

**UNIVERSIDADE FEDERAL DO ABC**

**CENTRO DE MATEMÁTICA, COMPUTAÇÃO E COGNIÇÃO**

**MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL - PROFMAT**

**ERICH STANO VALENÇA**

**LEIS DE POTÊNCIA**

**Santo André - SP,  
2016**





**ERICH STANO VALENÇA**

## **LEIS DE POTÊNCIA**

Dissertação apresentada ao Centro de Matemática, Computação e Cognição da Universidade Federal do ABC, como requisito parcial à obtenção do título de Mestre.

**Orientador: Prof. Dr. Rafael de Mattos Grisi**

**SANTO ANDRÉ,  
2016**

Sistema de Bibliotecas da Universidade Federal do ABC  
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC  
com os dados fornecidos pelo(a) autor(a).

Valença, Erich Stano  
Leis de Potência / Erich Stano Valença. — 2016.

159 fls.

Orientador: Rafael de Mattos Grisi

Dissertação (Mestrado) — Universidade Federal do ABC, Mestrado  
Profissional em Matemática em Rede Nacional - PROFMAT,  
Santo André, 2016.

1. Leis de Potência. 2. Estatística. 3. Probabilidade. 4. Regressão  
Linear. 5. Ensino Médio. I. Grisi, Rafael de Mattos. II. Mestrado  
Profissional em Matemática em Rede Nacional - PROFMAT,  
2016. III. Título.



**MINISTÉRIO DA EDUCAÇÃO**  
**Fundação Universidade Federal do ABC**  
**Programa de Pós-Graduação em Mestrado Profissional em Matemática**  
**em Rede Nacional**

Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP  
CEP 09210-580 · Fone: (11) 4996-0017  
profmat@ufabc.edu.br

**FOLHA DE ASSINATURAS**

Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Erich Stano Valença, realizada em 2 de março de 2016:

Prof.(a) Dr.(a) **Rafael de Mattos Grisi** (UFABC) – Presidente

Prof.(a) Dr.(a) **Marcus Antônio Mendonça Marrocos** (UFABC) – Membro Titular

Prof.(a) Dr.(a) **Gleiciane da Silva Aragão** (UNIFESP) – Membro Titular

Prof.(a) Dr.(a) **Ana Carolina Boero** (UFABC) – Membro Suplente

Prof.(a) Dr.(a) **Christophe Frédéric Gallesco** (UNICAMP) – Membro Suplente



**Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca no dia da defesa, sob responsabilidade única do autor e com a anuência de seu orientador.**

**Santo André, \_\_\_\_ de \_\_\_\_\_ de 20\_\_\_\_.**

**Assinatura do autor: \_\_\_\_\_**

**Assinatura do orientador: \_\_\_\_\_**





---

Dedico este trabalho à minha esposa Amanda, sem seu auxílio e compreensão jamais esta dissertação poderia ter sido concluída; aos meus filhos Ignácio, Henrique, José Luís, Guilherme e Fernando.



---

## AGRADECIMENTOS

---

Ao orientador deste trabalho, Prof. Dr. Rafael de Mattos Grisi, pela contribuição imprescindível, sobretudo nos pontos de maior dificuldade.

Aos membros da banca, Profa. Dra. Gleiciane da Silva Aragão e Prof. Dr. Marcus Antônio Mendonça Marrocos, pela leitura cuidadosa, atenta e pelas ótimas correções e sugestões realizadas.

Aos meus pais, Ademir e Mary, pelo apoio que sempre me deram em tudo, inclusive em relação aos estudos, propiciando as condições necessárias para que boa parte da minha vida acadêmica fosse possível.

Ao saudoso professor Orlando, homem de grande virtude e conhecimento. Seu caráter, ensinamentos e exemplo ainda ecoam. A segunda citação de minha epígrafe certamente a ele se aplica.

Aos colegas do PROFMAT, pela convivência e pelo auxílio.

Aos funcionários da secretaria acadêmica, em especial ao Thiago de Lima, sempre muito atencioso e solícito em esclarecer todas as dúvidas burocráticas e em resolver todos os problemas que surgiam.



---

*“Quando a razão percorre o céu e a terra, descobre que nada lhe agrada fora da beleza; e na beleza, as figuras; nas figuras as dimensões; nas dimensões os números.”*

(Santo Agostinho, "De Ordine", II, XV, 42)

*“Esta disciplina [matemática] é a mesma lei de Deus que, sempre permanecendo fixa e inabalável Nele, de certo modo se imprime nas almas dos sábios, de modo que tanto melhor sabem viver e com tanta maior elevação quanto mais perfeitamente a contemplam com sua inteligência e a guardam com sua vida.”*

(Santo Agostinho, "De Ordine", II, VIII, 25)



---

## RESUMO

---

O objetivo deste trabalho é apresentar uma abordagem das funções de probabilidade conhecidas como Leis de Potência para alunos do Ensino Médio. Para atingir este objetivo, foi necessário desenvolver outros temas. Apresentamos algumas noções básicas de estatística e probabilidade, incluindo funções de probabilidade. Obtivemos fórmulas para o cálculo da área sob certas curvas não retilíneas, sem o uso do cálculo integral. Tratamos de Regressões Lineares Simples, sem o uso de derivadas. Desenvolvemos o tema Leis de Potência do ponto de vista teórico e prático; quanto a isto, nos inspiramos em Newman (2006) [3]. Foram apresentadas, por fim, uma série de atividades, com explicações detalhadas, envolvendo Leis de Potência.

**Palavras-chave:** Leis de Potência, Estatística, Probabilidade, Regressão Linear, Ensino Médio.





---

## ABSTRACT

---

The aim of this paper is to show an approach to probability functions known as Power Law for High School students. To achieve this goal, it was necessary to study other topics. We will show some basic information on statistics and probability, including probability functions. We've got formulas to calculate the area under certain non-rectilinear curves, without using of the integral (calculus). We studied Simple Linear Regressions, without using derivatives. We talked about Power Laws under the theoretical and practical point of view; for that, we were inspired by Newman (2006 ). We presented, finally, a series of activities, with detailed explanations, involving Power Laws.

**Keywords:** Power Laws, Statistics, Probability, Linear Regression, High School.



---

# CONTEÚDO

---

Lista de Figuras	19
Lista de Tabelas	25
Introdução	27
1 NOÇÕES DE ESTATÍSTICA E PROBABILIDADE	29
1.1 O que é a Estatística	29
1.2 Tipos de variáveis	30
1.3 Organização dos dados	31
1.4 Medidas-Resumo	36
1.5 Probabilidade	38
1.6 Variáveis Aleatórias Discretas	40
1.6.1 Modelo Uniforme Discreto	41
1.6.2 Valor Médio e Variância para Variáveis Aleatórias Discretas	42
1.6.3 Distribuição Binomial	43
1.6.4 Distribuição Geométrica	44
1.7 Variáveis Aleatórias Contínuas	48
2 ÁREA SOB A CURVA	57
2.1 Área sob a curva - 1º Caso: $f(x) = x^m$ , com $m \geq 0$	60
2.1.1 1º Caso com $f(x) = x^0$ , $x \geq 0$	60
2.1.2 1º Caso com $f(x) = x^1$ , $x \geq 0$	61
2.1.3 1º Caso com $f(x) = x^2$ , $x \geq 0$	61
2.1.4 1º Caso com $f(x) = x^3$ , $x > 0$	67
2.2 2º Caso: $f(x) = x^{-m}$ , $x > 0$ , com $m > 1$	69
2.2.1 2º Caso com $f(x) = x^{-2}$ , $x > 0$	70
2.2.2 2º Caso com $f(x) = x^{-3}$ , $x > 0$	74
2.2.3 2º Caso com $f(x) = cx^{-m}$ , $x > 0$ , $c > 0$ e $m > 1$	79
2.3 3º Caso: $f(x) = x^{-1}$ , $x > 0$	79
2.4 Possibilidade da utilização de $f(x) = c.x^m$ , $x > 0$ , $c > 0$ , como função de densidade de probabilidade	86

## Conteúdo

3	REGRESSÃO LINEAR SIMPLES	89
3.1	Reta que passa por dois pontos distintos . . . . .	90
3.2	Reta a partir de três, ou mais, pontos distintos . . . . .	90
3.3	Fórmulas para Regressão Linear Simples . . . . .	93
4	LEIS DE POTÊNCIA	101
4.1	Caso das cidades brasileiras . . . . .	102
4.2	Frequência Acumulada . . . . .	110
4.3	Exemplos de Lei de Potência . . . . .	112
4.4	Valor médio e Variância . . . . .	117
4.5	Distribuições de cauda leve e Cauda pesada . . . . .	122
5	ATIVIDADES COM LEIS DE POTÊNCIA	125
5.1	Objetivo das três primeiras atividades . . . . .	125
5.2	Características dos dados das três primeiras atividades . . . . .	125
5.3	Fonte dos dados das três primeiras atividades e procedimentos iniciais .	126
5.4	1º Atividade: Populações das cidades agrupadas em histogramas com amplitudes iguais (20.000). . . . .	126
5.5	2º Atividade: Populações das cidades agrupadas em histogramas com amplitudes diferentes (Binning logarítmico). . . . .	131
5.6	3º Atividade: Distribuição das populações das cidades por meio da frequência acumulada. . . . .	136
5.7	4º Atividade: Frequência das palavras utilizadas no livro Moby Dick de Herman Melville. . . . .	139
	Considerações Finais	145
A	APÊNDICE	147
A.1	Área sob a curva de $f(x) = x^m$ , $x \geq 0$ , com $m \geq 0$ . . . . .	147
A.2	Área sob a curva de $f(x) = cx^{-m}$ , para $m > 1$ e $c > 0$ . . . . .	153
	Bibliografia	159

---

## LISTA DE FIGURAS

---

Figura 1	Tipos de variáveis. . . . .	31
Figura 2	Gráfico de barras. . . . .	33
Figura 3	Gráfico de dispersão. . . . .	34
Figura 4	Histograma dos salários dos funcionários de uma empresa . . .	35
Figura 5	Probabilidades na distribuição uniforme contínua. Reprodução de Morettin e Bussab (2013, p.170), [2]. . . . .	49
Figura 6	Gráfico da função real $f(x) = 2x$ . . . . .	51
Figura 7	Região de pouso do aeroplano. . . . .	52
Figura 8	Disco circular sobre a tábua graduada. . . . .	54
Figura 9	Triângulo equilátero inscrito em um círculo de raio 1. . . . .	54
Figura 10	Corda AB com distância para o centro do círculo inferior a $\frac{1}{2}$ . . .	54
Figura 11	Corda AB com distância para o centro do círculo superior a $\frac{1}{2}$ . . .	55
Figura 12	Área sob a curva de $f(x) = x^{-\alpha}$ cercada por retângulos inferiores e superiores. . . . .	58
Figura 13	Área sob a curva de $f(x) = x^2$ no intervalo $[s, t]$ . . . . .	59
Figura 14	Área sob a curva de $f(x) = x^2$ nos intervalos $[s, r]$ , $[r, t]$ , $[s, t]$ . . .	59
Figura 15	Área sob a curva de $f(x) = x^0$ em $[0, t]$ . . . . .	60
Figura 16	Área sob a curva de $f(x) = x$ em $[0, t]$ . . . . .	61
Figura 17	Área sob a curva de $f(x) = x^2$ em $[0, t]$ . . . . .	61
Figura 18	Área sob a curva de $f(x) = x^2$ aproximada por $n$ retângulos. . .	62
Figura 19	Área sob a curva em $[0, t]$ cercada por 5 retângulos superiores. .	63
Figura 20	Área sob a curva em $[0, t]$ cercada por 10 retângulos superiores. .	63
Figura 21	Área sob a curva em $[0, t]$ cercada por 20 retângulos superiores. .	64

LISTA DE FIGURAS

Figura 22	Área sob a curva em $[0, t]$ cercada por 40 retângulos superiores.	64
Figura 23	Área sob a curva em $[0, t]$ cercada por 80 retângulos superiores.	64
Figura 24	Área sob a curva em $[0, t]$ cercada por 160 retângulos superiores.	65
Figura 25	Área sob a curva de $f(x) = x^3$ , em $[0, t]$ , cercada por $n$ retângulos superiores. . . . .	67
Figura 26	Área sob a curva de $f(x) = \frac{1}{x^2}$ , em $[a, b]$ . . . . .	71
Figura 27	Área sob a curva de $f(x) = \frac{1}{x^2}$ , em $[a, b]$ , cercada por dois retângulos. . . . .	71
Figura 28	Área sob a curva de $f(x) = \frac{1}{x^2}$ , em $[a, b]$ , cercada por retângulos inferiores. . . . .	72
Figura 29	Área sob a curva de $f(x) = \frac{1}{x^2}$ , em $[a, b]$ , cercada por retângulos superiores. . . . .	72
Figura 30	Área sob a curva de $f(x) = \frac{1}{x^3}$ , em $[a, b]$ . . . . .	75
Figura 31	Área sob a curva de $f(x) = \frac{1}{x^3}$ , em $[a, b]$ , cercada por dois retângulos. . . . .	75
Figura 32	Área sob a curva de $f(x) = \frac{1}{x^3}$ , em $[a, b]$ , cercada por retângulos inferiores. . . . .	76
Figura 33	Área sob a curva de $f(x) = \frac{1}{x^3}$ , em $[a, b]$ , cercada por retângulos superiores. . . . .	76
Figura 34	Área sob a curva de $f(x) = \frac{1}{x}$ , em $[a, b]$ . . . . .	80
Figura 35	Áreas sob a curva de $f(x) = \frac{1}{x}$ , em $[a, b]$ e $[ak, bk]$ . . . . .	80
Figura 36	Áreas das regiões sob a curva de $f(x) = \frac{1}{x}$ , em $[a, b]$ e $[a \cdot k, b \cdot k]$ . São destacados dois retângulos no interior de ambas as regiões.	81
Figura 37	Áreas das regiões sob a curva de $f(x) = \frac{1}{x}$ , em $[a, b]$ e $[ak, bk]$ cercadas, ambas, por $n$ retângulos. . . . .	82
Figura 38	Áreas das regiões sob a curva de $f(x) = \frac{1}{x}$ , em $[1, a]$ , $[1, b]$ , $[1, ab]$ , $[a, ab]$ . . . . .	83
Figura 39	Para qualquer $x$ real, é sempre possível $k$ inteiro tal que $2^k \leq x < 2^{k+1}$ . . . . .	84

Figura 40	Área sob a região da curva de $f(x) = \frac{1}{x}$ no intervalo $[1, 2^k]$ . . . . .	85
Figura 41	Área sob a região da curva de $f(x) = \frac{1}{x}$ no intervalo $[1, x]$ . . . . .	85
Figura 42	Área sob a região da curva de $f(x) = \frac{1}{x}$ no intervalo $[1, 2^{k+1}]$ . . . . .	86
Figura 43	Três pontos não alinhados. . . . .	91
Figura 44	Três pontos não alinhados e a linha de tendência. . . . .	93
Figura 45	Oito pontos não alinhados. . . . .	94
Figura 46	Reta ajustada aos 8 pontos não alinhados. . . . .	99
Figura 47	Nº de cidades X População. . . . .	103
Figura 48	Logaritmo do Nº de Cidades X População. . . . .	105
Figura 49	Log do Nº de cidades X Log da População. . . . .	106
Figura 50	Log do Nº de cidades X Log da População, e a reta ajustada por Regressão Linear. . . . .	107
Figura 51	Log do Nº de cidades X Log da População. Incluindo todas as populações acima de 10.000. . . . .	109
Figura 52	Frequência relativa (probabilidade) acumulada das Populações X População. Em escala log-log. . . . .	111
Figura 53	Frequência absoluta (quantidade) acumulada das Populações X População. Em escala log-log. . . . .	112
Figura 54	Gráficos dos exemplos de (a) a (l). Extraído de Newman (2006, p.6), [3]. . . . .	116
Figura 55	Área do retângulo de base com medida $d - c$ e altura $f(x_i)$ . . . . .	118
Figura 56	Divisão do intervalo $[a, b]$ em $n$ intervalos iguais. . . . .	118
Figura 57	$n$ retângulos em torno à curva $f(x) = cx^{-\alpha}$ , no intervalo $[a, b]$ . . . . .	119
Figura 58	Etapa 1 da 1º atividade. . . . .	127
Figura 59	Etapa 2 da 1º atividade. . . . .	127
Figura 60	Etapa 3 da 1º atividade. . . . .	128
Figura 61	Etapa 4 da 1º atividade. . . . .	128
Figura 62	Gráfico incluindo os blocos de até 490.000. . . . .	129

LISTA DE FIGURAS

Figura 63	Gráfico incluindo todos os blocos. . . . .	129
Figura 64	Gráfico da densidade de frequência em escala log-log. . . . .	130
Figura 65	Etapa 1 da 2 <sup>o</sup> atividade. . . . .	133
Figura 66	Etapas de (2) a (6) da 2 <sup>o</sup> atividade. . . . .	133
Figura 67	Gráfico da densidade de frequência em escala log-log. Foi utilizada a técnica do binning logarítmico. . . . .	134
Figura 68	Etapa 7 da 2 <sup>o</sup> atividade. . . . .	134
Figura 69	Gráfico da 2 <sup>o</sup> atividade para obtenção do expoente da lei de potência. . . . .	135
Figura 70	Etapa 1 da 3 <sup>o</sup> atividade. . . . .	136
Figura 71	Etapa 2 da 3 <sup>o</sup> atividade. . . . .	137
Figura 72	Etapa 3 da 3 <sup>o</sup> atividade. . . . .	137
Figura 73	Etapa 4 da 3 <sup>o</sup> atividade. . . . .	138
Figura 74	Frequências de ocorrências das palavras no livro Moby Dick, de Herman Melville. . . . .	140
Figura 75	Etapa 2 da 4 <sup>o</sup> atividade. . . . .	140
Figura 76	Etapa 3 da 4 <sup>o</sup> atividade. . . . .	141
Figura 77	Etapa 4 da 4 <sup>o</sup> atividade. . . . .	141
Figura 78	Etapa 5 da 4 <sup>o</sup> atividade. . . . .	142
Figura 79	Log do número de palavras com ocorrência $\geq x$ pelo log do número de ocorrências. . . . .	142
Figura 80	Log do número de palavras com ocorrência $\geq x$ pelo log do número de ocorrências. Uma reta ajustada aos pontos foi acrescentada ao gráfico. . . . .	143
Figura 81	Área da região sob a curva de $f(x) = x^m, m \geq 0$ , em $[0, t]$ . . . . .	147
Figura 82	Área da região sob a curva de $f(x) = x^m, m \geq 0$ , em $[0, t]$ , cercada por $n$ retângulos. . . . .	148
Figura 83	Área da região sob a curva de $f(x) = x^m, m \geq 0$ , em $[0, t]$ , cercada por 5 retângulos. . . . .	149



Figura 84 Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 10 retângulos. . . . . 149

Figura 85 Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 20 retângulos. . . . . 150

Figura 86 Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 40 retângulos. . . . . 150

Figura 87 Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 80 retângulos. . . . . 151

Figura 88 Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 160 retângulos. . . . . 151

Figura 89 Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ . . . 153

Figura 90 Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ , cercada por dois retângulos. . . . . 154

Figura 91 Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ , cercada por retângulos inferiores. . . . . 155

Figura 92 Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ , cercada por retângulos superiores. . . . . 155



---

## LISTA DE TABELAS

---

Tabela 1	Frequência do número de filhos por casal. . . . .	32
Tabela 2	Variável número de filhos por casal e frequências. . . . .	32
Tabela 3	Salários dos funcionários de uma empresa. . . . .	35
Tabela 4	Três pontos não alinhados. . . . .	91
Tabela 5	Oito pontos não alinhados. . . . .	94
Tabela 6	Parâmetros para posterior aplicação das fórmulas. . . . .	98



---

## INTRODUÇÃO

---

Este é um trabalho na área de Educação Matemática. Certamente, o ensino de matemática está entre as coisas mais criticadas em nosso sistema escolar. As críticas, muitas vezes, são pertinentes, mas cabe a pergunta: qual é a disciplina que goza de aprendizado sólido por parte dos nossos alunos?

Esta pergunta, talvez um tanto provocativa, não é gratuita. Pensamos que certos problemas que em outras disciplinas podem ficar um pouco mais camuflados, em matemática se evidenciam. Isto se dá, pela maneira como o raciocínio é permanentemente exigido nesta matéria, a começar pela simples interpretação de problemas, tanto mais pela compreensão das fórmulas ou pela capacidade de relacionar as coisas.

Capacidade de relacionar as coisas! Os próprios conteúdos, dentro do ensino da matemática, poucas vezes são relacionados. O que é que área tem a ver com função? Função com probabilidade? Há relação entre progressão geométrica e probabilidade? E entre estatística e probabilidade?

Essas relações, dentre outras, são exploradas, não exaustivamente, nesta dissertação. O objetivo inicial era discutir um tipo de função de probabilidade conhecida como Lei de Potência. Ao perseguir este objetivo, todavia, muitas coisas apareceram no caminho - Como tratar de área sob a curva com alunos que não conhecem o cálculo integral?, por exemplo - e foram se agregando, e servindo de alicerce, ao tema original.

É neste contexto que o presente trabalho deve ser encarado. Cada capítulo tem uma certa autonomia perante os demais. O que dá unidade aos assuntos, às vezes aparentemente díspares, são as funções de Leis de Potência. E, neste sentido, este trabalho pode ser de utilidade para o professor do ensino médio, que é a quem nós nos dirigimos.

Apresentaremos a seguir, resumidamente, o que se encontra em cada capítulo.

No Capítulo 1: “Nocões de Estatística e Probabilidade”, são apresentados alguns conceitos básicos, outros nem tanto, em relação a estes dois assuntos para que se possa compreender, mais adiante, o conceito de Leis de Potência. Embora possa parecer óbvia para um matemático, a relação entre Estatística e Probabilidade normalmente

não é enfatizada no ensino médio, parecendo coisas sem ligação. Neste capítulo esta conexão se dá naturalmente, e ambos os assuntos são vistos como etapas de um processo científico mais amplo. Relações entre função, probabilidade, área e progressão geométrica também estão aqui presentes.

No Capítulo 2: “Área sob a curva”, são apresentadas algumas técnicas para o cálculo de áreas sob certas curvas não retilíneas, sem o auxílio de integrais. A relação entre este capítulo e as leis de potência se dá, pois sendo leis de potência funções de densidade de probabilidade, o valor da área sob a sua curva, em um certo intervalo, corresponde a probabilidade de ocorrência deste intervalo.

No Capítulo 3: “Regressão Linear Simples”, mostra-se, como o título sugere, como obter uma reta que melhor se ajusta a um conjunto de pontos no plano cartesiano. O uso do excel é ensinado para esta finalidade. A Regressão Linear será utilizada, posteriormente, para as leis de potência, pois como veremos, quando o gráfico de uma função de lei de potência é colocado em escala logarítmica (log-log) os pontos sugerem uma reta. E, a partir desta reta, um dos coeficientes da função é obtido.

No Capítulo 4: “Leis de Potência”, define-se as Leis de Potência; fornece-se alguns exemplos de aplicações desta lei e, em especial, é mostrada a aplicação dela ao caso da distribuição das cidades brasileiras em relação aos seus números de moradores. Mostra-se, ainda, como obtém-se a equação destas leis. Discute-se a conveniência da utilização da frequência acumulada para estas leis. Os cálculos da esperança e variância também são tratados. Discute-se brevemente a caracterização de uma função como de cauda leve ou pesada, e a relação destes termos com o conceito de momento.

No Capítulo 5: “Atividades com Leis de Potência”, são sugeridas diversas atividades, com explicações pormenorizadas, que podem ser realizadas com alunos. O Excel é a plataforma utilizada nestas atividades. Apresentamos também, em uma destas atividades, a técnica “logarithmic binning”.

No apêndice, são realizadas algumas demonstrações concernentes ao Capítulo 2.

---

## NOÇÕES DE ESTATÍSTICA E PROBABILIDADE

---

A finalidade deste capítulo é apresentar conceitos que serão necessários para a compreensão das chamadas Leis de Potência. Apresentaremos noções básicas de estatística e probabilidade tais como: tipos de variáveis; organização de dados em tabelas de frequência, histogramas, gráficos; medidas de resumo; funções de distribuição e densidade de probabilidade.

Para este capítulo, utilizamos como referência Morettin e Bussab (2013) [2] e Ross (2010) [6].

### 1.1 O QUE É A ESTATÍSTICA

A Estatística pode ser compreendida como uma parte da metodologia da ciência. É atribuída a ela a coleta, organização, redução, análise, modelagem e inferência de um conjunto de dados, extraídos de uma experiência, ou estudo, em qualquer área do conhecimento.

A coleta, organização, redução e análise dos dados fazem parte da chamada Estatística Descritiva. Como o nome já explicita, nesta primeira etapa os dados são descritos da melhor maneira possível, o que já propicia um certo conhecimento a respeito deles e prepara para as etapas seguintes.

A modelagem dos dados é a etapa em que, de posse dos dados organizados, procura-se um modelo matemático que melhor os descreva. Padrões são procurados. Para a modelagem de um fenômeno aleatório, ou seja, de um fenômeno sobre o qual não temos certeza de seu resultado, é feito uso da teoria das probabilidades. Esta teoria ocupa, portanto, um papel central em Estatística.

Já a inferência é a extrapolação dos dados, seja para a população da qual os dados (amostra) foram retirados, ou, ainda, para prever o que ocorrerá futuramente.

De todo modo, como pôde ser visto, a Estatística tem a incumbência de transformar dados em informações; compreender estas informações; e fazer previsões sobre elas.

## 1.2 TIPOS DE VARIÁVEIS

O trabalho de um estatístico começa, como vimos, na pesquisa/coleta de dados. Este trabalho é muito importante, pois dele depende todo o resto. No entanto, não trataremos deste ponto neste trabalho. Nosso foco estará no que ocorre após a coleta dos dados; sobretudo nas etapas da análise e modelagem dos dados.

O primeiro passo após a coleta é a organização/classificação dos dados. Neste passo os dados são classificados de acordo com algumas de suas **características, que recebem o nome de variáveis**. Tal classificação define o tipo de análise que podemos fazer posteriormente. Dados numéricos, por exemplo, serão tratados a partir de ferramentas distintas de dados que exprimem qualidades, como cor ou gênero. Naturalmente, dentro de cada uma das classificações podemos ter diversas subclassificações.

De modo geral as variáveis podem ser classificadas em dois grupos: **qualitativas ou quantitativas**. Dentro do primeiro grupo podemos ainda classificar as variáveis em nominais ou ordinais, e no segundo em discretas ou contínuas (Figura 1).

Uma variável é chamada de qualitativa quando seus resultados representam qualidades, atributos, e não valores numéricos. Dois exemplos: nível de instrução (ensino fundamental, ensino médio ou ensino superior) e sexo (masculino ou feminino). No primeiro exemplo, há uma ordem, gradação, hierarquia, entre os possíveis valores; variáveis deste tipo são chamadas de **qualitativas ordinais**. No segundo exemplo, não há uma ordem/gradação entre os valores; tais variáveis são chamadas de **qualitativas nominais**.

Uma variável é chamada de quantitativa quando seus resultados são valores numéricos. Dois exemplos: número de filhos e altura. No primeiro caso, os valores possíveis que tal variável pode assumir é finito ou contável (discreto); tal variável é chamada então de **quantitativa discreta**. No segundo exemplo, a obtenção dos valores (alturas) se dá por um processo de medição, sendo que neste caso os valores podem se



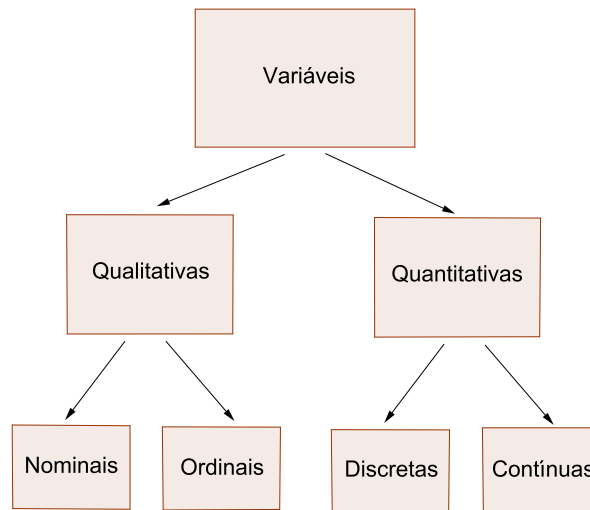


Figura 1: Tipos de variáveis.

encontrar em um intervalo de números reais; tem-se aqui uma variável **quantitativa contínua**.

Dado o tema deste trabalho, trataremos apenas das variáveis quantitativas.

### 1.3 ORGANIZAÇÃO DOS DADOS

Tendo colhido os dados, o pesquisador precisa organizá-los para poder compreendê-los, analisá-los e, finalmente, modelá-los, ou seja, encontrar uma maneira matemática de entender o comportamento deste conjunto de dados.

A maneira inicial como os dados são apresentados dependem diretamente de como a coleta foi feita. Podemos, por exemplo, estar interessados no total de filhos que cada casal tem em uma população. Neste caso os dados seriam apresentados em uma lista com a identificação do casal, e o total de filhos, como

{casal 1: 1; casal 2: 3; casal 3: 4, casal 4: 3; casal 5: 1; casal 6: 1; casal 7: 3;... }

Uma lista como esta é difícil de ser lida, e interpretada. Organizada desta forma ela mostra pouca informação sobre os dados. Antes de reorganizar os dados, note que os rótulos dados a cada casal (casal 1, casal 2, etc) são irrelevantes, e portanto agrupar os

casais pelo total de filhos que cada um tem seria mais interessante. Com isso podemos montar uma tabela com estes dados.

Tabela 1: Frequência do número de filhos por casal.

Número de filhos	Frequência $n_i$
1	15
2	20
3	20
4	10
5	9
6	8
7	7
8	5
Total	94

Na Tabela 1 temos a variável “Número de filhos por casal” em uma determinada amostra. A coluna “frequência  $n_i$ ” exibe a quantidade de casais que tem os respectivos números de filhos. Esta frequência é chamada de *frequência absoluta*, e corresponde ao total de observações com aquele valor.

Tabela 2: Variável número de filhos por casal e frequências.

Número de filhos	$n_i$	$f_i$	%	$f_{ac}$	$\%_{ac}$
1	15	0,16	16%	0,16	16%
2	20	0,21	21%	0,37	37%
3	20	0,21	21%	0,58	58%
4	10	0,11	11%	0,69	69%
5	9	0,10	10%	0,79	79%
6	8	0,09	9%	0,88	88%
7	7	0,07	7%	0,95	95%
8	5	0,05	5%	1,00	100%
<b>Total</b>	<b>94</b>	<b>1,00</b>	<b>100%</b>		

Na Tabela 2, além da frequência absoluta ( $n_i$ ), temos: frequência relativa ( $f_i$ ), porcentagem (%), frequência acumulada ( $f_{ac}$ ) e a porcentagem acumulada ( $\%_{ac}$ ).

A frequência relativa ( $f_i$ ), de determinado valor da variável, corresponde a frequência absoluta ( $n_i$ ), do valor, dividida pela soma das frequências absolutas (total de observações/amostras). A porcentagem (%) é simplesmente a frequência relativa apresentada de forma percentual.

A frequência acumulada ( $f_{ac}$ ), de um determinado valor, corresponde à somatória da frequência relativa deste valor com as frequências relativas dos valores anteriores, de modo, que, a frequência acumulada do último valor da variável será, sempre, igual a um. A porcentagem acumulada ( $\%_{ac}$ ) é a frequência acumulada em forma percentual.

Para facilitar a visualização dos dados, uma alternativa frequentemente utilizada é a representação dos dados em gráficos. Os gráficos apresentados nas Figuras 2 e 3 são chamados de gráfico de barras e gráfico de dispersão, respectivamente.

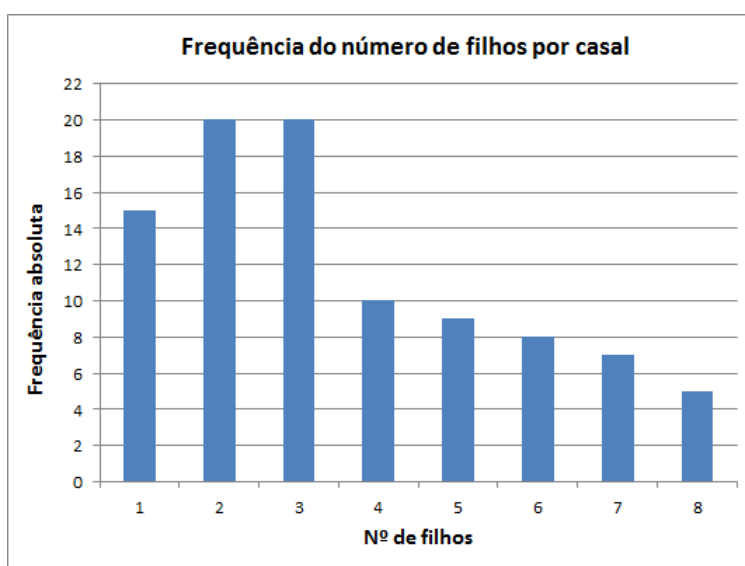


Figura 2: Gráfico de barras.

Note que no gráfico da Figura 2, os valores da variável “frequência do número de filhos por casal” foram destacados no eixo das abscissas, enquanto suas frequências absolutas foram destacadas no eixo das ordenadas. Poderia-se ter optado pelo uso da frequência relativa ou porcentagem.

No gráfico da Figura 3, tem-se algo similar ao gráfico de barras, com os dados representados por pontos em vez de barras.

O que fizemos no exemplo acima foi identificar a variável de interesse (número de filhos), e agrupar os resultados por valores possíveis da variável, verificando a frequên-

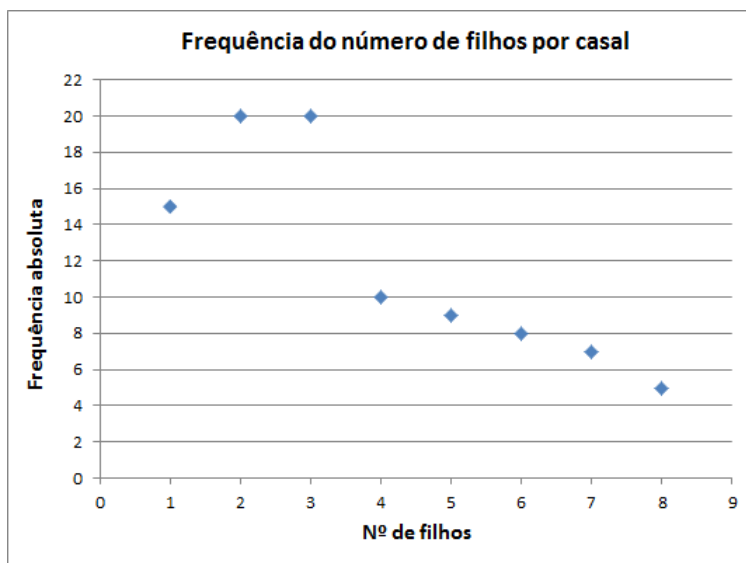


Figura 3: Gráfico de dispersão.

cia de cada valor. Mas nem sempre este procedimento é suficiente. Pense, por exemplo, no salário de cada funcionário em uma grande empresa em um dado mês. Os valores finais pagos a cada funcionário dependem de diversos fatores, como total de faltas no mês, horas extras, tempo de carreira, descontos em folha, gratificações, dentre outros.

Com isso, é pouco frequente que dois funcionários tenham exatamente o mesmo salário. Mesmo que ocupem o mesmo cargo, o salário no final do mês pode variar, mesmo que apenas na casa dos centavos. Neste caso, agrupar por salário seria apenas uma reordenação dos dados, acrescentando pouco ao que já era conhecido. Ao mesmo tempo, estas diferenças pequenas entre dois salários é uma informação pouco importante na análise final. É mais importante saber se o funcionário recebe entre 1300 e 1400 reais, do que saber que ele recebeu exatamente R\$1352,25 em um dado mês.

Uma solução para este problema é, portanto, agrupar os dados em intervalos. Podemos, por exemplo, dividir os salários em intervalos de amplitude de R\$1000,00. Na Tabela 3, tem-se a variável fictícia “Salários dos funcionários de uma empresa”, e, também, as frequências absolutas e relativas, porcentagens, frequências acumuladas e porcentagens acumuladas. Neste caso, os salários foram agrupados em faixas, considerando que as extremidades inferiores estão incluídas nos intervalos (1500, por exemplo, caso tenha ocorrido, está incluído no segundo intervalo, e não no primeiro).

Este procedimento é utilizado, com frequência, para variáveis quantitativas contínuas, ou para discretas, que apresentem valores em um intervalo de grande amplitude.

Tabela 3: Salários dos funcionários de uma empresa.

Salários	$n_i$	$f_i$	%	$f_{ac}$	$\%_{ac}$
[500;1500[	82	0,29	29%	0,29	29%
[1500;2500[	53	0,19	19%	0,49	49%
[2500;3500[	41	0,15	15%	0,63	63%
[3500;4500[	34	0,12	12%	0,76	76%
[4500;5500[	30	0,11	11%	0,86	86%
[5500;6500[	22	0,08	8%	0,94	94%
[6500;7500[	12	0,04	4%	0,99	99%
[7500;8500[	4	0,01	1%	1,00	100%
<b>Total</b>	<b>278</b>	<b>1,00</b>	<b>100%</b>		

Para a representação gráfica dos dados da Tabela 3, é conveniente a utilização de histogramas (Figura 4). Chama-se **gráfico de histogramas**, ou simplesmente **histogramas**, uma sequência de retângulos contíguos cujas bases (abscissas) são os intervalos das faixas e cujas alturas (ordenadas) são as densidades das frequências relativas dos valores.

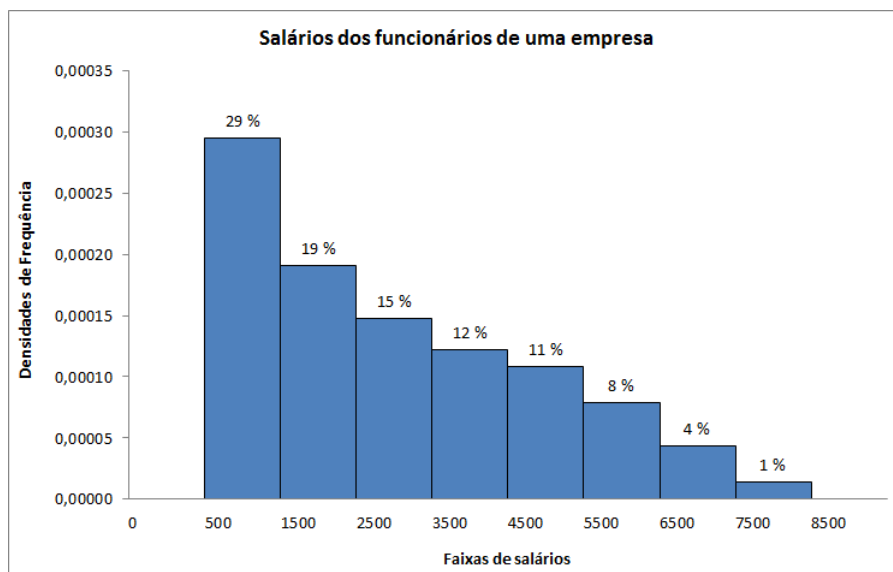


Figura 4: Histograma dos salários dos funcionários de uma empresa

## 1.4 MEDIDAS-RESUMO

Por meio da organização dos dados brutos, tal como foram colhidos, em uma tabela de frequências, consegue-se obter uma visão mais resumida e, portanto, mais organizada dos dados. No entanto, pode-se querer uma visão ainda mais enxuta dos dados, resumindo-os ainda mais.

Uma primeira maneira de fazer isso é tentar “localizar” os dados. Pense, por exemplo, que uma criança resolve brincar com um punhado de feijões, espalhando-os pela mesa. Quando dizemos que os feijões estão espalhados no meio da mesa, estamos um ponto exato da mesa (neste caso o centro) para descrever a posição dos feijões. Mas qual o melhor ponto para representar esta posição?

Existem diversas formas de se escolher um único valor que seja representativo de toda a série de dados. Podemos, por exemplo, escolher o valor mais frequente (*moda*). Para dados ordenados, podemos escolher o valor para o qual a frequência acumulada seja 50% (*mediana*). Mas a mais utilizada é a *média aritmética*.

A *média aritmética*, ou simplesmente *média*, corresponde à soma dos valores observados em cada dado, dividida pela quantidade total de observações. No exemplo da Tabela 2, a média é dada por:

$$\bar{x} = \frac{1 \times 15 + 2 \times 20 + 3 \times 20 + 4 \times 10 + 5 \times 9 + 6 \times 8 + 7 \times 7 + 8 \times 5}{94} = \frac{337}{94} \approx 3,6.$$

Assim, usando os dados brutos, sem organização em tabela de frequência, se para uma variável  $X$  foram feitas  $n$  observações, iguais ou não, dadas por  $x_1, x_2, \dots, x_n$ , podemos escrever a média de  $X$  como:

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

Ou ainda,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Agrupando os valores da variável  $X$  em uma tabela de frequência, teremos que o valor  $x_1$  ocorreu  $n_1$  vezes,  $x_2$  ocorreu  $n_2$  vezes, ...,  $x_k$  ocorreu  $n_k$  vezes, de modo que  $n_1 + n_2 + \dots + n_k = n$ , e podemos reescrever a média como:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k \frac{n_i}{n} x_i.$$

Considerando que a frequência relativa de  $x_i$  pode ser expressa por  $f_i = n_i/n$ , em que  $n_i$  é o número de ocorrências de  $x_i$ , então a média pode ser reescrita como:

$$\bar{x} = \sum_{i=1}^k f_i x_i. \quad (1.1)$$

Embora as medidas de posição (moda, mediana e média) tenham a vantagem de reduzir o conjunto de dados para apenas um número, naturalmente ocorre uma perda de informações. Considere por exemplo, os seguintes conjuntos de dados

$$X = \{-2, -1, 0, 1, 2\}$$

e

$$Y = \{-10, -5, 0, 5, 10\}.$$

A média aritmética dos dois conjuntos é 0, mas eles são claramente distintos. Talvez a principal diferença consiste em quanto os conjuntos estão *dispersos* em torno de suas médias. Os dados de  $X$  estão claramente mais *concentrados* em torno da média 0.

Com isso, fica claro que precisamos de uma espécie de *medida de dispersão* para nos dizer o quão espalhados nossos dados estão.

Uma primeira forma de fazer tal estimativa é medir a distância de cada dado para a média, calculando  $x_i - \bar{x}$ , obtendo assim o quanto o valor  $x_i$  está afastado, para baixo ou para cima, da média. Este cálculo nos fornece o desvio de  $x_i$ . Por exemplo, para o conjunto de valores: 1, 3, 4 e 8, cuja média é  $\bar{x} = 4$ , os desvios são:  $1 - 4 = -3$ ,  $3 - 4 = -1$ ,  $4 - 4 = 0$  e  $8 - 4 = 4$ . Podemos notar que a soma dos desvios é igual a zero! De fato,

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) = x_1 + x_2 + \dots + x_n - n \cdot \bar{x} = 0.$$

A solução para isso parece simples! Se queremos medir distância do dado para a média, precisamos de um valor positivo. Podemos, por exemplo, tomar o módulo da diferença, calculando  $|x_i - \bar{x}|$ , e depois calcular a média destes valores. Calculamos assim o **desvio médio** ( $dm$ ) dos dados, que para os valores  $\{1, 3, 4, 8\}$  acima teríamos

$$dm = \frac{|1 - 4| + |3 - 4| + |4 - 4| + |8 - 4|}{4} = 2.$$

No entanto a solução mais usada para o problema (soma dos desvios igual a zero), é elevar cada um dos desvios ao quadrado. Para em seguida calcular a média destes quadrados dos desvios. Obtemos assim o que chamamos de **variância** ( $var$ ) dos dados:

$$var = \frac{(1 - 4)^2 + (3 - 4)^2 + (4 - 4)^2 + (8 - 4)^2}{4} = 6,5.$$

Como a dimensão da variância é o quadrado da dimensão dos dados, uma vez que elevamos os desvios ao quadrado, para obter um número que nos forneça uma interpretação mais fácil em relação à dispersão dos dados, tiramos a raiz quadrada da variância. O número obtido é o **desvio padrão** ( $dp$ ):

$$dp = \sqrt{6,5} \approx 2,55.$$

Podemos interpretar o desvio padrão, 2,55, como sendo o erro médio cometido ao assumirmos a média, 4, como representativa dos dados.

