica, tabular e a análise dos dados foram realizadas conjuntamente com os alunos (usando a sala de informática) para que fizessem análises, inferências e apresentassem propostas para a abordagem do tema no espaço escolar.

Das treze questões, quatro foram sobre variáveis quantitativas, sendo altura, peso, circunferência do quadril e tempo diário destinado ao uso de eletrônicos. Nove ques-

tões foram qualitativas, sendo gênero do aluno, o que geralmente come no café da manhã, almoço e jantar, se possuem o hábito de se alimentar entre as refeições, se participa das atividades da disciplina de educação física, se pratica atividades físicas fora da escola, se ingeri alimentos usando eletrônicos e se dormem em horários regulares. As respostas das perguntas sobre refeições foram categorizadas, com o auxílio de professores de biologia, em 5 classes de nutrientes (carboidratos, lipídeos, proteínas, sais minerais e vitaminas) essenciais aos seres humanos. Foi considerado “*sim*” para um nutrientes, se o alimento possuir uma quantidade significativa deste nutriente. Por exemplo, se a quantidade de sais minerais em um alimento for insignificante comparada à quantidade necessária diariamente, não será considerado que o alimento possui sais minerais.

As variáveis quantitativas medidas foram, exclusivamente, para determinar o Índice de Massa Corporal (*IMC*), que relaciona o peso e a altura de uma pessoa, e o Índice de Adiposidade Corporal (*IAC*) que, segundo pesquisa da revista *Obesity* [41], mede de forma fidedigna a porcentagem de gordura no corpo humano relacionando a circunferência do quadril com a altura. Foi sugerido pelo professor de Ciências o método das dobras cutâneas, também chamado de método das “9 dobras” (ou indireto) [40], que divide o corpo humano em massa corporal magra (*MGM*) e massa gorda (*MG*). Este método leva em conta mensurações em dobras cutâneas, feitas com um aparelho denominado adipômetro, no subescapular (região do ombro), tríceps, bíceps, peitoral, axilar média, supra-iliaca (região do quadril), abdome, coxa e panturrilha medial (parte interna). O método é cientificamente muito confiável, seus procedimentos devem ser seguidos a risca. A unidade escolar não possuía um adipômetro (ou plicômetro), e seria necessária a permissão prévia dos pais/responsáveis pelos menores para a aplicação do método dado que se trata de uma avaliação de contato com o menor.

Com as mensurações foram calculados o *IMC* e o *IAC*, dados pelas expressões:

$$IMC = \frac{P}{h^2}; \quad IAC(\%) = \frac{c}{h \cdot \sqrt{h}} - 18.$$

onde  $P$ = Peso ( $Kg$ ),  $h$ =altura ( $m$ ) e  $c$ =circunferência do quadril ( $cm$ ).

O *IMC*, apesar de ser adotado como referencial pela Organização Mundial de Saúde (*OMS*), foi calculado, porém não foi utilizado como fator de decisão, pois há empecilhos em usá-lo para determinar se uma pessoa está acima do peso. Por exemplo, pessoas musculosas podem ter um índice de massa corporal alto e não serem gordas, porque este mede a massa corpórea. Além disso, o *IMC* não tem aplicações confiáveis em menores de 16 anos (nesse caso é melhor a análise gráfica e específica) e para idosos acima de 60 anos (é utilizado uma classificação diferenciada). Existe

ainda uma vinculação sobre o impacto para diferenças raciais e étnicas. Por exemplo, um grupo de assessoramento à Organização Mundial de Saúde concluiu que pessoas de origem asiática poderiam ser consideradas acima do peso com um *IMC* maior que 23 (saudável na tabela “padrão”). Os valores padronizados para *IMC* e *IAC* e suas respectivas classificações encontram-se na Tabela 5.1 [42].

Tabela 5.1: Tabela de Classificação de *IMC* e *IAC*.

IMC	Classificação	Homens	Gênero	Mulheres
< 16	Magreza grave	IAC (%)	Classificação	IAC (%)
16 até 17	Magreza moderada	< 10	Baixíssima	< 15
17 até 18,5	Magreza leve	10 até 14	Baixa	15 até 19
18,6 até 24,9	Saudável	14 até 18	Ideal	19 até 25
25 até 29,9	Sobrepeso	18 até 24	Moderada	25 até 29
30 até 34,9	Obesidade I	≥ 24	Excesso	≥ 29
35 até 39,9	Obesidade II (severa)			
≥ 40	Obesidade III (mórbida)			

Usaremos como referência a tabela à direita, por ser mais confiável conforme a breve abordagem acima.

### 5.3 Análise das informações

Após o preenchimento de uma tabela contendo todas as informações, foi feita a análise dos resultados, questão por questão, visando uma percepção geral dos dados. Depois, foram analisados *IMC* e *IAC* e detectados os possíveis casos de magreza e obesidade. Para estes casos fizemos uma análise específica, buscando correlações que não necessariamente serão descritas como causalidades destas condições. O questionário foi respondido por 36 alunas e 32 alunos.

Para análise, a variável altura foi dividida em 5 classes conforme a Figura 5.3.

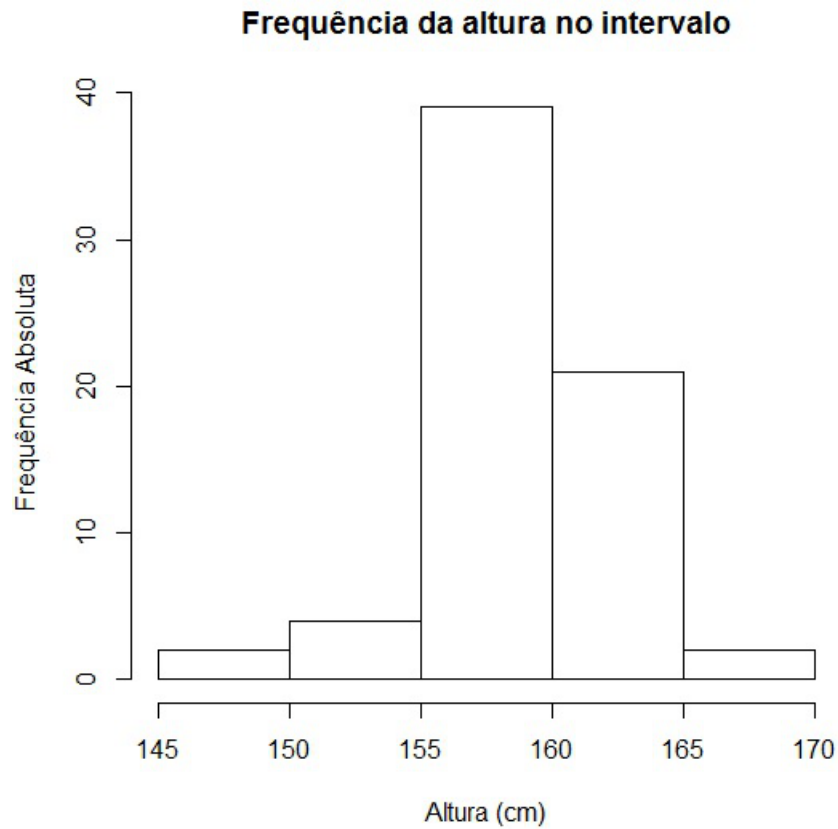


Figura 5.3: Frequência das alturas no intervalo,  $n = 68$ .

Tabela 5.2: Estatísticas descritivas da variável altura,  $n = 68$ .

$x_{(1)}$	148,7	$q_3 - q_1$	3,975
$q_1$	156,75	$x_{(68)} - x_{(1)}$	18,4
$\tilde{x}$	158,6	$\bar{x}$	158,67
$q_3$	160,725	$s$	3,432
$x_{(68)}$	167,1	$C_V$	0,0216

O olhar conjunto para a Figura 5.3 e a Tabela 5.2, mostra que a altura é bem concentrada em torno de  $\bar{x} \approx \tilde{x}$ , mais especificamente entre  $q_1$  e  $q_3$  que concentram 34 valores de altura em menos de 4cm. O valor de  $C_V$  corrobora, pois a variabilidade é de apenas 0,0216, pouco mais de 2% de variação de torno de  $\bar{x}$ .



A variável peso foi dividida em 6 classes e gerado o histograma da Figura 5.4.

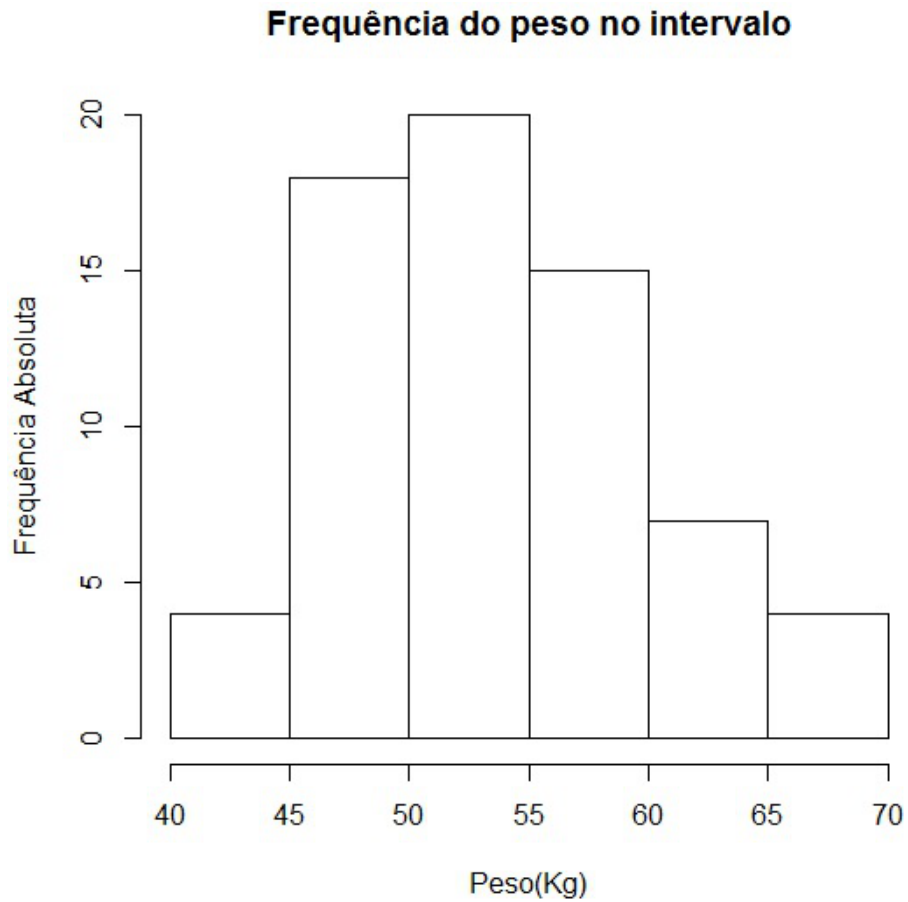


Figura 5.4: Frequência do peso no intervalo,  $n = 68$ .

Tabela 5.3: Estatísticas descritivas da variável peso,  $n = 68$ .

