

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

CLEOMAR TAUMATURGO DA SILVA

UMA EXPLORAÇÃO DA TEORIA DE PROBABILIDADE APLICADA EM FINANÇAS

CURITIBA

2022

CLEOMAR TAUMATURGO DA SILVA

UMA EXPLORAÇÃO DA TEORIA DE PROBABILIDADE APLICADA EM FINANÇAS

An exploration of probability theory applied in finances

Dissertação apresentada como requisito para obtenção do título de Mestre no Programa Mestrado Profissional em Matemática em Rede Nacional - PROFMAT da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientadora: Profa. Dra. Denise de Siqueira

Coorientador: Prof. Dr. Andres David Baez Sanchez

CURITIBA

2022



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite que outros remixem, adaptem e criem a partir do trabalho licenciado para fins não comerciais, desde que atribuam ao autor o devido crédito e que licenciem as novas criações sob termos idênticos.



**Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Curitiba**



CLEOMAR TAUMATURGO DA SILVA

UMA EXPLORAÇÃO DA TEORIA DE PROBABILIDADE APLICADA EM FINANÇAS

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Matemática.

Data de aprovação: 29 de Julho de 2022

Dra. Denise De Siqueira, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Isaac Pinheiro Dos Santos, Doutorado - Universidade Federal do Espírito Santo (Ufes)

Dr. Joao Luis Goncalves, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 29/07/2022.

*Dedico este trabalho a meu pai Antonio da
Silva*

AGRADECIMENTOS

- A minha mãe Araci que sempre priorizou o melhor pelos filhos.
- Ao meu falecido pai Antonio que sempre motivou a formação acadêmica.
- A minha esposa Patricia pelo apoio, atenção e dedicação em todos os momentos.
- A meu irmão Evilásio.
- A minha orientadora Profa. Dra. Denise de Siqueira e coorientador Prof. Dr. Andres David Baez Sanchez pela compreensão, dedicação, orientação e ensinamentos durante a realização desta dissertação.
- Aos colegas de curso Ricardo, Paulo, Felipe, Fábio, Enoque, Hugo, Isaías, João, Mariana, Jéssica e Victória.
- Aos professores do PROFMAT da UTFPR - câmpus Curitiba pelos ensinamentos.
- À Sociedade Brasileira de Matemática que, na busca da melhoria do ensino de matemática na Educação Básica, viabilizou a implementação do PROFMAT.
- À CAPES, pela recomendação do PROFMAT por meio do parecer do Conselho Técnico Científico da Educação Superior.

O risco vem de não saber o que você está fazendo. Warren Buffett (1930 -): presidente do conselho e diretor executivo da Berkshire Hathaway.

RESUMO

TAUMATURGO, CLEOMAR. **Uma exploração da teoria de probabilidade aplicada em finanças**. 77 f. Dissertação - Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Universidade Tecnológica Federal do Paraná. Curitiba, 2022.

O presente trabalho introduz alguns tópicos de probabilidade aplicados às finanças. Ao longo do trabalho é apresentada uma introdução a vários dos principais conceitos que envolvem o estudo de probabilidade, como, por exemplo, conceitos de variáveis aleatórias, distribuições de probabilidade além de uma breve introdução à análise de sobrevivência. Todos os conceitos são explorados em exemplos que remetem o mercado de ações.

Palavras-chave: Probabilidade; Análise de Sobrevivência; Mercado de ações.

ABSTRACT

TAUMATURGO, Cleomar. **An exploration of probability theory applied to finances.** 77 pg. Dissertation - Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Universidade Tecnológica Federal do Paraná. Curitiba, 2021.

The present work introduces some topics of probability applied to finance. Throughout the work, an introduction to several of the main concepts that involve the study of probability is presented, such as concepts of random variables, probability distributions, and a brief introduction to survival analysis. All ideas are explored in examples that refer to the stock market.

Keywords: Probability, Survival Analysis, Stock Market.

LISTA DE FIGURAS

Figura 2.1 – Diagrama de árvore para o preço após dois dias	19
Figura 2.2 – Diagrama de árvore para o preço após três dias	19
Figura 2.3 – Diagrama de árvore para o preço após quatro dias	20
Figura 2.4 – Diagrama de árvore para o preço após cinco dias	21
Figura 2.5 – Distribuição de probabilidade do preço da ação após 2 dias	22
Figura 2.6 – Distribuição de probabilidade do preço da ação após 3 dias	23
Figura 2.7 – Distribuição de probabilidade do preço da ação após 4 dias	24
Figura 2.8 – Distribuição de probabilidade do preço da ação após 5 dias	25
Figura 2.9 – Diagrama de árvore para a probabilidade após dois dias	26
Figura 2.10–Diagrama de árvore para a probabilidade após três dias	27
Figura 2.11–Variação dos índices, Selic, IBOV e IFIX no período de 05/02/2019 a 27/10/2021	35
Figura 2.12–Função de distribuição probabilidade de Bernoulli $f(k, p)$	39
Figura 2.13–Distribuição de probabilidade do tipo Bernoulli para n ensaios com $n \in$ $\{4, 5, 10, 20\}$ e $p = 0.5$	40
Figura 2.14–Distribuição de probabilidade do tipo Bernoulli para n ensaios com $k \in$ $\{5, 10, 15, 20\}$ e $p = 1/i, i = 1, \dots, 9$	41
Figura 2.15–Distribuição do tipo Poisson para diferentes valores de λ	44
Figura 2.16–Função de distribuição exponencial para diferentes valores de λ	48
Figura 2.17–Distribuição acumulada Exponencial para diferentes valores de λ no intervalo $[0, 10]$	49
Figura 3.1 – Ajuste linear para o preço do ativo ARZZ3	52
Figura 3.2 – Ajustes de diferentes ordens polinomiais para o preço do ativo ARZZ3	56
Figura 3.3 – Ajuste do tipo exponencial com $g(x) = 40.94 \cdot e^{0.0021x}$ para o preço do ativo ARZZ3	58
Figura 4.1 – Número de semanas levado, por cada empresa listada na Tabela 4.2 para retornar ao valor de referência	62
Figura 4.2 – Função de sobrevivência $S(t)$ (a esquerda) e função de risco $h(t)$ (a direita) para $\lambda = 5$	65
Figura 4.3 – Função de sobrevivência $S(t)$ (esquerda) e função de risco $h(t)$ (direita) para diferentes valores λ	65
Figura 4.4 – Tempo de retorno (em semanas) ao valor de referência, por cada empresa listada na Tabela 4.2	69
Figura 4.5 – Estimativa de Kaplan-Meier para função de sobrevivência	69
Figura 4.6 – Função de probabilidade acumulada $\hat{F}(t)$	70
Figura 4.7 – Risco acumulado para o grupo de ratos	71

Figura 4.8 – Função de sobrevivência $\hat{S}(t)$	74
Figura 4.9 – Probabilidade acumulada $\hat{F}(t)$	74
Figura 4.10–Função de risco acumulado $\hat{H}(t)$	74

LISTA DE TABELAS

Tabela 2.1 – Distribuição de probabilidade após 2 dias	22
Tabela 2.2 – Valores para a distribuição de probabilidade do ativo após 3 dias	23
Tabela 2.3 – Tabela de distribuição de probabilidade após 4 dias	24
Tabela 2.4 – Distribuição de probabilidade após 5 dias	25
Tabela 2.5 – Função de Probabilidade para Ω_1	25
Tabela 2.6 – Função de Probabilidade para Ω_2	26
Tabela 2.7 – Probabilidades para Ω_3	27
Tabela 2.8 – Coeficientes das distribuições de probabilidade até 5 dias	28
Tabela 2.9 – Coeficientes binomiais do triângulo de Pascal	28
Tabela 2.10–Probabilidades do preço do ativo ser maior que R\$ 12.00	30
Tabela 2.11–Probabilidades para o preço do ativo após 10, 15 e 20 dias	30
Tabela 2.12–Histórico dos índices Selic, IBOV e IFIX no período de 05/02/2019 a 27/10/2021	34
Tabela 2.13–Frequências	35
Tabela 2.14–Probabilidades Condicionais entre os índices IBOV, IFIX e Selic	37
Tabela 2.15–Frequências de ocorrências simultâneas entre os índices Selic, IBOV e IFIX	37
Tabela 2.16–Probabilidade das intersecções entre eventos envolvendo os índices IBOV, IFIX e Selic	38
Tabela 2.17–Preços e probabilidades	40
Tabela 2.18–Função de densidade binomial X Função de densidade de Poisson	44
Tabela 2.19–Valor semanal dólar no período de 03/05/2020 até 03/03/2022	45
Tabela 2.20–Probabilidades ao fim de 5 dias	46
Tabela 3.1 – Cotação do ativo AZZR3 no período de 04/05/2020 a 02/06/2020	51
Tabela 3.2 – Dados do ajuste linear $g(x) = 40.95 + 0.09x$ para o preço do ativo AZZR3	54
Tabela 3.3 – Erros dos ajustes	58
Tabela 3.4 – Tabela comparativa da previsão de preço do ativo ARZZ3	59
Tabela 4.1 – Preço Referência	61
Tabela 4.2 – Tempo de retorno ao preço de referência	61
Tabela 4.3 – Falhas (retorno ao preço referência) por período, para os dados da Tabela 4.2	67
Tabela 4.4 – Dados referentes ao estudo de tratamento de câncer num grupo de ratos . . .	68
Tabela 4.5 – Dados de análise de sobrevivência para o estudo de câncer em um grupo de ratos	68
Tabela 4.6 – Funções $\hat{S}(t)$ e $\hat{H}(t)$ para o grupo de ratos	71
Tabela 4.7 – Função de sobrevivência $\hat{S}(t)$ e de risco acumulado $\hat{H}(t)$ para as empresas da bolsa de valores	73

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Uma breve introdução ao mercado financeiro	14
2	PROBABILIDADE	16
2.1	Breve contexto histórico	16
2.2	Preliminares	16
2.3	Distribuição de probabilidade	17
2.3.1	Probabilidade condicional	31
2.4	Funções de distribuição	38
2.4.1	Ensaio de Bernoulli e distribuição binomial	38
2.4.2	Valor esperado	40
2.4.3	Distribuição do tipo Poisson	42
2.5	Função de distribuição acumulada para variáveis discretas	45
2.5.1	Distribuição de probabilidade do tipo exponencial	48
3	MODELOS DE REGRESSÃO	51
3.1	Introdução	51
3.1.1	Regressão linear	51
3.1.2	Ajustes polinomiais	54
3.1.3	Regressão exponencial	57
4	INTRODUÇÃO A ANÁLISE DE SOBREVIVÊNCIA	60
4.1	Introdução	60
4.2	Conceitos básicos	62
4.2.1	Estimador de Kaplan-Meier	66
4.3	Uma aplicação de análise de sobrevivência no mercado de ações	72
5	CONCLUSÃO	75
	REFERÊNCIAS	77

1 INTRODUÇÃO

No dia 11/03/2020 a Organização Mundial da Saúde decretou a propagação de Covid-19 como pandemia, o medo passou a tomar conta do mundo todo. O pânico permaneceu no mercado financeiro mundial por alguns dias. Nesse dia o principal índice da bolsa de valores de São Paulo fecharia em queda de 7.64% um dia após com queda de 14.78%. O IBOV que havia registrado 106625 pontos no dia 02/03/2020, caiu em 19/03/2020 aos 61691 pontos ¹.

Muitas das empresas que tiveram fortes quedas durante o período inicial da pandemia, ainda não retornaram aos seus preços de antes desse período, acumulando portanto um prejuízo aos seus investidores.

Na contramão, o número de investidores pessoas físicas na Bolsa de Valores de São Paulo, vem crescendo desde então. Em dezembro de 2020 o número de pessoas físicas que investem na bolsa atingiu quase 3.2 milhões, um aumento de cerca de 300% em relação ao mesmo período de 2018 ². A maioria dessas pessoas acaba entrando no mercado de ações sem um conhecimento prévio dos riscos e podem perder todo ou parte do dinheiro investido. Portanto é fundamental conhecer alguns conceitos relacionados ao mercado financeiro.

Neste contexto, o objetivo principal deste trabalho é introduzir conceitos básicos de probabilidade e modelos de regressão, usando como motivação, exemplos e situações inspiradas no mercado financeiro.

Um exemplo simples de probabilidade aplicada às finanças pode ser tentar estabelecer, sob determinadas condições, a probabilidade de aumento ou queda do preço de um determinado ativo. Situações como esta, em que se tem apenas dois resultados possíveis *sucesso ou fracasso*, estão relacionados com os conceitos de ensaios de Bernoulli e distribuição de probabilidade binomial. No Capítulo 2 são introduzidos os conceitos básicos de probabilidade, algumas distribuições de probabilidade de variável aleatória discreta como ensaio de Bernoulli, distribuição binomial e distribuição de Poisson, além da distribuição de probabilidade exponencial relacionada a uma variável aleatória contínua.

A variação de preços dos ativos negociados na bolsa de valores sempre gerou o interesse por parte de seus investidores de tentar prever os futuros preços. Uma forma de tentar estabelecer esta previsibilidade poderia ser usando modelos de regressão a partir de dados históricos. No Capítulo 3 tratamos de alguns modelos básicos de regressão como: regressões polinomiais e regressão exponencial e analisamos seu desempenho para prever preços futuros de um ativo específico.

¹ <<https://br.investing.com/indices/bovespa-historical-data>>

² <https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/>

As corretoras também disponibilizam a seus clientes a opção de vender ou comprar um ativo assim que atingir um determinado valor. Logo é de interesse analisar o tempo até que ocorra este evento. Neste contexto uma ferramenta para analisar este tipo de situação é a *Análise de sobrevivência*. A análise de sobrevivência, tem por objetivo estudar o tempo até a ocorrência de um determinado evento de interesse. No Capítulo 4 introduzimos alguns desses conceitos bem como o estimador de Kaplan-Meier e consideramos algumas aplicações inspiradas no mercado financeiro.

É importante mencionar que a Base Nacional Comum Curricular (BNCC) implementou a Educação Financeira como um tema transversal a ser abordado em diferentes disciplinas. No estado do Paraná desde 2021 a disciplina de Educação Financeira é ofertada na grade curricular do ensino médio e visa oferecer ferramentas para que os estudantes tenham uma vida financeira mais saudável e sejam mais conscientes sobre o consumo. Este trabalho pode ser utilizado como um material de apoio, para explorar a interdisciplinaridade entre a matemática básica, presente no ensino médio, e a Educação Financeira.

Para a elaboração desse trabalho utilizaram-se diversas ferramentas computacionais, em particular Google Sheets e sua função Google Finance foram utilizadas para a obtenção e organização das cotações das empresas e índices.

1.1 UMA BREVE INTRODUÇÃO AO MERCADO FINANCEIRO

Os primeiros registros de negociações similares às que hoje são realizadas nas bolsas de valores, datam do século XV e suas operações eram limitadas às negociações de moedas, letras de câmbio e metais preciosos. Anteriormente as transações eram realizadas nas ruas e a primeira sede física de que se tem registro estava localizada na cidade de Bruges, Bélgica, onde as operações eram realizadas no edifício do Senhor Van der Burse. Nessa época ainda não era comum a numeração nas residências, e sua identificação era feita por brasões. O brasão da família Burse eram três bolsas e dessa forma a casa ficou conhecida como casa das bolsas. Posteriormente outros locais onde eram realizadas transações comerciais passaram a ser identificadas como *Bolsas* (BORGES, 1997).

O conceito de *ações* de uma empresa surge em 1602, quando o governo holandês uniu seis empresas de navegação, criando a Companhia Holandesa das Índias Orientais, e começou a disponibilizar ao público geral participações na companhia. Cada participação daria direito a uma parte dos possíveis lucros ou prejuízos da companhia. A Companhia das Índias Orientais tornou-se a maior e mais lucrativa empresa da época. (PINHEIRO, 2019).

A história das bolsas no Brasil se remonta à primeira metade do século XIX com o surgimento de diversas bolsas regionais (BORGES, 1997). Mais recentemente, em 1986, foram inauguradas a Bolsa Mercantil e de Futuros (BM&F) e a Central de Custódia e Liquidação Financeira de Títulos (CETIP). Em 2000 a BOVESPA formaliza a integração de todas as bolsas

regionais do país e passa a ser a única bolsa de valores do Brasil. Finalmente, em 2017 cria-se a B3 (Brasil, Bolsa e Balcão), com a fusão entre BM&F, BOVESPA e CETIP³.

No Brasil, o índice IBOV é um indicador das ações mais importantes negociadas na B3. Podemos considerar esse índice como um termômetro do desempenho das empresas e consequentemente da economia do país. Criada em 02 de janeiro de 1968, representa uma carteira hipotética de uma quantidade pré determinada de ativos.

O cálculo de pontuação do índice é feito somando-se um lote de 100 ações de cada um dos ativos considerados, acrescentando dividendos, juros sobre capital próprio, proventos e bonificações. Por exemplo, se o índice estiver em 120000 pontos, isso representa a soma do preço de 100 ações de cada um dos ativos que compõem o índice. Assim, se um investidor deseja ter um lote com 100 ações de cada um dos ativos que compõem o índice, seria preciso investir R\$120.000,00.

Outro índice revelante no mercado brasileiro é o *Índice de Fundos de Investimentos Imobiliários - IFIX*, lançado em 02 de setembro de 2012. Assim como o IBOV, o IFIX representa um termômetro do desempenho dos fundos de investimentos imobiliários que o compõem.

De acordo com dados da B3, mais de 400 empresas estão listadas na bolsa de valores, e para as negociações de cada uma delas usam-se *Tickers*, que resumidamente, é o código usado para representar algum ativo. Se o número final for 3, este ativo é uma ação *Ordinária*, essas ações dão direito a voto para o acionista. Se o número final for 4, trata-se de uma ação *Preferencial*, não tem direito a voto, mas tem prioridade no recebimento de dividendos. *Tickers* com final 11 referem-se as *Units*, sendo uma mescla das duas anteriores, isto é, os acionistas têm direito a voto a recebem dividendos.

Outro indicador importante na economia geral é a taxa Selic (Sistema Especial de Liquidação e Custódia), estabelecida pelo Banco Central (BC), através do Comitê de Política Monetária (Copom). Em reuniões que ocorrem em média a cada 45 dias, que definem se a taxa será mantida, reduzida ou terá aumento.

Esta taxa serve como referência para os juros que serão cobrados pelas instituições financeiras e seus produtos oferecidos, como empréstimos e financiamentos. A Selic também tem papel importante na regulação da economia do país. Em momentos em que se deseja estimular a movimentação da economia, o BC pode diminuir a taxa dos juros, tornando o crédito mais barato, estimulando assim empréstimos e financiamentos. Em contrapartida, quando ocorre necessidade de controlar a inflação o BC pode subir a taxa de juros, freando assim o consumo.

Sendo assim a taxa Selic está diretamente relacionada com o momento presente da economia e com os caminhos que se deseja seguir. Taxas mais elevadas ou baixas podem afetar no desempenho de ativos da bolsa de valores, e consequentemente influenciar outros índices como o IBOV e o IFIX.

³ <<https://www.acervob3.com.br/historia-da-bolsa>>, última data de acesso 01/07/2022.

2 PROBABILIDADE

Neste capítulo inicialmente será abordado um breve histórico da origem da Teoria de Probabilidades, posteriormente conceitos básicos e por fim alguns modelos de distribuições de probabilidade, que serão utilizados no decorrer do trabalho. O desenvolvimento deste capítulo teve como referências Viali (2009) e Bussab e Morettin (2010).

2.1 BREVE CONTEXTO HISTÓRICO

A Teoria das Probabilidades pode ser definida como: “o ramo da matemática que pretende modelar fenômenos não determinísticos, isto é, aqueles fenômenos em que o ‘acaso’ representa um papel preponderante.” (VIALI, 2009, p. 143).

Neste contexto, o *acaso*, pode ser entendido como: “Um conjunto de forças, em geral, não determinadas ou controladas, que exercem individualmente ou coletivamente papel preponderante na ocorrência de diferentes resultados de um experimento ou fenômeno.” (VIALI, 2009, p. 144).

“A Teoria das Probabilidades como disciplina matemática originou-se das tentativas de quantificação dos riscos associados a sinistros (navrágios, acidentes, morte, etc.) e da quantificação das possibilidades de se ganhar em jogos de azar.” (VIALI, 2009, p. 145).

O desenvolvimento marítimo, a ocorrência de roubos e navrágios e o crescimento populacional nas cidades da Europa trouxe a necessidade e criação de seguros, e com eles surgem os primeiros estudos matemáticos relacionados a seguros e probabilidade.

O primeiro trabalho sobre seguros de vida data de 1693. Em Halley (1693), Edmond Halley, aquele mesmo do cometa, analisou a mortalidade e natalidade na cidade de Breslaw a fim de calcular o valor adequado da anuidade do seguro de vida, como função da expectativa de vida do indivíduo e da probabilidade de sobrevivência.

2.2 PRELIMINARES

Supondo o lançamento de um dado honesto, as possibilidades para a face superior são: 1, 2, 3, 4, 5, 6. Esse conjunto de valores será chamado de espaço amostral e denotado por Ω , dessa forma $\Omega = \{1, 2, 3, 4, 5, 6\}$. Considere que o dado caia com a face superior no número 5, esse evento será denotado por A .

Utilizaremos a definição clássica para determinar a probabilidade de ocorrer o evento A , esta definição considera uma razão que tem como numerador o número de casos favoráveis a ocorrência de A , representado por $n(A)$, e como denominador o número de casos possíveis, representado por $n(\Omega)$. A probabilidade de ocorrer o evento A será:

$$P(A) = \frac{\text{n}^\circ \text{ de casos favoráveis de ocorrer } A}{\text{n}^\circ \text{ de casos possíveis}} = \frac{n(A)}{n(\Omega)}.$$

No caso do lançamento do dado, como $\Omega = \{1, 2, 3, 4, 5, 6\}$ e o evento A corresponde somente a um desses casos, então temos que $n(A) = 1$, $n(\Omega) = 6$, logo a probabilidade de ocorrer o evento A , representado por $P(A)$ será:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{1}{6} = 16.66\%.$$

Observe que cada face do dado cair varado para cima e seu respectivo evento tem a mesma probabilidade de acontecer, dizemos que os eventos são equiprováveis.

A partir da definição anterior, obtende-se algumas propriedades básicas das probabilidades:

- i) Para todo evento A , tem-se que $0 \leq P(A) \leq 1$.
- ii) $P(\Omega) = 1$ e $P(\emptyset) = 0$, ou seja, quando o evento é o espaço amostral todo, todos os casos são favoráveis. Do mesmo modo, quando o evento é o conjunto vazio, nenhum caso é favorável.
- iii) Dados dois eventos A e B , tem-se que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- iii) Se A e B forem eventos mutuamente exclusivos, então $P(A \cup B) = P(A) + P(B)$.
- iv) Para qualquer evento A , $P(A) + P(A^c) = 1$ em que $A^c = \Omega \setminus A$ é o complementar do evento A .

2.3 DISTRIBUIÇÃO DE PROBABILIDADE

Considerando novamente a situação do lançamento de um dado, seja X a variável correspondente ao número da face superior.

A variável aleatória X pode assumir valores de 1 a 6 e cada um desses valores corresponde a uma probabilidade. Quando listamos cada um desses valores de uma variável aleatória com a sua probabilidade, formamos uma distribuição de probabilidade que será representado por $P(X = k)$. A expressão $X = k$ significa que a variável X assume o valor numérico k . Uma distribuição de probabilidade deve satisfazer as seguintes condições:

- i) A probabilidade de cada valor da variável está entre 0 e 1, ou seja, $0 \leq P(X = k) \leq 1$;
- ii) A soma de todas as probabilidades é igual a 1, $\sum_k P(X = k) = 1$.

As variáveis aleatórias podem ser discretas ou contínuas. Dizemos que uma variável aleatória é discreta se ela assume um número finito de valores ou um número infinito de valores que são contáveis. Podemos considerar que uma variável é discreta quando seus valores podem ser listados.

Por exemplo: o número de ligações recebidas em uma pizzaria, pode ser um valor igual a $0, 1, 2, 3, 4, 5, \dots$. Desta forma a variável aleatória X é definida como:

- X : número de ligações recebidas pela pizzaria.

Uma variável aleatória é contínua quando os valores que ela pode assumir podem ser representados como um intervalo na reta dos números reais.

Por exemplo, no lançamento de martelo nas Olimpíadas, os valores de medição do lançamento podem tomar qualquer valor positivo podendo assumir uma infinidade não enumerável de possibilidades. Neste caso, se X = distância do arremesso, então X seria uma variável aleatória contínua.

A seguir, consideremos um exemplo referente a variável aleatória discreta, inspirada no desempenho de um ativo hipotético.

