



UNIVERSIDADE FEDERAL DO MARANHÃO

Fundação Instituída nos termos da Lei 5.152 de 21/10/1966 - São Luís - MA

Centro de Ciências Exatas e Tecnologia
Curso de Matemática – Licenciatura

Aline Silva Dourado

Distribuições de Probabilidade Aplicadas em Dados do Futebol

São Luís - MA
2023

Aline Silva Dourado

Distribuições de Probabilidade Aplicadas em Dados do Futebol

Monografia apresentada à Coordenadoria dos cursos de Matemática, da Universidade Federal do Maranhão, como requisito parcial para obtenção do grau de Licenciado em Matemática.

Curso de Matemática – Licenciatura
Universidade Federal do Maranhão

Orientador: Prof. Dr. Josenildo de Souza Chaves

São Luís - MA

2023

Aline Silva Dourado

Distribuições de Probabilidade Aplicadas em Dados do Futebol/ Aline Silva
Dourado. – São Luís - MA, 2023-
38 p. : il.(alguma color.); 30 cm.

Orientador: Prof. Dr. Josenildo de Souza Chaves

Monografia – Curso de Matemática – Licenciatura
Universidade Federal do Maranhão, 2023.

1. Variáveis Aleatórias. 2. Algumas Distribuições de Probabilidade Importantes.
3. Aplicações. I. Orientador. II. Universidade Federal do Maranhão. III. Curso de
Matemática – Licenciatura. IV. Título

CDU XX:XXX:XXX.X

Aline Silva Dourado

Distribuições de Probabilidade Aplicadas em Dados do Futebol

Monografia apresentada à Coordenadoria dos cursos de Matemática, da Universidade Federal do Maranhão, como requisito parcial para obtenção do grau de Licenciado em Matemática.

Trabalho **SUBMETIDO** . São Luís - MA, ____/____/____

Prof. Dr. Josenildo de Souza Chaves
Orientador
DEMAT/UFMA

Prof.^a Dr.^a
Primeira Examinadora
DEMAT/UFMA

Prof. Dr.
Segundo Examinador
DEMAT/UFMA

Ao Deus de Abraão, Isaque e Jacó.

Agradecimentos

À Cristo que me salvou, me redimiou e me amou de tal maneira.

À minha querida família, que sempre esteve ao meu lado e me deu condições de chegar até aqui. A dona Silvana, o seu Airton e o Vinícius vocês são e sempre serão os amores da minha vida.

À meus amigos da Matemática, em especial Lázaro Reis, Ariany Pereira e Amanda Castro. Nunca me esquecerei do incentivo e companheirismo de ambos, vocês são parte dessa conquista. Devo citar também Rianderson Correa, Tiago Miranda, Douglas Guedes e José Vitor.

À todos os grupos de amigos que fiz parte, vocês me deram o abraço e o apoio que eu precisava para seguir em frente. Ao Orlando Neto pelo cuidado, paciência e ajuda, você foi providência divina na minha vida. Devo citar também meus amigos queridos Thalia Oliveira, Pedro Nunes e Ana Beatriz, obrigada por existirem e por fazerem meus dias mais leves.

À Aliança Bíblica Universitária do Brasil, por ter sido meu porto seguro, pelas experiências, pelas pessoas maravilhosas que tive o privilégio de conhecer e por ter me aproximado do melhor amigo, Jesus Cristo.

À Cassiane Reis, Willard Figueiredo e Reginaldo Louzeiro, certamente os melhores amigos que alguém poderia ter.

À meu excepcional orientador Josenildo Chaves, que eu seja capaz de ser tão paciente, honesta e disciplinada como você.

À meu querido coordenador do curso Cléber Araújo, obrigada pela paciência e atenção de sempre.

À todos os excelentes professores do Departamento de Matemática e todos que contribuíram na minha formação desde os anos iniciais.

E por fim, todos que pude ter o privilégio de partilhar da vida até aqui, meus sinceros e amistosos agradecimentos. Vida longa e não se esqueçam, a jornada é a melhor parte!

“e pelas suas pisaduras, fomos sarados.”

Isaías 53:5

Resumo

O futebol é o esporte mais praticado do Brasil e é marca registrada da identidade brasileira. Sendo pentacampeão da maior competição de equipes internacionais, é referência da alegria e simpatia do seu povo. Pensando nisso, este trabalho tem por objetivo apresentar as distribuições de probabilidade importantes fazendo aplicações em dados futebolísticos com uso do R.