$x_{(1)}$	42,1	$q_3 - q_1$	7,25
$q_1$	49,2	$x_{(68)} - x_{(1)}$	27,6
$\tilde{x}$	52,8	$\bar{x}$	53,59
$q_3$	56,45	$s$	6,067
$x_{(68)}$	69,7	$C_V$	0,1132

Pela Figura 5.4 e pela Tabela 5.3, temos  $\bar{x} - \tilde{x} = 0,79Kg$ . A diferença interquartil  $q_3 - q_1 > 7kg$  e o valor de  $C_V$  mostra que a variável peso tem maior variabilidade que a altura.

Para a variável Circunferência do quadril foi construído o gráfico da Figura 5.5.

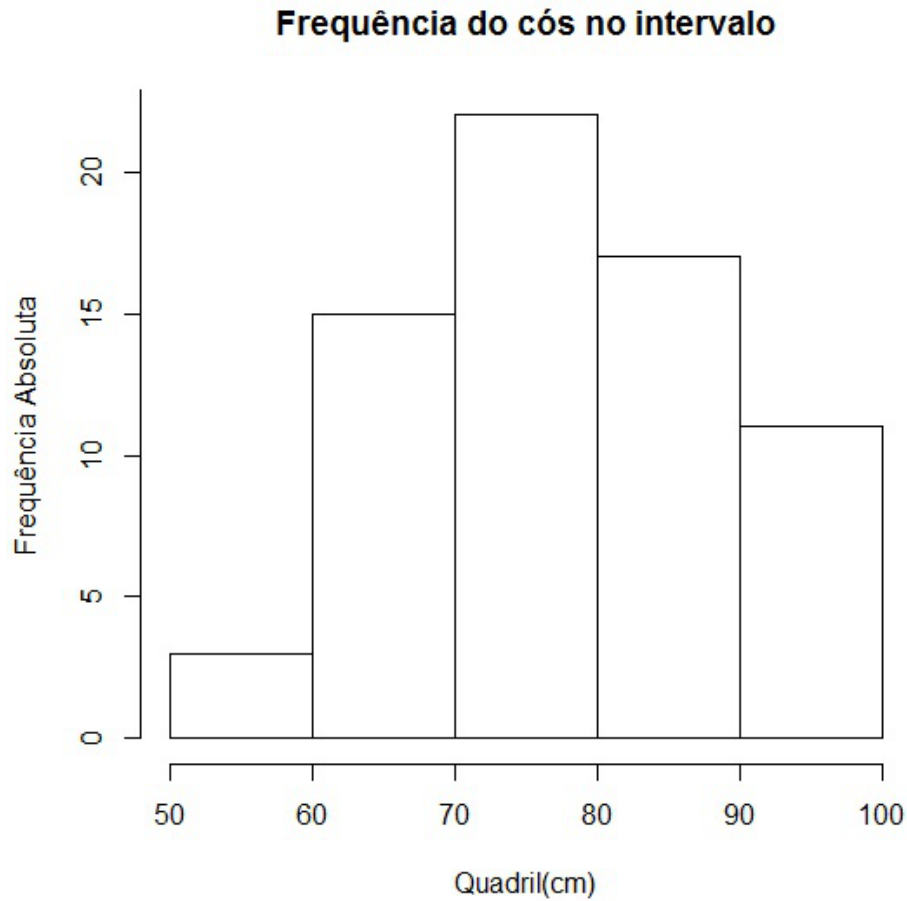


Figura 5.5: Frequência da circunferência do quadril no intervalo,  $n = 68$ .

Tabela 5.4: Estatísticas descritivas da variável quadril,  $n = 68$ .

$x_{(1)}$	58,7	$q_3 - q_1$	16,975
$q_1$	69,68	$x_{(68)} - x_{(1)}$	41,1
$\tilde{x}$	74,2	$\bar{x}$	77,39
$q_3$	86,65	$s$	10,76
$x_{(68)}$	99,8	$C_V$	0,139

Observando a Figura 5.5 e a Tabela 5.4, notamos que a classe modal é de 70 a 80cm de circunferência do quadril,  $\bar{x}$  e  $\tilde{x}$  estão nesta classe, porém  $q_1 < 70$  e  $q_3 > 80$  estão fora da classe modal e diferem em quase 17cm. Pelo valor do coeficiente de

variação  $C_V = 0,139$  a variável circunferência do quadril apresentou maior dispersão em torno da média do que as variáveis antropométricas altura e peso.

Depois analisamos as perguntas do questionário.

Para a pergunta 5, *O que você geralmente come no café da manhã?*, temos os seguintes dados.

Tabela 5.5: O que geralmente os alunos comem no café da manhã,  $n = 68$ .

Alimento líquido	Alunos	Alimento sólido	Alunos
Café com leite	21	Pão com margarina	43
Leite com achocolatado	18	Frutas	7
Leite com açúcar	13	Pão com frios	5
Café	5	Pão com geléia	2
Suco/Refrigerante	3	Pão integral	1
Iogurte	1	Total	58
Água de berinjela	1		
Total	61		

Além dos alimentos sólidos apresentados na Tabela 5.5, os alunos possuem o hábito de ingerir bolo, biscoito recheado ou salgado, torradas, pães de queijo e bisnagas. Pela Tabela 5.5, temos que 3 alunos ingerem apenas líquido no café da manhã e 7 alunos não fazem essa refeição.

Pela análise da questão 6, *O que geralmente comem no almoço?*, temos a Tabela 5.6.

Tabela 5.6: O que geralmente os alunos comem no almoço,  $n = 68$ .

Alimento sólido	Alunos	Alimento líquido	Alunos
Arroz, feijão, carne e salada	35	Suco/Refrigerante	22
Arroz, feijão e carne	23	Suco/Refrigerante	13
Macarrão	4	Suco/Refrigerante	3
Arroz e feijão	2	Suco/Refrigerante	2
Arroz com ovo	1	Suco/Refrigerante	1
Soja	1	Suco/Refrigerante	1
Arroz, carne e salada	1	Suco/Refrigerante	1
Total	67	Total	43

Vale ressaltar que o aluno que não tem o hábito de almoçar é um dos sete que não tomam café da manhã. O que definimos como carne entende-se por bovina, suína ou galinácea. Apenas um aluno tem o hábito de comer sobremesa e outro come fruta como complemento da refeição.

Salientamos que a quantidade de suco/refrigerante nas Tabelas 5.6 e 5.7, refere-se ao número de alunos que fazem a respectiva refeição tomando estes líquidos.

Para o jantar como refeição, temos a Tabela 5.7.

Tabela 5.7: O que geralmente os alunos comem no jantar,  $n = 68$ .

Alimento sólido	Alunos	Alimento líquido	Alunos
Arroz, feijão, carne e salada	40	Suco/Refrigerante	31
Arroz, feijão e carne	19	Suco/Refrigerante	14
Macarrão	7	Suco/Refrigerante	4
Pizza ou lanches	2	Suco/Refrigerante	2
Total	68	Total	51

Além dos 2 alunos que jantam pizza, lanche ou similar, temos 4 alunos que, eventualmente, comem pizzas, lanches, biscoitos, sopa de legumes e cuscuz.

Em relação à pergunta 8, *Além das três refeições básicas, você tem o hábito de se alimentar entre elas?*, temos a Tabela 5.8.

Tabela 5.8: Alimentação entre as refeições principais,  $n = 68$ .

Alimento	Alunos
Guloseimas	21
Merenda escolar	45
Naturais	30
Outros	3
Total	67

Vale ressaltar que um aluno faz somente as 3 refeições diárias. Para estes alunos, as refeições intermediárias entre as refeições são, especificamente, entre o almoço e o jantar.

**Outros** é caracterizado por alimentos como leite com achocolatado, danone, vitamina de frutas, sorvete, pão, bolo, biscoito recheado e barra de cereais.

Para melhor entendimento e visualização do que os alunos ingerem nestas refeições entre as refeições principais foi feito o diagrama de Venn (Veja mais em [5])<sup>1</sup>(assunto não abordado, apenas colocado na análise deste trabalho) da Figura 5.6.

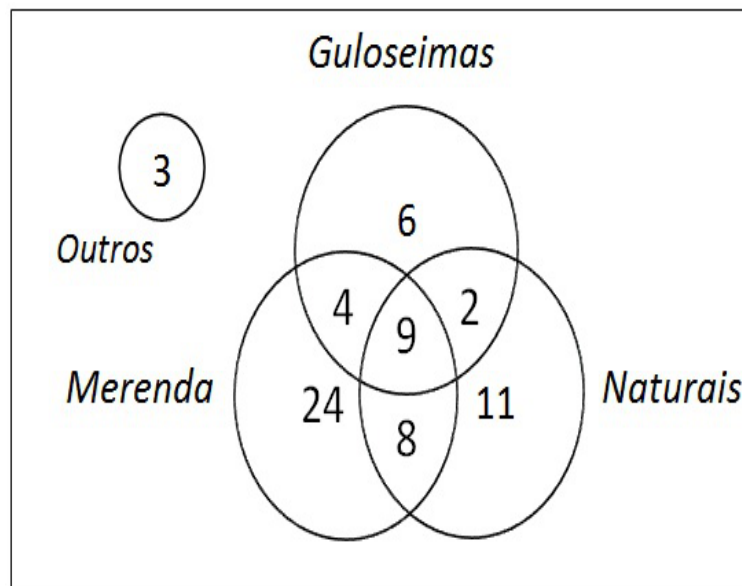


Figura 5.6: Diagrama de Venn para refeições intermediárias,  $n = 68$ .

Vemos no diagrama que, mais da metade (41 de 68) dos alunos observados comem apenas uma das três alternativas, sendo estas alimentos naturais como lanches, sucos e frutas; alimentos industrializados como salgadinhos e pães de queijo; e, guloseimas como bala, chiclete e chocolate, dentre outros. Dos alunos que ingerem apenas uma das alternativas, quase 60% se alimenta da merenda escolar. Vale ressaltar que dos 9 alunos ingerem as 3 opções, 3 tem o hábito de comer bolos, iogurtes e sorvetes.

---

<sup>1</sup>O diagrama de Venn é referente a teoria elementar de conjuntos, iniciada nos anos finais do ensino fundamental (6<sup>o</sup> ao 9<sup>o</sup> anos) e rerepresentada com mais rigor no ensino médio, geralmente, como pré requisito aos conceitos de introdução à probabilidade.

Para a pergunta 9, sobre a prática de atividades físicas na escola, temos a Tabela 5.9.

Tabela 5.9: Prática de atividades físicas no ambiente escolar,  $n = 68$ .

Realiza?	Alunos
Totalmente	50
Parcialmente	13
Não realiza	4
Dispensado*	1
Total	68

\* Motivo cirurgia.

A questão 10, referente à prática de atividades esportivas fora do ambiente escolar, apresenta como resultado a Tabela 5.10.

Tabela 5.10: Prática de atividades físicas fora do ambiente escolar,  $n = 68$ .

Realiza?	Total
Sim	60
Não	8

As atividades físicas praticadas fora da escola, são jogar futebol, vôlei, queimada, kinect, basquete, fazer caminhada, ginástica, karatê, boxe, academia (musculação e esteira), correr, pular corda, arrumar a casa, cuidar do irmão, brincar de rouba-bandeira, pega-pega, esconde-esconde, boneca, dançar, andar de bicicleta, patins, skate, soltar pipa, tocar instrumentos musicais. Fato interessante é que o aluno dispensado por cirurgia, joga futebol fora das dependências escolares. Destas atividades relacionadas as mais frequentes são apresentadas na Tabela 5.11.