Exemplo 2.1. *Suponha que um certo ativo se comporte da seguinte maneira: cada dia aumenta ou decresce R\$0.50 (não fica igual). Assim se o preço hoje for de R\$10.00, amanhã só pode ser R\$10.50 ou R\$9.50. Sendo assim, se o preço hoje for de R\$10.00, quais os preços possíveis em 2 dias, 5 dias e n dias?*

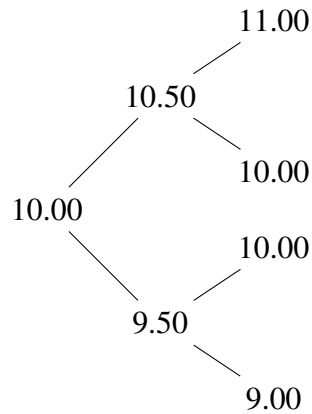
Veremos inicialmente os possíveis preços do ativo ao final do segundo dia. Como o ativo tinha preço inicial de R\$ 10.00 e o valor pode aumentar ou diminuir R\$ 0.50 a cada dia, no final do primeiro dia os possíveis preços serão de R\$ 9.50 ou R\$ 10.50. No segundo dia, partindo de R\$ 9.50 os possíveis preços podem ser de R\$ 9.00 ou R\$ 10.00, do mesmo modo, se o valor era de R\$ 10.50, agora os possíveis valores podem ser de R\$ 10.00 ou R\$ 11.00. Portanto, ao final do segundo dia os possíveis preços do ativo poderiam ser, R\$ 9.00, R\$ 10.00 ou R\$ 11.00. Denotaremos o conjunto de preços do ativo ao final do segundo dia por Ω_2 :

$$\Omega_2 = \{R\$ 9.00, R\$ 10.00, R\$ 11.00\}. \quad (2.1)$$

É possível organizar essas informações num diagrama conhecido como *diagrama de árvore*, que será apresentado na Figura 2.1.

Seguindo os preços obtidos ao final do segundo dia, se o valor foi de R\$ 9.00, os novos valores podem ser de R\$ 9.50 ou R\$ 8.50, com o preço de R\$ 10.00 pode-se obter R\$ 10.50 ou R\$ 9.50 e se o valor final no segundo dia foi de R\$ 11.00 os possíveis valores podem ser de R\$ 11.50 ou R\$ 10.50. Assim o conjunto de preços ao final do terceiro dia pode ser representado

Figura 2.1 – Diagrama de árvore para o preço após dois dias



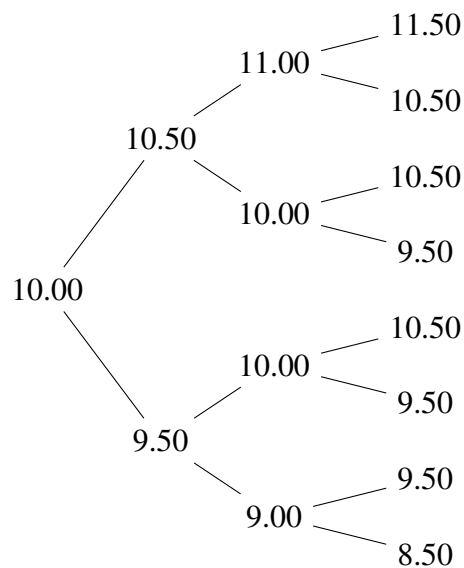
Fonte: Autor.

pelo conjunto:

$$\Omega_3 = \{R\$ 8.50, R\$ 9.50, R\$ 10.50, R\$ 11.50\}. \quad (2.2)$$

Todas as possibilidades de preço para o terceiro dia pode ser visualizadas no diagrama de árvore a seguir.

Figura 2.2 – Diagrama de árvore para o preço após três dias



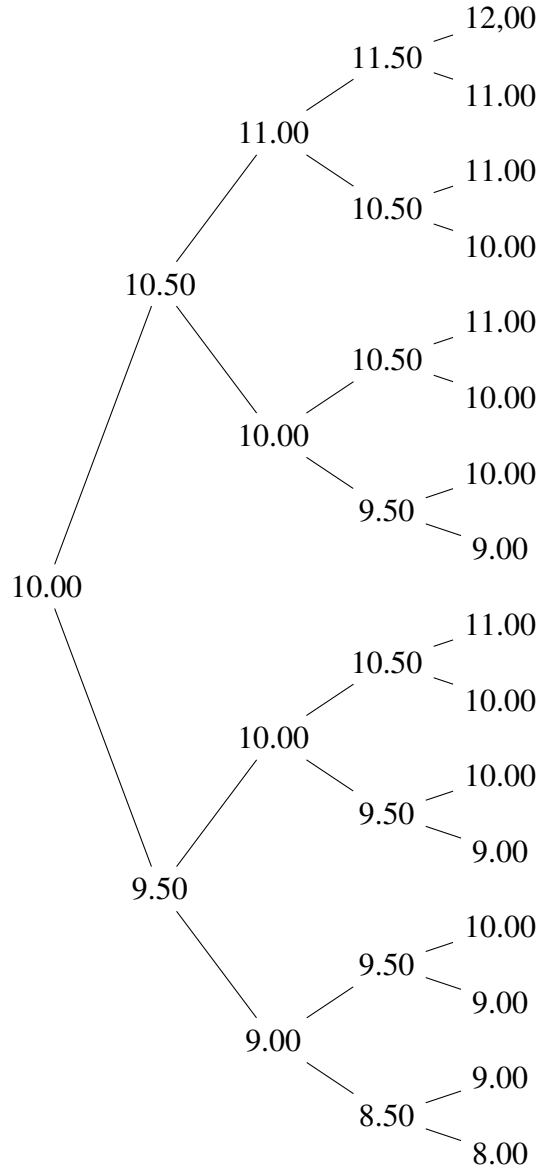
Fonte: Autor.

A partir dos valores observados no conjunto de valores em Ω_3 , se o valor ao final do terceiro dia foi de $R\$ 8.50$ então os novos possíveis valores podem ser de $R\$ 8.00$ ou $R\$ 9.00$. Se o valor foi de $R\$ 9.50$, os preços obtidos no 4º dia podem ser de $R\$ 10.00$ ou $R\$ 9.00$. Com o valor de $R\$ 10.50$ as possibilidades podem ser de $R\$ 11.00$ e $R\$ 10.00$ e finalmente, se o valor ao final do 3º dia foi de $R\$ 11.50$ então os valores poderam ser de $R\$ 12.00$ ou $R\$ 11.00$. E o conjunto de preços ao final do quarto dia pode ser representado pelo conjunto:

$$\Omega_4 = \{R\$ 8.00, R\$ 9.00, R\$ 10.00, R\$ 11.00, R\$ 12.00\}. \quad (2.3)$$

Todas as possibilidades de preços obtidas no 4º dia podem ser observadas no seguinte diagrama de árvore

Figura 2.3 – Diagrama de árvore para o preço após quatro dias



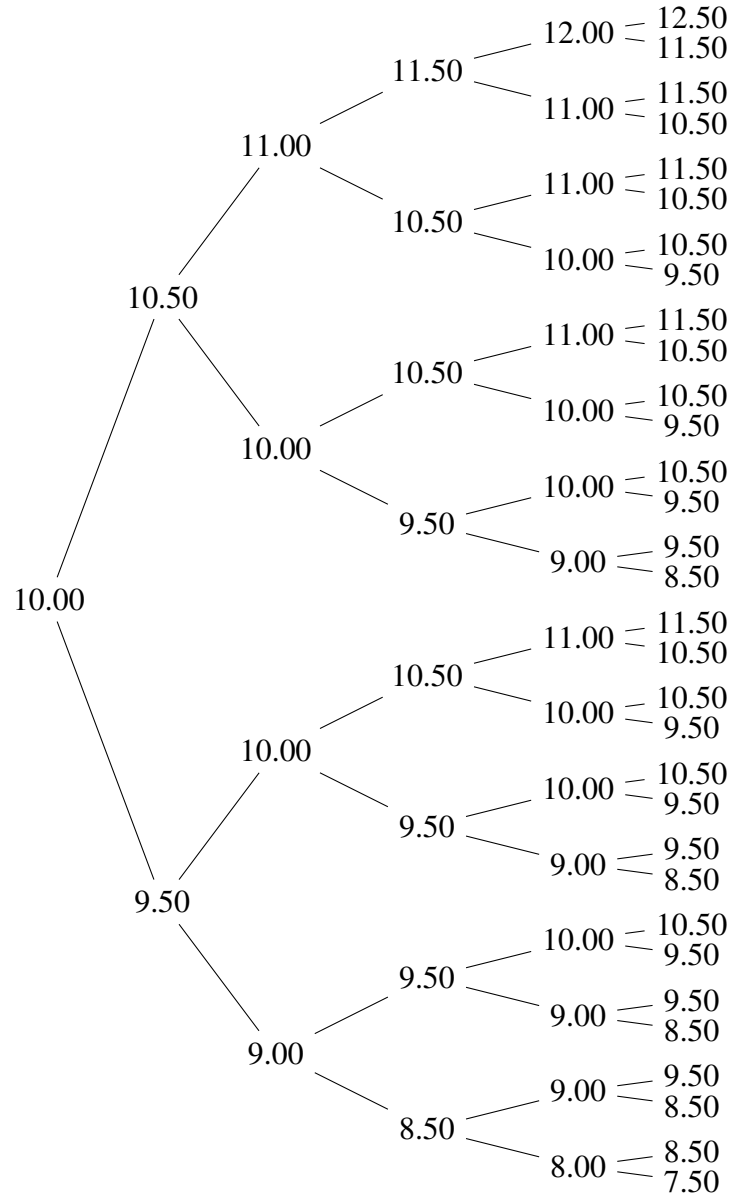
Fonte: Autor.

Com os valores obtidos em Ω_4 , podemos obter os possíveis valores ao final do quinto dia, iniciando com o valor de R\$ 8.00, os possíveis preços podem ser de R\$ 7.50 ou R\$ 8.50. Sendo o preço final no quarto dia de R\$ 9.00 os preços podem ser de R\$ 8.50 ou R\$ 9.50. Com o valor de R\$ 10.00 podemos obter R\$ 9.50 ou R\$ 10.50. Se o valor foi de R\$ 11.00 os possíveis preços podem ser de R\$ 10.50 ou R\$ 11.50. Finalmente, se o valor ao final do quarto dia foi de R\$ 12.00, então os possíveis novos valores podem ser de R\$ 11.50 ou R\$ 12.50. Assim o conjunto de preços ao final do quinto dia pode ser representado pelo seguinte conjunto de valores:

$$\Omega_5 = \{R\$ 7.50, R\$ 8.50, R\$ 9.50, R\$ 10.50, R\$ 11.50, R\$ 12.50\}. \quad (2.4)$$

As possibilidades de preços ao final do quinto dia do ativo, podem ser visualizados no seguinte diagrama de árvore:

Figura 2.4 – Diagrama de árvore para o preço após cinco dias



Fonte: Autor.

Observe que, para cada n dias temos $n + 1$ valores possíveis para o preço do ativo, ou seja, a cardinalidade, cujo símbolo matemático é $\#$, do conjunto Ω_n é $n + 1$, isto é $\#\Omega_n = n + 1$.

Pode-se observar também que o valor máxima do preço do ativo no dia n é de $10 + 0.50 \cdot n$ e o valor mínimo de $10 - 0.5 \cdot n$.

Além disso é possível escrever o conjuntos de possíveis valores para o n -ésimo dia como:

$$\Omega_n = \begin{cases} \{10 + i\} \cup \{10 - i\}, i = 0, \dots, n - 1 & \text{se } n \text{ par,} \\ \{10 + 0.50(2i + 1)\} \cup \{10 - 0.50(2i + 1)\}, i = 0, \dots, n - 1 & \text{se } n \text{ ímpar.} \end{cases}$$

Exemplo 2.2. Ainda com relação ao Exemplo 2.1, suponha que a probabilidade de aumentar ou cair, em cada dia, sejam de 50%. Determinar a distribuição de probabilidade para o preço do ativo nos seguintes casos: (a) 3 dias, (b) 4 dias, (c) 5 dias.

Considere inicialmente 2 dias, neste caso o espaço amostral (conjunto de valores possíveis para o preço do ativo) é $P_2 = \{9, 10, 11\}$. Observando o diagrama de árvore 2.1, o valor R\$ 9.00 pode ocorrer 1 dentre as 4 possibilidades, o valor R\$ 10.00 poderá ocorrer 2 entre as 4 possibilidades e por fim R\$ 11.00 terá ocorrência de 1 em 4. A probabilidade do preço do ativo ser 9 ao final do segundo dia será representado por $P(x = 9)$, 10 será representado por $P(x = 10)$ e a probabilidade de 11 será $P(x = 11)$. Obtendo as seguintes probabilidades:

1. $P(x = 9) = 1/4 = 0.25$,
2. $P(x = 10) = 2/4 = 0.50$,
3. $P(x = 11) = 1/4 = 0.25$.

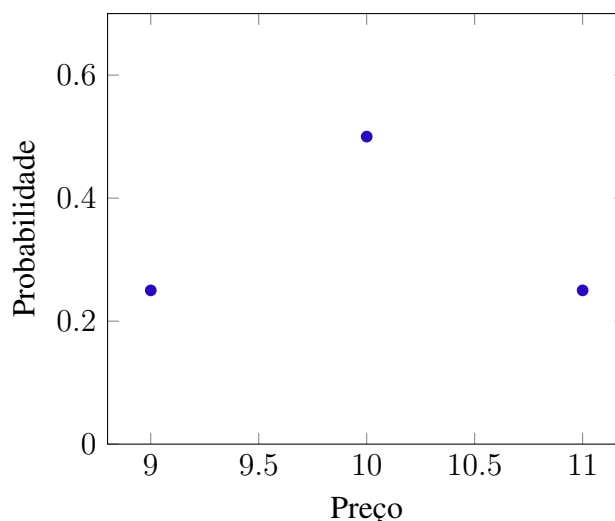
Essa distribuição de probabilidades pode ser organizada na Tabela 2.1 e representada na Figura 2.5.

Tabela 2.1 – Distribuição de probabilidade após 2 dias

k	9	10	11
$P(x = k)$	0.25	0.50	0.25

Fonte: Autor.

Figura 2.5 – Distribuição de probabilidade do preço da ação após 2 dias



Fonte: Autor.

Considere agora os possíveis valores após 3 dias, neste caso o espaço amostral será $P_3 = \{8.50, 9.50, 10.50, 11.50\}$. Podemos observar no diagrama de árvore 2.2, a ocorrência dos valores entre as 8 possibilidades, o valor de R\$ 8.50 ocorre 1 vez, R\$ 9.50 acontece 3 vezes, R\$ 10.50 também se repete 3 vezes e o preço de R\$ 11.50 pode ocorrer 1 vez, assim temos,

1. Probabilidade do preço ser de 8.50:

$$P(X = 8.50) = 1/8 = 0.125$$

2. Probabilidade do preço ser de 9.50:

$$P(X = 9.50) = 3/8 = 0.375$$

3. Probabilidade do preço ser de 10.50:

$$P(X = 10.50) = 3/8 = 0.375$$

4. Probabilidade do preço ser de 11.50:

$$P(X = 11.50) = 1/8 = 0.125$$

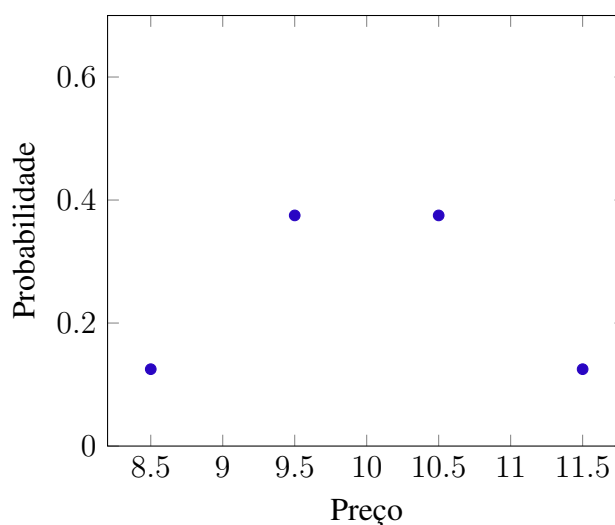
Essa distribuição de probabilidade esta resumida na Tabela 2.2 e ilustrada na Figura 2.6.

Tabela 2.2 – Valores para a distribuição de probabilidade do ativo após 3 dias

k	8.50	9.50	10.50	11.50
$P(X = k)$	0.125	0.375	0.375	0.125

Fonte: Autor.

Figura 2.6 – Distribuição de probabilidade do preço da ação após 3 dias



Fonte: Autor.

Considerando os possíveis valores após 4 dias, o espaço amostral será:

$$\Omega_4 = \{8.00, 9.00, 10.00, 11.00, 12.00\}.$$

Podemos obter a ocorrência de cada um desses valores dentre as 16 possibilidades observando o diagrama de árvore apresentado na Figura 2.3, ou seja,

1. Probabilidade do preço ser de 8.00:

$$P(x = 8.00) = 1/16 = 0.0625$$

2. Probabilidade do preço ser de 9.00:

$$P(x = 9.00) = 4/16 = 0.25$$

3. Probabilidade do preço ser de 10.00:

$$P(x = 10.00) = 6/16 = 0.375$$

4. Probabilidade do preço ser de 11.00:

$$P(x = 11.00) = 4/16 = 0.25$$

5. Probabilidade do preço ser de 12.00:

$$P(x = 12.00) = 1/16 = 0.0625.$$

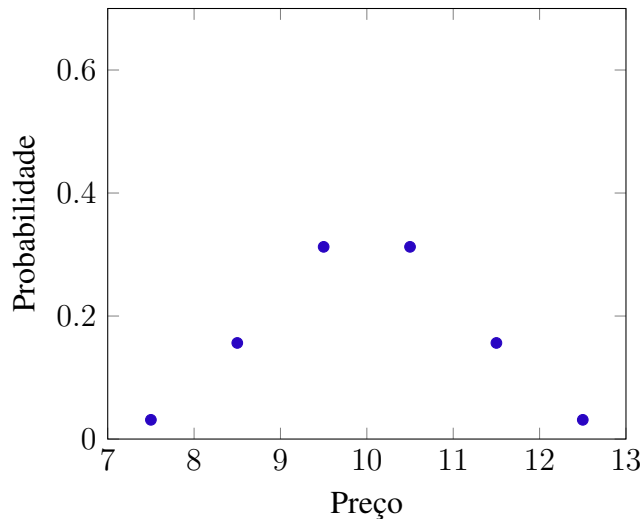
Essa distribuição de probabilidade está resumida na Tabela 2.3 e representada na Figura 2.7.

Tabela 2.3 – Tabela de distribuição de probabilidade após 4 dias

k	8.00	9.00	10.00	11.00	12.00
$P(X = k)$	0.0625	0.25	0.375	0.25	0.0625

Fonte: Autor.

Figura 2.7 – Distribuição de probabilidade do preço da ação após 4 dias



Fonte: Autor.

Finalmente, considerando os possíveis valores após 5 dias, o espaço amostral será:

$$\Omega_5 = \{7.50, 8.50, 9.50, 10.50, 11.50, 12.50\},$$

e sua distribuição de probabilidade é apresentada na Tabela 2.4 e seu gráfico na Figura 2.8.

O processo descrito acima pode ser generalizado para n dias, da seguinte forma:

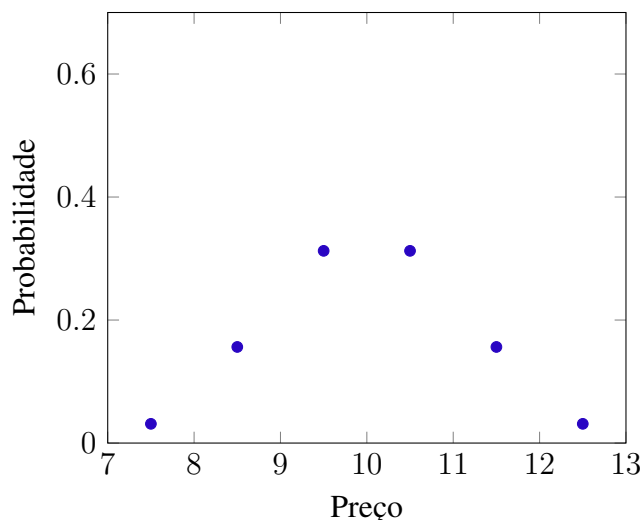
Para o aumento do ativo de R\$0.50 consideraremos sucesso e será representado por S e para a perda será considerado fracasso denotado por F, além disso, a probabilidade de sucesso

Tabela 2.4 – Distribuição de probabilidade após 5 dias

x	7.50	8.50	9.50	10.50	11.50	12.50
$P(x)$	0.03125	0.15625	0.3125	0.3125	0.15625	0.03125

Fonte: Autor.

Figura 2.8 – Distribuição de probabilidade do preço da ação após 5 dias



Fonte: Autor.

será representada por p e de fracasso, será q . Assim ao fim do primeiro dia de observação teremos apenas duas possibilidades: S ou F, com probabilidade de ocorrer de p e q , respectivamente. Dessa forma ao final do primeiro dia, o preço de R\$9.50 ocorre com 1 fracasso ou 0 sucesso e o preço de R\$10.50 ocorre com 1 sucesso. Dessa forma o espaço amostral pode ser escrito em relação ao número de sucessos, logo $\Omega_1 = \{0, 1\}$.

Note que se a probabilidade de sucesso é p logo a probabilidade de fracasso q pode ser escrita como $q = 1 - p$, pois são complementares. Tendo como probabilidades:

Tabela 2.5 – Função de Probabilidade para Ω_1

$k = n^\circ$ de sucessos (S)	0	1
$P(X = k)$	p	$1 - p$

Fonte: Autor.

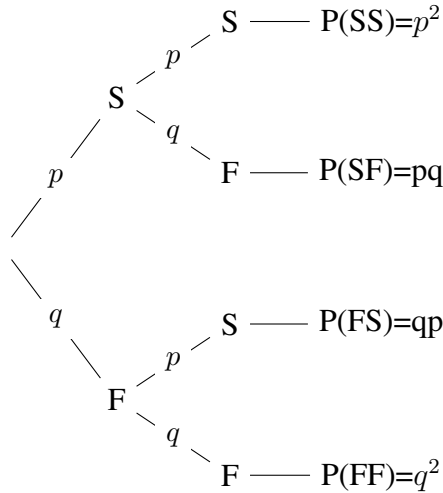
Observe que os valores da Tabela 2.5 representam a imagem da função de probabilidade associado ao espaço amostral $\Omega_1 = \{0, 1\}$, ou seja,

$$P(\Omega_1) = \{p, 1 - p\}. \quad (2.5)$$

Ao final do segundo dia o preço de R\$9.00 foi obtido com dois fracassos, representado por (FF) o preço de R\$10.00 foi obtido com uma alta e uma queda ou vice-versa, isto é, sucesso seguido de fracasso ou fracasso seguido de sucesso, representado por (SF) ou (FS) e o valor de R\$11.00 foi obtido com dois sucessos seguidos, denotado por (SS) . Perceba que no segundo dia as probabilidades de sucesso e fracasso permanecem as mesmas. Dessa forma a probabilidade

de ocorrer dois fracassos (FF) será igual a $q \cdot q = q^2$, a probabilidade de ocorrer sucesso e fracasso (SF) será de $p \cdot q$, a probabilidade de ocorrer fracasso e sucesso (FS) será de $q \cdot p$ e a probabilidade de ocorrer dois sucessos (SS) será de $p \cdot p = p^2$. O diagrama 2.9 pode exemplificar o caso:

Figura 2.9 – Diagrama de árvore para a probabilidade após dois dias



Fonte: Autor.

Observe ainda que para o caso de (FF) o número de sucessos é igual a 0, os casos (FS) o (SF) tem 1 sucesso e o caso (SS) representa 2 sucessos. No diagrama 2.9 obtemos as probabilidades p^2 , pq , qp e q^2 . Porém $qp = pq$, logo a ocorrência dessa probabilidade é de $2pq$, assim como feito no primeiro dia $q = 1 - p$. Os dados obtidos estão na seguinte tabela:

Tabela 2.6 – Função de Probabilidade para Ω_2

Preço (Real)	R\$9.00	R\$10.00	R\$11.00
Eventos	(FF)	(SF) ou (FS)	(SS)
$k = n^\circ$ de sucessos	0	1	2
$P(X = k)$	$(1 - p)^2$	$2p(1 - p)$	p^2

Fonte: Autor.

