Universidade Federal de Goiás
Instituto de Matemática e Estatística
Programa de Mestrado Profissional em
Matemática em Rede Nacional

Estatística Descritiva, Probabilidade e Estimação: Noções para o Ensino Básico

Evandro de Moura Rios

Goiânia

2014

sibi sistema de bibliotecas ufg

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR ELETRONICAMENTE OS TRABALHOS DE CONCLUSÃO DE CURSO NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **Trabalho de Conclusão de Curso de Mestrado Profissional**

2. Identificação do Trabalho

Autor (a):	**Evandro de Moura Rios**
E-mail:	**professorevandrorios@gmail.com**
Seu e-mail pode ser disponibilizado na página?	[**X**]Sim [ ] Não
Vínculo empregatício do autor	**Mestrando Bolsista**
Agência de fomento:	**Coord. Aperf. De Pessoal de Nível Superior** — Sigla: **CAPES**
País: **Brasil**	UF: **Go** — CNPJ: **00889834/0001-08**
Título:	**Estatística Descritiva, Probabilidade e Estimação: Noções para o Ensino Básico.**
Palavras-chave:	**Estatística Descritiva, Probabilidade, Inferência Estatística, População, Amostra, Educação Básica.**
Título em outra língua:	**Descriptive Statistics, Probability and Estimation: Concepts for Basic Education.**
Palavras-chave em outra língua:	**Descriptive Statistics, Probability, Statistical Inference, Population, Sample, Basic Education.**
Área de concentração:	**Matemática do Ensino Básico**
Data defesa: (dd/mm/aaaa)	**26/09/2014**
Programa de Pós-Graduação:	**PROFMAt/UFG**
Orientador (a):	**Prof. Dr. Valdivino Vargas Júnior**
E-mail	**vvjunior@gmail.com**
Co-orientador(a):*	
E-mail:	

*Necessita do CPF quando não constar no SisPG

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [**X**] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF ou DOC do trabalho de conclusão de curso.

O sistema da Biblioteca Digital de Teses e Dissertações garante aos autores, que os arquivos contendo eletronicamente as teses, dissertações ou trabalhos de conclusão de curso, antes de sua disponibilização, receberão procedimentos de segurança, criptografia (para não permitir cópia e extração de conteúdo, permitindo apenas impressão fraca) usando o padrão do Acrobat.

Assinatura do (a) autor (a) — Data: 26 / 10 / 2014

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Evandro de Moura Rios

Estatística Descritiva, Probabilidade e Estimação: Noções para o Ensino Básico

Trabalho de Conclusão de Curso apresentado ao Instituto de Matemática e Estatística da Universidade Federal de Goiás, como parte dos requisitos para obtenção do grau de Mestre em Matemática.
Área de Concentração: Matemática do Ensino Básico
Orientador: Prof. Dr. Valdivino Vargas Júnior

Goiânia

2014

Dados Internacionais de Catalogação na Publicação na (CIP)
GPT/BC/UFG

Rios, Evandro de Moura.
R586e Estatística Descritiva, Probabilidade e Estimação: Noções para o Ensino Básico [manuscrito] / Evandro de Moura Rios. - 2014.
109 f. : il., figs, tabs.

Orientador: Prof. Dr. Valdivino Vargas Júnior;
Dissertação (Mestrado) – Universidade Federal de Goiás, Instituto de Matemática e Estatística, 2014.
Bibliografia.
Inclui lista de figuras.
1. Estatística matemática 2. Probabilidade e estatística I. Título.

CDU: 519.2

Evandro de Moura Rios

Estatística Descritiva, Probabilidade e Estimação: Noções para o Ensino Básico

Trabalho de Conclusão de Curso defendido no Programa de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT/UFG, do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Matemática, área de concentração Matemática do Ensino Básico, aprovado no dia 26 de setembro de 2014, pela Banca Examinadora constituída pelos professores:

Valdivino Vargas Júnior

Prof. Dr. Valdivino Vargas Júnior
Instituto de Matemática e Estatística-UFG
Presidente da Banca, Orientador

Prof. Dr. André Krindges
Departamento de Matemática/ICET/UFMT

Profª. Drª. Tatiane Ferreira do Nascimento Melo e Silva
Instituto de Matemática e Estatística-UFG

Evandro de Moura Rios graduou-se em Matemática pela Universidade Federal de Goiás, especializou-se em Ensino de Matemática para Educação Básica e Superior pela Universidade Estadual de Goiás. É professor do ensino fundamental e médio em goiânia na rede pública e privada, e professor universitário na Faculdade Alfa.

Ao "Seu Gerson", meu pai, que embora ausente, se faz presente em todos os momentos pelo exemplo dado em vida aos seus...Estar presente é muito relativo!

Agradecimentos

Primeiramente a Deus, por todas as benções recebidas, pois tenho tido muito mais do que pedi ou esperava ter.

Aos meus pais Gerson e Marta, meus irmãos Miguel e Ester, meus filhos Matheus e Letícia, aos meus sogros Onofre e Maria José, cunhadas Karina e Camila, assim como os seus respectivos Rodrigo e Gustavo. Enfim, toda a minha família, pois se existe algo ou alguém em que podemos depositar toda nossa confiança, é na nossa família. Ela nos mostra o que é certo, indica os melhores caminhos, e nos proporciona um amor verdadeiro e incondicional, mesmo nos momentos mais críticos.

Ao Professor Valdivino pela paciência na orientação e incentivo que tornaram possível a conclusão desta monografia, e pricipalmente, pelas suas aulas que não apenas nos ensinaram, mas nos inspiraram.

A todos os colegas pelos anos de convívio e aprendizado, em especial Ricardo Ferreira da Cunha, Marcelo Honório dos Santos e meu ex-aluno (com muito orgulho) Normando Silva Júnior, companheiros de tardes e noites de estudo. Aprendi muito com vocês.

Aos professores Aline de Souza Lima, Jesus Carlos da Mota, José Yunier Bello Cruz, Lucimeire Alves de Carvalho, Mário José de Souza, Ole Peter Smith, Paulo Henrique de Azevedo Rodrigues e Rogério de Queiroz Chaves, pelos dias de dedicação ao nosso curso.

Agradeço à CAPES pelo suporte financeiro. A bolsa de mestrado oportunizou a redução da carga de horário de trabalho, que mesmo por breve período, viabilizou os estudos tão necessários para a formação sólida que obtive nestes dois anos junto à UFG.

E a você, Tatiana, meu amor...Essa conquista não é minha. É nossa! Obrigado...

Resumo

Neste trabalho descrevemos toda a Estatística Descritiva e boa parte da Estatítica Inferencial. Após este referencial teórico, procuramos fazer um paralelo entre o que é estudado desde o ensino fundamental e o que é feito nas graduações. Concluímos com algumas descrições de aulas práticas que trabalharam alguns conceitos básicos da Inferência Estatística.

Palavras-chave

Estatística Descritiva, Probabilidade, Inferência Estatística, População, Amostra, Educação Básica.

Abstract

In this paper we describe all the descriptive statistics and much of Statistical Inference. After this theoretical framework, we try to draw a parallel between what is studied since elementary school, and what is done graduations. We conclude with some descriptions of practical lessons that worked some basic concepts of statistical inference.

Keywords

Descriptive Statistics, Probability, Statistical Inference, Population, Sample, Basic Education.

Lista de Figuras

Sumário

1 Introdução

Partindo do pressuposto de que a Estatística deve ser apresentada como uma ferramenta de análise de dados imprescindível aos profissionais das mais diversas áreas, foi norte para este trabalho, a análise do que é estudado desde o ensino fundamental até parte das graduações acerca desta ciência.

Neste trabalho, procura-se primeiramente, realizar um estudo amplo de toda a parte de estatística descritiva e inferencial. Feito este estudo, como base teórica, é traçado um paralelo do que é feito no ensino básico em relação às ideias básicas da estatística e o que é feito depois somente nas graduações.

É indiscutível a importância da estatística nos mais diversos campos do conhecimento, e a prova disso, é a existência desta matéria nos mais diversos cursos de graduação, como por exemplo: administração, ciências contábeis, pedagogia, economia, engenharias, psicologia, biologia, medicina, agronomia, zootecnia, publicidade, direito, entre outras.

Considerando a importância da estatísitica nos mais diversos campos, surge uma dúvida em relação a falta de importância dada a esta ciência no ensino básico.

Percebe-se, no ensino básico, um distanciamento considerável no que diz respeito as ideias de análise combinatória, probabilidade, e noções básicas de estatística, tratadas nesta fase da educação como tratamento da informação.

Acredita-se que se pode dar muito mais significado a estes temas, ainda no ensino fundamental. Trabalhando com ideias e experimentos mais concretos, foi experimentado e verificado até em turmas mais jovens, a capacidade de entender mais sobre aleatoriedade, população e amostras, análise de amostras e a representatividade das conclusões tiradas em amostras sobre determinadas populações.

Encerra-se este trabalho, com descrições das aulas dadas e resultados obtidos em cada experimento, com o objetivo de motivacionar outros profissionais da educação básica a trabalhar de forma mais significativa a Estatística deste o Ensino Fundamental.

2 Um Breve Histórico

Primeiramente, vamos localizar a Estatística na linha do tempo. Embora a palavra Estatística ainda não existisse, há indícios de que antes da Era Cristã, já eram realizados censos na Babilônia, China e Egito.

Segundo Costa [5], a própria Bíblia nos traz referência a Estatística, no quarto livro de Números, quarto livro do antigo testamento, onde no primeiro versículo, é dada a Moisés a instrução de que fosse feito o recenseamento de toda a comunidade de Israel, buscando homens que pudessem servir o exército.

O Imperador César Augusto (Que nasceu Caio Júlio César Otaviano Augusto em 23 de setembro de 63 a.C.) foi o primeiro imperador romano, e solicitou em seu governo, que se fizesse o censo de todo Império Romano. A palavra "CENSO" é derivada da palavra "CENSERE", que em Latim significa "TAXAR".

Ainda, segundo Costa [5], em meados de 1085, após conquistar a Inglaterra, Guilherme I solicitou um levantamento de todas as propriedades e seus respectivos proprietários, com a intenção de estabelecer uma taxação. Todo este registro foi feito em um livro intitulado "Domesday Book".

Mas, mesmo que a prática de coletar dados sobre populações, propriedades, animais, produção e impostos fosse conhecida pelos egípcios e mesopotâmicos bem antes de Cristo, somente no século XVII a Estatística passou a ser considerada disciplina autônoma, tendo como objetivo básico a descrição dos Bens do Estado.

A palavra Estatística foi cunhada pelo historiador e jurista alemão Gottfried Achenwall (1719-1772), mas com significados e objetivos diferentes daqueles que fundamentam a Estatística Moderna. Na Enciclopédia Britânica, o verbete "STATISTICS" apareceu em 1797.

Na Inglaterra do século XVII, surgiram os aritméticos políticos, dentre os quais destacaram-se John Graunt (1620-1674) e William Petty (1623-1687). Graunt nasceu em 24 de Abril de 1620 em Londres, Inglaterra. Filho de um comerciante, ocupou diversos cargos civis. Embora não tenha estudado em nenhuma universidade, viveu um período de grande atividade intelectual. Em 1662, Graunt publicou o famoso Natural and Political Observations on the London Bills of Mortality (Observações Naturais e Políticas da taxa de mortalidade londrina). Este foi o seu primeiro tratamento estatístico de dados demográficos. Estudou a mortalidade da cidade de Londres e as incidências das causas naturais, sociais e políticas nesse fenômeno. Através das Tábuas de Mortalidade realizadas na altura da peste na cidade de Londres, Graunt fez uma análise exaustiva do número de pessoas que morriam de várias doenças e estimou o número de nascimentos de homens e mulheres. Foi o primeiro estatístico a fazer observações entre sexos e mostrou que nasciam mais homens que mulheres. Com o seu amigo William Petty (1623-1687) fundou a escola dos "Aritméticos Políticos" que se preocupava com o estudo numérico dos fenômenos sociais e políticos, na busca de leis quantitativas que pudessem explicá-los. Dessa forma, a escola dos "Aritméticos Políticos"pode ser considerada o berço da Demografia.

Na última metade do século XIX, os alemães Helmert (1843-1917) e Wilhelm Lexis (1837-1914), o dinamarquês Thorvald Nicolai Thiele (1838-1910) e o inglês Francis Ysidro Edgeworth (1845-1926), obtiveram resultados extremamente valiosos para o desenvolvimento da Inferência Estatística, muitos dos quais só foram completamente compreendidos mais tarde. Contudo, o impulso decisivo deve-se a Karl Pearson (1857-1936), William S. Gosset (1876-1937) e, em especial, a Ronald A. Fisher (1890-1962).

Karl Pearson (1857-1936) formou-se em 1879 pela Cambridge University e inicialmente dedicou-se ao estudo da evolução de Darwin, aplicando os métodos estatísticos aos problemas biológicos relacionados com a evolução e hereditariedade. Em 1896, Pearson foi eleito membro da Royal Society of London. Entre 1893 e 1912 escreveu um conjunto de 18 artigos denominado Mathematical Contribution to the Theory Evolution, com contribuições extremamente importantes para o desenvolvimento da teoria da Análise de Regressão e do Coeficiente de Correlação, bem como do teste de hipóteses de qui-quadrado. Em sua maioria, seus trabalhos foram publicados na revista Biometrika, que fundou em parceria com Walter Frank Raphael Weldon (1860-1906) e Francis Galton (1822-1911). Além da valiosa contribuição que deu para a teoria da regressão e da correlação, foi um grande contribuidor para o desenvolvimento da estatística como uma disciplina científica séria e independente. Foi o fundador do Departamento de Estatística Aplicada na University College London em 1911 que foi o primeiro departamento universitário dedicado à estatística em todo o mundo.

William Sealey Gosset (1876-1937) estudou Química e Matemática na New College Oxford. Em 1899 foi contratado como Químico da Cervejaria Guiness em Dublin, desenvolvendo um trabalho extremamente importante na área de Estatística. Devido à necessidade de manipular dados provenientes de pequenas amostras, extraídas para melhorar a qualidade da cerveja, Gosset derivou o teste t de Student baseado na distribuição de probabilidades. Esses resultados foram publicados em 1908 na revista Biometrika, sob o pseudônimo de Student, dando origem a uma nova e importante fase dos estudos estatísticos. Gosset usava o pseudônimo de Student, pois a Cervejaria Guiness não desejava revelar aos concorrentes os métodos estatísticos que estava empregando no controle de qualidade da cerveja.

A contribuição de Ronald Aylmer Fisher (1890-1962) para a Estatística Moderna é, sem dúvidas, a mais importante e decisiva de todas. Formado em astronomia pela Universidade de Cambridge em 1912, foi o fundador do célebre Statistical Laboratory da prestigiosa Estação Agronômica de Rothamsted, contribuindo enormemente tanto para o desenvolvimento da Estatística quanto da Genética. Ele apresentou os princípios de planejamento de experimentos, introduzindo os conceitos de aleatorização e da Análise da Variância, procedimentos muito usados atualmente. No princípio dos anos 20, estabeleceu o que a maioria aceita como a estrutura da moderna Estatística Analítica, através do conceito da verossimilhança.

Outra área de investigação extremamente importante para o desenvolvimento da Estatística é a Teoria das Probabilidades. Usualmente, costuma-se atribuir a origem do Cálculo de Probabilidades às questões relacionadas aos jogos de azar que o célebre cavaleiro Méré (1607-1684) encaminhou à Blaise Pascal (1623-1662).

No entanto, outros autores sustentam que o Cálculo de Probabilidades teve a sua origem na Itália, com especial referência para Luca Pacioli (1445-1517), Girolamo Cardano (1501-1576), Nicolo Fontana Tartaglia (1500-1557) e Galileo Galilei (1564-1642). Três anos depois de Pascal ter previsto que a "aliança do rigor geométrico" com a "incerteza do azar" daria lugar a uma nova ciência, Christiaan Huygens (1629-1695) publicou o trabalho denominado "De Raciociciis in Ludo Aleae", que é considerado o primeiro livro sobre o Cálculo de Probabilidades. Além disso, ainda teve a notável particularidade de introduzir o conceito de esperança matemática.

Gottfried Wilhelm von Leibniz (1646-1716) também dedicou-se ao estudo do Cálculo de Probabilidades, publicando um trabalho sobre a "arte combinatória" e outro sobre aplicações às questões financeiras. Leibniz também estimulou Jacques Bernoulli (1654-1705) ao estudo do Cálculo de Probabilidades, cuja grande obra, denominada "Ars Conjectandi", foi publicada oito anos após a sua morte. Em Ars Conjectandi de Jacques Bernoulli, foi publicada e rigorosamente provada a Lei dos Grandes Números de Bernoulli, considerada o primeiro teorema limite. Pode-se dizer que graças às contribuições de Bernoulli o Cálculo de Probabilidades adquiriu o status de ciência. Além da obra póstuma de Bernoulli, o início do século XVII foi marcado pelos estudos de Pierre Rémond de Montmort (1678-1719),e de Abraham De Moivre (1667-1754), intitulado The Doctrine of Chances.

É extremamente importante falar, também, do reverendo Thomas Bayes (1702-1761) a quem se deve o conceito de probabilidade inversa, relacionado com situações em que se caminha do particular para o geral. Bayes formula através do teorema que leva seu nome e do postulado que tantas vezes se lhe associa: a primeira tentativa de matematização da inferência Estatística.

Os estudos dos astrônomos Pierre-Simon Laplace (1749-1827), Johann Carl Friedrich Gauss (1777-1855) e Lambert Adolphe Jacques Quetelet (1796-1874) foram fundamentais para o desenvolvimento do Cálculo de Probabilidades. Devido aos novos métodos e idéias, o trabalho de Laplace de 1812, intitulado "Théorie Analytique des Probabilités", até o presente é considerado um dos mais importantes trabalhos sobre a matéria.

Johann Carl Friedrich Gauss, professor de astronomia e diretor do Observatório de Gottingen, em 1809 apresentou o estudo intitulado "Theoria combinationis Observatorium Erroribus Minimis Obnoxia", explanando uma teoria sobre a análise de observações aplicável a qualquer ramo da ciência, alargando o campo de aplicação do Cálculo de Probabilidades.

Antoine Augustin Cournot (1801-1877) percebeu a importância da Teoria das probabilidades na análise estatística, tendo sido o pioneiro no tratamento matemático dos fenômenos econômicos. Suas ideias foram publicadas em "Exposition de la théorie des chances et des probabilités".

Na segunda metade do século XIX a Teoria das Probabilidades atingiu um dos pontos mais altos com os trabalhos da escola russa fundada por Pafnuty Lvovich Chebyshev (1821-1894), que contou com representantes como Andrei Andreyevich Markov (1856-1922) e Aleksandr Mikhailovich Lyapunov (1857-1918). Contudo, o seu maior expoente foi Andrey Nikolayevich Kolmogorov (1903-1987), a quem se deve um estudo indispensável sobre os fundamentos da Teoria das Probabilidades, denominado "Grundbegrife der Warscheinlichkeitrechnung", publicado em 1933. Em 1950 foi traduzido para o Inglês sob o título "Foundations of Probability".

Segundo Pardal [16], no Brasil, em 1863, foi criada na Escola Central, sucessora da Escola Militar, a cadeira de Economia Política, Estatística e Princípios de Direito Administrativo. O primeiro a ocupar esta cadeira foi José da Silva Paranhos, que inclusive, realizou o primeiro censo geral do império. Apenas em 1925, ainda segundo Pardal [16], o estudo da estatística teórica, foi introduzida na Escola Politécnica, na cadeira assumida por Jorge Kafuri, chamada "Estatística, Economia Política e Finanças". Em 1974, tem-se o primeiro vestibular para o curso de bacharel em Estatística, na Universidade Estadual do Rio de Janeiro.

Este interesse em discutir um pouco acerca da história da Estatística, é nos posicionar e compreender melhor o conteúdo que é ensinado hoje nas escolas, tendo em vista que apenas no final do século passado, a Estatística passou a ocupar lugar de destaque no mundo acadêmico. Cabe aqui expor a intenção de buscar uma ponte mais concreta entre o que é proposto no ensino médio e o que é de fato aplicável na graduação, seja qual for o campo do conhecimento.

3 Sobre Inferência

Um questionamento frequente dos leigos em relação à estatística, é como os pesquisadores podem afirmar com convicção aquilo que poderá acontecer num momento futuro, como por exemplo o resultado de uma eleição envolvendo milhões de eleitores, se eles consultam apenas algumas centenas ou poucos milhares de eleitores? Ou como pode uma rádio local, ter definida a sua audiência, se apenas alguns ouvintes foram pesquisados?

O que para muitos parece mágica, para os mais estudiosos, é uma ciência, ou mais precisamente, um ramo da matemática, chamado Estatística.

Cabe aqui, esclarecer que a palavra estatística tem dois significados. No sentido mais amplo, estatística significa um conjunto de dados numéricos, como por exemplo, o desempenho de um time de futebol, o número de nascimentos e óbitos em um país, o PIB de um país, e assim por diante. A palavra estatística, como já foi citado anteriormente, também designa um ramo da matemática que analisa dados estatísticos.

A princípio, se estuda a estatística descritiva, que é o processo de aquisição de informações a partir de conjuntos de números, em geral demasiadamente grandes para serem trabalhados diretamente, como por exemplo, toda a população de um país. Em seguida, se estuda a inferência estatística, que consiste em estimar propriedades de uma grande população a partir de observações feitas em uma amostra desta.

> "O uso de informações de uma amostra para concluir sobre o todo faz parte da atividade diária da maioria das pessoas. Basta observar como uma cozinheira verifica se o prato que ela está preparando tem ou não a quantidade adequada de sal. Ou, ainda, quando um comprador, após experimentar um pedaço de laranja numa banca de feira, decide se vai comprar ou não as laranjas. Essas são decis$\tilde{(o)}$es baseadas em procedimentos amostrais".(Bussab e Morettin, 2012, p.261)[3]

Inferir significa de forma concisa, deduzir uma coisa de outra. A inferência pode ser dedutiva, quando se parte de premissas para se chegar à conclusões, ou indutiva, quando se vai do específico para o geral.

Costa [7], apresenta vários tipos de inferência: indução simples, analogia, inferência estatística e método hipotético-dedutivo e inferência probabilística.

A indução simples se resume no fato de que se, dada uma amostra de uma população A, observar que todos os elementos de A também são de B, e se desconhece qualquer elemento de B que não seja de A, então toda a população de A compõe-se de B.

Na analogia, supõe-se que se elementos de uma amostra: $x_1, x_2, x_3, ...x_n$, possuem certas propriedades P e Q, se x_{n+1} possuir a propriedade P, x_{n+1}também vai possuir a propriedade Q.

A inferência estatística, se refere a estimação parâmetros, o teste de hipóteses e a teoria da decisão. A forma elementar de inferência estatística, é o chamado silogismo estatístico: se $y\%$ da população com propriedade A tem a propriedade B, se x_n tem A, logo x_n tem B.

O método hipotético-dedutivo, proposto pelo filósofo austríaco Karl Popper, sugere uma abordagem que busca a eliminação dos erros de uma hipótese. Faz isso a partir da ideia de testar a falsidade de uma proposição, ou seja, a partir de uma hipótese, estabelece-se que situação ou resultado experimental nega essa hipótese e tenta-se realizar experimentos para negá-la. Assim, a abordagem do método hipotético-dedutivo é a de buscar a verdade eliminando tudo o que é falso.

A inferência probabilística estabelece uma relação entre as premissas e a conclusão baseada em probabilidades. Se as premissas forem verdadeiras, existe uma determinada probabilidade da conclusão também ser verdadeira.

Assim, a inferência estatística consiste de procedimentos para fazer generalizações sobre as características de uma população a partir da informação contida na amostra.