Tabela 5.11: Atividades físicas mais praticadas fora da escola,  $n = 68$ .

Atividade	Total	Meninas	Meninos
Futebol	29	5	24
Arrumar a casa	14	11	3
Bicicleta	11	2	9

Para melhor compreensão da pergunta 11, referente ao tempo que os alunos gastam com a utilização de equipamentos eletrônicos como celulares, tablet's, computadores, iPad, iPod, mp3 players, videogames, kinect, câmera digital, televisores, entre

outros, foi feita a Tabela 5.12, página 124, com as frequências absolutas, estatísticas de tendência central e dispersão.

Tabela 5.12: Tempo de utilização de eletrônicos,  $n = 68$ .

Tempo (h)	F.A.
0	1
1	11
2	10
3	8
4	7
5	18
6	12
Total	67
$x_{(1)}$	0
$q_1$	2
$\bar{x}$	3,67
$\tilde{x}$	4
$q_3$	5
$x_{(67)}$	6
$s$	1,84
$C_V$	0,5014

**Legenda:** F.A. = Frequência absoluta; h = Horas.

Pela Tabela 5.12, podemos observar que o tempo médio  $\bar{x}$  de utilização é de 3 horas e 40 minutos, com um desvio padrão  $s$  de 1 hora e 50 minutos e um coeficiente de variação  $C_V$  de mais de 50%. Se usarmos a mediana, vemos que o tempo de uso aumenta para 4 horas diárias. A diferença interquartil  $q_3 - q_1$  é de 3 horas compreendida entre 2 e 5 horas, pelos valores apresentados temos uma concentração maior de uso destes aparelhos eletrônicos entre os valores de  $\bar{x}$  e  $q_3$ . Por fim, vale ressaltarmos que apenas um aluno não utiliza equipamentos eletrônicos.

Para a pergunta 12, referente a ingestão de alimentos utilizando eletrônicos, de forma mais específica, televisão, computador e videogame, temos a Tabela 5.13.

Tabela 5.13: Alimentação utilizando aparelhos eletrônicos,  $n = 68$ .

Alimenta?	F.A.
Sim	46
Não	22

**Legenda:** F.A. = Frequência absoluta.

Pela Tabela 5.13, temos que 46 alunos (mais de 67%) se alimentam na frente dos eletrônicos ingerindo leite com achocolatado, suco, refrigerante, danone, sorvete, pipoca, bolo, salgadinhos fritos ou industrializados, pão, chocolate, frutas, balas, chicletes e pirulitos. Destes 22 alunos tem hábito de realizar as refeições diárias na frente dos equipamentos, mesmo número de alunos que não se alimentam durante a utilização dos aparelhos eletrônicos.

Para a pergunta 13, referente aos horários que os aluno dormem e acordam, temos a Tabela 5.14.

Tabela 5.14: Possui horários regulares para dormir e acordar,  $n = 68$ .

Horários regulares?	Alunos	Todos os dias?	Alunos
Sim	43	Sim	19
Não	25	Não	24

Os 24 alunos com horários regulares para dormir e acordar, porém não todos os dias, nos fins de semana, dormem e acordam mais tarde também em horários regulares.



Para melhor entendimento desta questão fizemos a Tabela 5.15 com o tempo que esses alunos dormem.

Tabela 5.15: Estatísticas do tempo que dorme os alunos com horários regulares.

(a) $n = 19$		(b) $n = 24$		(c) $n = 24$	
Todos os dias		Dias da semana		Fim de semana	
Tempo (h)	F. A.	Tempo (h)	F. A.	Tempo (h)	F. A.
9	11	8	3	8	4
10	8	9	8	9	7
$q_1$	9	10	7	10	4
$\tilde{x}$	9	11	4	11	7
$\bar{x}$	9,42	12	1	12	2
$q_3$	10	13	1	$q_1$	9
$s$	0,507	$q_1$	9	$\bar{x}$	9,833
$C_V$	0,0538	$\bar{x}$	9,792	$\tilde{x}$	10
		$\tilde{x}$	10	$q_3$	11
		$q_3$	11	$s$	1,274
		$s$	1,25	$C_V$	0,1277
		$C_V$	0,1296		

**Legenda:** F.A. = Frequência absoluta; h = Horas.

Na Tabela 5.15 (a), vemos que os alunos com horários regulares dormem, em média  $\bar{x}$ , 9 horas e 25 minutos por dia com desvio padrão  $s$  de cerca de 30 minutos,  $C_V$  mostra uma variabilidade de 5%. Já na Tabela 5.15 (b), alunos com horários semanais regulares possuem  $\bar{x}$  igual a 9 horas e 48 minutos e  $s$  de 1 hora e 15 minutos e  $C_V = 0,1296$ . Na Tabela 5.15 (c),  $\bar{x}$  é 9 horas e 50 minutos com  $s$  de 1 hora e 16 minutos e  $C_V = 0,1277$ . Portanto, o grupo de alunos que dormem em horários regulares todos os dias possuem menor variabilidade no tempo que dormem e estatísticas de tendência central ( $\bar{x}$ ,  $\tilde{x}$ ) mais próximas.

### 5.3.1 Análise bivariada

Para continuar a análise de forma mais detalhada, fizemos gráficos e tabelas visando avaliar tendências (se existirem) e obter conclusões.

Fizemos a Figura 5.7, que relaciona as variáveis altura e peso considerando o gênero.

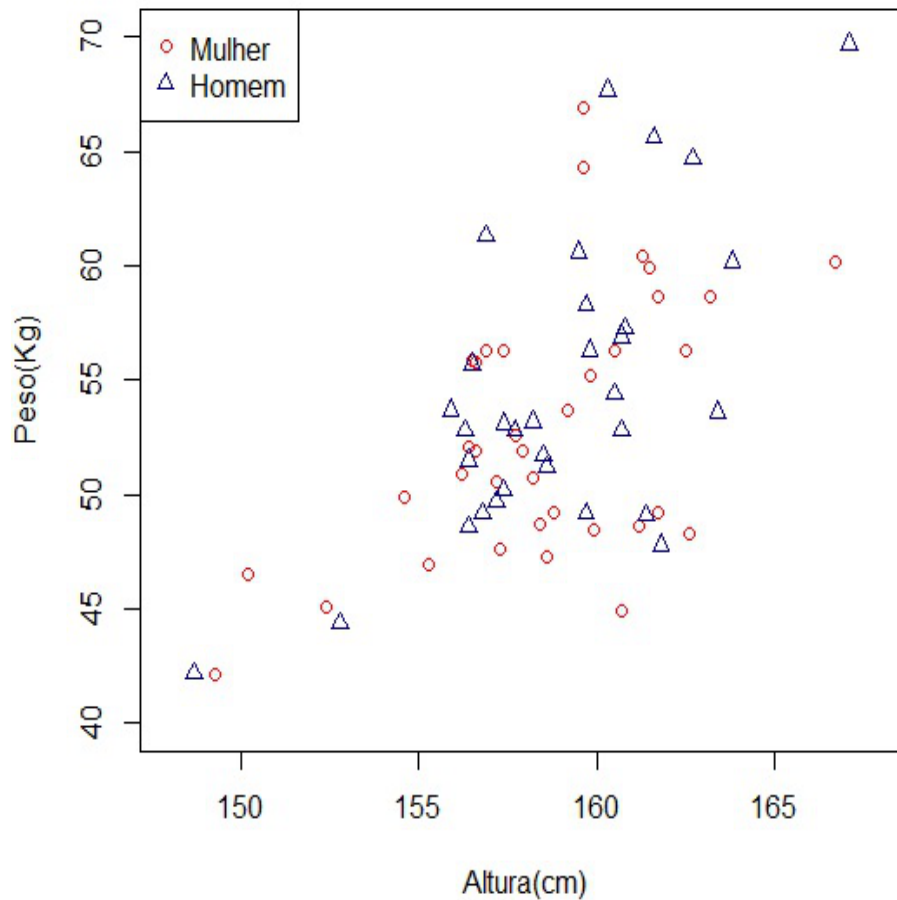


Figura 5.7: Dispersão da altura  $\times$  peso,  $n = 68$

Observando a Figura 5.7, temos evidências de que também existe correlação entre altura e peso. Calculamos o coeficiente de correlação linear  $r = 0,5887$ , o que corrobora com o visual e determina uma correlação moderada positiva. Ainda pela observação do gráfico, vemos que as nuvens de pontos para os gêneros possuem comportamento similar.

Para ajudar na compreensão das características altura e peso por gênero fizemos a Tabela 5.16.

Tabela 5.16: Estatísticas descritivas, por gênero,  $n = 68$ .

Altura (cm)			Peso (kg)		
Gênero/valor	Feminino	Masculino	Gênero/valor	Feminino	Masculino
$x_{(1)}$	149,3	148,7	$x_{(1)}$	42,1	42,2
$q_1$	156,6	156,9	$q_1$	48,58	50,08
$\bar{x}$	158,4	158,9	$\bar{x}$	52,73	54,55
$\tilde{x}$	158,5	159,1	$\tilde{x}$	51,9	53,15
$q_3$	160,8	160,7	$q_3$	56,3	57,55
$x_{(36)}$	166,7	167,1	$x_{(32)}$	66,9	69,7
$s$	3,5	3,4	$s$	5,68	6,43
$C_V$	0,0221	0,0214,	$C_V$	0,1077	0,1179

Pela observação da Tabela 5.16, notamos que para a variável altura as estatísticas são próximas, inclusive o valor de  $C_V$  para ambos os gêneros. Para a variável peso, notamos que os valores para o gênero feminino são inferiores aos do gênero masculino e o valor de  $C_V$  é inferior para o gênero feminino.

Na Tabela 5.17 estão as estatísticas descritivas para a circunferência do quadril, por gênero.

Tabela 5.17: Estatísticas descritivas, por gênero, da circunferência do quadril.

Quadril (cm), $n = 68$ .		
Gênero/valor	Feminino	Masculino
$x_{(1)}$	59,8	58,7
$q_1$	67,85	71,77
$\bar{x}$	76,54	78,34
$\tilde{x}$	72,95	74,95
$q_3$	86,65	86,68
$x_{(max)}$	99,2	99,8
$s$	11,2	10,33
$C_V$	0,1463	0,1319

Em relação a circunferência do quadril por gênero, pela Tabela 5.17, temos para o gênero feminino uma maior amplitude, maior diferença intequartil ( $q_3 - q_1$ ), maior diferença entre  $\bar{x}$  e  $\tilde{x}$  e maior variabilidade  $C_V$  do que para o gênero masculino.

A Figura 5.8 relaciona altura e circunferência do quadril por gênero.