De um modo geral, para uma variável  $X$ , temos

$$var(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \quad (1.2)$$

$$dp(X) = \sqrt{var(X)}. \quad (1.3)$$

Em relação ao exemplo considerado na Tabela 2, obtemos

$$var = \frac{15(1 - 3,6)^2 + 20(2 - 3,6)^2 + \dots + 7(7 - 3,6)^2 + 5(8 - 3,6)^2}{94} \approx 4,29.$$

$$dp = \sqrt{4,29} \approx 2,07.$$

Reescrevendo a expressão (1.2) usando frequência relativa temos

$$var(X) = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^2}{n} = \sum_{i=1}^k \frac{n_i}{n} (x_i - \bar{x})^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2. \quad (1.4)$$

**Observação 1.4.1.** No caso de dados agrupados em intervalos de valores, é necessário escolher um valor para representar o intervalo. O mais comum neste caso é usar o valor médio do  $i$ -ésimo intervalo como o valor de  $x_i$ , fazendo os demais cálculos normalmente.

## 1.5 PROBABILIDADE

Como comentamos anteriormente um dos objetivos ao analisar um conjunto de dados é ter um modelo matemático que o represente. O suporte teórico para tais modelos é dado pela chamada Teoria da Probabilidade.



A probabilidade se propõe a estudar o que chamaremos de *fenômeno aleatório*, ou seja, um acontecimento ou experimento no qual não temos controle total, e portanto não é possível ter certeza do resultado. Ao lançar uma moeda comum, por exemplo, não se tem certeza se sairá cara ou coroa. O mesmo acontece quando paramos uma pessoa no meio da rua para perguntar qual o candidato na próxima eleição. Este tipo de experimento é conhecido como experimento aleatório, e dá início ao processo de modelagem que descreveremos brevemente a seguir.

Apesar de não termos controle sobre o resultado final do experimento realizado, temos ao menos o conhecimento prévio do conjunto de resultados possíveis para tal experimento. Se rolarmos um dado, por exemplo, sabemos de antemão que obteremos um valor no conjunto  $\{1, 2, 3, 4, 5, 6\}$ . Ou ao sortear um funcionário em uma empresa e perguntar seu salário, sabemos que vamos obter um valor real positivo. Temos assim definido o que chamamos de Espaço Amostral ( $\Omega$ ), que é simplesmente o conjunto formado por todos os possíveis resultados de um experimento aleatório.

O próximo passo na nossa modelagem é associar aos resultados de um experimento o que chamaremos de probabilidade, ou seja, um valor numérico que represente a chance que temos de observar um dado resultado em uma realização do experimento. Se lançarmos um dado de seis faces para o alto, ao cair, uma de suas faces ficará voltada para cima. O espaço amostral desta experiência é o conjunto  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Admitindo que o dado é perfeitamente simétrico, com sua massa igualmente distribuída por todo o seu volume, ou seja, que nenhuma face tem mais chance do que outra de ocorrer, tem-se a mesma probabilidade para cada face, no valor de  $1/6$ . Ao atribuir a cada face a probabilidade de  $1/6$ , considera-se que ao repetir a experiência um número grande de vezes, a frequência relativa de cada face será um número, se não igual, ao menos bem próximo, de  $1/6$ .

Infelizmente nem todo experimento pode ser resolvido de maneira tão simples. Considere o experimento simples de jogar um feijão dentro de um caixa de base quadrada de lado  $1m$ . Estamos interessados em medir a distância do feijão para uma face específica da caixa. Neste caso o espaço amostral é todo o intervalo  $[0, 1]$ , composto por todos os números reais entre 0 e 1.

Se quisermos fazer o mesmo procedimento usado na rolagem de um dado, esbarraremos em um problema complicado. Antes de mais nada, não existe nada no experimento que nos faça crer que alguma distância é mais provável que as demais. Assim seria tentador falar que cada valor em  $[0, 1]$  tem a mesma probabilidade. Mas infelizmente isso não é possível!

Note que no caso dos dados a soma das probabilidades de cada resultado é 1 (ou 100%), que é compatível com o fato de não existir nenhum outro resultado possível. Mas no intervalo  $[0, 1]$  o total de resultados possíveis é infinito, e portanto qualquer valor positivo que fosse atribuído aos pontos individuais, levaria à uma soma infinita, que não tem como ser interpretada como probabilidade.

A solução para este problema é associar valores à intervalos e não à pontos. No caso do exemplo do feijão, diríamos que dois intervalos de mesmo tamanho teriam a mesma probabilidade, dada apenas pelo comprimento do intervalo. Isso daria ao intervalo  $[0, 1]$  uma probabilidade de 1, consistente com o experimento proposto.

Vamos poupar o leitor dos formalismos envolvidos na definição de probabilidade. Ao invés disso vamos passar direto para exemplos, tentando ilustrar o processo de modelagem.

Para isso vamos usar um objeto auxiliar chamado de *variável aleatória*. Tais variáveis são, assim como no caso da estatística, valores numéricos associados ao resultado do experimento. A estes valores iremos associar um valor positivo, e menor que 1, que chamaremos de probabilidade.

Assim como descrevemos na estatística, as variáveis aleatórias serão divididas em dois grandes grupos: as discretas e as contínuas. A relação entre tais variáveis e as descritas na estatística vão além do nome, como deixaremos mais claro nas próximas seções.

## 1.6 VARIÁVEIS ALEATÓRIAS DISCRETAS

Chamaremos de variável aleatória discreta, um observável numérico relacionado a um experimento aleatório que assume valores em um conjunto contável. Formalmente, uma variável aleatória discreta é uma função com domínio no espaço amostral, e cuja imagem é um conjunto contável.

**Exemplo 1.1.** Como exemplo de variável aleatória discreta, pode-se considerar, a variável  $X$  que representa o “Número de filhos”, como presente nas Tabelas 1 e 2. Ao sortearmos um casal aleatoriamente dentre os casais pesquisados, o total de filhos do casal será um valor aleatório dentro dos inteiros positivos. Mais precisamente,  $X$  pode assumir valores em  $\{1, 2, 3, 4, 5, 6, 7, 8\}$ .

Após identificar a variável, o próximo passo é definir as probabilidades da variável assumir cada valor na imagem. Isso é feito associando a cada valor  $x_i$  da imagem, um valor  $p(x_i)$ , definindo assim a chamada *função de probabilidade*.

Chamaremos de função de probabilidade da variável aleatória discreta  $X$ , uma função que associa uma probabilidade de ocorrência a cada um dos valores possíveis de  $X$ . Considerando cada valor particular de  $X$  como  $x_i$ , sendo  $i$  natural, indicamos o valor da função de probabilidade para cada valor de  $X$  de um dos seguintes modos

$$p(x_i) = P(X = x_i) = p_i, \quad i = 1, 2, \dots,$$

de modo que

$$p(x_1) + p(x_2) + p(x_3) + \dots = 1.$$

Considerando, ainda, a variável “Número de filhos”, das Tabelas 1 e 2, como poderíamos estabelecer uma função de probabilidade para esta variável? Considerando que cada casal da população tem a mesma probabilidade de ser sorteado, a probabilidade de sortearmos um casal com um filho seria dado pelo total de casais com 1 filho, dividido pelo total de casais da população. Ou seja, a proporção de casais da população com apenas 1 filho.

Normalmente, ao realizarmos este tipo de experimento, não sabemos exatamente o total de casais com  $i$  filhos, e normalmente esta é uma das informações que buscamos!

Dadas certas condições que não discutiremos neste trabalho, é possível assumir as frequências relativas da Tabela 2 como aproximação para as probabilidades de cada um dos valores da variável. Neste caso, a função de probabilidade associa os valores  $x_i$  desta variável às frequências relativas,  $f_i$ , desta tabela.

Mais adiante, será vista a aplicação de probabilidade para variáveis aleatórias contínuas; mas antes, serão considerados alguns modelos/distribuições mais comuns para variáveis aleatórias discretas (v.a. discretas).

### 1.6.1 Modelo Uniforme Discreto

Este modelo aplica-se às v.a. discretas em que cada valor possível tem a mesma probabilidade de ocorrer. Ou seja, uma variável aleatória discreta  $X$  tem uma distribuição

uniforme se assume valores em um conjunto finito  $\{x_1; \dots; x_k\}$  e para cada valor  $x_i$ , tem-se:

$$p(x_i) = \frac{1}{k}, \quad \text{para qualquer } i = 1, 2, \dots, k.$$

Novamente, pode-se evocar, como exemplo, o caso de um dado, não viciado, de seis faces ( $k = 6$ ). Para cada uma das faces, tem-se a mesma probabilidade de ocorrência ( $P = 1/6$ ). Logo, a função de probabilidade da variável aleatória que associa uma probabilidade de ocorrência à cada face é de distribuição uniforme.

Aproveitaremos o Modelo Uniforme Discreto para apresentar a ideia de valor médio esperado (esperança) e variância para uma variável aleatória discreta.

### 1.6.2 Valor Médio e Variância para Variáveis Aleatórias Discretas

Considerando a relação entre probabilidade e frequência relativa, podemos adaptar a fórmula (1.1), da média de valores observados, para o que chamamos de *valor esperado (esperança)* de uma variável aleatória  $X$ , cujos possíveis valores são  $x_1, x_2, \dots, x_n$ . Em (1.1) substituiremos  $f_i$  por  $p_i$ , obtendo

$$E(X) = \sum_{i=1}^n p_i x_i. \quad (1.5)$$

Qual seria, então, o valor esperado ao lançar um dado, não viciado, de seis faces ( $k = 6$ )?

A probabilidade de cada valor (face) de nossa variável aleatória ( $X$ ) é a mesma:

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$

Aplicando a fórmula da esperança encontra-se

$$E(X) = 1 \times \left(\frac{1}{6}\right) + 2 \times \left(\frac{1}{6}\right) + 3 \times \left(\frac{1}{6}\right) + 4 \times \left(\frac{1}{6}\right) + 5 \times \left(\frac{1}{6}\right) + 6 \times \left(\frac{1}{6}\right) = 3,5.$$

De forma análoga, define-se **variância para uma v.a. discreta**. Em (1.4) substituiremos  $f_i$  por  $p_i$ , e  $\bar{x}$  por  $E(X)$ , obtendo

$$var(X) = \sum_{i=1}^n p_i (x_i - E(X))^2. \quad (1.6)$$

A partir de (1.6), podemos deduzir uma outra fórmula para a variância, que nos será muito útil. Segue a dedução:

$$\begin{aligned}
 \text{var}(X) &= \sum_{i=1}^n p_i (x_i - E(X))^2 \\
 &= \sum_{i=1}^n p_i \cdot (x_i^2 - 2x_i E(X) + [E(X)]^2) \\
 &= \sum_{i=1}^n (p_i x_i^2 - 2p_i x_i E(X) + p_i [E(X)]^2) \\
 &= \sum_{i=1}^n p_i x_i^2 - \sum_{i=1}^n 2p_i x_i E(X) + \sum_{i=1}^n p_i [E(X)]^2.
 \end{aligned}$$

Denotando  $\sum_{i=1}^n p_i x_i^2 = E(X^2)$ , e usando (1.5), encontramos

$$\begin{aligned}
 \text{var}(X) &= E(X^2) - 2E(X) \cdot \sum_{i=1}^n p_i x_i + [E(X)]^2 \cdot \sum_{i=1}^n p_i \\
 &= E(X^2) - 2E(X) \cdot E(X) + [E(X)]^2 \cdot 1 \\
 &= E(X^2) - 2[E(X)]^2 + [E(X)]^2 \\
 &= E(X^2) - [E(X)]^2.
 \end{aligned}$$

Assim,

$$\text{var}(X) = E(X^2) - [E(X)]^2. \quad (1.7)$$

### 1.6.3 Distribuição Binomial

Tome um certo experimento, onde estamos interessados em observar um resultado específico, cuja probabilidade de ocorrência nós conhecemos, e é igual a  $p$ . Agora repita este experimento  $n$  vezes e defina a variável aleatória  $X$  como o total de vezes que o resultado em questão foi observado.

Assim,  $X$  é portanto uma variável aleatória discreta, assumindo valores em  $\{0, 1, 2, \dots, n\}$ , e estamos interessados em encontrar sua função de probabilidade. Ou seja, queremos encontrar  $P(X = k)$ .

Para isso, considere uma sequência de resultados específicos onde o evento de interesse foi observado  $k$  vezes. Por exemplo,

$$\underbrace{SSS \cdots SS}_{k \text{ vezes}} \underbrace{NNN \cdots NN}_{n-k \text{ vezes}}$$

onde  $S$  denota o acontecimento do resultado de interesse, e  $N$  a não ocorrência.

Esta sequência tem probabilidade  $p^k(1-p)^{n-k}$  de ocorrer, pois cada  $S$  tem probabilidade  $p$  enquanto cada  $N$  tem probabilidade  $1-p$ .

Mas cada permutação destes  $n$  elementos tem a mesma probabilidade de ocorrer, e em todas elas o valor observado de  $X$  é  $k$ .

Assim a função probabilidade de  $X$  é dada por

$$P(X = k) = p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{com } k = 0, 1, \dots, n.$$

**Exemplo 1.2.** Lança-se uma moeda, não viciada, cinco vezes para o alto. Qual é a probabilidade de ocorrência de, exatamente, três caras?

Consideraremos cara como “C”, e coroa como “R”. Depreendendo do problema os parâmetros  $k$ ,  $n$  e  $p$ , podemos escrever:

$$n = 5; \quad k = 3; \quad p = \frac{1}{2}; \quad 1 - p = 1 - \frac{1}{2} = \frac{1}{2}; \quad \text{e, portanto:}$$

$$P(X = 3) = \binom{5}{3} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^{5-3}.$$

Note que quando se procura a probabilidade de exatamente três ocorrências de cara, a ordem destas ocorrências pode ser qualquer uma. Algumas ordens possíveis: CCCRR, CRCCR, CCRRC. No cálculo da probabilidade, tem-se:

$\left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^{5-3}$ : A probabilidade de apenas três caras em uma certa ordem, por exemplo, CCCRR.

$\binom{5}{3}$ : A quantidade de casos em que se tem exatamente três caras considerando qualquer ordem possível. Esta quantidade corresponde a combinação de cinco, três a três.

#### 1.6.4 Distribuição Geométrica

Trataremos agora, de uma distribuição particularmente interessante para este trabalho, a distribuição geométrica.

Nas distribuições anteriores, nos limitamos a casos em que o número de valores assumidos pela variável aleatória era finito. As variáveis aleatórias que seguem uma distri-

buição geométrica assumem um total infinito enumerável de valores. Para introduzi-la, utilizaremos um exemplo.

**Exemplo 1.3.** Uma urna contém 6 bolas brancas e 4 bolas pretas. As bolas são retiradas aleatoriamente, uma de cada vez, até que saia uma bola preta. Se cada bola retirada é devolvida para a urna antes da próxima bola ser retirada, e  $X$  é o total de sorteios necessários, qual a função de probabilidade de  $X$ .

Observe que  $X$  é uma variável aleatória que assume valores no conjunto  $\{1, 2, 3, 4, \dots\}$ , e para que  $X = n$  teremos que sortear  $n - 1$  bolas brancas e na  $n$ -ésima jogada uma bola preta.

Como a probabilidade de sortear uma bola branca é de  $\frac{6}{10}$ , enquanto sortear uma bola preta tem probabilidade  $\frac{4}{10}$ , temos que

$$P(X = n) = \left(\frac{6}{10}\right)^{n-1} \cdot \frac{4}{10}.$$

A distribuição de probabilidade deste exemplo é chamada de geométrica.

Em uma distribuição geométrica, assim como na distribuição binomial, temos repetições, sucessivas e independentes, de um experimento envolvendo uma v.a. Em cada experimento, a probabilidade de um evento, chamado de sucesso, é  $p$ ,  $0 \leq p \leq 1$ . A probabilidade de não ocorrência do sucesso é chamada de fracasso e sua probabilidade é  $1 - p$ . O experimento é repetido diversas vezes até que um sucesso ocorra. Chamando de  $X$  o número de experiências até que ocorra um sucesso, podemos escrever:

$$P(X = n) = (1 - p)^{n-1} p, \quad \text{com } n = 1, 2, \dots \quad (1.8)$$

A justificativa da equação (1.8) é que até a ocorrência do sucesso, cuja probabilidade é  $p$ , são necessárias  $n - 1$  tentativas fracassadas, cada uma com probabilidade  $1 - p$ .

Pode pairar uma dúvida, quanto à legitimidade da distribuição geométrica como uma função de distribuição de probabilidade, já que para isto, é necessário que a soma das probabilidades associadas aos valores da v.a. seja igual 1. A variável aleatória pode assumir, em tese, infinitos valores. Logo, fica a impressão de que a soma das probabilidades pode crescer indefinidamente.

Para resolver este problema, vamos fazer um parêntese do texto para lembrar uma expressão conhecida de boa parte dos estudantes do ensino médio.

Seja  $a_n$  uma PG de razão  $r$ , ou seja,  $a_n = a_1 r^{n-1}$ . Queremos saber qual é a soma dos  $n$  primeiros termos desta PG. Para isso faça

$$\begin{aligned} S_n &= a_1 + a_2 + \cdots + a_n \\ &= a_1 + a_1 r + \cdots + a_1 r^{n-1}, \end{aligned}$$

e note que

$$r \cdot S_n = a_1 r + a_1 r^2 + \cdots + a_1 r^n.$$

Assim

$$S_n - r \cdot S_n = (a_1 + a_1 r + \cdots + a_1 r^{n-1}) - (a_1 r + a_1 r^2 + \cdots + a_1 r^n) = a_1 - a_1 r^n,$$

e portanto

$$(1 - r)S_n = a_1(1 - r^n),$$

e

$$S_n = \frac{a_1(1 - r^n)}{1 - r}.$$

Agora, para encontrar a soma de todos os termos da PG, basta fazer o valor de  $n$  crescer. Assim se  $0 \leq r < 1$ ,  $r^n$  fica cada vez mais próximo de zero quanto maior for o valor de  $n$ , de onde concluímos que

$$a_1 + a_1 r + a_1 r^2 + a_1 r^3 + \cdots = \frac{a_1}{1 - r},$$

para  $0 \leq r < 1$ .

Terminado este breve parêntes, voltemos ao problema em questão. Queremos mostrar que a soma de todos os valores da função de probabilidade de  $X$  é de fato 1. Para isso note que

$$\sum_{n=1}^{\infty} P(X = n) = \sum_{n=1}^{\infty} (1 - p)^{n-1} p = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1.$$

Temos, então, que a soma, embora de infinitos termos, é igual a 1.

Uma questão interessante que pode ser colocada é se há valor médio, esperança, para uma v.a. geométrica. A resposta é positiva. Mostraremos abaixo, que para uma v.a. geométrica podemos escrever:

$$E(X) = \frac{1}{p}. \quad (1.9)$$



Para isso note que

$$\begin{aligned} E(X) &= \sum_{n=1}^{\infty} n \cdot p(n) \\ \Rightarrow E(X) &= \sum_{n=1}^{\infty} n \cdot (1-p)^{n-1} \cdot p. \end{aligned}$$

Fazendo  $1-p=q$ , temos

$$\begin{aligned} E(X) &= \sum_{n=1}^{\infty} n \cdot q^{n-1} \cdot p \\ \Rightarrow E(X) &= \sum_{n=1}^{\infty} (n-1+1) \cdot q^{n-1} \cdot p \\ \Rightarrow E(X) &= \sum_{n=1}^{\infty} (n-1) \cdot q^{n-1} \cdot p + \sum_{n=1}^{\infty} q^{n-1} \cdot p. \end{aligned}$$

Notando que  $\sum_{n=1}^{\infty} q^{n-1} \cdot p = 1$ , por soma de PG, segue que

$$E(X) = \sum_{n=1}^{\infty} (n-1) \cdot q^{n-1} \cdot p + 1.$$

Fazendo  $n-1=m$ :

$$\begin{aligned} E(X) &= \sum_{m=0}^{\infty} m \cdot q^m \cdot p + 1 \\ \Rightarrow E(X) &= \sum_{m=0}^{\infty} m \cdot q^{m-1} \cdot q \cdot p + 1 \\ \Rightarrow E(X) &= q \sum_{m=0}^{\infty} m \cdot q^{m-1} \cdot p + 1. \end{aligned}$$

Notando que o termo da soma para  $m=0$  é nulo, temos

$$\begin{aligned} E(X) &= q \sum_{m=1}^{\infty} m \cdot q^{m-1} \cdot p + 1 \\ \Rightarrow E(X) &= q \cdot E(X) + 1 \\ \Rightarrow E(X) - q \cdot E(X) &= 1 \\ \Rightarrow E(X) \cdot (1-q) &= 1 \\ \Rightarrow E(X) \cdot [1-(1-p)] &= 1 \\ \Rightarrow p \cdot E(X) &= 1 \\ \Rightarrow E(X) &= \frac{1}{p}. \end{aligned}$$

## 1.7 VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Chamaremos de *variável aleatória contínua* um observável numérico que assuma valores em um intervalo de números reais. Mais formalmente, uma variável aleatória contínua é uma função com domínio em um espaço amostral ( $\Omega$ ), com imagem dada por um intervalo de número reais.

Para tentar entender melhor o que entendemos por variável aleatória contínua, vamos usar um exemplo, adaptado de Morettin e Bussab (2013, p.168), [2]. Considere um relógio elétrico, onde o ponteiro do segundo se mova continuamente, sem saltos. Após anos de funcionamento o relógio finalmente para. A posição em que o ponteiro dos segundos para é aleatória, e pode ser medida pelo ângulo entre o ponteiro dos segundos e o eixo imaginário que passa pelo centro e pelo ponto onde se encontra o XII.

No relógio em questão, elétrico, o ponteiro dos segundos pode parar em qualquer lugar, e portanto o ângulo  $X$  em questão é uma variável aleatória que assume valores no intervalo  $[0, 360)$ .

Como comentamos anteriormente, não podemos atribuir a cada ponto do intervalo  $[0, 360)$  um mesmo valor positivo. Assim, para contornar esta dificuldade, atribuímos para cada intervalo de  $[a, b)$  em  $[0, 360)$ , um valor representando a probabilidade da variável  $X$  estar naquele intervalo.

Neste caso em particular, intervalos de um mesmo tamanho devem ter a mesma probabilidade associada. Levando isto em conta, ao considerar, por exemplo, a chance do ponteiro estar compreendido entre  $0^\circ$  e  $90^\circ$ , esta probabilidade deve ser de  $1/4$ , pois

$$P(0 \leq X < 90) = P(90 \leq X < 180) = P(180 \leq X < 270) = P(270 \leq X < 360),$$

e

$$P(0 \leq X < 90) + P(90 \leq X < 180) + P(180 \leq X < 270) + P(270 \leq X < 360) = 1.$$

Mas isso é o mesmo que dizer que

$$P(0^\circ \leq X \leq 90^\circ) = \frac{1}{4} = \frac{90 - 0}{360}.$$

Para duas extremidades  $a$  e  $b$  de um intervalo em  $[0, 360)$ , a probabilidade de  $X \in [a, b)$ , é dada por:

$$P(a \leq X < b) = \frac{b - a}{360}.$$

É interessante notar que tais probabilidades podem ser calculadas como a área entre o eixo das abscissas e o gráfico da função  $f(x)$  dada por

$$f(x) = \begin{cases} 0 & , \text{ se } x < 0; \\ \frac{1}{360} & , \text{ se } 0 \leq x < 360; \\ 0 & , \text{ se } x \geq 360, \end{cases}$$

como ilustrado na Figura 5.

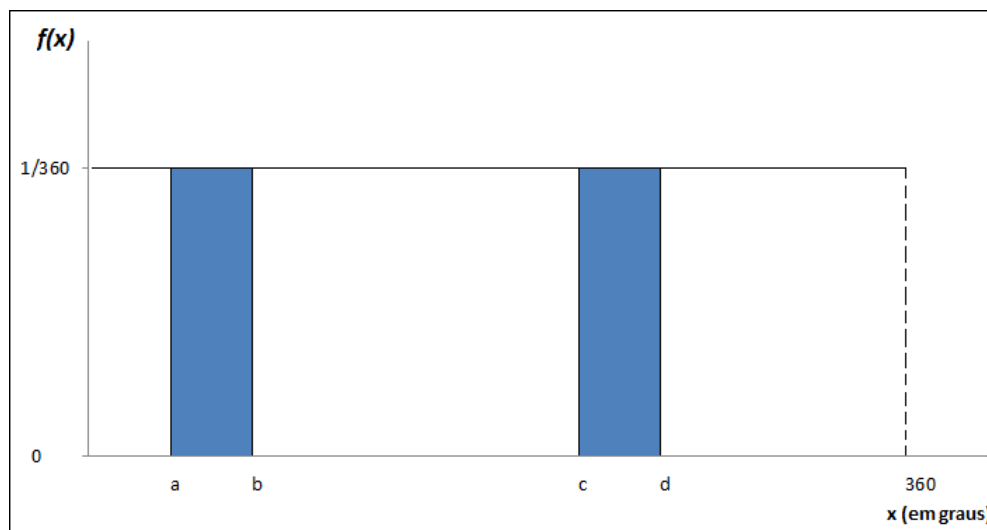


Figura 5: Probabilidades na distribuição uniforme contínua. Reprodução de Morettin e Bussab (2013, p.170), [2].

O interessante desta ideia é que ela pode ser generalizada, dando uma forma concreta de definir probabilidades para uma variável aleatória contínua. Em outras palavras, a probabilidade de uma variável aleatória contínua ser observada em um intervalo  $[a, b]$  pode ser definida como a área sob a curva da função  $f(x)$ , acima do eixo das abscissas, entre  $a$  e  $b$ .

Tal função  $f(x)$  é chamada de *função densidade de probabilidade* da v.a.  $X$ . No exemplo acima, onde a função  $f(x)$  é constante em um intervalo, diremos a variável  $X$  possui distribuição uniforme contínua.

Assim, dada uma função  $f(x)$ , se denotarmos por  $A_f[a, b]$  a área abaixo do gráfico de  $f$ , acima do eixo das abscissas, no intervalo  $[a, b]$ , temos que

$$P(a \leq X < b) = A_f[a, b].$$

Mas note que nem toda função pode ser usada como função densidade de probabilidade. De fato, se considerarmos a função constante  $f(x) = 1$ , para todo  $x \in \mathbb{R}$ , como função densidade de uma variável aleatória  $X$  teríamos que

$$P(0 \leq X \leq 2) = A_f[0, 2] = 2,$$

o que é claramente um absurdo.

Vamos então dar um pouco mais de precisão ao conceito de função densidade de probabilidade de uma v.a. contínua.

Para que uma função possa ser de *densidade de probabilidade de uma v.a. contínua*, é estritamente necessário que ela não assuma valores negativos, e que a área da região sob a sua curva, limitada também pelo eixo das abscissas, seja igual a um. Ou seja,

- $f(x) \geq 0$ , para todo  $x \in \mathbb{R}$ ;
- $A_f[\mathbb{R}] = 1$ .

**Exemplo 1.4.** Seria a função

$$f(x) = \begin{cases} 0 & , \quad \text{se } x < 0; \\ 2x & , \quad \text{se } 0 \leq x < a; \\ 0 & , \quad \text{se } x \geq a, \end{cases}$$

uma candidata função densidade de probabilidade? E, caso positivo, para que valor de  $a$ ? (Figura 6)

Note, antes de mais nada, que  $f(x) \geq 0$  para todo  $x \in \mathbb{R}$ . Falta portanto determinar a área  $A_f[\mathbb{R}]$ . Pra isso, note que para  $x < 0$  e  $x \geq a$ , temos  $f(x) = 0$ , e portanto a área desejada pode ser calculada apenas no intervalo  $[0, a)$ . Assim,

$$A_f[\mathbb{R}] = \frac{(a - 0) \cdot 2a}{2} = a^2.$$

E portanto, para que tenhamos  $A_f[\mathbb{R}] = 1$ , como  $a > 0$ , precisamos que  $a = 1$ .

Portanto,  $f(x) = 2x$ ,  $0 \leq x \leq 1$ , pode ser tomada como uma função densidade de probabilidade de uma v.a.  $X$ . Neste caso, as probabilidades dos intervalos  $[0; 0,5[$  e  $[0,5; 1]$  podem ser calculadas como abaixo:

$$P(0 \leq X < 0,5) = \frac{0,5 \cdot 1}{2} = 0,25.$$

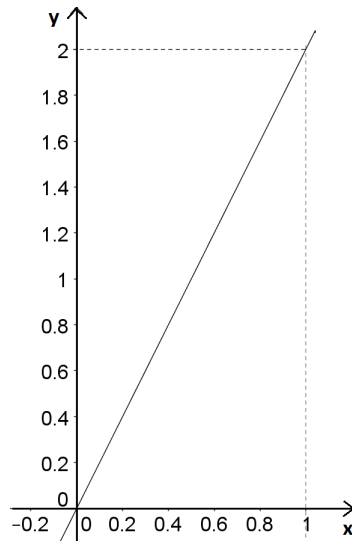


Figura 6: Gráfico da função real  $f(x) = 2x$ .

$$P(0,5 \leq X < 1) = \frac{(2+1) \cdot 0,5}{2} = 0,75.$$

É interessante notar que embora os intervalos tenham a mesma amplitude, as probabilidades são diferentes - diferentemente do que ocorreu no exemplo anterior em que a função densidade de probabilidade era constante:  $f(x) = \frac{1}{360}$ .

No Capítulo 2, serão desenvolvidas algumas técnicas para o cálculo da área sob as curvas de certas funções não retilíneas. Essas técnicas visarão, sobretudo, às distribuições que definiremos como leis de potência.

Vejamos, agora, dois outros exemplos envolvendo probabilidade para v.a. contínuas. Frisamos que nestes casos trabalhamos com intervalos que contém um número infinito de valores, uma vez que dizem respeito aos números reais.

**Exemplo 1.5.** <sup>1</sup>: Em uma competição de aerodelismo, vence o participante que conseguir pousar mais vezes seu aeroplano na área especificada. Esta área consiste em um triângulo equilátero, inscrito em um círculo. Sabendo que um aeroplano pousou dentro do círculo, qual é a probabilidade de ter pousado também dentro do triângulo?

<sup>1</sup> Disponível em: <<http://cdn.veduca.com.br/uploads/lecture/material/303-Aula01-Resolucao-dos-Quizzes-e-Exercicios.pdf>>, em <<http://www.veduca.com.br/>>. Acesso em: 20 jan.2016.

Suponha que a densidade de probabilidade em todos os pontos do círculo é a mesma. Obs: Lembre-se de considerar o brinquedo como um ponto material.

Primeiro, vamos representar a região de pouso do aeroplano por meio de um desenho (Figura 7).

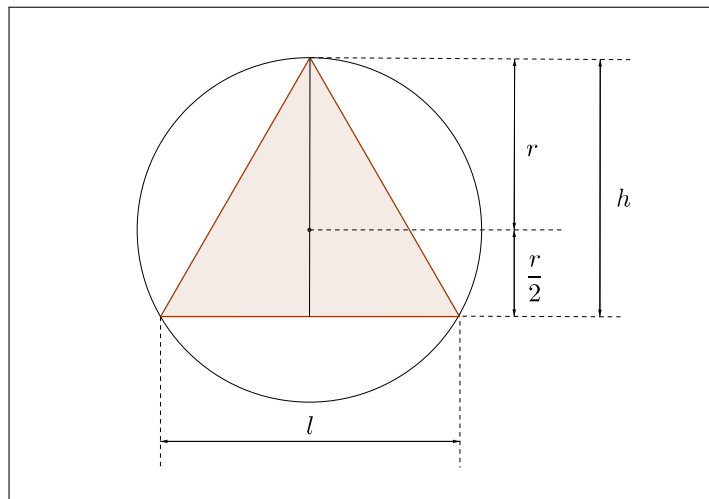


Figura 7: Região de pouso do aeroplano.

Tanto no círculo, quanto no triângulo equilátero, há infinitos pontos. Mas, como já fizemos antes, consideraremos não os pontos, mas as áreas das regiões.

O espaço amostral corresponde ao círculo. A v.a. é constituída pelas áreas das regiões contidas no círculo<sup>2</sup>. Podemos associar uma área para cada região. Como a densidade de probabilidade é a mesma para todos os pontos (tratando-se de uma v.a. de distribuição uniforme, portanto), temos que se duas regiões do círculo têm a mesma área então as probabilidades de ambas são iguais.

Para calcular a probabilidade do aeroplano ter pousado dentro do triângulo equilátero, basta fazer a razão entre a área do triângulo, que aqui corresponde ao evento, pela área do círculo. Para tal, primeiro expressaremos ambas as áreas em função de uma mesma medida.

Sendo o circuncentro do triângulo equilátero também baricentro deste mesmo triângulo, e considerando que a distância do baricentro até um vértice qualquer equivale ao raio, ( $r$ ), do círculo, e que a distância do baricentro até o lado oposto ao vértice

<sup>2</sup> Desde que as intersecções entre as regiões sejam vazias e a reunião de todas constitua o círculo.

anteriormente considerado equivale a  $\frac{r}{2}$  (Figura 7), temos a altura ( $h$ ) do triângulo:

$$h = r + \frac{r}{2} = \frac{3}{2}r.$$

Sabendo que a fórmula para obter a medida da altura de um triângulo equilátero de lado  $l$  é dada por  $h = \frac{l\sqrt{3}}{2}$ , temos:

$$\begin{aligned} h &= \frac{3}{2}r \\ \implies \frac{l\sqrt{3}}{2} &= \frac{3}{2}r \\ \implies l &= \sqrt{3}r. \end{aligned}$$

A área do círculo ( $A_c$ ):  $A_c = \pi r^2$ .

A área do triângulo equilátero ( $A_t$ ):  $A_t = \frac{l^2\sqrt{3}}{4} = \frac{3r^2 \cdot \sqrt{3}}{4}$ .

Logo, podemos obter a probabilidade desejada:

$$P = \frac{A_t}{A_c} = \frac{\frac{3r^2 \cdot \sqrt{3}}{4}}{\pi r^2} = \frac{3\sqrt{3}}{4\pi} \approx 0,4135.$$

**Exemplo 1.6.** Um disco circular de raio 1 é jogado em uma tábua graduada com linhas paralelas separadas por 2 de distância, de modo que uma dessas linhas determina uma corda no disco circular. Obter a probabilidade de que o comprimento desta corda seja maior que o comprimento do lado de um triângulo equilátero inscrito no disco.

Obs: A espessura do disco é desprezível. As massas, tanto do círculo, quanto da tábua, estão igualmente distribuídas.

Primeiro, vamos representar uma situação possível (Figura 8). Notando que há infinitas cordas que podem ser determinadas no disco pelas linhas.

Em um triângulo equilátero inscrito em um círculo de raio 1 (Figura 9), a distância do centro do círculo até um dos lados ( $l$ ) do triângulo é  $\frac{1}{2}$  (pela propriedade do baricentro do triângulo equilátero, ver o exemplo anterior).

Podemos comparar o tamanho da corda, em relação a  $l$ , a partir da sua distância ( $d$ ) ao centro do círculo. Note que  $0 < d < 1$ . Se  $0 < d < \frac{1}{2}$  a corda será maior que  $l$  (Figura 10). Se  $\frac{1}{2} < d < 1$  a corda será menor que  $l$  (Figura 11). Não há porque privilegiar qualquer valor de  $d$  em relação aos demais; qualquer valor de  $d$ , em  $[0, 1]$ , é igualmente provável, configurando esta v.a. como de distribuição uniforme.

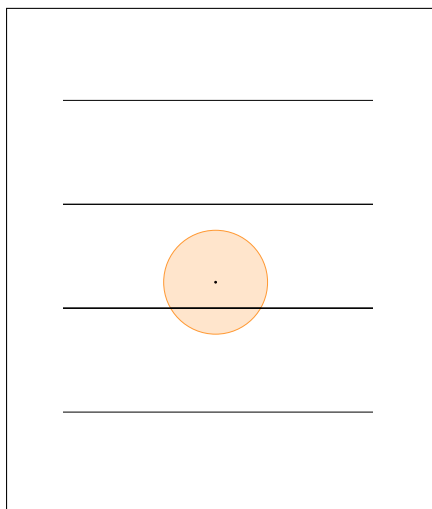


Figura 8: Disco circular sobre a tábua graduada.

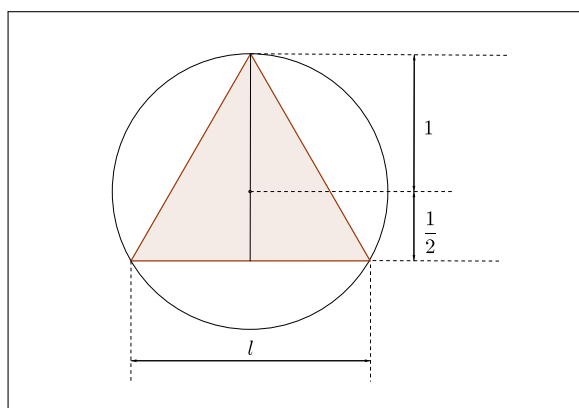


Figura 9: Triângulo equilátero inscrito em um círculo de raio 1.

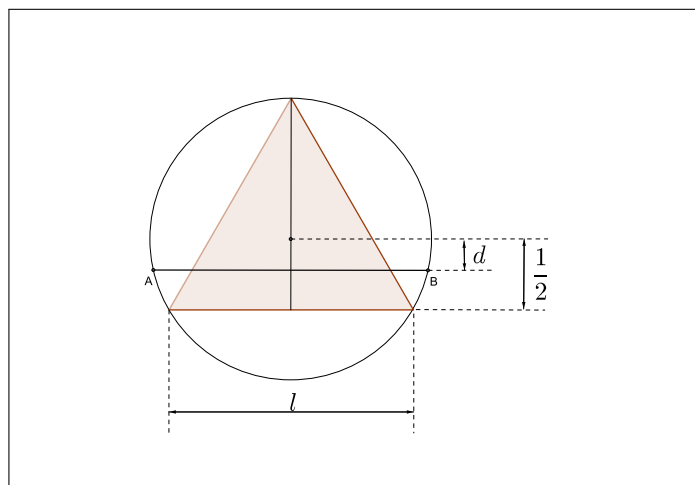


Figura 10: Corda AB com distância para o centro do círculo inferior a  $\frac{1}{2}$ .



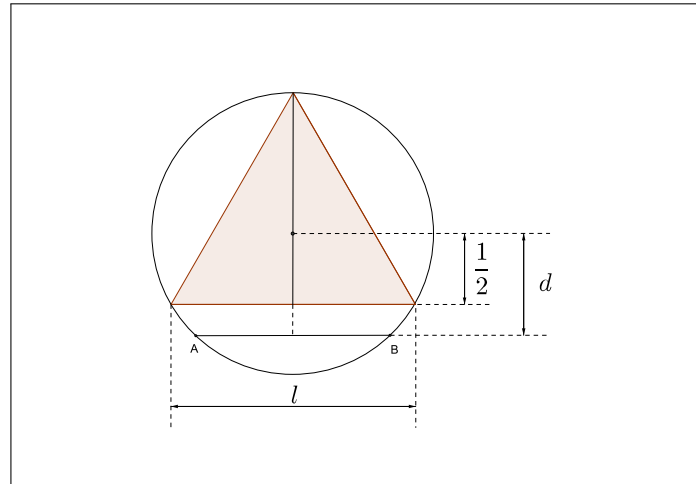


Figura 11: Corda AB com distância para o centro do círculo superior a  $\frac{1}{2}$ .

Logo, de todas as distâncias possíveis da corda ao centro:  $0 < d < 1$ , cuja amplitude do intervalo é 1, as favoráveis são aquelas em que  $0 < d < \frac{1}{2}$ , em que a amplitude do intervalo é  $\frac{1}{2}$ . Assim, a probabilidade da corda ser maior que  $l$  é:

$$P = \frac{1/2}{1} = \frac{1}{2}.$$



---

## ÁREA SOB A CURVA

---

No capítulo anterior, foram vistos alguns conceitos básicos de estatística. Dentre os conceitos estudados estava o de função de distribuição de probabilidade, para variáveis aleatórias (v.a.) discretas; e função densidade de probabilidade, que se aplica a v.a. contínuas, quando delas queremos obter, para determinado intervalo, a sua probabilidade. Tal probabilidade é determinada pela área sob a curva da referida função, que, na maioria dos casos, é obtida por meio do cálculo integral. Esbarramos, então, no problema de que integrais não são estudadas no ensino médio e foi prometido, para este presente capítulo, a dedução de fórmulas para obter a área sob a curva de certas funções não retilíneas, sem a utilização de integrais <sup>1</sup>.

Como comentamos anteriormente, estamos interessados em estudar variáveis aleatórias que obedecem leis de potência, ou seja, para  $X$  discreta, queremos:

$$P(x = k) = c_1 \cdot k^{-\alpha}, \quad (\text{função distribuição de probabilidade}) \quad (2.1)$$

e, para  $X$  contínua:

$$f(x) = c_2 \cdot x^{-\alpha}, \quad (\text{função densidade de probabilidade}). \quad (2.2)$$

Lembre-se que, para que (2.1) e (2.2) estejam bem definidas precisamos:

$$c_1 \cdot \sum_{k=1}^{+\infty} k^{-\alpha} = 1 \quad (\text{no caso de (2.1)}) \quad (2.3)$$

---

<sup>1</sup> Discussões sobre áreas sob curvas são normalmente realizadas em livros de Cálculo, como uma introdução ao estudo do Cálculo Integral. Ideias, como a aproximação da área sob uma curva por meio da utilização de retângulos, exploradas neste trabalho, podem ser estudadas, inclusive sob aspectos diferentes, nestes livros. Como uma referência de algo mais próximo ao que desenvolvemos aqui, indicamos Rogério, Badan e Silva (1994, Capítulo 8) [5].

e

$$A_{-\alpha}[x_{\min}, \infty] = 1 \quad (\text{no caso de (2.2)}). \quad (2.4)$$

Em (2.4), está escrito que a área sob a curva da função  $f(x)$ , a partir de um  $x$  mínimo e com  $x$  muito grande (tendendo ao infinito), ou seja, em toda a sua extensão, é igual a 1.

Neste capítulo, vamos estudar maneiras de calcular tais quantidades. Nos concentraremos, principalmente, no cálculo das áreas; isto porque, a soma de (2.3) está intimamente relacionada com a área de (2.4). Para ver isso, observe a Figura 12 .

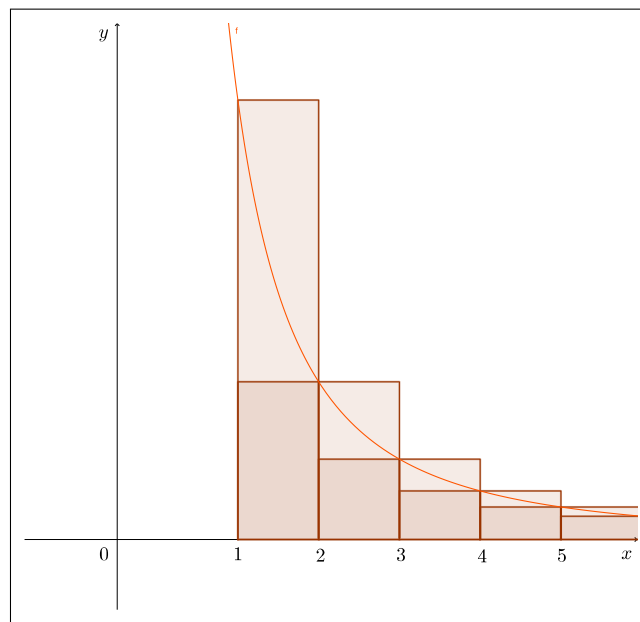


Figura 12: Área sob a curva de  $f(x) = x^{-\alpha}$  cercada por retângulos inferiores e superiores.