4 Referencial Teórico

4.1 Estatística Descritiva

Para que possamos entender e nos aprofundar na inferência estatística, precisamos entender vários conceitos e notações, inclusive da estatística descritiva.

Como já foi citado, a estatística descritiva estuda diversas relações estatísticas que descrevem um conjunto de dados representado por uma amostra.

Definição 4.1.1. (Dados Brutos) *O conjunto de dados numéricos obtidos após a crítica dos valores coletados em uma pesquisa, é o conjunto dos dados brutos.*

Exemplo 4.1.1. *As estaturas de um grupo de crianças de uma escola:*

Tabela 1: Estaturas dos Alunos (cm)

41	35	41	30	91	48	45	65	66	77
53	80	51	69	52	64	81	54	55	66
62	35	65	42	94	78	68	60	89	77
71	39	73	74	55	88	47	57	59	60
97	85	84	73	65	74	76	85	66	67

Fonte: Dados Hipotéticos

Definição 4.1.2. (Rol) *Rol é o arranjo dos dados brutos em ordem crescente ou decrescente.*

Exemplo 4.1.2. *Para os dados do Exemplo 4.1.1, temos:*

Tabela 2: Estaturas dos Alunos

30	35	35	39	41	41	42	45	47	48
51	52	53	54	55	55	57	59	60	60
62	64	65	65	65	66	66	66	67	68
69	71	73	73	74	74	76	77	77	78
80	81	84	85	85	88	89	91	94	97

Fonte: Dados Hipotéticos

Definição 4.1.3. (Estatística de Ordem) *Sejam $x_1; x_2; \ldots; x_n$ valores de um conjunto de dados. Denotamos por $x_{(i)}$ o i-ésimo valor apresentado pelo rol destes dados quando este apresenta os dados em ordem crescente. Assim:*

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n-1)} \leq x_{(n)}$$

Exemplo 4.1.3. *Para os dados do Exemplo 4.1.2, temos: $x_{(1)} = 30$; $x_{(2)} = 35$, e assim por diante.*

Definição 4.1.4. (Amplitute Total) *Amplitute total ou Range, (R), é a diferença entre o maior valor ($x_{(n)}$) e menor valor ($x_{(1)}$) observado, ou seja:*

$$R = x_{(n)} - x_{(1)}.$$

Exemplo 4.1.4. *No Exemplo 4.1.2, $R = 97 - 30 = 67$.*

Definição 4.1.5. (Distribuição de frequências) *Distribuição de frequências é o arranjo em uma tabela, dos valores e suas respectivas frequências.*

Ao estudar conjuntos de dados, é conveniente resumi-los em uma tabela, com suas respectivas frequências. Denominamos frequência o número que fica relacionado a um determinado valor da variável. Quando os dados são discretos, com ou sem valores repetidos, a simples identificação dos mesmos com as respectivas frequências, pode ser um procedimento adequado ao que damos o nome de distribuição de frequênciaas sem intervalos de classe.

Exemplo 4.1.5. *Considere as idades de 10 alunos de uma turma de graduação:*

Tabela 3: Idades dos Alunos (anos)

19	19	20	22	22
22	25	26	32	33

Fonte: Dados Hipotéticos

Temos a seguinte distribuição de frequências:

Tabela 4: Idades dos alunos (anos)

x_i	F_i
19	2
20	1
22	3
25	1
26	1
32	1
33	1

Fonte: Dados Hipotéticos

Quando o conjunto de dados é muito grande, ou os dados são contínuos, iniciamos o processo de distribuição de frequências com intervalos de classes. Classes de frequência, ou simplesmente classes, são intervalos de variação da variável.

Exemplo 4.1.6. *Seja $x_{(i)}$ a estatura da i-ésima criança no grupo de 50 crianças. Temos:*

Tabela 5: Estaturas dos Alunos (cm)

30	35	35	39	41	41	42	45	47	48
51	52	53	54	55	55	57	59	60	60
62	64	65	65	65	66	66	66	67	68
69	71	73	73	74	74	76	77	77	78
80	81	84	85	85	88	89	91	94	97

Fonte: Dados Hipotéticos

Tabela 6: Idades

Classes	F_i
$30 \vdash 40$	4
$40 \vdash 50$	6
$50 \vdash 60$	8
$60 \vdash 70$	13
$70 \vdash 80$	9
$80 \vdash 90$	7
$90 \vdash 100$	3

Fonte: Dados Hipotéticos

Definição 4.1.6. (Número de Classes - k) *Para se determinar o número de classes, não há uma fórmula exata.*

> "Em geral, uma distribuição de frequências deve ter pelo menos 5 classes, mas não mais do que 15 classes, uma vez que o fato de ter uma quantidade demasiadamente pequena ou demasiadamente grande de classes proporciona poucas informações novas"(Levine, Stephan, Krehbiel e Berenson, 2012, p.26)[13]

Ainda, segundo Bussab e Morettin [3]:

> A escolha dos intervalos é arbitrária e a familiaridade do perquisador com os dados é que lhe indicará quantas e quais classes (intervalos) devem ser usadas. Entretanto, deve-se observar que, com um pequeno número de classes, perde-se a informação, e com um número grande de classes, o objetivo de resumir os dados fica prejudicado(...). Normalmente, sugere-se o uso de 5 a 15 classes com a mesma amplitude.(Bussab e Morettin, 2012, p.13)[3]

Certo de que o numero de intervalos deve ser um valor entre 5 e 15, existem duas fórmulas sugeridas para o cálculo do número de classes, segundo Crespo [6]:

$$k = \begin{cases} \sqrt{n} & \text{se } n \geq 25; \\ 5 & \text{se } x < 25. \end{cases}$$

onde n é o número de dados da amostra, ou ainda, a fórmula de Sturges:

$$k = 1 + 3,22 \cdot log(n).$$

Exemplo 4.1.7. *Seja $n = 50$, teremos:*

$$k = \sqrt{50} \approx 7,$$

ou ainda:

$$k = 1 + 3,22 \cdot log(50) \approx 7.$$

Definição 4.1.7. (Amplitude das classes - h) *É a diferença entre o limite superior e o limite inferior de cada classe. É obtida através da razão entre a amplitude total e o número de classes, aproximando para o maior inteiro.*

Exemplo 4.1.8. *Seguindo o Exemplo 4.1.7, onde $n = 50$ e $k = 7$, teremos:*

$$h \approx \frac{R}{h} = \frac{50}{7} \approx 7.$$

Os limites das classes podem ser expressos de três diferentes maneiras, onde os limites superiores e inferiores podem ou não pertencerem ao intervalo.

Exemplo 4.1.9. $30 \vdash 40$ *: valores de 30 a 40, excluindo o 40.*

Exemplo 4.1.10. $30 \dashv 40$ *: valores de 30 a 40, excluindo o 30.*

Exemplo 4.1.11. $30 \vdash\dashv 40$ *: valores de 30 a 40.*

Definição 4.1.8. (Pontos médios das classes) *O ponto médio da i-ésima classe, com $i = 1, 2, \ldots, k$, é dado por:*

$$m_i = \frac{l_I + l_S}{2}, onde:$$

l_I é o limite inferior da classe

l_S é o limite superior da classe

Exemplo 4.1.12. *Para o intervalo* $30 \vdash 40$*, teremos:*

$$m_i = \frac{30+40}{2} = \frac{70}{2} = 35.$$

Definição 4.1.9. (Frequência Absoluta) *Denotada por* F_i*, é o número de vezes que o elemento aparece na amostra, ou o número de elementos da i-ésima classe.*

Exemplo 4.1.13. *Para o Exemplo 4.1.6, temos* $F_7 = 3$.

Definição 4.1.10. (Frequência Absoluta acumulada - F_{ac}) *É o total das frequências de todos os valores inferiores ou iguais ao valor dado.*

Exemplo 4.1.14. *Para distribuição de frequências sem intervalos de classes, temos como exemplo a tabela abaixo:*

Tabela 7: Idades dos alunos (anos)

x_i	F_i	F_{ac}
19	2	2
20	1	3
22	3	6
25	1	7
26	1	8
32	1	9
33	1	10

Fonte: Dados Hipotéticos

Esta frequência permite expressar, de maneira mais rápida, e sem precisar calcular, a frequência acima ou abaixo de um limite. Na Tabela 7, por exemplo, percebe-se que são 7 alunos com idade inferior a 26 anos, sem precisar efetuar a soma, já que esta já está feita.

Exemplo 4.1.15. *Para distribuição de frequências com intervalos de classes, temos como exemplo a tabela abaixo:*

Tabela 8: Estaturas (cm)

Classes	F_i	F_{ac}
$30 \vdash 40$	4	4
$40 \vdash 50$	6	10
$50 \vdash 60$	8	18
$60 \vdash 70$	13	31
$70 \vdash 80$	9	40
$80 \vdash 90$	7	47
$90 \vdash 100$	3	50
Total	50	-

Fonte: Dados Hipotéticos

Definição 4.1.11. (Frequência Relativa - f_i) *É a razão entre a frequência simples e a frequência total, da i-ésima classe com $i = 1, 2, \ldots, k$. É dada por:*

$$f_i = \frac{F_i}{n}.$$

A importância das frequências relativas é o de permitir a análise ou facilitar as comparações entre os valores.

Exemplo 4.1.16. *Vejamos a tabela a seguir:*

Tabela 9: Estaturas (cm)

Classes	F_i	f_i
$30 \vdash 40$	4	0,08
$40 \vdash 50$	6	0,12
$50 \vdash 60$	8	0,16
$60 \vdash 70$	13	0,26
$70 \vdash 80$	9	0,18
$80 \vdash 90$	7	0,14
$90 \vdash 100$	3	0,06
Total	50	1

Fonte: Dados Hipotéticos

Note que $\sum f_i = 1$.

Definição 4.1.12. (Frequência Relativa Acumulada - f_{ac}) *É a razão entre a frequência acumulada da da i-ésima classe com $i = 1, 2, \ldots, k$ e a frequência total da distribuição. É dada por:*

$$f_{ac_i} = \frac{F_i}{\sum_{i=1}^{n} f_i}.$$

Exemplo 4.1.17. *Vejamos a tabela a seguir:*

Tabela 10: Estaturas (cm)

Classes	F_i	f_i	F_{ac}	f_{ac}
$30 \vdash 40$	4	0,08	4	0,08
$40 \vdash 50$	6	0,12	10	0,20
$50 \vdash 60$	8	0,16	18	0,36
$60 \vdash 70$	13	0,26	31	0,62
$70 \vdash 80$	9	0,18	40	0,80
$80 \vdash 90$	7	0,14	47	0,94
$90 \vdash 100$	3	0,06	50	1
Total	50	1	-	-

Fonte: Dados Hipotéticos

Definição 4.1.13. (Frequência Relativa Percentual - $100f_i$) *É a razão entre a frequência simples e a frequência total em porcentagem.*

Definição 4.1.14. (Frequência Relativa Acumulada Percentual - $100f_{ac}$) *É a razão entre a frequência acumulada da classe e a frequência total da distribuição em porcentagem.*

Conhecer os tipos de frequência, ajuda a responder várias questões com relativa facilidade, como por exemplo:

- Quatro alunos tem entre 30 cm, inclusive, e 40 cm, segundo a **frequência acumulada**.
- 62% dos alunos têm estaturas inferiores a 70 cm, segundo a **frequência relativa porcentual**.
- Quarenta alunos tem estatura abaixo de 80 cm, segundo a **frequência relativa acumulada porcentual.**

Tabela 11: Estaturas (cm)

Classes	F_i	f_i	$100f_i$	F_{ac}	f_{ac}	$100f_{ac}$
$30 \vdash 40$	4	0,08	8%	4	0,08	8%
$40 \vdash 50$	6	0,12	12%	10	0,20	20%
$50 \vdash 60$	8	0,16	16%	18	0,36	36%
$60 \vdash 70$	13	0,26	26%	31	0,62	62%
$70 \vdash 80$	9	0,18	18%	40	0,80	80%
$80 \vdash 90$	7	0,14	14%	47	0,94	94%
$90 \vdash 100$	3	0,06	6%	50	1	100%
Total	50	1	-	-	-	-

Fonte: Dados Hipotéticos

Outra maneira de apresentar uma distribuição de frequência, é por meio de gráficos.

Definição 4.1.15. (Histograma) *É um gráfico formado por retângulos justapostos, tendo como base o intervalo de classe, sendo a área de cada retângulo proporcional à frequência da classe correspondente. Segundo Levine at al [13], este gráfico foi idealizado pelo geneticista e estatístico Karl Pearson.*

Ainda, segundo Bussab e Morettin [3]:

> O histograma é um gráfico de barras contíguas, com bases proporcionais aos intervalos de classes e a área de cada retângulo proporcional à respectiva frequência. Pode-se usar tanto a frequência absoluta F_i como a relativa f_i. Indiquemos a amplitude do i-ésimo intervalo por Δ_i. Para que a área do retângulo respectivo seja proporcional a frequência, a sua altura deve ser proporcional a f_i/Δ_i (ou F_i/Δ_i), que é chamada densidade de frequência da i-ésima classe. Com essa convenção, a área total do histograma será igual a 1. (Bussab e Morettin, 2012, p.18)

Exemplo 4.1.18. *A distribuição das notas de trinta alunos de Estatística de uma escola está representado abaixo:*

Tabela 12: Notas de Estatística

Classe	F_i
$0 \vdash 2$	4
$2 \vdash 4$	5
$4 \vdash 6$	12
$6 \vdash 8$	8
$8 \vdash 10$	1
Total	30

Fonte: Dados Hipotéticos

Representando as classes de distribição no eixo das abcissas e as frequências no eixo das ordenadas, temos o seguinte Histograma:

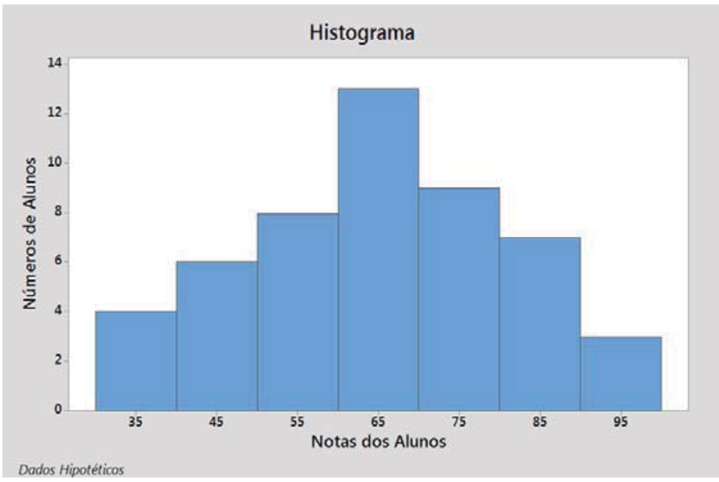


Figura 1: Histograma

É importante salientar a facilidade de leitura que o histograma traz em relação às frequências. É fácil notar o valor de maior frequência, ou o de menor frequência, sem precisar comparar todos os dados.

Definição 4.1.16. (Polígono de frequências) *Ao ligar os pontos médios da parte superior de cada retângulo do histograma, obtemos um polígono denominado* ***Polígono de Frequência****.*

Exemplo 4.1.19. *Retomando o histograma da Figura 1, temos:*

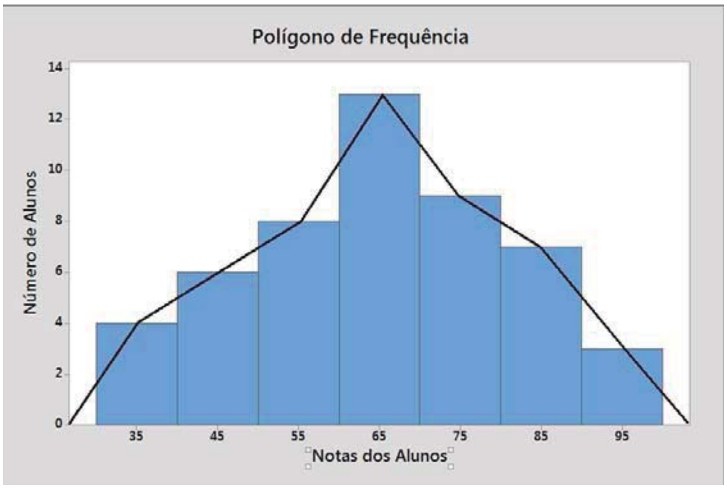


Figura 2: Polígono de Frequência

Definição 4.1.17. (Ogiva) *É um polígono de frequência acumulada, traçado marcando-se as frequências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.*

Exemplo 4.1.20. *Sejam os seguintes dados:*

Tabela 13: Estaturas dos Alunos (cm)

30	35	35	39	41	41	42	45	47	48
51	52	53	54	55	55	57	59	60	60
62	64	65	65	65	66	66	66	67	68
69	71	73	73	74	74	76	77	77	78
80	81	84	85	85	88	89	91	94	97

Fonte: Dados Hipotéticos

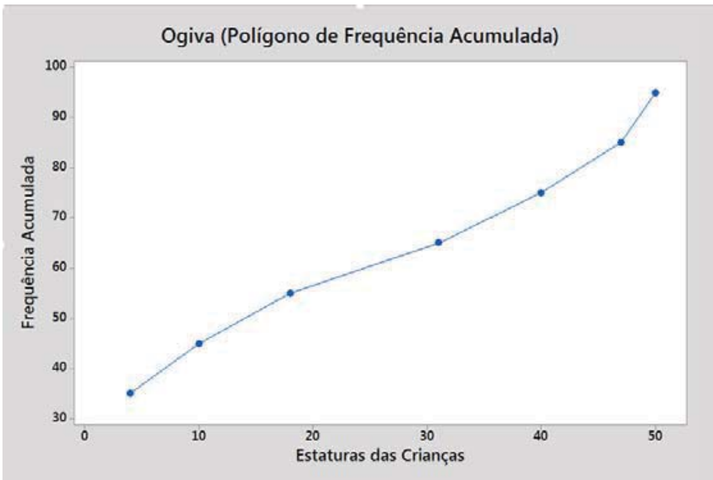


Figura 3: Ogiva

Definição 4.1.18. (Ramos e Folhas) *Uma disposição **Ramo e Folha** organiza dados em grupos, denominados **Ramos**, de tal modo que os valores (**Folhas**)em cada grupo (**Ramo**), se ramifiquem para a direita em cada uma das linhas. Esta disposição permite verificar como os dados estão distribuídos e onde há maior concentração de dados. Uma das vantagens, segundo Bussab e Morettin, é que não se perde, ou se perde pouca informação sobre os dados:*

> "Um procedimento alternativo para resumir um conjunto de valores, com o objetivo de se obter uma ideia da forma de sua distribuição, é o *ramo e folhas*. Uma vantagem deste diagrama sobre o histograma é que não perdemos (ou perdemos pouca) informação sobre os dados em si"(Bussab e Morettin, 2012, p.20)

Exemplo 4.1.21. *Sejam os dados abaixo, os gastos de 15 estudantes em uma lanchonete:*

Tabela 14: Gastos em R$

5,40	4,30	4,80	5,50	7,30
8,50	6,10	4,80	4,90	4,90
5,50	3,50	5,90	6,30	6,60

Fonte: Levine at al., p. 40 [13]

Para construir a disposição ramos e folhas para este exemplo, utilizamos as unidades como ramos e os decimais para folhas, obtendo a seguinte distribuição:

Tabela 15: Gastos em R$ - Ramo e Folhas

3	5
4	3 8 8 9 9
5	4 5 5 9
6	1 3 6
7	3
8	5

Fonte: Dados Hipotéticos

Da disposição ramo e folha acima, podemos retirar algumas conclusões, como:

- O menor gasto foi de R$3,50.
- O maior gasto foi de R$ 8,50.
- A maioria dos alunos gastaram em torno de R$4,00 e R$5,00.

Um resumo de dados, como já foi visto, pode ser representado por uma tabela de frequências ou por meio de gráficos, que mesmo de forma resumida, não deixa de fornecer informações sobre o comportamento de uma variável em questão. Embora tabelas e gráficos resumam a quantidade de dados, frequentemente, precisamos resumir ainda mais estes dados, apresentando um ou mais valores que representem todos os dados. Usualmente emprega-se as seguintes medidas, conhecidas como **Medidas de Posição Central**: Média, Moda e Mediana.

Segundo Ara et al [1], medidas de posição são valores que procuram indicar o centro da distribuição de frequências (média e mediana) e a região de maior concentração de frequências (moda).

Definição 4.1.19. (Média aritmética para variáveis discretas) *A média aritmética, é a soma das observações dividida pelo número delas, ou seja, se $x_1, x_2, ..., x_n$ são os n valores (distintos ou não) da variável X, a média aritmética, ou simplesmente média, de X pode ser escrita:*

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Agora, se tivermos n observações da variável X, das quais n_1 são iguais a x_1, n_2 são iguais a x_2,..., n_k são iguais a x_k, então a média pode ser escrita:

$$\overline{x} = \frac{n_1x_1 + n_3x_2 + ... + n_nx_n}{n} = \frac{1}{n}\sum_{i=1}^{n} n_ix_i.$$

Exemplo 4.1.22. *Sabendo que a produção leiteira diária de uma vaca **A**, durante uma semana, foi de 10, 14, 13, 15, 16, 18 e 12 litros, temos, para produção média da semana:*

$$\overline{x} = \frac{10 + 14 + 13 + 15 + 16 + 18 + 12}{7},$$

$$\overline{x} = \frac{98}{7},$$

$$\overline{x} = 14.$$

Exemplo 4.1.23. *Consideremos a distribuição relativa a 34 famílias de quatro filhos, tomando como variável o número de filhos do sexo masculino:*

Tabela 16: Número de filhos do sexo masculino

x_i	f_i	$x_i.f_i$
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
-	$\sum x_i$	$\sum x_i.f_i$

Fonte: Dados Hipotéticos

Neste caso, temos:

$$\overline{x} = \frac{0.2 + 1.6 + 2.10 + 3.12 + 4.16}{34} = \frac{78}{34} \cong 2,3$$

$$\overline{x} = \frac{78}{34}$$

$$\overline{x} \cong 2 \ \textit{filhos} \ .$$

Para a média aritmética de dados distribuídos em intervalos de classes, como em variáveis contínuas, por exemplo, convencionamos que todos os valores incluídos em um determinado intervalo de classe coincidem com seu ponto médio. Assim, determinamos a média aritmética ponderada por meio da fórmula:

$$\overline{x} = \frac{\sum m_i.F_i}{\sum F_i},$$

onde m_i é o ponto médio da classe.

Exemplo 4.1.24. *Seja a tabela utilizada no Exemplo 4.1.20:*

Tabela 17: Notas de Estatística

Classe	x_i	F_i	$x_i.F_i$
$0 \vdash 2$	1	4	4
$2 \vdash 4$	3	5	15
$4 \vdash 6$	5	12	60
$6 \vdash 8$	7	8	56
$8 \vdash 10$	9	1	9
Total	-	30	$\sum x_i.F_i = 144$

Fonte: Dados Hipotéticos

$$\overline{x} = \frac{144}{30} = 4,8.$$

Portanto, A nota média dos alunos de Estatística é $4,8$.

Definição 4.1.20. (Média Geométrica) *É a n-ésima raiz do produto de n valores.*

$$\overline{x}_g = (x_1 \times x_2 \times \ldots \times x_n)^{\frac{1}{n}},$$

$$\overline{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i},$$

ou ainda, se a cada variável x_k*, estiver associado uma frequência* F_k*, teremos:*

$$\overline{x}_g = \sqrt[n]{x_1^{F_1} \times x_2^{F_2} \times \ldots \times x_k^{F_k}},$$

onde $F_1 + F_2 + \ldots + F_k = n$.