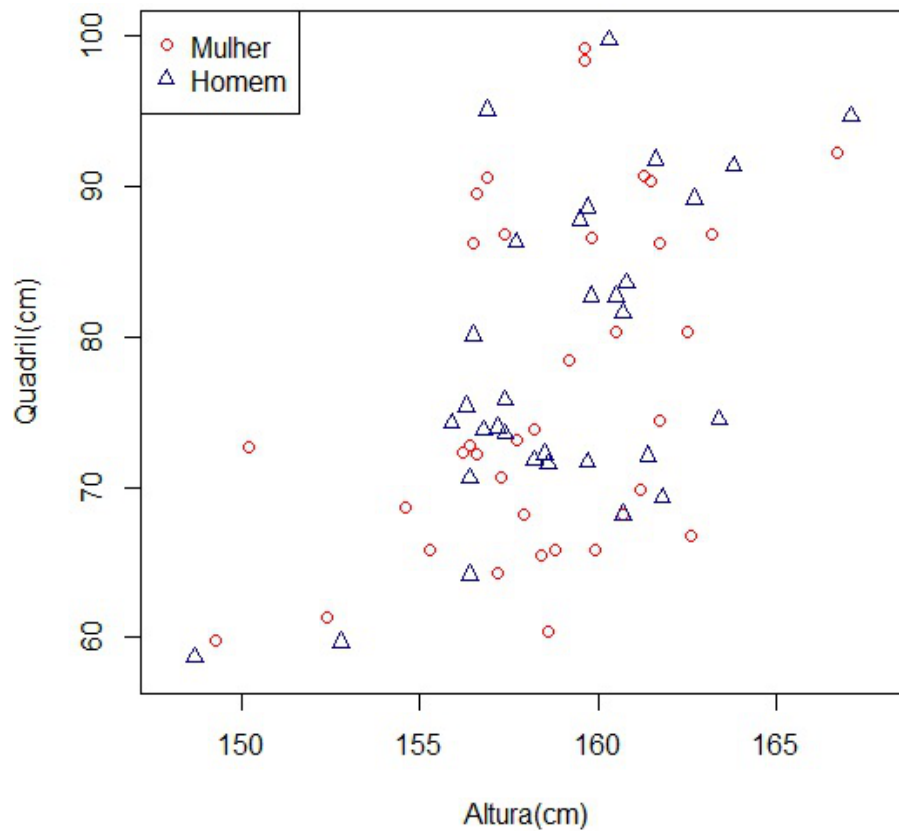


Figura 5.8: Dispersão da altura  $\times$  quadril,  $n = 68$ .

Pela observação da Figura 5.8, temos que as nuvens de pontos formada por gênero apresentam um padrão próximo de dispersão aparentemente. Notamos ainda, um ponto mais disperso da nuvem, localizado em ( $altura \approx 150$ ,  $quadril \approx 73$ ). Este ponto é um candidato à outlier, mas vale lembrar que deveríamos compará-lo ao padrão do grupo de garotas, já que remete ao gênero feminino, que possui 36 observações, número insuficiente para determinarmos um padrão.

Para análise das qualidades nutricionais da alimentação, consideramos os nutrientes carboidratos, proteínas, lipídios, sais minerais e vitaminas. A tabulação serviu para determinar qual (ou quais) refeição(ões) era(m) mais nutritiva(s). O critério de atribuição consistiu na criação de uma função indicadora sendo que: se a refeição tivesse uma quantidade significativa do nutriente, este receberia o valor 1, caso contrário, receberia o valor 0. Criamos uma contagem para indicar a quantidade dos nutrientes em cada refeição.

A Tabela 5.18, regista a quantidade de nutrientes por refeição <sup>2</sup>.

Tabela 5.18: Tabela de nutrientes por refeição.

Nutriente/Refeição	Carboidrato	Lipídio	Proteína	Sais minerais	Vitaminas
Café da manhã	61	60	56	30	11
Almoço	67	67	67	47	45
Jantar	68	68	67	54	53

Construímos a Figura 5.9 considerando a frequência relativa <sup>3</sup> de nutriente <sup>4</sup>.

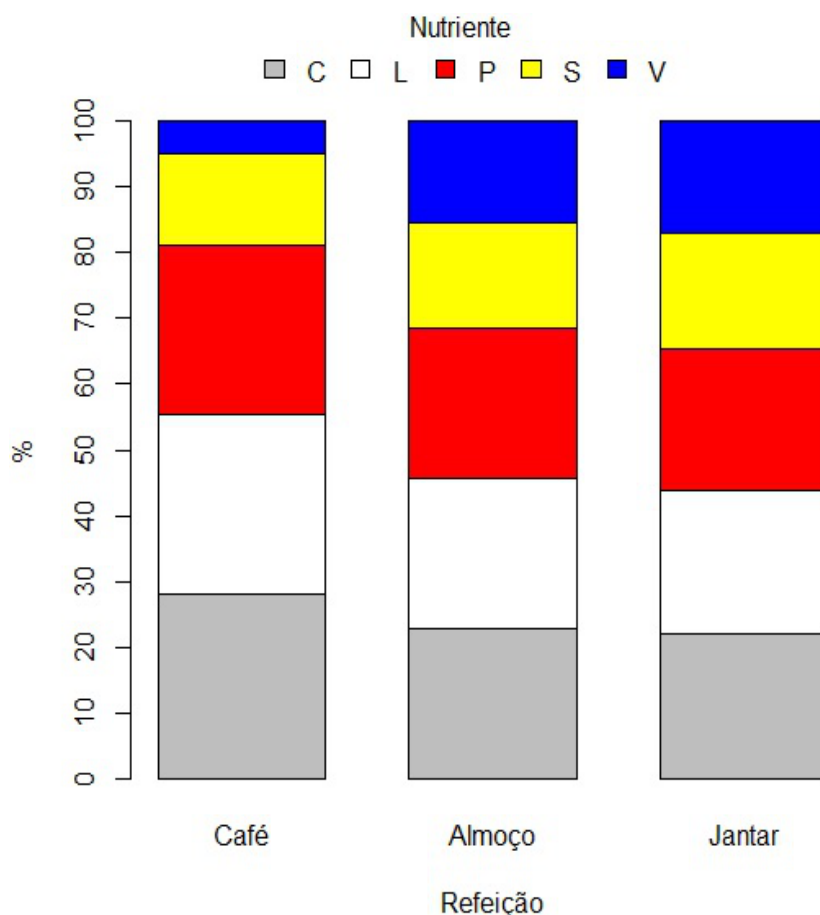


Figura 5.9: Frequência relativa de nutrientes por refeição.

**Legenda:** C = Carboidratos; L = Lipídios; P = Proteína; S = Sais minerais; V = Vitaminas.

<sup>2</sup>O valor máximo para um nutriente no café da manhã é 61; para o almoço, 67; e jantar, 68.

<sup>3</sup>Razão da quantidade do nutriente pela quantidade total na refeição.

<sup>4</sup>Note que uma refeição balanceada deve possuir 20% (ou 0,2) de cada um dos cinco nutrientes.

Pela análise da Figura 5.9, o café da manhã é a refeição menos provida de nutrientes, principalmente sais minerais e vitaminas, apenas o carboidrato é consumido em todas as refeições, além disto, apenas 8 alunos ingerem quantidades significantes dos cinco nutrientes diariamente no café da manhã. Para o almoço, vemos um aumento no consumo dos cinco nutrientes, sendo que carboidrato, lipídio e proteína atingem a totalidade dos casos para esta refeição. Temos ainda que 45 alunos ingerem quantidades consideráveis dos cinco nutrientes no almoço. No jantar, notamos que os nutrientes sais minerais e vitaminas são consumidos em cerca de 80% das refeições e 51 alunos ingerem quantidades significantes dos cinco nutrientes. A Figura 5.9 evidencia que o jantar é uma refeição mais balanceada, pois a barra tem uma mais equilibrada para cada nutriente.

Analizamos em qual(is) das refeições é ingerido mais nutrientes considerando-a(s) prioritária(s). Ou seja, se o aluno ingere quantidades significantes dos cinco nutrientes no café da manhã; de três deles no almoço e de quatro deles no jantar; o café da manhã será considerado a refeição mais “importante”. Estas informações são apresentadas na Tabela 5.19.

Tabela 5.19: Refeição mais provida de nutrientes.

Refeição mais nutritiva/balanceada	
Café	0
Café/Almoço	3
Café/Jantar	5
Almoço	14
Almoço/Jantar	28
Jantar	18
Total	68

Pela análise da Tabela 5.19, vemos que o café da manhã não é a mais nutritiva das refeições para nenhum dos alunos observados. Temos que café da manhã e almoço são prioritários em 4,41% do total; café da manhã e jantar são prioritários em 7,35%; o almoço em 20,59%; almoço e janta em 41,18% e janta em 26,47%.

Fica muito claro, pelos dados observados, que o café da manhã não é considerado uma refeição essencial pelos alunos, temos que mais de 10% (7 alunos) nem sequer se alimentam no café da manhã e apenas 11,76% (8 alunos) dos alunos se alimentam de maneira nutritiva nesta refeição.

A Figura 5.10 <sup>5</sup>, mostra peso  $\times$  tempo médio que os alunos dormem.

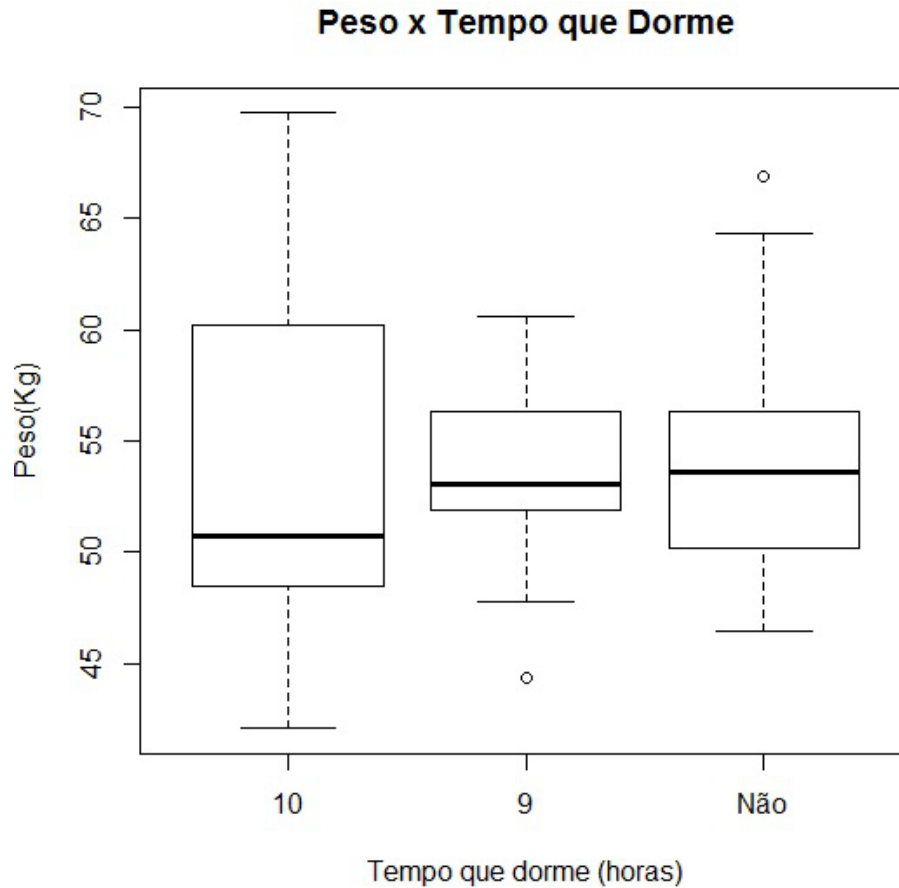


Figura 5.10: Tempo que dorme  $\times$  peso.