Palavras-chave: Futebol, Distribuição de Probabilildade, Aplicações.

Abstract

Soccer is the most practiced sport in Brazil and is a trademark of the Brazilian identity. Being a five-time champion of the biggest competition of international teams, it has become a reference of the joy and sympathy of its people. With this in mind, this work aims to present the important probability distributions making applications in soccer data using R.

Keywords: Soccer, Probability Distributions, Applications

Sumário

1	INTRODUÇÃO	10
1.1	Organização do Trabalho	10
2	VARIÁVEIS ALEATÓRIAS	11
2.1	Variáveis Aleatórias Discretas	11
2.1.1	Média, valor esperado ou esperança matemática de uma v.a. discreta	12
2.1.2	Variância de uma v.a. discreta	13
2.2	Variáveis Aleatórias Contínuas	15
2.2.1	Função distribuição acumulada de uma v.a. contínua	16
2.2.2	Média, valor esperado ou esperança matemática de uma v.a. contínua	18
2.2.3	Variância	19
2.2.4	Função distribuição acumulada	19
3	ALGUMAS DISTRIBUIÇÕES DE PROBABILIDADE IMPORTANTES	21
3.1	Distribuição de Bernoulli	21
3.2	Distribuição Binomial	22
3.3	Distribuição Geométrica	22
3.4	Distribuição Binomial Negativa	23
3.5	Distribuição Hipergeométrica	24
3.6	Distribuição de Poisson	25
3.7	Distribuição Normal	26
4	APLICAÇÕES	30
4.1	Introdução	30
4.2	Copa do Mundo 2022	30
4.3	Análise de dados da partida Brasil x Croácia	31
4.3.1	Número de Finalizações	32
4.3.1.1	Cruzamentos	33
4.3.1.2	Tempo de jogo	33
4.4	Brasileirão 2022	34
4.4.1	Gols	34
4.5	Clássico do Povo	35
4.5.1	Vitórias por estádio	35
5	CONSIDERAÇÕES FINAIS	37

REFERÊNCIAS 38

1 Introdução

Variáveis aleatórias estão associadas a experimentos aleatórios presentes em todas as áreas do conhecimento. Seguindo [Morettin e Bussab \(2017\)](#) e [Meyer \(1983\)](#) apresentamos primeiro as variáveis aleatórias discretas e em seguida as contínuas. As suas principais características estão exploradas e serão exemplificadas principalmente por [Ross \(2009\)](#), [Ara, Musetti e Schneiderman \(2003\)](#) e [??](#)). As distribuições de probabilidades importantes descrevem os comportamentos dessas variáveis aleatórias em determinadas condições e serão neste trabalho atribuídas em especial, ao futebol.

Segundo [Santos \(2014\)](#) o futebol é uma manifestação cultural que possui uma intensa relação com o povo brasileiro. É quase que umbilical, porém histórica, essa relação entre futebol e Brasil se expressa na realidade de muitas maneiras: seja na forma de lazer, com a famosa pelada do fim de semana, seja em uma conversa de bar, seja na compra de uma chuteira do jogador Neymar, ou na aquisição do ingresso que garante o passaporte para o espetáculo. A questão é que o futebol representa comportamentos, linguagens e interesses de muitas pessoas, independente de idade ou condições socioeconômicas. Pensando nisso, estamos interessados em aplicar os conceitos abordados e apresentar como a estatística está presente na vida cotidiana do brasileiro, até mesmo no seu time do coração.

1.1 Organização do Trabalho

Este trabalho está organizado em 5 capítulos. No Capítulo 2, apresentamos as definições e principais propriedades das variáveis aleatórias discretas e contínuas, facilitando sua compreensão através de exemplos objetivos. No Capítulo 3, estamos interessados em apresentar as distribuições de probabilidade importantes que são elas: Bernoulli, Binomial, Geométrica, Binomial Negativa, Hipergeométrica, Poisson e Normal; considerando principalmente suas particularidades.