Conforme podemos notar na Figura 12:

$$\begin{aligned} c_1 \cdot 2^{-\alpha} + c_1 \cdot 3^{-\alpha} + \dots &\leq A_{-\alpha}[x_{\min}, \infty] \leq c_1 \cdot 1 + c_1 \cdot 2^{-\alpha} + c_1 \cdot 3^{-\alpha} + \dots \\ c_1 \cdot (2^{-\alpha} + 3^{-\alpha} + \dots) &\leq A_{-\alpha}[x_{\min}, \infty] \leq c_1 \cdot (1 + 2^{-\alpha} + 3^{-\alpha} + \dots). \end{aligned}$$

Note que, no membro esquerdo da inequação está a soma das áreas dos retângulos inferiores; e, no membro direito da inequação, a soma das áreas dos retângulos superiores. A área sob a curva de  $f(x)$  está entre ambas as somas citadas.

Para maior facilidade, antes de tratarmos das funções  $f(x) = cx^m$ ,  $x \geq 0$  e  $c > 0$ , começaremos por  $f(x) = x^m$ ,  $x \geq 0$  ( $c$  igual a 1) . Para o cálculo da área sob  $f(x) = x^m$ ,

num intervalo positivo determinado no eixo das abscissas por  $s$  e  $t$  (intervalo  $[s, t]$ ), dividiremos o problema em três casos: Para  $m \geq 0$ ,  $m < 0$  (exceto  $m = -1$ ) e  $m = -1$ . O terceiro caso:  $m = -1$  é tratado separadamente devido ao fato dele produzir uma indeterminação quando a ele aplicamos a fórmula obtida nos casos anteriores. Quanto a este terceiro caso, nos limitaremos a mostrar que a área sob sua curva não converge a um número, mas diverge, quando fazemos  $x$  muito grande.

Denotaremos por  $A_m[s, t]$  a área sob a curva de  $f(x) = x^m$  no intervalo  $[s, t]$  (Figura 13).

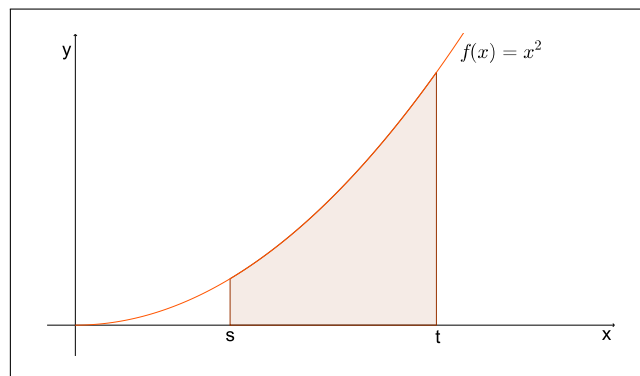


Figura 13: Área sob a curva de  $f(x) = x^2$  no intervalo  $[s, t]$ .

Adotando  $0 < s < r < t$ , a seguinte propriedade será muito utilizada (Figura 2):

$$A_m[s, t] = A_m[s, r] + A_m[r, t].$$

E, em consequência:

$$A_m[r, t] = A_m[s, t] - A_m[s, r]. \tag{2.5}$$

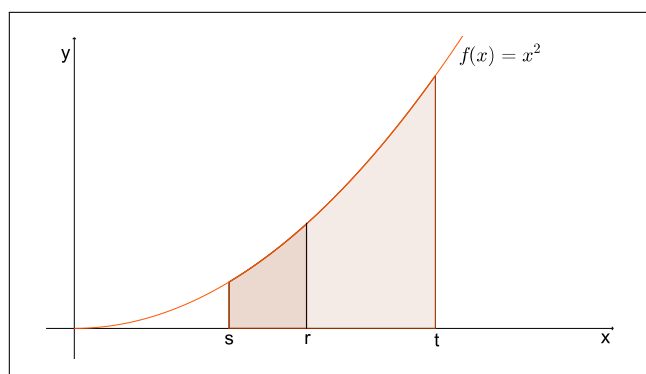


Figura 14: Área sob a curva de  $f(x) = x^2$  nos intervalos  $[s, r]$ ,  $[r, t]$ ,  $[s, t]$ .

Deste modo, basta saber calcular  $A_m[s, t]$  para um valor escolhido de  $s$  e poderemos obter a área para outros intervalos.

2.1 ÁREA SOB A CURVA - 1º CASO:  $f(x) = x^m$ , COM  $m \geq 0$

Faremos  $s = 0$  em  $A_m[s, t]$ . Como vimos anteriormente, sabendo calcular a área para um  $s$  escolhido, somos capazes de calcular  $A_m[r, t]$ , com  $s \leq r < t$ . Adotando  $s = 0$ , qualquer área poderá ser calculada posteriormente, nas condições impostas.

Para simplificar a notação, consideraremos  $A_m[0, t] = A_m(t)$ .

Começaremos analisando individualmente o que ocorre para  $m = 0$ ,  $m = 1$ ,  $m = 2$  e  $m = 3$ . A generalização para  $m \geq 0$  está desenvolvida no Apêndice A.1.

2.1.1 1º Caso com  $f(x) = x^0$ ,  $x \geq 0$

Para  $m = 0$ , como  $x^0 = 1$ , a região considerada é um retângulo de lados  $t$  e 1 (Figura 15) e, portanto:  $A_0(t) = t \cdot 1 = t$ .

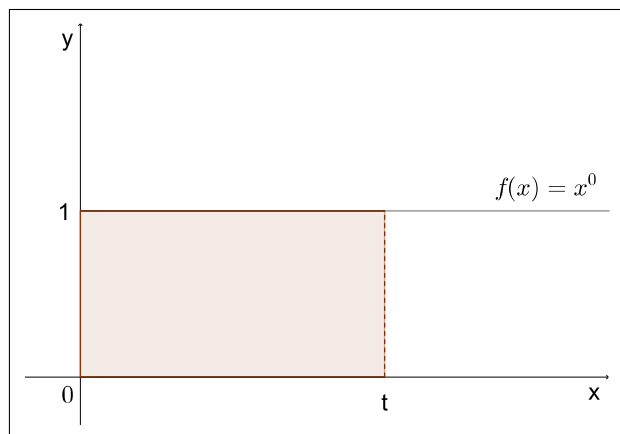


Figura 15: Área sob a curva de  $f(x) = x^0$  em  $[0, t]$ .

2.1.2 1º Caso com  $f(x) = x^1$ ,  $x \geq 0$

Para  $m = 1$ , a região delimitada é um triângulo retângulo com base  $t$  e altura  $t$  (Figura 16) e, portanto:  $A_1(t) = \frac{t \cdot t}{2} = \frac{t^2}{2}$ .

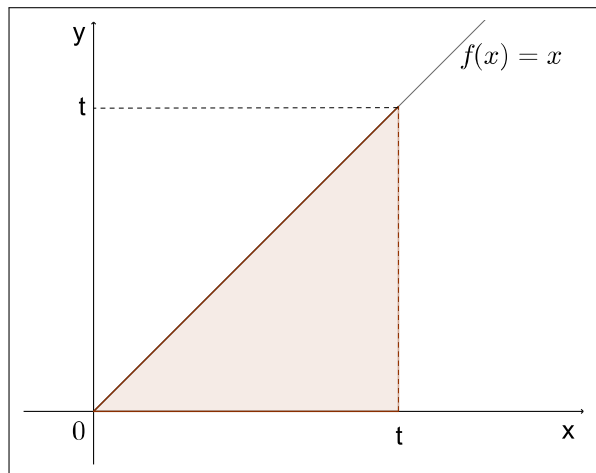


Figura 16: Área sob a curva de  $f(x) = x$  em  $[0, t]$ .

2.1.3 1º Caso com  $f(x) = x^2$ ,  $x \geq 0$

Para  $m = 2$ , os cálculos se complicam. Neste caso tem-se  $f(x) = x^2$  e o gráfico corresponde à parábola da Figura 17, com a área desejada em destaque.

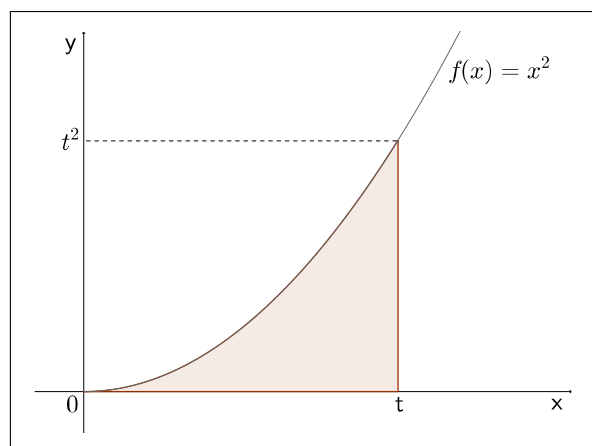


Figura 17: Área sob a curva de  $f(x) = x^2$  em  $[0, t]$ .

Como a área sob uma curva, não linear, é algo que foge do conteúdo do ensino médio, a aproximaremos com a utilização de  $n$  retângulos (Figura 18).

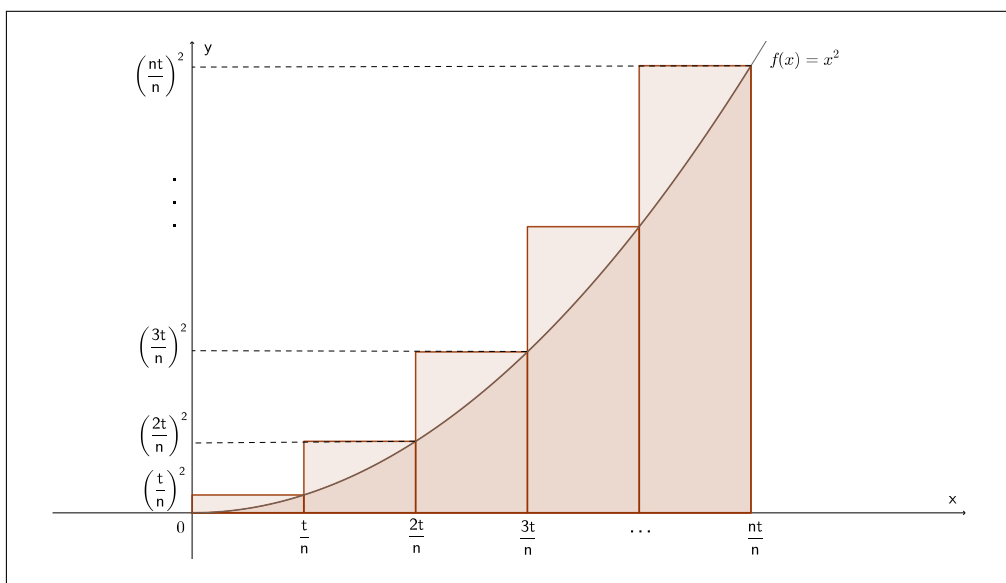


Figura 18: Área sob a curva de  $f(x) = x^2$  aproximada por  $n$  retângulos.

Cada retângulo tem como medida da base:  $\frac{t-0}{n} = \frac{t}{n}$ . E as alturas dos  $n$  retângulos em ordem crescente, do primeiro ao último, correspondem a:  $\left(\frac{t}{n}\right)^2$ ,  $\left(\frac{2t}{n}\right)^2$ ,  $\left(\frac{3t}{n}\right)^2$ , ...,  $\left(\frac{nt}{n}\right)^2$ . Note que  $f\left(\frac{t}{n}\right) = \left(\frac{t}{n}\right)^2$ ,  $f\left(\frac{2t}{n}\right) = \left(\frac{2t}{n}\right)^2$ , ...,  $f\left(\frac{nt}{n}\right) = \left(\frac{nt}{n}\right)^2$  (Figura 18).

Como pode ser percebido, a soma das áreas destes retângulos é superior à área sob a curva. Pode-se notar, também, que à medida que aumentamos o número de retângulos (valor de  $n$ ), a soma das áreas se aproxima da área procurada (Figuras de 19 a 24). Dizemos que, para  $n$  muito grande ( $n$  tendendo ao infinito), a soma das áreas dos retângulos é igual a área sob a curva.

Denotaremos a soma das áreas dos  $n$  retângulos por  $R_{m,n}(t)$ , em que:  $m$  é o expoente de  $x$ ,  $n$  é o número de retângulos (cada um com base de medida  $\frac{1}{n}$ ) e  $t$  é o limite à direita do intervalo.



2.1 ÁREA SOB A CURVA - 1º CASO:  $f(x) = x^m$ , COM  $m \geq 0$

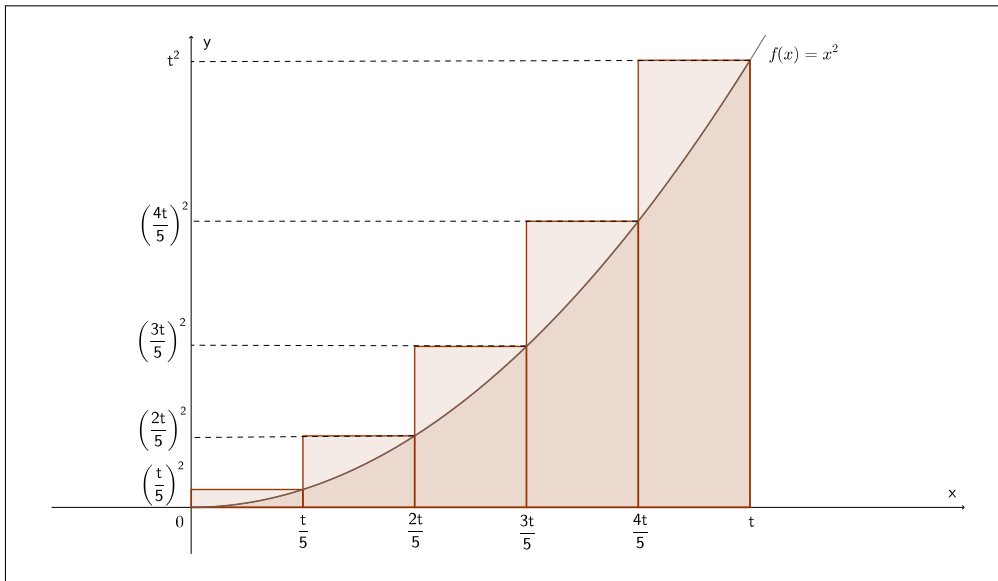


Figura 19: Área sob a curva em  $[0, t]$  cercada por 5 retângulos superiores.

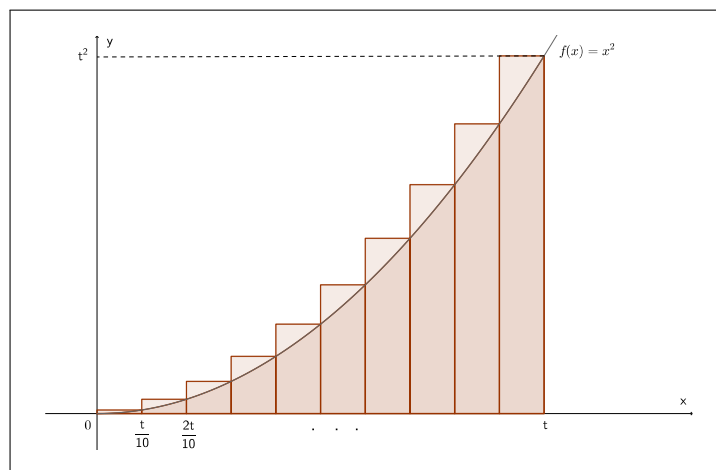


Figura 20: Área sob a curva em  $[0, t]$  cercada por 10 retângulos superiores.

Deste modo, para  $m = 2$  podemos escrever:

$$R_{2,n}(t) = \frac{t}{n} \cdot \left(\frac{t}{n}\right)^2 + \frac{t}{n} \cdot \left(\frac{2t}{n}\right)^2 + \frac{t}{n} \cdot \left(\frac{3t}{n}\right)^2 + \dots + \frac{t}{n} \cdot \left(\frac{nt}{n}\right)^2.$$

Colocando  $\left(\frac{t}{n}\right)^3$  em evidência:

$$R_{2,n}(t) = \left(\frac{t}{n}\right)^3 \cdot (1^2 + 2^2 + 3^2 + \dots + n^2).$$

## ÁREA SOB A CURVA

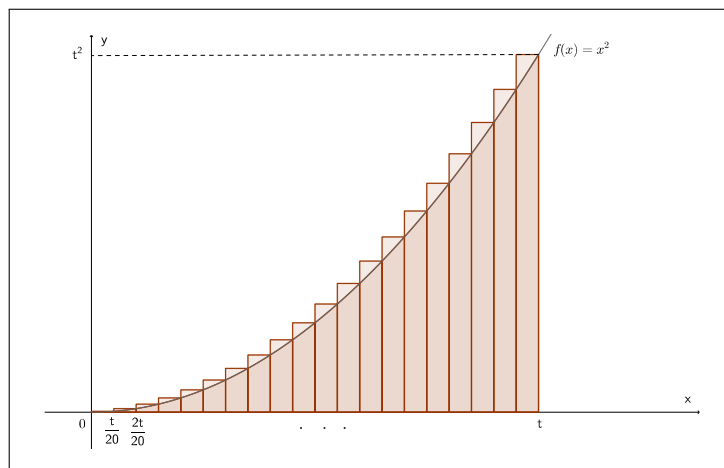


Figura 21: Área sob a curva em  $[0, t]$  cercada por 20 retângulos superiores.

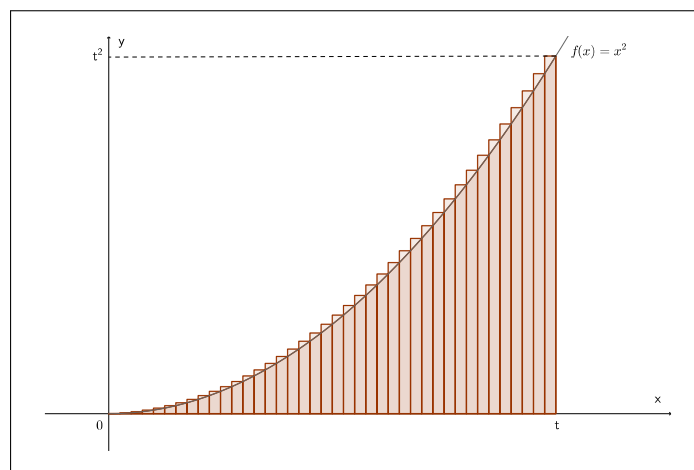


Figura 22: Área sob a curva em  $[0, t]$  cercada por 40 retângulos superiores.

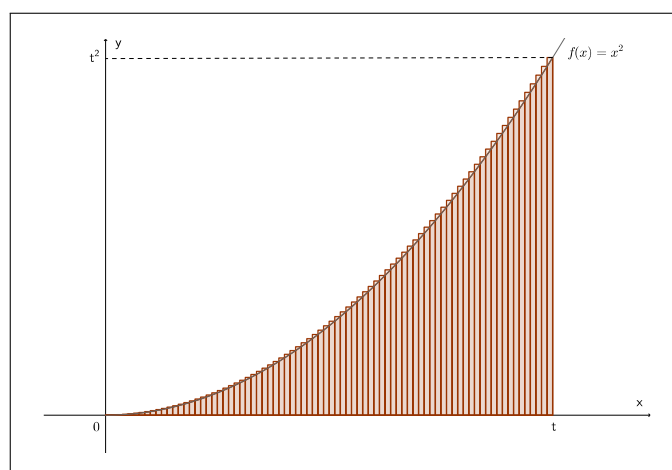


Figura 23: Área sob a curva em  $[0, t]$  cercada por 80 retângulos superiores.

2.1 ÁREA SOB A CURVA - 1º CASO:  $f(x) = x^m$ , COM  $m \geq 0$

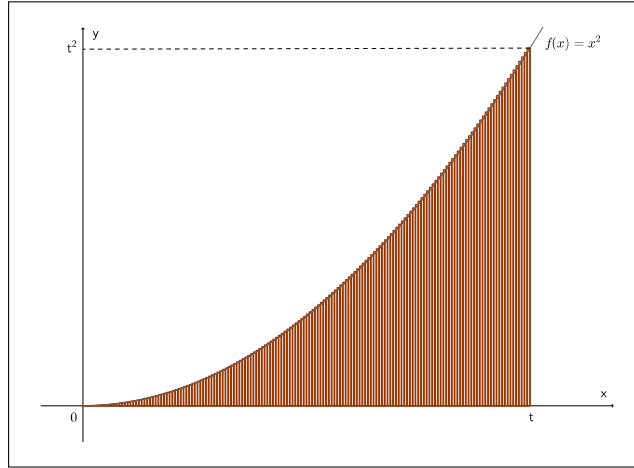


Figura 24: Área sob a curva em  $[0, t]$  cercada por 160 retângulos superiores.

Vamos definir  $S_k(n) = 1^k + 2^k + 3^k + \dots + n^k$ . Eventualmente, para simplificar a notação, deixaremos de explicitar a dependência em  $n$ , notando  $S_k(n)$  apenas por  $S_k$ .

Considerando  $S_2(n) = 1^2 + 2^2 + 3^2 + \dots + n^2$ , tem-se:

$$R_{2,n}(t) = \left(\frac{t}{n}\right)^3 \cdot (1^2 + 2^2 + 3^2 + \dots + n^2) = \frac{S_2(n) \cdot t^3}{n^3}. \quad (2.6)$$

Para  $n$  muito grande,  $R_{2,n}(t) \approx A_2(t)$ . E, em particular, para  $t = 1$  e  $n$  muito grande:

$$\frac{S_2(n) \cdot 1^3}{n^3} = R_{2,n}(1) \approx A_2(1). \quad (2.7)$$

Será importante notar que para  $k = 1, 2$ :

$$\frac{S_{2-k}(n)}{n^3} = \frac{1}{n^k} \cdot \frac{S_{2-k}(n)}{n^{3-k}} = \frac{1}{n^k} \cdot \frac{S_{2-k}(n)}{n^{(2-k)+1}}. \quad (2.8)$$

Em que  $S_{2-k}(n) = 1^{2-k} + 2^{2-k} + 3^{2-k} + \dots + n^{2-k}$ .

Quanto mais aumentarmos o valor de  $n$  mais próximo de zero estará o valor de  $\frac{1}{n^k}$ , sendo assim, para  $n$  muito grande, considera-se que o valor de  $\frac{1}{n^k}$  é nulo. Para  $n$  muito grande  $\frac{S_{2-k}(n)}{n^{(2-k)+1}}$  é aproximadamente  $A_{2-k}(1)$ . Como o produto de zero por um número é igual a zero, temos que, para  $n$  muito grande, o produto acima, (2.8), é nulo.

Temos pelo binômio de Newton:

$$(a + 1)^3 = a^3 + 3a^2 + 3a + 1.$$

Variando  $a$ , entre 0 e  $n$ , temos:

$$\begin{aligned}
 (0+1)^3 &= 0^3 + 3 \cdot 0^2 + 3 \cdot 0 + 1^3 \\
 (1+1)^3 &= 1^3 + 3 \cdot 1^2 + 3 \cdot 1 + 1^3 \\
 (2+1)^3 &= 2^3 + 3 \cdot 2^2 + 3 \cdot 2 + 1^3 \\
 (3+1)^3 &= 3^3 + 3 \cdot 3^2 + 3 \cdot 3 + 1^3 \\
 &\quad \cdot \\
 &\quad \cdot \\
 &\quad \cdot \\
 (n+1)^3 &= n^3 + 3 \cdot n^2 + 3 \cdot n + 1^3.
 \end{aligned}$$

Somando os membros correspondentes das equações de todas as linhas acima, obtemos<sup>2</sup>:

$$(n+1)^3 = 3 \cdot (1^2 + 2^2 + \dots + n^2) + 3 \cdot (1 + 2 + \dots + n) + n + 1.$$

Denotando  $S_1(n) = 1 + 2 + \dots + n$ ,  $S_2(n) = 1^2 + 2^2 + \dots + n^2$  e dividindo ambos os membros por  $n^3$ :

$$\frac{(n+1)^3}{n^3} = 3 \cdot \frac{S_2(n)}{n^3} + 3 \cdot \frac{S_1(n)}{n^3} + \frac{n}{n^3} + \frac{1}{n^3}.$$

Donde segue:

$$\left(1 + \frac{1}{n}\right)^3 = 3 \cdot R_{2,n}(1) + 3 \cdot \frac{S_1(n)}{n^3} + \frac{1}{n^2} + \frac{1}{n^3}.$$

Para  $n$  muito grande as parcelas:  $\frac{1}{n}$ ,  $3 \cdot \frac{S_1(n)}{n^3}$ ,  $\frac{1}{n^2}$  e  $\frac{1}{n^3}$  anulam-se e  $R_{2,n}(1)$  torna-se aproximadamente  $A_2(1)$ , obtendo-se assim a equação:

$$1 = 3 \cdot A_2(1) \implies A_2(1) = \frac{1}{3}.$$

Multiplicando-se ambos os membros por  $t^3$ :

$$A_2(1) \cdot t^3 = \frac{1}{3} \cdot t^3.$$

Aplicando a equação (2.7), na equação acima, para  $n$  grande, tem-se:

$$\frac{S_2(n)}{n^3} \cdot t^3 = \frac{t^3}{3}.$$

<sup>2</sup> Note que, ao somarmos, os cubos dos membros esquerdos das equações, com exceção de  $(n+1)^3$ , são cancelados com cubos dos membros direitos das equações seguintes.

Aplicando as equações (2.6) e (2.7) na equação acima:

$$A_2(t) = \frac{t^3}{3}.$$

2.1.4 1º Caso com  $f(x) = x^3$ ,  $x > 0$

Para  $m = 3$ , os passos serão análogos ao caso anterior ( $m = 2$ ). Aproximaremos a área sob a curva de  $f(x) = x^3$ , no intervalo  $[0, t]$ , por  $n$  retângulos de base  $\frac{t}{n}$  e alturas equivalentes a  $\left(\frac{t}{n}\right)^3$ ,  $\left(\frac{2t}{n}\right)^3$ ,  $\left(\frac{3t}{n}\right)^3$ , ...,  $\left(\frac{nt}{n}\right)^3$  (Figura 25).

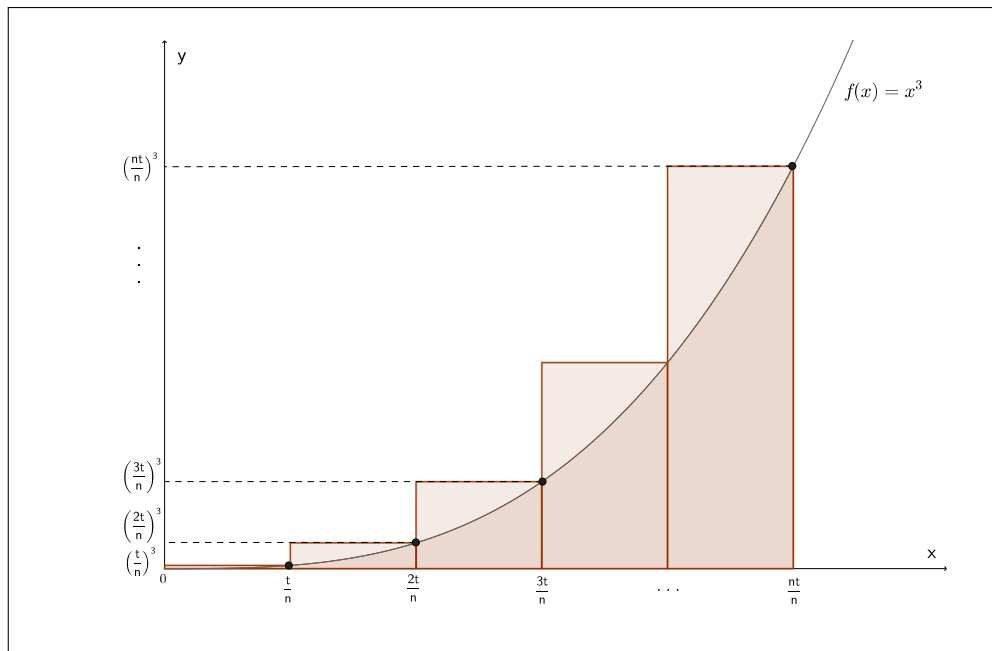


Figura 25: Área sob a curva de  $f(x) = x^3$ , em  $[0, t]$ , cercada por  $n$  retângulos superiores.

Denotando por  $R_{3,n}(t)$  a soma das áreas dos  $n$  retângulos, temos:

$$R_{3,n}(t) = \frac{t}{n} \cdot \left(\frac{t}{n}\right)^3 + \frac{t}{n} \cdot \left(\frac{2t}{n}\right)^3 + \frac{t}{n} \cdot \left(\frac{3t}{n}\right)^3 + \dots + \frac{t}{n} \cdot \left(\frac{nt}{n}\right)^3.$$

Colocando  $\left(\frac{t}{n}\right)^4$  em evidência:

$$R_{3,n}(t) = \left(\frac{t}{n}\right)^4 \cdot (1^3 + 2^3 + 3^3 + \dots + n^3).$$

Considerando  $S_3(n) = 1^3 + 2^3 + 3^3 + \dots + n^3$ , tem-se:

$$R_{3,n}(t) = \left(\frac{t}{n}\right)^4 \cdot (1^3 + 2^3 + 3^3 + \dots + n^3) = \frac{S_3(n) \cdot t^4}{n^4}.$$

Para  $n$  muito grande e  $t = 1$ :

$$\frac{S_3(n) \cdot 1^4}{n^4} = R_{3,n}(1) \approx A_3(1). \quad (2.9)$$

Temos pelo binômio de Newton:

$$(a + 1)^4 = a^4 + 4a^3 + 6a^2 + 4a + 1^4.$$

Variando  $a$ , entre 0 e  $n$ , temos:

$$\begin{aligned} (0 + 1)^4 &= 0^4 + 4 \cdot 0^3 + 6 \cdot 0^2 + 4 \cdot 0 + 1 \\ (1 + 1)^4 &= 1^4 + 4 \cdot 1^3 + 6 \cdot 1^2 + 4 \cdot 1 + 1 \\ (2 + 1)^4 &= 2^4 + 4 \cdot 2^3 + 6 \cdot 2^2 + 4 \cdot 2 + 1 \\ (3 + 1)^4 &= 3^4 + 4 \cdot 3^3 + 6 \cdot 3^2 + 4 \cdot 3 + 1 \\ &\cdot \\ &\cdot \\ &\cdot \\ (n + 1)^4 &= n^4 + 4 \cdot n^3 + 6 \cdot n^2 + 4 \cdot n + 1. \end{aligned}$$

Somando os membros correspondentes das equações de todas as linhas acima, obtemos<sup>3</sup>:

$$(n + 1)^4 = 4 \cdot (1^3 + 2^3 + \dots + n^3) + 6 \cdot (1^2 + 2^2 + \dots + n^2) + 4 \cdot (1 + 2 + \dots + n) + n + 1.$$

Denotando  $S_1(n) = 1 + 2 + \dots + n$ ,  $S_2(n) = 1^2 + 2^2 + \dots + n^2$ ,  $S_3(n) = 1^3 + 2^3 + \dots + n^3$  e dividindo ambos os membros por  $n^4$ :

$$\frac{(n + 1)^4}{n^4} = 4 \cdot \frac{S_3(n)}{n^4} + 6 \cdot \frac{S_2(n)}{n^4} + 4 \cdot \frac{S_1(n)}{n^4} + \frac{n}{n^4} + \frac{1}{n^4}.$$

Donde segue:

$$\left(1 + \frac{1}{n}\right)^4 = 4 \cdot R_{3,n}(1) + 6 \cdot \frac{S_2(n)}{n^4} + 4 \cdot \frac{S_1(n)}{n^4} + \frac{1}{n^3} + \frac{1}{n^4}.$$

<sup>3</sup> Note que, ao somarmos, as potências dos membros esquerdos das equações, com exceção de  $(n + 1)^4$ , são canceladas com potências à quarta dos membros direitos das equações seguintes.

2.2 2º CASO:  $f(x) = x^{-m}$ ,  $x > 0$ , COM  $m > 1$

Para  $n$  muito grande as parcelas:  $\frac{1}{n}$ ,  $6 \cdot \frac{S_2(n)}{n^4}$ ,  $4 \cdot \frac{S_1(n)}{n^4}$ ,  $\frac{1}{n^3}$  e  $\frac{1}{n^4}$  anulam-se e  $R_{3,n}(1)$  torna-se próximo a  $A_3(1)$ , obtendo-se assim a equação:

$$1 = 4 \cdot A_3(1) \implies A_3(1) = \frac{1}{4}.$$

Multiplicando-se ambos os membros por  $t^4$ :

$$A_3(1) \cdot t^4 = \frac{1}{4} \cdot t^4.$$

Da equação (2.9), para  $n$  grande, tem-se:

$$\frac{S_3(n)}{n^4} \cdot t^4 = \frac{t^4}{4}.$$

Donde segue:

$$A_3(t) = \frac{t^4}{4}.$$

Para  $m > 3$ , os resultados são análogos, de modo que podemos generalizar <sup>4</sup>:

$$\boxed{A_m(t) = \frac{t^{m+1}}{m+1}}. \quad (2.10)$$

Notando, que a equação (2.8) pode ser generalizada, com  $0 < k \leq m$ , como segue:

$$\frac{S_{m-k}(n)}{n^{m+1}} = \frac{1}{n^k} \cdot \frac{S_{m-k}(n)}{n^{(m+1)-k}} = \frac{1}{n^k} \cdot \frac{S_{m-k}(n)}{n^{(m-k)+1}}. \quad (2.11)$$

E, assim como argumentado após a equação (2.8), para  $n$  muito grande, o produto de (2.11) é nulo.

2.2 2º CASO:  $f(x) = x^{-m}$ ,  $x > 0$ , COM  $m > 1$

Neste 2º caso<sup>5</sup>, abordaremos o cálculo da área sob a curva de uma função do tipo  $f(x) = x^{-m}$ , com  $m > 1$ . Mostraremos que uma fórmula, decorrente da equação (2.10), do 1º caso, serve para esta presente situação. Veremos, nesta seção, os casos em que  $m = 2$  e  $m = 3$ ; uma generalização será feita, mas a sua demonstração foi deslocada para o Apêndice A.2.

<sup>4</sup> A demonstração encontra-se no Apêndice A.1.

<sup>5</sup> Note, que em vez de escrever  $f(x) = x^m$ , com  $m < -1$ , optamos por  $f(x) = x^{-m}$ , com  $m > 1$ . O fizemos para uma maior facilidade de escrita e compreensão.

Primeiramente, precisamos encontrar uma candidata à função que nos permita calcular a área sob a curva, em um determinado intervalo, de  $f(x) = x^{-m}$ , com  $m > 1$ . Podemos obter uma candidata partindo da equação (2.10):

$$A_m[0, t] = \frac{t^{m+1}}{m+1}.$$

Conforme vimos em (2.5):

$$\begin{aligned} A_m[s, t] &= A_m[0, t] - A_m[0, s] \\ \Rightarrow A_m[s, t] &= \frac{t^{m+1}}{m+1} - \frac{s^{m+1}}{m+1}. \end{aligned}$$

Temos, então, como hipótese:

$$\begin{aligned} A_{-m}[s, t] &= \frac{t^{-m+1}}{-m+1} - \frac{s^{-m+1}}{-m+1} \\ \Rightarrow A_{-m}[s, t] &= \frac{s^{-m+1}}{m-1} - \frac{t^{-m+1}}{m-1} \\ \Rightarrow A_{-m}[s, t] &= \frac{1}{m-1}(s^{-m+1} - t^{-m+1}) \\ \Rightarrow A_{-m}[s, t] &= \frac{1}{m-1} \left( \frac{1}{s^{m-1}} - \frac{1}{t^{m-1}} \right). \end{aligned} \quad (2.12)$$

A função da equação (2.12) será, então, a nossa candidata. Verificaremos em relação aos casos em que  $m = 2$  e  $m = 3$ , neste capítulo, e, o caso geral, no Apêndice A.2.

### 2.2.1 2º Caso com $f(x) = x^{-2}$ , $x > 0$

Inicialmente, trataremos do caso em que  $m = 2$  o que resulta em  $f(x) = x^{-2} = \frac{1}{x^2}$ .

Consideraremos, sob a curva, a área limitada ao intervalo  $[a, b]$ , com  $0 < a < b$  (Figura 26). Note que o valor desta área é superior a área do retângulo de base  $b - a$  e altura  $\frac{1}{b^2}$  e inferior a área do retângulo de base  $b - a$  e altura  $\frac{1}{a^2}$  (Figura 27). As áreas de retângulos, nestas condições, serão denotadas por  $r[a, b]$  e  $R[a, b]$ , respectivamente. Tem-se:

$$\begin{aligned} r[a, b] &= \frac{1}{b^2} \cdot (b - a), \\ R[a, b] &= \frac{1}{a^2} \cdot (b - a). \end{aligned}$$



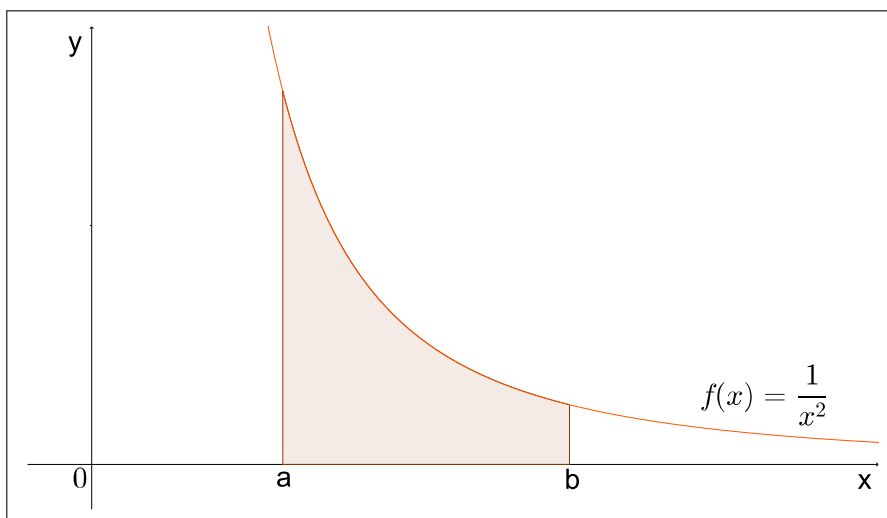


Figura 26: Área sob a curva de  $f(x) = \frac{1}{x^2}$ , em  $[a, b]$ .

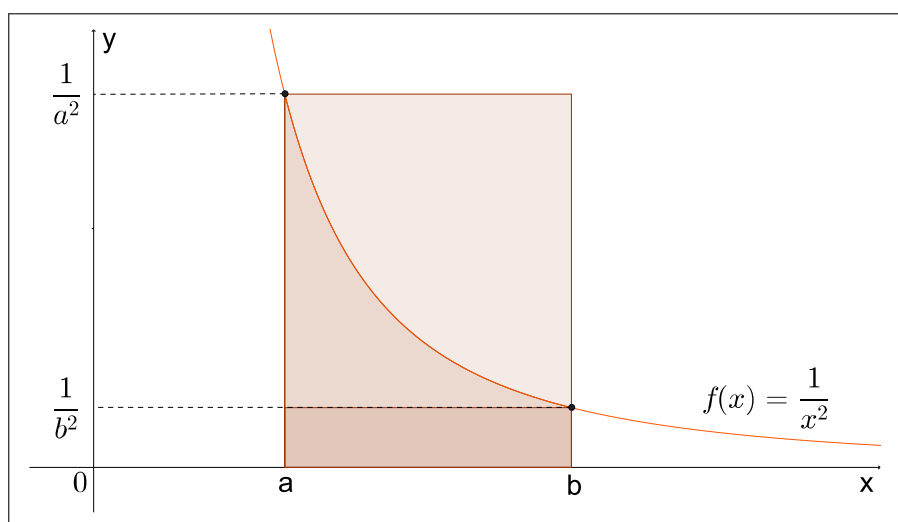


Figura 27: Área sob a curva de  $f(x) = \frac{1}{x^2}$ , em  $[a, b]$ , cercada por dois retângulos.

No 1º caso cercamos a área sob a curva por meio de  $n$  retângulos (Figuras 18 e 25). A soma das áreas destes retângulos, embora superior a área desejada, se aproximava do valor da área procurada, na medida que aumentávamos o número de retângulos (valor de  $n$ ). Neste 2º caso, a área sob a curva será cercada por meio de  $n$  retângulos cuja soma das áreas é inferior a área sob a curva (Figura 28), tal soma será denotada por  $r_n[a, b]$ , e por  $n$  retângulos cuja soma das áreas é superior a área sob a curva (Figura 29), e neste caso esta soma será denotada por  $R_n[a, b]$ . Tanto para  $r_n[a, b]$ , quanto para  $R_n[a, b]$ , para  $n$  muito grande, temos que ambas as somas convergem e se

aproximam de  $A_{-2}[a, b]$  (analogamente ao que foi observado no 1º caso, na sequência das Figuras 19 a 24).

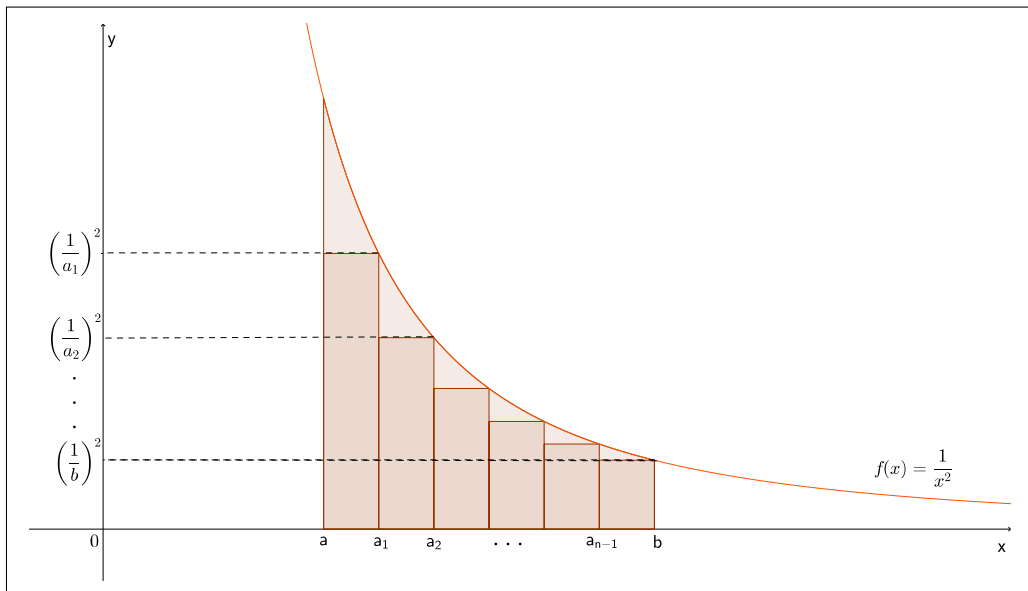


Figura 28: Área sob a curva de  $f(x) = \frac{1}{x^2}$ , em  $[a, b]$ , cercada por retângulos inferiores.