Exemplo 4.1.25. *A média geométrica entre os valores 4, 45 e 150, será:*

$$\overline{x}_g = \sqrt[3]{4 \times 45 \times 150} = \sqrt[3]{27000} = 30.$$

Definição 4.1.21. (Média Harmônica) *A média harmônica dos n números positivos* $x_1, x_2, \ldots, x_n$ *é definida por:*

$$\overline{x}_h = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots + \dfrac{1}{x_n}}.$$

Exemplo 4.1.26. *A média harmônica entre os valores 5, 10 e 20, será:*

$$\overline{x}_h = \frac{3}{\dfrac{1}{5} + \dfrac{1}{10} + \dfrac{1}{20}} = \frac{3}{\dfrac{7}{20}} = \frac{60}{7} \approx 8,6.$$

Definição 4.1.22 (Moda)**.** *A moda, denotada por M_o, é definida como a realização mais frequente do conjunto de valores observados.*

Para encontrarmos a moda, de acordo com a definição, basta identificar o valor que mais se repete. Entretanto, podemos encontrar séries de valores nas quais não exista valor modal, ou seja, nas quais nenhum valor apareça mais vezes que os outros (**amodal**). Em outros casos, pode haver dois ou mais valores de concentração. Dizemos então, que a série tem dois ou mais valores modais (**bimodal** ou **plurimodal**).

Para dados agrupados sem intervalos de classe, determina-se imediatamente a moda: basta fixar o valor da variável de maior frequência.

Exemplo 4.1.27. *Na Tabela 16, temos que a moda será 3, pois é o número de filhos do sexo masculino com maior frequência, ou seja, $M_o = 3$.*

Para dados agrupados com intervalos de classes, determinamos primeiro a **classe modal**, que é a classe que apresenta a maior frequência. A moda neste caso, é o valor que mais aparece entre os limites da classe modal. Um método mais simples, é calcular o ponto médio da classe modal, que denominamos de **moda bruta**.

$$M_o = \frac{l_i + L_i}{2}, \text{onde:}$$

l_i é o limite inferior da classe modal
L_i é o limite superior da classe modal

Exemplo 4.1.28. *Seguindo os valores da Tabela 11, teremos que a classe modal é a 4ª classe ($60 \vdash 70$), pois há 13 alunos com estaturas entre 60 e 70 cm. A moda, segundo o rol dos dados na Tabela 5 é 65 e 66, ou seja, é uma série bimodal. A moda bruta é o ponto médio da classe modal, ou seja, $M_o = 65$.*

Definição 4.1.23. (Mediana) *A mediana é a realização que ocupa a posição central da série de observações, quando estão ordenadas em ordem crescente. Quando o número de observações for par, usa-se como mediana a média aritmética das duas observações centrais.*

Como foi visto na Definição 4.1.3 , as observações assim ordenadas, sem intervalos de classes, são chamadas estatísticas de ordem. Com esta notação , a mediana da variável X pode ser definida como:

$$md(x) = \begin{cases} x_{(\frac{n+1}{2})} \text{ se n é ímpar,} \\ \dfrac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, \text{ se n é par.} \end{cases}$$

Para distribuições com intervalos de classes, seguiremos os seguintes passos:

1º. Determinamos as frequências acumuladas;

2º. Calculamos $\dfrac{\sum F_i}{2}$;

3º. Pela frequência absoluta acumulada identificamos a classe que contém a mediana, ou seja, a **classe mediana**;

4º. Aplicamos a fórmula:

$$M_d = l_{md} + \frac{\left[\dfrac{n}{2} - \sum f_{ant}\right] \times h}{F_{md}}, \text{onde:}$$

l_{md} é o limite inferior da classe mediana;
$\sum f_{ant}$ é a frequência acumulada da classe anterior à classe mediana;
F_{md} é a frequência simples da classe mediana;
h é a amplitude do intervalo da classe mediana.

Exemplo 4.1.29. *Sejam as notas de 30 estudantes de uma turma da Tabela 4.1.18:*

Tabela 18: Notas de Estatística

i	Classe	F_i	F_{ac}
1	$0 \vdash 2$	4	4
2	$2 \vdash 4$	5	9
3	$4 \vdash 6$	12	21
4	$6 \vdash 8$	8	29
5	$8 \vdash 10$	1	30
-	Total	30	-

Fonte: Dados Hipotéticos

Como $\frac{n}{2} = 15$*, então a classe mediana é a de ordem 3. Então, temos:*
$l_{md} = 4$,
$\sum f_{ant} = 9$,
$F_{md} = 12$ *e*
$h = 2$.
Substituindo estes valores na fórmula, temos:

$$M_d = 4 + \frac{[15 - 9] \times 2}{12} = 4 + 1 = 5.$$

No caso de existir uma frequência acumulada exatamente igual a $\frac{n}{2}$, a mediana será o limite superior da classe correspondente.

Percebemos aqui que a mediana caracteriza uma série de dados devido a sua posição central. Mas também, apresenta uma segunda característica: ela separa a série em dois grupos com o mesmo número de valores. Assim, como as medianas, existem outras medidas, que não são medidas de tendência central, mas também **separam** uma série em partes iguais, se basenado em suas respectivas posições. Estas medidas são as **separatrizes**, que são além das medianas, são os **quartis**, os **percentis** e os **decis**.

Definição 4.1.24. (quartis) *Denominamos* ***quartis****, os valores de uma série que a dividem em quatro partes iguais. Há portanto três quartis:*

- *O primeiro quartil (Q_1) - valor situado de tal modo que 25% dos dados é menor que ele e os 75% restantes são maiores.*

- *O segundo quartil (Q_2) - valor situado de tal modo que 50% dos dados é menor que ele e os 50% restantes são maiores, ou seja, é a mediana (M_d).*

- *O terceiro quartil (Q_3) - valor situado de tal modo que 75% dos dados é menor que ele e os 25% restantes são maiores.*

Quando os dados são agrupados, para determinar os quartis usamos a mesma técnica do cálculo da mediana, bastanto substituir na fórmula da mediana, $\frac{n}{2}$ por $\frac{n}{4}$ para o primeiro quartil e por $\frac{3n}{4}$ para o terceiro quartil. Ou seja:

$$Q_1 = l_{md} + \frac{\left[\frac{n}{4} - \sum f_{ant}\right] \times h}{F_{md}},$$

$$Q_3 = l_{md} + \frac{\left[\frac{3n}{4} - \sum f_{ant}\right] \times h}{F_{md}}.$$

Definição 4.1.25. (percentis) *Denominamos **percentis** os noventa e nove valores que separam uma série em 100 partes iguais. Indicamos por $P_1, P_2, \ldots, P_{99}$. Nota-se que $P_{25} = Q_1$,$P_{50} = M_d$ e $P_{75} = Q_3$.*

O cálculo do percentil, segue a mesma técnica do cálculo da mediana, porém a fórmula $\frac{n}{2}$ será substituída por $\frac{k.n}{100}$ sendo k a ordem do percentil.

Definição 4.1.26. (decis) *Denominamos **decis** os nove valores que separam uma série em 10 partes iguais. Indicamos por $D_1, D_2, \ldots, D_9$. Nota-se que $D_5 = M_d$. O cálculo do decil, segue a mesma técnica do cálculo da mediana, porém a fórmula $\frac{n}{2}$ será substituída por $\frac{k.n}{10}$ sendo k a ordem do decil.*

É importante saber, que a representação de um conjunto de dados por uma única medida de posição central, pode esconder informações sobre a variabilidade do conjunto de observações. Considerando por exemplo, dois grupos de estudantes que se submeteram a um teste:

Sejam as notas tiradas em um teste por alunos de dois grupos:

Grupo 1 (variável X): 3, 4, 5, 6, 7

Grupo 2 (variável Y): 1, 3, 5, 7, 9.

Calculando a média dos dois grupos, temos $\overline{x} = 5$ e $\overline{y} = 5$ também. Ou seja, a identificação das duas séries por sua média, nada informa sobre suas diferentes variabilidades.

Temos então a necessidade de medidas que determinem a variabilidade de um conjunto de observações, e que nos permita comparar conjuntos de valores como nos exemplos acima, sob algum critério. A essas medidas damos o nome de **Medidas de Dispersão ou Variabilidade**

Definição 4.1.27. (Amplitude Total) *É a diferença entre o maior e o menor valor da série de dados.*

Exemplo 4.1.30. *Sejam os valores do Grupo 1: 3, 4, 5, 6, 7:*

$$R = 7 - 3 = 4.$$

A amplitude tem utilização limitada como medida de dispersão, pois seu valor só depende dos extremos, fazendo com que a variabilidade dos outros valores não afete seu resultado.

Definição 4.1.28. (Desvio)*É a diferença entre o valor do dado e a média do conjunto* $(x_i - \overline{x})$*. Mede a dispersão de cada dado em torno da média.*

Exemplo 4.1.31. *No Grupo 1 acima, temos que os desvios* $(x_i - \overline{x})$ *são -2, -1, 0, 1 e 2, e no Grupo 2 temos -4, -2, 0, 2 e 4.*

É fácil perceber que o somatório destes desvios é sempre igual a zero, ou seja, $\sum_{i=1}^{n}(x_i - \overline{x})$ não é uma boa medida de dispersão para os conjuntos dados.

Uma primeira opção para este problema, é considerar o Desvio Médio:

Definição 4.1.29. (Desvio Médio - $dm(X)$) *É a média dos módulos dos desvios.*

$$dm(X) = \frac{\sum_{i=1}^{k} |x_i - \overline{x}|}{n}.$$

Se os valores da variável ocorrem com certas frequências $F_1, F_2, \ldots, F_k$, então:

$$dm(X) = \frac{\sum_{i=1}^{k} |x_i - \overline{x}|.F_i}{n}.$$

Exemplo 4.1.32. *Para o Grupo 1 teríamos*

$$dm(X) = \frac{\sum_{i=1}^{5} |x_i - \overline{x}|}{5} = \frac{|-2| + |-1| + |0| + |1| + |2|}{5} = 1,2,$$

e para o Grupo 2,

$$dm(X) = \frac{\sum_{i=1}^{5} |x_i - \overline{x}|}{5} = \frac{|-4| + |-2| + |0| + |2| + |4|}{5} = 2,4.$$

Os dois valores, 1,2 e 2,4, nos mostram o quanto a média para cada um dos grupos é mais ou menos representativa. Quanto mais próximo de zero, menor a dispersão (ou variabilidade) dos valores da amostra em relação à média.

A segunda opção é considerar a variância:

Definição 4.1.30. (Variância) *É a média dos quadrados dos desvios.*

$$\sigma_x^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2}{n}.$$

Exemplo 4.1.33. *Para o Grupo 1 teríamos*

$$\sigma_x^2 = \frac{\sum_{i=1}^{5} (x_i - \overline{x})^2}{5} = \frac{4 + 1 + 0 + 1 + 4}{5} = 2,$$

e para o Grupo 2,

$$\sigma_x^2 = \frac{\sum_{i=1}^{5} (x_i - \overline{x})^2}{5} = \frac{16 + 4 + 0 + 4 + 16}{5} = 8.$$

Como a variância é uma medida com dimensão igual ao quadrado da dimensão dos dados, esta pode ocasionar problemas na interpretação. Assim, costuma-se utilizar a terceira medida de dispersão dentre as citadas anteriormente, que é o *desvio padrão*.

Definição 4.1.31. (Desvio Padrão) *É a raíz quadrada da variância:*

$$dp(X) = \sqrt{\sigma_x^2}.$$

Para o Grupo 1 teremos $dp(X) = \sqrt{2} \approx 1,4$ *e para o Grupo 2,* $dp(Y) = \sqrt{8} \approx 2,8$.

As medidas de dispersão indicam em média qual será o erro cometido ao tentar substitutir cada observação pela medida resumo do conjunto de dados, no exemplo, a média.

4.2 Probabilidade

Os princípios da probabilidade são importantes por fazerem uma ponte entre a estatística descritiva e a inferência estatística, que é aplicada às mais variadas áreas do conhecimento, e em todas as situações em que se deseja tirar conclusões sobre dados numéricos.

Definição 4.2.1. (Experimentos Aleatórios) *Um experimento aleatório é aquele cujo resultado não pode ser previsto.*

Exemplo 4.2.1. *São fenômenos aleatórios:*

- *O resultado no lançamento de uma moeda;*
- *O resultado no lançamento de um dado;*
- *O número de peças defeituosas em um lote de 100 peças produzidas diariamente por uma indústria;*
- *O tempo de duração de uma lâmpada elétrica.*

Definição 4.2.2. (Espaço Amostral - Ω) *Espaço amostral de um evento aleatório é o conjunto de todos os reultados possíveis desse experimento, representado por* Ω. *Cada elemento desse conjunto é chamado de ponto amostral ou evento elementar ou evento simples.*

Exemplo 4.2.2. *Os espaços amostrais associados aos experimentos aleatórios citados no exemplo 4.1.33 são:*

- $\Omega = \{c; k\}$, *onde* c = *cara e* k = *coroa;*
- $\Omega = \{1, 2, 3, 4, 5, 6\}$*;*
- $\Omega = \{1, 2, 3, 4, \ldots, 100\}$*;*
- $\Omega = \{x \in \mathbb{R} \,|\, x \geq 0\}$*.*

Definição 4.2.3. (Evento) *Seja* Ω *o espaço amostral de um evento aleatório. Evento é qualquer subconjunto de* Ω*.*

Dois casos relevantes de eventos são o próprio espaço amostral Ω, chamado de evento certo, e o conjunto vazio $\emptyset$, chamado evento impossível. Eventos que contém um único ponto amostral são chamados eventos simples e eventos que possuem **pelo menos** dois pontos amostrais são chamados eventos compostos ou combinados. Dizemos que um evento ocorre se acontecer pelo menos um de seus pontos amostrais.

Definição 4.2.4. (Evento Aleatório) *Evento ao qual atribui-se uma probabilidade.*

Exemplo 4.2.3. *Uma indústria de peças automotivas produz um determinado produto que pode ser classificado como bom (B) ou defeituoso(D). Da linha de produção são retirados três peças. O espaço amostral deste experimento é:*

$$\Omega = \{BBB, BBD, BDB, DBB, DDB, DBD, BDD, DDD\}$$

Se A *é o evento que consiste em obter duas peças defeituosas, então:*

$$A = \{DDB, DBD, BDD\}$$

Definição 4.2.5. (Classe de eventos aleatórios) *Um evento ao qual atribui-se uma probabilidade, é chamado de evento aleatório. Define-se a classe de eventos aleatórios para um experimento com espaço amostral* Ω*. Representa-se essa classe por* $\mathcal{F}$*. Exige-se dessa classe as seguintes propriedades:*

i) $\Omega \in \mathcal{F}$,

ii) Se $A \in \mathcal{F}$, *então* $A^c \in \mathcal{F}$, *iii) Se* $A_1, A_2, \ldots \in \mathcal{F}$, *então* $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

No caso discreto, podemos tomar $\mathcal{F} = \mathcal{P}(\Omega)$.

Seja Ω um espaço amostral finito. Se $|\Omega| =$ n, então $|\mathcal{F}| = 2^n$.

Exemplo 4.2.4. *No caso do lançamento de um dado onde* $\Omega = \{1, 2, 3, 4, 5, 6\}$, *temos:*

$$\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1; 2\}, \ldots, \Omega\}$$

$$|\mathcal{F}| = 2^6 = 64.$$

Como eventos são conjuntos, podemos aplicar aos mesmos as operações usuais de conjuntos, ou seja:

- O evento $A \cup B$ ocorre se pelo menos um dos eventos A ou B ocorrer:

$$A \cup B = \{x \in \Omega \,|\, x \in A \textit{ ou } x \in B\}$$

- O evento $A \cap B$ ocorre se ambos os eventos A e B ocorrerem:

$$A \cap B = \{x \in \Omega \,|\, x \in A \textit{ e } x \in B\}$$

- O evento A^C ocorre se A não ocorrer:

$$A^C = \{x \in \Omega \mid x \notin A\}$$

- O evento $A - B$ ocorre se A ocorrer e B não ocorrer:

$$A - B = \{x \in \Omega \,|\, x \in A \textit{ e } x \notin B\}$$

Podemos generalizar a união e a interscção para n eventos:

- O evento $A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i$ ocorre se pelo menos um dos $A_i s$ ocorrerem.
- O evento $A_1 \cap A_2 \cap \ldots \cap A_n = \bigcap_{i=1}^{n} A_i$ ocorre se todos os $A_i s$ ocorrerem.

Exemplo 4.2.5. *Seja o experimento lançamento de um dado, com* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *e os eventos* A *que consiste em observar face ímpar e evento* B *que consiste em observar face maior que 2, ou seja,* $A = \{1, 3, 5\}$ *e* $B = \{3, 4, 5, 6\}$*. Temos:*

$$A \cup B = \{1, 3, 4, 5, 6\}$$

$$A \cap B = \{3, 5\}$$

$$A^C = \{2, 4, 6\}$$

$$B^C = \{1, 2\}$$

$$A - B = \{1\}$$

$$A \cup A^C = \Omega$$

$$A \cap A^C = \emptyset$$

Definição 4.2.6. (Eventos Mutuamente Exclusivos) *Dois eventos* A *e* B *são mutuamente exclusivos se* $A \cap B = \emptyset$

Exemplo 4.2.6. *Seja o experimento lançamento de um dado, com* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *e os eventos* A *que consiste em observar face par e evento* B *que consiste em observar face ímpar, ou seja,* $A = \{2, 4, 6\}$ *e* $B = \{1, 3, 5\}$*. Teremos:*

$$A \cap B = \emptyset$$

Definição 4.2.7. (Partição de um Espaço Amostral) *Se tivermos n eventos* $A_1, A_2, \ldots, A_n$ *mutuamente exclusivos, isto é, se* $A_i \cap A_j = \emptyset$*, para todo* $i \neq j$ *e se* $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$*, temos que esses eventos formam uma partição de* Ω*.*

Exemplo 4.2.7. *Seja o experimento lançamento de um dado, com* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *e os eventos* A *que consiste em observar face par e evento* B *que consiste em observar face ímpar, ou seja,* $A = \{2, 4, 6\}$ *e* $B = \{1, 3, 5\}$*. Teremos:* $A \cap B = \emptyset$ *e* $A \cup B = \Omega$*, ou seja,* A *e* B *formam uma partição de* Ω

Segundo Meyer [17], para os eventos são válidas ainda as seguintes propriedades:

Propriedade 4.2.1. $A \cup B = B \cup A$

Propriedade 4.2.2. $A \cup (B \cup C) = (A \cup B) \cup C$

Propriedade 4.2.3. $A \cap B = B \cap A$

Propriedade 4.2.4. $A \cap (B \cap C) = (A \cap B) \cap C$

Propriedade 4.2.5. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Propriedade 4.2.6. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Propriedade 4.2.7. $A - B = A \cap B^C$

Propriedade 4.2.8. $A \cup \emptyset = A$

Propriedade 4.2.9. $A \cup \Omega = \Omega$

Propriedade 4.2.10. $(A \cup B)^C = A^C \cap B^C$

Propriedade 4.2.11. $(A \cap B)^C = A^C \cup B^C$

Para experimentos aleatórios, existe sempre uma incerteza sobre a ocorrência ou não de um determinado evento. Para medir esta *chance* ou *probabilidade* com a qual se pode esperar que um evento ocorra, é conveniente atribuir um número entre 0 e 1 (0% e 100%). Quando existe uma certeza de que o evento acontecerá, é dado que a probabilidade é 1 (100%). Quando existe uma certeza de que o evento não acontecerá, é dado que a probabilidade é 0 (0%)

Segundo Spiegel et al [18] existem dois procedimentos importantes por meio dos quais podemos estimar a probabilidade de um evento: a abordagem clássica e a abordagem frequencista.

- **Abordagem Clássica**: Se um evento pode ocorrer de h maneiras diferentes em um número total de n maneiras possíveis, todas elas igualmente prováveis, então a probabilidade do evento é $\frac{h}{n}$
- **Abordagem Frequencista**: Se após n repetições de um experimento, onde n é um número muito grande, é observado que um evento ocorre em n destas repetições, então a probabilidade do evento é próxima a $\frac{h}{n}$ (Probabilidade Empírica do Evento)

As duas abordagens têm algumas restrições. A primeira porque o termo "igualmente provável" é vago, ou seja, qual a garantia, por exemplo, que no lançamento de uma moeda, esta não seja viciada? A segunda, porque o "número grande" também é vago. Quantas vezes, por exemplo, eu teria que jogar uma moeda para verificar a probabilidade de 50% para cara? Devido a essas dificuldades, os matemáticos têm se voltado para uma abordagem axiomática de probabilidade:

Definição 4.2.8. (Definição Clássica de Probabilidade) *Dado um experimento aleatório, sendo Ω o seu espaço amostral, vamos admitir que todos os elementos de Ω tenham a mesma chance de acontecer, ou seja, que Ω é um conjunto eqüiprovável. Definimos probabilidade de um evento $A \subset \Omega$ ao número real $P(A)$, tal que:*

$$P(A) = \frac{|A|}{|\Omega|}$$

Exemplo 4.2.8. *No lançamento de um dado, seja A o evento "ocorrer face par". Tem-se:*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$|\Omega| = 6$$

$$A = \{2, 4, 6\}$$

$$|A| = 3$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

Definição 4.2.9. (Definição Axiomática de Probabilidade (Kolmogorov)) *Uma probabilidade é uma função $P : \mathcal{F} \longrightarrow [0; 1]$ que satisfaz:*

A1) $0 \leqslant P(A) \leqslant 1, \forall\ A \in \mathcal{F}$;

A2) $P(\Omega) = 1$;

A3) Aditividade enumerável: para qualquer sequência $A_1, A_2, \ldots \in \mathcal{F}$ de eventos dois a dois mutuamente exclusivos,isto é, $A_i \cap A_j = \emptyset$ se $i \neq j$, $P\left(\bigcup_{n=1}^{\infty}\right) = \sum_{i=1}^{\infty} P(A_i)$.

Definição 4.2.10 (Probabilidade Condicional)**.** *Sejam A e B dois eventos tais que $P(A) > 0$. Denota-se por $P(B|A)$ a probabilidade de B dado que ocorreu A. Como é conhecido que A ocorreu, este se torna um novo espaço amostral, substituindo o original espaço Ω . Daí :*

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

ou

$$P(A \cap B) = P(A) \cdot P(B|A)$$

Exemplo 4.2.9. *No lançamento de um dado, seja A o evento "sair face par" e B o evento "sair face 6". Tem-se:*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{2, 4, 6\}$$

$$B = \{6\}$$

A probabilidade de B ocorrer tendo ocorrido A, isto é, a probabilidade de ter saído face 6 no lançamento do dado, sabendo-se que saiu um número par, é dada por:

$$P(B|A) = \frac{1}{3}.$$

Ou seja, obtém-se o resultedo pelo quociente entre o número de elementos da interscção de A e B, e o número de elementos de A, ou aplicando a definição:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.$$

Teorema 4.2.1. *Para quaisquer três eventos A_1, A_2, A_3, tem-se:*

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$$

O resultado é facilmente generalizado para n eventos:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

$$\forall A_1, A_2, \ldots, A_n \in \mathcal{F}, n = 1, 2, 3, \ldots$$

Exemplo 4.2.10. *Uma urna contém 10 bolas idênticas, sendo 5 verdes, 3 brancas e 2 vermelhas. Extraindo-se 3 bolas, ao acaso e sem reposição, qual a probabilidade de que a primeira bola selecionada seja verde, a segunda branca e a terceira vermelha?*
Sejam:
A_1 *: a primeira bola selecionada é verde;*
A_2 *: a segunda bola selecionada é branca;*
A_3 *: a terceira bola selecionada é vermelha;*
tem-se:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) = \frac{5}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{24}.$$

Definição 4.2.11. (Eventos Independentes) *Se $P(B|A) = P(B)$, isto é, aprobabilidade de B ocorrer não é afetada pela ocorrência ou não de A, então diz-se que A e B são eventos independentes. Isto é:*

$$P(A \cap B) = P(A).P(B).$$

De forma recíproca, se $P(A \cap B) = P(A).P(B)$ vale, então A e B são independentes. Dizemos que três eventos A_1, A_2, A_3 são independentes se eles são independentes aos pares, ou seja: $P(A_i \cap A_k) = P(A_i).P(A_j)$ $j \neq k$,com $j, k = 1, 2, 3$ e:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

O teorema a seguir é bastante útil em situações em que o experimento pode ser repartido em etapas:

Teorema 4.2.2. (Teorema da Probabilidade Total) *Sejam $A_1, A_2, \ldots, A_n$ uma partição de Ω, com $P(A_i) > 0$, para todo i, isto é, $A_1, A_2, \ldots, A_n$ é uma família de eventos mutuamente exclusivos, cuja reunião completa todo o espaço amostral e dos quais, em cada experiência, u e somente um deles ocorre. Então, para cada evento B de Ω, tem-se:*

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i).$$

Demonstração.