O Capítulo 4, trata das aplicações das distribuições em dados do futebol com o uso da linguagem R. Um exemplo de seção dentro do capítulo é a aplicação em dados da Copa do Mundo, que é a principal competição do desporto a nível mundial. Esse exemplo, familiariza torcedores de todas as idades, até mesmo aqueles que se atentam apenas a essa competição. O Capítulo 5, trata das considerações finais, promovendo continuidade dessa proposta de aplicação.

2 Variáveis Aleatórias

De acordo com [Morettin e Bussab \(2017\)](#), nos experimentos aleatórios há o interesse em quantificar variáveis dando a elas um valor numérico. Por exemplo, considere o caso de um questionário em que uma pessoa é indagada a respeito de uma proposição, e as respostas possíveis são sim ou não. Podemos associar ao problema uma variável que toma dois valores, 1 ou 0, por exemplo, correspondentes às respostas sim ou não, respectivamente. Dessa forma, transformamos a qualidade de um experimento em uma quantidade. Assim, temos que uma variável aleatória (v.a.) é uma função que associa os elementos do espaço amostral a valores da reta real.

Definição 2.1. *Seja Ω um espaço amostral associado a um experimento aleatório ε . Uma função X , que associa a cada elemento $\omega \in \Omega$ um número real, $X(\omega)$, é denominada variável aleatória.*

Notação. *Para representar as variáveis aleatórias usa-se, por exemplo, X , Y , Z e W (letras latinas maiúsculas) e respectivamente x , y , z e w (letras latinas minúsculas) para representar os valores que aquelas variáveis aleatórias assumem.*

2.1 Variáveis Aleatórias Discretas

Definição 2.2. *Uma variável aleatória X é discreta se o número de valores possíveis de X for um número finito ou um infinito enumerável.*

Exemplo 2.1. *São exemplos de v.a. discretas:*

- Número de alunos aprovados na disciplina de matemática no último ano do ensino médio em uma escola.
- Número de títulos de um time de futebol em um campeonato nacional.
- Número de atendimentos realizados por um médico por dia da semana em um hospital.

Seguindo [Meyer \(1983\)](#), temos a seguinte definição de função de probabilidade.

Definição 2.3. *Seja X uma v.a. Então, R_X , o contradomínio de X , será formado no máximo por um número infinito numerável de valores $x_1, x_2, \dots$. A cada possível resultado x associaremos um número $p(x_i) = P(X = x_i)$, denominado probabilidade de x_i . Os números $p(x_i)$, $i = 1, 2, \dots$ devem satisfazer às seguintes condições:*

$$p(x_i) \geq 0, \text{ para todo } i, \quad (2.1)$$

$$\sum_{i=1}^{\infty} p(x_i) = 1. \quad (2.2)$$

A função $(x, p(x))$ é chamada função de probabilidade (fp) da v.a. X ou apenas distribuição de probabilidade.

Notação. Utiliza-se a notação $P(X = x_i) = p(x_i) = p_i, i = 1, 2, \dots$ para representar as probabilidades que a variável X assume.

Exemplo 2.2. Seja n o número de valores que uma v.a. discreta X pode assumir. Então a fp dessa v.a. X pode ser representada pela tabela seguinte.

x_i	x_1	x_2	$\dots$	x_n	Σ
$p(x_i)$	$p(x_1)$	$p(x_2)$	$\dots$	$p(x_n)$	1

Fonte: Autora.

2.1.1 Média, valor esperado ou esperança matemática de uma v.a. discreta

Há situações em que é necessário obter uma ou mais medidas resumo para uma v.a. X . Podemos fazer isso, determinando valores que sejam representativos da v.a., como por exemplo, medidas de posição, variabilidade e assimetria. Uma medida de posição chamada de valor esperado é dada a seguir.

Definição 2.4. Seja X uma v.a. discreta, assumindo os valores $x_1, \dots, x_n$. Então, o valor esperado de X é dado por

$$E(X) = \sum_{i=1}^n x_i P(X = x_i) = \sum_{i=1}^n x_i p(x_i). \quad (2.3)$$

Notação. A notação $E(X) = \mu_X = \mu$ é usada para representar o valor esperado de uma variável aleatória X .

Segundo [Ara, Musetti e Schneiderman \(2003\)](#), o valor esperado de uma v.a. X tem as seguintes propriedades (assumindo k como constante):

- $E(k) = k$,
- $E(kX) = kE(X)$,
- $E(X + Y) = E(X) + E(Y)$,
- $E(X + k) = E(X) + k$.