Vemos no boxplot à esquerda uma maior concentração de peso dos alunos entre  $q_1$  e  $\tilde{x}$  (cerca de  $2Kg$ ) e maior dispersão entre  $\tilde{x}$  e  $q_3$  (quase  $10 kg$ ). Vale lembrar que temos 17 alunos neste grupo, portanto não é sensato inferir conclusões. Para o boxplot do centro temos dispersão maior entre a  $\tilde{x}$  e  $q_3$  e um candidato a outlier (ponto com valor distante do padrão apresentado pelas outras observações. Neste caso, dado por  $x^* < q_1 - 1,5(q_3 - q_1)$ ,  $x^* < q_1$ ). Este aluno pesa  $44,4 kg$ , contudo a quantidade de dados não permite tirar conclusões se o aluno tem o peso abaixo do padrão do grupo com 17 alunos. Para os 25 alunos que não dormem em horários regulares, a dispersão é maior entre  $q_1$  e  $\tilde{x}$ . Um aluno com  $66,9 kg$  é um candidato a outlier ( $x^* > q_3 + 1,5(q_3 - q_1)$ ,  $x^* > q_3$ ).

<sup>5</sup>As categorias 9 e 10 possuem 17 valores cada e a categoria Não possui 25 valores.

Vale salientar que, os alunos que dormem, em média, 8, 11 e 12 horas diárias não foram computados no gráfico porque possuem frequência insuficiente para gerar um boxplot.

Por fim, calculamos o *IMC* e *IAC*. Como já referido no texto deste capítulo, o *IMC* serviu apenas para embasamento, já que não é cientificamente confiável na faixa etária do público alvo. As análises ficaram concentradas no *IAC*. A título de informação, se fossemos levar em conta o *IMC*, teríamos 4 alunos com **obesidade I** e 3 alunos com **magreza leve**. Considerando o *IAC*, temos os dados apresentados na tabela 5.20.

Tabela 5.20: Índice de Adiposidade Corporal dos alunos,  $n = 68$ .

<i>IAC</i>	Alunos	<i>IAC</i>	Alunas
< 10	0	< 15	8
10 a 14	1	15 a 19	9
14 a 18	11	19 a 25	11
18 a 24	11	25 a 29	6
> 24	9	> 29	2
Total	32	Total	36

Lembrando que, segundo classificação da página 116, em ordem crescente de valores, as cinco classes do *IAC* são baixíssima, baixa, ideal, moderada e excesso de gordura. Temos na Tabela 5.20, 21 alunos (mais de 65%) e 25 alunas (quase 69%) fora do índice ideal. Considerando os garotos, temos que 20 estão com o Índice de Adiposidade Corporal, (*IAC*), acima do ideal e considerando as garotas, observamos 17 abaixo do ideal.

Pensando em probabilidade, temos que se escolhermos um aluno ao acaso, a probabilidade deste estar com *IAC* fora do ideal é superior a 0,65(65%), independentemente do gênero.

Estendendo para probabilidade condicional, temos que a probabilidade de um adolescente que participou do estudo ser homem dado que o índice é fora do ideal é de 0,4565; a probabilidade de ser uma garota dado que o *IAC* é abaixo do ideal é de 0,9444; A probabilidade de ser um garoto dado de que índice é acima do ideal é 0,7143; a probabilidade de ser uma garota dado que o adolescente tem excesso de gordura é de 0,3529; estes são alguns exemplos. Os resultados trazem evidências de que os problemas de obesidade e magreza acometem qualquer gênero. Nos dados observados, homens sofrem mais com obesidade e mulheres com magreza.



Para que pudéssemos visualizar do comportamento do Índice de Adiposidade Corporal em relação ao gênero, fizemos os histogramas apresentados nas Figuras 5.11 e 5.12.

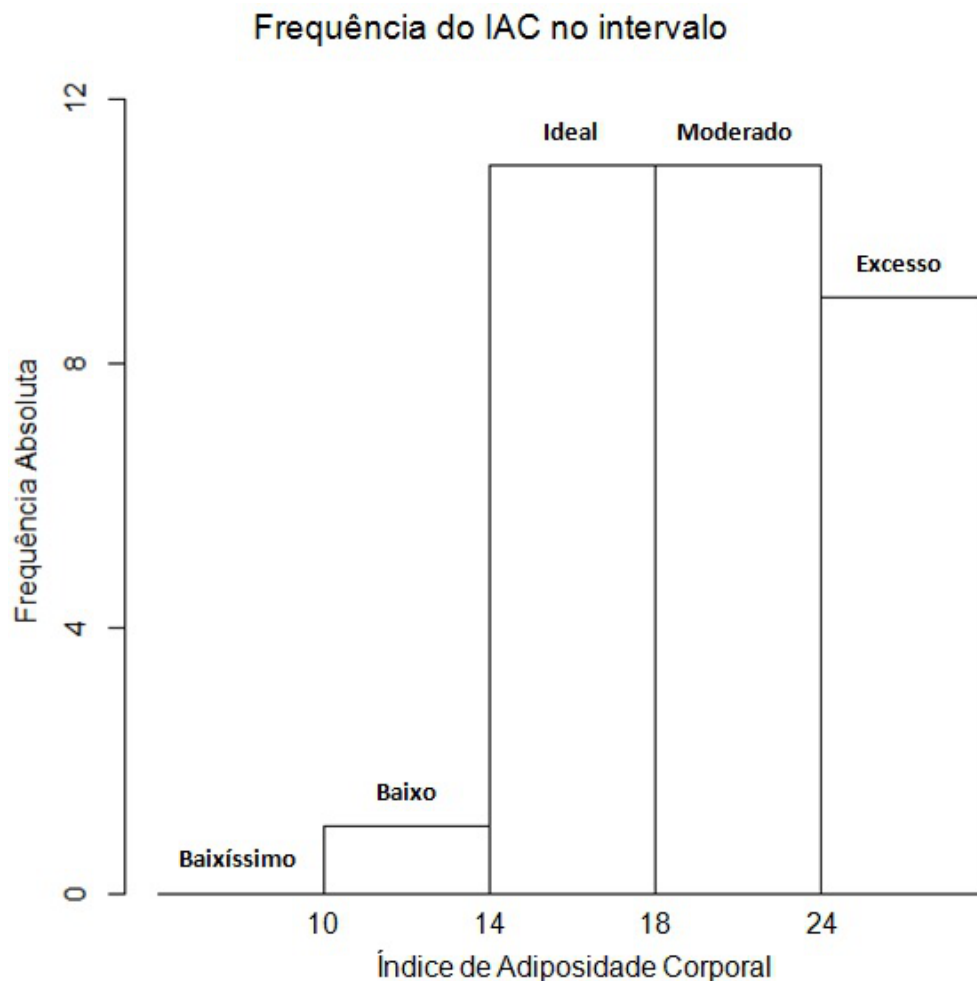


Figura 5.11: Frequência do *IAC* para o gênero masculino,  $n = 32$ .

A Figura 5.11 ilustra o que apresentamos nas duas primeiras colunas da Tabela 5.20. Através da Figura 5.11 podemos visualizar que os alunos do gênero masculino que participaram da pesquisa possuem uma propensão maior de estarem com o *IAC* nas categorias ideal e moderada. Temos 20 de 32 alunos (62,5%) com *IAC* acima do ideal, segundo dados da POF do IBGE, de 2008-9, os casos de obesidade infantil são preocupantes atingindo a 16,6% nos meninos e 11,8% nas meninas entre 5 e 9 anos, número superior a quatro vezes do que há 20 anos atrás.

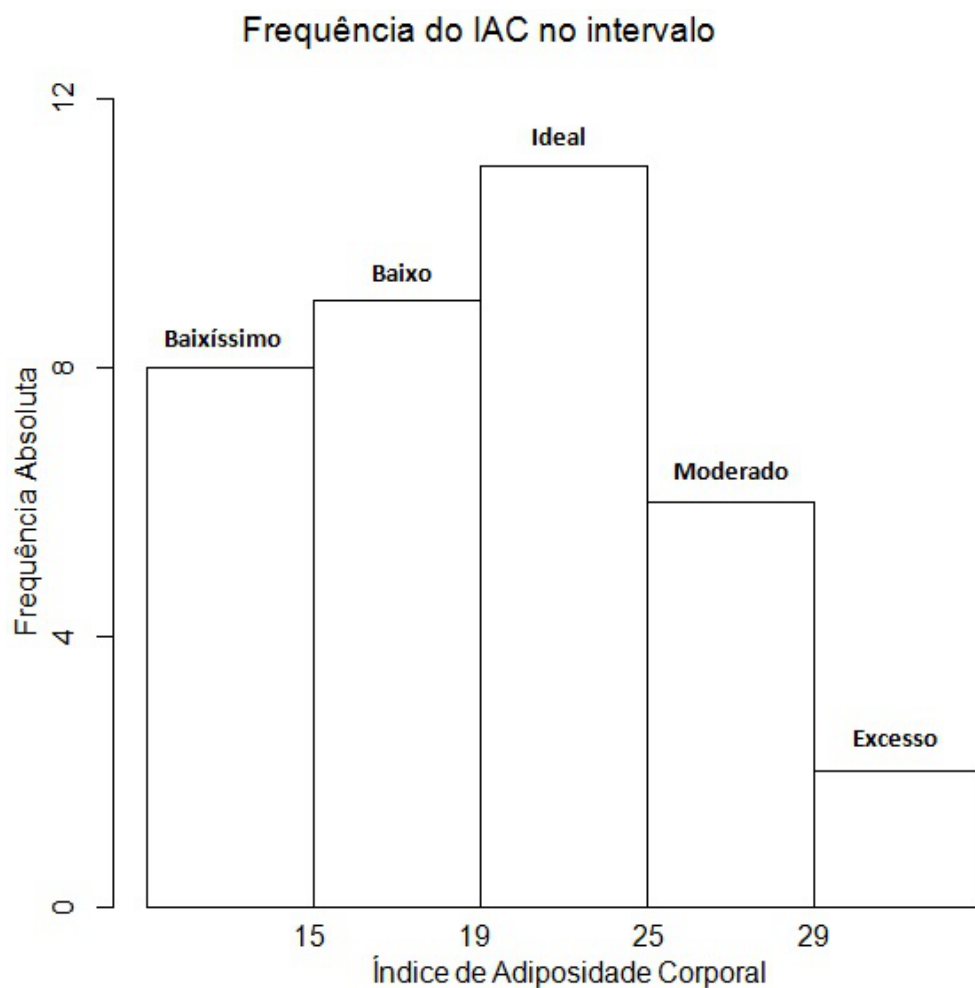


Figura 5.12: Frequência do *IAC* para o gênero feminino,  $n = 36$ .

Novamente a Figura 5.12 ilustra o que apresentamos nas duas últimas colunas da Tabela 5.20. Foram registrados 8 casos de *IAC* acima do ideal e uma 17 casos (quase metade do número de garotas que participaram da pesquisa) onde o índice calculado ficou abaixo do ideal.

A partir destes resultados, analisamos ficha por ficha para buscar entendimento dos índices observados.