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \ldots + P(B \cap A_n)$$

$$P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + \ldots + P(B|A_n) \cdot P(A_n)$$

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

□

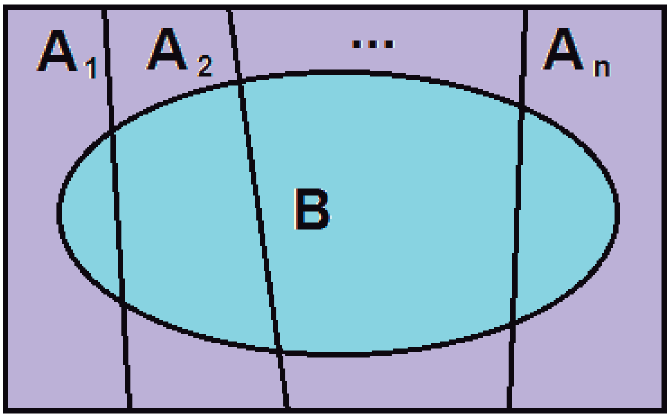


Figura 4: Partição de Ω

Em alguns casos, a probabilidade procurada não é *a posteriori*, ou seja, o que interessa não é $P(B|A)$, mas sim a probabilidade *a priori* $P(A|B)$. Sendo $P(B) > 0$, podemos inverter a probabilidade condicional, na seguinte forma:

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)}, i = 1, 2, 3, ..., n.$$

Desda forma, enuncia-se o Teorema de Bayes:

Teorema 4.2.3. (Teorema de Bayes)*: Sendo $A_1, A_2, \ldots, A_n$ partições de Ω, com $P(A_i) > 0$ para todo i, e B um evento de Ω com $P(B) > 0$, então:*

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)}, i = 1, 2, 3, ..., n.$$

Exemplo 4.2.11. *Uma indústria produz determinado tipo de peça em três máquinas M_1, M_2 e M_3. A máquina M_1 produz 40% das peças, enquanto M_2 e M_3 produz 30% cada uma. As porcentagens de peças defeituosas produzidas por essas máquinas são respectivamente iguais a 1%, 4% e 3%. Se uma peça é selecionada aleatoriamente da produção total, a probabilidade dessa peça ser defeituosa será 0,025. De fato, sendo D o evento "peça defeituosa", tem-se:*

$P(M_1) = 0,40 \qquad P(D|M_1) = 0,01$
$P(M_2) = 0,30 \qquad P(D|M_2) = 0,04$
$P(M_3) = 0,30 \qquad P(D|M_3) = 0,03$

Aplicando o Teorema 4.2.2, tem-se:

$$P(D) = P(D|M_1) \cdot P(M_1) + P(D|M_2) \cdot P(M_2) + P(D|M_3) \cdot P(M_3),$$

$$P(D) = 0,01 \cdot 0,40 + 0,04 \cdot 0,30 + 0,03 \cdot 0,30),$$

$$P(D) = 0,025.$$

*Agora, se uma peça escolhida ao acaso da **produção total** é defeituosa, a probabilidade dessa peça ter sido produzida pela máquina M_1 é 0,16. Aplicando o Teorema 4.2.3, tem-se:*

$$P(M_1|D) = \frac{P(D|M_1) \cdot P(M_1)}{P(D|M_1) \cdot P(M_1) + P(D|M_2) \cdot P(M_2) + P(D|M_3) \cdot P(M_3)}$$

$$P(M_1|D) = \frac{0,01 \cdot 0,40}{0,025} = 0,16$$

4.3 Análise Combinatória

Em muitos experimentos, o número de eventos do espaço amostral em questão não é muito grande, o que permite a contagem direta dos pontos necessários para o cálculo de probabilidades, porém, em certos experimentos, a contagem direta se torna impraticável. Nestes casos, aprende-se desde o ensino fundamental, algumas noções de Análise Combinatória, também chamado de Método Sofisticado de Contagem, por alguns autores como Spiegel [17].

Definição 4.3.1. (Princípio Fundamental da Contagem(PFC) ou Princípio Multiplicativo*: Suponha uma sequência formada por i elementos* $(a_1; a_2; a_3; \ldots; a_i)$*, em que:*

- a_1 *pode ser escolhido de* n_1 *maneiras distintas ;*
- a_2 *pode ser escolhido de* n_2 *maneiras distintas;*
- a_3 *pode ser escolhido de* n_3 *maneiras distintas;*

  $\vdots$

- a_i *pode ser escolhido de* n_i *maneiras distintas;*
- *e que o número de escolhas numa etapa não pode ser influenciada por escolhas anteriores.*

Então, o número de possibilidades para construir a sequência $(a_1; a_2; a_3; \ldots; a_i)$ *é:* $n_1 \cdot n_2 \cdot n_3 \cdot \ldots \cdot n_i$.

Segundo Iezzi et al [10], este resultado serve de base para vários problemas de contagem.

Exemplo 4.3.1. *Cinco estradas ligam as cidades X e Y e três estradas ligam as cidades Y e Z. Existem quinze maneiras diferentes de ir da cidade X à cidade Z, passando pela cidade Y, pois pelo PFC,* $5 \cdot 3 = 15$.

Na resolução de problemas de contagem utilizando o PFC, frequentemente aparecem multiplicações envolvendo números naturais consecutivos, como por exemplo, $5{\cdot}4{\cdot}3{\cdot}2{\cdot}1$. Nestes casos é possível representar estas multiplicações de uma forma mais resumida ,através do Fatorial de um número natural, definição apresentada a seguir:

Definição 4.3.2. (Fatorial) *Dado um número natural* n*, define-se o fatorial de* n*, denotando-se por* $n!$*, a relação* $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$ *para* $n \geq 2$*. Caso* $n = 1$*, tem-se* $1! = 1$ *e se* $n = 0$*, tem-se* $0! = 1$.

Exemplo 4.3.2. *O fatorial do número natural 6 será* $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$.

Em alguns casos, ainda segundo Iezzi et al [10], n sendo muito grande, o cálculo de $n!$ se torna muito cansativo. Torna-se necessário então, algumas simplificações a partir da relação $n! = n \cdot (n-1)!; n \in \mathbb{N}$ e $n \geq 2$.

Exemplo 4.3.3. $\dfrac{9!}{4!} = \dfrac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4!}{4!} = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 15120.$

Proposição 4.3.1. (Princípio Aditivo das Partes Disjuntas) *Se* $A_1, A_2, \ldots, A_i$ *são conjuntos 2 a 2 disjuntos, então:*

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i|.$$

Exemplo 4.3.4. *Em uma classe existem 28 rapazes e 12 moças. Podemos selecionar 1 estudante de 40 modos diferentes.*

O princípio fundamental da contagem (PFC) é a técnica mais utilizada nos problemas de contagem, porém, em alguns casos, este, sozinho, não se torna suficiente. No ensino médio, estuda-se três diferentes maneiras de agrupamentos simples, ou seja, agrupamentos de k elementos distintos escolhidos em conjuntos de n elementos: permutações, arranjos e combinações.

Proposição 4.3.2. (Permutação) *Dado um conjunto com n elementos distintos, chama-se permutação desses n elementos todo agrupamento ordenado (sequência) formado por n elementos. Denota-se por* P_n*, que calcula-se através da fórmula* $P_n = n!$.

Demonstração. Sejam n elementos distintos e P_n o número de permutações possíveis desses n elementos. Para encontrar o número de sequências formadas por esses n elementos, procede-se da seguinte forma:

- Escolhe o primeiro termo da sequência dentre os n elementos, ou seja, tem-se n possibilidades de escolha para o primeiro termo da sequência.
- Escolhe o segundo termo da sequência dentre os $n-1$ elementos que sobraram, ou seja, tem-se $n-1$ possibilidades de escolha para o segundo termo da sequência, já que não se pode escolher o primeiro novamente.
- Definidos os dois primeiros termos da sequência, escolhe o terceiro termo da sequência dentre os $n-2$ elementos que sobraram, ou seja, tem-se n_2 possibilidades de escolha para o segundo termo da sequência, já que não se pode escolher o primeiro e o segundo novamente.
- $\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$
- Definidos os $n-1$ primeiros termos da sequência, o termo que ocupa a última posição da sequência, só pode ser o único que sobrou, ou seja, fica determinado de maneira única, assim, uma possibilidade de escolha.

Assim, pelo PFC, tem-se: $P_n = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$ □

Exemplo 4.3.5. *Podemos escrever 24 números com 4 algarismos distintos escolhendo os algarismos dentre 3, 5, 7 e 9, pois* $P_4 = 4! = 24$.

Proposição 4.3.3. (Arranjo) *Dado um conjunto com n elementos distintos, chama-se arranjo desses n elementos, tomados k a k (com $k \leq n$) qualquer agrupamento ordenado de k elementos distintos escolhidos entre os n existentes. Denota-se por $A_{n;k}$, e calcula-se através da fórmula* $A_{n;k} = \dfrac{n!}{(n-k)!}$. .

Demonstração. Sejam n elementos distintos e $A_{n;k}$ o número de arranjos desses elementos tomados k a k. Usando novamente o PFC, tem-se:

- Escolhe o primeiro termo da sequência dentre os n elementos, ou seja, tem-se n possibilidades de escolha para o primeiro termo da sequência.

- Escolhe o segundo termo da sequência dentre os $n-1$ elementos que sobraram, ou seja, tem-se $n-1$ possibilidades de escolha para o segundo termo da sequência, já que não se pode escolher o primeiro novamente.

- Definidos os dois primeiros termos da sequência, escolhe o terceiro termo da sequência dentre os $n-2$ elementos que sobraram, ou seja, tem-se n_2 possibilidades de escolha para o segundo termo da sequência, já que não se pode escolher o primeiro e o segundo novamente.

- $\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$

- Para escolher o K-ésimo termo, a partir das $k-1$ escolhas anteriores, tem-se que sobraram $n-(k-1)$ elementos, ou seja, $n-k+1$ elementos.

Assim, pelo P.F.C., tem-se:

$$A_{n;k} = n \cdot (n-1) \cdot (n-2) \cdot \ldots (n-k+1) \tag{1}$$

Mas, multiplicando (1) por $(n-k)! = (n-k) \cdot (n-k+1) \cdot \ldots \cdot 2 \cdot 1$ tem-se:

$$A_{n;k} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots (n-k+1) \cdot (n-k) \cdot (n-k+1) \cdot \ldots \cdot 2 \cdot 1}{(n-k) \cdot (n-k+1) \cdot \ldots \cdot 2 \cdot 1}$$

$$A_{n;k} = \frac{n!}{(n-k)!}. \tag{2}$$

□

Problemas que envolvem contagem de arranjos podem ser resolvidos pela fórmula (2).

Exemplo 4.3.6. *Podemos escrever 72 números de dois algarismos distintos escolhidos dentre 1, 2, 3, 4, 5, 6, 7, 8 e 9, pois:*

$$A_{9;2} = \frac{9!}{(9-2)!} = \frac{9!}{7!} = \frac{9 \cdot 8 \cdot 7!}{7!} = 9 \cdot 8 = 72.$$

Proposição 4.3.4. (Combinação)*Dado um conjunto com n elementos distintos, chama-se combinação dos n elementos, tomados k a k (com $k \leq n$) qualquer subconjunto de k elementos distintos escolhidos entre os n existentes. Denota-se por $C_{n;k}$ ou $\binom{n}{k}$, e calcula-se através da fórmula $C_{n;k} = \dfrac{n!}{(n-k)! \cdot k!}$. .*

Demonstração. Sejam n elementos distintos de um conjunto e $C_{n;k}$ o número de combinações possíveis desses n elementos.

- Com o PFC, conta-se o número de arranjos formados por k elementos distintos, escolhidos dentre os n elementos:

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot [n-(k-1)] = A_{n;k}.$$

- Com o PFC, conta-se o número de sequências distintas que podem ser formadas com os k elementos escolhidos:

$$k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 2 \cdot 1 = P_k = k!$$

- Como toda e qualquer permutação dos elementos de uma sequência, dá origem a uma única combinação, o número de combinações dos n elementos tomados k a k é:

$$C_{n;k} = \frac{A_{n;k}}{P_k} = \frac{A_{n;k}}{k!}$$

Aplicando a fórmula para arranjo, tem-se:

$$C_{n;k} = \frac{\dfrac{n!}{(n-k)!}}{P_k} = \frac{n!}{(n-k)! \cdot k!}$$

□

Exemplo 4.3.7. *O número de maneiras nas quais 4 cartas podem ser escolhidas de um total de 10 cartas diferentes é:*

$$C_{10;4} = \frac{10!}{(10-4)4!} = \frac{10!}{6! \cdot 4!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = 210.$$

Muito útil para simplificar os cálculos vale resaltar a importância da seguinte propriedade:

Propriedade 4.3.1. (Igualdade de combinações Complementares) $C_{n;k} = C_{n;n-k}$.

Demonstração.

$$C_{n;k} = \frac{n!}{k! \cdot (n-k)!}$$

$$C_{n;k} = \frac{n!}{(n-k)! \cdot k!}$$

$$C_{n;k} = \frac{n!}{(n-k)! \cdot [n-(n-k)]!}$$
$$C_{n;k} = C_{n;n-k}$$

□

Exemplo 4.3.8.

$$C_{6;4} = C_{6;2}$$
$$\frac{6!}{4! \cdot 2!} = \frac{6!}{2! \cdot 4!}$$
$$15 = 15.$$

Segundo Dante [8], existem também, alguns agrupamentos que não são simples, ou seja, agrupamentos formados por elementos repetidos.

Proposição 4.3.5. (Permutações com Repetição) *Dado um conjunto com n elementos dos quais x_1 são de um tipo, x_2 de outro, x_3 de outro, e assim por diante até x_k de outro tipo com $x_1+x_2+x_3+\ldots+x_k = n$, chama-se permutação com repeticão desses n elementos todo agrupamento ordenado (sequência), formado por estes n elementos. Denota-se por $P_n^{x_1;x_2;\ldots x_k}$, que calcula-se através da fórmula $P_n^{x_1;x_2;\ldots;x_k} = \dfrac{n!}{x_1! \cdot x_2! \cdot \ldots \cdot x_k!}$.*

Justificativa: Se temos x_1 elementos iguais, as permutações entre estes não produzem um novo anagrama. Se temos x_2 elementos iguais, as permutações entre estes também não produzem um novo anagrama, e assim por diante até x_k elementos iguais. Por isso divide-se $n!$ por $x_1! \cdot x_2! \cdot \ldots \cdot x_k!$

Exemplo 4.3.9. *Sabe-se que anagramas são diferentes disposições das letras de uma palavra. A quantidade de anagramas da palavra BATATA é 60, pois $P_6^{3;2} = \dfrac{6!}{3! \cdot 2!}$.*

Justificativa: As permutações entre os três "As"não produzem um novo anagrama, assim como as dos dois "Ts"também não produzem. Se fossem todas as seis letras diferentes, teríamos 6! permutações, mas devido aos elementos iguais, dividimos 6! por 3! e 2!.

Proposição 4.3.6. (Permutações circulares) *O número de maneiras de colocar n elementos em círculo, de maneira que disposições adquiridas por rotações sejam consideradas iguais são permutações circulares. Segundo Lima et al [14], denota-se por $(PC)_n$, que calcula-se através da fórmula $(PC)_n = \dfrac{n!}{n} = (n-1)!$.*

Exemplo 4.3.10. *Seis crianças podem formar uma roda de ciranda de 120 maneiras diferentes, pois $(PC)_6 = \dfrac{6!}{6} = 5! = 120$.*

4.4 Distribuições de Probabilidades

O conceito de variável aleatória, que será definido a seguir, permitirá uma associação dos resultados de experimentos aleatórios a números reais, para que assim, utilizando o conceito de função, poder com mais praticidade calcular probabilidades dos vários eventos correspondentes a um experimento ou fenômeno aleatório. Segundo Lebenstayn e Coletti [12]:

Definição 4.4.1. (Variável Aleatória) *Uma função* $X : \Omega \longrightarrow \mathbb{R}$ *é variável aleatória se* $\{w \in \Omega, X(w) \leq x\} \in \mathcal{F}, \forall$ *x real.*

Ou seja, variável aleatória é uma variável cujo valor depende do resultado de um experimento aleatório, ou seja, uma função definida em Ω, que assume valores reais. As variáveis aleatórias são representadas por letras maiúsculas para se evitar confusão com símbolos usuais da álgebra.

Exemplo 4.4.1. *Seja o experimento "lançamento simultâneo de duas moedas". Considerando c para cara e k para coroa, tem-se* $\Omega = \{(c, c); (c, k), (k, c), (k, k)\}$, *e como variável aleatória, por exemplo,* ***X*** *representando o "número de caras". Daí, a cada ponto amostral, associamos um número real de acordo com a tabela a seguir:*

Tabela 19: X: Número de caras

Ponto Amostral	**X**
(c, c)	2
(c, k)	1
(k, c)	1
(k, k)	0

Uma variável aleatória é definida como uma variável que produz respostas numéricas, tais como o número de pontos em um jogo, ou a estatura de uma pessoa. Variáveis numéricas são classificadas como discretas ou contínuas. Variáveis contínuas produzem resultados produzidos em medições, como por exemplo as estaturas, enquanto as variáveis discretas produzem resultados de processos de contagem, como o exemplo dos pontos feitos em um jogo.

Definição 4.4.2. (Variável Aleatória Discreta - Distribuição de Probabilidades) *Uma distribuição de probabilidades para uma variável aleatória discreta é uma lista mutuamente excludente de todos os resultados numéricos possíveis, juntamente com a probabilidade de ocorrência de cada um dos resultados.*

Indicando as probabilidades de ocorrência de cada um dos valores da variável aleatória X por $p(x_i) = P(X = x_i)$, devem ser feitas as seguintes condições:

- $p(x_i) \geqslant 0$ para todo i,
- $\sum_i p(x_i) = 1$.

Exemplo 4.4.2. *Voltando ao exemplo 4.4.1 das moedas, teremos:*

Tabela 20: X (Número de caras)

Número de Caras(X)	$P(X = x_i)$
0	$P(X = 0) = 1/4$
1	$P(X = 1) = 2/4$
2	$P(X = 2) = 1/4$

Considerando-se por exemplo, um experimento que consiste em selecionar, ao acaso, uma peça de uma linha de produção, e determinar o valor do comprimento da peça em milímetros. Nesse caso, dentro de um determinado grau de precisão decorrente da limitação do instrumento de medição, a variável X : "comprimento da peça" pode assumir qualquer valor dentro de um intervalo em $\mathbb{R}$, sendo portanto, uma variável contínua.

Definição 4.4.3. (Variável Aleatória Contínua - Função Densidade de Probabilidade) *Como uma variável aleatória contínua pode assumir uma infinidade de valores em um intervalo real, a cada um dos infinitos valores da reta real é atribuída probabilidade nula, podendo-se apenas calcular probabilidades para valores em intervalos da reta real. Assim, as probabilidades de ocorrências de cada um dos possíveis resultados do experimento aleatório são determinados por uma função contínua $f(x)$, denominada função densidade de probabilidade, e que satisfaz as seguintes propriedades:*

i) $f(x) \geqslant 0; \forall x \in \mathbb{R}$,

ii) $P(a \leqslant X \leqslant b) = \displaystyle\int_a^b f(x)d(x)$,

iii) $\displaystyle\int_{-\infty}^{\infty} f(x)d(x) = 1$.

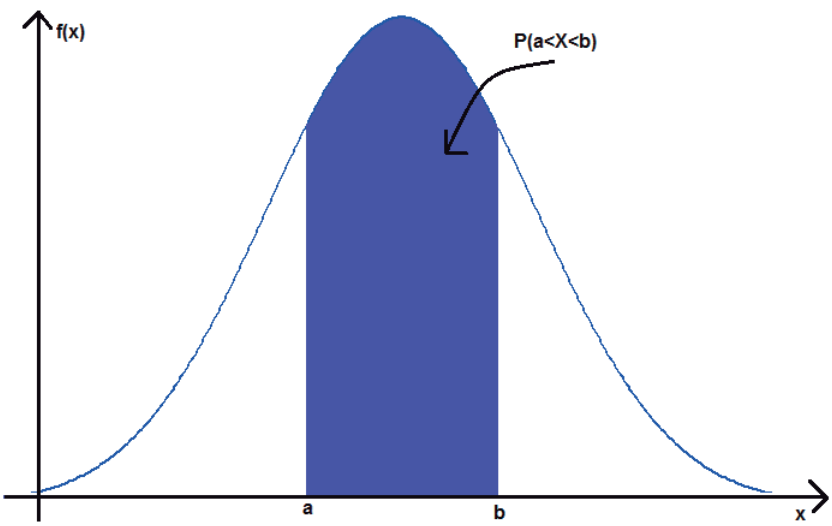


Figura 5: Função Densidade de Probabilidade

Exemplo 4.4.3. *Uma variável aleatória contínua X é dada pela função densidade de probabilidade $f(x) = 3x^2, 0 \leq x \leq 1$. A probabilidade $P(0 \leq X \leq 1/2) = \frac{1}{8}$.*

Demonstração.

$$P(0 \leq X \leq 1/2) = \int_0^{\frac{1}{2}} f(x)d(x) = \int_0^{\frac{1}{2}} 3x^2 d(x) = x^3 \Big|_0^{\frac{1}{2}} = \frac{1}{8}.$$

□

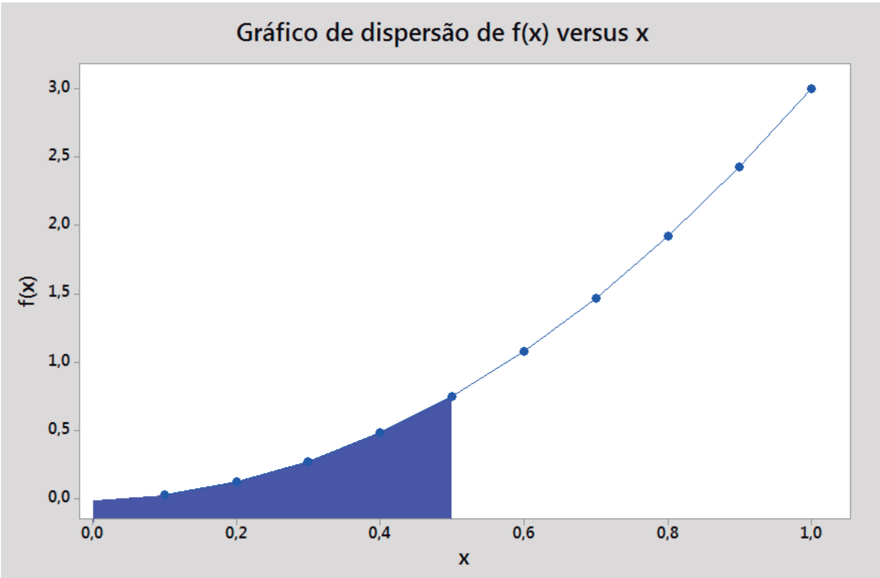


Figura 6: Função Densidade de Probabilidade (II)

Segundo Downing e Clark [9], intuitivamente as variáveis aleatórioas discretas são mais fáceis de entender. Todavia, matematicamente, as variáveis contínuas são de mais fácil manejo. Se uma distribuição discreta tem muitos valores possíveis muito próximos uns dos outros, pode em geral, ser aproximada por uma distribuição contínua.