2.1.2 Variância de uma v.a. discreta

Definição 2.5. Dada uma v.a. X discreta, que assume os valores $x_1, \dots, x_n$, chamamos de variância de X ao valor

$$\text{Var}(X) = \sum_{i=1}^n (x_i - \mu)^2 p(x_i), \quad (2.4)$$

em que, $\mu = E(X)$.

Notação. A notação $\text{Var}(X) = \sigma_X^2 = \sigma^2$ é usada para representar a variância de uma variável X .

O desvio padrão de X , σ_X , é definido como a raiz quadrada positiva da variância, $\sigma = \sigma_X = \sqrt{\sigma_X^2}$. Devemos observar que σ_X é dado na mesma unidade de medida de X .

Se X for uma v.a. e a uma constante, são válidas as seguintes propriedades.

$$\text{Var}(a + X) = \text{Var}(X) \quad (2.5)$$

e

$$\text{Var}(aX) = a^2 \text{Var}(X). \quad (2.6)$$

Se X e Y são variáveis aleatórias independentes, então:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y). \quad (2.7)$$

A equação (2.4) pode ser reescrita na seguinte forma:

$$\begin{aligned} \text{Var}(X) &= \sum (x_i - \mu)^2 p(x_i) = \sum (x_i^2 - 2x_i\mu + \mu^2) p(x_i) \\ &= \sum x_i^2 p(x_i) - \sum 2x_i\mu p(x_i) + \sum \mu^2 p(x_i) \\ &= E(X^2) - 2\mu \sum x_i p(x_i) + \mu^2 \sum p(x_i) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2. \end{aligned} \quad (2.8)$$

Exemplo 2.3. Dada a função de distribuição de probabilidade discreta,

X	0	1	2	3	Σ
$P(X = x)$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{4}{12}$	$\frac{5}{12}$	1

obtemos $E(X)$ e $\text{Var}(X)$ do seguinte modo: Temos que,

$$E(X) = 0 \times \frac{1}{12} + 1 \times \frac{2}{12} + 2 \times \frac{4}{12} + 3 \times \frac{5}{12} = \frac{25}{12}$$

e

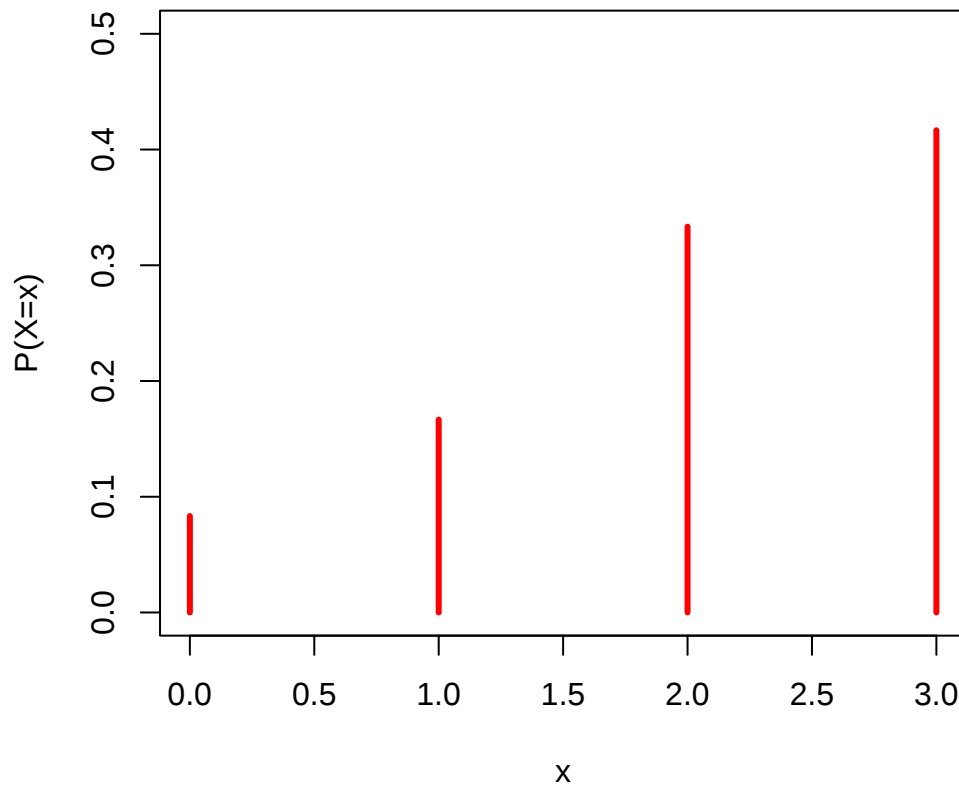
$$E(X^2) = 0^2 \times \frac{1}{12} + 1^2 \times \frac{2}{12} + 2^2 \times \frac{4}{12} + 3^2 \times \frac{5}{12} = \frac{63}{12}.$$