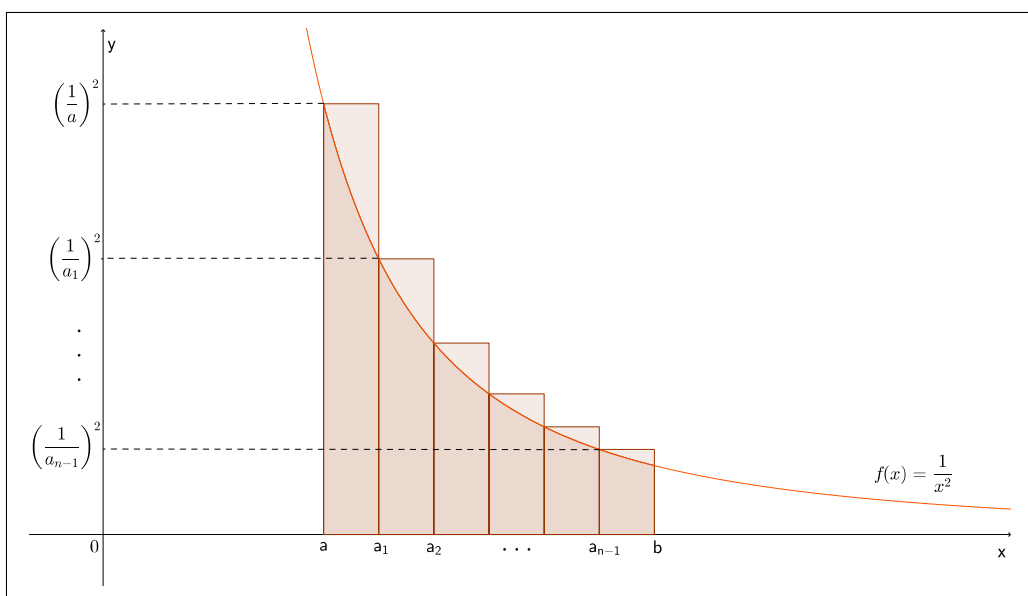


Figura 29: Área sob a curva de  $f(x) = \frac{1}{x^2}$ , em  $[a, b]$ , cercada por retângulos superiores.

Consideraremos a função  $g(t) = \left(\frac{1}{s} - \frac{1}{t}\right)$ , obtida da candidata (2.12), quando fazemos  $m = 2$ . Queremos demonstrar que esta função corresponde a área  $A_{-2}[a, b]$ ,

fazendo  $s = a$  e  $t = b$ . Para tal feito, demonstraremos que  $g(t)$  é sempre superior, ou igual, a  $r_n[a, b]$  e sempre inferior, ou igual, a  $R_n[a, b]$  (considerando  $s = a$  e  $t = b$ ). Como para  $n$  muito grande as somas  $r_n[a, b]$  e  $R_n[a, b]$  convergem a  $A_{-2}[a, b]$ , dado que  $g(t)$  estará “espremido” por ambas as somas,  $g(t) = A_{-2}[a, b]$ .

Para  $g(t)$ , tem-se:

$$g(b) - g(a) = \left(\frac{1}{s} - \frac{1}{b}\right) - \left(\frac{1}{s} - \frac{1}{a}\right) = \frac{1}{a} - \frac{1}{b} = \frac{b-a}{ab} = \frac{1}{ab} \cdot (b-a).$$

Como  $b > a$  e  $\frac{1}{b} < \frac{1}{a}$ , tem-se:

$$\begin{aligned} \frac{1}{b^2} \cdot (b-a) &\leq \frac{1}{ab} \cdot (b-a) \leq \frac{1}{a^2} \cdot (b-a) \\ \Rightarrow \frac{1}{b^2} \cdot (b-a) &\leq g(b) - g(a) \leq \frac{1}{a^2} \cdot (b-a) \\ \Rightarrow r[a, b] &\leq g(b) - g(a) \leq R[a, b]. \end{aligned} \tag{2.13}$$

Na função  $g(t)$ , faremos  $s = 1$  apenas para facilitar a compreensão, sem perda de generalidade. Lembrando que tal escolha, em relação ao intervalo  $[a, b]$ , da área, equivale a fazer  $a = 1$ .

$$g(t) = 1 - \frac{1}{t}. \tag{2.14}$$

Não esquecendo que:

$$g(b) - g(a) = \frac{1}{a} - \frac{1}{b}.$$

E, cercado a área sob a curva, em  $[1, b]$ , por meio de  $n$  retângulos inferiores e por meio de  $n$  retângulos superiores, com bases correspondentes aos intervalos  $[1, a_1]$ ,  $[a_1, a_2]$ ,  $[a_2, a_3]$ , ...,  $[a_{n-1}, b]$ , sendo as alturas dos inferiores:  $\left(\frac{1}{a_1}\right)^2$ ,  $\left(\frac{1}{a_2}\right)^2$ ,  $\left(\frac{1}{a_3}\right)^2$ , ...,  $\left(\frac{1}{b}\right)^2$ , respectivamente (Figura 28 com  $a = 1$ ). E as alturas dos superiores:  $1$ ,  $\left(\frac{1}{a_1}\right)^2$ ,  $\left(\frac{1}{a_2}\right)^2$ , ...,  $\left(\frac{1}{a_{n-1}}\right)^2$ , respectivamente (Figura 29 com  $a = 1$ ). Utilizando as abscissas das limitações dos intervalos das bases em  $g(t)$ , obtém-se:

$$\begin{aligned} [g(b) - g(a_{n-1})] + [g(a_{n-1}) - g(a_{n-2})] + [g(a_{n-2}) - g(a_{n-3})] + \dots + [g(a_1) - g(1)] \\ = g(b) - g(1) \\ = g(b). \end{aligned}$$

Note que na equação acima, utilizamos  $g(1) = 0$ , o que se segue da equação (2.14).

Com base em (2.13), consideremos as desigualdades:

$$\begin{aligned} r[1, a_1] &\leq g(a_1) - g(1) \leq R[1, a_1] \\ r[a_1, a_2] &\leq g(a_2) - g(a_1) \leq R[a_1, a_2] \\ &\cdot \\ &\cdot \\ &\cdot \\ r[a_{n-2}, a_{n-1}] &\leq g(a_{n-1}) - g(a_{n-2}) \leq R[a_{n-2}, a_{n-1}] \\ r[a_{n-1}, b] &\leq g(b) - g(a_{n-1}) \leq R[a_{n-1}, b]. \end{aligned}$$

Somando os membros correspondentes das desigualdades acima, obtém-se:

$$\begin{aligned} r_n[1, b] &\leq g(b) - g(1) \leq R_n[1, b] \\ \implies r_n[1, b] &\leq g(b) \leq R_n[1, b]. \end{aligned} \tag{2.15}$$

Em que:

$$\begin{aligned} r_n[1, b] &= r[1, a_1] + r[a_1, a_2] + \dots + r[a_{n-1}, b], \\ R_n[1, b] &= R[1, a_1] + R[a_1, a_2] + \dots + R[a_{n-1}, b]. \end{aligned}$$

Fazendo  $n$  muito grande na dupla desigualdade (2.15) temos que  $r_n[1, b]$  e  $R_n[1, b]$  convergem à  $A_{-2}[1, b]$ . Como  $g(b)$  está “ensanduichado” (espremido) por  $r_n[1, b]$  e  $R_n[1, b]$ , tem-se que  $g(b) = A_{-2}[1, b]$ , conforme queríamos mostrar. Embora tenhamos fixado  $a = 1$  em  $[a, b]$ , tal escolha foi feita apenas para facilitar a compreensão, conforme já dito; pode-se fazer  $a$  igual a qualquer valor maior que zero e substituir  $s$  por este mesmo valor em  $g(t)$ .

### 2.2.2 2º Caso com $f(x) = x^{-3}$ , $x > 0$

Analisaremos agora o valor da área sob a curva para  $m = 3$ , o que implica em  $f(x) = x^{-3} = \frac{1}{x^3}$ .

Analogamente ao que foi feito para  $m = 2$ , consideraremos a área sob a curva limitada ao intervalo  $[a, b]$ , com  $0 < a < b$  (Figura 30).

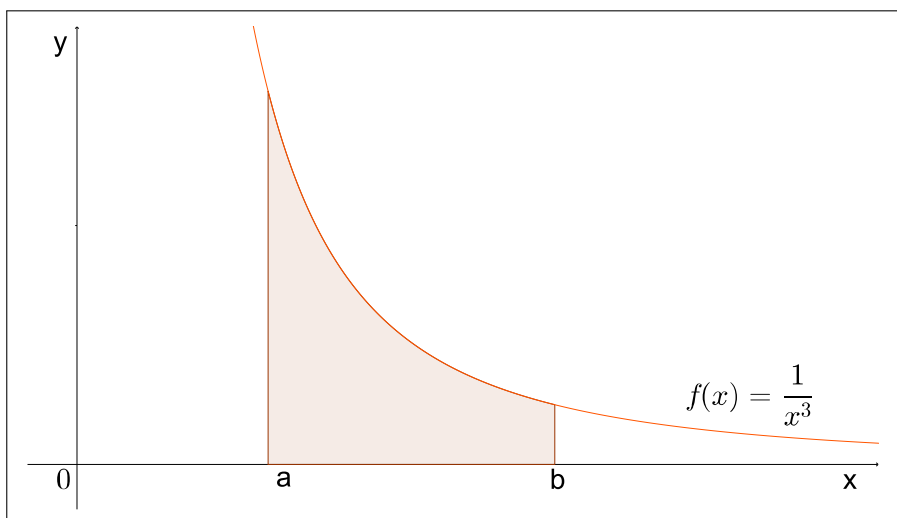


Figura 30: Área sob a curva de  $f(x) = \frac{1}{x^3}$ , em  $[a, b]$ .

Assim como feito anteriormente, denotaremos a área do retângulo inferior e do retângulo superior à curva (Figura 31) da seguinte maneira:

$$r[a, b] = \frac{1}{b^3} \cdot (b - a),$$

$$R[a, b] = \frac{1}{a^3} \cdot (b - a).$$

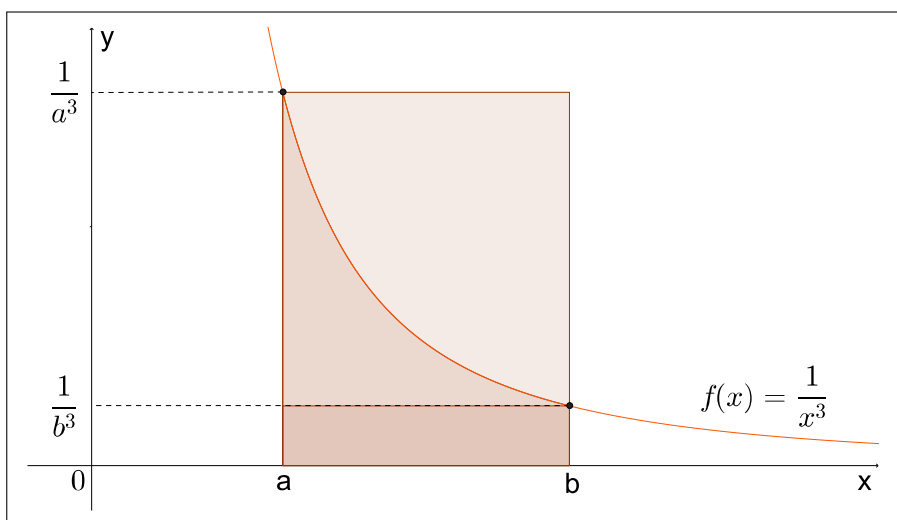


Figura 31: Área sob a curva de  $f(x) = \frac{1}{x^3}$ , em  $[a, b]$ , cercada por dois retângulos.

Como já realizado, cercaremos a área sob a curva por meio de  $n$  retângulos cuja soma das áreas é inferior a área sob a curva (Figura 32), esta soma será representada

por  $r_n[a, b]$ , e por  $n$  retângulos cuja soma das áreas é superior a área sob a curva (Figura 33), esta soma será representada por  $R_n[a, b]$ . Tanto para  $r_n[a, b]$ , quanto para  $R_n[a, b]$ , para  $n$  muito grande, temos que ambas as somas convergem à  $A_{-3}[a, b]$ .

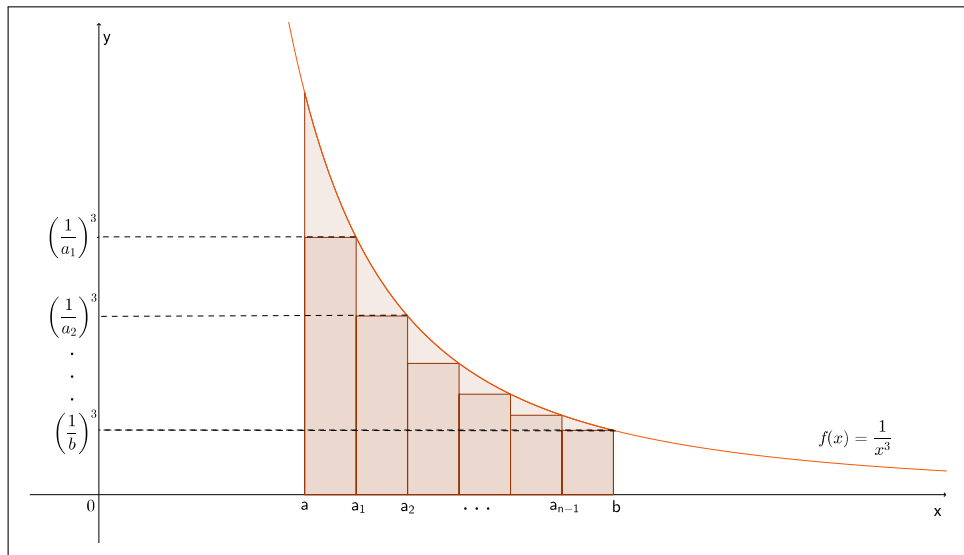


Figura 32: Área sob a curva de  $f(x) = \frac{1}{x^3}$ , em  $[a, b]$ , cercada por retângulos inferiores.

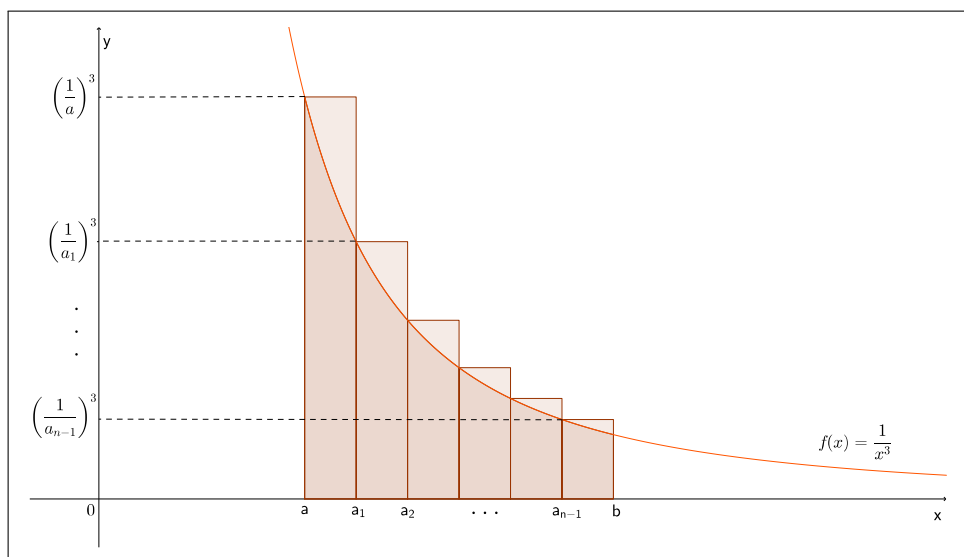


Figura 33: Área sob a curva de  $f(x) = \frac{1}{x^3}$ , em  $[a, b]$ , cercada por retângulos superiores.

Consideraremos a função  $g(t) = \frac{1}{2} \cdot \left( \frac{1}{s^2} - \frac{1}{t^2} \right)$ , obtida da candidata (2.12), quando fazemos  $m = 3$ . Queremos demonstrar que esta função corresponde a área  $A_{-3}[a, b]$ , fazendo  $s = a$  e  $t = b$ . Procederemos como no caso anterior, mostrando que  $g(t)$  é

sempre superior, ou igual, a  $r_n[a, b]$  e sempre inferior, ou igual, a  $R_n[a, b]$  e, como para  $n$  muito grande  $r_n[a, b]$  e  $R_n[a, b]$  convergem à  $A_{-3}[a, b]$ , isto implica em  $g(t) = A_{-3}[a, b]$ .

Para  $g(t)$ , tem-se:

$$\begin{aligned} g(b) - g(a) &= \frac{1}{2} \cdot \left( \frac{1}{s^2} - \frac{1}{b^2} \right) - \frac{1}{2} \cdot \left( \frac{1}{s^2} - \frac{1}{a^2} \right) \\ \Rightarrow g(b) - g(a) &= \frac{1}{2} \cdot \left( \frac{1}{a^2} - \frac{1}{b^2} \right) \\ \Rightarrow g(b) - g(a) &= \frac{1}{2} \cdot \left( \frac{1}{a} - \frac{1}{b} \right) \cdot \left( \frac{1}{a} + \frac{1}{b} \right) \\ \Rightarrow g(b) - g(a) &= \frac{1}{2} \cdot \left( \frac{b-a}{ab} \right) \cdot \left( \frac{1}{a} + \frac{1}{b} \right) \\ \Rightarrow g(b) - g(a) &= \frac{(b-a)}{2} \cdot \frac{1}{ab} \cdot \left( \frac{1}{a} + \frac{1}{b} \right). \end{aligned}$$

Como  $b > a$  e  $\frac{1}{b} < \frac{1}{a}$ , tem-se:

$$\begin{aligned} \frac{(b-a)}{2} \cdot \frac{1}{b^2} \cdot \frac{2}{b} &\leq \frac{(b-a)}{2} \cdot \frac{1}{ab} \cdot \left( \frac{1}{a} + \frac{1}{b} \right) \leq \frac{(b-a)}{2} \cdot \frac{1}{a^2} \cdot \frac{2}{a} \\ \Rightarrow \frac{(b-a)}{2} \cdot \frac{1}{b^2} \cdot \frac{2}{b} &\leq g(b) - g(a) \leq \frac{(b-a)}{2} \cdot \frac{1}{a^2} \cdot \frac{2}{a} \\ \Rightarrow \frac{b-a}{b^3} &\leq g(b) - g(a) \leq \frac{b-a}{a^3} \\ \Rightarrow r[a, b] &\leq g(b) - g(a) \leq R[a, b]. \end{aligned} \tag{2.16}$$

Na função  $g(t)$ , faremos  $s = 1$  para facilitar a compreensão, sem perda de generalidade. Em relação ao intervalo  $[a, b]$ , da área, isto equivale a fazer  $a = 1$ .

$$g(t) = \frac{1}{2} \cdot \left( 1 - \frac{1}{t^2} \right). \tag{2.17}$$

Lembrando que:

$$g(b) - g(a) = \frac{1}{2} \cdot \left( \frac{1}{a^2} - \frac{1}{b^2} \right).$$

E, cercando a área sob a curva, em  $[1, b]$ , por meio de  $n$  retângulos inferiores (Figura 32, com  $a = 1$ ), e por meio de  $n$  retângulos superiores (Figura 33, com  $a = 1$ ), com bases correspondentes aos intervalos  $[1, a_1]$ ,  $[a_1, a_2]$ ,  $[a_2, a_3]$ , ...,  $[a_{n-1}, b]$ , sendo as alturas dos inferiores:  $\left(\frac{1}{a_1}\right)^3$ ,  $\left(\frac{1}{a_2}\right)^3$ ,  $\left(\frac{1}{a_3}\right)^3$ , ...,  $\left(\frac{1}{b}\right)^3$ , respectivamente. E as alturas dos

superiores:  $1, \left(\frac{1}{a_1}\right)^3, \left(\frac{1}{a_2}\right)^3, \dots, \left(\frac{1}{a_{n-1}}\right)^3$ , respectivamente. Utilizando as abscissas das limitações dos intervalos das bases em  $g(t)$ , obtém-se:

$$\begin{aligned} [g(b) - g(a_{n-1})] + [g(a_{n-1}) - g(a_{n-2})] + [g(a_{n-2}) - g(a_{n-3})] + \dots + [g(a_1) - g(1)] \\ = g(b) - g(1) \\ = g(b). \end{aligned}$$

Com base em (2.16), consideremos as desigualdades:

$$\begin{aligned} r[1, a_1] &\leq g(a_1) - g(1) \leq R[1, a_1] \\ r[a_1, a_2] &\leq g(a_2) - g(a_1) \leq R[a_1, a_2] \\ &\cdot \\ &\cdot \\ &\cdot \\ r[a_{n-2}, a_{n-1}] &\leq g(a_{n-1}) - g(a_{n-2}) \leq R[a_{n-2}, a_{n-1}] \\ r[a_{n-1}, b] &\leq g(b) - g(a_{n-1}) \leq R[a_{n-1}, b]. \end{aligned}$$

Somando os membros correspondentes das desigualdes acima, obtém-se:

$$\begin{aligned} r_n[1, b] &\leq g(b) - g(1) \leq R_n[1, b] \\ \implies r_n[1, b] &\leq g(b) \leq R_n[1, b]. \end{aligned} \tag{2.18}$$

Em que:

$$\begin{aligned} r_n[1, b] &= r[1, a_1] + r[a_1, a_2] + \dots + r[a_{n-1}, b], \\ R_n[1, b] &= R[1, a_1] + R[a_1, a_2] + \dots + R[a_{n-1}, b]. \end{aligned}$$

Fazendo  $n$  muito grande na dupla desigualdade (2.18), temos que  $r_n[1, b]$  e  $R_n[1, b]$  convergem à  $A_{-3}[1, b]$ . Como  $g(b)$  está espremido por  $r_n[1, b]$  e  $R_n[1, b]$ , tem-se que  $g(b) = A_{-3}[1, b]$ , conforme queríamos mostrar. Embora tenhamos fixado  $a = 1$  em  $[a, b]$  para facilitar a compreensão, pode-se fazer  $a$  igual a qualquer valor maior que zero e substituir  $s$  por este mesmo valor em  $g(t)$ .



2.2.3 2º Caso com  $f(x) = cx^{-m}$ ,  $x > 0$ ,  $c > 0$  e  $m > 1$ 

Conforme pudemos verificar, a candidata em (2.12) “passou no teste” para  $m = 2$  e  $m = 3$ . Na verdade, ela serve para todo  $m > 1$ , o que pode ser verificado por meio de raciocínios análogos aos realizados para os valores de  $m$  mencionados. No Apêndice A.2, encontra-se uma demonstração para  $f(x) = cx^{-m}$ ,  $m > 1$  e  $c > 0$ , de onde podemos obter a equação (2.12) fazendo  $c = 1$ .

Assumiremos, então, que a área sob a curva de  $f(x) = x^{-m}$ ,  $m > 1$ , no intervalo  $[s, t]$  é:

$$A_{-m}[s, t] = g(t) = \frac{1}{m-1} \cdot \left( \frac{1}{s^{m-1}} - \frac{1}{t^{m-1}} \right). \quad (2.19)$$

E, conforme em A.2, a área sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$  e  $c > 0$ , no intervalo  $[s, t]$  é:

$$A_{-m}[s, t] = g(t) = \frac{c}{m-1} \cdot \left( \frac{1}{s^{m-1}} - \frac{1}{t^{m-1}} \right). \quad (2.20)$$

2.3 3º CASO:  $f(x) = x^{-1}$ ,  $x > 0$ 

É fácil verificar que, neste 3º caso, a fórmula (2.19) para a obtenção da área sob a curva, verificada para o 2º caso, não se verifica aqui. Se substituirmos, em (2.19), o valor de  $m$  por 1, caímos em uma indeterminação com o número zero ocorrendo no denominador da fração.

Nesta seção, em vez de obtermos uma fórmula para a área sob a curva de  $f(x) = x^{-1} = \frac{1}{x}$ ,  $x > 0$ , o que, aliás, seria possível, pretendemos mostrar que ao aumentar indiscriminadamente o valor de  $x$  ou, como costumamos dizer, fazê-lo tender ao infinito, a área sob a curva não converge para um valor, mas cresce indefinidamente, não servindo a função  $f(x)$  para uma função densidade de probabilidade. Isto é o suficiente para o nosso objetivo neste trabalho, devendo  $f(x)$  ser descartada como Lei de Potência.

Dados  $a$  e  $b$  reais, com  $0 < a < b$ , a área da região sob o gráfico de  $f(x) = x^{-1} = \frac{1}{x}$ , com  $a \leq x \leq b$ , acima do eixo  $x$ , será denotada por  $A[a, b]$  (Figura 34).

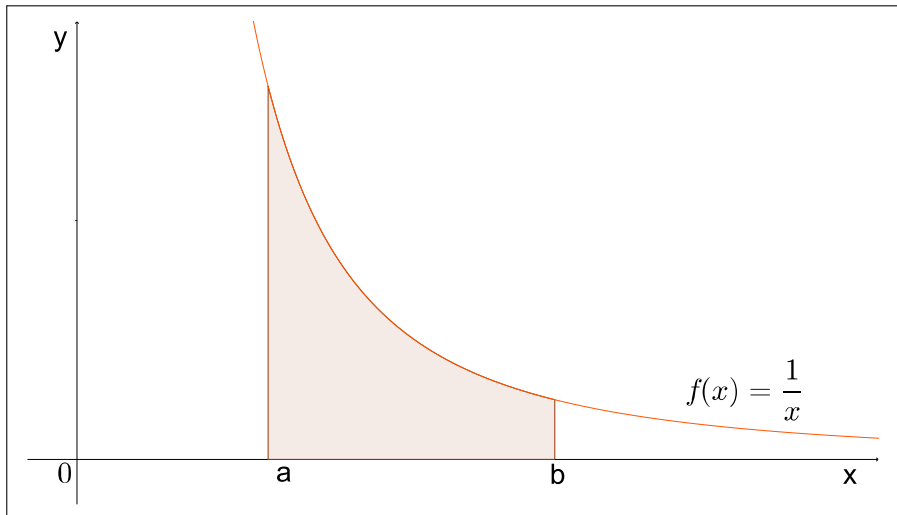


Figura 34: Área sob a curva de  $f(x) = \frac{1}{x}$ , em  $[a, b]$ .

Queremos mostrar, primeiramente, uma propriedade que será de grande importância. Dado  $k > 0$ , as áreas  $A[a, b]$  e  $A[ak, bk]$  são equivalentes (Ver Figura 35).

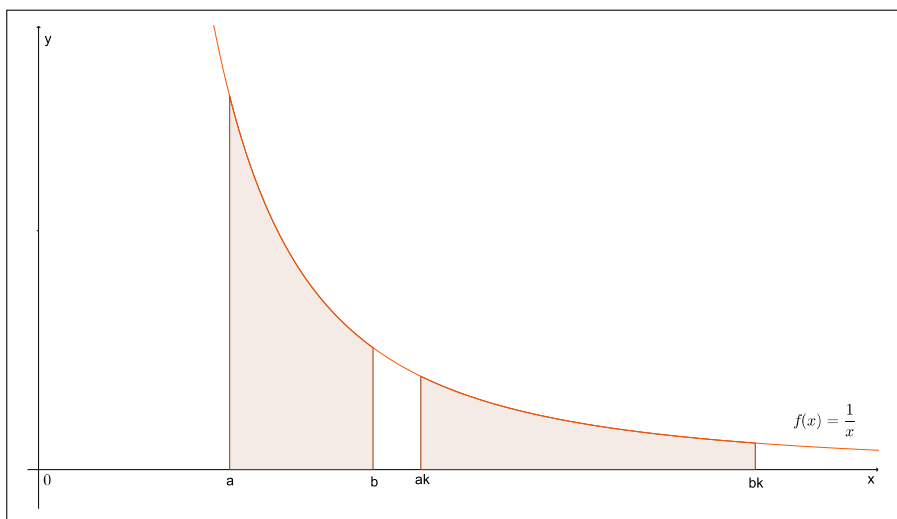


Figura 35: Áreas sob a curva de  $f(x) = \frac{1}{x}$ , em  $[a, b]$  e  $[ak, bk]$ .

Dado um retângulo, no interior da região sob a curva, em  $[a, b]$ , de base igual ao intervalo  $[a_1, a_2]$  e altura  $\frac{1}{a_2}$ , há um retângulo no intervalo  $[ak, bk]$  que possui a mesma área que o retângulo mencionado, é o retângulo de base igual a  $[a_1 \cdot k, a_2 \cdot k]$  e altura  $\frac{1}{a_2 \cdot k}$  (Figura 36). Denotando a área destes retângulos por  $r[a_1, a_2]$  e  $r[a_1 \cdot k, a_2 \cdot k]$ , respectivamente, tem-se:

$$r[a_1, a_2] = (a_2 - a_1) \cdot \frac{1}{a_2} = 1 - \frac{a_1}{a_2},$$

$$r[a_1 \cdot k, a_2 \cdot k] = (a_2 \cdot k - a_1 \cdot k) \cdot \frac{1}{a_2 \cdot k} = \frac{a_2 \cdot k}{a_2 \cdot k} - \frac{a_1 \cdot k}{a_2 \cdot k} = 1 - \frac{a_1}{a_2}.$$

Comprovando-se que  $r[a_1, a_2] = r[a_1 \cdot k, a_2 \cdot k]$ . (2.21)

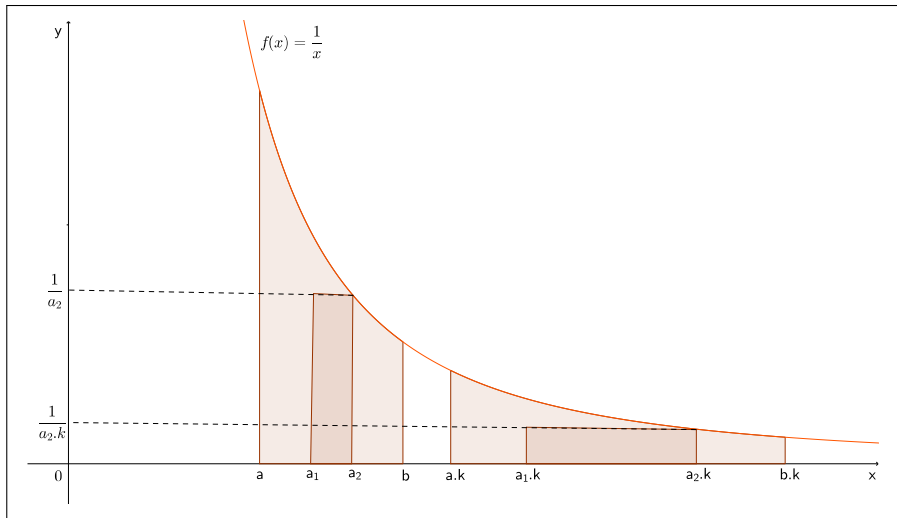


Figura 36: Áreas das regiões sob a curva de  $f(x) = \frac{1}{x}$ , em  $[a, b]$  e  $[a \cdot k, b \cdot k]$ . São destacados dois retângulos no interior de ambas as regiões.

Vamos agora aproximar  $A[a, b]$  por  $n$  retângulos com bases iguais a  $[a, a_1], [a_1, a_2], \dots, [a_{n-1}, b]$ , com  $a < a_1 < a_2 < \dots < b$ , e alturas  $\frac{1}{a_1}, \frac{1}{a_2}, \dots, \frac{1}{b}$ , respectivamente (Figura 37). Para cada um destes retângulos existe, de acordo com (2.21), um retângulo correspondente em  $[ak, bk]$ , de modo que a área  $A[ak, bk]$  também pode ser aproximada por  $n$  retângulos com bases iguais a  $[ak, a_1k], [a_1k, a_2k], \dots, [a_{n-1}k, bk]$  e alturas  $\frac{1}{a_1k}, \frac{1}{a_2k}, \dots, \frac{1}{bk}$ , respectivamente (Figura 37). Tem-se:

ÁREA SOB A CURVA

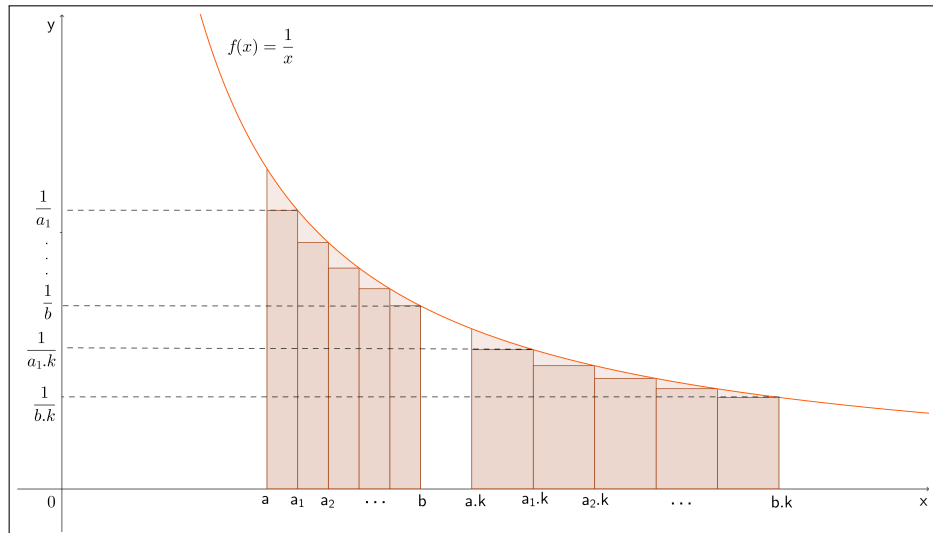


Figura 37: Áreas das regiões sob a curva de  $f(x) = \frac{1}{x}$ , em  $[a, b]$  e  $[ak, bk]$  cercadas, ambas, por  $n$  retângulos.

$$\begin{aligned}
 r[a, a_1] &= r[ak, a_1k] \\
 r[a_1, a_2] &= r[a_1k, a_2k] \\
 &\vdots \\
 &\vdots \\
 &\vdots \\
 r[a_{n-1}, b] &= r[a_{n-1}k, bk].
 \end{aligned}$$

Somando os membros correspondentes das equações acima, obtém-se:

$$r_n[a, b] = r_n[ak, bk]. \tag{2.22}$$

Em que:

$$\begin{aligned}
 r_n[a, b] &= r[a, a_1] + r[a_1, a_2] + \dots + r[a_{n-1}, b], \\
 r_n[ak, bk] &= r[ak, a_1k] + r[a_1k, a_2k] + \dots + r[a_{n-1}k, bk].
 \end{aligned}$$

Para um número muito grande de retângulos (valor muito grande de  $n$ ):

$$\begin{aligned}
 r_n[a, b] &= A[a, b], \\
 r_n[ak, bk] &= A[ak, bk].
 \end{aligned} \tag{2.23}$$

Das equações (2.22) e (2.23):

$$A[a, b] = A[ak, bk]. \quad (2.24)$$

Da propriedade (2.24), verificada acima, decorre uma outra, que para nós será, também, de grande importância.

Tomando  $a$  e  $b$  reais,  $1 < a \leq b$ , e considerando as áreas  $A[1, a]$ ,  $A[1, b]$ ,  $A[1, ab]$  e  $A[a, ab]$  (Figura 38), pode-se escrever<sup>6</sup>:

$$\begin{aligned} A[1, ab] &= A[1, a] + A[a, ab] \\ \Rightarrow A[1, ab] &= A[1, a] + A[1, b]. \end{aligned} \quad (2.25)$$

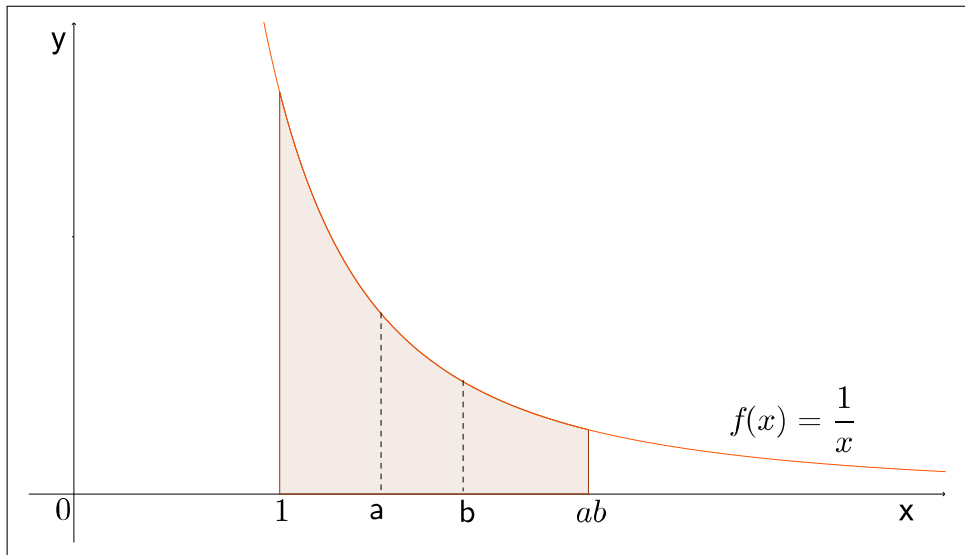


Figura 38: Áreas das regiões sob a curva de  $f(x) = \frac{1}{x}$ , em  $[1, a]$ ,  $[1, b]$ ,  $[1, ab]$ ,  $[a, ab]$ .

Como consequência, em (2.25), para  $b = a$ :

$$\begin{aligned} A[1, a \cdot a] &= A[1, a] + A[1, a] \\ \Rightarrow A[1, a^2] &= 2 \cdot A[1, a]. \end{aligned}$$

Ou ainda:

$$\begin{aligned} A[1, a^k] &= A[1, a] + A[1, a] + \dots + A[1, a] \quad (k \text{ parcelas iguais a } A[1, a]) \\ \Rightarrow A[1, a^k] &= k \cdot A[1, a]. \end{aligned} \quad (2.26)$$

<sup>6</sup> Note que  $A[1, b] = A[a, ba]$ , conforme a propriedade (2.24)

Munidos agora das propriedades (2.24) e (2.26), podemos mostrar o que queremos: que a área sob a curva de  $f(x) = \frac{1}{x}$  aumenta indefinidamente conforme aumentamos o valor de  $x$ , a área não converge para um número, não servindo  $f(x)$  para uma função densidade de probabilidade, pois para tal, a área da curva, para  $x$  tendendo ao infinito, deveria ser igual a 1.

Consideremos, que para qualquer  $x$  real fixado,  $x > 1$ , existe  $k \in \mathbb{N}$  tal que:

$$2^k \leq x < 2^{k+1} \quad (\text{Figura 39}).$$

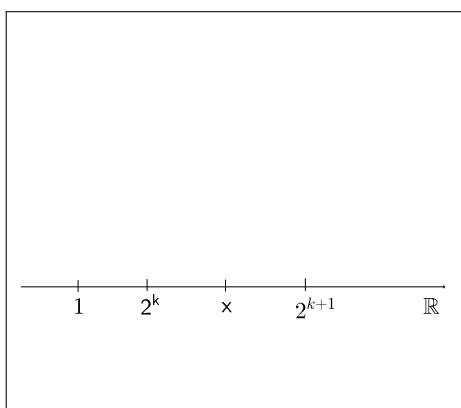


Figura 39: Para qualquer  $x$  real, é sempre possível  $k$  inteiro tal que  $2^k \leq x < 2^{k+1}$ .

Considerando valores de  $x$  cada vez maiores ( $x$  tendendo ao infinito), tem-se que  $2^k$  e  $2^{k+1}$  também devem crescer sem limites para “cercarem” o  $x$ , logo  $k$  e  $k + 1$  também crescem sem limites ( $k$  e  $k + 1$  tendem ao infinito)<sup>7</sup>.

Considerando agora as áreas sob a curva de  $f(x) = \frac{1}{x}$  em  $[1, 2^k]$ ,  $[1, x]$ ,  $[1, 2^{k+1}]$  (Figuras 40, 41 e 42), tem-se:

$$A[1, 2^k] \leq A[1, x] \leq A[1, 2^{k+1}].$$

Aplicando a propriedade (2.26) na dupla desigualdade acima:

$$k \cdot A[1, 2] \leq A[1, x] \leq (k + 1) \cdot A[1, 2].$$

<sup>7</sup> É fácil mostrar que não existe  $k$  fixo tal que  $2^k \leq x < 2^{k+1}$ , para qualquer  $x$ . Basta fazer  $x = 2^{k+2}$  e tem-se um valor de  $x$  que não verifica a desigualdade para o  $k$  determinado. Logo, para um aumento indiscriminado de  $x$  tem-se um aumento indiscriminado de  $k$ .

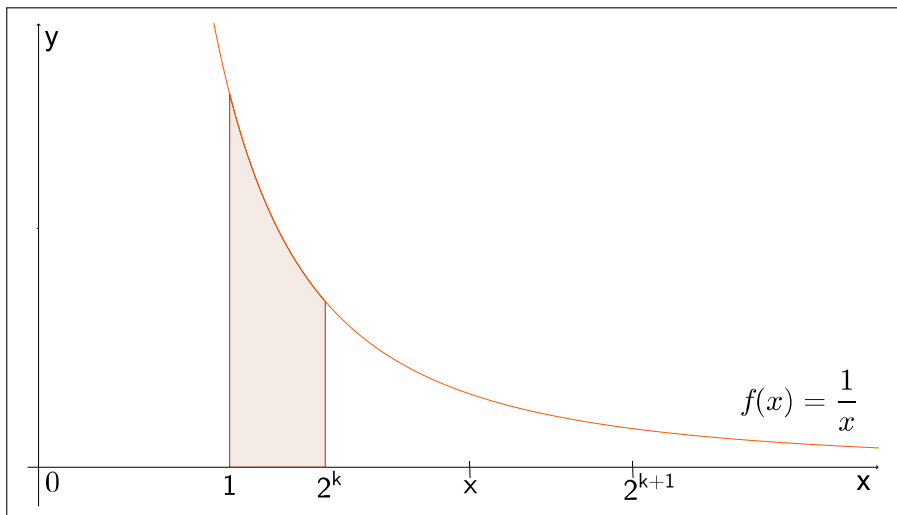


Figura 40: Área sob a região da curva de  $f(x) = \frac{1}{x}$  no intervalo  $[1, 2^k]$ .

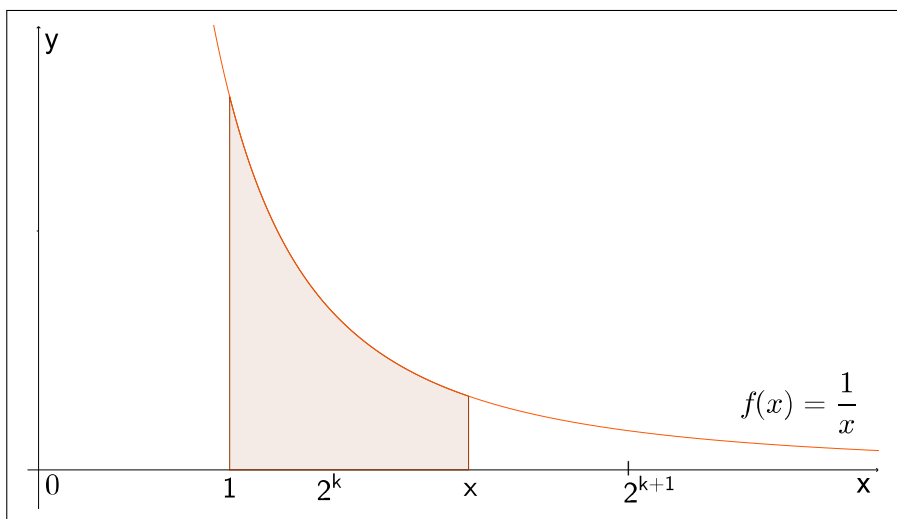


Figura 41: Área sob a região da curva de  $f(x) = \frac{1}{x}$  no intervalo  $[1, x]$ .

Para  $x$  muito grande tem-se  $k$  e  $k + 1$  muito grandes, conforme já discutido, e, conseqüentemente,  $k \cdot A[1, 2]$  e  $(k + 1) \cdot A[1, 2]$  tornam-se arbitrariamente grandes, não convergindo para um número. Logo,  $A[1, x]$  também cresce arbitrariamente para  $x$  tendendo ao infinito, não se aproximando de um número, conforme queríamos mostrar.