Definição 4.4.4 (Esperança ou Média ou Valor esperado de uma variável aleatória - μ)**.** *A média aritmética μ de uma distribuição de probabilidades é o **valor esperado** de sua respectiva variável aleatória. É o principal parâmetro de posição e fornece informação sobre o centro de sua distribuição de probabilidades.*

Para calcular o valor esperado, multiplica-se cada um dos resultados possíveis x_i pela sua probabilidade correspondente $p(x_i)$, e depois adicionam-se esses produtos.

$$\mu = E(X) = \sum_{i}^{n} x_i \cdot p(x_i),$$

onde:

$$X_i = \text{i-ésimo resultado para a variável aleatória discreta } X,$$

$$P(X_i) = \text{probabilidade de ocorrência do i-ésimo resultado de } X.$$

Para uma variável aleatória contínua, calculamos:

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot d(x).$$

Exemplo 4.4.4. *Um jogo consiste em retirar, ao acaso, uma bola de uma caixa contendo 5 bolas verdes, 3 bolas brancas e 2 bolas vermelhas. Se a bola selecionada for branca ganham-se R\$ 10,00 e se for preta ou vermelha perdem-se, respectivamente, R\$ 5,00 e R\$ 15,00. O lucro médio do jogo será de* $R\$0,50$.

Demonstração. Sendo X a variável aleatória lucro, temos a seguinte distribuição de probabilidades:

Tabela 21: X (Distribuição de Probabilidade de X)

Resultado	Probabilidade $p(x_i)$	Valor x_i
Bola verde	$1/2$	10
Bola branca	$3/10$	-5
Bola vermelha	$1/5$	-15

Então:

$$\mu = E(X) = \sum_i x_i \cdot p(x_i) = 10 \cdot \frac{1}{2} + (-5) \cdot \frac{3}{10} + (-15) \cdot \frac{1}{5} = 0,5.$$

Logo, o lucro médio será de R$ 0,50. □

Definição 4.4.5. (Variância de uma variável aleatória) *A variância de uma variável aleatória é o principal parâmetro de dispersão e mede a variabilidade dos valores em relação à sua média. A variância de uma variável aleatória, indicada por σ^2 ou $Var(X)$, é definida por:*

$$\sigma^2 = Var(X) = E(X-\mu)^2, \textit{onde: } \mu = E(X), \textit{que é a média de} X.$$

Proposição 4.4.1. $Var(X) = E(X^2) - [E(X)^2]$,

No caso de variável discreta, é dada por:

$$Var(X) = \sum_i x_i^2 \cdot P(X = x_i) - \mu^2,$$

e no caso da variável contínua, por:

$$Var(X) = \int_{-\infty}^{\infty} x^2 \cdot f(x) \cdot d(x) - \mu^2.$$

Exemplo 4.4.5. *Um jogo consiste em lançar uma moeda. Se sair cara ganham-se R\$10,00 e se sair coroa perdem-se R\$5,00. O lucro médio do jogo será de* R$\$2,50$ *e a sua variância de* $56,25$.

Sendo X a variável lucro, temos:

Tabela 22: Distribuição de Probabilidade de X

Resultado	Probabilidade $p(x_i)$	Valor x_i	x_i^2
Cara	$1/2$	10	100
Coroa	$1/2$	-5	25

Demonstração. A média é calculada por:

$$\mu = E(X) = \sum_i x_i \cdot p(x_i) = 10 \cdot \frac{1}{2} + (-5) \cdot \frac{1}{2} = 2,5.$$

Logo, o ganho médio do jogo é de $R\$2,50$. A variância é calculada por:

$$Var(X) = \sum_i x_i^2 \cdot p(x_i) - \mu^2 = 100 \cdot \frac{1}{2} + 25 \cdot \frac{1}{2} - (2,5)^2 = 62,5 - 6,25 = 56,25.$$

□

Como a variância é um número em unidade quadrada em relação à variável em questão, o que sob o ponto de vista prático, é um inconveniente, definimos o *desvio padrão*, indicado por σ ou $DP(X)$, como sendo a raíz quadrada da variância. O desvio padrão tem a vantagem de ser expresso na mesma unidade da variável.

No Exemplo 4.4.5, o desvio padrão é dado por:

$$\sigma = DP(X) = \sqrt{Var(X)} = \sqrt{56,25} = 7,5.$$

Logo, no jogo das moedas, o desvio padrão do lucro é R$7,50.

Supondo um experimento como lançar um dado, temos que cada lançamento é chamado um ensaio. Em cada ensaio, haverá uma probabilidade associada a um evento particular, como por exemplo, 3 no caso do dado. Em alguns casos, esta probabilidade não muda de um ensaio para o próximo. Tais ensaios são considerados independentes, e definidos como ensaios de Bernoulli, após Jacques Bernoulli te-lôs estudados no século $XVII$.

Definição 4.4.6. (Ensaio de Bernoulli) *Ensaio de Bernoulli é um experimento cujo resultado apresenta ou não determinada característica. Ao apresentar tal característica, usa-se o termo sucesso, cuja probabilidade é p, e ao não apresentar tal carcterística, usa-se o termo fracasso, cuja probabilidade é $q = 1 - p$.*

Exemplo 4.4.6. *Um dado é lançado. Deseja-se tirar face 3. Sair face 6 é sucesso e não sair face 3 é fracasso. Neste caso, temos $p = \frac{1}{6}$ e $q = \frac{5}{6}$.*

Definição 4.4.7. (Variável Aleatória de Bernoulli) *A variável aleatória X que assume apenas os valores 0 ou 1, com função de probabilidade $p(0) = P(X = 0) = p$ ou $p(1) = P(X = 1) = q = 1-p$ é chamada variável aleatória de Bernoulli com parâmetro p. A notação é $X \sim$ Bernoulli (p).*

Exemplo 4.4.7. *Em um lançamento de um dado, vamos considerar a variável X como o número de face 3 que pode sair. Temos duas possibilidades, que são $X = 0$ ou $X = 1$. Temos:*

$$P(X = 0) = \frac{5}{6},$$
$$P(X = 1) = \frac{1}{6}.$$

Definição 4.4.8. (Distribuição Binomial) *Seja p a probabilidade de um evento acontecer em um ensaio de Bernoulli qualquer (chamada probabilidade de sucesso). Então $q = 1 - p$ é a probabilidade do evento não acontecer em um ensaio qualquer (chamada probabilidade de fracasso). A probabilidade do evento vir a acontecer k vezes em n repetições do ensaio, isto é, ocorrer k sucessos e $n - k$ fracassos é dada pela função de probabilidade:*

$$P(X = k) = \begin{pmatrix} n \\ \\ k \end{pmatrix} p^k (1-p)^{n-k},$$

onde a variável X denota o número de sucessos em n ensaios e $k = 0, 1, \ldots, n$. A notação é $X \sim$ Binomial (n, p).

Exemplo 4.4.8. *Dois times de futebol, G e V, jogam entre si 6 vezes, e a chance de um time ganhar uma vez é $\frac{1}{3}$. A probabilidade do time G ganhar 4 jogos será:*

$$P(X = 4) = \begin{pmatrix} 6 \\ \\ 4 \end{pmatrix} \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^2 = 15 \cdot \frac{1}{81} \cdot \frac{4}{9} = \frac{20}{243} \approx 8,23\%,$$

onde:

$n = 6$ *é o número de jogos,*

$k = 4$ *é o número de vitórias,*

$p = \frac{1}{3}$ *é a probabilidade de sucesso, ou seja, a probabilidade de ganhar entre ganhar, empatar ou perder os jogos,*

$p = \frac{2}{3}$ *é a probabilidade de fracasso, ou seja, a probabilidade de empatar ou perder entre ganhar, empatar ou perder os jogos.*

Definição 4.4.9. (Distribuição de Poisson) *X é variável aleatória Poisson com parâmetro λ se sua função probabilidade é dada por:*

$$p(k) = P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}, k = 0, 1, 2, \ldots$$

A distribuição de Poisson é um caso especial, por não dizer, um caso limite de distribuição binomial, quando $n \longrightarrow \infty$, $p \longrightarrow 0$ e $\mu = np$ permanece constante, e se aplica no caso em que, ao invés de observar o número de sucessos em n realizações independentes de um experimento de Bernoulli, observa-se o número de sucessos em um intervalo contínuo de observação t. A notação é $X \sim$ Poisson (λ). Esse intervalo contínuo de observações pode ser um intervalo qualquer em que se vai observar a ocorrência de sucessos, como, por exemplo, um intervalo de tempo, de comprimento, de área ou de volume.

São exemplos de fenômenos que segue um distribuição de Poisson:

- Número de veículos que chegam a um posto de combustível em determinado período de um dia;
- número de chamadas telefônicas que chegam a uma central de *call center* em um determinado intervalo de tempo;
- número de defeitos encontrados em uma chapa de metal vendida em uma metalúrgica;
- número de impurezas encontradas em um determinado volume de uma substância.

Para que uma variável aleatória X tenha uma distribuição de Poisson, ela deve seguir os seguintes requisitos:

i) Para intervalos Δt muito pequenos, a probabilidade de ocorrência de mais de um sucesso é desprezível;

ii) Para intervalos Δt muito pequenos, a probabilidade de ocorrência de um sucesso é proporcional ao tamanho do intervalo e igual a $\lambda \Delta t$, onde $\lambda > 0$ é a taxa de sucessos por unidade de observação;

iii) as ocorrências de sucessos em intervalos disjuntos (não sobrepostos) são independentes.

Então, se uma variável aleatória X_t igual ao número de sucessos em um intervalo t de observação tem distribuição de Poisson, pode-se demonstrar que sua distribuição de probabilidades é dada por:

$$P(X_t = k) = \frac{e^{-\lambda t} \cdot (\lambda t)^k}{k!}, k = 0, 1, 2, \ldots$$

Sendo $\mu = \lambda \cdot t$ o número de ocorrências no intervalo t, a expressão acima pode ser escrita da forma:

$$P(X_t = k) = \frac{e^{-\mu} \cdot (\mu)^k}{k!}, k = 0, 1, 2, \ldots$$

Exemplo 4.4.9. *Sabe-se que um determinado telefone recebe em média 0,75 chamada por hora. Supondo que a distribuição de Poisson é adequada, a probabilidade de que este telefone receba em 4 horas exatamente duas chamadas, é aproximadamente 0,224.*

Demonstração. Sendo X a variável aleatória igual ao número de chamadas recebidas em um intervalo de 4 horas, teremos: $\lambda = 0,75$ chamadas por hora e $t = 4$ horas. Daí segue que:

$$\mu = \lambda \cdot t = 0,75 \cdot (4) = 3,$$

$$P(X = 2) = \frac{e^{-3} \cdot (3)^2}{2!} \approx 0,224.$$

□

Nas hipóteses do exemplo anterior, a probabilidade de que este telefone receba no máximo uma chamada é aproximadamente 0,199.

Demonstração.

$$P(X \leqslant 1) = P(X = 0) + P(X = 1) = \frac{e^{-3} \cdot (3)^0}{0!} + \frac{e^{-3} \cdot (3)^1}{1!} \approx 0,199.$$

□

E a probabilidade de que este mesmo telefone receba no mínimo três chamadas é aproximadamente 0,057.

Demonstração.

$$P(X \geqslant 3) = 1 - P(X < 3) =$$
$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] =$$
$$= 1 - [\frac{e^{-3}.(3)^0}{0!} + \frac{e^{-3}.(3)^1}{1!} + \frac{e^{-3}.(3)^2}{2!} \approx$$
$$\approx 0,0576.$$

□

Os casos de distribuição até agora descritos, são casos de distribuição de probabilidades para variáveis discretas. Apresenta-se agora, um caso de distribuição contínua de probabilidades, a distribuição normal, também conhecida como curva de Gauss, que é a mais importante das distribuições contínuas de probabilidade.

Definição 4.4.10. (Distribuição Normal) *Uma variável aleatória X contínua tem distribuição normal com parâmetros μ e σ^2 se sua função densidade de probabilidade for dada por:*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

onde e é a base dos logaritmos neperianos, cujo valor aproximado é 2,718 281 828. A notação é $X \sim$ Normal (μ, σ^2).

Demonstra-se que os parâmetros μ e σ^2 são respectivamente a média e a variância da distribuição normal, isto é, $E(X) = \mu$ e $Var(X) = \sigma^2$. O grafico da distribuição normal tem a forma de um sino, é simétrico em relação a μ e tem como pontos de inflexão $(\mu - \sigma)$ e $(\mu + \sigma)$. Além disso, $f(x) \longrightarrow 0$, quando $x \longrightarrow \pm\infty$. Para se calcular as probabilidades da variável X assumir valores em certos intervalos, probabilidades essas dadas pelas áreas sob a curva da distribuição normal nesses intervalos, utiliza-se uma distribuição normal particular de média $\mu = 0$ e variância $\sigma^2 = 1$, denominada *distribuição normal padronizada.* As áreas sob essas curvas estão tabeladas e fornecem diretamente essas probabilidades.

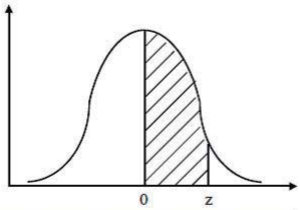
Utilizando-se as propriedades da média, da variância e da distribuição normal, pode-se demonstrar que, se uma variável aleatória X tem distribuição normal com média μ e variância σ^2, então a variável aleatória

$$Z = \frac{X - \mu}{\sigma}$$

tem distribuição normal padronizada, com média 0 e variância 1. Existem várias tabelas de distribuição normal padronizada, porém a utilizada no presente trabalho será a que fornece as áreas centrais entre 0 e Z.

Vale ressaltar que as probabilidades normais não podem ser calculadas analiticamente.

ÁREA SUBTENDIDA PELA CURVA NORMAL REDUZIDA DE 0 A Z



z	0	1	2	3	4	5	6	7	8	9
0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Figura 7: Tabela Z

Exemplo 4.4.10. *A duração de um certo componente eletrônico tem média de 850 dias e variância de 1600 dias. Definindo X como a duração do componente em horas, e sabendo que a duração é normalmente distribuída, a probabilidade desse componente durar mais de 800 dias será 0,8944.*

Demonstração.

$$P(X > 800) = P\left(\frac{X-\mu}{\sigma} > \frac{800-850}{\sqrt{1600}}\right) = P(Z > -1,25)$$

$$= P(-1,25 < Z < 0) + P(Z > 0) = P(0 < Z < 1,25) + P(Z > 0) =$$

$$\text{que seguindo a tabela Z em anexo, tem-se: } 0,3944 + 0,5 = 0,8944.$$

□

Os casos de distribuições descritos até agora, foram todos casos unidimensionais, ou seja, o resultado do experimento era registrado como um único número x qualquer. No entanto, existem casos, e que se interessa observar duas ou mais características simultaneamente, como por exemplo a estatura (X) e o peso (Y) de um adulto. Daí an ecessidade das seguintes definições:

Definição 4.4.11. (Variável Aleatória Bidimensional) *Sejam A um experimento e Ω um espaço amostral associado a A. Sejam $X = X(\omega)$ e $Y = Y(\omega)$ duas variáveis aleatórias, cada uma associando um número real a cada resultado $\omega \in \Omega$. Denomina-se $(X;Y)$ uma variável aleatória bidimensional, ou vetor aleatório.*

Definição 4.4.12. (Variável Aleatória n-dimensional) *Sejam A um experimento e Ω um espaço amostral associado a A. Sejam $X_1 = X_1(\omega), X_2 = X_2(\omega), \ldots, X_n = X_n(\omega)$, n variáveis aleatórias, cada uma associando um número real a cada resultado $\omega \in \Omega$. Denomina-se $(X_1, X_2, \ldots, X_n)$ uma variável aleatória n-dimensional, ou vetor aleatório n-dimensional.*

Assim como nos casos de variáveis aleatórias unidimensionais, distingue-se também dois tipos principais de variáveis aleatórias bidimensionais: variáveis aleatórias discretas e contínuas.

Definição 4.4.13. (Variável Aleatória Discreta Bidimensional) $(X;Y)$ *é uma variável aleatória discreta bidimensional se os possíveis valores de* $(X;Y)$ *forem finitos, ou infinitos enumeráveis, isto é, os possíveis valores* $(X;Y)$ *podem ser representados por* $(x_i;y_j)$, $i = 1,2,\ldots,n,\ldots; j = 1,2,\ldots,m,\ldots$

Definição 4.4.14. (Variável Aleatória Contínua Bidimensional) $(X;Y)$ *é uma variável aleatória contínua bidimensional se há uma função* $f(x,y)$ *tal que :*

- $f(x,y) \geq 0$,
- $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$,

ou seja, se $(X;Y)$ *puder assumir todos os valores em algum conjunto não numerável do plano euclidiano.*

Assim como se define independência entre dois eventos de um espaço amostral, é importante definir a independência entre variáveis aleatórias:

Definição 4.4.15. (Densidades Condicionais) *As funções* $f_{X|Y}(x|y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$ *e* $f_{Y|X}(y|x) = \dfrac{f_{XY}(x,y)}{f_X(x)}$ *são chamadas de densidade marginal de* X *dado* Y *e de de densidade marginal de* Y *dado* X.

Definição 4.4.16. (Probabilidades Condicionais) *As probabilidades condicionais de* X *dado* Y *e de* Y *dado* X *são respectivamente* $P(X \in A|Y = y) = \displaystyle\int_A f_{X|Y}(x,y)dx$ *e* $P(Y \in A|X = x) = \displaystyle\int_A f_{Y|X}(y,x)dy$.

Definição 4.4.17. (Independência de Variáveis Aleatórias) *As variáveis aleatórias* $X_1, X_2, \ldots, X_n$ *são independentes se para quaisquer conjuntos* $A_i \subset \mathbb{R}$*(Borelianos),* $i = 1,2,\ldots,n$, *temos:*

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

Proposição 4.4.2. *(a) Seja* $(X;Y)$ *uma variável aleatória discreta bidimensional.* X *e* Y *são variáveis aleatórias independentes se, e somente se,* $P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j)$ *para quaisquer* i *e* j. *(b) Seja* $(X;Y)$ *uma variável aleatória contínua bidimensional* X *e* Y *são variáveis aleatórias independentes se, e somente se,* $f(x;y) = f_X(x)f_Y(y)$ *para todo* $(x;y)$ *, onde* f *é a função de probabilidade conjunta, e* f_X *e* f_Y *são as funções de probabilidade marginais de* X *e* Y *, respectivamente.*

Outra maneira de tratar a independência entre duas variáveis aleatórias, é através do seguinte teorema:

Teorema 4.4.1. *(a) Seja $(X;Y)$ uma variável aleatória discreta bidimensional. X e Y são variáveis aleatórias independentes se, e somente se, $P(X = x_i|Y = y_j) = P(X = x_i)$ para todo i e j [Ou, o que é equivalente se, e somente se, $P(Y = y_i|x_i) = P(Y = y_j)$ para todo i e j]. (b) Seja $(X;Y)$ uma variável aleatória contínua bidimensional. X e Y são variáveis aleatórias independentes se, e somente se, $f_{X|Y}(x|y) = f_X(x)$, ou equivalente, se e somente se, $f_{Y|X}(y|x) = f_Y(y)$.*

Propriedade 4.4.1. *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias com função de distribuição conjunta $F(x_1, x_2, \ldots, x_n)$ e funções de distribuição marginais $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$ respectivamente. Então, as variáveis aleatórias $X_1, X_2, \ldots, X_n$ são independentes se, e somente se $F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$.*

Segundo Spiegel [18], um teorema de grande importância para o estudo da probabilidade e da estatística, que fornece uma propriedade geral de variáveis aleatórias discretas ou contínuas tendo média e variância finitas, é conhecido como desigualdade de Chebyshev.

Teorema 4.4.2. (Desigualdade de Chebyshev) *Suponha que X seja uma variável aleatória (discreta ou contínua) tendo média μ e variância σ^2, as quais finitas. Então se ϵ for um número positivo qualquer,*

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

O seguinte teorema, é consequência importante da Desigualdade de Chebychev:

Teorema 4.4.3. (Lei Fraca dos Grandes Números) *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias mutualmente independentes (discretas ou contínuas), cada uma tendo média μ e variância σ^2 finitas. Então, se $S_n = X_1 + X_2 + \ldots + X_n$, com $n = 1, 2, 3, \ldots$,*

$$\lim_{n \to +\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq e\right) = 0.$$

Como S_n/n é a média aritmética de $X_1, X_2, \ldots, X_n$, este teorema estabelece que a probabilidade da diferença da média S_n/n e seu valor esperado μ ser mais do que e, se aproxima de zero quando $n \to +\infty$. Um resultado mais forte, que poderíamos supor como verdadeiro, é que $\lim_{x \to +\infty} S_n/n = \mu$, mas isto é falso. Entretanto, podemos provar que $\lim_{x\to+\infty} S_n/n = \mu$ com probabilidade 1, resultado conhecido como a Lei forte dos grandes números, e por isto, o teorema 4.3.3 é chamado de Lei fraca dos grandes números.

Teorema 4.4.4. (Teorema Central do Limite) *Sejam $(X_1, X_2, \ldots, X_n)$, variáveis aleatórias independentes e identicamente distribuídas (isto é, todas têm a mesma função de probabilidade no caso discreto ou função densidade no caso contínuo) com média μ e variância σ^2 finitas. Então se $S_n = X_1 + X_2 + \ldots + X_n$, com $n \in \mathbb{N}^*$,*

$$\lim_{n\to+\infty} P\left(a \leq \frac{S_n - n\cdot\mu}{\sigma\cdot\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b \epsilon^{-u^2/2}du,$$

isto é, a variável $(S_n - n\cdot\mu)/\sigma\cdot\sqrt{n}$, que é a variável aleatória padronizada correspondente a S_n, é assintoticamente normal.

Ou seja, se $(X_1, X_2, \ldots, X_n)$, é uma amostra aleatória simples retirada de uma população com médias μ e variâncias σ^2 finitas, a distribuição amostral da média $\overline{X}$ aproxima-se, para n grande, de uma distribuição normal, com média μ e variância σ^2.

Corolário 4.4.1. (Aproximação Normal à Binomial) *Seja X uma variável aleatória comparâmetros n e p. Se n é grande e nem p nem $q = 1 - p$ estão muito próximos de zero, a distribuição binomial pode ser aproximada por uma distribuição normal através da variável aleatória padronizada dada por:*

$$Z = \frac{X - n\cdot p}{\sqrt{n\cdot p\cdot q}},$$

onde X é a variável aleatória dando o número de sucessos em n ensaios de Bernoulli e p é a probabilidade de sucesso. A aproximação se torna melhor com n crescendo e é exata no caso limite.

Este fato pode ser descrito como:

$$\lim_{n\to+\infty} P\left(a \leq \frac{X - n\cdot p}{\sqrt{n\cdot p\cdot q}} \leq b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-u^2/2}du,$$

que é uma aplicação direta do Teorema Central do Limite.