Logo,

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{63}{12} - \left(\frac{25}{12}\right)^2 = \frac{131}{144}.$$

A sua função de distribuição de probabilidade é representada pela Figura 1.

Figura 1 – Função de probabilidade do Exemplo 2.3.



Fonte: A Autora.

O exemplo seguinte apresenta a v.a. uniforme discreta, muito importante na geração de números aleatórios.

Exemplo 2.4. *Uma v.a. discreta X , assumindo os valores $x_1, \dots, x_k$, tem distribuição uniforme se, e somente se,*

$$p(x_i) = \frac{1}{k}. \quad (2.9)$$

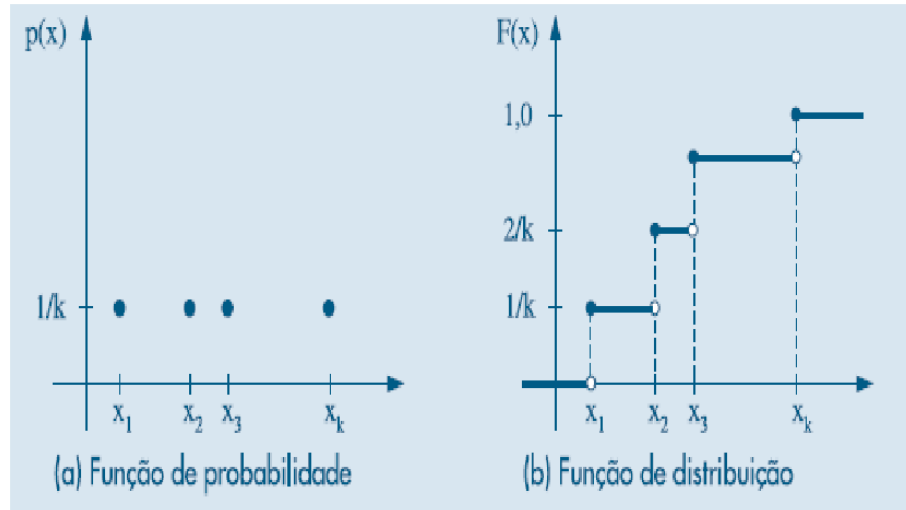
para todo $i = 1, 2, \dots, k$. De imediato, podemos observar que

$$E(X) = \frac{1}{k} \sum_{i=1}^k x_i \quad (2.10)$$

e

$$Var(X) = \frac{1}{k} \left(\sum_{i=1}^k x_i^2 - \frac{(\sum_{i=1}^k x_i)^2}{k} \right). \quad (2.11)$$

Figura 2 – (a) fp e (b) fda do Exemplo 2.4 .



Fonte: Morettin e Bussab (2017).

2.2 Variáveis Aleatórias Contínuas

Definição 2.6. Uma função X , definida sobre o espaço amostral Ω e assumindo valores num intervalo de números reais, é chamada de variável aleatória contínua.

Exemplo 2.5. São exemplos de v.a.'s contínuas:

- A altura de jogadores de um time de futebol.
- O tempo de falha de um equipamento eletrônico.
- O peso de animais em uma localidade.
- A quantidade de chuva que ocorre em uma região.