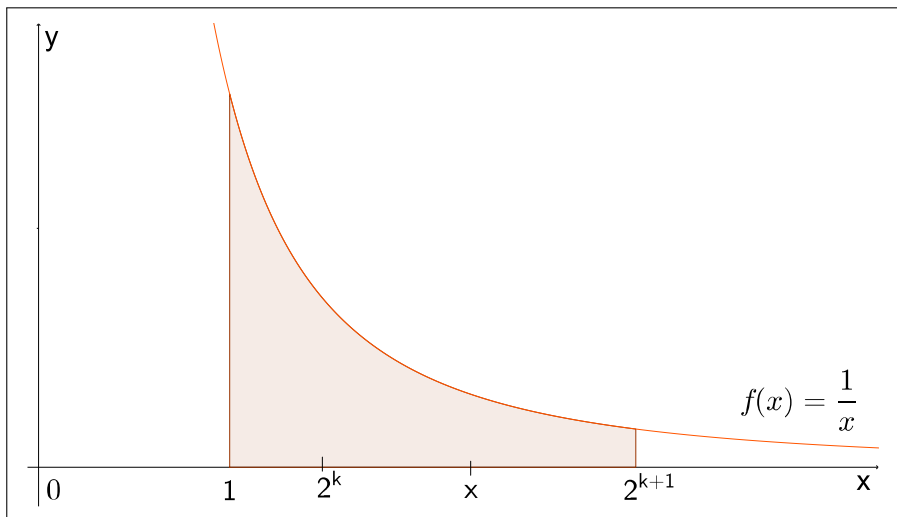


Figura 42: Área sob a região da curva de  $f(x) = \frac{1}{x}$  no intervalo  $[1, 2^{k+1}]$ .

#### 2.4 POSSIBILIDADE DA UTILIZAÇÃO DE $f(x) = c \cdot x^m$ , $x > 0$ , $c > 0$ , COMO FUNÇÃO DE DENSIDADE DE PROBABILIDADE

Nesta seção, retomaremos algumas conclusões precedentes na intenção de avaliar para quais casos/exponentes a função  $f(x) = cx^m$ ,  $x > 0$ ,  $c > 0$ , serve como função densidade de probabilidade. De certa forma esta discussão já esteve presente anteriormente, porém, convém organizá-la aqui, dada a sua importância para o assunto deste trabalho.

Lembremos que uma característica essencial de uma função densidade de probabilidade, conforme discutido no Capítulo 1, na seção 1.7, é que a área sob a sua curva, em seu intervalo, é igual a 1. Isto se deve ao fato de que como esta área corresponde à probabilidade de todos os valores da função, seu valor, tal como a soma destas probabilidades, deve ser igual a 1.

Em relação à função  $f(x) = x^m$ ,  $x > 0$ , vimos três casos com  $m \geq 0$ ,  $m < 0$  (exceto  $m = -1$ ) e  $m = -1$ . Para  $m = -1$ , já verificamos que para valores de  $x$  muito grandes a área não se aproxima de um número e, portanto, tal função não serve para densidade de probabilidade. Desta forma, tampouco para  $m = -1$  a função  $f(x) = cx^m$ ,  $c > 0$ , serve, pois o acréscimo da constante  $c$  não muda o caráter divergente da área sob sua curva. Resta-nos verificar os outros dois casos, referentes aos outros intervalos de  $m$ .

Para  $m \geq 0$ , a área sob a curva em  $[0, t]$  pode ser expressa, conforme em (2.10), por:



$$A_m(t) = \frac{t^{m+1}}{m+1}.$$

Para valores muito grandes de  $t$  ( $t$  tendendo ao infinito) é fácil ver que  $t^{m+1}$  também aumenta indiscriminadamente (tende ao infinito) e, portanto, a fração  $\frac{t^{m+1}}{m+1}$  também tende ao infinito. Desta forma, tal função também não se presta à densidade de probabilidade.

Para  $m < 0$ , exceto  $m = -1$ , ou, de modo equivalente, para  $m > 0$ , exceto  $m = 1$ , em  $f(x) = cx^{-m}$ ,  $c > 0$ , como vimos em (2.20):

$$A_{-m}[s, t] = \frac{c}{m-1} \cdot \left( \frac{1}{s^{m-1}} - \frac{1}{t^{m-1}} \right).$$

Pode-se notar que para  $t$  muito grande (tendendo ao infinito), a fração  $\frac{1}{t^{m-1}}$  anula-se, somente se  $m - 1 > 0$ , o que ocorre se  $m > 1$ , conforme assumimos neste caso. Fazendo  $\frac{1}{t^{m-1}} = 0$  na equação acima, obtém-se:

$$A_{-m}[s, t] = \frac{c}{m-1} \cdot \left( \frac{1}{s^{m-1}} \right).$$

Nesta última equação a área equivale a um número que depende dos valores de  $c$ ,  $m$  e  $s$  da função. Pode-se, dados  $m$  e  $s$ , por exemplo, obter um valor para o  $c$  que “normaliza” a função, ou seja, que proporciona que a área sob a sua curva seja igual a 1.

Toda esta discussão aplica-se, também, para o caso dos valores de  $m$  pertencentes ao conjunto dos racionais; porém, como os cálculos dependem de algo técnico, cuja complexidade está além do escopo deste trabalho, eles não serão aqui expostos.

Este último caso, em que o expoente é menor que  $-1$ , incluindo os expoente racionais neste intervalo, é no que consistirá uma “Lei de Potência”, assunto que desenvolveremos mais no Capítulo 4.



---

## REGRESSÃO LINEAR SIMPLES

---

Um tema importante em matemática e muitas vezes negligenciado no ensino básico é o da modelagem matemática. A modelagem matemática consiste em construir um modelo matemático que se ajuste, o melhor possível, a um conjunto de dados de que se dispõe. Estes dados, por sua vez, são geralmente obtidos em uma ou várias experiências reais, ou simuladas. Ao construir um modelo que se adequa à experiência, tem-se uma melhor compreensão do fenômeno, além de possibilitar previsões acerca de dados futuros.

Dentro da modelagem, a Regressão Linear Simples é o método pelo qual, dadas duas variáveis relacionadas  $X$  e  $Y$ , obtém-se uma reta que melhor se ajusta, de acordo com algum critério, aos dados disponíveis. A variável  $X$  é chamada de independente ou explicativa. A variável  $Y$  é chamada de variável dependente ou resposta. Ao aplicar o método da Regressão Linear Simples sobre as variáveis  $X$  e  $Y$ , assume-se que o fenômeno constituído por ambas pode ser bem ajustado por uma reta, e busca-se prever os valores de  $Y$  por meio dos valores de  $X$ , ou, em outras palavras, conhecer o quanto as alterações nos valores de  $X$  alteram os valores de  $Y$ .

O critério que será adotado para a obtenção da reta será o Método dos Mínimos Quadrados (MMQ).

Para um estudo mais detalhado do assunto, pode-se consultar, por exemplo, Morettin e Bussab (2013, p.454), [2].

## 3.1 RETA QUE PASSA POR DOIS PONTOS DISTINTOS

A equação de reta é algo bem estudado no ensino médio, aparecendo no estudo de funções polinomiais do 1º grau e em geometria analítica. Como bem se sabe, dados dois pontos distintos só há uma reta que os contém, e dadas as coordenadas destes dois pontos é possível determinar a equação desta reta. De forma genérica, a equação de uma reta será representada por:

$$y = ax + b \quad (3.1)$$

Sendo  $a$  o coeficiente de  $x$  ou coeficiente angular e  $b$  o coeficiente linear ou intercepto, uma vez que  $b$  é a ordenada do ponto em que a reta intercepta o eixo das ordenadas.

**Exemplo 3.1.** Dados dois pontos A(1,2) e B(3,4), qual é a equação da reta que os contém?

Substituindo na equação (3.1),  $x$  e  $y$  pelas coordenadas  $x$  e  $y$  dos pontos  $A$  e  $B$ , obtém-se:

$$\begin{cases} a + b = 2 \\ 3a + b = 4 \end{cases} .$$

Tal sistema fornece os valores  $a = 1$  e  $b = 1$ , e portanto:  $y = x + 1$ .

Tem-se aqui, para os dois pontos considerados, um ajuste perfeito; a reta dá conta perfeitamente de ambos os pontos.

## 3.2 RETA A PARTIR DE TRÊS, OU MAIS, PONTOS DISTINTOS

Considerando três ou mais pontos distintos, a situação é diferente. Caso os pontos estejam alinhados, tem-se um caso tal como visto na seção anterior, bastando encontrar a equação da reta por meio de dois dos pontos e, caso necessário, utilizando os demais apenas para conferir a equação obtida. Caso os três pontos, ou mais, não estejam alinhados, evidentemente não há uma equação de reta que dê conta de todos os pontos simultaneamente e neste caso busca-se a reta que, segundo algum critério, aproxime-se mais dos pontos em questão.

**Exemplo 3.2.** Dados os pontos da Tabela 4, qual é a equação da reta que melhor se ajusta a estes pontos?

Tabela 4: Três pontos não alinhados.

Variável X	Variável Y
1	2
3	5
4	8

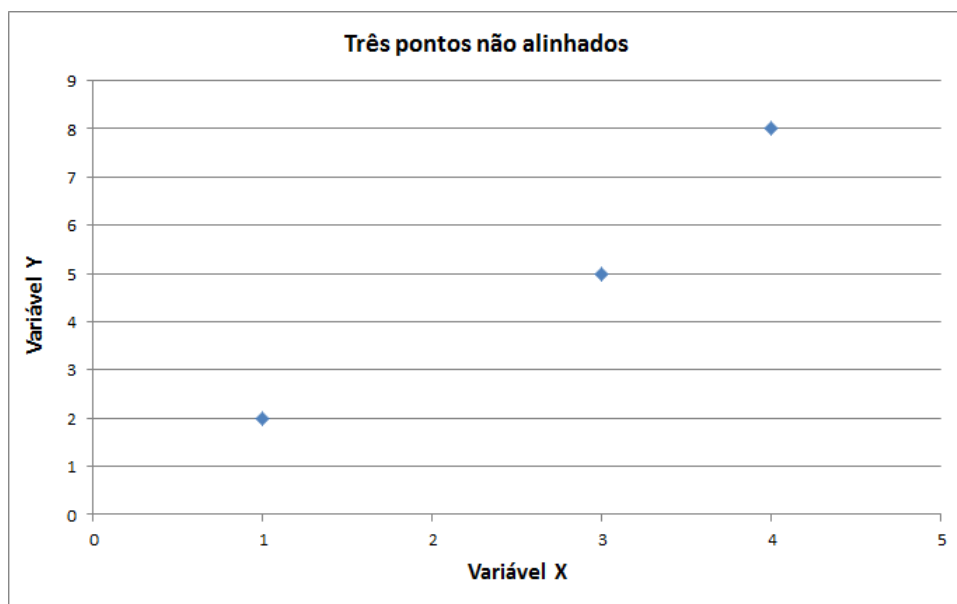


Figura 43: Três pontos não alinhados.

Podemos colocar estes três pontos em um plano cartesiano (Figura 43).

*Pelo MMQ, a reta que melhor se ajusta aos pontos da Figura 43 é aquela em que a soma dos quadrados dos resíduos (diferença entre os respectivos valores de Y observados e os estimados pela equação da reta) é a menor possível.*

Adotando:

$y_i$ : Valor observado (real) da variável Y no  $i$ -ésimo ponto.

$\hat{y}_i$ : Valor da variável Y estimado pela equação da reta no  $i$ -ésimo ponto.

$e_i = y_i - \hat{y}_i$  (resíduo do  $i$ -ésimo ponto).

$y = ax + b$  : Equação da reta obtida pelo MMQ.

$\sum_{i=1}^n e_i^2$  : Soma dos quadrados dos resíduos dos  $n$  pontos.

Tem-se:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Encontrar a equação da reta procurada é encontrar os valores dos coeficientes  $a$  e  $b$ .  
Aplicando o MMQ nos dados da Figura 43:

$$\begin{aligned} \sum_{i=1}^3 (y_i - ax_i - b)^2 &= (2 - 1a - b)^2 + (5 - 3a - b)^2 + (8 - 4a - b)^2 \\ &= 26a^2 + 3b^2 - 98a - 30b + 16ab + 93. \end{aligned} \quad (3.2)$$

Procuram-se os valores de  $a$  e  $b$  que minimizam a equação (3.2), para encontrá-los serão postos em evidência, num primeiro caso o  $a$ , e, posteriormente, o  $b$ , desta forma, e utilizando a ideia do vértice da parábola, serão obtidas duas equações e pela resolução do sistema assim formado, se chegará aos valores de  $a$  e  $b$  procurados.

1º Caso: Colocando o  $a$  em evidência:

$$26a^2 + (16b - 98)a + 3b^2 - 30b + 93.$$

Considerando  $a$  como a variável da expressão, e  $b$  como uma constante, tem-se uma função polinomial do 2º grau em  $a$ . Neste caso, para o valor mínimo da função, o valor de  $a$ , ( $a_0$ ), corresponde a abscissa do vértice:

$$a_0 = -\frac{16b - 98}{2 \cdot 26} \Rightarrow 26a_0 + 8b = 49. \quad (3.3)$$

2º Caso: Colocando  $b$  em evidência:

$$3b^2 + (16a - 30)b + 26a^2 - 98a + 93.$$

Considerando, agora,  $b$  como a variável e  $a$  como constante, tem-se uma função polinomial do 2º grau em  $b$ . O valor mínimo da função ocorre para o valor de  $b$ ,  $b_0$ , que corresponde a abscissa do vértice:

$$b_0 = -\frac{16a - 30}{2 \cdot 3} \Rightarrow 8a + 3b_0 = 15. \quad (3.4)$$

Tomando as equações (3.3) e (3.4), e levando em consideração que se quer encontrar  $a$  e  $b$  que correspondem a  $a_0$  e  $b_0$ , respectivamente, tem-se o seguinte sistema:

$$\begin{cases} 26a_0 + 8b_0 = 49 \\ 8a_0 + 3b_0 = 15. \end{cases}$$

Do que decorre:  $a_0 = \frac{27}{14} \Rightarrow a_0 \approx 1,929$ . E,  $b_0 \approx -0,143$ .

Logo, a equação da reta é:  $y = 1,929x - 0,143$ . Na Figura 44, temos a sua representação no gráfico.

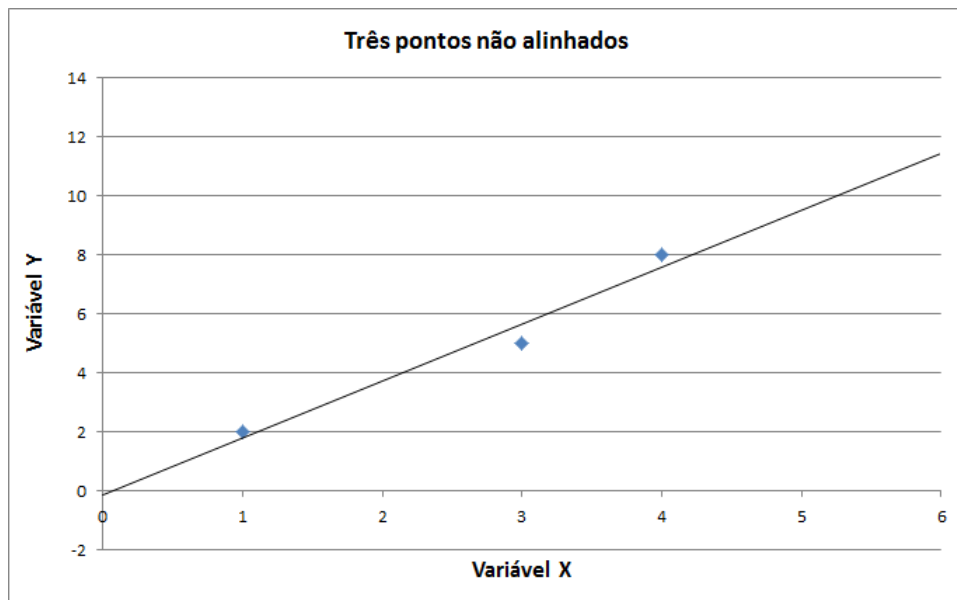


Figura 44: Três pontos não alinhados e a linha de tendência.

### 3.3 FÓRMULAS PARA REGRESSÃO LINEAR SIMPLES

Como proceder para mais de três pontos não alinhados? Vejamos, através do próximo exemplo:

Tabela 5: Oito pontos não alinhados.

Variável X	Variável Y
1	8
2	12
4	10
6	14
8	15
9	23
12	20
14	26

**Exemplo 3.3.** Dados os pontos da Tabela 5, qual é a equação da reta que melhor se ajusta a estes pontos?

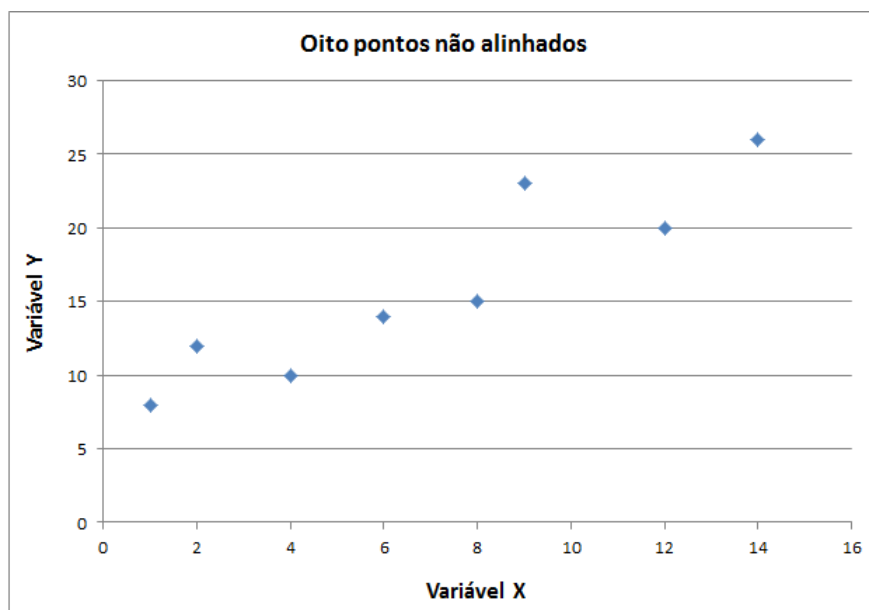


Figura 45: Oito pontos não alinhados.

Colocando estes pontos no plano cartesiano, obtemos o gráfico da Figura 45.

Poderão ser aplicados, aqui, os mesmos raciocínios e cálculos aplicados no exemplo anterior, porém o trabalho será muito maior.



Antes de prosseguirmos com o exemplo, deduziremos algumas fórmulas para o cálculo dos coeficientes  $a$  e  $b$  da reta de regressão linear pelo MMQ. Para essas deduções, será utilizada a mesma abordagem do exemplo anterior, generalizando-a para  $n$  pontos.

Expressando a soma dos quadrados dos resíduos:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (ax_i + b)]^2 = \sum_{i=1}^n (y_i - ax_i - b)^2.$$

Desenvolvendo:

$$\sum_{i=1}^n (y_i^2 + a^2 + b^2 x_i^2 - 2ay_i - 2by_i x_i + 2abx_i).$$

1º Caso: Colocando  $a$  em evidência:

$$\begin{aligned} & \sum_{i=1}^n [a^2 + (2bx_i - 2y_i)a + y_i^2 - 2by_i x_i + b^2 x_i^2] \\ &= \sum_{i=1}^n a^2 + \sum_{i=1}^n (2bx_i - 2y_i)a + \sum_{i=1}^n (y_i^2 - 2by_i x_i + b^2 x_i^2) \\ &= na^2 + a \sum_{i=1}^n (2bx_i - 2y_i) + \sum_{i=1}^n (y_i^2 - 2by_i x_i + b^2 x_i^2). \end{aligned}$$

Considerando  $b$  como constante, e tratando a expressão como uma função polinomial do 2º grau em  $a$ , o valor de  $a$ , para o qual o valor da função é mínimo, corresponde a abscissa do vértice ( $a_0$ ):

$$\begin{aligned} a_0 &= \frac{-\sum_{i=1}^n (2bx_i - 2y_i)}{2n} \\ \Rightarrow 2n \cdot a_0 &= -\sum_{i=1}^n (2bx_i - 2y_i) \\ \Rightarrow 2n \cdot a_0 &= -\sum_{i=1}^n (2bx_i) - \sum_{i=1}^n (-2y_i) \\ \Rightarrow 2n \cdot a_0 &= -2b \sum_{i=1}^n x_i + 2 \sum_{i=1}^n y_i. \end{aligned}$$

Dividindo ambos os membros da equação por 2, e isolando  $a_0$ :

$$a_0 = \frac{-b}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n y_i.$$

Considerando que  $\frac{\sum_{i=1}^n x_i}{n} = \bar{x}$  e  $\frac{\sum_{i=1}^n y_i}{n} = \bar{y}$ , pode-se reescrever a equação:

$$a_0 = \bar{y} - b\bar{x}. \quad (3.5)$$

Esta equação será utilizada para encontrar o valor do coeficiente  $a$  da reta de regressão linear. Note que  $\bar{y}$  é a média dos valores observado de  $Y$  e  $\bar{x}$  é a média dos valores observados de  $X$ .

Agora, no 2º caso, o procedimento será semelhante, colocando o  $b$  em evidência. Será obtida uma equação que permitirá a obtenção do valor do coeficiente  $b$ .

**2º Caso:** Colocando  $b$  em evidência:

$$\begin{aligned} & \sum_{i=1}^n [x_i^2 b^2 + (2ax_i - 2y_i x_i)b + y_i^2 - 2ay_i + a^2] \\ &= \sum_{i=1}^n x_i^2 b^2 + \sum_{i=1}^n (2ax_i - 2x_i y_i)b + \sum_{i=1}^n (y_i^2 - 2ay_i + a^2) \\ &= b^2 \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n (2ax_i - 2x_i y_i) + \sum_{i=1}^n (y_i^2 - 2ay_i + a^2). \end{aligned}$$

Considerando  $a$  como constante, e pensando na expressão como a de uma função polinomial em  $b$ , tem-se que o seu valor mínimo ocorre quando  $b$  corresponde a abscissa do vértice ( $b_0$ ):

$$\begin{aligned}
b_0 &= \frac{-\sum_{i=1}^n (2ax_i - 2x_i y_i)}{2 \sum_{i=1}^n x_i^2} \\
\Rightarrow 2b_0 \sum_{i=1}^n x_i^2 &= -\sum_{i=1}^n (2ax_i - 2x_i y_i) \\
\Rightarrow 2b_0 \sum_{i=1}^n x_i^2 &= -\sum_{i=1}^n 2ax_i - \sum_{i=1}^n -2x_i y_i \\
\Rightarrow 2b_0 \sum_{i=1}^n x_i^2 &= -2a \sum_{i=1}^n x_i + 2 \sum_{i=1}^n x_i y_i.
\end{aligned}$$

Dividindo ambos os membros da equação por 2, e fazendo  $a = a_0 = \bar{y} - b_0 \bar{x}$ :

$$\begin{aligned}
b_0 \sum_{i=1}^n x_i^2 &= -(\bar{y} - b_0 \bar{x}) \sum_{i=1}^n x_i + \sum_{i=1}^n x_i y_i \\
\Rightarrow b_0 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i + b_0 \bar{x} \sum_{i=1}^n x_i \\
\Rightarrow b_0 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i - \bar{y} \cdot n \cdot \bar{x} + b_0 \cdot \bar{x} \cdot n \cdot \bar{x} \\
\Rightarrow b_0 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} + n \cdot b_0 \cdot \bar{x}^2.
\end{aligned}$$

Colocando  $b_0$  em evidência no membro esquerdo da equação:

$$\begin{aligned}
b_0 \left( \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) &= \sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} \\
\Rightarrow b_0 &= \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2}. \tag{3.6}
\end{aligned}$$

A equação (3.6) fornece o coeficiente  $b$  da função em termos dos valores observados das variáveis  $X$  e  $Y$ .

Voltaremos, agora, ao Exemplo 3.3. Na Tabela 6, tem-se os dados da Tabela 5 e algumas colunas que foram acrescentadas com o auxílio do Excel; ela será utilizada para auxiliar na aplicação das fórmulas (3.5) e (3.6).

Tabela 6: Parâmetros para posterior aplicação das fórmulas.

Pontos ( $i$ )	$x_i$	$y_i$	$x_i * y_i$	$x_i^2$
1	1	8	8	1
2	2	12	24	4
3	4	10	40	16
4	6	14	84	36
5	8	15	120	64
6	9	23	207	81
7	12	20	240	144
8	14	26	364	196
<b>SOMA</b>			<b>1087</b>	<b>542</b>

Considerando, ainda:  $n = 8$ ,  $\bar{x} = 7$  e  $\bar{y} = 16$ , tem-se:

$$b = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} = \frac{1087 - 8 \cdot 7 \cdot 16}{542 - 8 \cdot 7^2} \approx 1,273,$$

$$a = \bar{y} - b\bar{x} \approx 16 - 1,273 \cdot 7 \approx 7,087.$$

Logo, a reta que melhor se ajusta aos pontos da Tabela 5, de acordo com o MMQ tem equação:

$$y = 1,273x + 7,087.$$

Esta reta pode ser observada no gráfico da Figura 46.

Embora as fórmulas (3.5) e (3.6) sirvam para a obtenção da equação da reta, o mais prático é obtê-la por meio de softwares estatísticos ou, ainda, pelo uso do Excel, como feito para a produção do gráfico da Figura 46. Ao utilizar o Excel, obtém-se a reta, chamada de linha de tendência, clicando com o botão direito do mouse sobre um dos pontos no gráfico e, em seguida, selecionando a opção “Adicionar Linha de Tendência...”, abre-se uma janela, na qual há vários campos. No primeiro destes campos

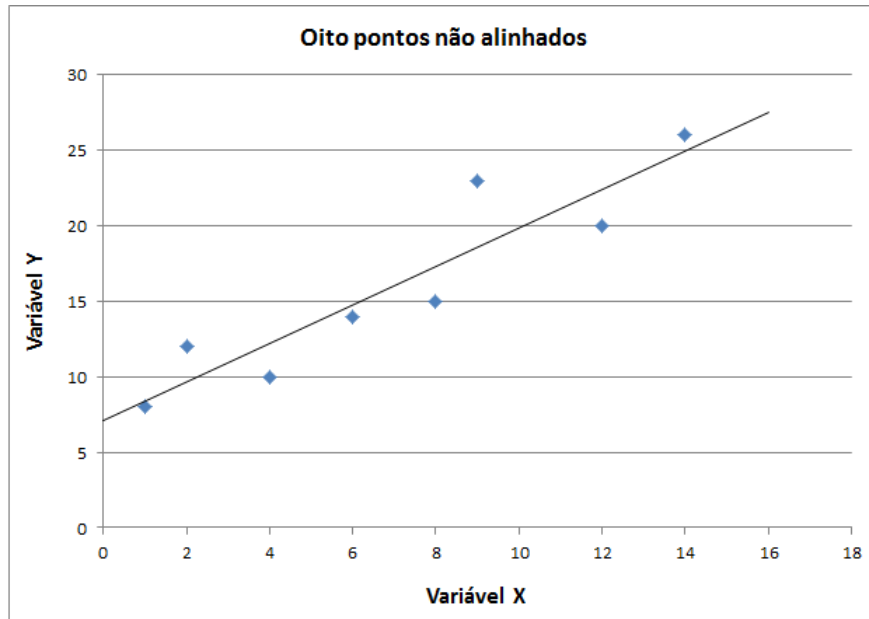


Figura 46: Reto ajustada aos 8 pontos não alinhados.

deve-se selecionar a opção “Linear”. No último dos campos tem-se a opção de exibir a equação da reta no gráfico.

No Capítulo 5, serão desenvolvidas várias atividades em que retas serão ajustadas aos pontos dos gráficos pelo método dos mínimos quadrados. Estas atividades serão realizadas no Excel, de modo que a teoria vista aqui, poderá ser, então, praticada.



---

## LEIS DE POTÊNCIA

---

Serviu-nos de inspiração para este capítulo o artigo de Newman (2006), [3]; todavia, a nossa abordagem não seguirá rigorosamente a abordagem utilizada neste artigo.

Conforme discutido no Capítulo 1, em muitos casos, é possível descrever o comportamento de um conjunto de dados obtidos por meio de uma experiência, ou pela análise teórica de uma situação, por meio de um modelo matemático. Este modelo, dependendo do caso, dá conta dos dados apenas de modo aproximado e sua escolha nem sempre é tão fácil.

Conforme visto na Seção 1.7, pode-se, por meio de uma função de densidade de probabilidade, estimar quais são as probabilidades envolvidas em relação às variáveis aleatórias contínuas, em cada intervalo de valores no eixo das abscissas, obtendo-se para isto a área delimitada pelo gráfico de tal função, pelo eixo  $x$  e pelas retas  $x = a$  e  $x = b$ , sendo  $a$  e  $b$  as extremidades do intervalo. Considerando, como veremos, que as leis de potência não apresentam gráfico retilíneo, obtivemos alguns resultados que nos permitem calcular a área sob a curva de certas funções nesta condição (Capítulo 2), dentre as quais, estão estas leis.

*Dizemos que uma variável obedece a uma lei de potência quando a densidade de probabilidade<sup>1</sup> de cada um de seus valores particulares varia de acordo com uma potência inversa deste valor. Como será visto mais adiante, considerando  $p(x)$  a densidade de probabilidade de ocorrência do valor  $x$ ,  $C$  e  $\alpha$  constantes convenientemente determinadas, pode-se expressar a referida lei de potência como:*

$$p(x) = Cx^{-\alpha}.$$

---

<sup>1</sup> Quando se trata de uma distribuição das frequências acumuladas, tem-se como valores da sua função lei de potência, não as densidades de probabilidade, mas as próprias probabilidades. Retomaremos isto mais adiante.

## 4.1 CASO DAS CIDADES BRASILEIRAS

Queremos encontrar um modelo matemático que melhor se adapte à distribuição dos tamanhos das cidades brasileiras em relação às suas populações. Sabemos que os tamanhos das populações das cidades apresentam grande diferença; algumas têm poucos moradores; outras, muitos. Poderíamos, então, encontrar uma função matemática, por exemplo, que descreva estas variações?

A primeira coisa, é obter dados confiáveis sobre aquilo que queremos analisar. Neste caso, a melhor fonte é o Instituto Brasileiro de Geografia e Estatística (IBGE). Consultando a página eletrônica do IBGE, temos acesso aos dados do censo demográfico, por ele realizado, de 2010. Estes são os dados que utilizaremos <sup>2</sup>.

Observando os dados, notamos que a população do Brasil, no referido censo, é de 190.755.799 de habitantes, sendo que, destes, 11.253.503 encontram-se na cidade de São Paulo. Temos, ainda, que 177.839.519 pertencem a municípios com mais de 10.000 habitantes, o que corresponde a aproximadamente 93% do total. Em nossas análises, utilizaremos somente estes municípios que possuem mais de 10.000 habitantes; pois há muitos pequenos municípios que, embora independentes, não são auto-sustentáveis<sup>3</sup>, o que comprometeria, sem necessidade, a nossa análise.

Pretendemos colocar estas populações em um gráfico, mas para isto, iremos organizá-las em histogramas<sup>4</sup>, conforme vimos na Seção 1.3. Cada histograma terá amplitude 20.000. O primeiro, portanto, incluirá as cidades com população de 10.000 a 30.000; o segundo, de 30.000 a 50.000; e assim por diante. Nos concentraremos, inicialmente, nas cidades com população de até 810.000 moradores.

---

2 Deve-se acessar o Banco de Dados Agregados do Sistema IBGE de Recuperação Automática (SIDRA). Disponível em: <<http://www.sidra.ibge.gov.br>>, na seção "Demográfico 2010- Sinopse na tabela "608 - População residente, por situação do domicílio e sexo". Nesta tabela, selecionamos, na seção "Unidade Territorial", a opção "Município". Obtivemos, assim, a população de cada cidade brasileira no referido censo. Acesso em: 08 jan.2016.

3 Para mais considerações a respeito, pode-se consultar SILVA; STOSIC, T.; STOSIC, B. (2008, p.1 e 2), [7]. Em seu artigo, Newman (2006, p.1), [3], também discute o caso e descarta cidades estadunidenses muito pequenas.

4 Certos pormenores práticos serão omitidos aqui e explicitados na atividade 5.4, quando analisarmos todos os dados, não nos limitando a 810.000. Consulte-a, agora, se desejar.



Criando um gráfico de dispersão para os histogramas<sup>5</sup>, acima descritos, considerando o valor médio de cada histograma e o número de cidades por ele contido, obtemos a Figura 47. Note que no gráfico da Figura 47, o primeiro valor destacado está bem afastado do eixo das abscissas, e na medida em que os valores da variável População aumentam, a frequência das cidades diminui; após um certo valor de  $x$ , a variação em  $y$  fica indiscernível pelo gráfico, com valores muito próximos do eixo das abscissas.

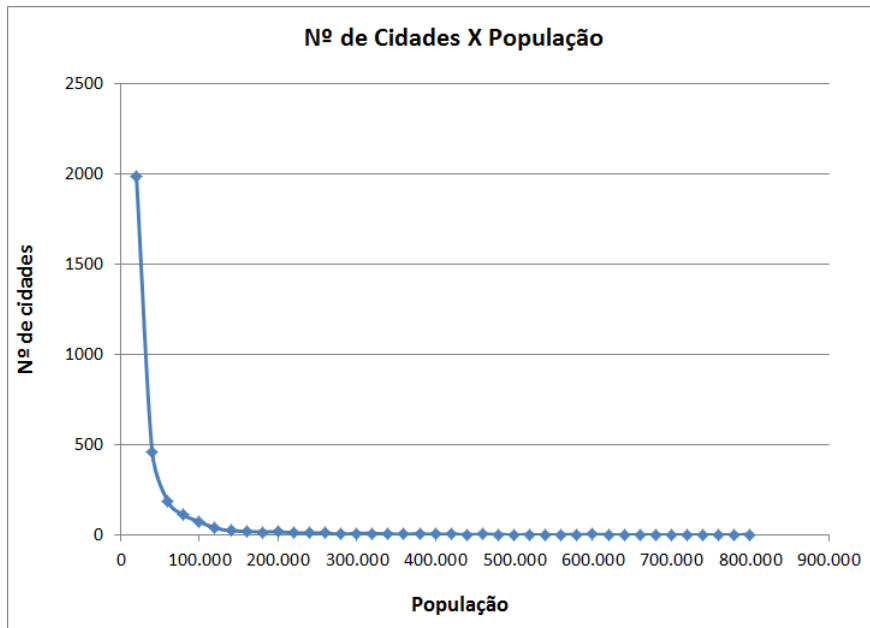


Figura 47: Nº de cidades X População.

Como fazemos para descobrir qual é a melhor função para relacionar as duas variáveis, Número de Cidades e População?

Para uma reta, a função é bem conhecida:  $y = ax + b$ . No entanto, pelo gráfico da Figura 47, vemos claramente que ele não sugere uma reta. Não sendo linear, não podemos, apenas por este gráfico, discernir qual é a sua função.

Duas funções candidatas, no entanto, são as funções reais:  $y = c \cdot e^{-x}$  (função exponencial) e  $y = c \cdot x^{-\alpha}$  (lei de potência).

Como saber de qual destas funções se trata?

<sup>5</sup> Embora seja conveniente a utilização de densidades de frequência, em vez das frequências absolutas, como vimos na Seção 1.3, optamos por utilizar, neste exemplo, a frequência absoluta. O fizemos para melhorar a visualização dos gráficos e, porque os histogramas possuem a mesma amplitude, a forma do gráfico, que é o que nos interessará neste momento, não se altera, seja com a frequência absoluta ou com a densidade de frequência.

Como não há dúvida em relação ao gráfico, quando este diz respeito à uma função linear, procuraremos tratar os dados, buscando uma disposição linear para os pontos.

Partindo da função exponencial, como podemos obter uma reta? Logaritmos! Vamos extrair o logaritmo em ambos os membros da equação, para  $c > 0$ :

$$\begin{aligned}y &= c \cdot e^{-x} \\ \Rightarrow \ln y &= \ln(c \cdot e^{-x}) \\ \Rightarrow \ln y &= \ln c + \ln e^{-x} \\ \Rightarrow \ln y &= \ln c - x.\end{aligned}$$

Fazendo  $\ln y = z$ :

$$z = -x + \ln c.$$

Como  $\ln c$  é uma constante, temos uma relação linear entre  $\ln y$  e  $x$ .

Vamos, agora, retornando ao nosso problema, extrair os logaritmos dos valores de  $y$  (nº de cidades), e verificar se ao plotarmos o gráfico com estes valores,  $\ln y$ , e os valores de  $x$  (população), a disposição dos pontos sugere uma reta. Observemos o resultado no gráfico da Figura 48.

Como podemos visualizar neste gráfico (Figura 48), a disposição dos pontos não nos sugere uma reta. Ao menos, não de modo claro.

E se aplicarmos a mesma ideia para uma função lei de potência, ou seja, também extrair os logaritmos <sup>6</sup> em ambos os membros de sua equação? Vejamos, para  $c > 0$ :

$$\begin{aligned}y &= c \cdot x^{-\alpha} \\ \Rightarrow \log y &= \log(c \cdot x^{-\alpha}) \\ \Rightarrow \log y &= \log c + \log x^{-\alpha} \\ \Rightarrow \log y &= -\alpha \log x + \log c.\end{aligned}$$

Fazendo  $\log y = z$  e  $\log x = w$ , temos:

<sup>6</sup> A escolha da base do logaritmo não é importante. Utilizaremos base 10 aqui.

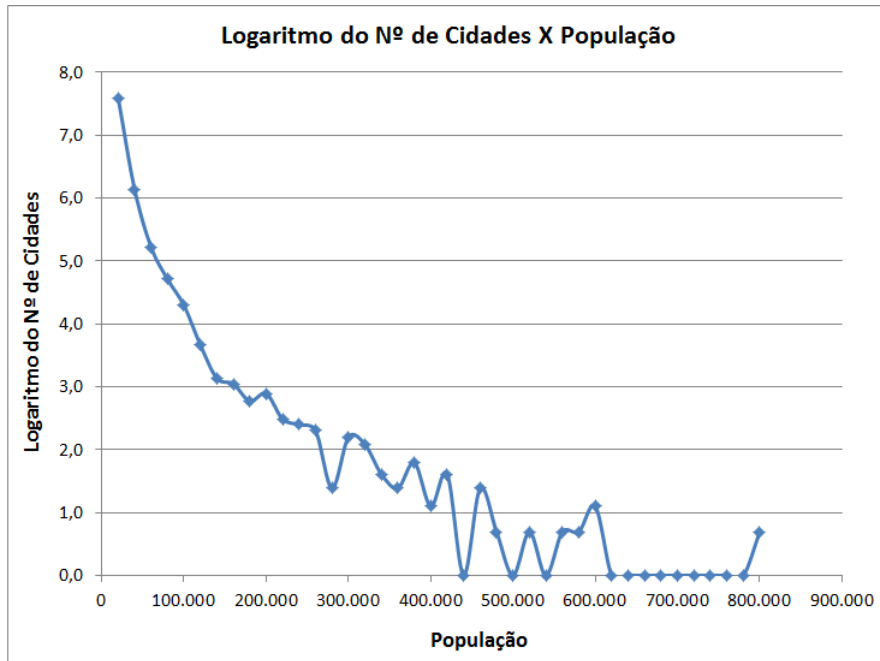


Figura 48: Logaritmo do Nº de Cidades X População.

$$z = -\alpha w + \log c. \quad (4.1)$$

Esta equação nos diz que quando temos uma lei de potência, existe uma relação linear entre os logaritmos das variáveis  $x$  e  $y$ .

Temos, então, que, se ao plotarmos o gráfico com as variáveis  $\log x$  (logaritmos das populações) e  $\log y$  (logaritmo do nº de cidades) os pontos nos sugerirem uma reta, podemos, então, modelar os nossos dados com uma lei de potência. Vejamos o gráfico, Figura 49, com as variáveis  $\log x$  e  $\log y$ .

Como podemos observar no gráfico da Figura 49, os pontos nos sugerem uma reta; assumiremos, então, se tratar de uma lei de potência. Este gráfico, por ter nos eixos os valores dos logaritmos das variáveis, é conhecido como  $\log\text{-}\log$ <sup>7</sup>. Um outro modo de obtê-lo, seria alterar, no Excel, as escalas dos eixos  $X$  e  $Y$ , no caso, para base 10 (outra base poderia ter sido utilizada).

<sup>7</sup> É uma característica fundamental, como vimos, de leis de potência, a relação linear entre as variáveis no gráfico  $\log\text{-}\log$ . Outra conveniência deste tipo de gráfico é que ele permite a visualização de dados que se encontram em uma faixa de grande amplitude.

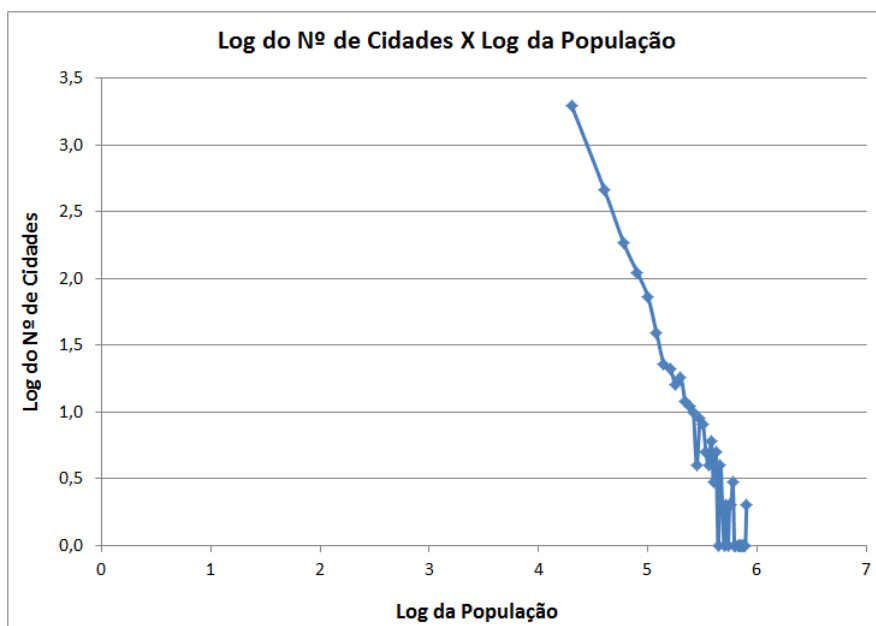


Figura 49: Log do Nº de cidades X Log da População.

Uma vez que assumimos se tratar de uma lei de potência, como devemos proceder para encontrar as constantes,  $\alpha$  e  $c$ , desta função?

Notando, pela equação (4.1), que  $-\alpha$  é o coeficiente angular da função linear desta equação, obtemos daí o valor de  $\alpha$ . Temos, portanto, que o valor de  $\alpha$  corresponde ao coeficiente angular da reta adaptada aos pontos do gráfico log-log. Lembrando, ainda, que para obter uma reta ajustada a estes pontos, podemos utilizar uma regressão linear segundo o método dos mínimos quadrados, conforme vimos no Capítulo 3; e, para isto, utilizamos algum software, como o Excel. Vejamos no gráfico da Figura 50 a reta ajustada aos pontos, por meio do Excel, e a equação desta reta.

Pela equação exibida no gráfico da Figura 50, temos que  $\alpha = 2,0917$ .

Já a constante  $c$  é obtida depois de determinado o  $\alpha$ . Para isto, consideramos que sendo a lei de potência uma função de densidade de probabilidade, é necessário que a área sob a curva de seu gráfico (em nosso exemplo, o gráfico da Figura 47), em todo o eixo das abscissas a partir de um certo valor de  $x$ , seja igual a 1. Para este processo de normalização, é necessário escolher um  $x$  mínimo ( $x_{min}$ ) a partir do qual a função se comporta de modo mais fidedigno a uma lei de potência. A escolha deste  $x_{min}$  nem sempre é trivial. Em nosso exemplo, esta escolha se deu quando optamos por considerar apenas as cidades com população igual, ou superior, a 10.000 habitantes.