Notamos que dos 18 alunos que possuem o índice abaixo do ideal, 7 não tomam café da manhã (um destes não almoça). Os 11 restantes tem o almoço como refeição mais nutritiva, sendo que 6 ingerem os cinco nutrientes essenciais. Temos que 14 desses alunos se alimentam apenas da merenda escolar nas refeições intermediárias

e praticam atividades esportivas de maneira integral nas aulas de educação física além de atividades físicas fora da escola. Vale ressaltar que 8 alunos fazem cursos como dança, informática, música; ou aulas de karatê, boxe e judô. Esses 18 alunos, gastam em média 2 horas com a utilização de aparelhos eletrônicos e 11 destes não se alimentam durante o uso. Por fim, 12 desses alunos possuem horários regulares e dormem, em média, 9 horas e 40 minutos por dia. Dessa forma, percebemos que estes 18 alunos, de forma geral, possuem uma rotina repleta de atividades extra escolares, tais afazeres, possivelmente, impedem o café da manhã ou faz que o mesmo não tenha os devidos nutrientes, além disso parecem gastar mais energia do que consome.

No entanto, o principal foco desta análise não é identificar a eventual causalidade, caso exista, da situação do aluno, e sim propôr, na medida do possível, orientações para ajudá-los a viver de maneira saudável.

Para os alunos que estão com níveis de adiposidade acima do ideal, vimos que 8 alunos costumam ingerir os cinco nutrientes em todas as refeições, sendo que os outros 20 possuem almoço e jantar com quantidades significativas de todos os nutrientes. Temos que 13 alunos tem o hábito de se alimentar bebendo suco ou refrigerante. Todos tem o hábito de realizar refeições intermediárias. Nove alunos ingerem guloseimas, merenda escolar, produtos da cantina escolar e alimentos naturais, 14 ingerem pelo menos duas dessas opções e os outros ingerem alimentos trazidos de casa ou comprados a caminho da escola. Vale salientar que, 4 alunos não praticam atividades físicas na escola e 13 as praticam parcialmente. Oito alunos não praticam atividades físicas fora da escola. Esses 28 alunos gastam, em média, 4 horas e 30 minutos utilizando aparelhos eletrônicos e 22 (quase 79%) destes realizam refeições utilizando-os, especificamente, televisão, computador e videogame. Por fim, 22 alunos não possuem horários regulares para dormir e acordar e os outros alunos dormem, em média, 11 horas e 20 minutos por dia. Percebemos que os alunos desse grupo possui uma rotina mais tranquila do que os alunos com *IAC* abaixo do ideal.

## 5.4 Conclusão

Após os resultados obtidos apresentamos e discutimos todas as informações construídas, mostrando aos alunos os benefícios de uma alimentação balanceada e prática de atividades físicas regulares. Uma das propostas foi sugerir aos alunos que criassem, e realizassem, um planejamento com horários diários para se alimentar de maneira saudável, estudar conteúdos escolares e momentos para lazer e diversão.

Sem apontar casos específicos nem dados de qualquer aluno, fizemos uma conscientização da importância de manter o peso e os índices de massa e adiposidade corporal ideais. Salientamos os benefícios à saúde que uma alimentação equilibrada

e a prática de atividades físicas acarretam, como satisfação, auto estima elevada, confiança, imunidade fisiológica, entre outros fatores. A melhor forma encontrada para realizar esta conscientização, foi mostrar os problemas que podem ser ocasionados por uma alimentação desprovida de nutrientes e falta de atividades físicas, como doenças causadas pelo desequilíbrio físico-energético que acarretam distúrbios físicos e mentais e vulnerabilidade fisiológica. Frisamos que as principais consequências são doenças como obesidade, colesterol alto, gastrite, diabetes e hipertensão, desnutrição, anemia, amenorréia (parada do ciclo menstrual), anorexia, bulimia, entre outras.

Incentivamos o consumo de alimentos naturais e diminuição no consumo dos alimentos processados e industrializados. Foi salientado que o exercício físico é sinônimo de melhora do sistema cardiorrespiratório, condicionamento físico, força, agilidade, coordenação motora, flexibilidade e equilíbrio, além de trazer confiança e satisfação pelo alívio das tensões emocionais, redução pressão arterial, estímulo ao emagrecimento, aumento da densidade óssea e massa muscular. Deixando evidente, aos que não praticam atividades físicas de maneira regular, que é importante começar essa prática com exercícios suaves para que o corpo se adapte e crie a fisiologia necessária para exercícios mais intensos.

Foi notória a melhora no senso crítico do aluno, tornando-o capaz de entender a importância das disciplinas do currículo e que todas estão relacionadas. Mais especificamente, que a matemática tem aplicação na vida cotidiana e é útil tanto para todas as profissões, além de ter uma linguagem universal, entendida em todos os países.

A pesquisa contribuiu para a aplicação de conceitos matemáticos conciliados com outras disciplinas, mais diretamente educação física e ciências biológicas, mostrando a utilidade das ferramentas aprendidas nas aulas destas disciplinas.

## Considerações Finais

Este trabalho teve como uma das principais finalidades mostrar a aplicação de conteúdos de matemática no cotidiano do aluno, mais especificamente tratamento de informações e probabilidade. Alguns alunos, geralmente desmotivados em aprender matemática fazem a seguinte pergunta: *“Para que eu preciso saber isso?”*. Baseados na ausência da resposta para essa pergunta, os alunos perdem o entusiasmo pelo aprendizado da matemática, pois se não possui utilidade então não é necessário aprender determinado conceito. Por isso, justificamos a proposta diferenciada construindo situações-problema que exigissem a utilização de conceitos aprendidos ao longo da vida escolar, visando responder de forma ativa, esta questão. Além de estimular o desenvolvimento do raciocínio lógico, matemático e científico, a execução desse trabalho visou tornar mais próximo do aluno o conhecimento matemático que, às vezes, é apresentado de forma extremamente abstrata, deixando-os inerte ao conhecimento. Vale ressaltar que, em momento algum, o conhecimento técnico e o vocabulário matemático formal foi dispensado, pelo contrário, sua importância foi frisada a todo momento, porém associada a uma forma prática de apresentação. A aplicação se deu, sempre, após a introdução e entendimento da teoria, não limitando este último a resolução de problemas e exercícios. Foi uma maneira de mostrar o uso de ferramentas matemáticas e raciocínio para resolver situações-problema na prática.

Para alcançarmos o objetivo de ensinar conceitos matemáticos de forma simples e prática, executamos experimentos com abordagens de diversas.

No experimento “Dobra a língua e coça a orelha” estimulamos a capacidade de interpretação com uma abordagem multidisciplinar. Nele associamos a hereditariedade de características próprias dos alunos, promovendo o estudo e interpretação de forma interativa e exploratória de características humanas. Primeiro, chamamos atenção para a classificação das características estudadas: dobrar a língua em forma de U e lóbulo da orelha colado; segundo, ensinamos uma maneira lógica de classificar estas características usando os próprios alunos como geradores de informação; depois, atribuindo aos mesmos a responsabilidade de serem pesquisadores e avaliadores e adquirir as informações necessárias com parentes consanguíneos; por fim, organi-

zando as informações em tabelas e gráficos e discutindo quais as conclusões que os dados permitiriam chegar.

No experimento “Jankenpon”, propomos uma forma lúdica de entender conceitos de probabilidade, associando-os com a habilidade de tratar informações por meio de tabelas e raciocínio lógico, olhar crítico e formulação de estratégias a partir do dissernimento matemático. Usamos ideias de uma forma de disputa muito conhecida pelos alunos.

No experimento que utiliza o método de Monte Carlo para aproximação de área de qualquer figura, apresentamos o conceito de probabilidade geométrica de forma multidisciplinar, pois a figura utilizada tinha o contorno do mapa do Brasil, e portanto, estimamos área da extensão territorial brasileira. Desta maneira, usamos conteúdos da disciplina de geografia e aspectos históricos das cidades do Monte Caburaí, RR, até Arroio Chuí, RS.

No experimento “Variáveis Antropométricas”, novamente chamamos atenção para aspectos corporais, usando a análise das informações para estabelecer e entender como estão relacionadas medidas corporais., visando a participação de todos, sem expor nomes e apresentando conclusões gerais.

Após a execução deste último experimento, houve uma preocupação acerca de alunos baixos com quadris largos e alunos altos com quadris finos, por isso fizemos uma pesquisa, que além da aplicação dos conceitos, teve uma preocupação com a qualidade da alimentação e a prática de atividades físicas, com a finalidade de orientá-los à alimentação balanceada e prática de exercícios físicos de maneira adequada.

Tais experimentos, juntamente com suas interpretações, análises e conclusões foram feitos sob o ponto de vista dos conteúdos teóricos, mostrando aplicação de conceitos como razão, proporção, frequências absoluta e relativa, porcentagem, construção e interpretação de tabelas e gráficos, raciocínio lógico, crítico e matemático, definições de variáveis aleatórias, probabilidade e probabilidade geométrica, proposições, teoremas, corolários, entre outros, construindo a base necessária para o aprendizado e entendimento dos objetivos da aplicação destes experimentos.

Outro ponto a ser destacado na construção deste trabalho, foi tornar o aluno mais participativo, se sentindo peça fundamental para que a aula possa transcorrer. Dois pontos notórios da realização deste trabalho são, primeiro alunos altamente colaborativos, participativos, curiosos e interessados quando a proposta de aula acopla pontos pessoais ou de interesse majoritário do corpo discente, sendo protagonistas com sugestões e ideias. Segundo, o fato de construir o ensino matemático baseado na multi e interdisciplinaridade desenvolve nos alunos a capacidade de relacionar conteúdos e disciplinas, entendendo que as disciplinas são ministradas de forma isolada, porém são parte de um todo com objetivos de formar um cidadão crítico, criativo,

participativo, capaz de relacionar informações, construir novas informações e tomar decisões que transforme a sociedade na qual estão inseridos.

Do primeiro ponto notório supra citado, que é chamar a atenção dos alunos para particularidades próprias. Mostramos dois lados de uma moeda, uma face com aluno-pesquisador, protagonista, articulador, produtor de dados e informações; e outra de aluno-pesquisado, objeto de estudo, coadjuvante, fonte de dados e informações. Ser essa “moeda informativa” aumentou o ímpeto dos alunos, principalmente dos que não participavam da aula. O “aprender mais de si” desperta uma inquietação nos alunos, estimula seu poder interrogativo e exclamativo, propaga o senso de estabelecer relações. Em diversos momentos, a associação com o aprendizado teórico foi feita pelos próprios alunos, como por exemplo: *“Professor, para podermos dizer se, proporcionalmente, mais meninas ou meninos fazem lição, precisamos calcular frequências relativas não é?”*. A proposta executada despertou interesse na maioria dos alunos. Observamos que as atividades em aulas tradicionais foram executadas por, no máximo, 12 de 40 alunos, enquanto nos experimentos, a realização chegou a 38 alunos. Para trabalhos extra escolares, a entrega era de 5, no máximo, enquanto todos (até os que se ausentaram) entregaram a pesquisa do experimento “dobra a língua e coça a orelha”. Acreditamos que a propagação do conhecimento aconteceu de forma mais homogênea e contemplativa nos experimentos.