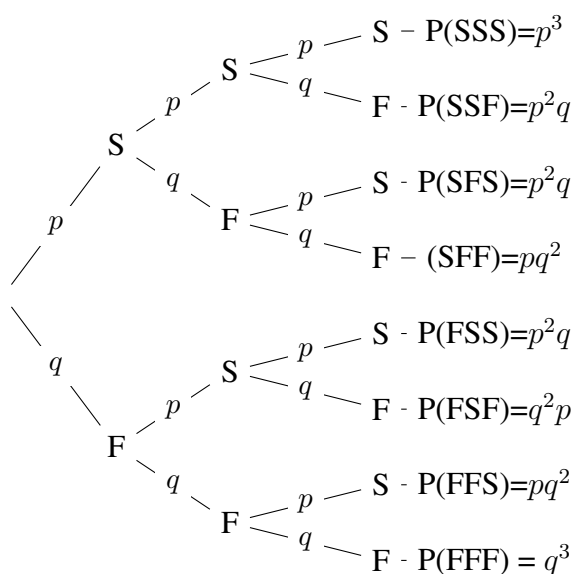
Observe que para $k = 0$ sucesso, o expoente de p é 0 e $q = 1 - p$ tem expoente 2, para $k = 1$ sucesso, o expoente de p é 1 e de $(1 - p)$ também é 1, e para $k = 2$ sucessos o expoente de p é 2 e de $(1 - p)$ é 0. Os valores de $P(X = k)$ da Tabela 2.6 representam a imagem da função de probabilidade associado ao espaço amostral $\Omega_2 = \{FF, SF, FS, SS\}$, ou seja,

$$P(\Omega_2) = \{(1 - p)^2, 2(1 - p)p, p^2\}. \tag{2.6}$$

De forma similar podemos obter as probabilidades para o dia 3 de observação do experimento, as sequências de fracasso(F) e sucesso(S) podem ser observadas no diagrama 2.10.

O espaço amostral dos preços do ativo de $\Omega_3 = \{8.50, 9.50, 10.50, 11.50\}$, o valor de 8.50 representa o evento de três fracassos seguidos isto é FFF, o valor de 9.50 pode ocorrer

Figura 2.10 – Diagrama de árvore para a probabilidade após três dias



Fonte: Autor.

com SFF, FSF ou FFS, isto é em 3 dias obter apenas 1 sucesso. O preço de 10.50 ocorre com 2 sucessos em 3 dias, logo podem ser: SSF, SFS ou FSS.

Podemos representar essas probabilidades com respectivos preços e número de sucessos com a seguinte tabela. (Utilizando $q = 1 - p$):

Tabela 2.7 – Probabilidades para Ω_3

Preço em Real	R\$8.50	R\$9.50	R\$10.50	R\$11.50
Eventos	(FFF)	(FFS), (FSF), (SFF)	(SSF), (FSS), (SFS)	(SSS)
$k = n^\circ$ de sucessos	0	1	2	3
$P(X = k)$	$(1 - p)^3$	$3p(1 - p)^2$	$3p^2(1 - p)$	p^3

Fonte: Autor.

Observe que para $k = 0$ o expoentes de p e $(1 - p)$ são respectivamente 0 e 3, para $k = 1$ os expoentes de p e $(1 - p)$ serão 1 e 2, quando $k = 2$ o expoente de p é 2 e de $(1 - p)$ será 1 e por fim para $k = 3$ o expoente de p será 3 e o expoente de $(1 - p)$ é 0. Dessa forma em n dias e k sucessos teremos que o expoente de p será k , obtendo p^k e o expoente de $(1 - p)$ será $n - k$, logo.

Os valores da Tabela 2.7 representam a imagem da função de probabilidade associada ao espaço amostral de $\Omega_3 = FFF, FFS, FSF, SFF, FSS, SFS, SSF, SSS$

$$P(\Omega_3) = \{(1 - p)^3, 3p(1 - p)^2, 3p^2(1 - p), p^3\}. \quad (2.7)$$

Observe também que os coeficientes de p e $1 - p$, podem ser organizados da seguinte maneira:

Tabela 2.8 – Coeficientes das distribuições de probabilidade até 5 dias

Dia 0	→			1				
Dia 1	→			1	1			
Dia 2	→		1	2	1			
Dia 3	→		1	3	3	1		
Dia 4	→	1	4	6	4	1		
Dia 5	→	1	5	10	10	5	1	

Fonte: Autor.

O triângulo da Figura 2.8 mostra os coeficientes encontrados nas distribuições de probabilidade para o dia 1, 2 e 3 do experimento. Mas ele também é conhecido como triângulo aritmético de Pascal, e está intimamente associado com o desenvolvimento da distribuição binomial. Cada número em qualquer linha do triângulo de Pascal é a soma dos dois números imediatamente acima da linha anterior. Observe que no terceiro dia, o 1 é obtido com os valores vizinhos da linha de cima, somente o número 1. O 3 é obtido somando os seus vizinhos de cima que são 1 e 2, o outro 3 resulta da soma de 2 e 1 e novamente o 1 tem somente como vizinho 1. De maneira similar obtemos os valores referentes ao quarto dia, 1 que tem somente o 1 adjacente a ele, $4 = 1 + 3$, $6 = 3 + 3$, $4 = 3 + 1$ e novamente o 1. Assim no quarto dia os coeficientes serão $\{1, 4, 6, 4, 1\}$, de forma similar obtemos os coeficientes para o quinto dia $\{1, 5, 10, 10, 5, 1\}$.

Note que os coeficientes dos elementos de $P(\Omega_n)$ representam números binomiais, isto é, consistem no número de combinações de “ n ” termos, k a k . Sendo este “ k ” a quantidade de sucessos desejados ao fim de cada “ n ” dias, representado por $\binom{n}{k}$.

Tabela 2.9 – Coeficientes binomiais do triângulo de Pascal

Dia 0	→			$\binom{0}{0}$				
Dia 1	→			$\binom{1}{0}$	$\binom{1}{1}$			
Dia 2	→		$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$			
Dia 3	→		$\binom{3}{0}$	$\binom{3}{1}$	$\binom{3}{2}$	$\binom{3}{3}$		
Dia 4	→	$\binom{4}{0}$	$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$		
Dia 5	→	$\binom{5}{0}$	$\binom{5}{1}$	$\binom{5}{2}$	$\binom{5}{3}$	$\binom{5}{4}$	$\binom{5}{5}$	

Fonte: Autor.

Portando a função de distribuição de probabilidade pode ser definida como:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}. \tag{2.8}$$

Assim conseguimos obter as probabilidades para o ativo, por exemplo, para o segundo dia. Observe que neste caso, dentre as possibilidades podemos ter, inicialmente nenhum sucesso, ou seja, temos (FF) , podemos ter 1 sucesso (FS) e (SF) e também dois sucessos, (SS) . Neste caso as probabilidades são:

1. Probabilidade de nenhum sucesso em 2 dias.

$$P(X = 0) = \binom{2}{0} \cdot p^0 \cdot (1 - p)^2 = 1 \cdot 1 \cdot (1 - p)^2 = (1 - p)^2 = 0.25 = 25\%.$$

2. Probabilidade de 1 sucesso em 2 dias.

$$P(X = 1) = \binom{2}{1} \cdot p^1 \cdot (1 - p)^1 = 2 \cdot p \cdot (1 - p) = (1 - p) = 0.50 = 50\%.$$

3. Probabilidade de 2 sucessos em 2 dias.

$$P(X = 2) = \binom{2}{2} \cdot p^2 \cdot (1 - p)^0 = 1 \cdot p^2 \cdot 1 = p^2 = 0.25 = 25\%.$$

Exemplo 2.3. *Nas mesmas condições do Exemplo 2.2, qual a probabilidade, em cada caso, do preço do ativo seja maior que R\$12.00 no final do período? E qual a probabilidade de ser exatamente R\$12.00? E após 20 dias?*

Para o primeiro caso de 3 dias, como a melhor possibilidade será de 3 altas sucessivas, assim o limite de valorização é de R\$ 1.50, e o maior valor será de R\$ 11.50, impossibilitando atingir ou passar o valor de R\$ 12.00.

Para 4 dias, o ativo poderá atingir o valor máximo de R\$ 12.00, se conseguir obter 4 aumentos de R\$ 0.50 nos 4 dias, isto é, 4 sucessos em 4 dias. Utilizando a Equação 2.8, temos:

$$P(X = 4) = \binom{4}{4} \cdot 0.5^4 \cdot 0.5^0 = 1 \cdot 0.0625 \cdot 1 = 0.0625 = 6.25\%.$$

No quinto dia, não temos a possibilidade de preço do ativo atingir R\$ 12.00, porém existe a possibilidade de ter 5 aumentos sucessivos e obter R\$ 12.50.

Para atingir o valor maior que R\$ 12.00 é necessário ter 5 aumentos sucessivos, para um experimento de 5 tentativas, logo $p = 0.50$, $n = 5$, $k = 5$, $q = 1 - p = 0.50$, $n - k = 0$.

$$P(X = 5) = \binom{5}{5} \cdot 0.5^5 \cdot 0.5^0 = 1 \cdot 0.03125 \cdot 1 = 0.03125 = 3.125\%.$$

Para o ativo atingir o valor de R\$ 12.00 ao final de 20 dias, devemos ter 4 sucessos a mais que fracassos, logo para esse caso deve-se ter 12 aumentos e 8 quedas. Logo, com $p = 0.50$, $n = 20$, $k = 12$ na Equação 2.8:

$$P(X = 12) = \binom{20}{12} \cdot 0.5^{12} \cdot 0.5^8 = 0.1201 = 12.01\%.$$

No dia 20 para o ativo atingir preço maior que R\$ 12.00, será preciso mais que 12 sucessos dentre as 20 tentativas, os resultados para este caso estão resumidos na Tabela 2.10.

A Tabela 2.10 fornece as probabilidades para o preço do ativo ser maior que R\$ 12.00, somando as probabilidades tem-se a probabilidade do ativo ser maior que R\$ 12.00.

Tabela 2.10 – Probabilidades do preço do ativo ser maior que R\$ 12.00

$P(X = 13) = 0,07393$	$P(X = 17) = 0.00108719$
$P(X = 14) = 0.03696$	$P(X = 18) = 0.00018120$
$P(X = 15) = 0.01479$	$P(X = 19) = 0.00001907$
$P(X = 16) = 0.00462$	$P(X = 20) = 0.00000095$

Fonte: Autor.

Portanto, a probabilidade do ativo ser maior que R\$ 12.00 é dada por

$$P(X > 12) = \sum_{i=13}^{20} P(X = i) = 0.1315, \quad (2.9)$$

ou seja, aproximadamente 13.15%.

Por fim, com a ajuda de uma planilha de cálculo, é possível calcular a distribuição de probabilidade para os preços em determinados n dias. A Tabela 2.11, foi gerada utilizando a planilha *Google Sheets* para quando $n = 10, 15$ e 20 .

Tabela 2.11 – Probabilidades para o preço do ativo após 10, 15 e 20 dias

DIA-10		DIA-15		...	DIA-20	
Probabilidade	Preço	Probabilidade	Preço		Probabilidade	Preço
0.09765625	5.00	0.0031	2.50		0.0001	0.00
0.9765625	6.00	0.0458	3.50		0.0019	1.00
4.39453125	7.00	0.3204	4.50		0.0181	2.00
11.71875	8.00	1.3885	5.50		0.1087	3.00
20.5078125	9.00	4.1656	6.50		0.4621	4.00
24.609375	10.00	9.1644	7.50		1.4786	5.00
20.5078125	11.00	15.2740	8.50		3.6964	6.00
11.71875	12.00	19.6381	9.50		7.3929	7.00
4.39453125	13.00	19.6381	10.50		12.0134	8.00
0.9765625	14.00	15.2740	11.50		16.0179	9.00
0.09765625	15.00	9.1644	12.50		17.6197	10.00
		4.1656	13.50		16.0179	11.00
		1.3885	14.50		12.0134	12.00
		0.3204	15.50		7.3929	13.00
		0.0458	16.50		3.6964	14.00
		0.0031	17.50		1.4786	15.00
					0.4621	16.00
					0.1087	17.00
					0.0181	18.00
					0.0019	19.00
					0.0001	20.00

Fonte: Autor.

Nas mesmas condições, ao final de 20 dias, qual a probabilidade do preço ser menor que R\$12.00? Qual o preço com maior probabilidade?

1. Qual a probabilidade do preço ser menor que R\$ 12.00?

A probabilidade do preço do ativo ser menor que R\$ 12 pode ser calculada somando-se as probabilidades para os preços menores que R\$ 12.00 na Tabela 2.11, logo

$$P(X < 12) = \sum_{i=0}^{11} P(X = i) = 0.748278.$$

Portanto, a probabilidade do ativo estar com preço abaixo de R\$ 12.00 é de aproximadamente 75%.

2. Qual o preço com maior probabilidade?

Observe pela Tabela 2.11 que o preço com maior probabilidade de ocorrência é R\$10.00, para isto acontecer será necessário obter 10 sucessos em 20 tentativas. Logo:

$$P(X = 10) = \binom{20}{10} \cdot 0.5^{10} \cdot 0.5^{10} = 0.176197 = 17.6197\%.$$

2.3.1 PROBABILIDADE CONDICIONAL

Em muitos casos, o experimento pode estar separado em etapas, e a informação ocorrida em uma etapa pode influenciar nas probabilidades das etapas seguintes. Dizemos que os eventos posteriores estão *condicionados* aos eventos anteriores, e para o cálculo dessas probabilidades usaremos a Probabilidade Condicional.

Para entender melhor a ideia de probabilidade condicional, considere a seguinte situação:

Um dado foi lançado, qual é a probabilidade de ter ocorrido na sua face superior o número 4, sabendo que ocorreu face par?

Neste caso o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$, composta pelas faces do dado. Consideremos os seguintes eventos:

A : ocorrer número 4,

B : ocorrer número par.

A probabilidade da face superior ser igual a 4 é $P(A) = 1/6$. Agora, dado que as faces pares são 2, 4 ou 6 segue que a probabilidade da face ser par é igual a $P(B) = 3/6$.

Sabendo que ocorreu face com número par, isto é, a face é 2, 4 ou 6, o evento A corresponde a 1 caso de 3 possibilidades, ou seja, a probabilidade de ocorrer o evento A dado que ocorreu o evento B , será de $1/3$, que denotaremos por $P(A|B) = 1/3 = 0.333\%$.

Observe que caso o evento B tenha ocorrido, o espaço amostral será reduzido, pois, todos os resultados possíveis são aqueles correspondentes ao evento B .

Com esse exemplo seguiremos para a definição de probabilidade condicional.

Definição 2.1. Dados os eventos A e B , a probabilidade do evento A , dado o evento B , denotada por $P(A|B)$ é,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.10)$$

Do mesmo modo, a probabilidade do evento B , dado o evento A , denotada por $P(B|A)$ é,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (2.11)$$

Utilizando a Definição 2.1 no exemplo do lançamento do dado, temos $A \cap B = \{4\}$ portanto $P(A \cap B) = 1/6$, e assim,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}. \quad (2.12)$$

Outra forma de utilizar a probabilidade condicional, é conhecida como *regra da multiplicação*, que fornece a probabilidade da intersecção dos eventos de interesse. Segue das Equações (2.10) e (2.11) que,

$$P(A \cap B) = P(B) \cdot P(A|B), \quad (2.13)$$

$$P(A \cap B) = P(A) \cdot P(B|A). \quad (2.14)$$

Substituindo a probabilidade da intersecção obtida na Equação (2.14) em (2.10) a probabilidade de A dado B é expresso por:

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}. \quad (2.15)$$

Da mesma forma, substituindo probabilidade da intersecção obtida na Equação (2.13) na Equação (2.11), a probabilidade de B dado A é expresso por:

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}. \quad (2.16)$$

As Equações (2.15) e (2.16) são conhecidas como **Teorema de Bayes**.

O Teorema de Bayes, que aparentemente poderia ser encarado como mais um resultado na teoria de probabilidades, tem importância fundamental, pois fornece a base para uma abordagem da inferência estatística conhecida como inferência bayesiana. (BUSSAB; MORETTIN, 2010, p. 119).

O Teorema de Bayes nos mostra como probabilidades posteriores podem ser afetadas pelas probabilidades a priori, isto é, anteriores. A fim de ilustrar este resultado, consideremos o seguinte exemplo correspondente ao Exemplo 5.16 em Bussab e Morettin (2010).

Exemplo 2.4. *A administração de um fundo de investimentos em ações pretende divulgar, após o encerramento do pregão, a probabilidade de queda do índice da bolsa no dia seguinte, baseando-se nas informações disponíveis até aquele momento. Suponha que a previsão inicial seja de 10%. Após encerrado o pregão, nova informação sugere uma alta do dólar frente ao real. Dados históricos indicam que, quando houve queda da bolsa no dia seguinte, 20% das vezes foram precedidas por informações de alta no dólar, enquanto, nos dias em que a bolsa esteve em alta, apenas em 5% das vezes houve esse tipo de notícia no dia anterior.*

A probabilidade inicial de queda para o índice da bolsa é de 10%, logo, a probabilidade do índice subir é de 90%. Nos dias de queda, historicamente, 20% desses dias foram precedidos por previsão que o dólar subiria. Nos dias de alta do índice, 5% das vezes foram precedidas pela previsão de alta no dólar.

Vale ressaltar que, estamos interpretando os dados históricos como probabilidades. Com isso estamos considerando que, eventos anteriores podem ser usados para analisar eventos futuros.

Queremos saber a probabilidade de queda do índice da bolsa, dada a informação que sugere alta do dólar. Assim será determinado pelos eventos:

A : queda do índice da bolsa,

B : previsão de alta do dólar.

Supondo que a probabilidade inicial de queda do índice seja $P(A) = 0.10$, então a probabilidade inicial de alta do índice será $P(A^c) = 0.90$. Pelo **Teorema de Bayes** será necessário determinar a probabilidade do dólar subir dado que a bolsa teve queda, segundo dados históricos essa probabilidade é igual a 0.20, isto é, $P(B|A) = 0.20$. E a probabilidade de alta do dólar, denotada por $P(B)$ ocorre nas seguintes situações: alta do índice e alta do dólar, com probabilidade de $0.90 \cdot 0.05$, outra possibilidade é a queda do índice e alta do dólar, com probabilidade de $0.10 \cdot 0.20$. Portanto a probabilidade de alta do dólar será: $P(B) = 0.90 \cdot 0.05 + 0.10 \cdot 0.20 = 0.065$.

Assim, pelo Teorema de Bayes descrito na Equação (2.15) temos,

$$P(A|B) = \frac{0.10 \cdot 0.20}{0.065} = \frac{0.02}{0.065} = 0.31 = 31\%.$$

Logo, a previsão de possível alta no dólar, atualiza a probabilidade de queda do índice da bolsa de 10% para 31%.

Exemplo 2.5. *Nesse exemplo iremos abordar a possível relação da taxa Selic em comparação com os índices IBOV e IFIX.*

A comparação será feita no mesmo período de 05/02/2019 até 27/10/2021. Os dados da taxa Selic foram extraídos diretamente da página do Banco Central¹, os dados referente a IFIX e IBOV foram obtidos usando a ferramenta do Google Finance. A análise foi realizada por 22 períodos e a variação dos três índices se encontra na Tabela 2.12 e ilustrado na Figura 2.11.

Partindo da data referida acima, a taxa Selic estava em 6.5%, o IBOV estava em 98311.20 pontos e o IFIX em 2404.14 pontos, o Copom manteve inalterada a taxa em 20/03/2019, nesta data o IBOV fechou em 98041.37 pontos e o IFIX em 2483.05 pontos, desta forma a variação dos três respectivamente foi de 0%, 0.10% e 3.28%.

Tabela 2.12 – Histórico dos índices Selic, IBOV e IFIX no período de 05/02/2019 a 27/10/2021

PERÍODO	DATA	SELIC	IBOV	IFIX
1	20/03/2019	0.00%	0.10%	3.28%
2	08/05/2019	0.00%	-2.49%	0.54%
3	20/06/2019	0.00%	4.92%	3.37%
4	31/07/2019	0.00%	1.50%	3.06%
5	18/09/2019	-7.69%	2.67%	-0.02%
6	30/10/2019	-8.33%	3.71%	4.48%
7	11/12/2019	-9.09%	2.36%	6.50%
8	05/02/2020	-10.00%	4.56%	3.03%
9	18/03/2020	-5.56%	-42.35%	-28.25%
10	06/05/2020	-11.76%	18.19%	18.24%
11	17/06/2020	-20.00%	20.85%	7.84%
12	05/08/2020	-25.00%	7.59%	-2.46%
13	16/09/2020	-11.11%	-3.04%	3.02%
14	28/10/2020	0.00%	-4.32%	-1.06%
15	09/12/2020	0.00%	18.49%	-0.08%
16	20/01/2021	0.00%	5.88%	3.59%
17	17/03/2021	0.00%	-2.59%	-1.85%
18	05/05/2021	37.50%	2.59%	1.34%
19	16/06/2021	27.27%	8.11%	-1.38%
20	04/08/2021	21.43%	-5.77%	-0.81%
21	22/09/2021	23.53%	-7.82%	-3.23%
22	27/10/2021	19.05%	-5.27%	-0.80%

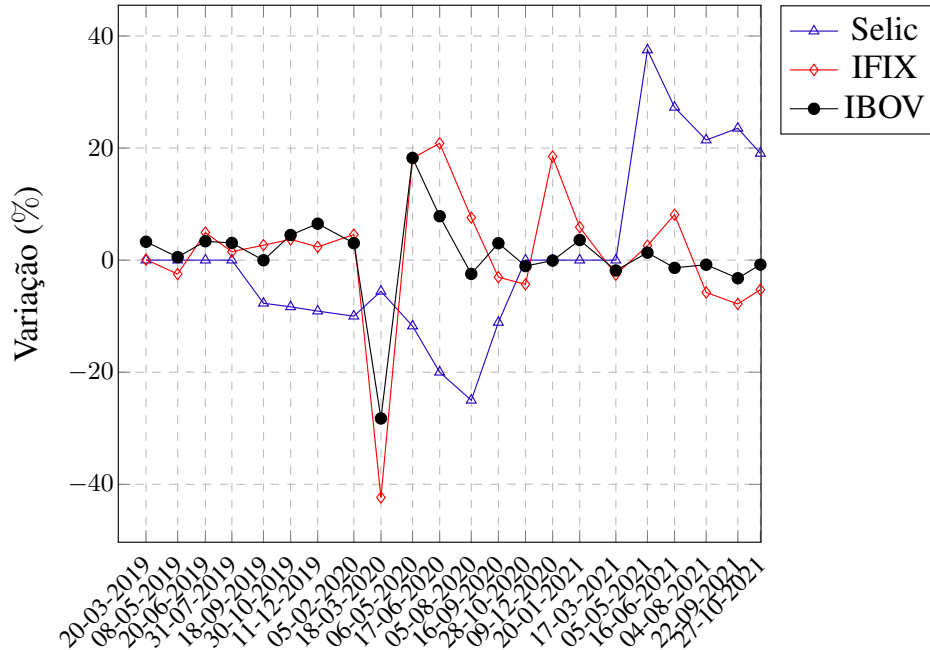
Fonte: Autor.

A partir da Tabela 2.12, podemos identificar que nos períodos 18 e 19, ambos os índices IBOV E Selic subiram. Do mesmo modo, em 7 períodos IBOV subiu enquanto a taxa Selic caiu. Observamos também que em três vezes o IBOV subiu em períodos de alta da Selic, e em dois períodos ambos caíram. Da mesma forma podemos fazer a análise comparativa para o IFIX e a taxa Selic. Essas frequências de ocorrência estão resumidas na Tabela 2.13.

A partir dessas informações poderemos estudar algumas relações entre esses índices. Por exemplo:

¹ <<https://www.bcb.gov.br/controleinflacao/historicotaxasjuros>>

Figura 2.11 – Variação dos índices, Selic, IBOV e IFIX no período de 05/02/2019 a 27/10/2021



Fonte: Autor.

Tabela 2.13 – Frequências

	IBOV SOBE	IBOV DESCE	IFIX SOBE	IFIX DESCE
SELIC SOBE	2	3	1	4
SELIC DESCE	7	2	6	3
TOTAL	9	7	7	7

Fonte: Autor.

a) Qual a probabilidade do IFIX cair dado que a Selic aumentou?

Para responder essa pergunta será utilizado o Teorema de Bayes. Para isso definimos os seguintes eventos

- A: queda do IFIX,
- B: alta da Selic,
- C: alta do IBOV.