5 Inferência Estatística

Tendo já estudado na seção 4.1 os procedimentos necessários para resumir descritivamente as variáveis associadas a pelo menos um conjunto de dados, e nas seções 4.2

e 4.3 os modelos probabilísticos capazes de representar o comportamento de algumas variáveis, tem-se nesta seção a intenção de aprofundar-se nos argumentos necessários para que se possa afirmar características sobre uma população a partir das observações amostrais, ou seja, inferir estatisticamente.

Como discutiu-se no capítulo 3, o uso de informações de uma amostra para concluir sobre o todo, faz parte do cotidiano das pessoas, mesmo que informalmente e de maneira intuitiva. O objetivo nesta seção, norteia-se em conceituar formalmente os princípios utilizados no dia a dia, para que possam ser utilizados de forma científica, em aplicações nas mais diversas áreas do conhecimento.

Como já foi dito, tem-se conhecimento de modelos probabilísticos que medem, ou pelo menos, tendem a medir a variabilidade de fenômenos de acordo com suas ocorrências, que são as distribuições de probabilidades. Mas na prática, os estudiosos podem até ter ideia de qual forma de distribuição, mas não dos valores dos parâmetros que a especifica.

Pode-se pensar que a distribuição das alturas dos alunos de uma escola pode ser representada por um modelo normal, mas daí a conhecer os parâmetros (média e variância) já é outro contexto que na maioria das vezes não é acessível. Mesmo porque, para se conhecer os parâmetros, seria necessário medir todos os alunos da escola, o que além de dispendioso trabalho, descarta a utilidade da inferência.

> "Raramente se consegue obter a distribuição exata de alguma variável, ou porque isso é muito dispendioso, ou muito demorado ou às vezes porque consiste num processo destrutivo. Por exemplo, se estivéssemos observando a durabilidade de lâmpadas e testássemos todas até queimarem, não restaria nenhuma para ser vendida. Assim, a solução é selecionar parte dos elementos (amostra), analisá-la e inferir propriedades para o todo (população)."(Bussab e Morettin, 2012, p.261)[3]

Alguns conceitos básicos:

Definição 5.0.18. (População) *Conjunto de todos os elementos ou resultados sob investigação.*

Definição 5.0.19. (Amostra) *Qualquer subconjunto da população.*

Exemplo 5.0.11. *Seja uma pesquisa para estudar as estaturas de 500 alunos de uma escola. Seleciona-se uma amostra de 50 alunos, e anota-se suas estaturas. A variável* *a ser observada é "estatura". A população é o conjunto formado pelas 500 estaturas correspondentes aos 500 alunos, e a amostra é o conjunto das 50 estaturas dos alunos selecionados. Ao estudar a distribuição das alturas da amostra, espera-se que esta reflita a distribuição das alturas de todos os alunos.*

Reiterando, o objetivo da Inferência estatística, é produzir afirmações sobre uma dada característica da população, a partir de informações colhidas de uma parte dessa população, sendo essa característica, representada por uma variável aleatória. Se houvesse conhecimento completo sobre a função de probabilidade no caso discreto, ou sobre a função densidade de probabilidade no caso contínuo, não seria necessário a escolha de uma amostra, mas isso raramente, como já foi dito, acontesse ou é possível. Em todo caso, o uso de uma amostra ajuda a formar uma opinião acerca da população.

Mesmo sendo de grande importância a identificação e a descrição de uma população no processo de inferência, é comum aos estudiosos dedicarem especial atenção ao selecionar uma amostra. O modo de se obter uma amostra é tão importante, que existem vários modos de fazê-lo, a ponto desses procedimentos serem uma especialidade dentro da estatística. porém, neste estudo, será trabalhado apenas a **Amostragem Aleatória Simples (AAS)**, que é a maneira mais fácil de se obter uma amostra probabilística de uma população.

Podemos obter uma amostra nessas condições, escrevendo cada elemento da população em um cartão, misturando e sorteando a quantidade escolhida para a amostra. Obviamente, este procedimento é inviável para o caso de populações muito grandes. Como processo alternativo, os elementos podem ser numerados e sorteados por meio de uma tabela de números aleatórios ou por meio de programas computacionais que geram números aleatórios. Utilizando-se um procedimento aleatório, sorteia-se um elemento da população, sabendo que todos os elementos são equiprováveis. Repete-se até que sejam sorteadas n unidades para a amostra. Pode-se ter uma AAS com reposição ou sem reposição. Considerando a quantidade de informações contidas em uma amostra, amostrar sem reposição é mais vantajoso, porém, com reposição, o tratamento teórico será mais simples, devido a independência entre as unidades selecionadas.

Definição 5.0.20. (Amostra Aleatória da Variável X) *Seja X uma variável aleatória com uma distribuição de probabilidade especificada. Sejam n variáveis aleatórias $X_1; X_2; \ldots; X_n$ independentes e tendo cada uma delas a mesma distribuição que X. Nesse caso, denomina-se $(X_1; X_2; \ldots; X_n)$ uma amostra aleatória da variável aleatória X.*

Ou seja, a amostra será a n-upla ordenada (ou vetor aleatório) $(X_1; X_2; \ldots; X_n)$, onde X_i indica a observação do i-ésimo elemento sorteado.

Obtida uma amostra, deseja-se comumente, utilizá-la para produzir alguma característica específica, como por exemplo, a média da amostra $(X_1; X_2; \ldots; X_n)$ que será dada por:

$$\overline{X} = \frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)$$

Observa-se que X também é uma variável aleatória, mas existe o interesse em qualquer outra carcterística da amostra, que será sempre uma função do vetor aleatório $(X_1; X_2; \ldots; X_n)$.

Definição 5.0.21. (Estatística) *Uma estatística é uma característica da amostra, ou seja, uma estatística T é uma função de $(X_1; X_2; \ldots; X_n)$.*

As estatísticas mais comuns são:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \ \text{ : média da amostra,}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}) \ \text{ : variância da amostra,}$$

$$X_1 = min(X_1; X_2; \ldots; X_n) \ \text{ : menor valor da amostra,}$$

$$X_n = max(X_1; X_2; \ldots; X_n) \ \text{ : maior valor da amostra,}$$

$$W = X_n - X_1 \ \text{ amplitude da amostra.}$$

Para facilitar a linguagem utilizada em Inferência estatística, diferencia-se as características da amostra e da população.

Definição 5.0.22. (Parâmetro) *Um parâmetro é uma medida para descrever uma característica da população.*

Os símbolos mais comuns são dados na tabela a seguir:

Tabela 23: Símbolos mais comuns para amostras e populações

Denominação	População	Amostra
Média	$\mu = E(X)$	$\overline{X} = \sum X_i/n$
Mediana	$Md = Q_2$	$md = q_2$
Variância	$\sigma^2 = Var(X)$	$S^2 = \sum(X_i - \overline{X})^2/(n-1)$
Número de elementos	N	n
Proporção	p	$\hat{p}$
Quantil	$Q(p)$	$q(p)$
Quartis	Q_1, Q_2, Q_3	q_1, q_2, q_3
Intervalo inter-quartil	$d_Q = Q_3 - Q_1$	$d_q = q_3 - q_1$
Função densidade	$f(x)$	Histograma
Função de distribuição	$F(x)$	$F_e(x)$

Fonte: Bussab e Morettin, 2012, p. 272 [3]

Definição 5.0.23. (Distribuição Amostral) *Uma estatística amostral calculada de* X_1*;* X_2*; ...;* X_n *é uma função destas variáveis aleatórias, e é portanto, ela mesma, uma variável aleatória. A distribuição de probabilidade desta estatística amostral é chamada Distribuição Amostral.*

Definição 5.0.24. (Distribuição Amostral da Média) *Seja* $F(x)$ *a função de distribuição de probabilidade de alguma população da qual extraímos uma amostra de tamanho* n*. A distribuição de probabilidade da estatística amostral* X *é chamada Distribuição Amostral da Média Amostral ou Distribuição Amostral das Médias.*

Exemplo 5.0.12. *Seja a população* $1; 3; 5; 5; 7$ *e todas as amostras de tamanho 2 coletadas com reposição. Temos na Tabela 24:*

Tabela 24: Distribuição amostral da Estatística $\overline{X}$

$\overline{x}$	1	2	3	4	5	6	7	Total
$P(\overline{X} = \overline{x})$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{5}{25}$	$\frac{6}{25}$	$\frac{6}{25}$	$\frac{4}{25}$	$\frac{1}{25}$	1

Fonte: Bussab e Morettin, 2012, p. 274 [3]

Calculando a partir da população, tem-se média $\mu = 4,2$ e variância $\sigma^2 = 4,16$. Da distribuição amostral, tem-se:

$$E(\overline{X}) = 1 \cdot \frac{1}{25} + 2 \cdot \frac{2}{25} + 3 \cdot \frac{5}{25} + 4 \cdot \frac{6}{25} + 5 \cdot \frac{6}{25} + 6 \cdot \frac{4}{25} + 7 \cdot \frac{1}{25} = 4,2,$$

$$Var(\overline{X}) = 2,08.$$

Este exemplo ilustra o teorema a seguir:

Teorema 5.0.5. *Seja X uma variável aleatória com média μ e variância σ^2, e seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória simples de X. Então, $E(X) = \mu$ e $Var(X) = \dfrac{\sigma^2}{n}$*

Teorema 5.0.6. *(TLC)Para amostras aleatórias simples $(X_1, X_2, \ldots, X_n)$, retiradas de uma população com média $\mu < \infty$ e variância $0 < \sigma^2 < +\infty$, a distribuição amostral da média $\bar{X}$ aproxima-se, para n grande, de uma distribuição normal, com média μ e variância σ^2/n, isto é, se $(X_1, X_2, \ldots, X_n)$ for uma amostra aleatória simples da população X, com média μ e variância σ^2 finita, e $\bar{X} = (X_1 + X_2 + \ldots + X_n)/n$, então:*

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

tem distribuição aproximadamente normal padrão para n grande.

Definição 5.0.25. (Erro Amostral da Média) *Denotado por e, é a variável aleatória que mede a diferença entre a estatística $\overline{X}$ e o parâmetro μ, isto é: $\overline{X} - \mu$.*

Corolário 5.0.2. *A distribuição de e se aproxima de uma distribuição normal com média 0 e variância μ e variância σ^2/n, isto é:*

$$\frac{\sqrt{n}e}{\sigma} \sim N(0,1), \text{ para } n \text{ grande.}$$

Considerando agora, uma população em que a proporção dos elementos portadores de uma certa característica é p, pode-se definir uma variável aleatória X da seguinte maneira:

$$X = \begin{cases} 1, & \text{se o elemento for portador da característica,} \\ 0, & \text{se o elemento não for portador da característica.} \end{cases}$$

Assim:

$$\mu = E(X) = p,$$
$$\sigma^2 = Var(X) = p(1-p).$$

Retirando uma amostra aleatória simples dessa população, e indicando por Y_n o total de indivíduos portadores da cacterística na amostra, teremos que:

$$Y_n \sim Binomial(n, p).$$

Definição 5.0.26. *(Distribuição Amostral da Proporção) Denotando por $\hat{p}$, definimos a proporção de indivíduos portadores da carcterística na amostra, isto é:*

$$\hat{p} = \frac{Y_n}{n},$$

então:

$$P(Y_n = k) = P(Y_n/n = k/n) = P(\hat{p} = k/n).$$

ou seja, a distribuição amostral de $\hat{p}$ é obtida da distribuição de Y_n.

Foi citado na Seção 4.3, que a distribuição binomial pode ser aproximada pela distribuição normal, justificado pelo Teorema do Limite Central. Observa-se que:

$$Y_n = X_1 + X_2 + \ldots + X_n,$$

onde cada X_i, tem distribuição de Bernoulli, com média $\mu = p$ e variância $\sigma^2 = p(1-p)$, e são duas a duas independentes, podendo escrever:

$$Y_n = n \cdot \overline{X}.$$

Mas, pelo Teorema do limite Central, $\bar{X}$ terá distribuição aproximadamente normal, com média p e variância $\dfrac{p(1-p)}{n}$, ou seja:

$$\overline{X} \sim N\left(p, \frac{p(1-p)}{n}\right), \text{ para n grande.}$$

Assim, a transformação $Y_n = n \cdot \overline{X}$ terá distribuição:

$$Y_n \sim N(np, np(1-p)), \text{ para n grande.}$$

Mas como $\bar{X}$ é a própria variável $\hat{p}$, então podemos considerar a distribuição de $\hat{p}$ como aproximadamente normal:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right), \text{ para n grande.}$$

5.1 Estimação

Já vimos que a Inferência Estatística tem como objetivo realizar afirmações acerca de uma população, a partir dos dados de uma amostra. Dois processos básicos da Inferência Estatística são:

- Estimação de Parâmetros,
- Teste de Hipóteses.

Lembrando que **parâmetros** são funções de valores populacionais e **estatísticas** são funções de valores amostrais.

Neste trabalho, iremos discutir apenas os problemas envolvendo estimações, devido ao fato de que o objetivo principal deste é aferir o que se pode da inferência, ser estudado, mesmo que de forma intuitiva, no ensino básico.

Com o intuito de ilustrar as ideias básicas de estimação, segue um exemplo prático:

Exemplo 5.1.1. *Uma amostra de* $n = 200$ *pessoas de uma cidade é escolhida, e a cada pessoa é feita uma pergunta a respeito da construção de um viaduto na entrada da cidade. A resposta à pergunta poderá ser SIM (favorável à construção) ou NÂO (contrária à construção). Deseja-se estimar a proporção de pessoas na cidade favoráveis à construção do viaduto. Se 120 cidadãos respondem SIM à pergunta, então uma estimativa natural para esta proporção seria* $120/200$ *ou* 60%.

Esta resposta é baseada supondo que a amostra representa realmente a população, da mesma forma que outra amostra poderia nos levar a outra estimativa. Um dos propósitos da Inferência Estatística é conhecer as propriedades desses **estimadores**.

No caso particular acima, podemos definir as variáveis aleatórias $X_1, X_2, \ldots, X_n$ tais que:

$$X_i = \begin{cases} 1, & \text{se a i-ésima pessoa da amostra responder SIM}, \\ 0, & \text{se a i-ésima pessoa da amostra responder NÃO}. \end{cases}$$

e fazer $p = P$ (sucesso), onde aqui sucesso significa responder SIM à pergunta.

Assim, se $Y_n = \sum_{i=1}^{n} X_i$, temos que Y_n tem distribuição binomial com parâmetros n e p, e o problema será estimar p. Logicamente, Y_n representa o número de cidadãos que são favoráveis a construção do viaduto, ou seja, que responderam SIM à pergunta, portanto, um possível **estimador** de P é:

$$\hat{p} = \frac{Y_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\text{número de SIM}}{\text{número de indivíduos}}.$$

Então, se $Y_n = k$, isto é, observamos o valor k da variável Y_n, teremos $\hat{p} = \dfrac{k}{n}$ como uma estimativa de p. Considere que $\hat{p}$ é uma variável aleatória, e que k/n é um número, ou seja, um valor da variável aleatória. No exemplo acima, uma estimativa é 60%. Na seção 5 foi visto que o estimador $\hat{p}$ tem distribuição aproximadamente normal com parâmetros $E(\hat{p}) = p$ e $Var(\hat{p}) = p(1-p)/n$ o que indica que o estimador $\hat{p}$ em média acerta p, ou seja, $\hat{p}$ é um estimador **não-viesado** ou **não-viciado** de p. Este fato nos indica também, que para grandes amostras, a diferença entre $\hat{p}$ e p tende a ser pequena, pois se $n \longrightarrow \infty$, $Var(\hat{p}) \longrightarrow 0$. Neste caso, diz-se que $\hat{p}$ é um estimador consistente de p.

Definição 5.1.1. (Estimador) *Um estimador T do parâmetro θ é qualquer função das observações da amostra, ou seja $T = g(X_1, X_2, \ldots, X_n)$.*

Definição 5.1.2. (Estimador Não Viesado) *O estimador T é não viesado para θ, se $E(T) = \theta$, para todo θ.*

Definição 5.1.3. (Estimativa) *Estimativa é o valor assumido pelo estimador em uma particular amostra.*

Exemplo 5.1.2. *No Exemplo 5.1.1, $\hat{p}$ é um estimador de p, enquanto 60% é uma estimativa de p.*

Definição 5.1.4. (n-ésimo momento de X) *Seja X uma variável aleatória. Para cada inteiro positivo n, o n-ésimo momento de X, denotado por μ_n, é dado por:*

$$\mu_n = E(X^n),$$

desde que $E(X^n)$ exista.

Além disso, definimos o n-ésimo momento central como sendo $E[(X-E(X))^n]$, caso exista.

Em particular, se X é uma variável aleatória discreta com função de probabilidade $p(x)$, temos que:

$$\mu_n = \sum x^n p(x),$$

e se X é uma variável aleatória contínua com função de probabilidade $f(x)$, temos que:

$$\mu_n = \int_{-\infty}^{\infty} x^n p(x) d(x).$$

Proposição 5.1.1. *Seja X uma variável aleatória com média μ e variância σ^2. Neste caso, as seguintes relações são válidas para os dois primeiros momentos populacionais:*

$$E(X) = \mu,$$

$$E(X^2) = \sigma^2 + \mu^2.$$

A primeira igualdade é imediata e a segunda, segue do fato de que:

$$Var(X) = E(x^2) - E(X)^2 \Longleftrightarrow E(x^2) = Var(X) + E(X)^2 = \sigma^2 + \mu^2.$$

Definição 5.1.5. (n-ésimo momento amostral) *Considere uma amostra de tamanho n da população $(X_1, X_2, \ldots, X_n)$. Define-se o n-ésimo momento amostral por:*

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^n, \text{ com } n = 1, 2, \ldots$$

5.2 Estimadores de Momentos

Definição 5.2.1. (Estimadores de Momentos) *Dizemos que* $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r$ *são estimadores obtidos pelo método dos momentos se eles forem soluções das equações:*

$$m_n = \mu_n, \quad n = 1, 2, \ldots, r.$$

Exemplo 5.2.1. *Seja* $(X_1, X_2, \ldots, X_n)$ *uma amostra aleatória simples de uma distribuição binomial de parâmetros* n *e* p*, ou seja:*

$$P(X = k|n, p) = \left(\begin{array}{c} n \\ \\ k \end{array} \right) \cdot p^k \cdot (1-p)^{n-k},$$

onde a variável X *denota o número de sucessos em* n *ensaios e* $k = 0, 1, \ldots, n$.

Assumindo que os parâmetros n *e* p *sejam desconhecidos, vamos encontrar estimadores para ambos os parâmetros a partir do método dos momentos:*

A partir do Exemplo 5.2.1 e usando o fato de que X_i *tem média* np *e variância* $np(1-p)$*, temos que os dois primeiros momentos populacionais são dados, respectivamente, por:*

$$\mu_1 = E(X) = np \ \text{ e } \ \mu_2 = E(X^2) = np(1-p) + n^2p^2.$$

Igualando os dois primeiros momentos amostrais m_1 *e* m_2 *aos dois primeiros momentos populacionais, temos o seguinte sistema de equações:*

$$\begin{cases} \overline{X} = np, \\ \dfrac{1}{m} \displaystyle\sum_{i=1}^{m} X_i^2 = np(1-p) + n^2p^2. \end{cases}$$

Resolvendo em n *e* p*, obtemos os seguintes estimadores pelo método dos momentos:*

$$\hat{n} = \frac{\overline{X}^2}{\overline{X} - (1/m)\sum_{i=1}^{m}(X_i - \overline{X})^2},$$

$$\hat{p} = \frac{\overline{X}}{\hat{n}}.$$

Este é um típico exemplo em que os estimadores não são os melhores para os parâmetros populacionais de interesse. Na verdade, utilizando estes estimadores, podemos ter estimativas negativas para n e p, o que não pode acontecer, já que estes devem ser números positivos.

5.3 Estimadores de Mínimos Quadrados

Definição 5.3.1. (Estimadores de Mínimos Quadrados) *O método de estimação por mínimos quadrados consiste em minimizar o quadrado das diferenças entre os valores observados de uma amostra e seus respectivos valores esperados.*

Exemplo 5.3.1. *Considera-se o seguinte procedimento em que se interessa no estudo da resistência Y de uma cabo de aço em função de seu diâmetro X. A partir de uma amostra coletada, percebe-se que as variáveis são, aproximadamente, proporcionais, isto é, $Y \approx \theta X$ em que θ é o coeficiente de proporcionalidade. O objetivo é estimar o parâmetro θ, baseado nas medidas disponíveis em uma amostra de 10 unidades mostradas na tabela a seguir:*

Tabela 25: Estudo da resistência Y em função do diâmetro X

X	0,50	0,60	0,75	0,80	0,90	1,05	1,20	1,30	1,50	1,65
Y	2,07	2,24	3,28	3,35	3,81	4,14	4,64	5,13	6,05	6,57

Fonte: Dados Hipotéticos

A partir dessas informações, pode-se concluir que, aparentemente, $\hat{\theta} = 4$ parece ser uma estimativa razoável para o parâmetro θ. Como pode-se verificar a qualidade desta estimativa? Uma forma de fazer isso é verificar as diferenças entre os valores observados Y e os valores esperados utilizando a estimativa, ou seja, $4X$. Na tabela a seguir, tem-se os valores da amostra, os valores esperados, a diferença $Y - 4X$ e as diferenças ao quadrado $(Y - 4X)^2$:

Tabela 26: Estudo da resistência Y em função do diâmetro X

X	Y	$4X$	$Y-4X$	$(Y-4X)^2$
0,50	2,07	2,00	0,07	0,0049
0,60	2,24	2,40	-0,16	0,0256
0,75	3,28	3,00	0,28	0,0784
0,80	3,35	3,20	0,15	0,0225
0,90	3,81	3,60	0,21	0,0441
1,05	4,14	4,20	-0,06	0,0036
1,20	4,64	4,80	-0,16	0,0256
1,30	5,13	5,20	-0,07	0,0049
1,50	6,05	6,00	0,05	0,0025
1,65	6,57	6,60	-0,03	0,0009
	Total		0,28	0,213

Fonte: Dados Hipotéticos

A ideia principal do método baseia-se em minimizar o erro quadrático total da amostra. Para a estimativa $\hat{\theta} = 4$, este erro é dado por $0,213$, porém, pode ser que exista alguma outra estimativa com erro quadrático total menor do que 0,213. Desta forma, o objetivo é minimizar a função:

$$S(\theta) = \sum_{i=1}^{10} (Y_i - \theta X_i)^2.$$

O mínimo da função é obtido derivando a função em relação a θ e igualando o resultado a zero, ou seja, encontrar $\hat{\theta}$ para o qual:

$$\frac{dS(\theta)}{d\theta} = \sum_{i=1}^{10} (Y_i - \hat{(\theta)} X_i)(-2X_i) = 0.$$

Resolvendo essa equação, obtemos o estimador:

$$\hat{\theta}_{MQ} = \frac{\sum_{i=1}^{10}(X_i \cdot Y_i)}{\sum_{i=1}^{10} X_i^2}.$$

Utilizando os dados de X e Y , encontramos $\hat{\theta}_{MQ} = 4,011625$, ou seja, a estimativa que minimiza o erro quadrático total da amostra é dada por $\hat{\theta} = 4,011625$. De fato, utilizando este valor, temos que o erro quadrático total é $0,2114015$.