Sobre o segundo ponto notório, acreditamos que preparar aulas inter e multidisciplinares com abordagens de temas da atualidade, como tecnologia e suas funcionalidades, mais especificamente, internet, facebook, instagram, entre outros, mostrando aos alunos que a matemática está presente no cotidiano deles e compreendê-la é importante e necessário. Enfim, às vezes, ensinamos a matemática de forma isolada, sem mostrar a sua aplicação, principalmente, à realidade do aluno, isto pode torná-la sem atratividade, quando mostramos a aplicabilidade e utilidade dos conteúdos de matemática em outras disciplinas e/ou no cotidiano do aluno, despertamos interesse em todos os alunos. Podemos planejar aulas que contextualize o aprendizado, usando como foco, por exemplo, o estudo de equipamentos eletrônicos, visando gerar entusiasmo nos alunos, além da troca de informações, pois, provavelmente, eles manejam com mais habilidade e conhecimento tais equipamentos.

Outro fator que tornou-se central, conforme o presente trabalho foi sendo executado, foi a necessidade de orientar os alunos a ter uma vida mais saudável, se alimentando de forma nutritiva e praticando regularmente atividades físicas e de lazer. Estas orientações foram dadas ao longo do 4º bimestre utilizando notícias, artigos e afins com informações sobre qualidade de vida. Também com auxílio dos professores de ciências/biologia e educação física, instruindo o aluno para o hábito de ter uma alimentação nutritiva e praticar atividades físicas para ficar livre ou menos

propenso a adquirir diversas doenças e evitar o sedentarismo.

Os alunos que participaram da construção deste, ensinaram que é importante acreditar no potencial deles, que vale a pena uma aula interativa, onde o aluno deixe de ser espectador e seja protagonista ou coadjuvante, participando da construção do ensino e aprendizado. Ensinaram que é preciso considerar as diferentes óticas sobre um mesmo tema e valorizá-los no planejamento das aulas, pois são mais capazes de realizar tarefas do que julgamos.

Enfim, pensamos ser necessárias e produtivas abordagens diversas e multidisciplinares dos conceitos matemáticos, mostrando a aplicação em atividades e experimentos que auxiliem os alunos na compreensão destes conceitos contribuindo para formar alunos críticos, capazes de relacionar informações, elaborar estratégias, resolver problemas do cotidiano e tomar decisões lógicas baseadas nestes conceitos matemáticos.



# Referências Bibliográficas

- [1] RIFO, L. L. R. **Dobra a língua e coça a orelha.** Universidade Estadual de Campinas - Unicamp - Matemática, Projeto Condigital MEC - MCT. 2010. Disponível em: <http://m3.ime.unicamp.br/recursos/1370>. Acesso em: 27/07/2013.
- [2] RIFO, L. L. R. **Jankenpon.** Universidade Estadual de Campinas - Unicamp - Matemática, Projeto Condigital MEC - MCT. 2010. Disponível em: <http://m3.ime.unicamp.br/recursos/1016>. Acesso em: 27/07/2013.
- [3] OLIVEIRA, S. R. **O método de Monte Carlo.** Universidade Estadual de Campinas - Unicamp - Matemática, Projeto Condigital MEC - MCT. 2011. Disponível em: <http://m3.ime.unicamp.br/recursos/1371>. Acesso em: 27/07/2013.
- [4] RIFO, L. L. R. **Variáveis antropométricas.** Universidade Estadual de Campinas - Unicamp - Matemática, Projeto Condigital MEC - MCT. 2011. Disponível em <http://m3.ime.unicamp.br/recursos/1038>. Acesso em: 27/07/2013.
- [5] ZEGARELLI, M. **Matemática para leigos.** 2. ed. Editora Alta books. 2011.
- [6] BUSSAB, W. O.; Morettin, P., A. **Estatística básica.** 8. ed. Editora Saraiva. 2013.
- [7] BOLFARINE, H.; Bussab, W. O. **Elementos de amostragem.** Editora Edgard Blucher, ABE - Projeto Fisher. 2005.
- [8] NETO, P. L. O. C. **Estatística.** 2. ed. Editora Edgard Blücher. 2002.
- [9] MEYER, P. L. **Probabilidade: Aplicações à Estatística.** Editora Livros Técnicos e Científicos. 2003.
- [10] SILVA, N. N. **Amostragem probabilística.** Editora da USP, São Paulo. 1998.
- [11] MURRAY. R. S. **Schaum's outline of theory and problems of statistics.** Editora McGraw-Hill, New York. 1965.

- [12] SANTOS, C. **Estatística descritiva: manual de auto-aprendizagem**. Editora Silabo, Lisboa. 2007.
- [13] TREES, H. L. V. **Detection, Estimation and Modulation Theory - Part I**. Editora Wiley. 2001.
- [14] BORNIA, A. C.; Barbetta, P. A.; Reis, M. M. **Estatística para cursos de engenharia e informática**. 3. ed. Editora Atlas. 2010.
- [15] MOORE, D. S. **The basic practice of statistics**. Editora Freeman, New York. 2007.
- [16] SOARES, J. L. **Biologia - volume único**. 2. ed. Editora Scipione. 1988.
- [17] REIS, G. M.; JÚNIOR, J. I. R. **Ferramentas estatísticas básicas da qualidade - guia prático do R**. Disponível em: <http://goo.gl/Z1O4Rl>. Acesso em: 12/12/2014. 2008.
- [18] TRIOLA, M. F. **Introdução à Estatística**. 7. Ed. Editora LTC, Rio de Janeiro. 1999.
- [19] SANTOS, G. E. O. **Cálculo amostral: calculadora on-line**. Disponível em: <http://www.calculoamostral.vai.la>. Acesso em: 20/12/2014.
- [20] CASELLA, G.; BERGER, R. L. **Inferência estatística - tradução da 2. edição norte americana**. Editora Centage Learning, Coleção Schaum, São Paulo. 2010.
- [21] MURRAY. R. S. **Estatística. Tradução de Pedro Cosentino**. Editora Ao livro técnico S.A., Rio de Janeiro. 1969.
- [22] NETTO, R. C. M. **Conceitos de genética - Dominante ou recessivo?**. *Revista Genética na Escola*. Publicada pela Sociedade Brasileira de Genética, Departamento de genética e biologia evolutiva, Instituto de Biociências, Universidade de São Paulo, v. VII, n. 2. 2012. Disponível em: <http://goo.gl/AuBskn>.
- [23] FELLER, W. **Introdução à teoria das probabilidades e suas aplicações**. Editora Edgard Blücher, São Paulo. 1976.
- [24] ROSS, S. **Probabilidade, um curso moderno com aplicações**. 8. ed. Editora Bookman Companhia Editora Ltda, divisão da Artmed Editora S.A., Copyright ©, 2010.

- [25] WAGNER, J. C.; HAGHIGHAT, A.; PETROVIC, B. G. **Monte Carlo transport calculations and analysis for reactor pressure vessel neutron fluence.** *Nuclear technology*. Nuclear Engineering Department, University Park, Pennsylvania, v. 114. 1996.
- [26] PARIS, M. W.; PANDHARIPANDE, V. R. **Quantum Monte Carlo calculations of six-quark states.** *Tese de Ph.D., University of Illinois at Urbana-Champaign*. 2000.
- [27] FRAASS, B. A.; SMATHERS, J.; DEYE, J. **Summary and recommendations of a National Cancer Institute workshop of Monte Carlo dose calculation algorithms for megavoltage external beam radiation therapy.** *Med Phys*, v. 30. 2003.
- [28] KAUPUZZ, J.; MAHNKEL, R. **A stochastic multi-cluster mode of freeway traffic.** *European Physical Journal B*, v. 14. 2000.
- [29] GIERSZ, M. **Monte Carlo simulations of star clusters - I. First results.** *Monthly Notices of the Royal Astronomical Society*, v. 298, n. 4, p. 1239-1248. 1998.
- [30] BINDER, K. **Monte Carlo Methods in Statistical Physics.** Editora Springer-Verlag, Berlin-Heidelberg-New York. 1979.
- [31] SCOTT, L. O. **Option pricing when the variance changes randomly: theory, estimation and an application.** *The Journal of Financial and Quantitative Analysis*, v. 22, n. 4, Copyright © School of Business Administration, University of Washington. 1987.
- [32] Wagner, E. **Probabilidade geométrica - O problema do macarrão e um paradoxo famoso.** *Revista do Professor de Matemática*, v. 34, p. 28-35, São Paulo. 1997.
- [33] STEINBRUCH, A.; WINTERLE, P. **Geometria analítica.** 2. ed. Editora Makron. 1987.
- [34] ÁVILA, G. S. S. **Cálculo das funções de uma variável - v. 1.** 7. ed. LTC - Livros Técnicos e Científicos Editora S.A., Rio de Janeiro. 2003.
- [35] ÁVILA, G. S. S. **Cálculo das funções de uma variável - v. 2.** 7. ed. LTC - Livros Técnicos e Científicos Editora S.A., Rio de Janeiro. 2004.

- [36] ÁVILA, G. S. S. **Cálculo com funções de várias variáveis - v. 3.** 7. ed. LTC - Livros Técnicos e Científicos Editora S.A., Rio de Janeiro. 2006.
- [37] CRUZ, A.; FILIPE, E.; ALMEIDA, G.; VALADARES, J.; PELLEGRINO, O. **Medições e incerteza de medição: Um contributo baseado nas convenções e resoluções internacionais.** *Caderno Brasileiro de Ensino de Física*, v. 26, n. 1, p. 125-135. 2009.
- [38] CHARNET, R.; BONVINO, H.; FREIRE, C. A. L.; CHARNET, E. M. R. **Análise de modelos de regressão linear com aplicações.** 2. ed. Editora da Unicamp. 2008.
- [39] SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear.** Editora Embrapa. 1998.
- [40] MACHADO, A. F. **Dobras Cutâneas: Localização e Procedimentos.** *Revista Motricidade*, v. 4, n. 2., p. 41-45. 2008.
- [41] BERGMAN, R.N.; STEFANOVSKI, D.; BUCHANAN, T.A.; SUMNER, A. E.; REYNOLDS, J.C.; SEBRING, N.G.; XIANG, A.H.; WATANABE, R.M. **A better index of body adiposity.** *Revista Obesity (Silver Spring)*, v. 19, n. 5, p. 1083-9. 2011.
- [42] VILARTA, R. **Alimentação Saudável, Atividade Física e Qualidade de Vida.** Editora Ipes Editorial. 2007.
- [43] DUFTY, W. **Sugar Blues: o Gosto Amargo do Açúcar.** Editora Ground. 2005.
- [44] LIRA, S. A. **Análise de correlação: abordagem teórica e de construção de coeficientes com aplicações.** Curitiba: UFP, 2004. 209 f. *Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Métodos Numéricos em Engenharia dos Setores de Ciências Exatas e de Tecnologia, Universidade Federal do Paraná, Curitiba.* 2004.
- [45] CARMO, I. **Alimentação Saudável: Alimentação Segura.** 3. ed. Editora Dom Quixote, coleção *Temas de Hoje - Viver Melhor.* 2004.
- [46] HALPERN, A. **Obesidade - Mitos e Verdades.** Editora Contexto. 1997.
- [47] BRASIL, Ministério da Saúde. **Guia alimentar para a população brasileira: promovendo a alimentação saudável.** Brasília, DF. p. 236. 2005.

# Apêndice

## Termo de Consentimento Livre e Esclarecido

Este estudo está dirigido pelo professor André da Silva Coura e visa analisar dados de saúde infantil com fins puramente acadêmicos para sua dissertação de conclusão do PROFMAT.

A metodologia da pesquisa será a aplicação de um questionário sem riscos previsíveis, custo ou remuneração ao respondente, que poderá parar de responder as perguntas em qualquer momento, sem prejuízo algum para nenhuma das partes. O sujeito têm o direito de se retirar do projeto a qualquer momento.