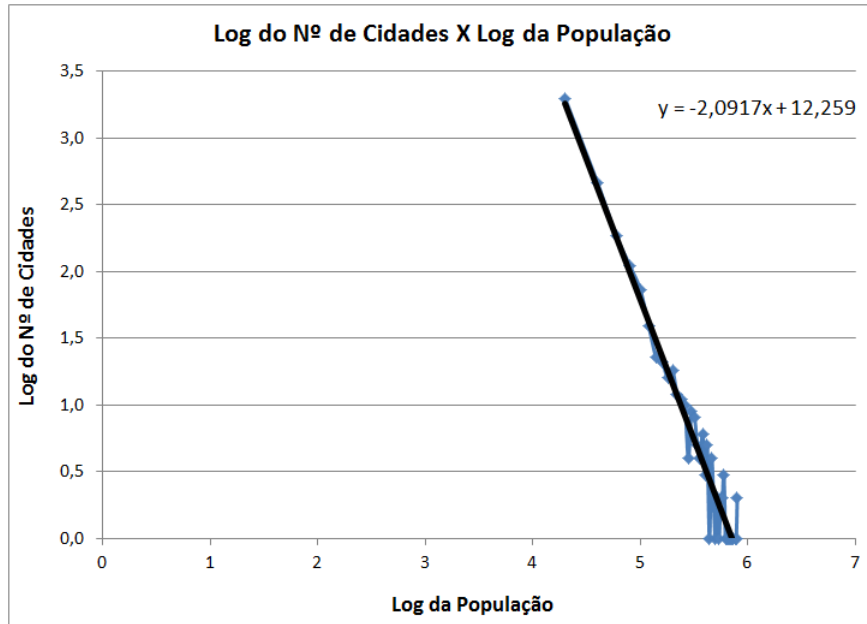


Figura 50: Log do Nº de cidades X Log da População, e a reta ajustada por Regressão Linear.

Conforme vimos, no Capítulo 2, a área sob a curva de  $p(x) = cx^{-\alpha}$ ,  $x > 0$  e  $c > 0$ , se aproxima de um número (converge para um número) para valores muito grandes de  $x$  ( $x$  tendendo ao infinito) apenas para  $\alpha > 1$ . Assumindo  $\alpha > 1$ , e impondo que a área sob a curva, a partir de  $x_{min}$ , com  $x$  tendendo ao infinito, deve ser igual a 1, obteremos o valor da constante  $c$ , escrevendo:

$$A_{-\alpha}[x_{min}, x] = 1, \quad \text{com } x \text{ tendendo ao infinito.}$$

Conforme a fórmula (2.20), do Capítulo 2, aplicada na equação acima:

$$\frac{c}{\alpha - 1} \cdot \left( \frac{1}{x_{min}^{\alpha-1}} - \frac{1}{x^{\alpha-1}} \right) = 1. \quad (4.2)$$

Para  $x$  muito grande a fração  $\frac{1}{x^{\alpha-1}}$  anula-se, tornando a equação (4.2), acima, equivalente a:

$$\frac{c}{\alpha - 1} \cdot \left( \frac{1}{x_{min}^{\alpha-1}} \right) = 1.$$

Isolando o  $c$  no membro esquerdo da equação:

$$\boxed{c = (\alpha - 1) \cdot x_{min}^{\alpha-1}}. \quad (4.3)$$

Obtém-se através de (4.3), o valor da constante  $c$  da função lei de potência, que denominaremos, doravante,  $p(x)$ . Podemos, ainda, reescrever a função  $p(x)$  com base em (4.3):

$$\begin{aligned} p(x) &= Cx^{-\alpha} \\ \Rightarrow p(x) &= (\alpha - 1) \cdot x_{min}^{\alpha-1} \cdot x^{-\alpha} \\ \Rightarrow p(x) &= (\alpha - 1) \cdot \frac{x_{min}^{-1}}{x_{min}^{-\alpha}} \cdot x^{-\alpha} \\ \Rightarrow p(x) &= \frac{(\alpha - 1)}{x_{min}} \cdot \left( \frac{x}{x_{min}} \right)^{-\alpha}. \end{aligned} \quad (4.4)$$

A equação (4.4) corresponde à Lei de Potência dados  $\alpha$  e  $x_{min}$ .

Aplicando a equação (4.3) em nosso problema:

$$\begin{aligned} C &= (2,0917 - 1) \cdot 10000^{2,0917-1} \\ \Rightarrow C &= 25404 \quad (\text{valor aproximado}). \end{aligned}$$

Portanto, a lei de potência, para o caso estudado, pode ser expressa da seguinte forma:

$$p(x) = 25404 \cdot x^{-2,0917}.$$

Já a probabilidade dos valores de  $x$  em um intervalo  $[a, b]$ , em relação a uma lei de potência  $p(x) = cx^{-\alpha}$ , sendo a área sob a curva, pode ser calculada por:

$$A_{-a}[a, b] = \frac{c}{\alpha - 1} \cdot \left( \frac{1}{a^{\alpha-1}} - \frac{1}{b^{\alpha-1}} \right),$$

Conforme a fórmula (2.20) do Capítulo 2.

Embora a utilização de histogramas, com larguras iguais, nos auxilie a visualizar melhor os dados, ela nos traz alguns inconvenientes. Vejamos o gráfico log-log com *todos* os dados do número de cidades em relação à população, a partir de 10.000 moradores, em histogramas de amplitude 20.000 (Figura 51).

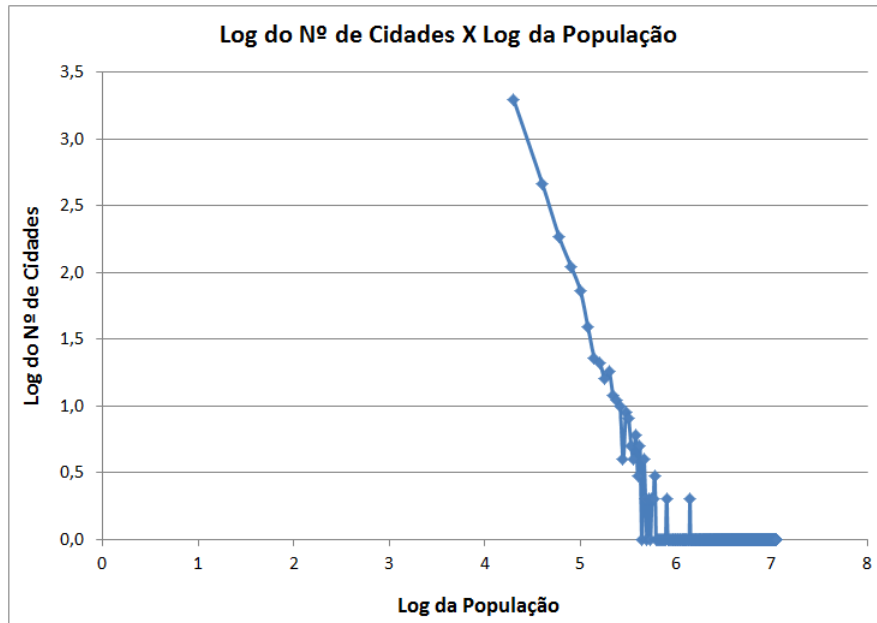


Figura 51: Log do Nº de cidades X Log da População. Incluindo todas as populações acima de 10.000.

Como é de se notar, há uma grande presença de “ruído” na cauda. Há dois fatores que contribuem para isto. Como os intervalos possuem o mesmo tamanho, aqueles mais à direita acabam tendo uma quantidade muito pequena de elementos. Alguns, inclusive, com nenhum elemento, valor que sequer pode ser plotado em um gráfico log-log (não calculamos logaritmo de zero). Ocorre, também, que embora os intervalos sejam fixos, as abscissas dos pontos não ficam igualmente espaçadas, dada a escala logarítmica, ficando muito próximas no final da cauda. Este ruído, naturalmente, prejudica o ajuste de uma reta.

Uma das alternativas para contornar este problema é a utilização de um procedimento chamado de binning logarítmico, em que as larguras dos histogramas não são iguais, mas aumentam progressivamente. Discorreremos mais sobre isto nas atividades do Capítulo 5, mais especificamente na Seção 5.5.

O melhor procedimento, no entanto, é utilizar a frequência acumulada, conforme explicaremos a seguir.

## 4.2 FREQUÊNCIA ACUMULADA

Terminamos a seção anterior, comentando as desvantagens proporcionadas pela utilização de histogramas, quando aplicados aos dados do número de cidades em relação ao tamanho da população. Vimos a grande presença de ruídos na cauda.

Uma outra alternativa, para organizarmos os dados, é a utilização da função de distribuição das frequências acumuladas. Ela apresenta, logo de início, uma grande vantagem; como veremos, não há na distribuição acumulada das frequências perda das informações individuais. Quando utilizamos histogramas, os dados dentro de um histograma são substituídos pelo valor médio dos valores do histograma, o que faz com que as informações individuais sejam perdidas.

Para aplicar a distribuição de frequência acumulada, em nosso caso, da distribuição do número de cidades em relação ao tamanho da população, realizamos as seguintes etapas: Primeiro listamos todas as populações acima de 10.000 moradores (não nos limitamos a 810.000 moradores) em ordem crescente. Essas populações constituem a variável  $X$  da nossa função. Os valores da função  $P(X)$  (com “ $P$ ” maiúsculo), de distribuição das frequências relativas (probabilidades) acumuladas, são então obtidos dividindo a soma das frequências dos valores maiores ou iguais a  $x$  pela frequência total (1 ou 100%). Também, é usual plotar as somas das frequências absolutas dos valores maiores ou iguais a  $x$ , ou seja, as quantidades acumuladas, em vez das probabilidades acumuladas. Ambos os gráficos se encontram nas Figuras 52 e 53. Todo o processo prático é desenvolvido na Atividade 5.6; nela obtemos, também, a função lei de potência para a distribuição dos dados.

Como é de se notar pelo gráfico (Figura 52), a função de distribuição das frequências acumuladas também segue uma lei de potência, porém, com expoente  $\alpha$  diferente do expoente obtido para a função não acumulada (por exemplo, pelo método do binning). No caso da função de distribuição acumulada, o expoente equivale a  $\alpha - 1$ . Vejamos como chegar a este resultado.

No caso da função  $P(x)$ , de distribuição das frequências relativas (probabilidades) acumuladas, associamos a cada valor de  $x$  a probabilidade dos valores maiores ou iguais a este  $x$ . Para um dado  $x$ , o valor de  $P(x)$  pode ser obtido, portanto, em relação à  $p(x) = Cx^{-\alpha}$ , considerando a área sob  $p(x)$ , como segue:

$$P(x) = A_{-\alpha}[x, m], \quad \text{com } m \text{ tendendo ao infinito.}$$



$$\text{Logo, } P(x) = \frac{C}{\alpha - 1} \cdot \left( \frac{1}{x^{\alpha-1}} - \frac{1}{m^{\alpha-1}} \right).$$

Para  $m$  muito grande, a fração  $\frac{1}{m^{\alpha-1}}$  anula-se (lembrando que  $\alpha > 1$ ). Com isto, a equação acima, pode ser escrita como:

$$\begin{aligned} P(x) &= \frac{C}{\alpha - 1} \cdot \left( \frac{1}{x^{\alpha-1}} \right) \\ \Rightarrow P(x) &= \frac{C}{\alpha - 1} \cdot x^{-(\alpha-1)}. \end{aligned} \quad (4.5)$$

Podemos constatar, em (4.5), que o expoente de  $P(x)$  é dado por  $\alpha - 1$  (em vez de  $\alpha$ ). Pode-se, então, fazer o caminho inverso e obter  $p(x)$  a partir da distribuição acumulada  $P(x)$ , obtendo, primeiramente, o valor de  $\alpha$  e, depois, o valor de  $C$ , conforme em (4.3).

Esta função de distribuição das frequências acumuladas possui muitas vantagens em relação ao agrupamento dos dados em histogramas. Pelo fato de que não é preciso agrupar os dados em histogramas, as informações individuais das populações não são perdidas, há valor para cada ocorrência de  $X$ . Evita-se a discussão sobre qual seria o tamanho mais conveniente dos histogramas (sejam eles de mesmo tamanho ou não), simplificando o processo. O ruído na cauda se torna menor, permitindo um melhor ajuste da reta.

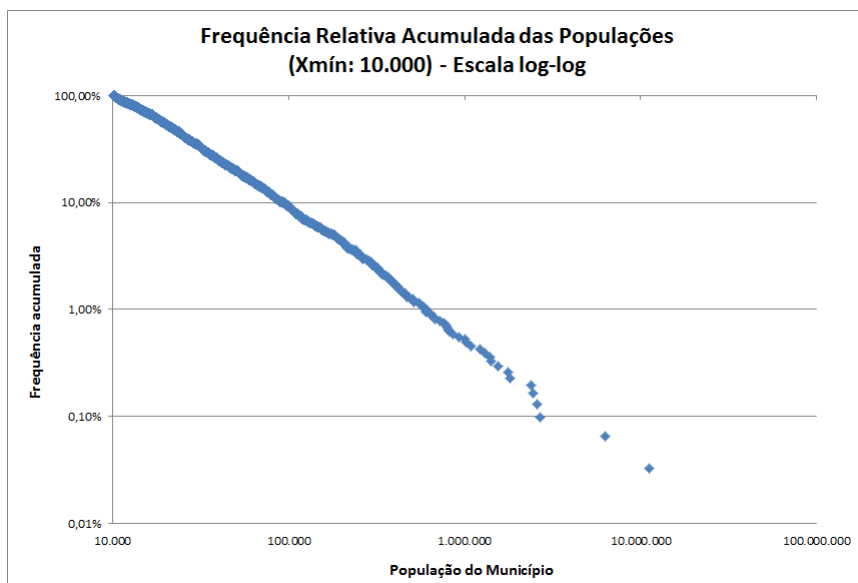


Figura 52: Frequência relativa (probabilidade) acumulada das Populações X População. Em escala log-log.

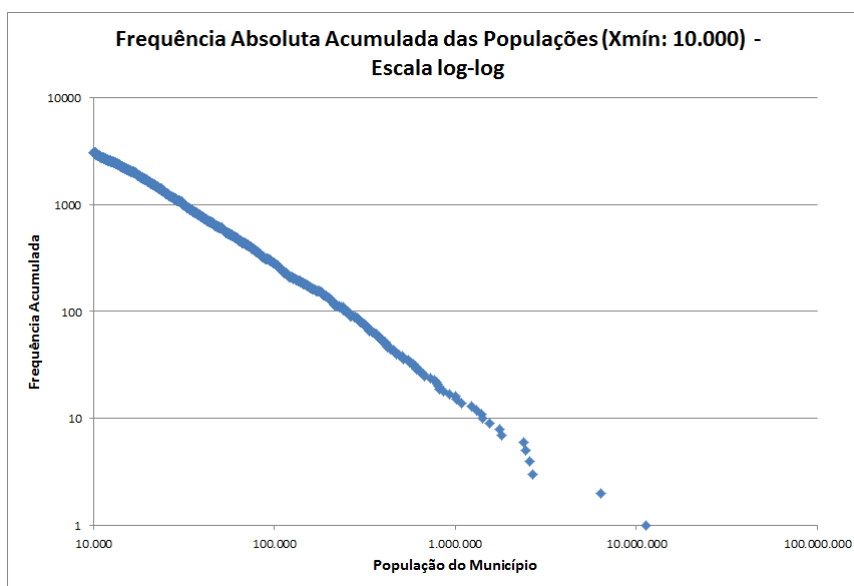


Figura 53: Frequência absoluta (quantidade) acumulada das Populações X População. Em escala log-log.

Distribuições de frequências acumuladas, que seguem a uma lei de potência, às vezes são chamadas de *Lei de Zipf* ou *Distribuição de Pareto*. A única distinção entre a lei de Zipf e a distribuição de Pareto é o modo como os dados são plotados<sup>8</sup>. Na lei de Zipf os dados são plotados com os valores de  $X$  no eixo horizontal e os valores de  $P(x)$  no eixo vertical. Na distribuição de Pareto os dados são plotados do modo inverso em relação aos eixos.

#### 4.3 EXEMPLOS DE LEI DE POTÊNCIA

Em Newman (2006, p. 5), [3], há diversos exemplos de aplicações de leis de potência em física, biologia, tecnologia e ciências sociais, os quais reproduziremos adiante<sup>9</sup>. Os gráficos encontram-se na Figura 54 e foram, igualmente, extraídos de Newman (2006, p.6). Em cada caso há um valor mínimo ( $x_{min}$ ) a partir do qual consideramos a lei de potência. Os valores de  $x$  abaixo do  $x_{min}$ , que foram descartados, encontram-se nas faixas cinzas dos gráficos.

(a) **Frequência de palavras em um texto:** A frequência com que as palavras ocorrem em um texto costuma obedecer a uma lei de potência. Esta tese foi examinada e

<sup>8</sup> De acordo com Newman (2006, p. 4)

<sup>9</sup> Os exemplos citados, a seguir, foram traduzidos/adaptados de NEWMAN (2006), p. 5.

confirmada por Zipf. Na Figura 54(a) a lei é aplicada em relação ao número de palavras (eixo das ordenadas) do livro *Moby Dick*, de Herman Melville, de acordo com a frequência acumulada de suas ocorrências (eixo das abscissas). Neste caso as palavras mais comuns são “the”, “of”, “and”, “a” e “to”.

Na quarta atividade, Seção 5.7, reproduzimos esta análise, fazendo-a passo a passo.

Para uma aplicação da lei de potência no livro “*Memórias Póstumas de Brás Cubas*” de Machado de Assis, conferir em Bortolossi, Queiroz, Silva (2012), [1].

(b) **Número de citações de artigos científicos:** O número de citações recebidas por artigos científicos parece também seguir a uma lei de potência. Os dados da Figura 54(b), foram obtidos do Índice de Citações Científicas, reunidas por Redner<sup>10</sup> e referem-se aos artigos publicados em 1981. A Figura 54(b) mostra o número de artigos (eixo das ordenadas) de acordo com a frequência acumulada do número de citações por eles recebidas (eixo das abscissas) entre seu ano de publicação, 1981, e junho de 1997.

(c) **Número de acessos a websites:** A distribuição acumulada do número de acessos recebidos pelos websites (servidores, não as páginas) durante um único dia, 1 de dezembro de 1997, por 60000 usuários do AOL (America Online Internet)<sup>11</sup>. O site com o maior número de acessos foi “yahoo.com”. A Figura 54(c) mostra o número de sites (eixo das ordenadas) em relação à frequência acumulada do número de acessos recebidos por estes sites (eixo das abscissas).

(d) **Número de cópias de livros vendidos:** A frequência acumulada do número de cópias vendidas na América dos 633 livros que venderam 2 milhões ou mais entre 1895 e 1965<sup>12</sup>. A Figura 54(d) mostra o número de livros (eixo das ordenadas) em relação à frequência acumulada do número de cópias por eles vendidas (eixo das abscissas).

(e) **Chamadas telefônicas:** A distribuição acumulada do número de chamadas recebidas em um dia por 51 milhões de usuários da AT&T, operadora telefônica americana de chamadas de longa distância<sup>13</sup>. O maior número de chamadas recebidas por um

10 REDNER, S. *How popular is your paper? An empirical study of the citation distribution.* *The European Physical Journal B* 4, 131-134 (1998), em Newman (2006, p.5), [3].

11 ADAMIC, L. A., HUBERMAN, B. A. *The nature of markets in the World Wide Web.* *Quarterly Journal of Electronic Commerce* 1, 512 (2000), em Newman (2006, p.5), [3].

12 HACKETT, A. *P.70 Years of Best Sellers, 1895-1965.* R. Bowker Company, New York, NY (1967), em Newman (2006, p.5), [3].

13 Aiello, W; Chung, F. e Lu, L. *A random graph model for massive graphs.* Em *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, p. 717 - 180, Association of Computing Machinery, New York (2000), em Newman (2006, p.5), [3].

usuário naquele dia foi 375746 (naturalmente, se trata do número telefônico de uma empresa). Ocorre distribuição similar para o número de mensagens de e-mail enviadas e recebidas. A Figura 54(e) mostra a quantidade de números de telefone (de usuários ou empresas) (eixo das ordenadas) em relação à frequência acumulada do número de chamadas por eles recebidas em um dia (eixo das abscissas).

(f) **Magnitude de terremotos:** A distribuição acumulada das magnitudes (em escala Richter) dos terremotos ocorridos na Califórnia entre janeiro de 1910 e maio de 1992, conforme registrado no catálogo de terremotos de Berkeley<sup>14</sup>. A magnitude, em escala Richter, de um terremoto, é proporcional ao logaritmo decimal da amplitude máxima do movimento detectado no terremoto, o que dispensa o uso de escala logarítmica para o eixo das abscissas. A lei de potência, neste caso, é, portanto, uma relação entre a amplitude e a frequência de ocorrência dos terremotos. Na Figura 54(f) tem-se a frequência acumulada de ocorrência dos terremotos (eixo das ordenadas) em relação à magnitude dos terremotos na escala Richter (eixo das abscissas).

(g) **Diâmetro das crateras lunares:** A distribuição acumulada dos diâmetros das crateras lunares<sup>15</sup>. No eixo das ordenadas, em vez de considerar a distribuição acumulada de todas as crateras da superfície da lua, considerou-se o número de crateras por quilômetro quadrado (normalizou-se). Isso explica porque neste eixo há valores abaixo de 1. No eixo das abscissas considerou-se os diâmetros das crateras (Figura 54(g)).

(h) **Intensidade das erupções solares:** A distribuição acumulada das intensidades dos picos de raios gama das erupções solares<sup>16</sup>. O gráfico mostra esta distribuição acumulada das intensidades dos picos contadas por segundo. As medições foram feitas a partir da órbita terrestre, por satélite, entre fevereiro de 1980 e novembro de 1989 (Figura 54(h)).

(i) **Intensidade de guerras:** A distribuição acumulada da intensidade de 119 guerras de 1816 a 1980<sup>17</sup>. A Intensidade de uma guerra é definida como o número de mortos em batalha dentre todos os participantes das nações envolvidas, dividido pela

14 Dados do National Geophysical Data Center: [www.ngdc.noaa.gov](http://www.ngdc.noaa.gov), em Newman (2006, p.5), [3].

15 NEUKUM, G. e IVANOV, B. A. **Crater size distributions and impact probabilities on Earth from lunar, terrestrial-planet and asteroid cratering data**. Em *T. Gehrels (ed.) Hazards Due to Comets and Asteroids*, p. 359 – 416, University of Arizona Press, Tucson, AZ (1994), em Newman (2006, p.5), [3].

16 Nasa Goddard Space Flight Center, [umbra.nascom.nasa.gov/smm/hxrbs.html](http://umbra.nascom.nasa.gov/smm/hxrbs.html). Acesso em: 24 jan. 2015, em Newman (2006, p.7), [3].

17 SMALL, M. e SINGER, J. D. **Resort to Arms: International and Civil Wars, 1816-1980**. Sage Publications, Beverly Hills (1982), em Newman (2006, p.7), [3].

soma das populações dessas nações e multiplicado por 10.000. Como exemplo, as intensidades da primeira e da segunda guerra mundial são, respectivamente, 141,5 e 106,3, o que significa que para cada 10.000 cidadãos das nações envolvidas há 141,5 e 106,3 mortos em batalha em cada uma destas guerras, respectivamente. A pior guerra do período analisado foi entre Paraguai e Bolívia de 1932 a 1935 com uma intensidade de 382,4. O gráfico da Figura 54(i), exibe o número de guerras (eixo das ordenadas) em relação à intensidade acumulada das guerras (eixo das abscissas).

(j) **A riqueza das pessoas mais ricas:** A distribuição acumulada do total de riqueza das pessoas mais ricas dos Estados Unidos<sup>18</sup>. A riqueza, ou “rede de valor agregado”, de alguém é definida como o valor total, em dólares, dos preços de mercado de todas as suas posses menos seus débitos. Por exemplo, em 2003, o americano mais rico foi William H. Gates III (Bill Gates), com uma riqueza de 46 bilhões de dólares. Muito desta riqueza estava em forma de ações da companhia por ele fundada: Microsoft Corporation. Estas ações não são a mesma coisa que dinheiro “vivo”. Inclusive, se Bill Gates quiser vendê-las é possível que seus valores se depreciem, tornando-se menores que seus atuais valores de mercado. A Figura 54(j) mostra o número de pessoas (eixo das ordenadas) em relação à riqueza acumulada, em dólares (eixo das abscissas).

(k) **Frequência dos nomes de família:** Distribuição acumulada da frequência de ocorrência, nos Estados Unidos, dos 89000 nomes de família mais comuns, conforme registrado pelo Censo americano de 1990 (US Census Bureau)<sup>19</sup>. Distribuições similares são observadas para nomes em algumas outras culturas (exemplo japonesa), porém não em todas (exemplo Coreana). Na figura 54(k) tem-se o número de nomes (eixo das ordenadas) em relação à frequência acumulada dos mesmos (eixo das abscissas).

(l) **Populações das cidades:** Distribuição acumulada dos tamanhos das populações das cidades dos Estados Unidos, conforme registrado pelo Censo americano em 2000<sup>20</sup>. A Figura 54(l) exibe o número de cidades (eixo das ordenadas) em relação à frequência acumulada do tamanho da população (eixo das abscissas).

18 *Forbes Magazine*, 6 out. 2003, em Newman (2006, p.7), [3].

19 Newman (2006, p.7), [3].

20 NEWMAN (2006, p.7), [3]

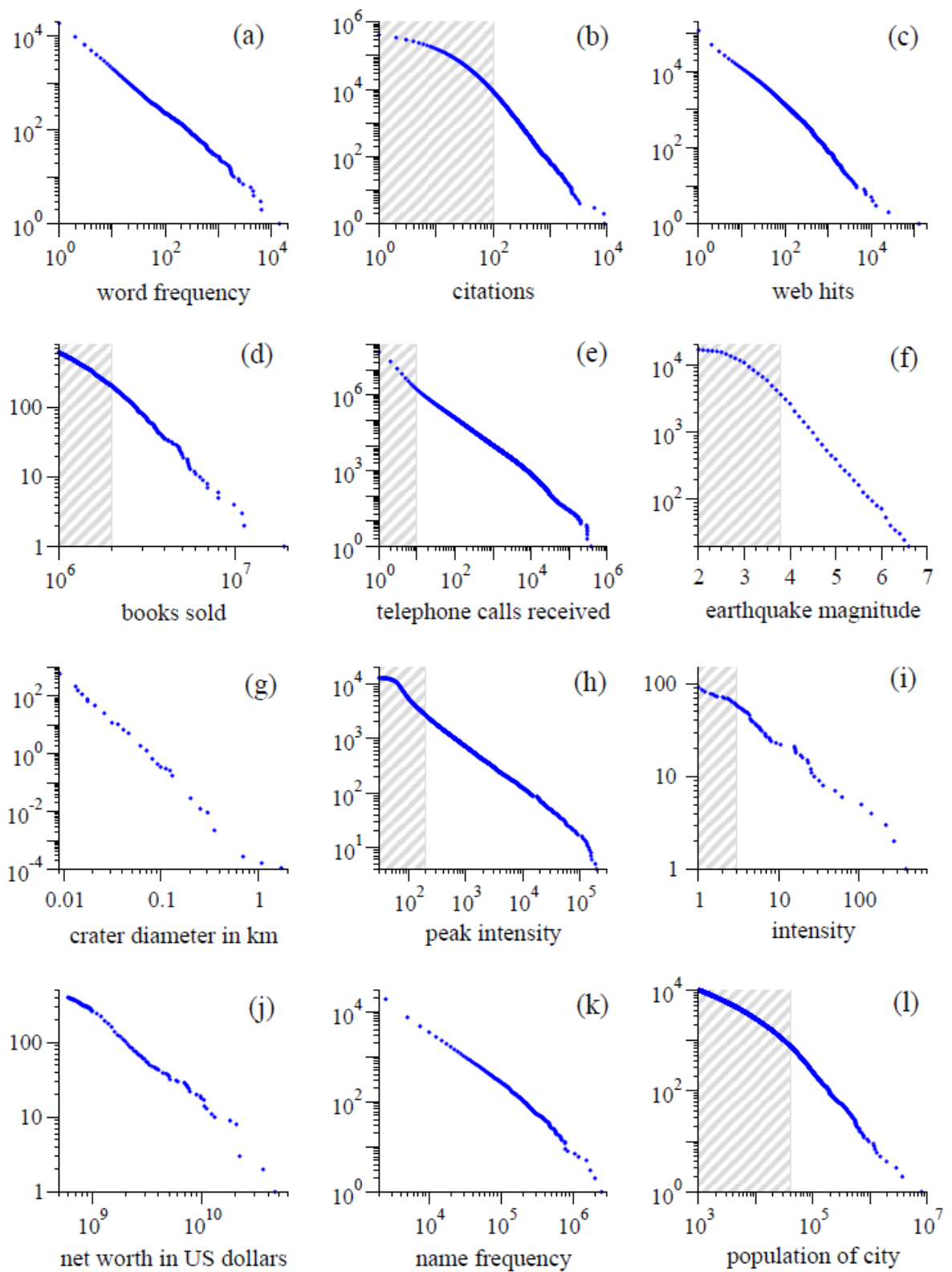


Figura 54: Gráficos dos exemplos de (a) a (l). Extraído de Newman (2006, p.6), [3].

## 4.4 VALOR MÉDIO E VARIÂNCIA

Na subseção 1.6.2, vimos o valor médio (esperança) e a variância/desvio padrão de v.a. discretas. Adaptaremos estes conceitos para leis de potência. Talvez esta seção não seja indicada para aplicação a alunos do ensino médio, todavia, dado o interesse do assunto, julgamos que não poderíamos deixá-lo “passar em branco”.

Primeiro, recordemos a equação da esperança para v.a. discretas, de 1.5, para uma v.a. discreta  $X$ :

$$E(X) = \sum_{i=1}^n p_i x_i,$$

onde  $p_i$  é a probabilidade de ocorrência do valor  $x_i$ .

Consideremos, agora, uma v.a. contínua  $X$  com função densidade de probabilidade  $f(x)$ , tal que:

$$f(x) = cx^{-\alpha}, \quad \text{com } x \geq x_{\min} > 0, \quad c > 0 \quad \text{e} \quad \alpha > 1 \quad (4.6)$$

A v.a.  $X$  segue uma lei de potência, portanto.

A probabilidade de ocorrência dos valores de  $X$ , em um intervalo  $[c, d]$ , corresponde, como vimos na Seção 1.7, à área sob a curva de  $f(x)$  em  $[c, d]$ . Esta área pode ser aproximada por  $f(x_i) \cdot h$  (Figura 55), em que  $x_i = \frac{c+d}{2}$  (ponto médio do intervalo) e  $h = d - c$  (amplitude do intervalo). Quanto menor o valor de  $h$ , tanto melhor será a aproximação da área.

Consideremos, agora, um intervalo  $[a, b]$ . Neste intervalo,  $a$  corresponde ao  $x$  mínimo da lei de potência ( $x_{\min}$ ); o valor de  $b$  o consideraremos, posteriormente, muito grande, ou seja,  $b$  tendendo ao infinito, como convém a uma lei de potência. Este intervalo,  $[a, b]$ , será subdividido em  $n$  intervalos iguais, cada um com amplitude  $h = \frac{b-a}{n}$  (Figura 56). As limitações à direita, em cada intervalo, serão denominadas  $a_1, a_2, a_3, \dots, a_i, \dots, a_n$ .

Em cada um destes  $n$  intervalos, tomaremos  $x_i = \frac{a_i + a_{i-1}}{2}$ , sendo, então,  $x_i$  o ponto médio do  $i$ -ésimo intervalo (Figura 57). A área sob a curva em cada intervalo,  $A_{-a}[a_{i-1}, a_i]$ , será aproximada da seguinte forma:

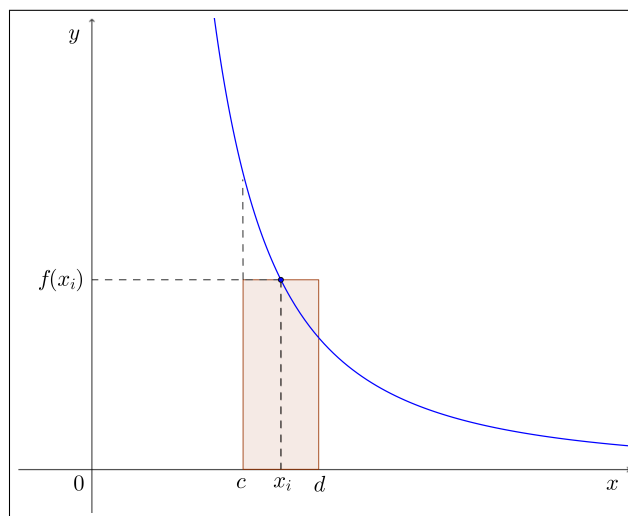


Figura 55: Área do retângulo de base com medida  $d - c$  e altura  $f(x_i)$ .

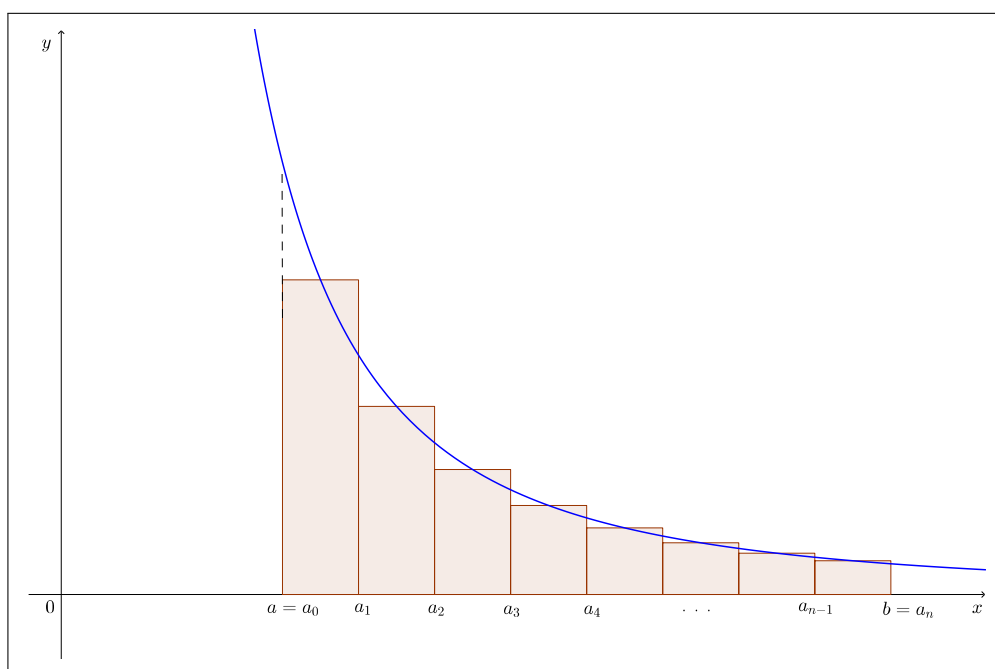


Figura 56: Divisão do intervalo  $[a, b]$  em  $n$  intervalos iguais.



$$A_{-a}[a_{i-1}, a_i] \approx f(x_i) \cdot h.$$

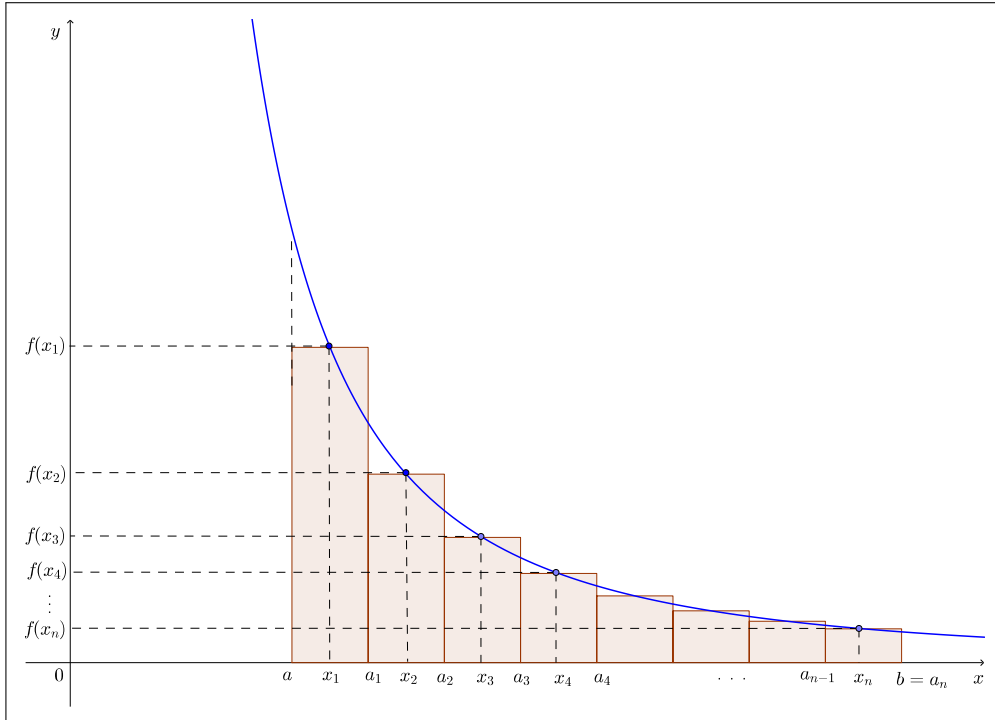


Figura 57:  $n$  retângulos em torno à curva  $f(x) = cx^{-\alpha}$ , no intervalo  $[a, b]$ .

Vamos considerar a v.a. discreta  $Y$  que assume os valores  $x_i$ .  $P(Y)$  será a função de probabilidade que associa os valores  $x_i$  com as probabilidades destes valores, assumidas como sendo  $f(x_i) \cdot h$ :

$$P(y_i = x_i) = f(x_i) \cdot h. \quad (4.7)$$

A esperança de  $Y$  pode ser escrita, conforme em (1.5):

$$\begin{aligned}
E(Y) &= \sum_{i=1}^n p_i x_i \\
\Rightarrow E(Y) &= \sum_{i=1}^n f(x_i) \cdot h \cdot x_i \\
\Rightarrow E(Y) &= \sum_{i=1}^n c x_i^{-\alpha} \cdot x_i \cdot h \\
\Rightarrow E(Y) &= \sum_{i=1}^n c x_i^{-\alpha+1} \cdot h.
\end{aligned}$$

Para  $n$  muito grande ( $n$  tendendo ao infinito), ou seja, para uma quantidade muito grande de intervalos, assumimos que a esperança de  $Y$  se torna igual à esperança de  $X$ :

$$E(X) = E(Y), \quad \text{quando } n \text{ tende ao infinito.} \quad (4.8)$$

Por outro lado, para  $n$  muito grande, a  $E(Y)$  torna-se equivalente à área sob a curva de  $g(x) = cx^{-\alpha+1}$  em  $[a, b]$ ; pois os intervalos tornam-se cada vez menores e seus pontos médios ( $x_i$ ) estão associados aos valores  $c x_i^{-\alpha+1}$ , da função  $g(x)$ . Então, para  $n$  muito grande:

$$E(Y) = A_{-\alpha+1}[a, b], \quad (4.9)$$

onde  $A_{-\alpha+1}[a, b]$  é a área sob  $g(x) = cx^{-\alpha+1}$  em  $[a, b]$ .

Para  $n$  muito grande, de (4.8) e (4.9) :

$$E(X) = A_{-\alpha+1}[a, b] = A_{-(\alpha-1)}[a, b].$$

E, conforme a equação (2.20):

$$\begin{aligned}
E(X) &= \frac{c}{(\alpha-1)-1} \left( \frac{1}{a^{(\alpha-1)-1}} - \frac{1}{b^{(\alpha-1)-1}} \right) \\
\Rightarrow E(X) &= \frac{c}{\alpha-2} \left( \frac{1}{a^{\alpha-2}} - \frac{1}{b^{\alpha-2}} \right). \quad (4.10)
\end{aligned}$$

Como consideramos as leis de potência a partir de um  $x_{min}$ , e com  $X$  assumindo valores cada vez maiores, sem limite, podemos fazer, na equação acima,  $a = x_{min}$  e  $b$  muito grande ( $b$  tendendo ao infinito). Para  $\alpha > 2$ , a expressão  $\frac{1}{b^{\alpha-2}}$  se anula e, neste caso:

$$E(X) = \frac{c}{\alpha - 2} \left( \frac{1}{x_{min}^{\alpha-2}} \right). \quad (4.11)$$

É importante frisar que  $E(X)$  só existirá, ou seja, convergirá para um número, se  $\alpha > 2$ .

Fazendo  $c = (\alpha - 1) \cdot x_{min}^{\alpha-1}$  (equação (4.3)) em (4.11):

$$\begin{aligned} E(X) &= \frac{(\alpha - 1) \cdot x_{min}^{\alpha-1}}{\alpha - 2} \cdot \left( \frac{1}{x_{min}^{\alpha-2}} \right) \\ \Rightarrow E(X) &= \frac{\alpha - 1}{\alpha - 2} \cdot x_{min}. \end{aligned}$$

Conforme discutimos,  $E(X)$  só convergirá para um número se  $\alpha > 2$ . Isto significa que a lei de potência só terá um valor médio se  $\alpha > 2$ . Para  $1 < \alpha \leq 2$ , a lei de potência não terá valor médio, pois (4.10) não convergirá para um número quando  $b$  tende ao infinito. O que fazer, então, neste caso? Para  $1 < \alpha \leq 2$ , o valor médio deverá ser calculado por meio dos dados reais, observados, da amostra. Porém, este valor médio não poderá ser generalizado, pois para outras amostras ele deverá assumir valores muito diferentes, dada a divergência em (4.10).

Podemos obter a variância da v.a. contínua  $X$  de modo análogo ao que fizemos para  $E(X)$ . Omitiremos esta dedução, bem como um resultado que dela se segue e que é equivalente à equação (1.7) (variância de uma v.a. discreta):

$$Var(X) = E(X^2) - [E(X)]^2. \quad (4.12)$$

Logo, para a variância precisamos de  $E(X^2)$ , valor, este, chamado de segundo momento<sup>21</sup>. Procedendo de modo análogo ao realizado para obter  $E(X)$ , temos:

<sup>21</sup>  $E(X^2)$  é conhecido como o segundo momento.  $E(X)$  é o primeiro momento. Analogamente,  $E(X^3)$  é o terceiro momento,  $E(X^4)$  é o quarto momento, e assim por diante. Os momentos são importantes na caracterização de distribuições de probabilidade.

$$E(X^2) = A_{-\alpha+2}[a, b] = A_{-(\alpha-2)}[a, b].$$

Logo,

$$\begin{aligned} E(X^2) &= \frac{c}{(\alpha-2)-1} \left( \frac{1}{a^{(\alpha-2)-1}} - \frac{1}{b^{(\alpha-2)-1}} \right) \\ \Rightarrow E(X^2) &= \frac{c}{\alpha-3} \left( \frac{1}{a^{\alpha-3}} - \frac{1}{b^{\alpha-3}} \right). \end{aligned} \quad (4.13)$$

Para  $b$  muito grande e  $\alpha > 3$ , a expressão  $\frac{1}{b^{\alpha-3}}$  anula-se, e, obtemos:

$$E(X^2) = \frac{c}{\alpha-3} \left( \frac{1}{a^{\alpha-3}} \right).$$

Fazendo  $a = x_{min}$  e  $c = (\alpha-1)x_{min}^{\alpha-1}$ , conforme a equação (4.3), obtemos:

$$\begin{aligned} E(X^2) &= \frac{(\alpha-1) \cdot x_{min}^{\alpha-1}}{\alpha-3} \left( \frac{1}{x_{min}^{\alpha-3}} \right) \\ \Rightarrow E(X^2) &= \frac{\alpha-1}{\alpha-3} \cdot x_{min}^2. \end{aligned}$$

Conforme dissemos, a expressão em (4.13) convergirá para um número se, e somente se,  $\alpha > 3$ . Logo, só haverá variância se  $\alpha > 3$ . Para  $1 < \alpha \leq 3$  a variância só poderá ser calculada com os valores observados, porém, tal como discutimos em relação à esperança, tal valor não deverá ser generalizado para além desta amostra.