Logo, baseado nos dados históricos da Tabela 2.12 e nas frequências da Tabela 2.13, obtemos os seguintes dados:

- i) $P(A)$ = Probabilidade do IFIX cair = $7/14$,
- ii) $P(B)$ = Probabilidade da Selic subir = $5/14$,
- iii) $P(B|A)$ = Probabilidade da Selic subir dado que o IFIX caiu = $4/7$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{4/7 \cdot 7/14}{5/14} = \frac{28/98}{5/14} = \frac{28 \cdot 14}{98 \cdot 5} = 0.8 = 80\%. \quad (2.17)$$

Logo, a probabilidade do IFIX cair nos respectivos períodos que a Selic subiu é de 80%

b) Qual a probabilidade do IFIX subir dado que a Selic aumentou?

Para este caso, perceba que a probabilidade do IFIX subir pela Tabela 2.13 será de $7/14$, que também pode ser representado como probabilidade complementar do evento A , ou $P(A^c) = 1 - P(A) = 7/14$.

Logo:

i) $P(A^c) = \text{Probabilidade do IFIX subir} = 7/14$

ii) $P(B) = \text{Probabilidade da Selic subir} = 5/14$

iii) $P(B|A^c) = \text{Probabilidade da Selic subir dado que o IFIX subiu} = 1/7$

$$P(A^c|B) = \frac{P(B|A^c)P(A^c)}{P(B)} = \frac{1/7 \cdot 7/14}{5/14} = \frac{1/14}{5/14} = \frac{1 \cdot 14}{14 \cdot 5} = 0.20 = 20\% \quad (2.18)$$

Portanto, a probabilidade do IFIX subir nos respectivos períodos que a Selic subiu é de 20%.

Note ainda que a probabilidade do IFIX cair dado que a Selic subiu e a probabilidade do IFIX subir dado que a Selic subiu são complementares, isto é, $P(A|B) + P(A^c|B) = 1$.

c) Qual a probabilidade do IBOV subir dado que a Selic caiu?

Observe que como o evento B é a alta da Selic, o evento queda da taxa Selic pode ser representado por B^c , e sua probabilidade por $P(B^c)$. Dessa forma tem-se:

i) $P(C) = \text{Probabilidade do IBOV subir} = 9/14$

ii) $P(B^c) = \text{Probabilidade da Selic cair} = 9/14$

iii) $P(B^c|C) = \text{Probabilidade da Selic cair dado que o Ibov subiu} = 7/9$

$$P(C|B^c) = \frac{P(B^c|C)P(C)}{P(B^c)} = \frac{7/9 \cdot 9/14}{9/14} = \frac{63/126}{9/14} = \frac{63 \cdot 14}{126 \cdot 9} = 0.777 = 77.7\% \quad (2.19)$$

Portanto, a probabilidade do índice IBOV subir dado que a taxa Selic teve queda é de aproximadamente 77.7%.

Como visto, a probabilidade do índice IBOV cair dado que a taxa Selic teve queda representado por $P(C^c|B^c)$ será dado por $1 - P(C|B^c) = 0.222$. Portanto, a probabilidade do índice IBOV cair dado que a taxa Selic teve queda é de aproximadamente 22.22%.

Todos os valores calculados das probabilidades condicionais estabelecidas entre IBOV e Selic, IFIX e Selic se encontram na Tabela 2.14

Observe que as mesmas probabilidades podem ser obtidas com o auxílio da Tabela 2.15, por exemplo, a probabilidade do IBOV subir dado que a Selic subiu pode ser dado pela razão

Tabela 2.14 – Probabilidades Condicionais entre os índices IBOV, IFIX e Selic

	IBOV SOBE	IBOV DESCE	IFIX SOBE	IFIX DESCE
SELIC SOBE	40%	60%	20%	80%
SELIC DESCE	77.7%	22.2%	66.6%	33.3%

Fonte: Autor.

entre períodos que ambos subiram em simultâneo, pela quantidade de períodos de alta da Selic. Logo $2/5 = 0.40 = 40\%$. A probabilidade do IBOV subir dado que a Selic teve queda é dado pela razão entre os períodos que IBOV subiu e Selic caiu, dividido pelo número de vezes que a Selic teve queda, logo $7/9 = 0.777 = 77.7\%$. Da mesma forma obtemos as probabilidades restantes para o IBOV e IFIX.

Tabela 2.15 – Frequências de ocorrências simultâneas entre os índices Selic, IBOV e IFIX

	IBOV SOBE	IBOV DESCE	TOTAL	IFIX SOBE	IFIX DESCE	TOTAL
SELIC SOBE	2	3	5	1	4	5
SELIC DESCE	7	2	9	6	3	9
TOTAL	9	5	14	7	7	14

Fonte: Autor.

Note porém que os numeradores representam as intercessões dos eventos de interesse, a probabilidade dessas, pode ser obtido através do Teorema de Bayes, onde seu numerador nos fornece a probabilidade de intercessão de eventos de interesse. Através da Equação (2.20), definimos as probabilidades dessas intercessões.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(B) \cdot P(A|B). \quad (2.20)$$

Desta forma podemos obter a probabilidade do IFIX cair e Selic subir ao mesmo tempo:

Nesse caso a probabilidade da Selic aumentar, $P(B)$ é de $5/14$, e $P(A|B) = 0.8$, logo:

$$P(A \cap B) = P(B) \cdot P(A|B) = 5/14 \cdot 0.8 = 0,2857 = 28.57\%. \quad (2.21)$$

Assim, a probabilidade de ambos os eventos acontecerem simultaneamente é de aproximadamente 28.57%.

Vejamos agora, a probabilidade do IFIX subir e Selic aumentar, isto é $P(A^c \cap B)$.

A probabilidade da Selic subir é de $5/14$ e a probabilidade $P(A^c|B)$ é de 0.20, logo

$$P(A^c \cap B) = P(B) \cdot P(A^c|B) = 5/14 \cdot 0.2 = 0.071 = 7.1\%. \quad (2.22)$$

Assim, a probabilidade dos eventos ocorrerem simultaneamente é de aproximadamente 7.1%.

Baseado no que foi exposto, podemos obter as probabilidades das intersecções que envolvem os índices IBOV, IFIX e Selic, conforme resumidos na Tabela 2.16.

Tabela 2.16 – Probabilidade das intersecções entre eventos envolvendo os índices IBOV, IFIX e Selic

$P(A \cap B)$	IBOV subir	IBOV cair	IFIX subir	IFIX cair
SELIC subir	14.28%	21.42%	7.1%	28.57%
SELIC cair	50%	14.28%	42.85%	21.42%

Fonte: Autor.

No mercado financeiro, a alta da taxa Selic é vista como um controlador da inflação, e um atrativo para o investimento em títulos públicos, sendo um investimento de baixo risco. Da mesma forma, uma taxa mais elevada da Selic diminuiu os financiamentos e compras de imóveis, afastando assim os investidores dos fundos de investimentos imobiliários, das ações e dos índices da bolsa de valores, através das cotações históricas. Os dados acima obtidos evidenciam essas correlações. Por exemplo, a probabilidade do IFIX cair dado uma alta da Selic foi de 80%, e para o índice IBOV cair dado uma alta da Selic é de 60%.

Até o momento discutimos e apresentamos exemplos que nos auxiliaram a entender sobre o espaço amostral, cálculo de probabilidade de um evento ocorrer bem como a noção de distribuição de probabilidade, a seguir conheceremos algumas distribuições de probabilidade frequentes na literatura.

2.4 FUNÇÕES DE DISTRIBUIÇÃO

Ao longo desta seção apresentaremos algumas funções de distribuição de probabilidade relacionadas aos Exemplos 2.1 e 2.2.

2.4.1 ENSAIOS DE BERNOULLI E DISTRIBUIÇÃO BINOMIAL

Nos Exemplos 2.1 e 2.2 consideramos que o ativo a cada dia de observação apresentava apenas duas opções para o preço, ter sucesso e aumentar R\$0.50 ou fracasso e perder R\$0.50 além disso, o aumento (ou perda) não dependia do que tivesse acontecido no dia anterior. Este tipo de situação é conhecido como *Ensaio de Bernoulli*.

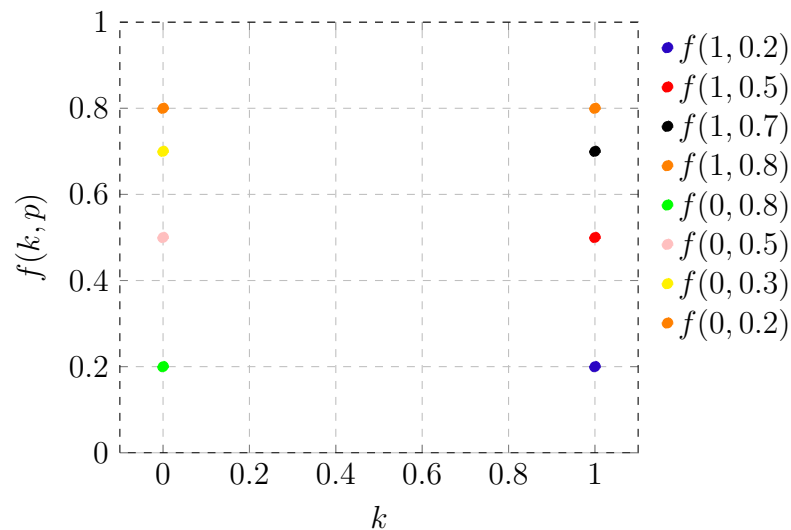
Considere um experimento que consiste em uma sequência de ensaios ou tentativas independentes, isto é, ensaios nos quais a probabilidade de um resultado em cada ensaio não depende dos resultados anteriores, nem afetará os resultados posteriores, por exemplo, o lançamento sucessivo de uma moeda. Cada lançamento é independente do anterior e, além disso, em cada ensaio, podem ocorrer apenas dois resultados, cara ou coroa. Em geral, esses eventos serão chamados de *sucesso* ou *fracasso*. Como visto anteriormente, a probabilidade de acontecer o sucesso será denotado por p , e a probabilidade de fracasso por q , sendo este complementar de p , isto é, $q = 1 - p$.

Denotando por $f(k, p)$ a probabilidade de obter k sucessos em um ensaio de Bernoulli tem-se,

$$f(k, p) = \begin{cases} p, & \text{se } k = 1 \\ 1 - p, & \text{se } k = 0 \end{cases} \quad (2.23)$$

denominada *Distribuição de Bernoulli*. A Figura 2.12, apresenta o gráfico de $f(k, p)$ para diferentes valores de p .

Figura 2.12 – Função de distribuição probabilidade de Bernoulli $f(k, p)$



Fonte: Autor.

No Exemplo 2.2, consideramos situações como, determinar a probabilidade de obter 3 altas do ativo ao final de 5 dias e na Equação (2.8), determinamos uma expressão para probabilidade de obter k sucessos ao final de n dias.

Em geral, pode-se considerar a probabilidade de se obter k sucessos ao final de n ensaios de Bernoulli. Denotando por $f(n, k, p)$ a probabilidade de obter k sucessos em n repetições de um ensaio de Bernoulli tem-se,

$$f(n, k, p) = P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \quad (2.24)$$

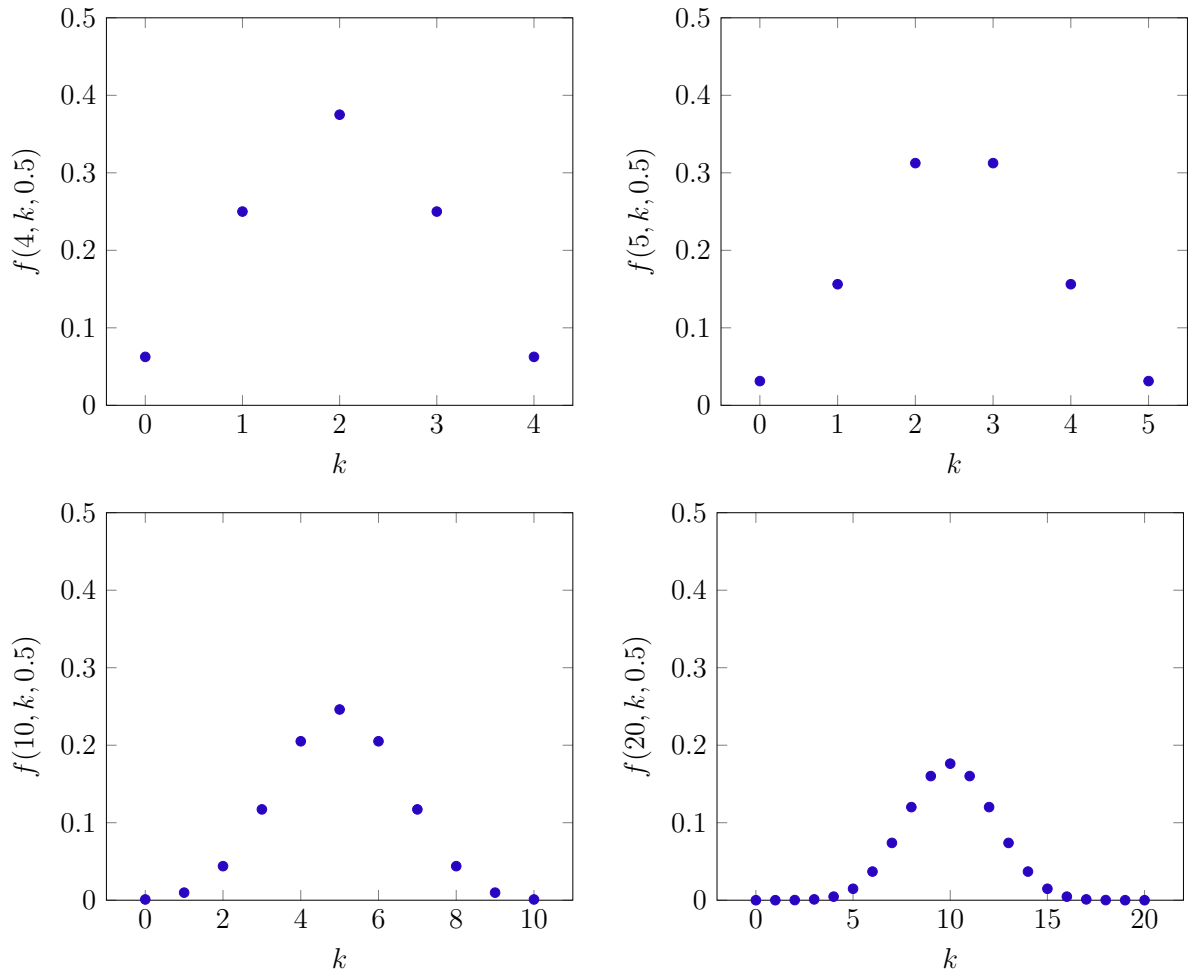
denominada *função de distribuição Binomial*.

A Figura 2.13 apresenta o gráfico da função $f(n, k, 0.5)$ com $n \in \{4, 5, 10, 20\}$ e $k = 0, \dots, n$.

Consideremos ainda um exemplo em que foram realizados 25 tentativas e queiramos obter 5, 10, 15 e 20 sucessos. Faremos a seguir uma simulação para os valores encontrados, sendo a probabilidade $p = \frac{i}{10}$ com $i = 1, \dots, 9$. Os valores obtidos se encontram na Figura 2.14.

Observe nos gráficos da Figura 2.14 que nos casos abordados, com $n = 25$ a probabilidade de obter exatamente 5 sucessos é maximizada quando $p = 0.2$, e nesse caso $f(25, 0.2, 5) = 0.1960$. Para obter exatamente 10 sucessos a probabilidade será maximizada em

Figura 2.13 – Distribuição de probabilidade do tipo Bernoulli para n ensaios com $n \in \{4, 5, 10, 20\}$ e $p = 0.5$



Fonte: Autor.

$p = 0.4$, e $f(25, 0.4, 10) = 0.0916$. E para obter $k = 15$ e $k = 20$ sucessos, as probabilidades serão maximizadas para $p = 0.6$ e $p = 0.8$ respectivamente.

2.4.2 VALOR ESPERADO

No Exemplo 2.2 foram determinadas as probabilidades para o preço do ativo após alguns dias. Relacionando os possíveis preços e suas probabilidades após 5 dias de observação obtemos a Tabela 2.17. Qual o valor médio dos possíveis preços do ativo ao final de 5 dias?

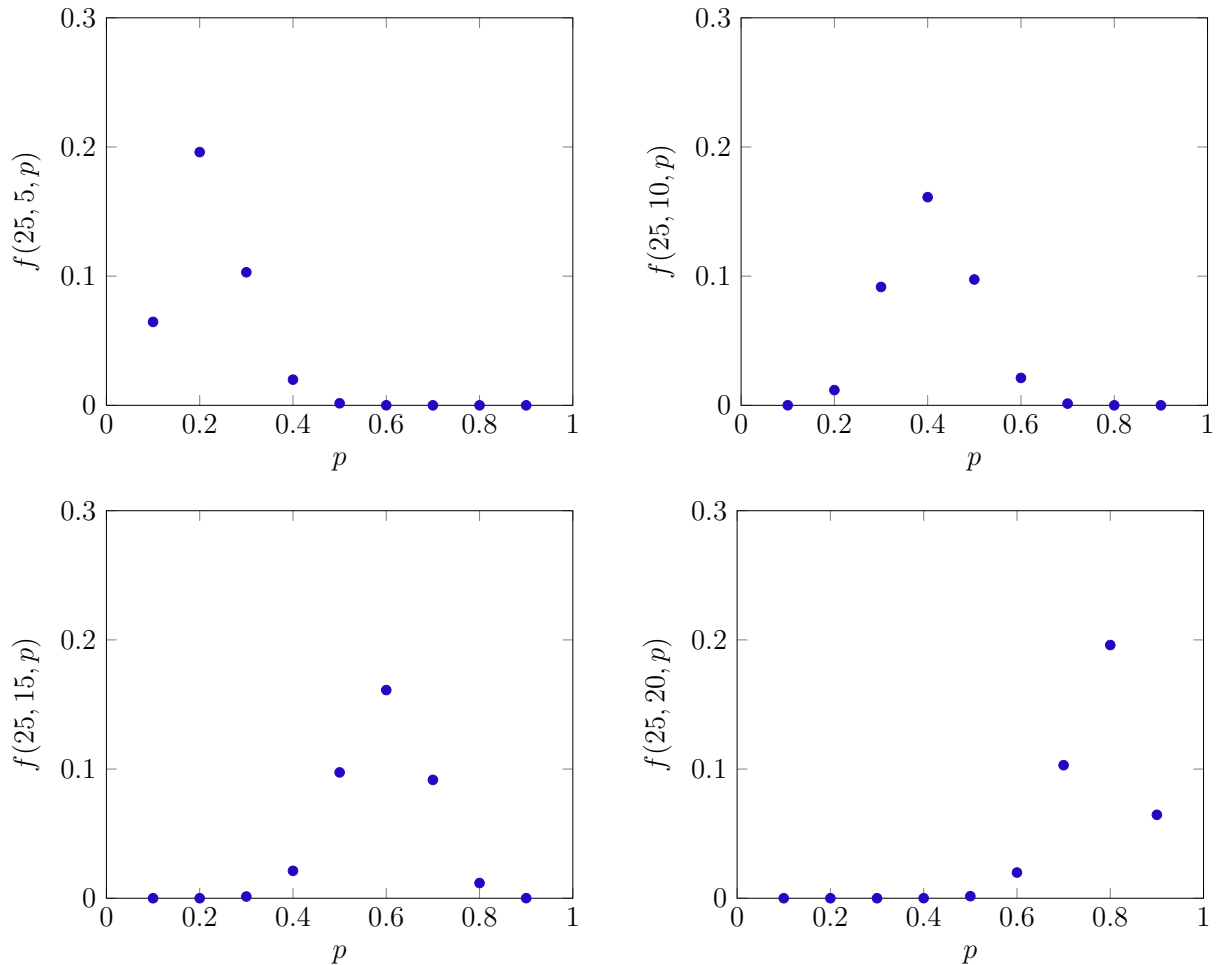
Tabela 2.17 – Preços e probabilidades

Preço	7.50	8.50	9.50	10.50	11.50	12.50
Probabilidade	1/32	5/32	10/32	10/32	5/32	1/32

Fonte: Autor.

Note que as probabilidades representam as frequências dos preços dentre 32 possibilida-

Figura 2.14 – Distribuição de probabilidade do tipo Bernoulli para n ensaios com $k \in \{5, 10, 15, 20\}$ e $p = 1/i, i = 1, \dots, 9$



Fonte: Autor

des. Logo podemos obter a média dos 32 possíveis resultados como:

$$\frac{7.50 + 5 \times 8.50 + 10 \times 10.50 + 5 \times 11.50 + 12.50}{32} = \frac{320}{32} = 10 \quad (2.25)$$

Em geral, esta noção de média pode ser definida como o valor esperado ou esperança matemática e para variáveis discretas é definida como:

$$E(X) = \sum_{i=1}^n x_i \cdot P(X = x_i) = \sum_{i=1}^n x_i \cdot p_i, \quad (2.26)$$

onde x_i representam os possíveis resultados para a variável aleatória e p_i suas respectivas probabilidades.

Assim, o valor de R\$ 10.00, encontrado em (2.25), é o valor esperado para o ativo ao final de 5 dias.

De modo geral, a distribuição binomial, dada na Equação (2.24) tem como valor esperado:

$$E(X) = \sum_{i=1}^n i \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}. \quad (2.27)$$

Utilizando a definição de número binomial, a Equação (2.27), pode ser reescrita como:

$$\begin{aligned}
 E(X) &= \sum_{i=1}^n i \frac{n}{i!(n-i)!} \cdot p^i \cdot (1-p)^{n-i} \\
 &= \sum_{i=1}^n i \frac{n(n-1)!}{i(i-1)!(n-i)!} \cdot p \cdot p^{i-1} \cdot (1-p)^{n-i} \\
 &= \sum_{i=1}^n np \frac{(n-1)!}{(i-1)!(n-i)!} \cdot p^{i-1} \cdot (1-p)^{n-i} \\
 &= np \sum_{i=1}^n \frac{(n-1)!}{(i-1)!(n-i)!} \cdot p^{i-1} \cdot (1-p)^{n-i}.
 \end{aligned} \tag{2.28}$$

Definindo $a = i - 1$ e $b = n - 1$, logo $a + 1 = i$ e $b + 1 = n$ e utilizando o Binômio de Newton, a Equação (2.28) pode ser reescrita como:

$$\begin{aligned}
 E(X) &= np \sum_{a=0}^b \frac{b!}{a!(b-a)!} \cdot p^a \cdot (1-p)^{b-a} \\
 &= np \sum_{a=0}^b \binom{b}{a} (1-p)^{b-a} \cdot p^a \\
 &= np((1-p) + p)^b = np.
 \end{aligned} \tag{2.29}$$

2.4.3 DISTRIBUIÇÃO DO TIPO POISSON

A distribuição de Poisson é uma distribuição de probabilidade discreta de uma variável aleatória X , com as seguintes características:

1. X se refere ao número de vezes, k , que um evento ocorre em um dado intervalo,
2. A probabilidade de um evento ocorrer é a mesma para cada intervalo,
3. O número de ocorrências em um intervalo é independente do número de ocorrências em outro intervalo.

Note que estas características são similares as da distribuição de probabilidade binomial e podemos considerar a função de distribuição de probabilidade de Poisson como uma função binomial onde o número de tentativas n aumenta indefinidamente e o número de sucessos k

permanece constante. Da Equação (2.24), temos,

$$\begin{aligned}
 f(n, k, p) = P(X = k) &= \frac{n!}{k!(n-k)!} \cdot p^k \cdot (1-p)^{n-k} \\
 &= \frac{n!}{k!(n-k)!} \cdot \frac{n^k}{n^k} p^k \cdot \left(1 - \frac{np}{n}\right)^{n-k} \\
 &= \frac{n!}{k!(n-k)!} \cdot \frac{(np)^k}{n^k} \cdot \left(1 - \frac{np}{n}\right)^{n-k} \\
 &= \frac{(np)^k}{k!} \cdot \frac{n!}{n^k(n-k)!} \cdot \left(1 - \frac{np}{n}\right)^{n-k} \\
 &= \frac{(np)^k}{k!} \cdot \frac{n!}{n^k(n-k)!} \cdot \left(1 - \frac{np}{n}\right)^n \cdot \left(1 - \frac{np}{n}\right)^{-k}. \quad (2.30)
 \end{aligned}$$

Denotando $np = \lambda$ a Equação (2.30) pode ser reescrita como:

$$P(X = k) = \frac{\lambda^k}{k!} \cdot \frac{n!}{n^k(n-k)!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Aplicando limite quando n tende ao infinito na expressão anterior tem-se:

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P(X = k) &= \lim_{n \rightarrow \infty} \frac{n!}{n^k(n-k)!} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &= \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \quad (2.31)
 \end{aligned}$$

Como:

- i) $\lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots(n-k+1)}{n^k} = 1,$
- ii) $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1,$
- iii) $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$

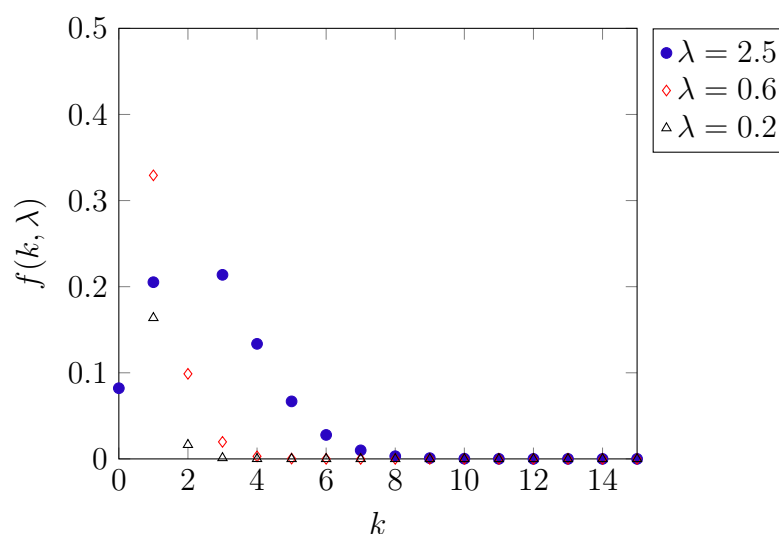
Segue que a Equação (2.31) é escrita como:

$$f(k, \lambda) : \lim_{n \rightarrow \infty} P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (2.32)$$

A função $f(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$ é denominada como função de densidade ou distribuição de Poisson.