Suas informações serão tratadas de forma confidencial para uma análise anônima dos dados obtidos. Com o preenchimento do questionário, você não tem a obrigação de participar em nenhum estudo ou projeto futuro da Unicamp.

Uma cópia deste termo ficará em posse do responsável pelo menor.

Em caso de dúvidas, por favor, entre em contato com a professor André da Silva Coura, pelo telefone (19) 3227-3424, em horário do período escolar do menor (13:00 às 18:20).

Para denúncias e/ou reclamações referentes aos aspectos éticos desta pesquisa, procure o Comitê de Ética em Pesquisa da FCM/UNICAMP.

Rua: Tessália Vieira de Camargo, 126. C.E.P.: 13083-887, Campinas, SP.

Fone (019) 3521-8936 ou 3521-7187

e-mail: cep@fcm.unicamp.br

-----  
Eu, \_\_\_\_\_,  
pai/mãe/responsável por \_\_\_\_\_,  
aluno(a) da E.E. Elvira de Pardo Mêo Muraro, autorizo sua participação neste estudo.

Campinas, novembro de 2013.



## Questionário aos Alunos

Este questionário tem por objetivo analisar aspectos de uma vida saudável de alunos do ensino fundamental da EE Elvira de Pardo Mêo Muraro. Faz parte de um projeto a ser desenvolvido no curso de pós-graduação da UNICAMP e não haverá identificação dos respondentes. São perguntas simples e quaisquer dúvidas podem ser esclarecidas com o Prof. André da Silva Coura, responsável pelo projeto. Desde já agradeço sua ajuda e cooperação.

1. Gênero: ( ) Feminino                      ( ) Masculino
2. Altura (em cm): \_\_\_\_\_
3. Peso (em Kg): \_\_\_\_\_
4. Quadril (em cm): \_\_\_\_\_
5. O que você geralmente come/bebe no café da manhã?  
\_\_\_\_\_
6. O que você geralmente come/bebe no almoço?  
\_\_\_\_\_
7. O que você geralmente come/bebe no jantar?  
\_\_\_\_\_
8. Além dessas três refeições, você tem o hábito de se alimentar no intervalo entre elas?  
( ) Não.  
( ) Sim, alimentos como:  
    ( ) Guloseimas (balas, chicletes, pirulitos, chocolates, entre outros);  
    ( ) Lanches, Biscoitos, Merenda ou produtos da cantina escolar, exceto, guloseimas;  
    ( ) Frutas, sucos e lanches naturais;  
    ( ) Outros. Quais? \_\_\_\_\_
9. Nas aulas de Educação Física, você:  
( ) Realiza ativamente todas as atividades;  
( ) Realiza parcialmente as atividades;  
( ) Não realiza as atividades;  
( ) Dispensado(a) das aulas. Motivo: \_\_\_\_\_

10. Fora da escola, você pratica atividades físicas?

( ) Não.

( ) Sim. Quais? \_\_\_\_\_

**(Exemplos de atividades físicas:** tocar instrumento musical, arrumar a casa, jogar videogame, praticar algum esporte, brincar (andar de bicicleta, correr, pular), etc.)

11. Quantas horas por dia, aproximadamente, você gasta com: TV, computador, videogame e outros aparelhos eletrônicos?

Resposta: \_\_\_\_\_

12. Você come/bebe enquanto está usando: TV, computador, videogame ou outros aparelhos eletrônicos?

( ) Não.

( ) Sim. O quê? \_\_\_\_\_

13. Você tem horários regulares para dormir e acordar?

( ) Sim, todos os dias, vou dormir às \_\_\_\_ horas e acordo às \_\_\_\_ horas.

( ) Sim, dias de semana, vou dormir às \_\_\_\_ horas e acordo às \_\_\_\_ horas, e, fins de semana: durmo às \_\_\_\_ horas e acordo às \_\_\_\_ horas.

( ) Não, durmo e acordo que horas quiser, sem horários pré-determinados.



## Anexos

### Termo de Ciência e concordância do Coordenador Pedagógico/Diretor

Através desse termo, eu, Angélica Carnicelli Adam, coordenadora pedagógica do ensino fundamental II, da unidade escolar E.E. Elvira de Pardo Mêo Muraro, estou ciente que, o professor André da Silva Coura, realizará a pesquisa de 13 questões, com alunos do ensino fundamental II, para fins puramente acadêmicos para a sua dissertação de mestrado do PROFMAT, que será entregue a Unicamp, sobre seu projeto "Vida com Qualidade: Fusão de alimentação saudável e atividades físicas", no mês de novembro de 2013.

A metodologia de pesquisa será aplicação de um questionário, e não tem riscos previsíveis nem custos para o respondente, que poderá parar de responder às perguntas em qualquer momento, sem prejuízo algum.

Suas informações serão tratadas de forma confidencial, para uma análise anônima dos dados obtidos. Com o preenchimento do questionário, o aluno não tem a obrigação de participar em nenhum estudo ou projeto futuro da Unicamp.

Uma cópia deste termo ficará em posse do Coordenador Pedagógico do Ensino da escola EE Elvira de Pardo Mêo Muraro, assim como uma cópia do Termo de Consentimento Livre e Esclarecido encaminhado aos pais dos alunos participantes da pesquisa.

Em caso de dúvidas, por favor, entre em contato com a Prof.<sup>a</sup> André da Silva Coura, pelo telefone (19) 3227-3424, em horário do período escolar do menor (13:00 às 18:20).

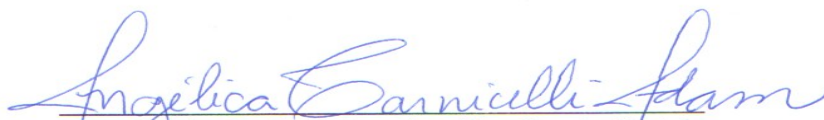
Para denúncias e/ou reclamações referentes aos aspectos éticos desta pesquisa, procure o Comitê de Ética em Pesquisa da FCM/UNICAMP.

Rua: Tessália Vieira de Camargo, 126 – CEP 13083-887 Campinas – SP

Fone (019) 3521-8936 ou 3521-7187

e-mail: cep@fcm.unicamp.br

Sem mais esse Termo de Ciência assinado abaixo



Coordenador Pedagógico: Angélica Carnicelli Adam

EE Elvira de Pardo Mêo Muraro, Campinas-SP

Rua Vicente March, s/nº, Jardim Florence I

Telefone para contato: (19) 3227-3424



Campinas, 06 de novembro de 2013



## Autorização do diretor do IMECC

Campinas, 4 de novembro de 2013.

Prof. Dr. Caio Negreiros  
Diretor  
IMECC

Caro Professor Caio,

Venho por meio desta solicitar autorização para a realização da pesquisa associada à dissertação de meu orientando, Prof. André da Silva Coura, aluno do curso de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT).

O projeto de pesquisa inclui a aplicação de questionário em alunos da escola onde André trabalha, motivo pelo qual este documento é necessário para a apresentação ao Comitê de Ética em Pesquisa da FCM, Unicamp.

O projeto intitulado: "Vida com Qualidade: Fusão de alimentação saudável e atividades físicas" encontra-se em anexo para sua apreciação.

Agradecendo desde já sua resposta, despeço-me cordialmente,



Prof. Dra. Laura Leticia Ramos Rifo  
Departamento de Estatística

Ciente.  
De acordo



Prof. Dra. Verônica Andrea González López  
Diretora Associada  
IMECC - UNICAMP  
Matric. 28485-3



## Parecer consubstanciado do CEP

FACULDADE DE CIENCIAS  
MEDICAS - UNICAMP  
(CAMPUS CAMPINAS)



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Vida com Qualidade: Fusão de alimentação saudável e atividades físicas.

**Pesquisador:** Andre da Silva Coura

**Área Temática:**

**Versão:** 2

**CAAE:** 24534014.1.0000.5404

**Instituição Proponente:** Instituto de Matemática, Estatística e Computação Científica

**Patrocinador Principal:** Universidade Estadual de Campinas - UNICAMP

#### DADOS DO PARECER

**Número do Parecer:** 550.920

**Data da Relatoria:** 17/03/2014

#### Apresentação do Projeto:

O professor dos alunos do ensino fundamental II de um escola pública é o proponente da pesquisa no âmbito de seu projeto de mestrado no IMECC/UNICAMP. Sua proposta é estabelecer uma amostra para as aulas de estatística com a própria classe de estudantes, onde, a partir das respostas de um questionário para avaliar aspectos de uma vida saudável, serão estudados conceitos e ferramentas da disciplina.

#### Objetivo da Pesquisa:

- Estimular hábitos de uma vida saudável.
- Ensinar conceitos e o uso de ferramentas de estatística.

#### Avaliação dos Riscos e Benefícios:

O autor esclareceu pontos importantes sobre a forma de aplicação do questionário, porém os riscos que se referem-se "tratamento que será dado às conclusões da pesquisa em sala de aula do ponto de vista dos alunos" permanecem. Isto deverá ser tratado adequadamente pelo autor no momento de discutir o assunto em sala com os alunos. Havendo preocupação de evitar qualquer tipo de identificação, os riscos poderão ser minimizados e os danos evitados.

#### Comentários e Considerações sobre a Pesquisa:

Embora o autor demonstre preocupação com a não identificação da letra dos alunos e do sigilo na

**Endereço:** Rua Tessália Vieira de Camargo, 126

**Bairro:** Barão Geraldo

**CEP:** 13.083-887

**UF:** SP

**Município:** CAMPINAS

**Telefone:** (19)3521-8936

**Fax:** (19)3521-7187

**E-mail:** cep@fcm.unicamp.br

FACULDADE DE CIENCIAS  
MEDICAS - UNICAMP  
(CAMPUS CAMPINAS)



Continuação do Parecer: 550.920

hora do preenchimento, enfatizamos que no momento da análise dos dados em classe, deve-se tomar cuidado para que os alunos com características marcantes não sejam identificados pelos demais e possam vir a sofrer qualquer tipo de assédio.

**Considerações sobre os Termos de apresentação obrigatória:**

Foram apresentados o projeto original, o formulário gerado pela Plataforma Brasil, Folha de Rosto devidamente assinada e o TCLE que estão condizentes com as premissas da Resolução 466-2012, CNS/MS. Cronograma adequado.

**Recomendações:**

---

**Conclusões ou Pendências e Lista de Inadequações:**

Projeto aprovado após resolução de pendências.

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

**Considerações Finais a critério do CEP:**

Cabe ao pesquisador desenvolver o projeto conforme apresentado nesta plataforma, elaborar e apresentar os relatórios parcial e final, bem como encaminhar os resultados para publicação, com os devidos créditos aos pesquisadores associados e ao pessoal técnico participante do projeto (Resolução 466/2012 CNS/MS).

CAMPINAS, 10 de Março de 2014

---

**Assinador por:**  
**Fátima Aparecida Bottcher Luiz**  
**(Coordenador)**

Endereço: Rua Tessália Vieira de Camargo, 126  
Bairro: Barão Geraldo CEP: 13.083-887  
UF: SP Município: CAMPINAS  
Telefone: (19)3521-8936 Fax: (19)3521-7187 E-mail: cep@fcm.unicamp.br