Deste modo, se assume que, para um dado valor da variável X, os valores da variável Y seguem uma distribuição de probabilidade $f_Y(y)$ centrada em θX, o que é equivalente a dizer que, para cada X, o desvio $\epsilon = Y - \theta X$ segue uma distribuição centrada em zero e, desta forma, é comum escrever:

$$Y = \theta x + \epsilon,$$

com ϵ seguindo a distribuição f_ϵ com média zero. Desta forma, é razoável escolher θ que minimiza a soma dos quadrados dos erros:

$$\sum_{i=1}^{n} \epsilon^2 = \sum_{i=1}^{n}(Y_i - \theta X_i)^2.$$

Se observa que o modelo pode ser generalizado. Isto é, pode-se considerar funções mais gerais dos parâmetros θ, ou seja,

$$Y = g(X, \theta) + \epsilon,$$

e, da mesma forma do exposto acima, deve-se encontrar o valor de θ que minimize a função:

$$S(\theta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n}(Y_i - g(X_i, \theta))^2,$$

para uma amostra $(X_1, Y_1), \ldots, (X_n; Y_n)$ das variáveis X e Y. A solução $\hat{\theta}_{MQ}$ é chamada de estimador de mínimos quadrados (EMQ) de θ.

5.4 Estimadores de Máxima Verossimilhança

Definição 5.4.1. (Estimadores de Máxima Verossimilhança) *Consideremos uma população e uma variável aleatória X, relacionada a essa população, com função de probabilidade (se X é uma variável aleatória discreta) ou função densidade de probabilidade* *(se X é uma variável aleatória contínua) $f(x, \theta)$, sendo θ o parâmetro desconhecido. Retiremos uma amostra aleatória simples de X, de tamanho n, $X_1, \ldots, X_n$, e sejam $x_1, \ldots, x_n$ os valores efetivamente observados. A função de verossimilhança L é definida por:*

$$L(\theta; x_1, \ldots, x_n) = f(x_1; \theta) \times \ldots \times f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

Se X é uma variável aleatória discreta com função de distribuição $p(x, \theta)$, a função de verossimilhança é dada por:

$$L(\theta; x_1, \ldots, x_n) = p(x_1; \theta) \times \ldots \times p(x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta).$$

que deve ser interpretada como uma função de θ. O estimador de máxima verossimilhança de θ é o valor que maximiza $L(\theta; x_1, \ldots, x_n)$.

O princípio de máxima verossimilhança é um dos procedimentos usados para se obter estimadores. O estimador de máxima verossimilhança (EMV) pode ser encontrado seguindo os passos abaixo:

- Encontrar a função de verossimilhança,
- Aplicar a função ln,
- Derivar em relação ao parâmetro θ,
- Igualar o resultado a zero,
- Verificar que este estimador é ponto de máximo.

Exemplo 5.4.1. *Seja X uma variável aleatória com distribuição Bernoulli(p). Tomemos uma amostra aleatória $X_1, \ldots, X_n$ de X. O estimador de máxima verossimilhança para p é $\hat{p} = \frac{1}{n} \cdot \overline{X}$.*

Demonstração. Como $X \sim$ Bernoulli (p), a função de probabilidade de X é:

$$p_X(x) = p^x (1-p)^{1-x}, x = 0, 1.$$

Dessa forma, a função de verossimilhança é dada por:

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} (1-p)^{\sum_{i=1}^{n} (1-x_i)}$$

Para encontrar o estimador de máxima verossimilhança para p, devemos encontrar o valor de p para o qual a função de verossimilhança $L(p; x_1, \ldots, x_n)$ é máxima.

Aplicando a função logaritmo natural (ln) na função de verossimilhança $L(p; x_1, \ldots, x_n)$, temos que:

$$L(p; x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i ln(p) + \sum_{i=1}^{n} (1 - x_i) ln(1 - p),$$

e, derivando em relação a p, segue que:

$$\frac{dlnL(p; x_1, \ldots, x_n)}{dp} = \frac{(1-p)\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n}(1 - x_i)}{p(1-p)}.$$

Igualando o resultado a zero, obtemos que:

$$\frac{(1-\hat{p})\sum_{i=1}^{n} x_i - \hat{p}\sum_{i=1}^{n}(1 - x_i)}{\hat{p}(1-\hat{p})} = 0 \Longleftrightarrow \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

Utilizando o teste da segunda derivada, verifica-se que $\hat{p} = \frac{1}{n} \cdot \overline{X}$ é realmente um estimador de máxima verossimilhança para p.

$\square$

5.5 Intervalos de Confiança

Definição 5.5.1. (Intervalos de Confiança) *Um intervalo de confiança (IC) é um intervalo estimado de um parâmetro de interesse de uma população. Em vez de estimar o parâmetro por um único valor, é dado um intervalo de estimativas prováveis. O quanto estas estimativas são prováveis será determinado pelo coeficiente de confiança* $(1 - \alpha)$, *para* $\alpha \in (0, 1)$.

Intervalos de confiança são usados para indicar a confiabilidade de uma estimativa. Por exemplo, um intervalo de confiança pode ser usado para descrever o quanto os resultados de uma pesquisa são confiáveis. Sendo todas as estimativas iguais, uma pesquisa que resulte num intervalo de confiança pequeno é mais confiável do que uma que resulte num intervalo de confiança maior.

Se G e H são estatísticas (isto é, funções da amostra) cuja distribuição de probabilidade dependa do parâmetro θ, e

$$P(G < \theta < H|\theta) = 1 - \alpha$$

então o intervalo aleatório $(G; H)$ é um intervalo de confiança com nível $100(1 - \alpha)\%$ para θ. Portanto, pode-se interpretar o intervalo de confiança como um intervalo que contém os valores "plausíveis"que o parâmetro θ pode assumir. Assim, a amplitude do intervalo está associada a incerteza que se tem a respeito do parâmetro.

Considere $X_1, X_2, \ldots, X_n$ uma amostra aleatória retirada de uma população com distribuição f_θ que depende do parâmetro θ. Por exemplo, toma-se $X_1, X_2, \ldots, X_n$ uma amostra aleatória com distribuição normal com média μ desconhecida e desvio padrão conhecido $\sigma = 1$. Para se propor um intervalo de confiança para o parâmetro θ, é necessário introduzir o conceito de quantidade pivotal:

Definição 5.5.2. (Quantidade pivotal) *Uma função Q da amostra $(X_1, X_2, \ldots, X_n)$ e do parâmetro θ cuja distribuição de probabilidade não depende do parâmetro θ é denominada quantidade pivotal.*

Desta forma, dado o nível de confiança $1 - \alpha$, toma-se:

$$1 - \alpha = P(q_1 \leq Q(X_1, X_2, \ldots, X_n; \theta) \leq q_2)$$

Se a quantidade pivotal Q for inversível, pode-se resolver a inequação acima em relação a θ e obter um intervalo de confiança.

Exemplo 5.5.1. *Suponha que se queira estimar a média μ de uma população com distribuição normal com variância σ^2 conhecida. O estimador de máxima verossimilhança para a média populacional μ é dado pela média amostral $\overline{X}$ de uma amostra de tamanho n. Assim, tem-se a seguinte quantidade pivotal $e = (\overline{X} - \mu) \sim N(0; \sigma^2/n)$.*

$$P(|e| < 1,96 \cdot \sigma^2/n) = 0,95$$
$$P(|\overline{X} - \mu| < 1,96 \cdot \sigma^2/n) = 0,95$$
$$P(\overline{X} - 1,96 \cdot \sqrt{\sigma^2/n} < \mu < \overline{X} + 1,96 \cdot \sqrt{\sigma^2/n}) = 0,95$$

Para interpretar o intervalo de confiança da média, se assume que os valores foram amostrados de forma independente e aleatória de um população com distribuição normal com média μ e variância σ^2. Dado que estas suposições são válidas, tem-se 95%

de confiança do intervalo conter o verdadeiro valor da média populacional. Em outras palavras, se forem produzidos diversos intervalos de confiança provenientes de diferentes amostras independentes de mesmo tamanho, pode-se esperar que aproximadamente 95% *destes intervalos devem conter o verdadeiro valor da média populacional.*

6 Estocástica no Ensino Fundamental e Médio

A Estatística é cada vez mais presente no dia a dia das pessoas, principalmente pelo excesso crescente de quantidade de informações a que são expostas.

Sendo assim, é necessário que conceitos básicos da Estatística como amostra, população, parâmetro, estimativa, censo, margem de erro, nível de confiança, entre outros, sejam entendidos pela sociedade. Daí um importante motivo pelo qual a Estatística deva ser trabalhada desde as séries iniciais até o ensino médio, mesmo porque, dado continuidade aos estudos, será raro alguma área de atuação que não utilize a Estatística como ferramenta básica.

Analisando os Parâmetros Curriculares Nacionais (PCNs), verifica-se que o ensino da Estatística está inserido no Bloco "Tratamento da Informação".

Para o primeiro ciclo do ensino fundamental 1, tem-se como conteúdos conceituais e procedimentais, a leitura e interpretação de informações contidas em imagens, a coleta e organização de informações, a criação de registros pessoais para comunicação das informações coletadas, a exploração da função do número como código na organização de informações (linhas de ônibus, telefones, placas de carros, registros de identidade, bibliotecas, roupas, calçados), a interpretação e elaboração de listas, tabelas simples, de dupla entrada e gráficos de barra para comunicar a informação obtida e a produção de textos escritos a partir da interpretação de gráficos e tabelas.

Para o segundo ciclo do ensino fundamental 1, tem-se como conteúdos conceituais e procedimentais, a coleta, organização e descrição de dados, a leitura e interpretação de dados apresentados de maneira organizada (por meio de listas, tabelas, diagramas e gráficos) e construção dessas representações, a interpretação de dados apresentados por meio de tabelas e gráficos, para identificação de características previsíveis ou aleatórias de acontecimentos, a produção de textos escritos, a partir da interpretação de gráficos e tabelas, a construção de gráficos e tabelas com base em informações contidas em textos jornalísticos, científicos ou outros, a obtenção e interpretação de média aritmética, a exploração da ideia de probabilidade em situações-problema simples, identificando sucessos possíveis, sucessos seguros e as situações de "sorte", a utilização de informações dadas para avaliar probabilidades e a identificação das possíveis maneiras de combinar elementos de uma coleção e de contabilizá-las usando estratégias pessoais.

Para o terceiro ciclo, tem-se como conteúdos conceituais e procedimentais, a coleta, organização de dados e utilização de recursos visuais adequados (fluxogramas, tabelas e gráficos) para sintetizá-los, comunicá-los e permitir a elaboração de conclusões, a leitura e interpretação de dados expressos em tabelas e gráficos, a compreensão do significado da média aritmética como um indicador da tendência de uma pesquisa, a representação e contagem dos casos possíveis em situações combinatórias e a construção do espaço amostral e indicação da possibilidade de sucesso de um evento pelo uso de uma razão.

Findando o ensino fundamental 1, tem-se para o quarto ciclo, como conteúdos conceituais e procedimentais, leitura e interpretação de dados expressos em gráficos de colunas, de setores, histogramas e polígonos de freqüência, organização de dados e construção de recursos visuais adequados, como gráficos (de colunas, de setores, histogramas e polígonos de freqüência) para apresentar globalmente os dados, destacar aspectos relevantes, sintetizar informações e permitir a elaboração de inferências, compreensão de termos como freqüência, freqüência relativa, amostra de uma população para interpretar informações de uma pesquisa, a distribuição das freqüências de uma variável de uma pesquisa em classes de modo que resuma os dados com um grau de precisão razoável, a obtenção das medidas de tendência central de uma pesquisa (média, moda e mediana), compreendendo seus significados para fazer inferências, a construção do espaço amostral, utilizando o princípio multiplicativo e a indicação da probabilidade de um evento por meio de uma razão, a elaboração de experimentos e simulações para estimar probabilidades e verificar probabilidades previstas.

Observa-se que somente no final do quarto ciclo, ou seja, no 9o ano, é que se propõe dar significado, mesmo que de forma intuitiva, a inferência estatística.

Como os conteúdos do bloco Análise de dados e probabilidade têm sido recomendados para todos os níveis da educação básica, assim como é apresentado no fundamental, este também é, em especial retomado no ensino médio.

Um dos motivos está na importância das ideias de incerteza e de probabilidade, associadas aos chamados fenômenos aleatórios, presentes de forma essencial tanto no cotidiano como no meio científico. A retomada deste conteúdo possibilita aos alunos ampliarem e formalizarem seus conhecimentos sobre o raciocínio combinatório, probabilístico e estatístico. Para que os alunos tenham uma visão concreta da importância dos modelos probabilísticos, é importante que estes tenham oportunidade de ver esses modelos em ação. Por exemplo, é possível simular o que ocorre em certa pesquisa de opinião estimando, com base em uma amostra, a fração de balas de determinada cor em uma caixa, procedimento feito na prática em uma das propostas de aula apresentadas neste trabalho.

O estudo da estatística foca também a aprendizagem da formulação de perguntas que podem ser respondidas com uma coleta de dados, organização e representação destes, conhecimentos já adquiridos no fundamental, e retomados no ensino médio. Nesta etapa, reitera-se, os alunos devem aprimorar as habilidades adquiridas no ensino fundamental no que se refere à coleta, à organização e à representação de dados. Motivacional, sugere-se um trabalho com ênfase na construção e na representação de tabelas e gráficos mais elaborados, analisando sua conveniência e utilizando tecnologias, como o computador. Problemas reais de estatística realísticos sempre começam com uma pergunta, e resultam em uma apresentação de resultados que se apóiam em inferências tomadas de uma amostra para uma população.

Ainda nesta etapa, seguindo os PCNs, os alunos precisam entender sobre o objetivo e a lógica das investigações estatísticas, bem como sobre o processo de investigação. Deve-se permitir aos alunos o entendimento intuitivo e depois, formal, das principais ideias matemáticas das representações estatísticas.

Um exemplo, é propiciar, através de exemplos práticos (onde a própria turma é um grande laboratório) que os estudantes sejam capazes de visualizar como o ponto médio é influenciado por valores das extremidades de um intervalo de dados, e o que acontece com o ponto médio e a mediana em relação a esses extremos.

É importante ressaltar a necessidade de se intensificar a compreensão sobre as medidas de posição (média, moda e mediana) e as medidas de dispersão (desvio médio, variância e desvio padrão), abordadas de forma mais intuitiva no ensino fundamental, e sendo possível, formalizadas no ensino médio.

Os estudantes devem desenvolver o senso crítico na discussão de resultados de pesquisas estatísticas e na avaliação de argumentos probabilísticos que se dizem baseados em alguma informação. A construção de argumentos racionais baseados em informações e observações, veiculando resultados convincentes, exige o apropriado uso de terminologia estatística e probabilística. É com este conhecimento, que os jovens se tornam capazes de questionarem a validade de interpretações de dados, as generalizações feitas com base em um estudo de uma amostra.

O estudo da combinatória e da probabilidade é essencial nesse bloco de conteúdo, pois os alunos precisam adquirir conhecimentos sobre o levantamento de possibilidades e a medida da chance de cada uma delas. A combinatória não tem apenas a função de auxiliar o cálculo das probabilidades, mas tem inter-relação estreita entre as idéias de experimento composto a partir de um espaço amostral discreto e as operações combinatórias.

A utilização do diagrama de árvores é importante para clarear a conexão entre os experimentos compostos e a combinatória. Ao estudar probabilidade e chance, os alunos precisam entender conceitos e palavras relacionadas à chance, incerteza e probabilidade, que aparecem no cotidiano.

Outras ideias importantes incluem a compreensão de que a probabilidade é uma medida de incerteza, que os modelos são úteis para simular eventos, para estimar probabilidades, e que algumas vezes nossas intuições são incorretas e podem nos levar a uma conclusão equivocada no que se refere à probabilidade e à chance. Nos experimentos aleatórios, os alunos precisam aprender a descrevê-los em termos de eventualidades, associá-los a um conjunto de eventos elementares e representá-los de forma esquemática. Os alunos necessitam também dominar a linguagem de eventos, levantar hipóteses de eqüiprobabilidade, associar a estatística dos resultados observados e as freqüências dos eventos correspondentes, e utilizar a estatística de tais freqüências para estimar a probabilidade de um evento dado. Percebe-se nas descrições acima que os conteúdos de Estatística estão bem distribuídos durante todo o ensino básico, com apenas a ressalva, de que a intuição de inferência estatística, não permeia em nenhum momento, ou em poucos momentos, o que é estudado do fundamental e médio.

Segundo as Orientações Curriculares Para o Ensino Médio [2], é sugerido a possibilidade de organizar atividades em que os alunos têm a oportunidade de lidar com as diversas etapas do trabalho de análise de dados reais: tabular, manipular, classificar, obter medidas como média e desvio padrão e obter representações gráficas variadas. Em nenhum momento é sugerido a estimação de estatísticas a partir de parâmetros.

Ainda, segundo estas orientações [2]:

> "Os conteúdos do bloco Análise de dados e probabilidade têm sido recomendados para todos os níveis da educação básica, em especial para o ensino médio. Uma das razões desse ponto de vista reside na importância das idéias de incerteza e de probabilidade, associadas aos chamados fenômenos aleatórios, presentes de forma essencial nos mundos natural e social. O estudo desse bloco de conteúdo possibilita aos alunos ampliarem e formali-

zarem seus conhecimentos sobre o raciocínio combinatório, probabilístico e estatístico"(Orientações Curriculares para o Ensino Médio, 2006, p.78)[2]

Analisando alguns livros adotados em escolas públicas e particulares, nota-se a falta deste elo entre a estatística descritiva e a probabilidade, que culmina na inferência estatística, mas através da leitura de alguns artigos e dissertações na área, percebe-se que o ensino da Estatística e da Probabilidade apresentavam-se sempre interligados nos currículos internacionais, tratados por um termo europeu específico, que é Estocástica. Segundo Lopes, [14], Estocástica é o termo utilizado para tratar a probabilidade como inseparável da estatística. Acredita-se que algumas propostas de atividades concretas, a seguir apresentadas, podem ajudar o aluno a abstrair conceitos indispensáveis ao completo entendimento da inferência estatística, que é o objetivo final de todo o estudo da estatística descritiva e das probabilidades construídos desde o ensino fundamental. Segundo Imenes e Lellis [11], a noção de "possibilidade" e "chance" surgem na vida de maneira natural, forma que é imprecisa e leva a enganos, e só é útil quando quantificada. Desta maneira, os autores propõem em sua coleção, ações práticas que levam os alunos a aprofundadrem a ideia de maneira lúdica e concreta. Algumas destas propostas colocamos em prática nas aulas propostas a seguir.

7 Propostas de Aula para o Ensino Básico

As propostas a seguir foram aplicadas em turmas das seguintes instituições de Goiânia: Colégio Santo Agostinho, Colégio Estadual Damiana da Cunha, Instituto Presbiteriano de Educação e Faculdades Alfa.

7.1 Pense em um país e um animal!

Objetivo:

Levar o aluno a entender a diferença entre o que é determinístico e o que é aleatório.

Descrição da Atividade:

Primeiramente, é escolhido um aluno, que perante a turma, executa alguns comandos dados pelo professor, que neste momento tem seus olhos vendados. Os comandos são os seguintes:

1º) Pense em um número natural de 1 a 10;

2º) Multiplique este número por 9;

3º) Adicione os algarismos formadores do resultado (produto) anterior;

4º) Subtraia 5 do resultado (soma) anterior;

5º) Associe o último resultado a uma letra do alfabeto seguindo o padrão: $1 \longrightarrow A; 2 \longrightarrow B; 3 \longrightarrow C; \ldots; 10 \longrightarrow J; \ldots$

6º) Após a associação do número à letra, é pedido ao aluno que escreva no quadro o nome de um país da Europa que comece com a determinada letra;

7º) Após a escrita do nome do país é pedido ao aluno que escreva no quadro o nome de um animal com a quinta letra do nome do país;

8º) Para finalizar, o professor que esteve com os olhos vendados durante todo o processo, pergunta ao aluno se na Dinamarca existe muito Macaco, onde a probabilidade de acerto é muito alta para o nome do animal e certo para o nome do país.

É importante salientar o aspecto lúdico do processo, que desperta a curiosidade, servindo assim como importante fator motivacional para o despertar do interesse da turma perante o assunto. Neste jogo, a probabilidade de acerto por parte do professor é alta, mas deve ficar claro aos alunos que o professor pode errar. Na pior das hipóteses, considerando que o aluno executou todos os comandos corretamente, o professor acerta pelo menos o país, o que por si já desperta a curiosidade do aluno.

A lógica do jogo é o fato de que até o momento da 5a letra do nome do país, temos um experimento determinístico, e isso é explicado aos alunos da seguinte maneira:

1º) Qualquer número escolhido de 1 a 10, quando multiplicado por 9 resulta em um produto cuja soma de seus algarismos é sempre 9. Isto é um evento **certo**.

2º) Ao subtrair 5 de 9, é **certo** que a diferença será 4, e isto também é **certo**, assim como a letra associada ao 4 é sempre a letra D, o que também é **certo**.

3º) Só existe um país da Europa com a letra **D**, que é a **Dinamarca**, cuja 5ª letra é sempre **M**! Até neste momento, mostramos ao aluno que o experimento é **determinístico**, pois não haveria outro resultado.

4º) Para terminar, perguntamos a turma qual animal cada um pensou, e percebemos que a maioria pensa em macaco, embora alguns alunos sugerem outros animais. Neste momento é que se explica que nesta parte do jogo, o experimento é aleatório, pois existem outras respostas. Neste momento introduzimos de forma intuitiva o conceito de probabilidade, através de perguntas relacionadas a escolha do macaco pela maioria.

Resultados da Aula:

Nas oito turmas aplicadas, o acerto por parte do professor foi de $87,5\%$, considerando que na única turma onde não houve acerto, durante a devolutiva, a maior parte dos alunos assumiram ter pensado em macaco. Outro fator importante, foi o fato de vários alunos terem comentado a tentativa de repetirem a experiência em casa com os pais, o que comprovou a função motivacional norteadora do jogo.

Turmas participantes da experiência:

- 8ºA e 8ºC do Colégio Santo Agostinho;
- 7ºA, 7ºB, 8ºA e 8ºB do Colégio Estadual Damianna da Cunha;
- 9ºA e 9ºC do Instituto Presbiteriano de Educação.

7.2 Jogo com Dado I - Jogo da Soma

Objetivo:

Introduzir o conceito de probabilidade.

Descrição da Atividade:

Na turma, forme 11 times, cada um com um número natural correspondente, de 2 a 12. O professor, ou os próprios alunos de forma alternada, lançam dois dados simultaneamente. Se a soma dos pontos obtidos for 4, ponto para o time 4, se a soma for 9, ponto para o time 9, e assim por diante.

É importante que antes do jogo começar, seja instigado aos alunos quem vai ganhar, e o professor aqui, pode arriscar a "adivinhar"quem vai ganhar, aguçando a curiosidades dos alunos caso acerte, ou pelo menos se aproxime do resultado.

No quadro, é anotado o número de pontos de cada time. Após 50 lançamentos, acaba o jogo e ganha o time com maior número de pontos.

Neste momento, discute-se com os alunos a questão da sorte de quem venceu, com algumas perguintas como:

1º) Antes de iniciar o jogo, alguém percebeu que alguns times tinham mais chances que outros?

2º) A vitória só pode ser associada à sorte?

3º) Qual é a probabilidade de cada soma?

Resultados da Aula:

Na turma aplicada e registrada, o time vencedor foi o time 6, seguido pelo time 7, o qual o professor havia sugerido. Foram realizados 109 lançamentos, onde o time 6 fez 20 pontos e o time 7 fez 19. O time 2, como poderia prever, não fez pontos. Comparando os resultados, o time 7 fez $17,43\%$, muito próximo dos $16,67\%$ previstos pela probabilidade. Como era de se esperar, a curiosidade dos alunos foi despertada em relação ao "acerto"do professor, que neste momento definiu o que é probabilidade, de forma adequada a faixa etária trabalhada.

Turmas participantes da experiência:

- 8ºA do Colégio Santo Agostinho;

7.3 Jogo com Dado II - Jogo do Par ou Ímpar

Objetivo:

Introduzir o conceito de probabilidade.