A Figura 2.15 mostra o gráfico da função de densidade de Poisson para diferentes valores de λ .

Escolhendo $n = 1000$ e $p = 0.0001$, verificamos que as funções de distribuição de probabilidade Binomial e Poisson tomam valores próximos, como mostra a Tabela 2.18.

Figura 2.15 – Distribuição do tipo Poisson para diferentes valores de λ 

Fonte: Autor.

Tabela 2.18 – Função de densidade binomial X Função de densidade de Poisson

k	Binomial	Poisson
1	0.0904923385897146	0.0904837418035960
2	0.0045205443669929	0.0045241870901798
3	0.0001503984824569	0.0001508062363393
4	0.0000037490570809	0.0000037701559085
5	0.0000000746886859	0.0000000754031182
6	0.0000000012387112	0.0000000012567186
7	0.0000000000175915	0.0000000000179531
8	0.0000000000002184	0.0000000000002244

Fonte: Autor.

Exemplo 2.6. Durante o período de 03/05/2020 até 03/03/2022 foram coletados os valores de fechamento do Dólar em relação ao Real no Ibovespa. Durante o período de observação, o preço do ativo atingiu ou superou R\$ 5.50, em média, 10 dias por mês. Na Tabela 2.19 são apresentados os preços de fechamento semanais do dólar no período considerado.

Considerando que a variável aleatória X seja o número de vezes que o ativo atinge ou supera R\$ 5.50 durante um mês e que tenha distribuição de Poisson com média $\lambda = 10$, podemos calcular as probabilidades de alguns eventos.

1. Qual a probabilidade do dólar não atingir ou superar R\$ 5.50 nenhuma vez durante 1 mês?

Nesse caso queremos saber a probabilidade do ativo não atingir nem superar R\$ 5.50, logo $X = 0$. Temos ainda que $\lambda = 10$. Pela Equação 2.32:

$$f(0;10) : (X = 0) = \frac{10^0}{0!} e^{-10} = e^{-10} = 0.00004539 = 0.004539\%. \quad (2.33)$$

A probabilidade do dólar não atingir ou superar R\$ 5.50 no período de 1 mês é de 4.978%.

2. Qual a probabilidade do dólar atingir ou superar o valor de R\$ 5.50 duas vezes no mês?

$$f(2.10) : (X = 2) = \frac{10^2}{2!} e^{-10} = \frac{100}{2} e^{-10} = 50 \cdot e^{-10} = 0.002269 = 0.22\%. \quad (2.34)$$

A probabilidade do dólar atingir ou superar R\$ 5.50 duas vezes é de 0.22% no período de 1 mês.

3. Qual a probabilidade do dólar atingir ou superar o valor de R\$ 5.50 dez vezes durante o mês?

$$f(10.10) : (X = 10) = \frac{10^{10}}{10!} e^{-10} = 0.1251 = 12.51\%. \quad (2.35)$$

A probabilidade do dólar atingir ou superar R\$ 5.50, dez vezes no mês é de 12.51% .

Tabela 2.19 – Valor semanal dólar no período de 03/05/2020 até 03/03/2022

Data	Preço do ativo	Data	Preço do ativo	Data	Preço do ativo	Data	Preço do ativo
29/05/2020	5.33	06/11/2020	5.36	16/04/2021	5.60	24/09/2021	5.33
05/06/2020	4.96	13/11/2020	5.46	23/04/2021	5.47	01/10/2021	5.36
12/06/2020	5.05	20/11/2020	5.38	30/04/2021	5.44	08/10/2021	5.51
19/06/2020	5.31	27/11/2020	5.34	07/05/2021	5.23	15/10/2021	5.46
26/06/2020	5.48	04/12/2020	5.15	14/05/2021	5.27	22/10/2021	5.65
03/07/2020	5.31	11/12/2020	5.06	21/05/2021	5.36	29/10/2021	5.64
10/07/2020	5.32	18/12/2020	5.10	28/05/2021	5.22	05/11/2021	5.54
17/07/2020	5.38	25/12/2020	5.22	04/06/2021	5.05	12/11/2021	5.46
24/07/2020	5.23	01/01/2021	5.19	11/06/2021	5.11	19/11/2021	5.61
31/07/2020	5.22	08/01/2021	5.41	18/06/2021	5.09	26/11/2021	5.61
07/08/2020	5.43	15/01/2021	5.29	25/06/2021	4.93	03/12/2021	5.65
14/08/2020	5.42	22/01/2021	5.47	02/07/2021	5.06	10/12/2021	5.61
21/08/2020	5.62	29/01/2021	5.46	09/07/2021	5.26	17/12/2021	5.70
28/08/2020	5.39	05/02/2021	5.37	16/07/2021	5.11	24/12/2021	5.67
04/09/2020	5.30	12/02/2021	5.37	23/07/2021	5.20	31/12/2021	5.57
11/09/2020	5.32	19/02/2021	5.38	30/07/2021	5.20	07/01/2022	5.64
18/09/2020	5.39	26/02/2021	5.60	06/08/2021	5.24	14/01/2022	5.53
25/09/2020	5.56	05/03/2021	5.69	13/08/2021	5.25	21/01/2022	5.46
02/10/2020	5.68	12/03/2021	5.55	20/08/2021	5.38	28/01/2022	5.37
09/10/2020	5.53	19/03/2021	5.49	27/08/2021	5.20	04/02/2022	5.33
16/10/2020	5.64	26/03/2021	5.75	03/09/2021	5.19	11/02/2022	5.25
23/10/2020	5.62	02/04/2021	5.71	10/09/2021	5.24	18/02/2022	5.14
30/10/2020	5.74	09/04/2021	5.69	17/09/2021	5.29	25/02/2022	5.16

Fonte: Autor.

2.5 FUNÇÃO DE DISTRIBUIÇÃO ACUMULADA PARA VARIÁVEIS DISCRETAS

No Exemplo 2.2, foram determinadas as probabilidades para os possíveis preços do ativo ao final de alguns dias. Agora utilizaremos o caso de 5 dias. O preço de R\$ 7.50 representa 0 sucessos, R\$ 8.50 representa 1 sucesso, sucessivamente até o preço de R\$ 12.50 que corresponde à 5 sucessos. Esta relação, de preço, número de sucessos e probabilidade pode se observada na Tabela 2.20.

Tabela 2.20 – Probabilidades ao fim de 5 dias

x	7.50	8.50	9.50	10.50	11.50	12.50
k (Sucessos)	0	1	2	3	4	5
$P(X = k)$	0.03125	0.15625	0.3125	0.3125	0.15625	0.03125

Fonte: Autor.

Com base nos dados da Tabela 2.20, podemos considerar as seguintes situações:

1. Qual a probabilidade do preço ser menor ou igual à R\$ 10.50?

Neste caso, para o ativo atingir até o preço desejado, será necessário obter até 3 sucessos, desta forma o ativo pode atingir 0, 1, 2 e 3 sucessos. Isto é, devemos calcular a probabilidade para k sucesso ser menor ou igual a 4, denotado por $P(X \leq 3)$. Porém $P(X \leq 3)$ pode ser obtida somando-se a probabilidade de sucessos para $k = 0$, $k = 1$, $k = 2$ e $k = 3$. Logo,

$$\begin{aligned}
 P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\
 &= 0.03125 + 0.15625 + 0.3125 + 0.3125 \\
 &= 0.8125 = 81.25\%.
 \end{aligned}
 \tag{2.36}$$

Portanto a probabilidade do preço do ativo ser menor ou igual a R\$ 10.50 é de 81.25%.

2. Qual a probabilidade do preço ser maior que R\$ 10.50 ?

Observe que para do preço o ativo estar no preço desejado são apenas duas opções: R\$ 11.50 e R\$ 12.50, em função do número de sucessos, podemos ter: $k = 4$ ou $k = 5$. Dessa forma, para calcular $P(X > 3)$, podemos somar as probabilidades para $k = 4$ e $k = 5$ sucessos:

$$\begin{aligned}
 P(X > 3) &= P(X = 4) + P(X = 5) \\
 &= 0.15625 + 0.03125 = 0.1875 = 18.75\%.
 \end{aligned}
 \tag{2.37}$$

Observe que $P(X \leq 3) + P(X > 3) = 1$, logo

$$P(X \leq 3) = 1 - P(X > 3) \text{ e } P(X > 3) = 1 - P(X \leq 3).
 \tag{2.38}$$

Da mesma forma ao final do quinto dia de estudo do ativo, se quisermos saber qual a probabilidade de obter 0 sucessos ou menos, representado por $P(X \leq 0)$, como o menor valor possível é de zero sucesso, teremos $P(X \leq 0) = 0.03125$.

Para calcular a probabilidade de ocorrer até 1 sucesso, temos que as possibilidades são de ocorrer 0 e 1 sucessos, logo,

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.03125 + 0.15625 = 0.1875.
 \tag{2.39}$$

A seguir, obtemos as probabilidades de se obter dois sucessos ou menos, representado por $P(X \leq 2)$; três sucessos ou menos, $P(X \leq 3)$; 4 sucessos ou menos, $P(X \leq 4)$ e por fim até 5 sucessos, $P(X \leq 5)$:

- a) $P(X \leq 0) = P(X = 0) = 0.03125$;
- b) $P(X \leq 1) = P(X = 0) + P(X = 1) = 0.03125 + 0.15625 = 0.1875$;
- c) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.03125 + 0.15625 + 0.3125 = 0.50$;
- d) $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0.8125$;
- e) $P(X \leq 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0.96875$;
- f) $P(X \leq 5) = P(X = 0) + P(X = 1) + \dots + P(X = 4) + P(X = 5) = 1$.

Observe que, em geral, para obter $P(X \leq n)$, teremos:

$$P(X \leq n) = P(X = 0) + P(X = 1) + \dots + P(X = n) = \sum_{i=0}^n P(X = i), \quad (2.40)$$

sendo uma função que acumula os valores das probabilidades.

Dada a variável aleatória X , chamaremos de função de distribuição acumulada, simplificada por (f.d.a.), a função $F(x)$ dada por

$$F(x) = P(X \leq x). \quad (2.41)$$

O domínio de F é o conjunto dos reais e o contradomínio pertence ao intervalo $[0, 1]$, pois $0 \leq P(X \leq x) \leq 1$.

Para a função de densidade discreta do tipo Poisson, a distribuição acumulada é uma função que acumula os valores das probabilidades. Dada a função de distribuição de probabilidade do tipo Poisson (2.32), $f(\lambda, k) = \frac{\lambda^k}{k!} e^{-\lambda}$, a função de distribuição acumulada $F(k)$ será:

$$\begin{aligned} F(k) &= P(X = 0) + P(X = 1) + \dots + P(X = k) \\ &= \frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \dots + \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \left(1 + \lambda + \dots + \frac{\lambda^k}{k!} \right). \end{aligned} \quad (2.42)$$

A seguir apresentamos um exemplo clássico para o estudo de distribuição de Poisson.

Exemplo 2.7. *A emissão de partículas radioativas tem sido modelada através de uma distribuição de Poisson, com o valor do parâmetro dependendo da fonte utilizada. Suponha que o número de partículas alfa, emitidas por minuto, seja uma variável aleatória seguindo o modelo da Poisson com parâmetro 5, isto é, a taxa média de ocorrências é de 5 emissões a cada minuto. Calcular a probabilidade de haver mais de duas emissões em um minuto.*

Observe neste caso que $\lambda = 5$ emissões/ minuto e, além disso,

$$\begin{aligned} P(X > 2) &= P(X = 3) + P(X = 4) + P(X = 5) + \dots \\ &= 1 - P(X \leq 2) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - e^{-5} \left[1 + 5 + \frac{25}{2} \right] = 0.8753. \end{aligned}$$

Logo, a probabilidade de haver mais duas emissões de partículas radioativas em um minuto é de 87.53%.

2.5.1 DISTRIBUIÇÃO DE PROBABILIDADE DO TIPO EXPONENCIAL

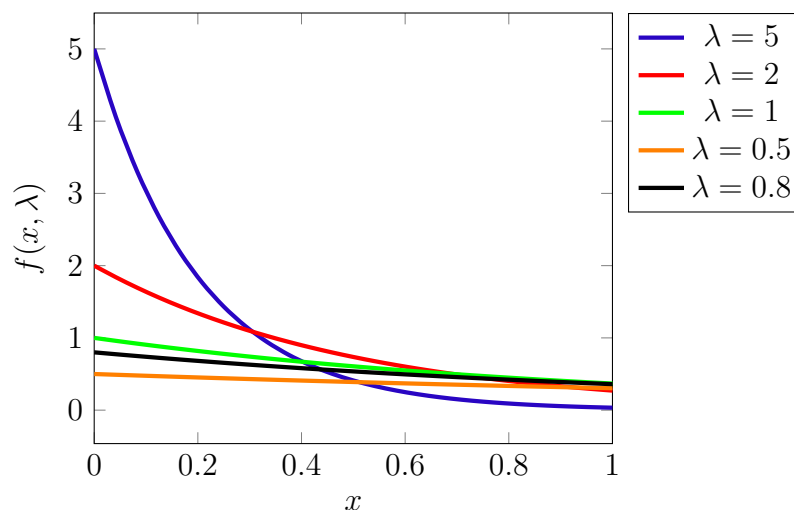
Apresentamos a seguir um modelo de distribuição de probabilidade para variáveis aleatórias contínuas, isto é, variáveis cujos possíveis valores pertencem a um intervalo de números reais.

Dizemos que uma variável aleatória contínua X segue distribuição exponencial com parâmetro $\lambda > 0$ se sua função de densidade de probabilidade (f.d.p) é dada por:

$$f(x; \lambda) = \begin{cases} \lambda \cdot e^{-\lambda x}, & \text{se } x \geq 0, \\ 0, & \text{se } x < 0. \end{cases} \quad (2.43)$$

Este modelo é bastante utilizado quando, a probabilidade de uma determinada variável aleatória é alta no início da análise e decai no decorrer do tempo. Um exemplo pode ser a venda de ingressos para shows de grande público, onde a procura pelos ingressos é maior no início das vendas se comparado pela procura pelos ingressos próximo do evento. A Figura 2.16, mostra o gráfico da distribuição do tipo exponencial para diferentes valores de λ .

Figura 2.16 – Função de distribuição exponencial para diferentes valores de λ



Fonte: Autor.

Em geral, para uma variável aleatória contínua X , com função de distribuição $f(x)$, definimos a probabilidade do evento X no intervalo a e b por:

$$P(a \leq X \leq b) = \int_a^b f(x)dx. \quad (2.44)$$

Observe que, a probabilidade do evento corresponde a área abaixo da curva de $f(x)$ no intervalo $[a, b]$.

Aplicando a Equação 2.44, temos que para uma função de distribuição $f(x)$ podemos definir a *função de distribuição acumulada* $F(x)$, para $x \geq 0$ como:

$$P(0 \leq X \leq x) = F(x) = \int_0^x \lambda \cdot f(u)du. \quad (2.45)$$

Note que, da definição anterior, tem-se que $F'(x) = f(x)$.

Para a função do tipo exponencial, temos que:

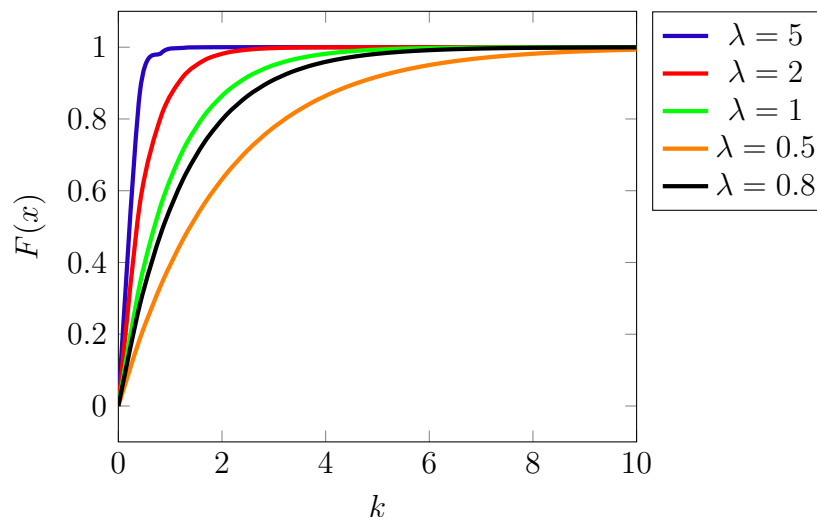
$$F(x) = \int_0^x \lambda \cdot e^{-\lambda u} du = 1 - e^{-\lambda x}. \quad (2.46)$$

Observe, portanto, que a probabilidade associada ao intervalo $[a, b]$ é dada por:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0. \end{cases} \quad (2.47)$$

A Figura 2.17, mostra o gráfico da distribuição acumulada do tipo exponencial para diferentes valores de λ .

Figura 2.17 – Distribuição acumulada Exponencial para diferentes valores de λ no intervalo $[0, 10]$



Fonte: Autor.

Observe ainda que, dada a função de distribuição de probabilidade do tipo exponencial, a probabilidade da variável aleatória estar no intervalo $[a, b]$ é dada por:

$$P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}. \quad (2.48)$$

O parâmetro λ para a função de distribuição exponencial está relacionado com o valor esperado da distribuição exponencial.

Em geral, para variáveis aleatórias contínuas, a esperança matemática é definida como:

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

No caso da distribuição exponencial, tem-se que:

$$\begin{aligned} E(X) &= \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} x \cdot e^{-\lambda x} dx \\ &= \lambda \left\{ \left[\frac{x e^{-\lambda x}}{-\lambda} \right]_0^{\infty} - \int_0^{\infty} \frac{e^{-\lambda x}}{-\lambda} dx \right\} \\ &= \lambda \left\{ \left[\frac{0 - 0}{-\lambda} \right] + \frac{1}{\lambda} \left[\frac{e^{-\lambda x}}{-\lambda} \right]_0^{\infty} \right\} \\ &= \lambda \left\{ \frac{1}{\lambda} \left[\frac{0 - 1}{-\lambda} \right] \right\} \\ &= \frac{1}{\lambda}. \end{aligned} \tag{2.49}$$

Exemplo 2.8. *O tempo médio de vida útil de determinada marca de celular é 2 anos. Qual será a probabilidade que ele dure mais que um ano, mas menos do que dois anos?*

Supondo que a vida útil do celular tenha distribuição exponencial, usando o fato de que a média de duração é 2, ou seja, $E(X) = 2$ temos que $\lambda = \frac{1}{E(X)} = \frac{1}{2} = 0.5$ portanto, da Equação (2.48):

$$P(1 < x < 2) = e^{-0.5} - e^{-1} = 0.2386. \tag{2.50}$$

Assim, a probabilidade do aparelho celular durar mais que um ano e menos que dois anos é de 23.86%.

3 MODELOS DE REGRESSÃO

3.1 INTRODUÇÃO

Os modelos de regressão são utilizados para obter uma função capaz de relacionar os resultados encontrados e suas variáveis correspondentes. Também são utilizadas em diversas áreas do conhecimento e possibilitam a obtenção de uma previsão de valores, neste capítulo será abordada sua aplicação no mercado financeiro, mais precisamente na obtenção de uma função que forneça os preços futuros dos ativos de interesse. Este capítulo teve como referências Ruggiero e Lopes (2000), Stewart (2017), Burden e Faires (2008) e Bortolossi (2002).

3.1.1 REGRESSÃO LINEAR

Durante o período de 04/05/2020 até 02/06/2020 observou-se o preço de abertura da empresa Arezzo (ARZZ3) na Bolsa de valores (Tabela 3.1). Associando as 22 datas de acompanhamento de preço do ativo com os números 1, 2, 3, ..., 22, podemos representar estes dados como na Figura 3.1. Observe que, não há uma relação precisa entre data e preço, no entanto, podemos observar uma tendência de crescimento do preço ao longo dos dias. Podemos nos perguntar se existe uma função que melhor aproxime a relação entre preço e dia para este ativo.

Tabela 3.1 – Cotação do ativo AZZR3 no período de 04/05/2020 a 02/06/2020

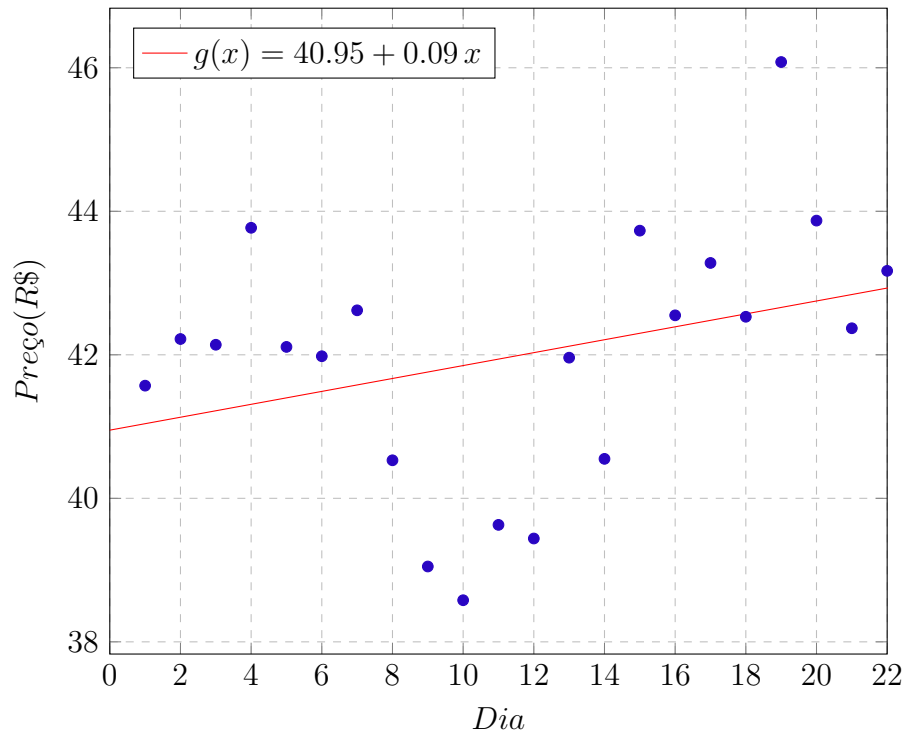
Data	Preço (R\$)	Data	Preço (R\$)
04/05/2020	41.57	19/05/2020	39.44
05/05/2020	42.22	20/05/2020	41.96
06/05/2020	42.14	21/05/2020	40.55
07/05/2020	43.77	22/05/2020	43.73
08/05/2020	42.11	25/05/2020	42.55
11/05/2020	41.98	26/05/2020	43.28
12/05/2020	42.62	27/05/2020	42.53
13/05/2020	40.53	28/05/2020	46.08
14/05/2020	39.05	29/05/2020	43.87
15/05/2020	38.58	01/06/2020	42.37
18/05/2020	39.63	02/06/2020	43.17

Fonte: Autor.

Considere inicialmente uma função linear $g(x) = ax + b$ que melhor se ajusta aos dados da Tabela 3.1, como ilustrado na Figura 3.1. Como então determinar os coeficientes a e b ?