Notando que para existir  $E(X)$  é necessário que  $\alpha > 2$ , e para existir  $E(X^2)$  é necessário que  $\alpha > 3$ , para existir  $E(X^m)$  demonstra-se que :  $\alpha > m + 1$ , ou, de modo equivalente,  $m < \alpha - 1$ . Com relação à fórmula:

$$E(X^m) = \frac{\alpha-1}{\alpha-1-m} \cdot x_{min}^m. \quad (4.14)$$

#### 4.5 DISTRIBUIÇÕES DE CAUDA LEVE E CAUDA PESADA

Caracterizamos uma v.a. contínua  $X$  (ou sua função densidade de probabilidade) como de cauda leve quando todos os seus momentos existem. Em oposição, uma

v.a. contínua  $X$  (ou sua função densidade de probabilidade) é considerada de cauda pesada quando nem todos os seus momentos existem.

Como vimos na seção precedente, 4.4, os momentos são expressos como  $E(X^m)$ . Para  $m = 1$ , temos o primeiro momento; para  $m = 2$ , o segundo momento; e assim por diante. O primeiro momento corresponde à esperança da v.a.  $X$ . O segundo momento é determinante para a variância.

No caso das leis de potência, a existência dos momentos está condicionada, como vimos na Seção 4.4, ao valor de  $\alpha$ . Para existir o  $i$ -ésimo momento  $E(X^m)$  é necessário que  $m < \alpha - 1$ ; portanto, para qualquer lei de potência a existência de seus momentos é limitada e depende do valor de seu  $\alpha$ , sendo, então, uma distribuição de cauda pesada.

Uma discussão aprofundada sobre distribuições de cauda pesada encontra-se em Pinho (2012), [4].



---

## ATIVIDADES COM LEIS DE POTÊNCIA

---

Neste capítulo serão dadas algumas atividades concernentes a Lei de Potência. Essas atividades são sugeridas para serem aplicadas aos alunos do Ensino Médio, conduzidas pelo professor, com as adaptações que forem julgadas necessárias. As três primeiras atividades correspondem às análises que foram feitas no Capítulo 4; elas se referem às Leis de Potência aplicadas à distribuição dos municípios em relação à quantidade de habitantes. Na primeira atividade, os dados são agrupados em histogramas com amplitude 20.000. Na segunda atividade, foi aplicada a técnica do Binning Logarítmico, em que os dados são agrupados em histogramas com amplitudes variadas. Na terceira atividade, é utilizada a frequência acumulada dos dados. Na quarta atividade, é aplicada a Lei de Potência para a distribuição acumulada das frequências das palavras de um livro.

### 5.1 OBJETIVO DAS TRÊS PRIMEIRAS ATIVIDADES

O objetivo das três primeiras atividades é aplicar Lei de Potência para descrever a distribuição dos municípios brasileiros em relação à quantidade de habitantes.

### 5.2 CARACTERÍSTICAS DOS DADOS DAS TRÊS PRIMEIRAS ATIVIDADES

Há, ao todo, 5.565 municípios perfazendo um total de 190.755.799 habitantes. Destes municípios, o mais numeroso é São Paulo (SP) com 11.253.503 habitantes. Conforme discutido na Seção 4.1, foram considerados na análise somente os municípios com

mais de 10.000 habitantes, o que inclui 3.052 municípios, que correspondem a 93% da população total, somando juntos 177.839.519 habitantes.

### 5.3 FONTE DOS DADOS DAS TRÊS PRIMEIRAS ATIVIDADES E PROCEDIMENTOS INICIAIS

Os dados que serão utilizados são do censo demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE) do ano 2010. Eles foram extraídos do site [www.sidra.ibge.gov.br](http://www.sidra.ibge.gov.br), da seguinte forma: No campo “Seções”, na 1ª coluna à esquerda, foi selecionado: “Demográfico e Contagem”. Na nova página aberta, no campo “Demográfico 2010”, foi selecionado “Sinopse”. Abriu-se a tabela 608: “População residente, por situação do domicílio e sexo”. No campo “Unidade Territorial (31.532)” selecionou-se em “Município (5.565)” a opção “Tudo”. Foi clicado o botão “Ok”, no final do quadro. Na tabela aberta foram selecionados todos os dados; os quais foram copiados e colados em uma planilha do Excel. Nesta planilha foi excluída a 1ª linha dos dados: “Brasil - 190.755.799”, por não se tratar de um município. As cidades foram dispostas em ordem decrescente de tamanho das populações. Foram excluídos os municípios com menos de 10.000 habitantes.

### 5.4 1ª ATIVIDADE: POPULAÇÕES DAS CIDADES AGRUPADAS EM HISTOGRAMAS COM AMPLITUDES IGUAIS (20.000).

Antes de realizar esta atividade, convém a leitura da Seção 4.1.

Organizaremos esta atividade em etapas. Vamos a elas:

(1) Na planilha dos dados, criar uma coluna “Bloco” com uma sequência de valores de 10.000 a 11.270.000 de 20.000 em 20.000, conforme a Figura 58.

(2) Na parte superior da tela, clicar em “Dados” e posteriormente em “Análise de Dados”. Na janela aberta, selecionar a opção “Histograma” e clicar em “Ok”. Na janela aberta, preencher o campo “Intervalo de entrada” selecionando os dados da coluna “População” (ver Figura 58 da etapa 1), e preencher o campo “Intervalo do bloco” selecionando os dados da coluna “Bloco” (Figura 58 da etapa 1). Clicar em “Ok”. Uma nova planilha surgirá (Figura 59).



5.4 1º ATIVIDADE: POPULAÇÕES DAS CIDADES AGRUPADAS EM HISTOGRAMAS COM AMPLITUDES IGUAIS (

	A	B	C	D	E
1	<b>Município</b>	<b>População</b>		<b>Intervalos</b>	
2	São Paulo - SP	11.253.503		10.000	
3	Rio de Janeiro - RJ	6.320.446		30.000	
4	Salvador - BA	2.675.656		50.000	
5	Brasília - DF	2.570.160		70.000	
6	Fortaleza - CE	2.452.185		90.000	
7	Belo Horizonte - MG	2.375.151		110.000	
8	Manaus - AM	1.802.014		130.000	
9	Curitiba - PR	1.751.907		150.000	
10	Recife - PE	1.537.704		170.000	
11	Porto Alegre - RS	1.409.351		190.000	
12	Belém - PA	1.393.399		210.000	
13	Goiânia - GO	1.302.001		230.000	
14	Guarulhos - SP	1.221.979		250.000	

Figura 58: Etapa 1 da 1º atividade.

	A	B	C
1	<i>Bloco</i>	<i>Frequência</i>	
2	10.000	0	
3	30.000	1982	
4	50.000	462	
5	70.000	185	
6	90.000	111	
7	110.000	73	
8	130.000	39	
9	150.000	23	
10	170.000	21	
11	190.000	16	
12	210.000	18	

Figura 59: Etapa 2 da 1º atividade.

O valor do bloco na primeira coluna indica qual é o limite superior do intervalo do histograma correspondente. O bloco 30.000 tem frequência 1.982, o que significa que há 1.982 municípios com população entre 10.000 e 30.000 habitantes, e assim por diante.

(3) Inserir entre as colunas “Bloco” e “Frequência” uma coluna com o nome “Valor Médio”. Esta coluna será preenchida com os valores médios dentro de cada bloco. Na célula B3, aplique a fórmula  $(A3 + A2)/2$ ; com isto, preenchamos B3 com a média entre 10.000 e 30.000, que são os extremos do primeiro bloco. Para o restante desta coluna “Valor Médio”, arrastar a fórmula utilizada em B3. A coluna deverá estar preenchida conforme a Figura 60.

Ao lado da coluna “Frequência”, acrescentar uma nova coluna: “Freq. Relativa” com as frequências relativas, obtidas da divisão de cada frequência (coluna C) pela soma das frequências (Figura 60).

	A	B	C	D
1	<i>Bloco</i>	<i>Valor Médio</i>	<i>Freqüência</i>	<i>Freq. Relativa</i>
2	10.000		0	
3	30.000	20.000	1982	0,649410223
4	50.000	40.000	462	0,151376147
5	70.000	60.000	185	0,06061599
6	90.000	80.000	111	0,036369594
7	110.000	100.000	73	0,023918742
8	130.000	120.000	39	0,012778506
9	150.000	140.000	23	0,007536042
10	170.000	160.000	21	0,006880734
11	190.000	180.000	16	0,005242464
12	210.000	200.000	18	0,005897772
13	230.000	220.000	12	0,003931848

Figura 60: Etapa 3 da 1º atividade.

(4) Ao lado da coluna “Freq. Relativa”, acrescentar a coluna “Densidade de Freq.” com as densidades de cada frequência, obtidas da divisão de cada frequência relativa pela amplitude do intervalo correspondente (no caso as amplitudes são todas iguais a 20.000) (Figura 61).

	A	B	C	D	E
1	<i>Bloco</i>	<i>Valor Médio</i>	<i>Freqüência</i>	<i>Freq. Relativa</i>	<i>Densidade de Freq.</i>
2	10.000		0		
3	30.000	20.000	1982	0,649410223	0,0000324705
4	50.000	40.000	462	0,151376147	0,0000075688
5	70.000	60.000	185	0,06061599	0,0000030308
6	90.000	80.000	111	0,036369594	0,0000018185
7	110.000	100.000	73	0,023918742	0,0000011959
8	130.000	120.000	39	0,012778506	0,0000006389
9	150.000	140.000	23	0,007536042	0,0000003768
10	170.000	160.000	21	0,006880734	0,0000003440
11	190.000	180.000	16	0,005242464	0,0000002621
12	210.000	200.000	18	0,005897772	0,0000002949
13	230.000	220.000	12	0,003931848	0,0000001966

Figura 61: Etapa 4 da 1º atividade.

(5) Selecionar na coluna “Valor Médio” os valores de 20.000 a 480.000 e, simultaneamente, selecionar na coluna “Densidade de Freq.” os valores correspondentes. Na parte superior da tela, clicar em “Inserir” e, no campo “Gráficos”, selecionar “Dispersão” e a opção “Dispersão com Linhas Suaves e Marcadores” (Figura 62). Fazer, depois, um gráfico nos mesmos moldes, porém, incluindo todos os blocos e todas as densidades de frequência (Figura 63).

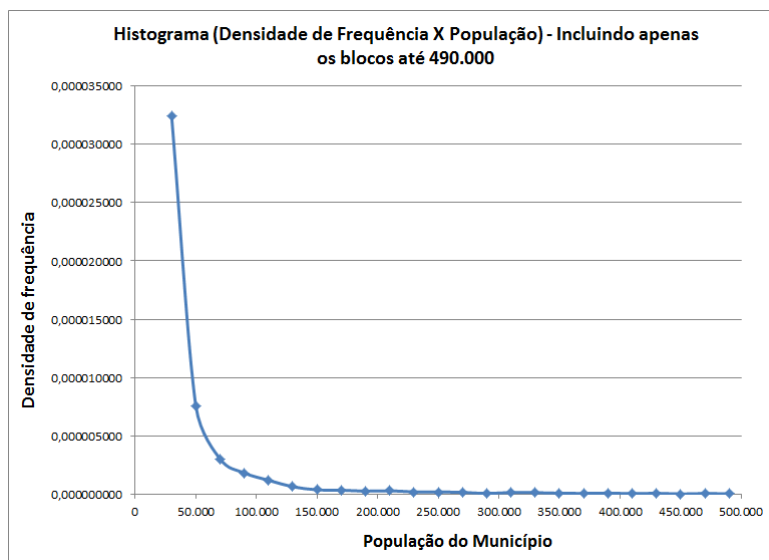


Figura 62: Gráfico incluindo os blocos de até 490.000.

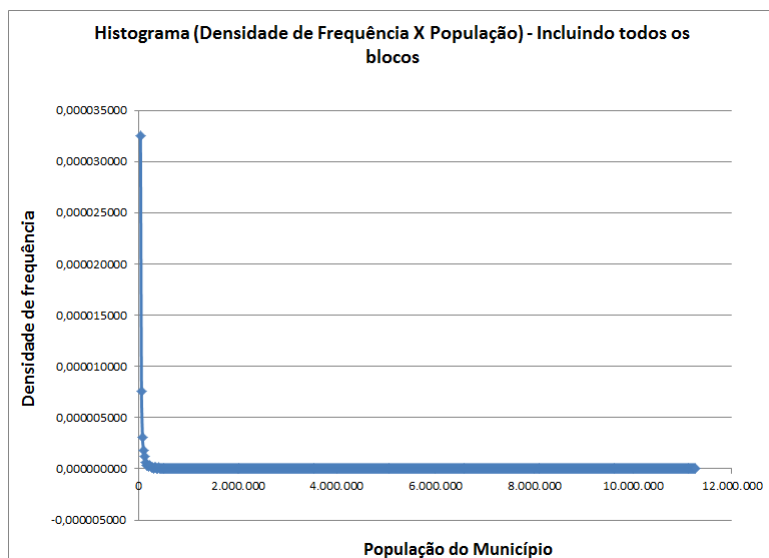


Figura 63: Gráfico incluindo todos os blocos.

O professor pode comentar sobre estes gráficos com os alunos, mostrando a relação entre a densidade de frequência e o número de habitantes dos municípios. Mostrar que há muitos municípios com poucos habitantes e poucos municípios com muitos habitantes. Pode, também, mostrar que alguns blocos foram deixados de fora do gráfico da Figura 62 para evitar a poluição visual, presente no gráfico da Figura 63, e que esta dificuldade em inserir todos os dados no gráfico se dá pela grande amplitude desses valores, variando de 10.000 a mais de 10 milhões.

(6) Para melhorar a visualização do gráfico da Figura 63, sem exclusão de dados, serão alteradas as escalas dos dois eixos X e Y para logarítmica (gráfico da Figura 64). Primeiramente, deve-se copiar e colar o gráfico da Figura 63 na mesma planilha. Clicando com o botão direito do mouse sobre o eixo X, selecionar a opção “Formatar Eixo”, na janela aberta selecionar “Escala Logarítmica”. Em “Opções de Eixo”, na opção “Mínimo”, selecionar “Fixo” e digitar 10.000. Na mesma janela, em “Rótulos do eixo”, selecionar “Inferior”. Pressionar o botão “Fechar”. Clicar o botão direito do mouse sobre o eixo Y, selecionar “Formatar Eixo” e na janela aberta selecionar “Escala Logarítmica”.

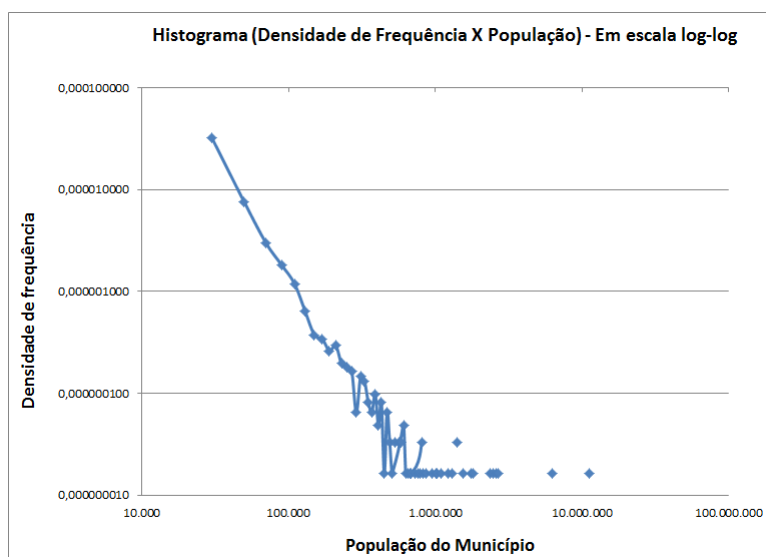


Figura 64: Gráfico da densidade de frequência em escala log-log.

O professor pode aproveitar para explicar aos alunos como funciona um gráfico em escala logarítmica; pode, ainda, comentar de outras situações em que esta escala é utilizada; como no mercado financeiro, por exemplo.

Nesta primeira atividade não obteremos uma lei de potência para os dados, devido o excesso de ruído na cauda. Este excesso de ruídos inviabiliza o ajuste de uma reta para os pontos do gráfico. Uma discussão sobre isto foi realizada no final da Seção 4.1, ali apontamos duas soluções; a primeira, utilizando o método Binning Logarítmico; e a segunda, por meio do uso das frequências acumuladas.

Na próxima atividade discutiremos o Binning Logarítmico, e na terceira atividade, o uso de frequências acumuladas.

#### 5.5 2º ATIVIDADE: POPULAÇÕES DAS CIDADES AGRUPADAS EM HISTOGRAMAS COM AMPLITUDES DIFERENTES (BINNING LOGARÍTMICO).

Aplicaremos, nesta atividade, a técnica conhecida como Logarithmic Binning, que traduziremos por Binning Logarítmico. Fizemos menção a ela no final da Seção 4.1 e deixamos, para esta presente atividade, a explicação de como ela funciona. Portanto, antes da atividade propriamente dita, faremos algumas considerações sobre esta técnica.

*O binning logarítmico é o procedimento que consiste em variar as larguras dos histogramas e, para isto, costuma-se determinar uma largura para o primeiro histograma e, do segundo em diante, multiplicamos a largura do histograma anterior por um fator constante.* Em nosso caso, utilizaremos para o primeiro histograma uma largura de 10.000 (população) e como fator 1,8. Ao realizar tal procedimento, não podemos nos esquecer de normalizar as frequências em densidades de frequências, ou seja, dividir a frequência relativa de cada intervalo pela sua respectiva largura (amplitude).

Uma pergunta cabível é o porquê de tal procedimento. Não bastaria plotar os dados agrupados em histogramas com larguras iguais, tal como fizemos no gráfico da Figura 47, apenas mudando os valores das variáveis para os seus logaritmos (gráfico log-log)? Porém, como vimos no gráfico da Figura 51, há uma grande quantidade de ruído na cauda, que impede um ajuste adequado para uma reta. Como discutimos, este ruído é gerado, sobretudo, por dois fatores. Com larguras fixas, os histogramas mais à direita ficam com uma quantidade muito pequena de elementos, por vezes nula. Por outro lado, com larguras iguais para os histogramas, o gráfico em log-log tende a ficar ruim, pois os pontos não ficam igualmente espaçados em relação ao eixo das abscissas.

O procedimento do binning logarítmico evita ambos inconvenientes. Como as larguras dos intervalos variam, aumentando exponencialmente, nos últimos intervalos,

embora a quantidade de elementos seja pequena, não ocorre a oscilação que vimos na Figura 51. Não há intervalos com nenhum elemento. Além disso, independentemente da base escolhida para o logaritmo (em nosso caso utilizaremos base 10), as abscissas dos pontos ficam igualmente espaçadas. Tudo isto contribui para um melhor ajustamento da reta, conforme veremos nesta atividade.

Esta atividade será parecida com a primeira, porém, nela, os histogramas terão amplitudes diferentes; o primeiro com largura 10.000, e cada um dos demais com amplitude equivalente à amplitude do histograma anterior multiplicada por 1,8. Aplicaremos a Lei de Potência para modelar as densidades de frequência.

Após estas explicações, vamos às etapas:

(1) Escreveremos quais serão os limites superiores dos intervalos dos histogramas. Na planilha dos dados (veja Seção 5.3), acrescentar a coluna “Bloco” (Figura 65). Chamaremos os valores de cada célula desta coluna de  $c_1, c_2, c_3, \dots, c_n, \dots$ . Faça  $c_1 = 10.000$ ,  $c_2 = 20.000$ , e para os demais valores ( $n > 2$ ) aplicar  $c_n = c_{n-1} + (c_{n-1} - c_{n-2}) \cdot 1,8$ . O procedimento deve ser repetido (“arrastado”) até  $c_{13} = 14.457.892$ . Nesta fórmula, tem-se:

$c_n$ : Limite superior do intervalo do histograma  $n - 1$  ( $n > 2$ ).

$c_{n-1}$ : Limite superior do intervalo do histograma  $n$ .

$c_{n-1} - c_{n-2}$ : Amplitude do intervalo do histograma  $n - 2$ .

1,8: Aumento da amplitude do intervalo de um histograma para o histograma seguinte.

5.5 2º ATIVIDADE: POPULAÇÕES DAS CIDADES AGRUPADAS EM HISTOGRAMAS COM AMPLITUDES DIFERENTES

	A	B	C	D
1	Município	População		Intervalos
2	São Paulo - SP	11.253.503		10.000
3	Rio de Janeiro - RJ	6.320.446		20.000
4	Salvador - BA	2.675.656		38.000
5	Brasília - DF	2.570.160		70.400
6	Fortaleza - CE	2.452.185		128.720
7	Belo Horizonte - MG	2.375.151		233.696
8	Manaus - AM	1.802.014		422.653
9	Curitiba - PR	1.751.907		762.775
10	Recife - PE	1.537.704		1.374.995
11	Porto Alegre - RS	1.409.351		2.476.991
12	Belém - PA	1.393.399		4.460.584
13	Goiânia - GO	1.302.001		8.031.051
14	Guarulhos - SP	1.221.979		14.457.892
15	Campinas - SP	1.080.113		
16	São Luís - MA	1.014.837		
17	São Gonçalo - RJ	999.728		

Figura 65: Etapa 1 da 2º atividade.

(2) a (6): Repetir as etapas de (2) a (6) da 1º atividade, considerando que na etapa 4 as amplitudes dos intervalos são diferentes, e na etapa 5 são considerados todos os blocos a partir de 20.000 (Figura 66). O gráfico, em escala log-log, está na Figura 67.

	A	B	C	D	E
1	Bloco	Valor Médio (x)	Freqüência	Freq. Relativa	Densidade de freq. (y)
2	10.000		0		
3	20.000	15.000	1.401	0,45904325	0,0000459043
4	38.000	29.000	851	0,27883355	0,0000154908
5	70.400	54.200	381	0,12483617	0,0000038530
6	128.720	99.560	217	0,07110092	0,0000012192
7	233.696	181.208	93	0,03047182	0,0000002903
8	422.653	328.174	63	0,02064220	0,0000001092
9	762.775	592.714	23	0,00753604	0,0000000222
10	1.374.995	1.068.885	12	0,00393185	0,0000000064
11	2.476.991	1.925.993	7	0,00229358	0,0000000021
12	4.460.584	3.468.788	2	0,00065531	0,0000000003
13	8.031.051	6.245.818	1	0,00032765	0,0000000001
14	14.457.892	11.244.472	1	0,00032765	0,0000000001

Figura 66: Etapas de (2) a (6) da 2º atividade.

O professor deve comentar com os alunos que o conjunto dos pontos obtidos sugere uma reta e, neste caso, é conveniente modelar os dados com uma Lei de Potência. Conforme visto na Seção 4.1, pode-se considerar:

$$\log p(x) = -\alpha \cdot \log x + b.$$

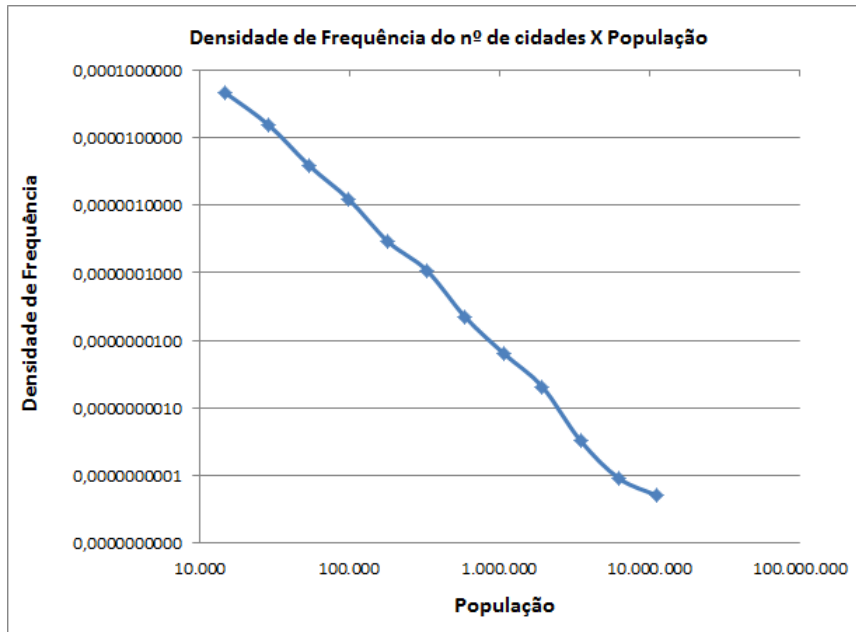


Figura 67: Gráfico da densidade de frequência em escala log-log. Foi utilizada a técnica do binning logarítmico.

Do que decorre  $-\alpha$ , tal que:

$$p(x) = k \cdot x^{-\alpha} \quad \text{e} \quad k = 10^b.$$

(7) Para descobrir o expoente  $\alpha$ , vamos acrescentar duas colunas: uma com os logaritmos dos valores de  $x$  na base 10 (poderia ser outra base), e outra com os logaritmos dos valores de  $y$  (Figura 68).

	A	B	C	D	E	F	G
1	Bloco	Valor Médio (x)	Freqüência	Freq. Relativa	Densidade de freq. (y)	Log (x)	Log (y)
2	10.000		0				
3	20.000	15.000	1.401	0,45904325	0,0000459043	4,1760912591	-4,3381463940
4	38.000	29.000	851	0,27883355	0,0000154908	4,4623979979	-4,8099274743
5	70.400	54.200	381	0,12483617	0,0000038530	4,7339992865	-5,4142045638
6	128.720	99.560	217	0,07110092	0,0000012192	4,9980848879	-5,9139423107
7	233.696	181.208	93	0,03047182	0,0000002903	5,2581773671	-6,5371916011
8	422.653	328.174	63	0,02064220	0,0000001092	5,5161046999	-6,9616065053
9	762.775	592.714	23	0,00753604	0,0000000222	5,7728451268	-7,6544917239
10	1.374.995	1.068.885	12	0,00393185	0,0000000064	6,0289310053	-8,1923108190
11	2.476.991	1.925.993	7	0,00229358	0,0000000021	6,2846547271	-8,6816665301
12	4.460.584	3.468.788	2	0,00065531	0,0000000003	6,5401777057	-9,4810070795
13	8.031.051	6.245.818	1	0,00032765	0,0000000001	6,7955893004	-10,0373095803
14	14.457.892	11.244.472	1	0,00032765	0,0000000001	7,0509390582	-10,2925820854

Figura 68: Etapa 7 da 2ª atividade.



Agora, será criado um gráfico de dispersão com os valores das colunas  $\log x$  e  $\log y$ . Em seguida, conforme visto no capítulo 3, adicionaremos a reta de regressão linear que melhor se ajusta aos pontos. Deve-se clicar com o botão direito do mouse sobre um dos pontos do gráfico, selecionar “Adicionar Linha de Tendência” e, na janela aberta, selecionar a opção “Linear”, selecionar também “Exibir Equação no Gráfico” e clicar em “Fechar” (Figura 69).

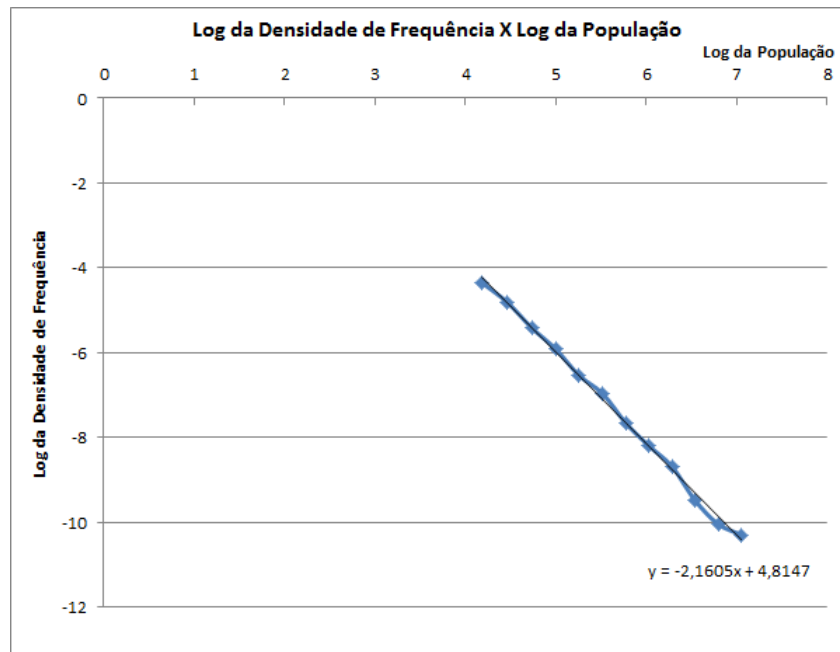


Figura 69: Gráfico da 2ª atividade para obtenção do expoente da lei de potência.

O expoente  $-\alpha$  da lei de potência é o coeficiente de  $x$ :  $-\alpha = -2,16$ .

Para obter o coeficiente “ $c$ ” da lei de potência, é necessário aplicar a equação (4.3):

$$c = (\alpha - 1) \cdot x_{min}^{\alpha-1}.$$

Notando que  $x_{min} = 10.000$ , tem-se:

$$\begin{aligned} c &= (2,16 - 1) \cdot 10000^{2,16-1} \\ \Rightarrow c &= 1,16 \cdot 10000^{1,16} \\ \Rightarrow c &\approx 50636. \end{aligned}$$

Logo, a lei de potência é dada por:  $p(x) = 50636 \cdot x^{-2,16}$ .

Embora o binning logarítmico forneça uma alternativa interessante para os problemas apontados, ainda assim, há alguns inconvenientes em sua aplicação. Um deles é que, por vezes, ainda permanece algum ruído na cauda. Outro inconveniente, bem mais relevante, é que ao agruparmos os dados em histogramas as informações individuais acabam perdidas. Consideramos os dados de um mesmo histograma como se fossem iguais. A solução mais interessante para estes problemas é a utilização das frequências acumuladas. Na próxima atividade aplicaremos este procedimento.

5.6 3<sup>o</sup> ATIVIDADE: DISTRIBUIÇÃO DAS POPULAÇÕES DAS CIDADES POR MEIO DA FREQUÊNCIA ACUMULADA.

Uma discussão sobre a utilização de frequências acumuladas, incluindo considerações teóricas, foi realizada na Seção 4.2.

Vamos às etapas da atividade:

(1) Na planilha dos dados das populações por cidade (veja Seção 5.3), ordenar os dados em ordem decrescente em relação às populações. A coluna “População” será a variável X (Figura 70).

	A	B	
1	<b>Município</b>	<b>População (x)</b>	
2	São Paulo - SP	11.253.503	
3	Rio de Janeiro - RJ	6.320.446	
4	Salvador - BA	2.675.656	
5	Brasília - DF	2.570.160	
6	Fortaleza - CE	2.452.185	
7	Belo Horizonte - MG	2.375.151	
8	Manaus - AM	1.802.014	
9	Curitiba - PR	1.751.907	
10	Recife - PE	1.537.704	
11	Porto Alegre - RS	1.409.351	

Figura 70: Etapa 1 da 3<sup>o</sup> atividade.

(2) Criar a coluna “Ranking”. Atribuir 1 para a maior população, 2 para a segunda maior até 3.052 que corresponde a cidade com menor população das que foram consideradas (10.004 habitantes) (Figura 71).

5.6 3º ATIVIDADE: DISTRIBUIÇÃO DAS POPULAÇÕES DAS CIDADES POR MEIO DA FREQUÊNCIA ACUMULADA

	A	B	C	D
1	<b>Município</b>	<b>População</b>	<b>Ranking</b>	
2	São Paulo - SP	11.253.503	1	
3	Rio de Janeiro - RJ	6.320.446	2	
4	Salvador - BA	2.675.656	3	
5	Brasília - DF	2.570.160	4	
6	Fortaleza - CE	2.452.185	5	
7	Belo Horizonte - MG	2.375.151	6	
8	Manaus - AM	1.802.014	7	
9	Curitiba - PR	1.751.907	8	
10	Recife - PE	1.537.704	9	
11	Porto Alegre - RS	1.409.351	10	
12	Belém - PA	1.393.399	11	
13	Goiânia - GO	1.302.001	12	
14	Guarulhos - SP	1.221.979	13	
15	Campinas - SP	1.080.113	14	

Figura 71: Etapa 2 da 3º atividade.

Cada posição em “Ranking” pode ser compreendida como a frequência absoluta acumulada das cidades com população maior ou igual a  $x_i$ . Por exemplo, só há uma cidade com população maior ou igual 11.253.503 (São Paulo - SP); só há duas cidades com população maior ou igual 6.320.446 e assim por diante.

(3) Criar a coluna “Freq. Rel. Acumulada” (Frequência Relativa Acumulada); esta coluna conterá os dados da variável  $P(X)$ . O valor de cada célula desta coluna será obtido dividindo a posição correspondente no ranking pelo total de cidades: 3.052 (Figura 72).

	A	B	C	D
1	<b>Município</b>	<b>População</b>	<b>Ranking</b>	<b>Freq. Rel. Acumulada</b>
2	São Paulo - SP	11.253.503	1	0,0003277
3	Rio de Janeiro - RJ	6.320.446	2	0,0006553
4	Salvador - BA	2.675.656	3	0,0009830
5	Brasília - DF	2.570.160	4	0,0013106
6	Fortaleza - CE	2.452.185	5	0,0016383
7	Belo Horizonte - MG	2.375.151	6	0,0019659
8	Manaus - AM	1.802.014	7	0,0022936
9	Curitiba - PR	1.751.907	8	0,0026212
10	Recife - PE	1.537.704	9	0,0029489
11	Porto Alegre - RS	1.409.351	10	0,0032765
12	Belém - PA	1.393.399	11	0,0036042
13	Goiânia - GO	1.302.001	12	0,0039318
14	Guarulhos - SP	1.221.979	13	0,0042595
15	Campinas - SP	1.080.113	14	0,0045872

Figura 72: Etapa 3 da 3º atividade.

(4) Criar um gráfico de dispersão com as variáveis/colunas  $X$  e  $P(X)$ . Alterar as escalas dos eixos para logarítmicas (conforme realizado nas atividades anteriores) (Figura 73).

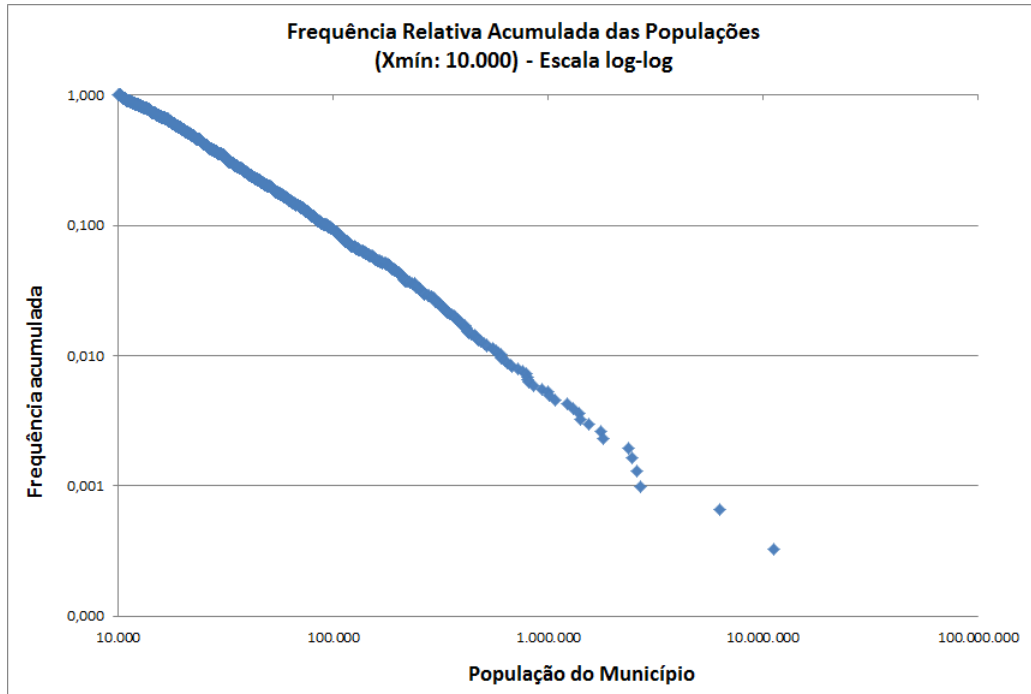


Figura 73: Etapa 4 da 3ª atividade.

(5) Notando a adaptabilidade dos pontos a uma reta, acrescentaremos ao gráfico a reta obtida por meio de uma regressão linear simples (Capítulo 3), utilizando, para isto, a função “Adicionar Linha de Tendência” conforme feito na etapa 7 da segunda atividade. Obtém-se, assim,  $-\delta = -1,097$  (o expoente está representado por  $\delta$  para diferenciar do  $\alpha$  que é o expoente da lei de potência para a distribuição não acumulada). Para obter o expoente  $\alpha$  da lei de potência para a distribuição não acumulada, conforme a equação (4.5), aplicamos:

$$\alpha - 1 = \delta \Rightarrow \alpha - 1 = 1,097 \Rightarrow \alpha = 2,097.$$

Para obter o valor da constante  $c$  da lei de potência da distribuição não acumulada, conforme a equação (4.3), fazemos:

#### 5.7 4º ATIVIDADE: FREQUÊNCIA DAS PALAVRAS UTILIZADAS NO LIVRO MOBY DICK DE HERMAN MELVILLE.

$$\begin{aligned}c &= (\alpha - 1) \cdot x_{min}^{\alpha-1} \\ \Rightarrow c &= (2,097 - 1) \cdot 10000^{2,097-1} \\ \Rightarrow c &= (1,097) \cdot 10000^{1,097} \\ \Rightarrow c &\approx 26.804.\end{aligned}$$

Obtemos, assim, a lei de potência:  $p(x) = 26804 \cdot x^{-2,097}$ .

#### 5.7 4º ATIVIDADE: FREQUÊNCIA DAS PALAVRAS UTILIZADAS NO LIVRO MOBY DICK DE HERMAN MELVILLE.

Leis de potência também costumam aparecer quando são analisadas as ocorrências de palavras em um livro. De modo geral, há algumas poucas palavras que correspondem a uma grande quantidade do texto, e muitas palavras que aparecem pouco. Em Newman (2006, p.5), tem-se o exemplo do livro Moby Dick de Herman Melville e um link: <http://tuvalu.santafe.edu/~aronc/powerlaws/data.htm> com os dados do livro para a análise.

Vamos às etapas desta atividade:

(1) Acessar o link: <http://tuvalu.santafe.edu/~aronc/powerlaws/data.htm> e na página aberta clicar em “Download the data” em “Empirical Data Sets - 1. The frequency of occurrence of unique words in the novel Moby Dick by Herman Melville”.

Os dados que aparecem na tela são as frequências de todas as palavras que aparecem no livro. Estes dados devem ser copiados e colados em uma tabela do excel. A coluna formada por estes valores será chamada de “Dados” (Figura 74).

Note que o primeiro valor, 14.086, quer dizer que há uma palavra que aparece esta quantidade de vezes no texto e o mesmo ocorrendo para os demais valores.

(2) Selecionar os valores da coluna “Dados”, na parte superior da tela clicar em “Inserir” e em “Tabela Dinâmica” e novamente em “Tabela Dinâmica”. Na janela aberta, clicar em “Ok”. Na nova planilha criada, na seção aberta na parte direita da tela, arrastar “Dados” para o campo “Rótulos de Linha”. Arrastar novamente “Dados”, agora para o campo “Valores”. No campo “Valores”, em “Soma de Dados”, clicar na setinha que está apontando para baixo e selecionar “Configurações do Campo de Valor” e, na

	A	E
1	<b>Dados</b>	
2	14086	
3	6414	
4	6260	
5	4573	
6	4484	
7	4040	
8	2917	
9	2483	
10	2374	
11	1942	
12	1792	

Figura 74: Frequências de ocorrências das palavras no livro Moby Dick, de Herman Melville.

janela aberta, selecionar “Contagem”, clicar em “Ok”. Aparecerá a planilha da Figura 75.

	A	B
1		
2		
3	<b>Rótulos de Linha</b>	<b>Contagem de Dados</b>
4	1	9161
5	2	3085
6	3	1629
7	4	926
8	5	627
9	6	469
10	7	361
11	8	300
12	9	232
13	10	179
14	11	165
15	12	146
16	13	115
17	14	90

Figura 75: Etapa 2 da 4ª atividade.

(3) Selecionar apenas os dados das duas colunas e colocá-los em uma nova planilha, com as colunas com os títulos conforme a planilha da Figura 76.

Note, que na primeira coluna, se encontram os números de ocorrências e, na segunda, o número de palavras que ocorrem na quantidade de vezes indicada na primeira coluna. Por exemplo, há 9.161 palavras que ocorrem apenas uma vez, 3.085 palavras que ocorrem duas vezes, e assim por diante.

5.7 4<sup>o</sup> ATIVIDADE: FREQUÊNCIA DAS PALAVRAS UTILIZADAS NO LIVRO MOBY DICK DE HERMAN MELVILLE.

	A	B	Barra
1	<b>Nº de Ocorrências (x)</b>	<b>Nº de Palavras</b>	
2	1	9161	
3	2	3085	
4	3	1629	
5	4	926	
6	5	627	
7	6	469	
8	7	361	
9	8	300	
10	9	232	
11	10	179	
12	11	165	
13	12	146	
14	13	115	

Figura 76: Etapa 3 da 4<sup>o</sup> atividade.

(4) Criar a coluna “Nº de Palavras com ocorrência  $\geq x$ ”, nela, cada célula  $c_i$  ( $i$  indica o número da linha) deve ser preenchida com a soma dos valores da coluna “Nº de Palavras” desde a célula  $B_i$ , até a última célula desta coluna (Figura 77).

	A	B	C
1	<b>Nº de Ocorrências (x)</b>	<b>Nº de Palavras</b>	<b>Nº de Palavras com ocorrência <math>\geq x</math></b>
2	1	9161	18855
3	2	3085	9694
4	3	1629	6609
5	4	926	4980
6	5	627	4054
7	6	469	3427
8	7	361	2958
9	8	300	2597
10	9	232	2297
11	10	179	2065
12	11	165	1886
13	12	146	1721
14	13	115	1575
15	14	90	1460
16	15	100	1370

Figura 77: Etapa 4 da 4<sup>o</sup> atividade.

(5) Criar duas colunas:  $\log x$  e  $\log y$ . A coluna  $\log x$  será preenchida com os logaritmos (utilizamos base 10) dos valores da coluna “Nº de Ocorrências”. A coluna  $\log y$  será preenchida com logaritmos (utilizamos base 10) dos valores da coluna “Nº de Palavras com ocorrência  $\geq x$ ” (Figura 78) .

ATIVIDADES COM LEIS DE POTÊNCIA

	A	B	C	D	E
1	<b>Nº de Ocorrências (x)</b>	<b>Nº de Palavras</b>	<b>Nº de Palavras com ocorrência <math>\geq x</math></b>	<b>log x</b>	<b>log y</b>
2	1	9.161	18.855	0,0000	4,2754
3	2	3.085	9.694	0,3010	3,9865
4	3	1.629	6.609	0,4771	3,8201
5	4	926	4.980	0,6021	3,6972
6	5	627	4.054	0,6990	3,6079
7	6	469	3.427	0,7782	3,5349
8	7	361	2.958	0,8451	3,4710
9	8	300	2.597	0,9031	3,4145
10	9	232	2.297	0,9542	3,3612
11	10	179	2.065	1,0000	3,3149
12	11	165	1.886	1,0414	3,2755
13	12	146	1.721	1,0792	3,2358
14	13	115	1.575	1,1139	3,1973
15	14	90	1.460	1,1461	3,1644
16	15	100	1.370	1,1761	3,1367
17	16	74	1.270	1,2041	3,1038
18	17	56	1.196	1,2304	3,0777
19	18	70	1.140	1,2553	3,0569

Figura 78: Etapa 5 da 4ª atividade.