Descrição da Atividade:

Separe a turma em dois times. Um time será PAR e o outro time será ÍMPAR.

Lançam-se três dados e multiplicam-se os pontos de cada dado. Se o produto for par, ponto para o time PAR; Se for ímpar, ponto para o time ÍMPAR.

Anotam-se os resultados da partida (a totalidade dos pontos dos times PAR e ÍMPAR)

Após 90 lançamentos, realiza-se algumas discussões com os alunos acerca da probabilidade de ganhar ou perder, seguindo as seguinte pergunta: "O time vencedor teve sorte, ou já era possível prever o resultado?"

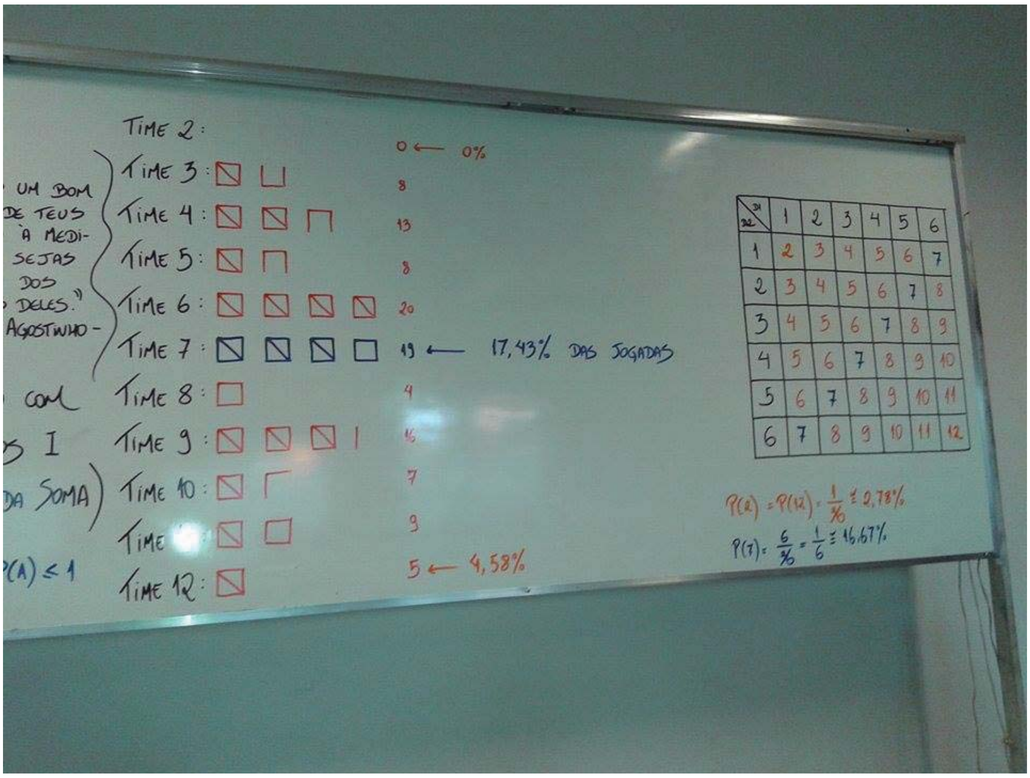


Figura 8: Jogo com dados 1

Para começar, se faz com que os alunos percebam que, dos 6 resultados possíveis, 3 são pares e 3 são ímpares, de modo que, em um dado honesto, a probabilidade de sortear um número par é igual a chance de sortear um número ímpar. O passo seguinte é fazer com que os alunos percebam que quando é lançado três dados, e multiplicados seus pontos, a probabilidade de resultado par não é mais igual a probabilidade de resultado ímpar.

Os estudantes devem visualizar que o produto de três números pares é par, o produto de dois pares e um ímpar é par e o produto de um par e dois ímpares também é par. Somente o produto de três ímpares é ímpar.

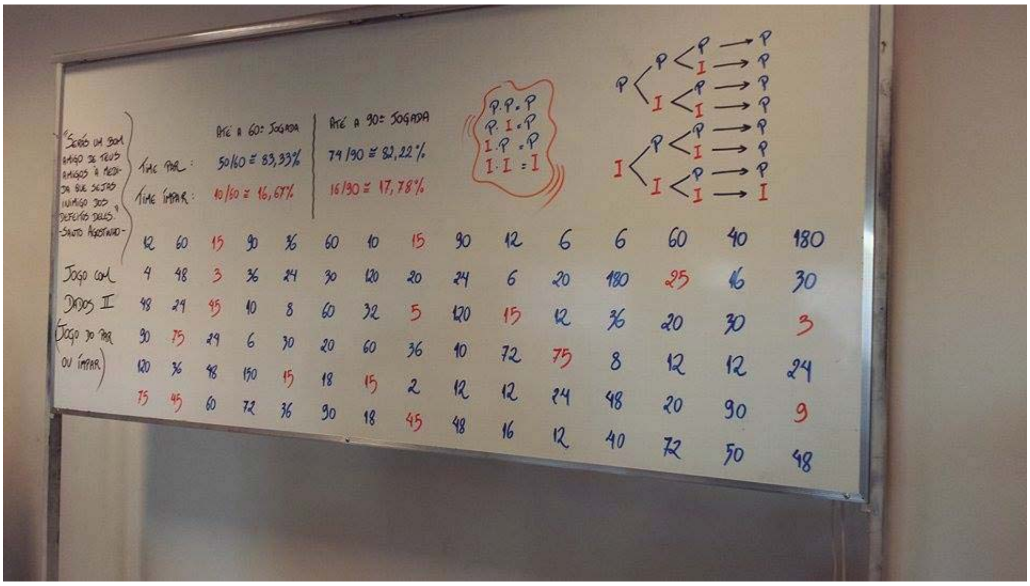


Figura 9: Jogo com dados 2

Neste momento pode-se ensinar ou retomar, o processo da árvore de possibilidades:

Do total de 8 possibilidades, 7 são produtos pares. Assim, nesse jogo, a chance de o produto ser par é $\frac{7}{8}$, e consequentemente, a chance de ser ímpar é de $\frac{1}{8}$. É importante salientar à turma que o ímpar pode ganhar algumas vezes.

Resultados da Aula:

Este jogo foi aplicado na mesma turma que foi aplicado o jogo da soma. Neste jogo, o time vencedor, como era de se esperar, foi o time PAR, com $82,22\%$ de acertos após 90 jogadas. Comparando os resultados com a probabilidade calculada a priori, tendo em vista que os alunos já haviam aprendido sobre probabilidade, o resultado foi bem próximo dos $87,50\%$ previstos pelo cálculo da probabilidade.

Algo que era de se esperar e não aconteceu no tempo pensado, era que parte da turma percebesse, através do que já havíamos estudado, a tendência do time par ganhar, o que não aconteceu antes do término do jogo e da explicação sobre a paridade dos produtos.

Turmas participantes da experiência:

$$
Par \left\{ \begin{array}{ll}
Par & \left\{ \begin{array}{l} Par \longrightarrow Par \\ Impar \longrightarrow Par \end{array} \right. \\[2em]
Impar & \left\{ \begin{array}{l} Par \longrightarrow Par \\ Impar \longrightarrow Par \end{array} \right.
\end{array} \right.
$$

$$
Impar \left\{ \begin{array}{ll}
Par & \left\{ \begin{array}{l} Par \longrightarrow Par \\ Impar \longrightarrow Par \end{array} \right. \\[2em]
Impar & \left\{ \begin{array}{l} Par \longrightarrow Par \\ Impar \longrightarrow Impar \end{array} \right.
\end{array} \right.
$$

Figura 10: Árvore de possibilidades

- $8^{o}A$ do Colégio Santo Agostinho;

7.4 Conclusões a partir de uma amostra

Objetivo:

Introduzir o conceito de Inferência.

Descrição da Atividade:

Peça que a turma forme grupos de três alunos, e qua cada grupo prepare 300 pedaços pequenos de papel, dividindo, por exemplo, uma folha de chamex em colunas e linhas de 3 cm. Desta maneira, com apenas uma folha, já obtem-se aproximadamente 50 pedaços de papel.

Em 200 dos papéis, o grupo escreve A, e nos restantes, o grupo escreve B. Misture bem, e coloque dentro de um saco. Peça que o grupo sorteie 30 papéis, contem a quantidade de A e de B, anotando os resultados e respondendo as seguintes questões:

1º) Nesta simulação, a população é formada pelos 300 papéis. Como se distribui essa população em relação a A e a B?

2º) A amostra sorteada, reflete, com alguma aproximação, o que acontece na população?

3º) Há alguma chance de se sortear uma amostra com 30 papéis A e nenhum B? Em sua opinião esta chance é grande ou pequena?

Depois de cada grupo elaborar suas respostas, compila-se os resultados de todos os grupos. Isso equivale, até um certo ponto, a considerar uma população bem maior, com uma amostra também maior. Se considerar 10 grupos, por exemplo, tem-se uma população de 3000 elementos com uma amostra de 300. Como cada grupo calculou a porcentagem de papéis marcados com A, calcula-se a média desses resultados.

Comparando os resultados, verifica-se que na amostra maior, existe um propabilidade alta desta refletir o que ocorre na população ($\approx 66,7\%$ de A) com maior precisão do que reflete a amostra menor.

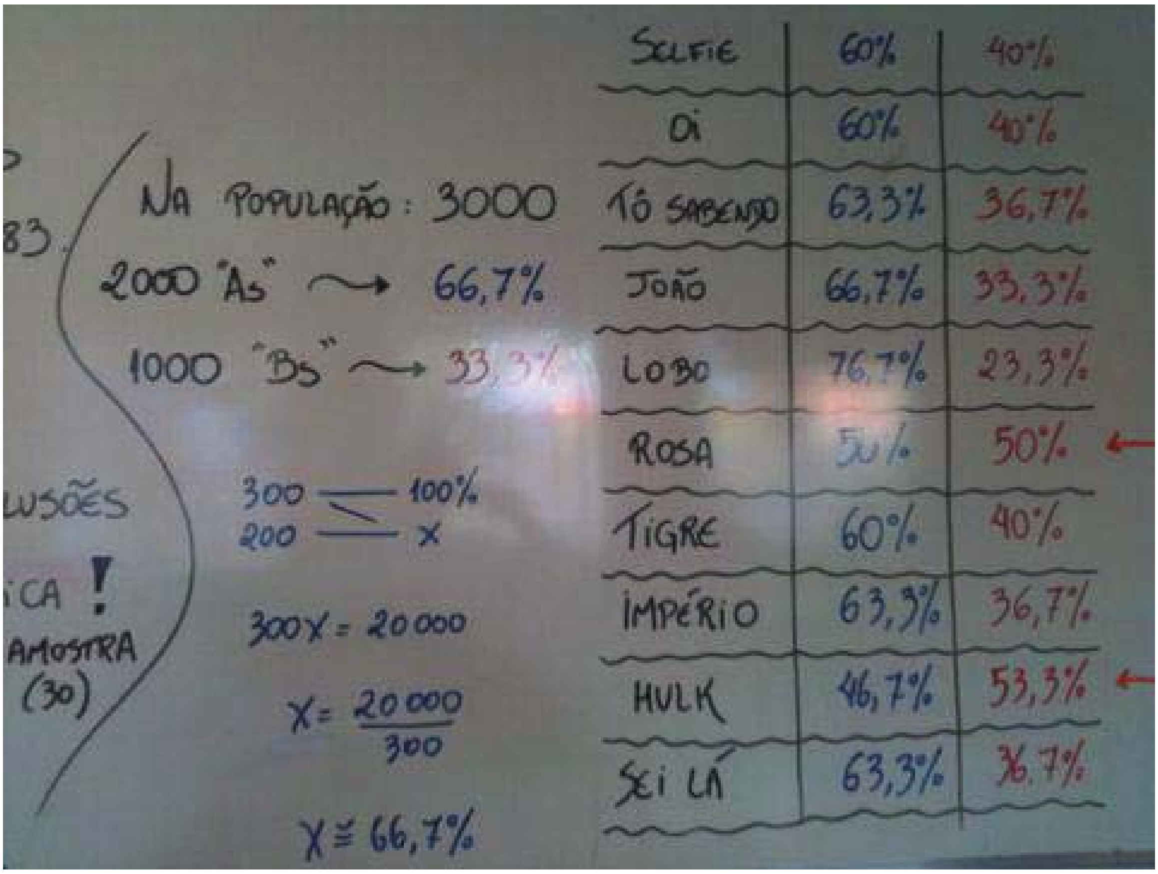


Figura 11: Conclusões a partir de uma amostra

Resultados da Aula:

Na primeira turma, foram formados 11 grupos, nos quais a menor proporção de A foi de 50% e a maior proporção foi de aproximadamente 83,33%. Após compilarmos todos os resultados, ou seja, 330 papéis de 3300, chegamos a proporção de aproximadamente 64%.

Na segunda turma, foram formados 7 grupos, nos quais a menor proporção de A foi de aproximadamente 53,33% e a maior proporção foi de aproximadamente 86,67%. Após compilarmos todos os resultados, ou seja, 330 papéis de 3300, chegamos a proporção de aproximadamente 64%.

Na terceira turma, foram formados 9 grupos, nos quais a menor proporção de A foi de aproximadamente 46,67% e a maior proporção foi de aproximadamente 76,67%. Após compilarmos todos os resultados, ou seja, 270 papéis de 2700, chegamos a proporção de aproximadamente 61,1%.

Nas três turmas, os alunos perceberam que aumentando a amostra, mesmo que isto signifique o aumento proporcional da população, a chance de aproximação foi bem maior.

Turmas participantes da experiência:

- Turmas 1301 e 3301 do 2º período de administração das Faculdades Alfa. São alunos que já estudaram Estatística Descritiva, e vão começar a estudar a parte de inferência. Esta aula foi introdutória.
- 8º A do Colégio Santo Agostinho;

7.5 Quantos peixes tem uma lagoa?

Objetivo:

Introduzir os conceitos de Amostra, População e Estimação de Parâmetros Populacionais.

Descrição da Atividade:

A atividade proposta está baseada numa oficina idealizada por Cordani (2006)[4].

Esta se baseia na estimação do total de uma população, no caso estimação de peixes em uma lagoa com o objetivo de trazer algo mais concreto e lúdico para sala de aula.

A oficina aplica o método de captura e recaptura, comum na área das ciências biológicas, para estimação da população. A partir deste experimento espera-se despertar os alunos para os conceitos de amostra, população e estimação de parâmetros populacionais.

A proposta de aprendizagem apresentada nessa pesquisa é de estimação de uma população utilizando o método de captura-recaptura. Esse método é bastante utilizado nas áreas de Ecologia e Epidemiologia. O método de captura-recaptura foi desenvolvido e utilizado por Laplace para estimar a população da França. A maior parte das aplicações do método de captura-recaptura diz respeito a inferência sobre o tamanho de populações animais, embora este passou a ser utilizado em outros estudos.

Os alunos envolvidos correspondem a uma turma de oitavo ano do ensino fundamental, do Colégio Santo Agostinho de Goiânia.

Nesta atividade, foram utilizados palitos de picolé representando os "peixes", que foram marcados com caneta hidrográfica após a "captura". No dia da aplicação estavam presentes 34 alunos. Foi mostrado aos alunos os palitos em uma sacola transparente e pedido que os mesmos sugerissem a quantidade de "peixes", ou seja, de palitos na sacola. Algumas respostas foram anotadas como mais extremas, cem palitos e 2000 palitos. Apenas o professor sabia a quantidade, que era de 400 palitos.

Após colocar todos os palitos em uma sacola escura (não transparente), foi pedido que cada aluno retirasse um palito do saco, e marcasse o "peixe" com canetinha preta, e depois de todos os 34 peixes serem marcados, os mesmos foram devolvidos à sacola preta. Este primeiro processo ficou definido como a "captura".

Depois desta fase, cada aluno retirou novamente um "peixe", de modo que foram anotados quantos marcados foram recapturados dentre os 34, que nesta primeira recaptura, foram 3, na segunda 2, na terceira 3, na quarta 2 e na quinta 3 novamente.

Todos os dados foram marcados e através do princípio fundamental da proporção, estimamos quantos "peixes"haviam na "lagoa":

$$\frac{34}{x} = \frac{3}{34}$$
$$3x = 1156$$

$x \approx 385$ para a 1ª recaptura.

O processo de recaptura foi repetido algumas vezes, dando as seguintes estimativas:

$$\frac{34}{x} = \frac{2}{34}$$
$$2x = 1156$$

$x \approx 578$ para a 2ª recaptura.

$$\frac{34}{x} = \frac{3}{34}$$
$$3x = 1156$$

$x \approx 385$ para a 3ª recaptura.

$$\frac{34}{x} = \frac{2}{34}$$
$$2x = 1156$$

$x \approx 578$ para a 4ª recaptura.

$$\frac{34}{x} = \frac{3}{34}$$
$$3x = 1156$$

$x \approx 385$ para a 5ª recaptura.

Após as cinco recapturas, calculamos a média desses resultados:

$$\overline{x} = \frac{385 + 578 + 385 + 578 + 385}{5} \approx 462$$

que é uma boa **estimativa** do total de "peixes"na lagoa.

Resultados da Aula:

Os alunos perceberam a aproximação do valor, após ser pedido que alguns alunos contassem juntos o total de palitos. Houve também a discussão sobre a quantidade de recapturas, explicando aos alunos que quanto maior a quantidade de recapturas, mais aproximaríamos do total de peixes.

Turma participante da experiência:

- $8^{o}A$ do Colégio Santo Agostinho;

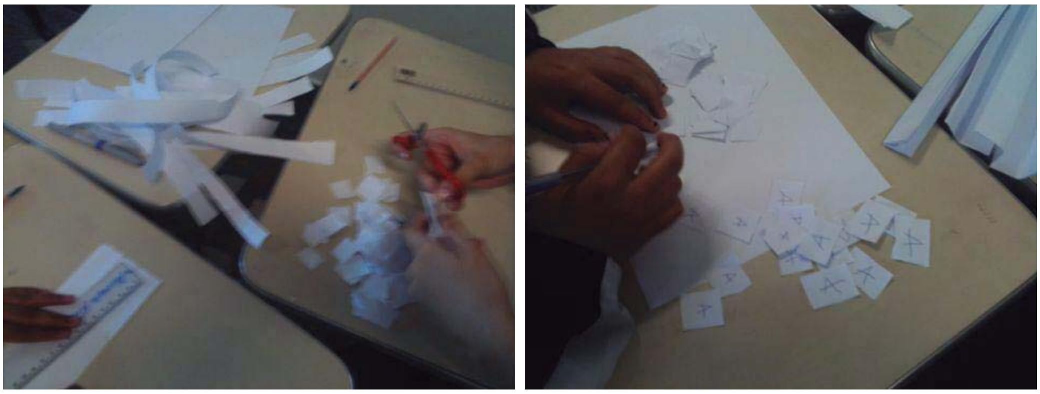


Figura 12: Quantos peixes tem na lagoa?

Figura 13: Quantos peixes tem na lagoa? (II)

8 Considerações finais

O que verificamos neste estudo, foi o quanto a Estatística é um campo da matemática muito mais amplo do que é apresentado no ensino básico.

Observamos, que tendo como referência a grandiosidade desta ciência, muito pouco é trabalhado no ensino basico, e dada a importância desta nos mais diversos campos do conhecimento técnico e científico, torna-se necessário a ampliação do que é estudado nas séries que antecedem as graduações.

No presente, os alunos estão saindo do Ensino Médio com uma quantidade imensa de conteúdos, deixando de lado, conceitos, mesmo que intuitivos, muito mais necessários e significativos do que outros, sem menosprezar a importância de cada um.

Em certa aula ministrada no IMPA, o professor Morgado, comentou sobre o absurdo que acontece no Brasil, onde um aluno depois de 11 anos de escola, entra na faculdade tendo aprendido sobre números complexos, mas não sabe com segurança, escolher entre uma compra a prazo, ou um desconto à vista. Certo de que o professor tratava da importância da matemática financeira, acrescenta-se aqui a Estatística. Os alunos chegam na Universidade sem noções básicas e importantes da Estatística, muitas vezes até mais importantes que outros assuntos menos relevantes.

Concluímos neste estudo, que muito mais pode ser trabalhado no ensino básico, inclusive pontos da inferência estatística quase nunca trabalhados nesta fase do ensino, e o mais importante, de maneira lúdica, motivacional e significativa para o aluno.

Referências

[1] ARA, A.B.; MUSETTI, A.V.; SCHNEIDERMAN,B.; ***Introdução à Estatística*** - 1.ed. - São Paulo; Edgard Blucher; 2003.

[2] BRASIL; Ciências da natureza, matemática e suas tecnologias / Secretaria de Educação Básica; ***Orientações Curriculares para o Ensino Médio; Volume 2*** - Brasília; Ministério da Educação; Secretaria de Educação Básica; 2006; 135p.

[3] BUSSAB, W.O.; MORETTIN, P.A.; ***Estatística Básica*** - 7.ed. - São Paulo; Saraiva; 2012.

[4] CORDANI, L.K.; ***Oficina: Estatística para todos*** ; Disponível em *http : //www.redeabe.ime.unicamp.br/pt$_B$R/files/CEduc/OficinaSite.pdf*; Acesso em: 11 de Maio de 2014.

[5] COSTA, S.F.; ***Introdução Ilustrada à Estatística*** - 5.ed. - São Paulo; Harbra; 2013.

[6] CRESPO, A.A.; ***Estatística Fácil*** - 11.ed. - São Paulo; Saraiva; 1994.

[7] DA COSTA, N.C.A.; ***Lógica indutiva e probabilidade*** - São Paulo; 1981.

[8] DANTE, L.R.; ***Matemática(Contexto e Aplicações)*** - 1.ed. - São Paulo; Ática; 2003.

[9] DOWNING, D.; CLARK, J.; ***Estatística Aplicada*** - 2.ed. - São Paulo; Saraiva; 2005.

[10] IEZZI, G.; DOLCE, O.; DEGENSZAJN, D.; PÉRIGLO, R.; ***Matemática (Ensino Médio)*** - 5.ed. - São Paulo; Atual; 2011.

[11] IMENES, L.M.; LELLIS, M.; ***Matemática*** - 1.ed. - São Paulo; Moderna; 2010.

[12] LEBENSZTAYN, E.; COLETTI, C.F.; ***NOTAS DE AULA - Probabilidade: Teoria e Exercícios*** - 1.ed. - São Paulo; USP; 2008.

[13] LEVINE, D.M.; STEPHAN, D.F.; KREHBIEL, T.C.; BERENSON, M.L.; ***Estatística - Teoria e Aplicações*** - 6.ed. - Rio de Janeiro; LTC; 2012.

[14] LIMA, E.L.; CARVALHO, P.C.; WAGNER, E.; MORGADO, A.C.; ***Temas e Problemas*** - 1.ed. - Rio de Janeiro; SBM; 2001.

[15] LOPES, C.A.E.; ***A Probabilidade e a Estatística no Ensino Fundamental: uma análise curricular.***; Campinas, SP; Faculdade de Educação da UNICAMP; 1998. 125p. (Dissertação, Mestrado em Educação).

[16] PARDAL, P.; *Primórdios do Ensino de Estatística no Brasil e na UERJ*, ***Revista do Instituto Histórico e Geográfico Brasileiro***, Rio de Janeiro; 154(378); 1-152; 1993.

[17] MEYER, P.L.; ***Probabilidade - Aplicações à Estatística*** - 2.ed. - [Reimp.] - Rio de Janeiro; LTC; 2011.

[18] SPIEGEL,M.R.;SCHILLER,J.;SRINIVASAN,A.,***Probabilidade e Estatística*** - 2.ed. - Porto Alegre; Bookman; 2004.