Observe que, nos dias 13, 16, 18 e 22, os pontos correspondentes estão mais próximos da reta dada pela função linear $g(x)$. Podemos dizer, portanto, que nestes dias o preço y_i no dia x_i ,

Figura 3.1 – Ajuste linear para o preço do ativo ARZZ3



Fonte: Autor.

está próximo do valor de $g(x_i)$, ou ainda, que o erro entre o preço do ativo e o valor de $g(x_i)$ é pequeno.

Dado um conjunto de pontos tabelados (x_i, y_i) , queremos determinar os valores de a e b de tal forma que a soma dos erros $e_i = |y_i - g(x_i)|$ seja o menor possível, ou seja, queremos resolver o seguinte problema,

$$\min \sum_{i=1}^n e_i = \min_{a,b \in \mathbf{R}} \sum_{i=1}^n |y_i - (a + bx_i)|. \quad (3.1)$$

Gostaríamos de determinar os valores de a e b de forma satisfazer (3.1). Como a função modular da Equação 3.1 não é diferenciável em todos os pontos, não é possível utilizar as ferramentas de cálculo diferencial para resolver este problema de minimização. No entanto, podemos considerar o problema de minimizar a soma dos quadrados dos erros, ou seja:

$$\min \sum_{i=1}^n e_i^2 = \min_{a,b \in \mathbf{R}} \sum_{i=1}^n (y_i - a - bx_i)^2. \quad (3.2)$$

Este problema de minimização pode ser interpretado como uma minimização da soma dos quadrados das distâncias dos pontos à reta e neste caso é possível usar as ferramentas de cálculo diferencial. Denotando por $M(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2$, o ponto de mínimo que fornece o menor erro possível deve ser um ponto crítico de $M(a, b)$, portanto deve satisfazer

$$\frac{\partial M}{\partial a} = 0, \quad \frac{\partial M}{\partial b} = 0, \quad (3.3)$$

ou seja,

$$\begin{cases} \frac{\partial M}{\partial a} = 2 \sum_{i=1}^n (y_i - a - bx_i)(-1) = 0, \\ \frac{\partial M}{\partial b} = 2 \sum_{i=1}^n (y_i - a - bx_i)(-x_i) = 0. \end{cases} \quad (3.4)$$

Ou ainda, reorganizando os termos, o sistema de equações (3.4) pode ser reescrito como

$$\begin{cases} a \sum_{i=1}^n 1 + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i, \end{cases} \quad (3.5)$$

que pode ser reescrito na forma matricial como:

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}. \quad (3.6)$$

O sistema (3.6) é usualmente denominado de *Sistema Normal*. Resolvendo o sistema em termos de a e b , obtém-se:

$$\begin{aligned} a &= \frac{n \sum x_i y_i - \sum x \sum y}{n \sum x_i^2 - (\sum x_i)^2}, \\ b &= \frac{\sum x_i \sum x_i y_i - \sum y_i \sum x_i^2}{(\sum x_i)^2 - n \sum x_i^2}. \end{aligned} \quad (3.7)$$

É possível mostrar que de fato a e b , como definidos na Equação (3.7) correspondem ao ponto de mínimo da função $M(a, b)$, isto é, minimizam a soma dos quadrados dos erros. Para maiores detalhes ver Bortolossi (2002).

Dessa forma obtemos os coeficientes a e b e a função linear $g(x)$ que melhor se aproxima dos dados (x_i, y_i) com $i = 1, \dots, n$. Usualmente a função $g(x)$ é chamada de *função de regressão linear*.

No caso do ativo Arezzo, temos que, $n = 22$,

$$\sum_{i=1}^{22} 1 = 22, \quad \sum_{i=1}^{22} x_i = 253, \quad \sum_{i=1}^{22} x_i^2 = 3795, \quad \sum_{i=1}^{22} y_i = 923.73, \quad \sum_{i=1}^{22} x_i y_i = 10702.64.$$

Logo o sistema (3.6) se resume a:

$$\begin{bmatrix} 22 & 253 \\ 253 & 3795 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 923.73 \\ 10702.64 \end{bmatrix}. \quad (3.8)$$

Resolvendo o sistema, encontramos, $a = 40.95$ e $b = 0.09$, e assim a função de regressão linear $g(x) = 40.95 + 0.09 \cdot x$, que está representado na Figura 3.1.

É possível encontrar o erro desta aproximação, para isso considere a Tabela 3.2, com os valores dos preços y_i e de sua aproximação $g(x_i)$. Como o erro é dado por $E = \sum_{i=1}^{22} e_i^2$ com $e_i = |y_i - g(x_i)|$, segue que $E = 60.2962$.

Tabela 3.2 – Dados do ajuste linear $g(x) = 40.95 + 0.09x$ para o preço do ativo AZZR3

x_i	$y(x_i)$	$g(x_i)$	e_i^2	x_i	$y(x_i)$	$g(x_i)$	e_i^2
1	41.57	41.04	0.2809	12	39.44	42.03	6.7081
2	42.22	41.13	1.1881	13	41.96	42.12	0.0256
3	42.14	41.22	0.8464	14	40.55	42.21	2.7556
4	43.77	41.31	6.0516	15	43.73	42.30	2.0449
5	42.11	41.40	0.5041	16	42.55	42.39	0.0256
6	41.98	41.49	0.2401	17	43.28	42.48	0.6400
7	42.62	41.58	1.0816	18	42.53	42.57	0.0016
8	40.53	41.67	1.2996	19	46.08	42.66	11.6964
9	39.05	41.76	7.3441	20	43.87	42.75	1.2544
10	38.58	41.85	10.6929	21	42.37	42.84	0.2209
11	39.63	41.94	5.3361	22	43.17	42.93	0.0576

Fonte: Autor.

O processo é similar quando fazemos o ajuste por outras curvas que não são necessariamente a reta, para encontrar a melhor aproximação vamos sempre iniciar o processo a partir do sistema normal.

3.1.2 AJUSTES POLINOMIAIS

Para obter regressões com polinômios não lineares, ou seja, de grau $m > 1$, o método dos mínimos quadrados pode ser generalizado.

Dado um conjunto de n pontos tabelados, obtidos geralmente de uma coleta experimental, tem-se por objetivo encontrar um polinômio de ordem m com $m \leq n$, que melhor se ajusta aos dados, ou seja, encontrar $g(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m$ tal que a soma dos quadrados dos erros seja o menor possível.

Para isso devemos determinar os coeficientes $a_0, a_1, a_2, \dots, a_m$ que resolvem o seguinte problema:

$$\min_{a_1, a_2, \dots, a_m \in \mathbb{R}} \sum_{i=1}^n (y_i - a_0 - a_1x_i - a_2x_i^2 - \dots - a_mx_i^m)^2. \quad (3.9)$$

Denotando por $M(a_0, \dots, a_m) = \sum_{i=1}^n (y_i - a_0 - a_1x_i - a_2x_i^2 - \dots - a_mx_i^m)^2$ sabemos

que os pontos críticos desta função devem satisfazer o seguinte sistema de equações,

$$\begin{cases} \frac{\partial M}{\partial a_0} = 2 \sum_{i=1}^n (y_n - a_0 - a_1 x_i - a_2 x_i^2 - \dots - a_m x_i^m)(-1) = 0 \\ \frac{\partial M}{\partial a_1} = 2 \sum_{i=1}^n (y_n - a_0 - a_1 x_i - a_2 x_i^2 - \dots - a_m x_i^m)(-x_i) = 0 \\ \vdots \\ \frac{\partial M}{\partial a_m} = 2 \sum_{i=1}^n (y_n - a_0 - a_1 x_i - a_2 x_i^2 - \dots - a_m x_i^m)(-x_i^m) = 0. \end{cases} \quad (3.10)$$

O sistema de equações definido em (3.10) é equivalente a

$$\begin{cases} \frac{\partial M}{\partial a_0} = \sum_{i=1}^n (-y_i + a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_m x_i^m) = 0 \\ \frac{\partial M}{\partial a_1} = \sum_{i=1}^n (-y_i x_i + a_0 x_i + a_1 x_i^2 + a_2 x_i^3 + \dots + a_m x_i^{m+1}) = 0 \\ \vdots \\ \frac{\partial M}{\partial a_m} = \sum_{i=1}^n (-y_n + a_0 x_i^m + a_1 x_i^{m+1} + a_2 x_i^{m+2} + \dots + a_m x_i^{2m}) = 0. \end{cases} \quad (3.11)$$

Reorganizando os termos, o sistema (3.11) pode ser escrito como:

$$\begin{cases} a_0 \sum_{i=1}^n 1 + a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 + \dots + a_m \sum_{i=1}^n x_i^m = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 + \dots + a_m \sum_{i=1}^n x_i^{m+1} = \sum_{i=1}^n y_i x_i \\ \vdots \\ a_0 \sum_{i=1}^n x_i^m + a_1 \sum_{i=1}^n x_i^{m+1} + \dots + a_m \sum_{i=1}^n x_i^{2m} = \sum_{i=1}^n y_i x_i^m \end{cases} \quad (3.12)$$

e portanto na forma matricial:

$$\begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \dots & \sum_{i=1}^n x_i^m \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \dots & \sum_{i=1}^n x_i^{m+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^n x_i^m & \sum_{i=1}^n x_i^{m+1} & \sum_{i=1}^n x_i^{m+2} & \dots & \sum_{i=1}^n x_i^{2m} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \\ \vdots \\ \sum_{i=1}^n y_i x_i^m \end{bmatrix} \quad (3.13)$$

A seguir serão apresentados alguns modelos de regressão não lineares, para ajustar os preços do ativo ARZZ3 considerados na Tabela 3.1, comparando os erros obtidos em cada caso.

Inicialmente considere um ajuste polinomial de grau 3, isto é, $g_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$. Nesse caso o sistema matricial (3.13) é da forma,

$$\begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 & \sum_{i=1}^n x_i^5 \\ \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 & \sum_{i=1}^n x_i^5 & \sum_{i=1}^n x_i^6 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \\ \sum_{i=1}^n y_i x_i^2 \\ \sum_{i=1}^n y_i x_i^3 \end{bmatrix}, \quad (3.14)$$

cujas entradas são:

$$n = 22, \quad \sum_{i=1}^{22} x_i = 253, \quad \sum_{i=1}^{22} x_i^2 = 3795, \quad \sum_{i=1}^{22} x_i^3 = 64009, \quad \sum_{i=1}^{22} x_i^4 = 1151403,$$

$$\sum_{i=1}^{22} x_i^5 = 21571033, \quad \sum_{i=1}^{22} x_i^6 = 415601835, \quad \sum_{i=1}^{22} y_i = 923.73, \quad \sum_{i=1}^{22} x_i y_i = 10702.64,$$

$$\sum_{i=1}^{22} x_i^2 y_i = 161853.06, \quad \sum_{i=1}^{22} y_i x_i^3 = 2747051.06.$$

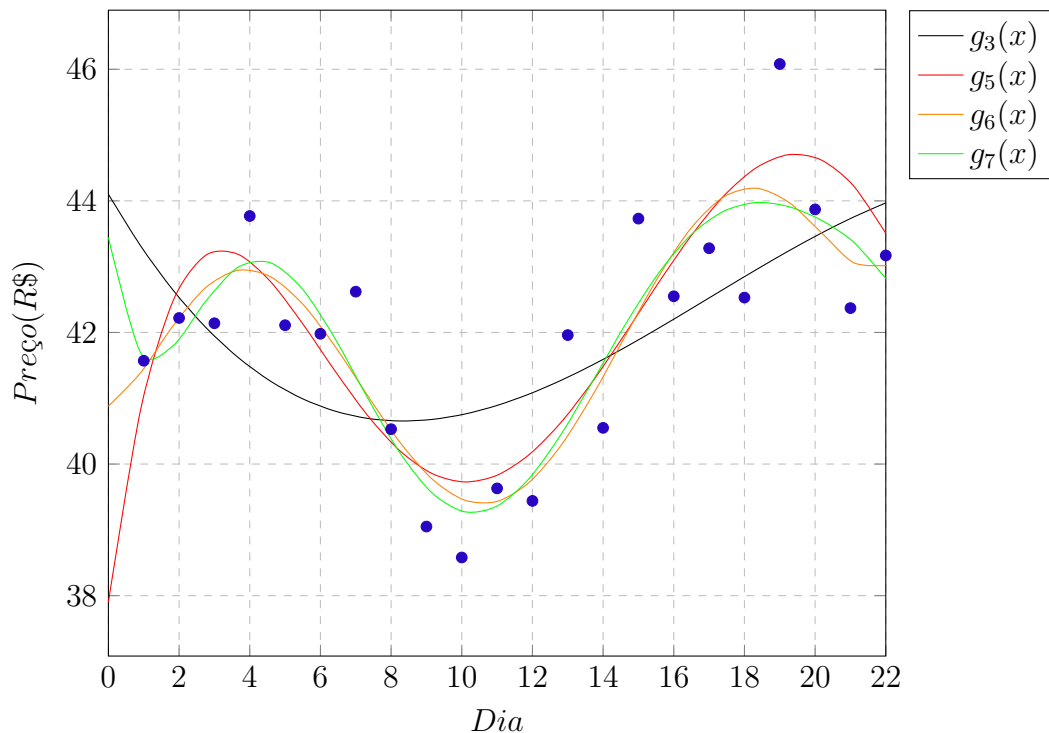
Resolvendo o sistema, neste caso utilizando a ferramenta online *Symbolab*¹, obtêm-se os seguintes coeficientes,

$$a_0 = 44.13930, \quad a_1 = -0.92572, \quad a_2 = 0.07354, \quad a_3 = -0.00144,$$

e, portanto o ajuste polinomial é dado pela função, $g_3(x) = 44.13 - 0.92x + 0.07x^2 - 0.00144x^3$, ilustrado na Figura 3.2, com erro $E = \sum_{i=1}^{22} (y_i - g_3(x_i))^2 = 42.4046$.

De forma análoga, é possível obter ajustes polinomiais de outras ordens. Na Figura 3.2 ilustramos ajustes polinomiais de ordem $g_i(x)$ de ordem i com $i = 5, 6, 7$.

Figura 3.2 – Ajustes de diferentes ordens polinomiais para o preço do ativo ARZZ3



Fonte: Autor.

Existem também outros tipos de ajustes que não são os polinomiais, a seguir apresentamos alguns desses tipos.

¹ <<https://pt.symbolab.com>>

3.1.3 REGRESSÃO EXPONENCIAL

Considere um ajuste por uma função exponencial na forma $g(x) = a \cdot b^x$. Para obter os coeficientes a e b usaremos as ferramentas já desenvolvidas de regressão linear. Observe que, se $y = a \cdot b^x$, com $y > 0$, tem-se que

$$\begin{aligned} y &= a \cdot b^x \\ \ln(y) &= \ln(a \cdot b^x) \\ \ln(y) &= \ln(a) + \ln(b^x) \\ \ln(y) &= \ln(a) + x \cdot \ln(b) \\ \bar{y} &= \bar{A} + \bar{B} \cdot x, \end{aligned} \tag{3.15}$$

em que $\bar{A} = \ln(a)$, $\bar{B} = \ln(b)$ e $\bar{y} = \ln(y)$.

Assim, dado um conjunto de pontos (x_i, y_i) , com $y_i > 0$, transformando estes dados na forma $(x_i, \ln(y_i))$, é possível encontrar os coeficientes \bar{A} e \bar{B} de um ajuste linear dos dados modificados e conseqüentemente, obter os coeficientes $a = e^{\bar{A}}$ e $b = e^{\bar{B}}$ correspondentes a um ajuste exponencial dos dados originais.

Considere, portanto, um ajuste da forma $g(x) = ab^x$ para o ativo ARZZ3, cujos preços estão tabelados em 3.1.

O sistema normal correspondente ao sistema (3.6), para os dados modificados é dado por

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \cdot \begin{bmatrix} \bar{A} \\ \bar{B} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n \ln(y_i) \\ \sum_{i=1}^n x_i \ln(y_i) \end{bmatrix}, \tag{3.16}$$

cujas entradas são

$$n = 22, \quad \sum_{i=1}^{22} x_i = 253, \quad \sum_{i=1}^{22} x_i^2 = 3795, \quad \sum_{i=1}^{22} \ln y_i = 82.2030, \quad \sum_{i=1}^{22} x_i \ln y_i = 947.2018,$$

e portanto o seguinte sistema matricial:

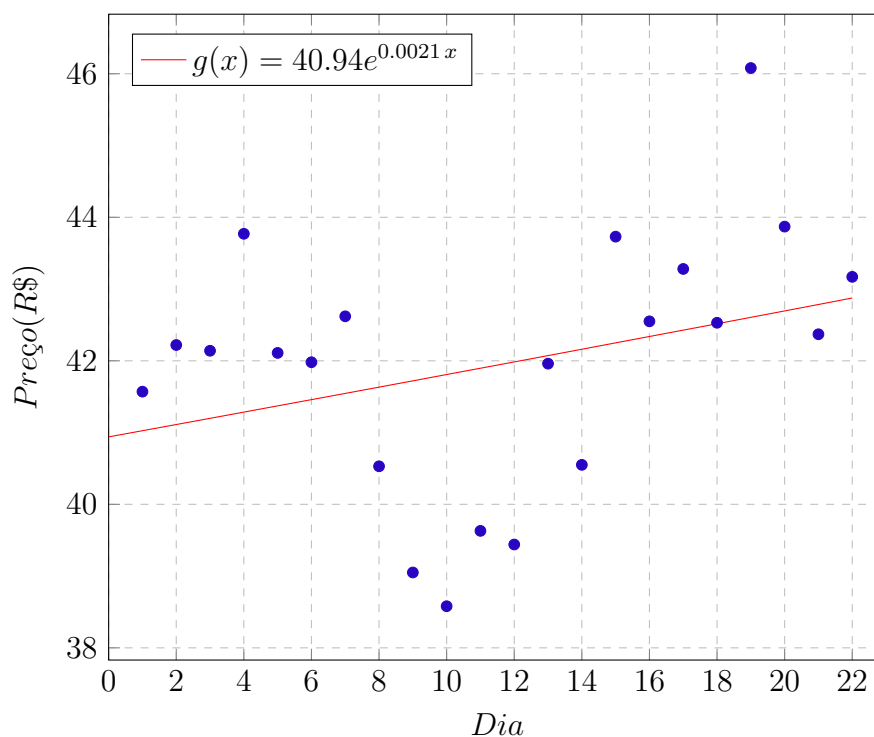
$$\begin{bmatrix} 22 & 253 \\ 253 & 3795 \end{bmatrix} \cdot \begin{bmatrix} \bar{A} \\ \bar{B} \end{bmatrix} = \begin{bmatrix} 82.2030 \\ 947.2018 \end{bmatrix}.$$

Resolvendo o sistema temos $\bar{A} = 3.7122$ e $\bar{B} = 0.0021$. Logo $a = e^{\bar{A}} \approx 40.94$ e $b = e^{\bar{B}}$ e, portanto, a função de regressão exponencial é $g(x) = 40.94 \cdot e^{0.0021x}$ e ilustrado na Figura 3.3. O erro correspondente a este ajuste é $E = \sum_{i=1}^{22} (y_i - g(x_i))^2 = 60.19$.

Por fim, reorganizamos os modelos de ajustes apresentados ao longo desta sessão na Tabela 3.3 bem como seus respectivos erros.

Um dos objetivos do estudo de regressão apresentado neste capítulo é a previsibilidade dos valores obtidos em períodos futuros de ativos, através das funções ajustadas $g(x)$. Dessa

Figura 3.3 – Ajuste do tipo exponencial com $g(x) = 40.94 \cdot e^{0.0021x}$ para o preço do ativo ARZZ3



Fonte: Autor.

Tabela 3.3 – Erros dos ajustes

Tipo de ajuste	Erros
Linear	60.29
3° Grau	42.40
5° Grau	22.50
6° Grau	17.90
7° Grau	17.18
Exponencial	60.19

Fonte: Autor.

forma, utilizaremos as funções de ajustes encontradas ao longo deste capítulo para *prever* o preço deste ativo, bem como comparar com seu real valor.

Faremos aqui uma determinação de valores para 8 dias após o período de observação do ativo ARZZ3 de acordo com cada função $g(x)$ obtida anteriormente. Os dados obtidos podem ser visualizados na Tabela 3.4

Tabela 3.4 – Tabela comparativa da previsão de preço do ativo ARZZ3

Data	x	Exponencial	7° grau	6° grau	5° grau	3° grau	Linear	ARZZ3
03/06/2020	23	42.98	41.65	44.21	41.22	44.22	43.02	47.85
04/06/2020	24	43.07	39.12	47.92	39.46	44.37	43.11	47.04
05/06/2020	25	43.16	33.79	55.86	37.36	44.45	43.20	46.76
08/06/2020	26	43.25	23.28	70.31	35.04	44.46	43.29	46.64
09/06/2020	27	43.34	3.89	94.25	32.64	44.40	43.38	47.59
10/06/2020	28	43.44	-29.78	131.45	30.37	44.25	43.47	47.35
12/06/2020	29	43.53	-85.37	186.54	28.49	44.01	43.56	44.58
15/06/2020	30	43.62	-173.29	265.23	27.32	43.66	43.65	42.97

Fonte: Autor.

Observe que, conforme a Tabela 3.4 os menores erros são dados por ajustes com funções polinomiais de grau 6 e 7. No entanto, quando utilizados para prever o preço do ativo, as funções de grau 3, linear e exponencial tiveram melhor desempenho. Por exemplo, no dia 25 o preço do ativo era de R\$ 46.76 e as funções de grau 6 e 7 preveram os preços de R\$ 55.86 e R\$ 33.79 respectivamente, enquanto que as previsões com as funções de grau 3 e linear foram R\$ 44.45 e R\$ 43.20 respectivamente.

Note ainda que, o ajuste de grau 7 forneceu um preço negativo nos dias 28, 29 e 30 o que não faz sentido no contexto aplicado.

Por fim, observa-se que quando aplicado ajuste da forma exponencial ou linear, estes ajustes não permitem projetar mudanças na tendência de crescimento ou queda dos preços, pois são funções monótonas, o que pode não refletir a dinâmica dos preços das ações.

Baseado no que foi exposto, vemos que informações de dados passados não garantem previsões precisas para o preço das ações no futuro, mas que os modelos matemáticos podem ser utilizados como ferramentas de apoio na tomada de decisões no curto prazo, que podem ser aprimorados com a inclusão de mais informações.

4 INTRODUÇÃO A ANÁLISE DE SOBREVIVÊNCIA

Ao longo deste capítulo, faremos uma breve introdução à *Análise de Sobrevivência*, baseado no conceitos apresentados em Colosimo (2006).

4.1 INTRODUÇÃO

Em análise de sobrevivência o que se pretende estudar é o tempo até a ocorrência de um evento de interesse, denominado *tempo de falha* que pode ser em segundos, dias, semanas, anos, etc, dependendo do contexto. Esse tipo de processo pode aparecer em diferentes áreas do conhecimento tais como:

1. Ciências atuariais: Tempo até um usuário de seguro fazer o uso do mesmo.
2. Engenharia : Período de tempo ou ciclo até que um dispositivo deixe de funcionar.
3. Finanças: Tempo até que um cliente se torne inadimplente, opte por desbloquear um cartão, ou feche uma conta. Tempo até um ativo voltar ou atingir certo valor.
4. Saúde: Reinfecção por uma doença, tempo até um medicamento cause efeito colateral, tempo de sobrevida de pacientes em tratamento por certa doença.

Uma característica relevante da análise de sobrevivência é a possível presença de *censura*, que ocorre quando não observamos o evento de interesse no período de observação. Por exemplo na área da saúde, se o evento de interesse for a morte do paciente por determinada doença, a censura acontece se o paciente sobrevive até o final do estudo ou quando ocorre a morte do paciente por outras causas ou não é possível estabelecer o estado do paciente no final do estudo por algum motivo (mudança de endereço, desistência do paciente, etc.).

Neste capítulo utilizamos as ideias descritas acima no contexto do mercado financeiro. Para isso iremos analisar o desempenho de 80 empresas que compõem o índice BOVESPA, no período de 02/03/2020 a 05/02/2021, que corresponde ao período de operações na Bolsa pós início da pandemia de COVID-19.

O evento considerado, será o tempo (em semanas) que o ativo levou para retornar ao preço que estava na semana que se encerrou em 06/03/2020. Os tickers das empresas e preços de referência estão na Tabela 4.1, o tempo que cada uma delas demorou para atingir o valor desejado estão na Tabela 4.2 e ilustrados na Figura 4.4.

O preço de referência será o maior valor do ativo observado no período de 02/03/2020 a 06/03/2020, por exemplo, o ativo ABEV3 teve nessa semana o maior preço R\$15.55.