(6) Selecionar os dados das colunas  $\log x$  e  $\log y$  e criar um gráfico de dispersão (conforme executado nas atividades anteriores)(Figura 79).

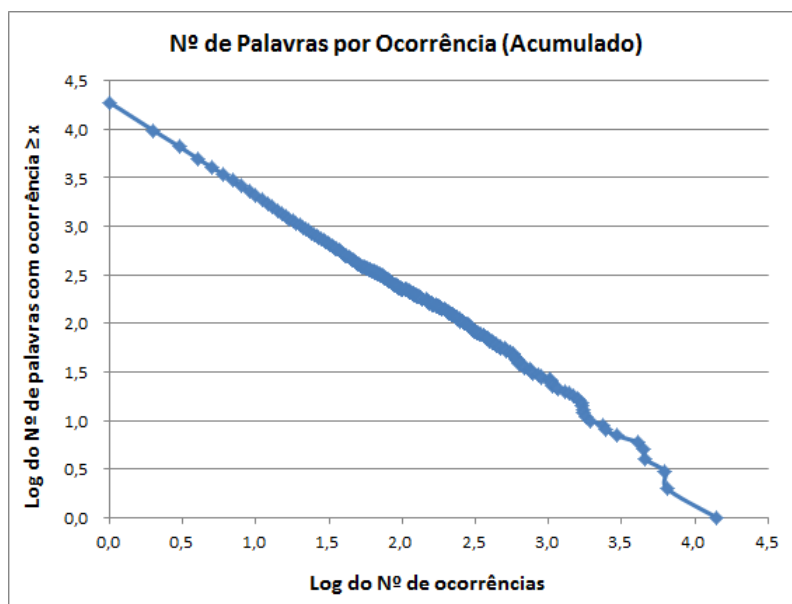


Figura 79: Log do número de palavras com ocorrência  $\geq x$  pelo log do número de ocorrências.

(7) Adicionar uma “Linha de Tendência Linear” ao gráfico (Figura 80). Note como a reta se ajusta bem aos pontos do gráfico, confirmando se tratar de uma lei de potência.



5.7 4<sup>o</sup> ATIVIDADE: FREQUÊNCIA DAS PALAVRAS UTILIZADAS NO LIVRO MOBY DICK DE HERMAN MELVILLE.

Por esta reta ajustada e, em particular, pelo seu coeficiente angular, obteremos o valor da constante  $\alpha$  da lei de potência.

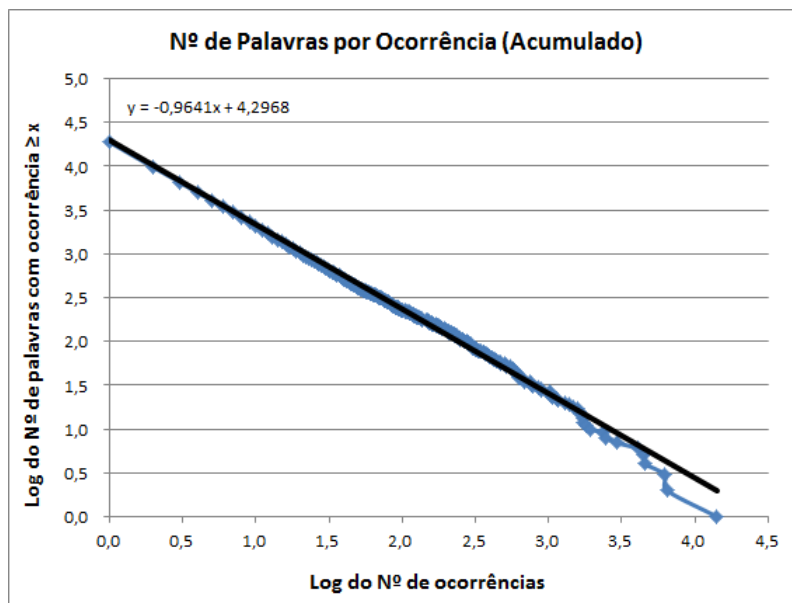


Figura 80: Log do número de palavras com ocorrência  $\geq x$  pelo log do número de ocorrências. Uma reta ajustada aos pontos foi acrescentada ao gráfico.

(8) Como se trata de uma distribuição de frequências acumuladas, conforme vimos na Seção 4.2, o expoente  $\alpha$  pode ser assim obtido:

$$\alpha - 1 = 0,964 \Rightarrow \alpha = 1,964.$$

Para obter a constante  $c$  da lei de potência, utilizamos a equação (4.3), com  $x_{min} = 1$  (pois a menor ocorrência considerada foi 1):

$$\begin{aligned} C &= (\alpha - 1) \cdot x_{min}^{\alpha-1} \\ \Rightarrow C &= (1,964 - 1) \cdot 1^{1,964-1} \\ \Rightarrow C &= (0,964) \cdot 1^{0,964} \\ \Rightarrow C &= 0,964. \end{aligned}$$

Logo, podemos escrever a lei de potência para esta distribuição:

$$p(x) = 0,964 \cdot x^{-1,964}.$$

Encerram-se aqui as atividades propostas. Outras atividades podem ser pensadas pelo professor e, para isto, podem ser utilizados os exemplos da Seção 4.3.

---

## CONSIDERAÇÕES FINAIS

---

Como dissemos na introdução, a motivação inicial para esta dissertação foram as Leis de Potência; mais especificamente, a abordagem destas leis para alunos de Ensino Médio. O artigo que nos apresentou o tema e, ao mesmo tempo, uma abordagem interessante do mesmo, foi Newman (2006), [3]. No Capítulo 4, esta abordagem foi fielmente seguida.

Também, como dito na introdução, esta motivação inicial, Leis de Potência, suscitou uma série de outras considerações que foram necessárias para que o assunto principal fosse compreendido. Essas considerações acabaram ganhando “vida própria” e, por vezes, tornaram-se capítulos.

Alguns desafios enfrentados foram: o cálculo da área sob a curva de certas funções não retilíneas, sem o uso de integrais; encontrar uma reta que se adapta, o melhor possível, a um conjunto de pontos. Estes temas não são abordados no Ensino Médio e precisaram ser desenvolvidos sem as ferramentas que são tradicionalmente utilizadas nestes casos: integrais e derivadas. Cremos que conseguimos lograr o nosso intento.

A parte prática também não foi negligenciada. Muitos exemplos foram utilizados e, naturalmente, são sugeridos para exposição em aula. No Capítulo 5, foram explicadas, passo a passo, diversas atividades, em que uma série de conceitos tratados em outros capítulos foram ilustrados. Essas explicações servem como roteiro para aplicação em outros casos, como os exemplos fornecidos no Capítulo 4.

Creemos, também, que esta dissertação conseguiu mostrar como assuntos, aparentemente não relacionáveis, se relacionam; pois na vida real, ao abordar algo, diversos conhecimentos e técnicas se complementam para que tenhamos compreensão do objeto de estudo. Estas conexões entre os assuntos facilitam, inclusive, a assimilação dos mesmos.

Ainda, com relação à parte prática, o software Excel foi largamente utilizado nas atividades propostas e na exposição presente no Capítulo 3. Acreditamos ser importante, para o aluno, ter nos recursos tecnológicos um aliado para facilitar trabalhos que, sem

## CONSIDERAÇÕES FINAIS

estes recursos, seriam demasiadamente trabalhosos. Isto dá ao estudante uma outra dimensão para a tecnologia, deixando de ser vista como mero entretenimento.

Por fim, cabe ao professor do Ensino Médio, leitor desta dissertação, julgar, do que foi exposto, o que deverá ser apresentado ao seus alunos e quais serão as adaptações necessárias. A transposição didática deve ser sempre levada em consideração. Uma aula expositiva nunca corresponderá, exatamente, à exposição escrita.

Esperamos ter contribuído, ainda que modestamente, para o aperfeiçoamento do ensino de matemática em nosso país.

---

## APÊNDICE

---

Aqui serão demonstradas algumas fórmulas do Capítulo 2.

### A.1 ÁREA SOB A CURVA DE $f(x) = x^m$ , $x \geq 0$ , COM $m \geq 0$

A demonstração a seguir segue a mesma estrutura dos casos em que  $m = 2$  e  $m = 3$  já tratados no Capítulo 2.

Faremos  $s = 0$ . Para simplificar a notação, denominamos a área sob a curva no intervalo  $[0, t]$  como  $A_m[0, t]$ , ou, de forma mais enxuta,  $A_m(t)$ . (Figura 81)

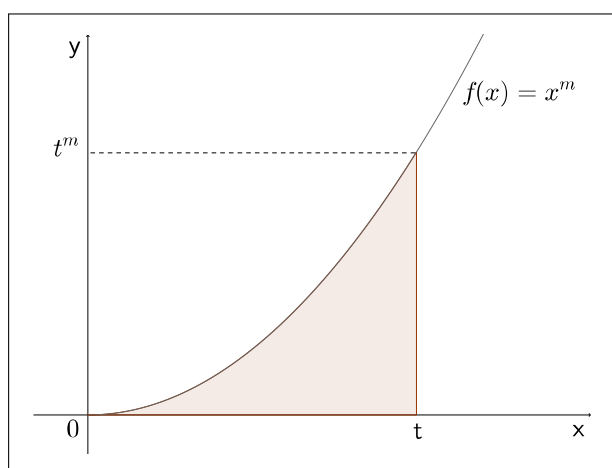


Figura 81: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ .

A área  $A_m(t)$  será aproximada pela utilização de  $n$  retângulos (Figura 82). Cada retângulo tem como medida da base:  $\frac{t-0}{n} = \frac{t}{n}$ . E as alturas dos  $n$  retângulos em

ordem crescente, do primeiro ao último, correspondem a:  $\left(\frac{t}{n}\right)^m$ ,  $\left(\frac{2t}{n}\right)^m$ ,  $\left(\frac{3t}{n}\right)^m$ , ...,  $\left(\frac{nt}{n}\right)^m$ . Note que  $f\left(\frac{t}{n}\right) = \left(\frac{t}{n}\right)^m$ ,  $f\left(\frac{2t}{n}\right) = \left(\frac{2t}{n}\right)^m$ , ...,  $f\left(\frac{nt}{n}\right) = \left(\frac{nt}{n}\right)^m$  (Figura 82).

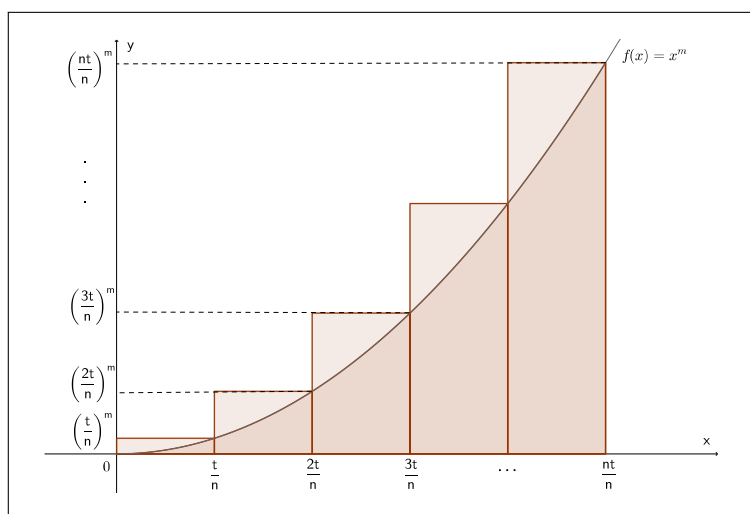


Figura 82: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por  $n$  retângulos.

Como pode ser percebido, a soma das áreas destes retângulos é superior à área sob a curva. Pode-se notar, também, que à medida que aumentamos o número de retângulos (valor de  $n$ ), a soma das áreas se aproxima da área procurada (Figuras de 83 a 88). Dizemos que para  $n$  muito grande ( $n$  tendendo ao infinito), a área sob os retângulos é equivalente à área sob a curva.

Denotaremos a soma das áreas dos  $n$  retângulos por  $R_{m,n}(t)$ , em que:  $m$  é o expoente de  $x$ ,  $n$  é o número de retângulos (cada um com base de medida  $\frac{1}{n}$ ), e  $t$  é a limitação à direita do intervalo.

$$R_{m,n}(t) = \frac{t}{n} \cdot \left(\frac{t}{n}\right)^m + \frac{t}{n} \cdot \left(\frac{2t}{n}\right)^m + \frac{t}{n} \cdot \left(\frac{3t}{n}\right)^m + \dots + \frac{t}{n} \cdot \left(\frac{nt}{n}\right)^m.$$

Colocando  $\left(\frac{t}{n}\right)^{m+1}$  em evidência:

A.1 ÁREA SOB A CURVA DE  $f(x) = x^m$ ,  $x \geq 0$ , COM  $m \geq 0$

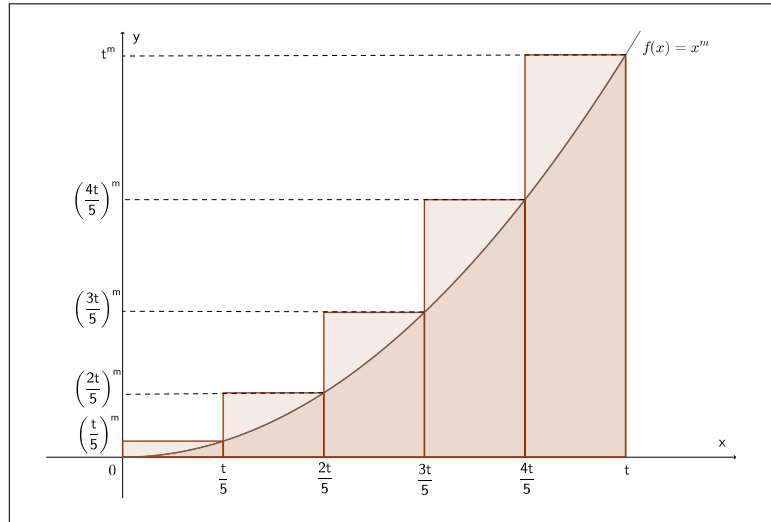


Figura 83: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 5 retângulos.

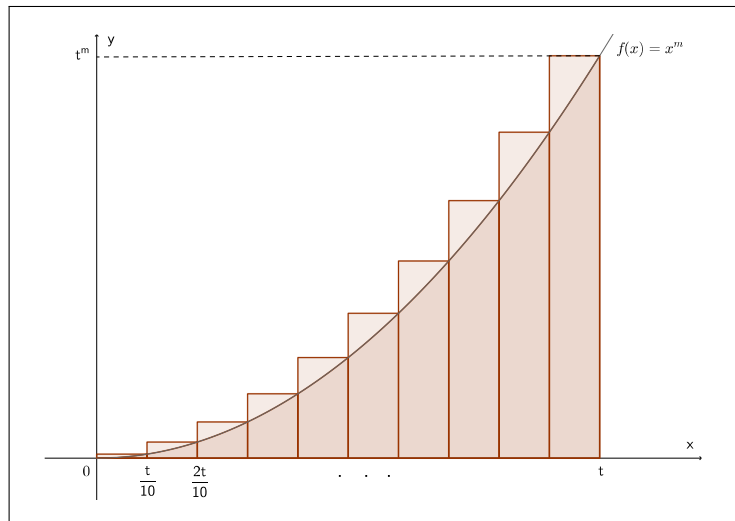


Figura 84: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 10 retângulos.

$$R_{m,n}(t) = \left(\frac{t}{n}\right)^{m+1} \cdot (1^m + 2^m + 3^m + \dots + n^m).$$

Considerando  $S_m(n) = 1^m + 2^m + 3^m + \dots + n^m$ , tem-se:

$$R_{m,n}(t) = \left(\frac{t}{n}\right)^{m+1} \cdot (1^m + 2^m + 3^m + \dots + n^m) = \frac{S_m(n) \cdot t^{m+1}}{n^{m+1}}.$$

Para  $n$  muito grande,  $R_{m,n}(t) = A_m(t)$ . E, em particular, para  $t = 1$  e  $n$  muito grande:

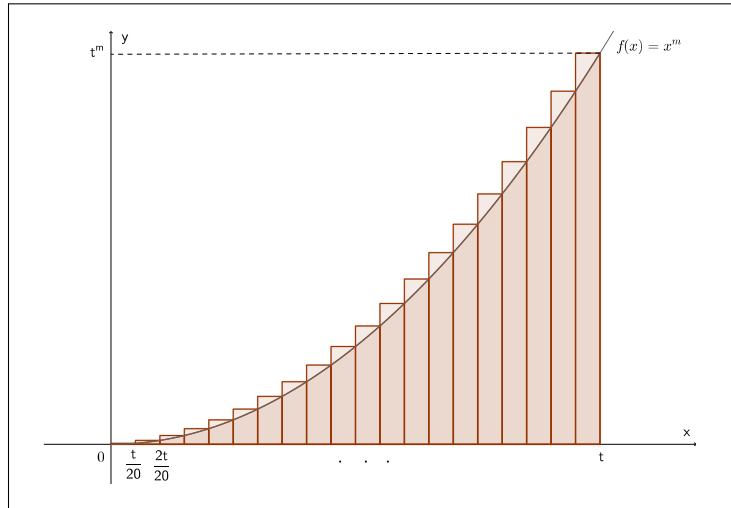


Figura 85: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 20 retângulos.

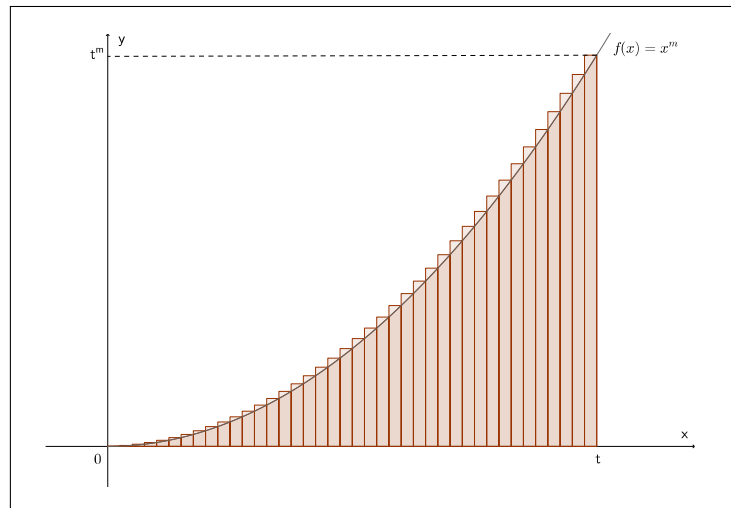


Figura 86: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 40 retângulos.

$$\frac{S_m(n) \cdot 1^{m+1}}{n^{m+1}} = R_{m,n}(1) \approx A_m(1). \quad (\text{A.1})$$

Será importante notar que para  $0 < k \leq m$  :

$$\frac{S_{m-k}(n)}{n^{m+1}} = \frac{1}{n^k} \cdot \frac{S_{m-k}(n)}{n^{(m+1)-k}} = \frac{1}{n^k} \cdot \frac{S_{m-k}(n)}{n^{(m-k)+1}}. \quad (\text{A.2})$$



A.1 ÁREA SOB A CURVA DE  $f(x) = x^m$ ,  $x \geq 0$ , COM  $m \geq 0$

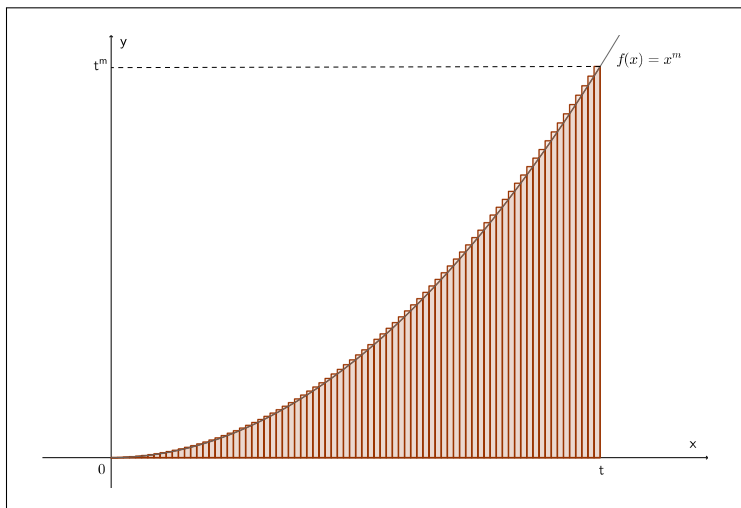


Figura 87: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 80 retângulos.

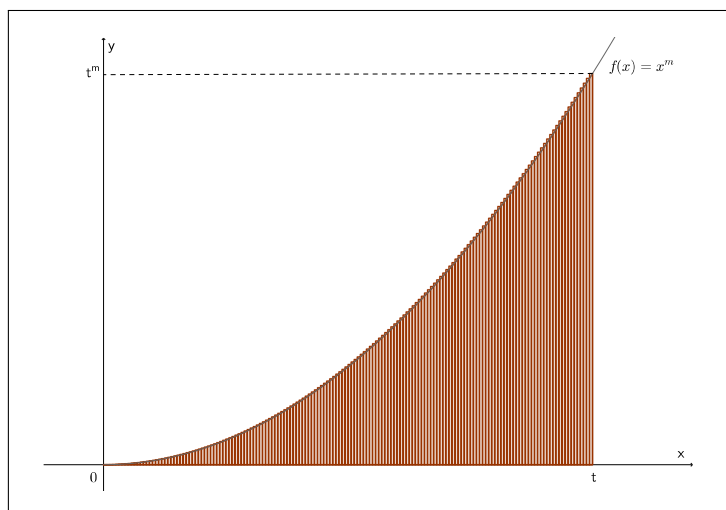


Figura 88: Área da região sob a curva de  $f(x) = x^m$ ,  $m \geq 0$ , em  $[0, t]$ , cercada por 160 retângulos.

Quanto mais aumentarmos o valor de  $n$  mais próximo de zero estará o valor de  $\frac{1}{n^k}$ , sendo assim, para  $n$  muito grande, considera-se que o valor de  $\frac{1}{n^k}$  é nulo. Para  $n$  muito grande,  $\frac{S_{m-k}(n)}{n^{(m-k)+1}}$  equivale a  $A_{m-k}(1)$ . Como o produto de zero por um número é igual a zero, temos que, para  $n$  muito grande, o produto acima (A.2) é nulo.

Temos pelo binômio de Newton:

$$(a + 1)^{m+1} = a^{m+1} + \binom{m+1}{1} \cdot a^m + \binom{m+1}{2} \cdot a^{m-1} + \dots + \binom{m+1}{m} \cdot a^1 + \binom{m+1}{m+1} \cdot a^0.$$

Variando  $a$  entre 0 e  $n$  temos:

$$\begin{aligned}
 (0+1)^{m+1} &= 0^{m+1} + \binom{m+1}{1} \cdot 0^m + \binom{m+1}{2} \cdot 0^{m-1} + \dots + \binom{m+1}{m} \cdot 0^1 + 1 \\
 (1+1)^{m+1} &= 1^{m+1} + \binom{m+1}{1} \cdot 1^m + \binom{m+1}{2} \cdot 1^{m-1} + \dots + \binom{m+1}{m} \cdot 1^1 + 1 \\
 (2+1)^{m+1} &= 2^{m+1} + \binom{m+1}{1} \cdot 2^m + \binom{m+1}{2} \cdot 2^{m-1} + \dots + \binom{m+1}{m} \cdot 2^1 + 1 \\
 (3+1)^{m+1} &= 3^{m+1} + \binom{m+1}{1} \cdot 3^m + \binom{m+1}{2} \cdot 3^{m-1} + \dots + \binom{m+1}{m} \cdot 3^1 + 1 \\
 &\vdots \\
 &\vdots \\
 (n+1)^{m+1} &= n^{m+1} + \binom{m+1}{1} \cdot n^m + \binom{m+1}{2} \cdot n^{m-1} + \dots + \binom{m+1}{m} \cdot n^1 + 1.
 \end{aligned}$$

Somando os membros correspondentes das equações de todas as linhas acima, obtemos<sup>1</sup>

$$(n+1)^{m+1} = \binom{m+1}{1} \cdot S_m(n) + \binom{m+1}{2} \cdot S_{m-1}(n) + \dots + \binom{m+1}{m} \cdot S_1(n) + n + 1.$$

Dividindo ambos os membros por  $n^{m+1}$ :

$$\frac{(n+1)^{m+1}}{n^{m+1}} = \binom{m+1}{1} \cdot \frac{S_m(n)}{n^{m+1}} + \binom{m+1}{2} \cdot \frac{S_{m-1}(n)}{n^{m+1}} + \dots + \binom{m+1}{m} \cdot \frac{S_1(n)}{n^{m+1}} + \frac{n}{n^{m+1}} + \frac{1}{n^{m+1}}.$$

Seguindo-se:

$$\left(1 + \frac{1}{n}\right)^{m+1} = \binom{m+1}{1} \cdot R_{m,n}(1) + \binom{m+1}{2} \cdot \frac{1}{n} \cdot \frac{S_{m-1}(n)}{n^m} + \dots + \binom{m+1}{m} \cdot \frac{1}{n^{m-1}} \cdot \frac{S_1(n)}{n^2} + \frac{1}{n^m} + \frac{1}{n^{m+1}}.$$

Para  $n$  muito grande, as parcelas:  $\binom{m+1}{2} \cdot \frac{1}{n} \cdot \frac{S_{m-1}(n)}{n^m}, \dots, \binom{m+1}{m} \cdot \frac{1}{n^{m-1}} \cdot \frac{S_1(n)}{n^2}, \frac{1}{n^m}, \frac{1}{n^{m+1}}$  anulam-se, conforme já justificado, e  $R_{m,n}(1)$  torna-se equivalente a  $A_m(1)$ , obtendo-se assim a equação:

$$\begin{aligned}
 1 &= \binom{m+1}{1} \cdot A_m(1) \\
 \implies 1 &= (m+1) \cdot A_m(1) \\
 \implies A_m(1) &= \frac{1}{m+1}.
 \end{aligned}$$

<sup>1</sup> Note que, ao somarmos as equações, as potências dos membros esquerdos, com exceção de  $(n+1)^{m+1}$ , são canceladas com as primeiras potências dos membros direitos das equações seguintes.

A.2 ÁREA SOB A CURVA DE  $f(x) = cx^{-m}$ , PARA  $m > 1$  E  $c > 0$

Multiplicando-se ambos os membros por  $t^{m+1}$ :

$$A_m(1) \cdot t^{m+1} = \frac{1}{m+1} \cdot t^{m+1}.$$

Da equação (A.1), tem-se:

$$\frac{S_m(n)}{n^{m+1}} \cdot t^{m+1} = \frac{t^{m+1}}{m+1}.$$

Donde segue:

$$A_m(t) = \frac{t^{m+1}}{m+1}.$$

A.2 ÁREA SOB A CURVA DE  $f(x) = cx^{-m}$ , PARA  $m > 1$  E  $c > 0$

A demonstração a seguir segue a mesma estrutura dos casos em que  $m = 2$  e  $m = 3$  já tratados no Capítulo 2.

Consideraremos a área sob a curva limitada ao intervalo  $[a, b]$ , com  $0 < a < b$  (Figura 89). Esta área será denotada por  $A_{-m}[a, b]$ . Note que o valor desta área é superior a área do retângulo de base  $b - a$  e altura  $f(b) = \frac{c}{b^m}$  e inferior a área do retângulo de base  $b - a$  e altura  $f(a) = \frac{c}{a^m}$  (Figura 90). As áreas dos retângulos nestas condições serão denotadas por  $r[a, b]$  e  $R[a, b]$ , respectivamente.

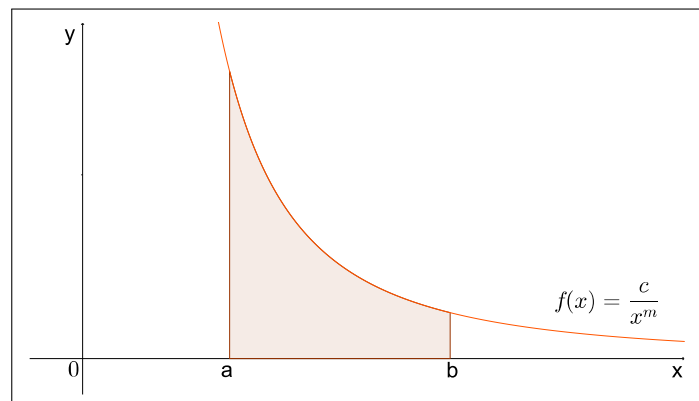


Figura 89: Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ .

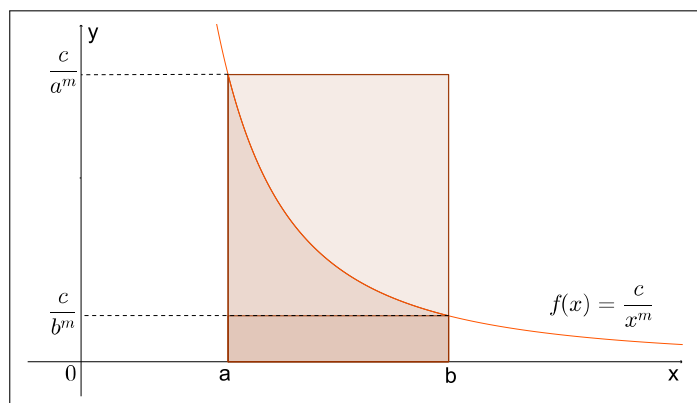


Figura 90: Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ , cercada por dois retângulos.

Tem-se:

$$r[a, b] = \frac{c}{b^m} \cdot (b - a),$$

$$R[a, b] = \frac{c}{a^m} \cdot (b - a).$$

A área  $A_{-m}[a, b]$  será cercada por  $n$  retângulos cuja soma das áreas é inferior a área sob a curva (Figura 91), esta soma será representada por  $r_n[a, b]$ , e por  $n$  retângulos cuja soma das áreas é superior a área sob a curva (Figura 92), esta soma será representada por  $R_n[a, b]$ . Tanto para  $r_n[a, b]$ , quanto para  $R_n[a, b]$ , para  $n$  muito grande, temos que ambas as somas convergem e se equivalem a  $A_{-m}[a, b]$ .

Consideraremos a função  $g(t) = \frac{c}{m-1} \left( \frac{1}{s^{m-1}} - \frac{1}{t^{m-1}} \right)$ , conforme conjecturamos em (2.12). Queremos mostrar que esta função corresponde a área  $A_{-m}[a, b]$ , fazendo  $s = a$  e  $t = b$ . Mostraremos que  $g(t)$  é sempre superior, ou igual, a  $r_n[a, b]$  e sempre inferior, ou igual, a  $R_n[a, b]$  e, como para  $n$  muito grande  $r_n[a, b]$  e  $R_n[a, b]$  convergem à  $A_{-m}[a, b]$ , isto implica em  $g(t) = A_{-m}[a, b]$ .

Para  $g(t)$ , tem-se:

A.2 ÁREA SOB A CURVA DE  $f(x) = cx^{-m}$ , PARA  $m > 1$  E  $c > 0$

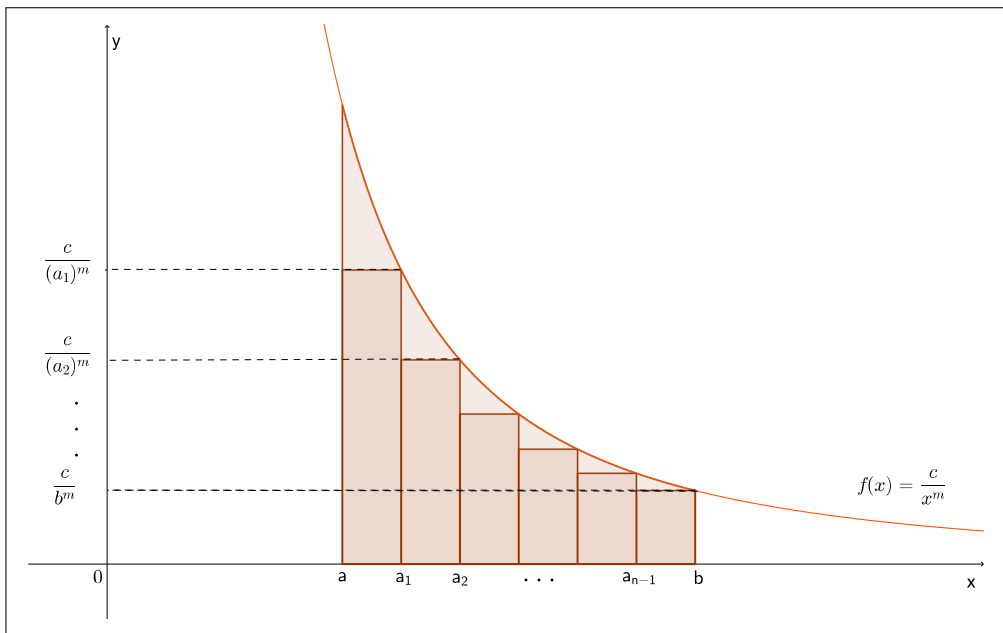


Figura 91: Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ , cercada por retângulos inferiores.

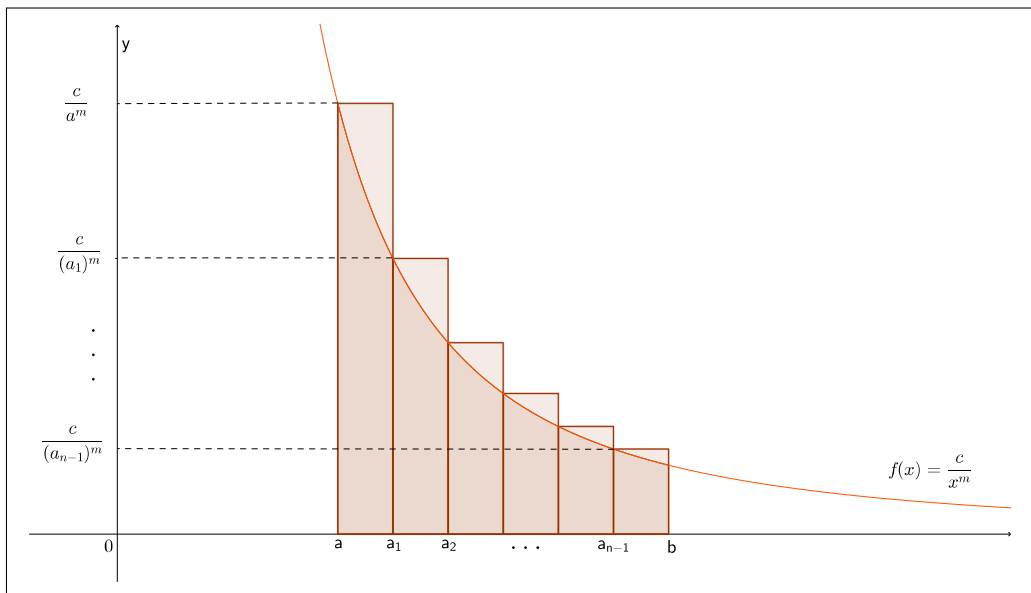


Figura 92: Área da região sob a curva de  $f(x) = cx^{-m}$ ,  $m > 1$ , em  $[a, b]$ , cercada por retângulos superiores.

$$\begin{aligned}
 g(b) - g(a) &= \frac{c}{m-1} \cdot \left( \frac{1}{s^{m-1}} - \frac{1}{b^{m-1}} \right) - \frac{c}{m-1} \cdot \left( \frac{1}{s^{m-1}} - \frac{1}{a^{m-1}} \right) \\
 \Rightarrow g(b) - g(a) &= \frac{c}{m-1} \cdot \left( \frac{1}{a^{m-1}} - \frac{1}{b^{m-1}} \right) \\
 \Rightarrow g(b) - g(a) &= \frac{c}{m-1} \cdot \left( \frac{1}{a} - \frac{1}{b} \right) \cdot \left( \frac{1}{a^{m-2}} + \frac{1}{a^{m-3}} \cdot \frac{1}{b} + \frac{1}{a^{m-4}} \cdot \frac{1}{b^2} + \dots + \frac{1}{b^{m-2}} \right) \\
 \Rightarrow g(b) - g(a) &= \frac{b-a}{m-1} \cdot \frac{c}{ab} \cdot \left( \frac{1}{a^{m-2}} + \frac{1}{a^{m-3}} \cdot \frac{1}{b} + \frac{1}{a^{m-4}} \cdot \frac{1}{b^2} + \dots + \frac{1}{b^{m-2}} \right).
 \end{aligned}$$

Como  $c > 0$ ,  $b > a$  e, conseqüentemente,  $\frac{1}{b} < \frac{1}{a}$ , tem-se:

$$\begin{aligned}
 \frac{b-a}{m-1} \cdot \frac{c}{b^2} \cdot \left( \frac{1}{b^{m-2}} + \frac{1}{b^{m-3}} \cdot \frac{1}{b} + \dots + \frac{1}{b^{m-2}} \right) &\leq g(b) - g(a) \leq \\
 \frac{b-a}{m-1} \cdot \frac{c}{a^2} \cdot \left( \frac{1}{a^{m-2}} + \frac{1}{a^{m-3}} \cdot \frac{1}{a} + \dots + \frac{1}{a^{m-2}} \right). &
 \end{aligned}$$

Donde segue:

$$\begin{aligned}
 \frac{b-a}{m-1} \cdot \frac{c}{b^2} \cdot \left( \frac{1}{b^{m-2}} + \frac{1}{b^{m-2}} + \dots + \frac{1}{b^{m-2}} \right) &\leq g(b) - g(a) \leq \\
 \frac{b-a}{m-1} \cdot \frac{c}{a^2} \cdot \left( \frac{1}{a^{m-2}} + \frac{1}{a^{m-2}} + \dots + \frac{1}{a^{m-2}} \right). &
 \end{aligned}$$

Dentro do primeiro par de parênteses, no membro esquerdo da inequação, há  $m-1$  parcelas iguais a  $\frac{1}{b^{m-2}}$  e, dentro do par de parênteses do membro direito da inequação, há  $m-1$  parcelas iguais a  $\frac{1}{a^{m-2}}$ . Do que, se segue:

$$\begin{aligned}
 \frac{b-a}{m-1} \cdot \frac{c}{b^2} \cdot \frac{m-1}{b^{m-2}} &\leq g(b) - g(a) \leq \frac{b-a}{m-1} \cdot \frac{c}{a^2} \cdot \frac{m-1}{a^{m-2}} \\
 \Rightarrow \frac{c}{b^m} \cdot (b-a) &\leq g(b) - g(a) \leq \frac{c}{a^m} \cdot (b-a)
 \end{aligned}$$

$$\Rightarrow r[a, b] \leq g(b) - g(a) \leq R[a, b]. \tag{A.3}$$

Aproximaremos a área sob a curva em  $[a, b]$  por meio de  $n$  retângulos inferiores e  $n$  retângulos superiores com bases correspondentes aos intervalos  $[a, a_1]$ ,  $[a_1, a_2]$ ,

$[a_2, a_3], \dots, [a_{n-1}, b]$ , sendo as alturas dos inferiores:  $\left(\frac{c}{a_1}\right)^m, \left(\frac{c}{a_2}\right)^m, \left(\frac{c}{a_3}\right)^m, \dots, \left(\frac{c}{b}\right)^m$ , respectivamente (Figura 91). E as alturas dos superiores:  $\left(\frac{c}{a}\right)^m, \left(\frac{c}{a_1}\right)^m, \left(\frac{c}{a_2}\right)^m, \dots, \left(\frac{c}{a_{n-1}}\right)^m$ , respectivamente (Figura 92). Utilizando as abscissas das limitações dos intervalos das bases em  $g(t)$  e fazendo, nesta função,  $s = a$ , obtém-se:

$$\begin{aligned} [g(b) - g(a_{n-1})] + [g(a_{n-1}) - g(a_{n-2})] + [g(a_{n-2}) - g(a_{n-3})] + \dots + [g(a_1) - g(a)] \\ = g(b) - g(a) \\ = g(b). \end{aligned}$$

Notando que na equação acima  $g(a) = 0$ , pois em  $g(t)$  fizemos  $s = a$ .

Com base em (A.3), consideremos as desigualdades:

$$\begin{aligned} r[a, a_1] &\leq g(a_1) - g(a) \leq R[a, a_1] \\ r[a_1, a_2] &\leq g(a_2) - g(a_1) \leq R[a_1, a_2] \\ r[a_2, a_3] &\leq g(a_3) - g(a_2) \leq R[a_2, a_3] \\ &\cdot \\ &\cdot \\ &\cdot \\ r[a_{n-1}, b] &\leq g(b) - g(a_{n-1}) \leq R[a_{n-1}, b]. \end{aligned}$$

Somando os membros correspondentes das desigualdades acima, obtém-se:

$$\begin{aligned} r_n[a, b] &\leq g(b) - g(a) \leq R_n[a, b] \\ \implies r_n[a, b] &\leq g(b) \leq R_n[a, b]. \end{aligned} \tag{A.4}$$

Em que:

$$\begin{aligned} r_n[a, b] &= r[a, a_1] + r[a_1, a_2] + \dots + r[a_{n-1}, b], \\ R_n[a, b] &= R[a, a_1] + R[a_1, a_2] + \dots + R[a_{n-1}, b]. \end{aligned}$$

Fazendo  $n$  muito grande na dupla desigualdade (A.4), temos que  $r_n[a, b]$  e  $R_n[a, b]$  convergem à  $A_{-m}[a, b]$ . Como  $g(b)$  está espremido por  $r_n[a, b]$  e  $R_n[a, b]$ , tem-se que  $g(b) = A_{-m}[a, b]$  (fazendo  $s = a$ ) conforme queríamos demonstrar.



---

## BIBLIOGRAFIA

---

- [1] BORTOLOSSI, QUEIROZ, SILVA; *A Lei de Zipf e Outras Leis de Potência em Dados Empíricos. Projeto Klein de Matemática em Língua Portuguesa*, (2012). Disponível em: <<http://klein.sbm.org.br/wp-content/uploads/2012/12/Zipt-bortolossi-queiroz-dasilva-lpp-projeto-klein.pdf>>. Acesso em: 22 jan. 2015.
- [2] MORETTIN, Pedro A.; BUSSAB, Wilton De O. *Estatística Básica*. 8.ed. São Paulo: Saraiva, 2013.
- [3] NEWMAN, M. E. J.; *Power laws, Pareto distributions and Zipf's law. Contemporary Physics*, v. 46, p. 323-351, (2005) [3.ed (2006)] .Disponível em: <<http://arxiv.org/pdf/cond-mat/0412004.pdf>>. Acesso em: 08 jan. 2015.
- [4] PINHO, Frank Magalhães de. *Modelos de espaço de estados não Gaussianos - Distribuições de Caudas Pesadas*. Belo Horizonte, 2012. Dissertação (Doutorado em Estatística) - Universidade Federal de Minas Gerais. Disponível em: <[http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/BUOS-978HAV/frank\\_magalh\\_es\\_de\\_pinho\\_\\_\\_tese\\_de\\_doutorado.pdf?sequence=1](http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/BUOS-978HAV/frank_magalh_es_de_pinho___tese_de_doutorado.pdf?sequence=1)>. Acesso em: 08 jan. 2015.
- [5] ROGERIO, Mauro Urbano; BADAN, Ana Amelia Fleury de Almeida; SILVA, Helio Correa da. *Calculo diferencial e integral: funcoes de uma variavel.*. 2.ed. Goiania: Universidade Federal de Goias, 1994. Disponível em: <<https://jhacruz.mat.ufg.br/n/44832-calculo-diferencial-e-integral>>. Acesso em: 06 mai. 2016.
- [6] ROSS, Sheldon. *PROBABILIDADE, Um curso moderno com aplicações.*.Tradução de Alberto Resende De Conti. 8.ed. Bookman Companhia Editora Ltda, 2010.
- [7] SILVA, Adilton; STOSIC, Tatijana; STOSIC, Borko. *Lei de potência para cidades brasileiras. 53º Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria*, 2008 .Disponível em: <<http://www.dex.ufla.br/53rbras/trabalhos/43.pdf>>. Acesso em: 08 jan. 2015.