Tabela 4.1 – Preço Referência

Ticker	Preço	Ticker	Preço	Ticker	Preço	Ticker	Preço
ABEV3	15.55	CPFE3	37.08	GOAU4	8.73	PCAR3	75.98
AMER3	69.51	CPLE6	7.75	HAPV3	11.4	PETR3	28.86
AZUL4	46.96	CRFB3	22.97	HGTX3	23.51	PETR4	27.48
B3SA3	16.87	CSAN3	19.69	HYPE3	42.8	PRI03	9
BBAS3	47.49	CSNA3	12.69	IGTA3	53.2	QUAL3	37.62
BBDC3	22.94	CVCB3	24.61	IRBR3	30.89	RADL3	26.1
BBDC4	25.69	CYRE3	32.24	ITSA4	12.39	RAIL3	22.41
BBSE3	32.04	ECOR3	16.58	ITUB4	32.29	RENT3	51.6
BEEF3	12.52	EGIE3	50.88	JBSS3	24.49	SANB11	39.79
BIDI11	15.72	ELET3	36.5	JHSF3	7.3	SBSP3	62.86
BPAC11	17.27	ELET6	37.13	KLBN11	22.66	SULA11	51.08
BRAP4	36.09	EMBR3	17.55	LAME4	11.44	SUZB3	44.28
BRDT3	27.31	ENBR3	21.81	LCAM3	22.01	TAE11	31.31
BRFS3	28.7	ENEV3	12.4	LREN3	54.96	TIMS3	17.41
BRKM5	29.73	ENGI11	60.44	LWSA3	6.36	TOTS3	25.17
BRML3	17.34	EQTL3	26.92	MGLU3	13.7	USIM5	8.42
CCRO3	17.08	EZTC3	51.96	MRF3	13.24	VALE3	49.09
CIEL3	7.47	FLRY3	32.15	MRVE3	19.49	VIVT3	53.78
CMIG4	12.99	GGBR4	18.54	MULT3	33	VVAR3	15.13
COGN3	10.79	GNDI3	66.17	NTCO3	45.96	WEGE3	24.9

Fonte: Autor.

A partir dos dados de referência contamos quantas semanas cada ativo levou para retornar o seu preço de referência . Por exemplo, para o ativo ABEV3, cujo preço referência era de R\$15.55, este preço foi atingido novamente 21 semanas depois. De forma similar, foi realizado para as 80 empresas. Os dados estão organizados na Tabela 4.2 onde os espaços em branco indicam que o preço de referência não foi atingido no período observado.

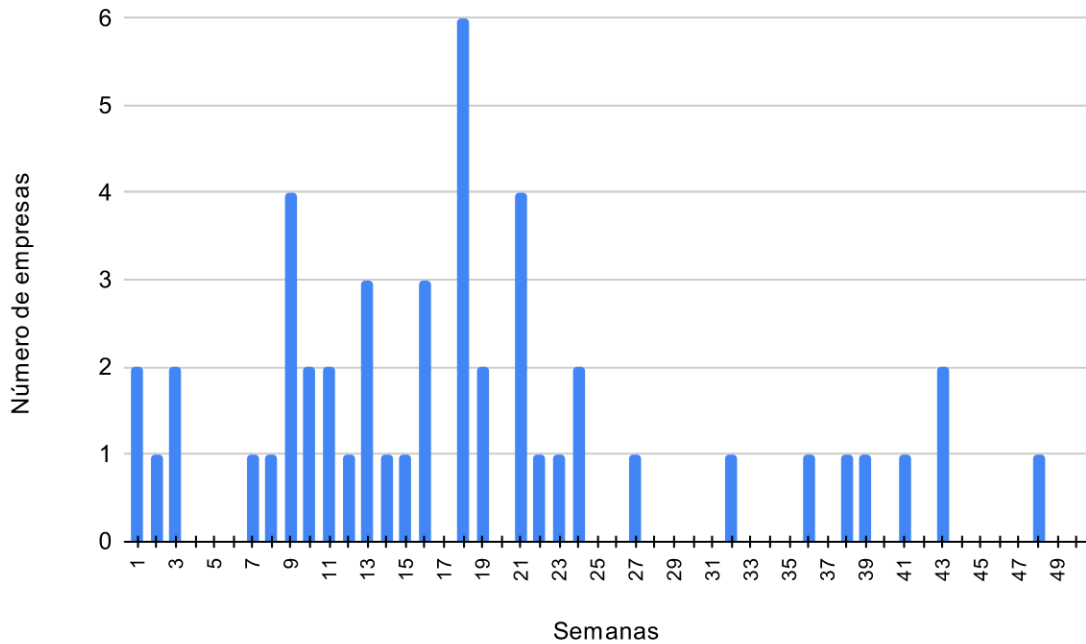
Tabela 4.2 – Tempo de retorno ao preço de referência

Ticker	Tempo	Ticker	Tempo	Ticker	Tempo	Ticker	Tempo
ABEV3	21	CPLE6		HYPE3		PETR3	41
AZUL4		CSAN3	18	IGTA3		PETR4	39
AMER3	7	CPFE3		GNDI3	16	BRDT3	
B3SA3	13	CVCB3	14	IRBR3		PRI03	23
BIDI11	18	CYRE3		ITSA4		QUAL3	
BBSE3		ECOR3		ITUB4	43	RADL3	32
BRML3		ELET3	18	JBSS3	3	RAIL3	12
BBDC3		ELET6	18	JHSF3	16	SBSP3	19
BBDC4		EMBR3		KLBN11	10	SANB11	38
BRAP4	11	ENBR3		RENT3	21	CSNA3	21
BBAS3		ENGI11		LCAM3	27	SULA11	
BRKM5	48	ENEV3	19	LWSA3	11	SUZB3	9
BRFS3		EGIE3		LAME4	9	TAE11	36
BPAC11	15	EQTL3		LREN3		VIVT3	1
CRFB3	2	EZTC3		MGLU3	9	TIMS3	20
CCRO3		FLRY3		MRF3	9	TOTS3	18
CMIG4	43	GGBR4	22	BEEF3	8	USIM5	24
HGTX3		GOAU4	24	MRVE3	18	VALE3	10
CIEL3		NTCO3	21	MULT3		VVAR3	13
COGN3		HAPV3	13	PCAR3	3	WEGE3	16

Fonte: Autor.

A partir das informações da Tabela 4.2 construímos um gráfico de colunas, representado na Figura 4.1 em que a primeira coluna representa o número de empresas que levaram 1 semana para retornar ao preço de referência, a segunda coluna, o número de empresas que levaram 2 semanas para retornar ao preço de referência e assim sucessivamente.

Figura 4.1 – Número de semanas levado, por cada empresa listada na Tabela 4.2 para retornar ao valor de referência



Fonte: Autor.

Ao considerar o evento *retorno ao preço de referência*, gostaríamos de estimar qual a probabilidade do evento acontecer após t semanas. Para isso introduziremos alguns conceitos gerais sobre análise de sobrevivência.

4.2 CONCEITOS BÁSICOS

Seja T uma variável aleatória que representa o tempo até que aconteça um determinado evento de interesse, que chamaremos de *falha*, durante um experimento ou observação.

Definimos a *função de sobrevivência* como a probabilidade de o experimento ou observação não falhar ou sobreviver até determinado tempo t , que pode ser escrito como:

$$S(t) = P(T \geq t). \quad (4.1)$$

Portanto, a função de distribuição acumulada, definida na Equação (2.45), pode ser escrita como:

$$F(t) = P(T \leq t) = 1 - P(T \geq t) = 1 - S(t), \quad (4.2)$$

que representa a probabilidade do experimento ou observação não sobreviver, ou falhar até o tempo t .

Introduziremos a seguir o conceito de *função de risco*.

Considere inicialmente que desejamos determinar a probabilidade de que ocorra a falha no intervalo $[t_1, t_2]$, dado que não aconteceu antes de t_1 .

Considerando:

A : a falha ocorre entre $[t_1, t_2]$,

B : não houve falha antes de t_1 ,

temos que pela Equação (2.10), a probabilidade de ocorrer A dado que ocorreu B , ou seja $P(A|B)$, é dada por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (4.3)$$

em que:

i) $P(B) = P(T \geq t_1) = S(t_1)$,

ii) $P(A \cap B) = P(t_1 \leq T \leq t_2) = F(t_2) - F(t_1) = S(t_1) - S(t_2)$.

Dessa forma, a probabilidade de ocorrer falha no intervalo $[t_1, t_2]$, dado que não aconteceu falha antes de t_1 pode ser expresso por,

$$\frac{S(t_1) - S(t_2)}{S(t_1)}. \quad (4.4)$$

Observe que esta probabilidade pode ser interpretada como uma medida do risco de falha até o instante t_2 , a partir do instante t_1 . Se considerarmos a razão entre esta probabilidade e o comprimento do intervalo $[t_1, t_2]$, o que temos é a taxa de variação da probabilidade de falha, a partir de t_1 , no intervalo $[t_1, t_2]$.

Com base nisso, dado $\Delta t = t_2 - t_1$, definimos a *função de risco* em t_1 por,

$$h(t_1) = \lim_{\Delta t \rightarrow 0} \frac{S(t_1) - S(\Delta t + t_1)}{\Delta t S(t_1)} \quad (4.5)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{F(t_1 + \Delta t) - F(t_1)}{\Delta t S(t_1)} \quad (4.6)$$

$$= \frac{1}{S(t_1)} \lim_{\Delta t \rightarrow 0} \frac{F(t_1 + \Delta t) - F(t_1)}{\Delta t}. \quad (4.7)$$

Como $\lim_{\Delta t \rightarrow 0} \frac{F(t_1 + \Delta t) - F(t_1)}{\Delta t} = F'(t_1) = f(t_1)$, segue que a função de risco no instante t pode ser escrita como,

$$h(t) = \frac{f(t)}{S(t)}, \quad (4.8)$$

que representa a taxa de variação instantânea da probabilidade de falha.

Exemplo 4.1. *Seja T uma variável aleatória com função de distribuição de probabilidade exponencial $f(x) = \lambda e^{-\lambda x}$. Neste caso, de acordo com o descrito na Seção 2.5.1 temos,*

$$i) F(t) = \int_0^t \lambda e^{-\lambda u} du = 1 - e^{-\lambda t},$$

$$ii) S(t) = 1 - F(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t},$$

$$iii) h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda,$$

ou seja, para a distribuição exponencial o risco é constante, não depende de t e é igual o inverso do valor esperado da distribuição.

Note que a função de risco (4.8) também pode ser considerada para distribuições discretas, mesmo que não seja possível encontrar expressões fechadas para $F(t)$, $S(t)$ ou $h(t)$, como ilustrado no próximo exemplo.

Exemplo 4.2. *Determinar a função de risco para uma variável aleatória discreta com distribuição de Poisson $f(t) = \frac{\lambda^t}{t!} e^{-\lambda}$ e média $\lambda = 5$.*

Neste caso, de acordo com a Equação (2.40) temos que,

$$i) F(t) = P(X = 0) + P(X = 1) + \dots + P(X = t) = \frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \dots + \frac{\lambda^k}{k!} e^{-\lambda},$$

$$iii) S(t) = 1 - \left[\frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \dots + \frac{\lambda^t}{t!} e^{-\lambda} \right],$$

$$iv) h(t) = \frac{f(t)}{S(t)} = \frac{\frac{\lambda^t}{t!} e^{-\lambda}}{1 - \left[\frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \dots + \frac{\lambda^t}{t!} e^{-\lambda} \right]}.$$

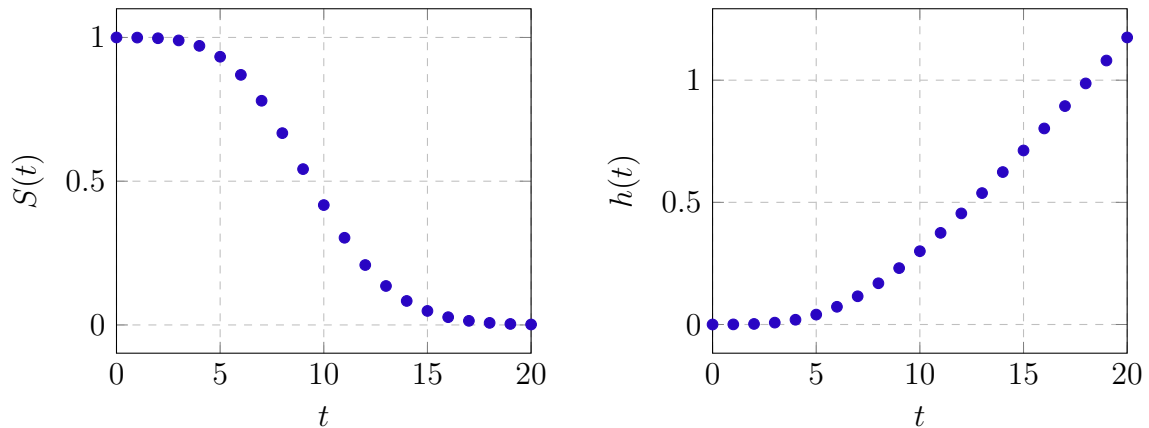
Dessa forma a probabilidade de que a variável seja maior do que 3 será de:

$$S(3) = 1 - \left[\frac{5^0}{0!} e^{-5} + \frac{5^1}{1!} e^{-5} + \frac{5^2}{2!} e^{-5} + \frac{5^3}{3!} e^{-5} \right] = 0.7349,$$

e a função de risco em $t = 3$ é dada por

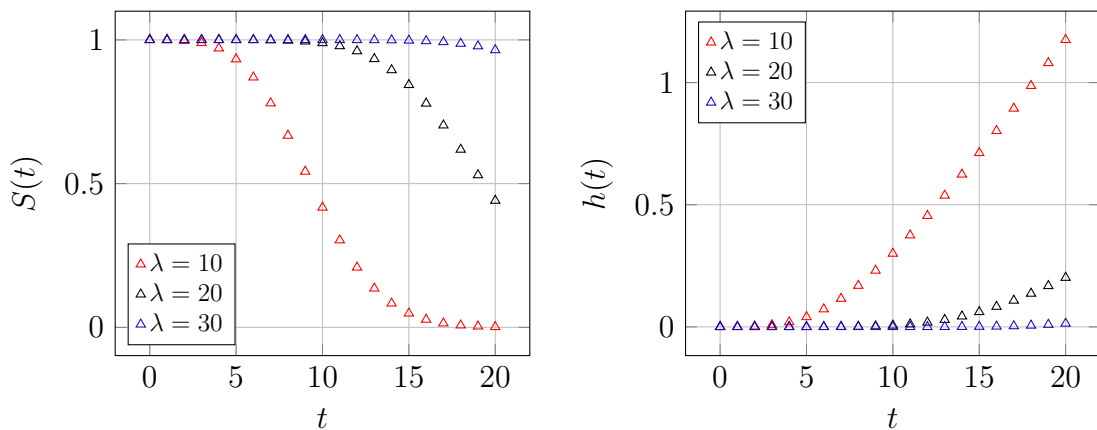
$$h(3) = \frac{f(3)}{S(3)} = \frac{0.1403}{0.7347} = 0.1909.$$

Figura 4.2 – Função de sobrevivência $S(t)$ (a esquerda) e função de risco $h(t)$ (a direita) para $\lambda = 5$



Fonte: Autor.

Figura 4.3 – Função de sobrevivência $S(t)$ (esquerda) e função de risco $h(t)$ (direita) para diferentes valores λ



Fonte: Autor.

A Figura 4.2 ilustra as funções de sobrevivência $S(t)$ e de risco $h(t)$ para diferentes valores de t e média $\lambda = 5$.

A Figura 4.3 ilustra as funções de sobrevivência $S(t)$ e função de risco $h(t)$ para diferentes valores de λ .

Exemplo 4.3. Observe na Tabela 4.2 que das empresas acompanhadas durante o período, 49 obtiveram retorno ao preço de referência, com tempo médio de volta ao preço, de 18.38 semanas. Supondo que a distribuição do tempo de retorno ao preço de referência seja do tipo Poisson, com $\lambda = 18.38$ determinar a função de risco $h(t)$ para $t = 20$.

De acordo com a Equação (2.40) temos que,

$$S(20) = 1 - \left[\frac{18.38^0}{0!} e^{-18.38} + \frac{18.38^1}{1!} e^{-18.38} + \dots + \frac{18.38^{20}}{20!} e^{-18.38} \right] = 0.3.$$

Como $f(20) = 0.082$, segue que $h(20) = \frac{f(20)}{S(20)} = 0.276$, ou seja, probabilidade de

uma determinada empresa retornar ao preço na semana 20, dado que não retornou antes dessa semana, será de 27.6%.

Da mesma forma é possível obter $h(21) = 0.319$, assim a probabilidade de retorno ao preço aumentou entre a semana 20 e a semana 21. Esta situação é observada, visto que assumimos que a distribuição de probabilidade é do tipo Poisson, que não é necessariamente uma hipótese realista.

Observe que, para calcular tanto a função de risco quanto a função de sobrevivência, para distribuições do tipo exponencial e Poisson só foi possível, pois conhecemos $f(t)$ explicitamente. No entanto, em muitos casos $f(t)$ é desconhecida e para estimar tais funções, recorremos a alguns estimadores. Na próxima seção abordaremos um estimador da função de sobrevivência, amplamente utilizado que é, o estimador de *Kaplan-Meier*.

4.2.1 ESTIMADOR DE KAPLAN-MEIER

Quando não conhecemos a função de probabilidade associada aos dados obtidos, podemos fazer uma estimativa para distribuição de probabilidade e para a função de sobrevivência. Um modelo amplamente utilizado é o modelo de *Kaplan-Meier* que permite obter uma estimativa da função de sobrevivência, para uma variável aleatória contínua.

Considere inicialmente que o tempo de observação é dividido em subintervalos $(0, t_1)$, (t_1, t_2) , \dots , (t_{i-1}, t_i) . Denote por d_i , o número de indivíduos que falharam em cada intervalo e n_i o número de indivíduos em observação até este período.

Por exemplo, na situação do tempo de retorno ao preço referência considerada no início do capítulo, o tempo de observação é de 48 semanas divididas em períodos de 1, 2 e 3 semanas. Ainda neste caso, d_i é o número de empresas que retornaram ao valor de referência no momento i e n_i o número de empresas que não retornaram ao preço referência até esse período. Estas informações estão resumidas na Tabela 4.3.

Em cada intervalo i , a probabilidade de ocorrer a falha pode ser estimada por $\hat{q}_i = \frac{d_i}{n_i}$, logo a probabilidade de não ocorrer a falha no intervalo é $1 - \hat{q}_i$.

Note que, para sobreviver até ao intervalo i , o indivíduo ou experimento precisa sobreviver aos períodos anteriores. Sendo assim, a probabilidade de sobreviver até o intervalo i é dada pelo produto das probabilidades de sobreviver aos intervalos $(0, t_1)$, (t_1, t_2) , \dots , (t_{i-1}, t_i) . Baseado nisso, definimos o estimador de *Kaplan-Meier* para a função de sobrevivência por,

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i} \right). \quad (4.9)$$

A seguir ilustramos o estimador de Kaplan-Meier para um exemplo hipotético de estudo de tratamento de câncer em um grupo de ratos ¹.

¹ <https://www.statsdirect.com/help/Default.htm#survival_analysis/kaplan_meier.htm>

Tabela 4.3 – Falhas (retorno ao preço referência) por período, para os dados da Tabela 4.2

Data (t_i)	Período	Empresas (n_i)	Falhas (d_i)
06/03/2020	0-1	80	0
20/03/2020	1-3	80	2
03/04/2020	3-5	78	2
17/04/2020	5-7	76	0
30/04/2020	7-9	76	2
15/05/2020	9-11	74	6
29/05/2020	11-13	68	3
12/06/2020	13-15	65	4
26/06/2020	15-17	61	4
10/07/2020	17-19	57	6
24/07/2020	19-21	51	3
07/08/2020	21-23	48	5
21/08/2020	23-25	43	3
04/09/2020	25-27	40	0
18/09/2020	27-29	40	1
02/10/2020	29-31	39	0
16/10/2020	31-33	39	1
30/10/2020	33-35	38	0
13/11/2020	35-37	38	1
27/11/2020	37-39	37	1
11/12/2020	39-41	36	1
23/12/2020	41-43	35	1
08/01/2021	43-45	34	3
22/01/2021	45-47	31	0
05/02/2021	47-48	31	1

Fonte: Autor.

Exemplo 4.4. *Suponha que a morte por câncer, após a exposição a um determinado agente cancerígeno tenha sido medida em um grupo de ratos. O tempo desde o pré-tratamento até a morte é registrado. Se um rato ainda estava vivo no final do experimento ou morreu por uma causa diferente, essa observação é considerada censura e recebe o valor 1 na coluna censura, caso contrário recebe o valor 0. Os dados estão apresentados na Tabela 4.4. Estime a função de sobrevivência utilizando o estimador de Kaplan-Maier.*

Da Equação (4.9), temos

$$\hat{S}(t_1 = 142) = \left(1 - \frac{d_1}{n_1}\right) = \left(1 - \frac{1}{22}\right) = 0.9545,$$

$$\hat{S}(t_2 = 157) = \left(1 - \frac{d_1}{n_1}\right) \left(1 - \frac{d_2}{n_2}\right) = \left(1 - \frac{1}{21}\right) \cdot 0.9545 = 0.9090,$$

$$\hat{S}(t_3 = 198) = \left(1 - \frac{1}{19}\right) \cdot 0.8636 = 0.8181.$$

Para $t_4 = 204$, como não houve nenhuma morte do grupo de risco, $\hat{S}(t)$ se mantém igual ao período anterior, ou seja,

$$\hat{S}(t_4 = 204) = \prod_{t_3 \leq t} \left(1 - \frac{1}{19}\right) \cdot 0.8636 = 0.8181.$$

Tabela 4.4 – Dados referentes ao estudo de tratamento de câncer num grupo de ratos

t_i	n_i	d_i	<i>censura</i>
142	22	1	0
157	21	1	0
163	20	1	0
198	19	1	0
204	18	0	1
205	17	1	0
232	16	3	0
233	13	4	0
239	9	1	0
240	8	1	0
261	7	1	0
280	6	2	0
295	4	2	0
323	2	1	0
344	1	0	1

Fonte: Autor.

Para $t_5 = 205$ o número de indivíduos em observação é reduzido pois ocorreu uma censura no período anterior, portanto

$$\hat{S}(t_5 = 205) = \prod_{t_i \leq t} \left(1 - \frac{1}{17}\right) \cdot 0.8181 = 0.77,$$

e assim sucessivamente estimamos a função de sobrevivência para os outros intervalos.

Observe que, da Equação (4.2) é possível obter uma estimativa para a probabilidade acumulada considerando $\hat{F}(t) = 1 - \hat{S}(t)$. Estes dados estão resumidos na Tabela 4.5 e ilustrados nas Figuras 4.5 e 4.6.

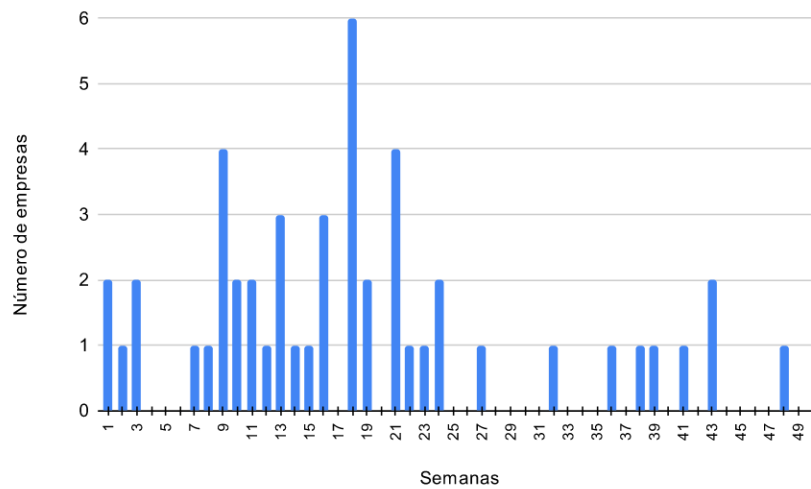
Tabela 4.5 – Dados de análise de sobrevivência para o estudo de câncer em um grupo de ratos

t_i	n_i	d_i	<i>censura</i>	$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$	$\hat{F}(t) = 1 - \hat{S}(t)$
142	22	1	0	0.9546	0.0455
157	21	1	0	0.9090	0.0910
163	20	1	0	0.8636	0.1364
198	19	1	0	0.8182	0.1818
204	18	0	1	0.8182	0.1818
205	17	1	0	0.7701	0.2300
232	16	3	0	0.6257	0.3743
233	13	4	0	0.4332	0.5669
239	9	1	0	0.3850	0.6150
240	8	1	0	0.3369	0.6631
261	7	1	0	0.2888	0.7112
280	6	2	0	0.1925	0.8075
295	4	2	0	0.0963	0.9037
323	2	1	0	0.0481	0.9519
344	1	0	1	0.0481	0.9519

Fonte: Autor.

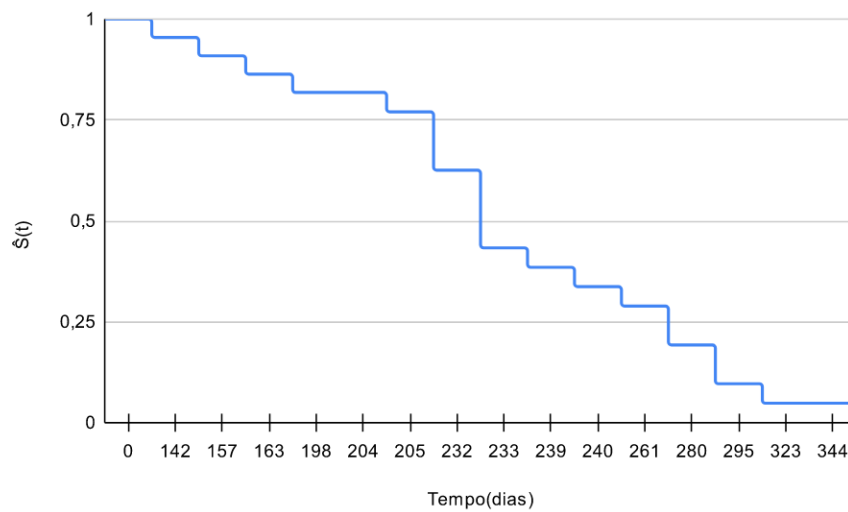
Note que, antes da primeira observação do evento, não houve nenhuma falha, ou seja, $S(t) = 1$ e $F(t) = 0$.

Figura 4.4 – Tempo de retorno (em semanas) ao valor de referência, por cada empresa listada na Tabela 4.2



Fonte: Autor.

Figura 4.5 – Estimativa de Kaplan-Meier para função de sobrevivência



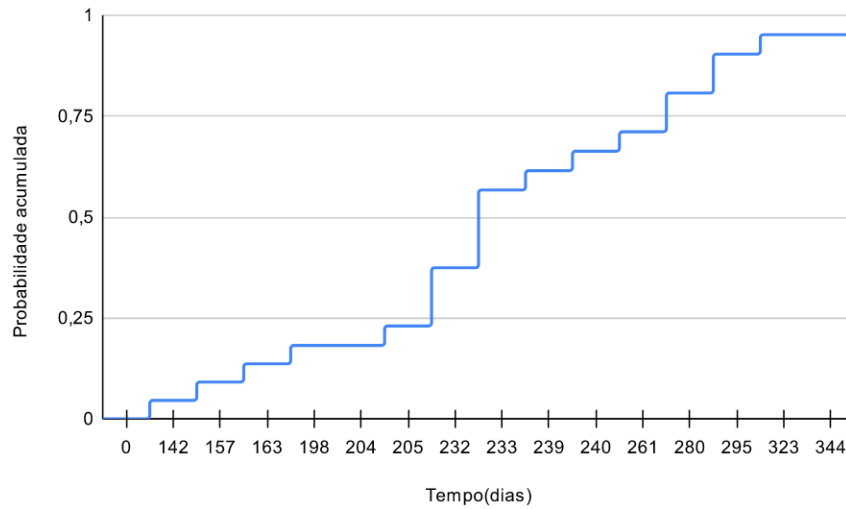
Fonte: Autor.

Observe que, a partir de 233 dias a probabilidade dos ratos sobreviverem está abaixo de 50%, além disso, essa probabilidade diminuiu com o decorrer do tempo. Em contrapartida, a probabilidade dos ratos morrerem no decorrer do tempo aumenta, como observado pelo gráfico da função $\hat{F}(t)$ na Figura 4.6.

Note que, próximo do dia 233 ocorreu um aumento considerável na probabilidade dos indivíduos morrerem. Baseado nisso podemos nos questionar se é possível estimar o risco dos indivíduos morrerem em determinado período.

Da Equação (4.8), temos que $h(t) = \frac{f(t)}{S(t)}$. No entanto em casos como do Exemplo 4.4 e do desempenho das empresas na bolsa de valores, não existe uma função de densidade de

Figura 4.6 – Função de probabilidade acumulada $\hat{F}(t)$



Fonte: Autor.

probabilidade $f(t)$ pré determinada. Desta forma, em princípio não podemos determinar $h(t)$. A seguir veremos que é possível estimar uma função do risco acumulado.

Da Equação (2.45) a função de distribuição de probabilidade acumulada é definida por $F(t) = \int_0^t f(u) du$, logo

$$\begin{aligned} f(t) &= \frac{d}{dt} F(t) \\ &= \frac{d}{dt} [1 - S(t)] \\ &= -\frac{d}{dt} S(t). \end{aligned} \quad (4.10)$$

Substituindo (4.10) em (4.8) temos,

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{-\frac{d}{dt} S(t)}{S(t)} \\ &= -\frac{d}{dt} \ln[S(t)]. \end{aligned} \quad (4.11)$$

Integrando ambos os lados de (4.11) obtemos,

$$\int_0^t h(u) du = -\ln(S(t)) + \ln(S(0)) = -\ln(S(t)), \quad (4.12)$$

pois, $S(0)$ diz respeito a probabilidade de sobreviver no instante $t = 0$, ou seja, sobreviver no início de observação do evento, logo $S(0) = 1$ e portanto $\ln[S(0)] = 0$.

Da Equação (4.12), definindo $H(t) = \int_0^t h(u) du$ como a função de risco acumulada, utilizando a estimativa de Kaplan-Meier da função de sobrevivência, obtém-se uma estimativa

da função de risco acumulada dada por

$$\hat{H}(t) = -\ln \hat{S}(t). \quad (4.13)$$

Observe que a estimativa do risco acumulado depende apenas da função estimada de sobrevivência $\hat{S}(t)$.

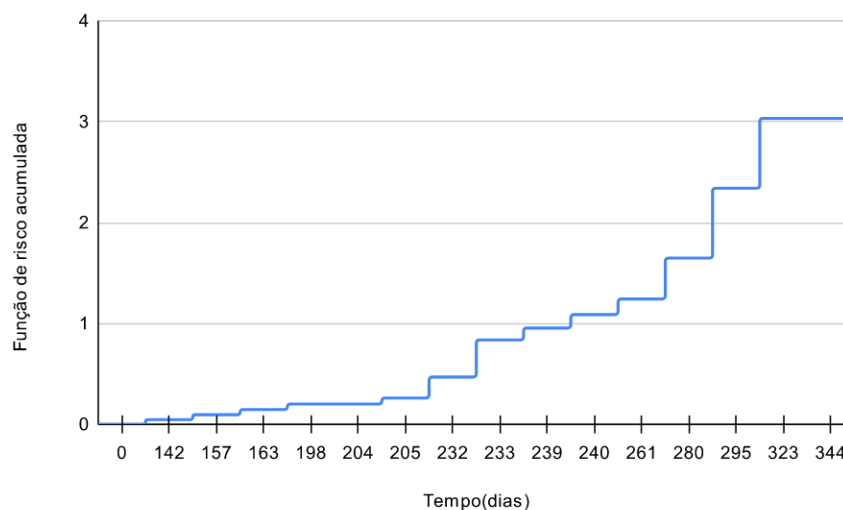
Voltando ao Exemplo 4.4, temos que, o risco de morte acumulado para os ratos até o dia 142 é dado por $\hat{H}(t_1 = 142) = -\ln 0.9546 = 0.0465$, assim como o risco acumulado de morte até o dia 157 é de $\hat{H}(t_2 = 157) = -\ln 0.9090 = 0.0954$ e assim sucessivamente obtemos os valores do risco acumulado. Os riscos acumulados obtidos podem ser observados na Tabela 4.6 e representados na Figura 4.7.

Tabela 4.6 – Funções $\hat{S}(t)$ e $\hat{H}(t)$ para o grupo de ratos

Dias	$\hat{S}(t)$	$\hat{H}(t) = -\ln \hat{S}(t)$
142	0.9546	0.0465
157	0.9090	0.0954
163	0.8636	0.1466
198	0.8182	0.2007
204	0.8182	0.2007
205	0.7701	0.2613
232	0.6257	0.4689
233	0.4332	0.8367
239	0.3850	0.9544
240	0.3369	1.0880
261	0.2888	1.2421
280	0.1925	1.6476
295	0.0963	2.3407
323	0.0481	3.0338
344	0.0481	3.0338

Fonte: Autor.

Figura 4.7 – Risco acumulado para o grupo de ratos



Fonte: Autor.

No Exemplo 4.3 supomos que o tempo de retorno ao preço de referência das empresas seguia um modelo de distribuição de Poisson e calculamos sua função de risco $h(t)$. Observe, no entanto que a melhor maneira é estimarmos esses valores utilizando o modelo de Kaplan-Meier. Retornamos então ao problema proposto no início do capítulo analisando o desempenho das empresas que compõem o IBOV.

4.3 UMA APLICAÇÃO DE ANÁLISE DE SOBREVIVÊNCIA NO MERCADO DE AÇÕES

A Tabela 4.3 apresenta o número de semanas que, 80 empresas listadas na Bolsa de Valores levaram para recuperar o valor que estavam no começo de março de 2020 (preço de referência em 4.1). Essa data foi escolhida, pois, apresentou o período de maiores baixas no mercado financeiro mundial, em decorrência da declaração da Organização Mundial da Saúde (OMS) com respeito à pandemia de COVID-19 em 11/03/2020.

No estudo do desempenho das empresas iniciaremos com a obtenção do estimador $S(t)$ de Kaplan-Meier, a análise começa com 80 empresas, durante a primeira semana não ocorreu falha, ou seja, nenhuma empresa retornou ao preço de referência, permanecendo assim a mesma quantidade de ativos para a segunda semana. No período entre a primeira e a terceira semana 3 ativos retornaram ao valor de referência, restando assim 77 empresas para o estudo no seguinte período. Prosseguindo dessa maneira até o último período analisado, que corresponde a semana 48. Estas informações estão organizadas na Tabela 4.3.

Utilizando as ferramentas desenvolvidas na Seção anterior é possível estimar a função de sobrevivência $\hat{S}(t)$ e $\hat{H}(t)$ que estão resumidos na Tabela 4.7 e ilustrados nas Figuras 4.8, 4.9 e 4.10.

Observe da Figura 4.8 que a probabilidade de um ativo se manter abaixo do preço de referência até a 19^a semana, por exemplo, é de 60% e essa probabilidade diminui ao longo das semanas. Além disso, a mediana do tempo de retorno das ações analisadas está entre 21 e 23 semanas, ou seja, o tempo estimado para que a probabilidade de retorno ao preço de referência seja 50%.

Note ainda que, entre a 7^a e a 19^a semana há um decaimento acentuado da probabilidade de manter-se abaixo do preço de referência. Comparando com a coluna d_i da Tabela 4.7 nestas semanas ocorrem uma maior incidência de retornos ao preço de referência.

Como $\hat{S}(t)$ fornece a probabilidade do preço das ações permanecerem abaixo do preço de referência, $F(t) = 1 - \hat{S}(t)$, dá a probabilidade do preço do ativo atingir os valores de referência. Com isso observamos que, até a 19^a semana a probabilidade do ativo voltar ao preço de referência é de 40% e essa probabilidade aumenta ao longo das semanas, como ilustrado na Figura 4.9.

Tabela 4.7 – Função de sobrevivência $\hat{S}(t)$ e de risco acumulado $\hat{H}(t)$ para as empresas da bolsa de valores

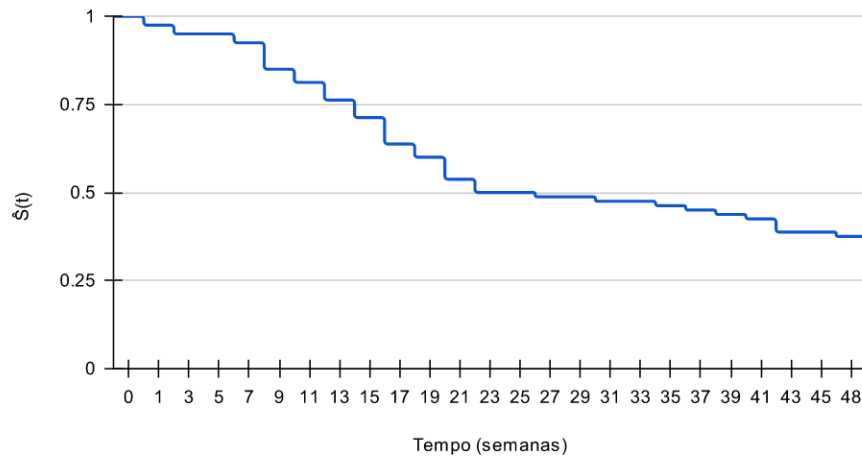
t_i	n_i	d_i	$\hat{S}(t)$	$\hat{F}(t)$	$\hat{H}(t)$
0	80	0	1.0000	0.0000	0.0000
1	80	2	0.9750	0.0250	0.0253
3	78	2	0.9500	0.0500	0.0513
5	76	0	0.9500	0.0500	0.0513
7	76	2	0.9250	0.0750	0.0780
9	74	6	0.8500	0.1500	0.1625
11	68	3	0.8125	0.1875	0.2076
13	65	4	0.7625	0.2375	0.2712
15	61	4	0.7125	0.2875	0.3390
17	57	6	0.6375	0.3625	0.4502
19	51	3	0.6000	0.4000	0.5108
21	48	5	0.5375	0.4625	0.6208
23	43	3	0.5000	0.5000	0.6931
25	40	0	0.5000	0.5000	0.6931
27	40	1	0.4875	0.5125	0.7185
29	39	0	0.4875	0.5125	0.7185
31	39	1	0.4750	0.5250	0.7444
33	38	0	0.4750	0.5250	0.7444
35	38	1	0.4625	0.5375	0.7711
37	37	1	0.4500	0.5500	0.7985
39	36	1	0.4375	0.5625	0.8267
41	35	1	0.4250	0.5750	0.8557
43	34	3	0.3875	0.6125	0.9480
45	31	0	0.3875	0.6125	0.9480
48	31	1	0.3750	0.6250	0.9808

Fonte: Autor.

Se nos perguntarmos qual o risco acumulado das empresas retornarem ao seu preço de referência até a 10^a semana, este risco é abaixo de 0.2, como observado na Figura 4.10. O risco acumulado de retorno ao preço de referência tem um aumento considerável após a 20^a semana e se mantém com pouca variação até a última semana.

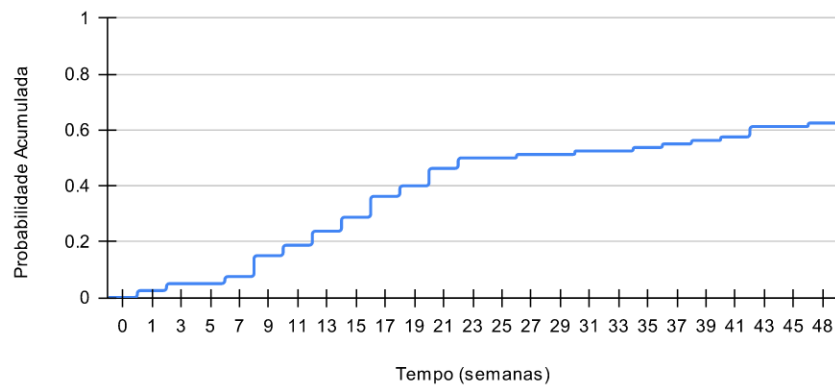
É importante observar que a função de risco $\hat{H}(t)$ não representa uma probabilidade e sim uma taxa. Representa a soma de todos os riscos em todos os tempos até o tempo t . Esta função é difícil de interpretar, mas pode ser considerada como uma medida da força acumulativa da mortalidade, ou interpretada também como o número de eventos esperados para cada indivíduo até o tempo t , caso o evento pudesse ocorrer repetidamente. Esta e outras considerações sobre a função de risco acumulado são discutidas com mais detalhe em Clark et al. (2003).

Figura 4.8 – Função de sobrevivência $\hat{S}(t)$



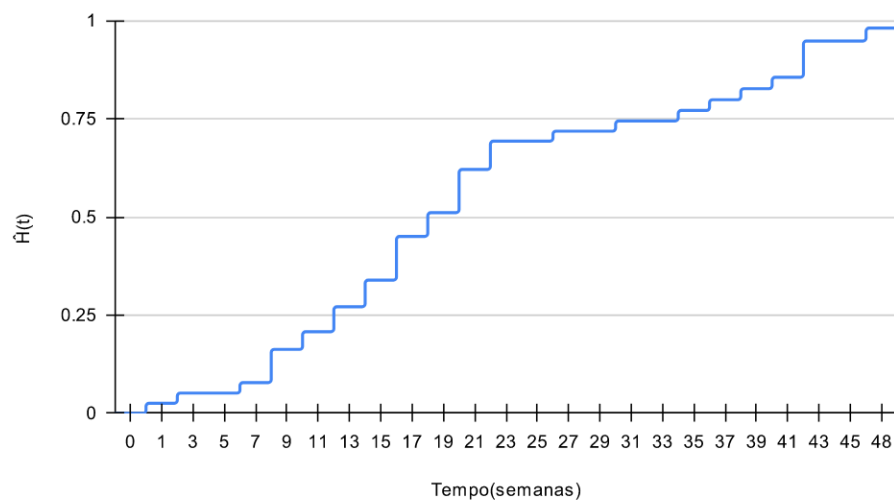
Fonte: Autor.

Figura 4.9 – Probabilidade acumulada $\hat{F}(t)$



Fonte: Autor.

Figura 4.10 – Função de risco acumulado $\hat{H}(t)$



Fonte: Autor.

5 CONCLUSÃO

Ao longo deste trabalho foram considerados conceitos de probabilidade e estatística aplicados no mercado financeiro.

Aplicamos o conhecimento de probabilidade condicional para estudar a relação entre a taxa básica Selic com dois dos principais índices do mercado financeiro, IBOV e IFIX. Por exemplo, no período de 05/02/2019 até 27/10/2021, sob condições específicas estimamos que a probabilidade do IFIX cair dado uma alta da Selic foi de 80%, e do índice IBOV cair dado uma alta da Selic foi de 60%.

Apresentamos algumas distribuições de probabilidade e em particular, consideramos um exemplo sobre o desempenho de um ativo fictício, com condições especiais, para introduzir a distribuição binomial. Observou-se que, se o ativo segue uma distribuição binomial (sucesso=aumento de R\$0.50 e fracasso=queda de R\$0.50) com a probabilidade de sucesso $p = 0.5$ com o decorrer do tempo, o mais provável é que o investimento permaneça próximo ao valor do investimento inicial.

Estudamos alguns modelos de regressão (polinomial e exponencial) e utilizamos para fazer uma projeção dos preços deste ativo, com base nos preços de abertura, no período de 04/05/2020 até 02/06/2020. Observou-se que, para os modelos apresentados, a projeção foi mais próxima da realidade quando consideramos o curto prazo. Além disso, os melhores modelos foram o de regressão linear e de regressão exponencial. No entanto, no longo prazo, esses modelos não se mostraram adequados para prever o preço do ativo.

Por fim, introduzimos alguns conceitos básicos de análise de sobrevivência (função de sobrevivência, função de risco, etc . . .) para funções de distribuição de probabilidade conhecida, como distribuição do tipo exponencial e de Poisson. Para quando a distribuição de probabilidade não é conhecida, introduzimos o estimador da função de sobrevivência, conhecido como estimador de *Kaplan-Meier*. Aplicamos esses conceitos para estudar a probabilidade do tempo de retorno do preço de um ativo a um valor de referência. Esse estudo levou em consideração informações de 80 empresas que compõem o índice IBOV, no período de 02/03/2020 a 05/02/2021. A partir da aplicação de alguns conceitos da análise de sobrevivência, o estudo mostrou, por exemplo, que a probabilidade do preço de um ativo se manter abaixo do preço de referência até 19^a semana é de 60% e que a probabilidade do preço do ativo atingir o preço desejado até esse período é de 40%. Mostrou-se também que, o tempo estimado para que a probabilidade de retorno ao preço de referência seja 50% é de 21 a 23 semanas.

Para o desenvolvimento deste trabalho foi necessário o uso de planilhas eletrônicas do *Google*, incluindo as funções do *Google Finance*, exploração do software *GeoGebra* para ilustrações gráficas além de uso ferramentas computacionais para formatação de texto usando

LaTeX (Overleaf, Tablesgenerator). O uso destas tecnologias, representou um desafio mas também um ganho de conhecimento próprio que terá impacto no desenvolvimento profissional, além do PROFMAT.

Notou-se durante a realização deste trabalho a escassez de referências em português sobre análise de sobrevivência aplicado no mercado financeiro.

Os conceitos abordados neste trabalho estão relacionados com diversos conteúdos do ensino médio (probabilidade, funções, esperança matemática, etc) e em particular, os conceitos relacionados ao mercado financeiro podem ser utilizados na disciplina de Educação Financeira, incluída na matriz curricular do Ensino Médio no Estado do Paraná desde 2021.

Como sequência para este trabalho sugere-se o aprofundamento de outras funções de distribuição de probabilidade com variável aleatória contínua, como por exemplo, distribuição de *Weibull* que possibilita determinar o tempo de vida médio e taxa de falhas ao longo do tempo. Esta distribuição de probabilidade tem diversas aplicações e em particular é amplamente utilizada em engenharia de confiabilidade.

Como, em diversas situações, o estudo de análise de sobrevivência envolve covariáveis que podem estar relacionadas com o tempo de sobrevivência, sugere-se um estudo que incorpore tais informações e neste caso um modelo amplamente utilizado é o modelo de risco proporcional de *COX*.

Profissionalmente o PROFMAT contribuiu para um aprofundamento do conhecimento proporcionado pelas disciplinas do curso o que reflete na atividade profissional, uma vez que desperta o interesse pela descoberta da origem de conceitos, fórmulas e aplicações no cotidiano.

Em relação ao desenvolvimento deste trabalho destacam-se os conhecimentos adquiridos no contexto de mercado financeiro e dos aprendizados de probabilidade, que serão levadas e utilizadas como investidor e aplicados também em sala de aula.

REFERÊNCIAS

- BORGES, F. R. A bolsa de valores brasileira como fonte de informações financeiras. **Perspectivas em Ciência da Informação**, v. 2, n. 2, p. 189–202, 1997. Disponível em: <<https://www.brapci.inf.br/index.php/res/v/32087>>. Acesso em: 19 Agosto 2022. 14
- BORTOLOSSI, H. J. **Cálculo Diferencial a Várias Variáveis: Uma introdução à Teoria de Otimização**. [S.l.]: Puc Rio, 2002. 618 p. 51, 53
- BURDEN, R.; FAIRES, J. D. **Análise Numérica**. 1a. ed. [S.l.]: Cengage Learning, 2008. 51
- BUSSAB, W.; MORETTIN, P. **Estatística Básica**. [S.l.]: Editora Saraiva, 2010. 16, 32
- CLARK, T. G. et al. Survival analysis part i: Basic concepts and first analyses. **British Journal of Cancer**, v. 89, p. 232–238, 2003. 73
- COLOSIMO, E. A. **Análise de Sobrevida Aplicada**. [S.l.]: Editora Blucher, 2006. 60
- HALLEY, E. An estimate of the degrees of the mortality of mankind; drawn from curious tables of the births and funerals at the city of breslaw; with an attempt to ascertain the price of annuities upon lives. **Philosophical Transactions of the Royal Society of London**, v. 17, n. 196, p. 596–610, 1693. Disponível em: <<https://royalsocietypublishing.org/doi/abs/10.1098/rstl.1693.0007>>. 16
- PINHEIRO, J. L. **Mercado de Capitais**. 9. ed. [S.l.]: Atlas, 2019. 289 p. Acesso em: 27 jun. 2022. 14
- RUGGIERO, M. A. G.; LOPES, V. L. da R. **Cálculo Numérico - Aspectos Teóricos e Computacionais**. 2. ed. [S.l.]: Pearson, 2000. 51
- STWEART, J. **Cálculo**. 8. ed. [S.l.]: Cengage Learning, 2017. v. 1. 51
- VIALI, L. Algumas considerações sobre a origem da teoria da probabilidade. **Revista Brasileira de História da Matemática**, v. 8, n. 16, p. 143–153, 2009. Disponível em: <<https://www.rbhm.org.br/index.php/RBHM/article/view/177/163>>. 16