Definição 2.7. Uma variável aleatória X é contínua se existir uma função f , chamada função densidade de probabilidade (fdp) de X satisfazendo as seguintes condições

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

e

$$P(a \leq X \leq b) = \int_a^b f(x) dx, \quad -\infty < a < b < \infty.$$

A probabilidade de uma *v.a* contínua X pertencer a um intervalo $(a, b]$ é dada por

$$P(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a).$$

Note que,

$$P(X \leq b) - P(X < a) = F(b) - F(a)$$

e que,

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

2.2.1 Função distribuição acumulada de uma *v.a.* contínua

Definição 2.8. A função de distribuição acumulada (*fda*) de uma (*v.a.*) X com *fdp* $f(x)$ é dada por

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(s)ds. \quad (2.12)$$

Devemos observar que $F(x)$ segue as seguintes propriedades:

- (i) $F(x)$ é não-decrescente. Isto é, se $x_1 \leq x_2$, então $F(x_1) \leq F(x_2)$.
- (ii) $\lim_{x \rightarrow +0} F(x) = 0$ e $\lim_{x \rightarrow -0} F(x) = 1$.
- (iii) Pelo teorema fundamental do cálculo $\frac{d}{dx}F(x) = f(x)$, para todo x no qual f seja derivável.

Exemplo 2.6. Seja X uma *v.a* contínua com *fdp*

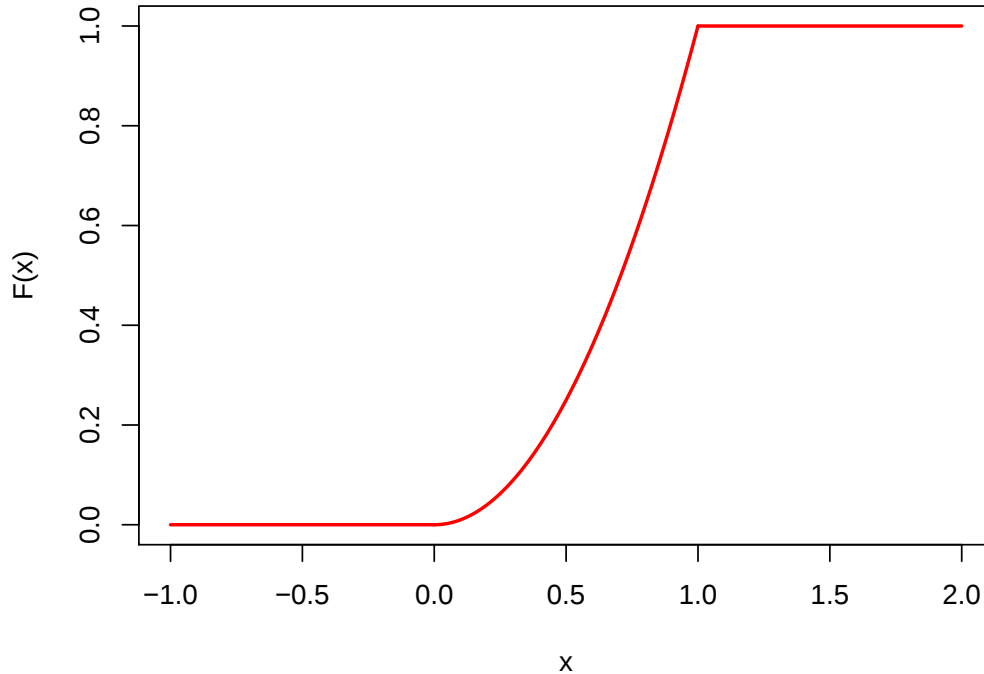
$$\begin{aligned} f(x) &= 2x, \quad 0 < x < 1. \\ &= 0, \quad x \notin (0, 1). \end{aligned}$$

Portanto a *fda* da *v.a.* X , $F(x)$ é dada por:

$$\begin{aligned} F(x) &= 0, \quad \text{se } x \leq 0 \\ &= x^2, \quad \text{se } 0 < x < 1 \\ &= 1, \quad \text{se } x \geq 1. \end{aligned}$$

Na Figura 3 temos uma representação gráfica de $F(x)$.

Figura 3 – Distribuição acumulada do Exemplo 2.6.



No exemplo a seguir, adaptado de [Ross \(2009\)](#), temos uma aplicação da Definição 2.7.

Exemplo 2.7. Suponha que X seja uma variável aleatória contínua cuja função densidade de probabilidade é dada por

$$f(x) = \begin{cases} c(4x - 2x^2) & \text{se } 0 < x < 2 \\ 0, & \text{caso contrário} \end{cases}$$

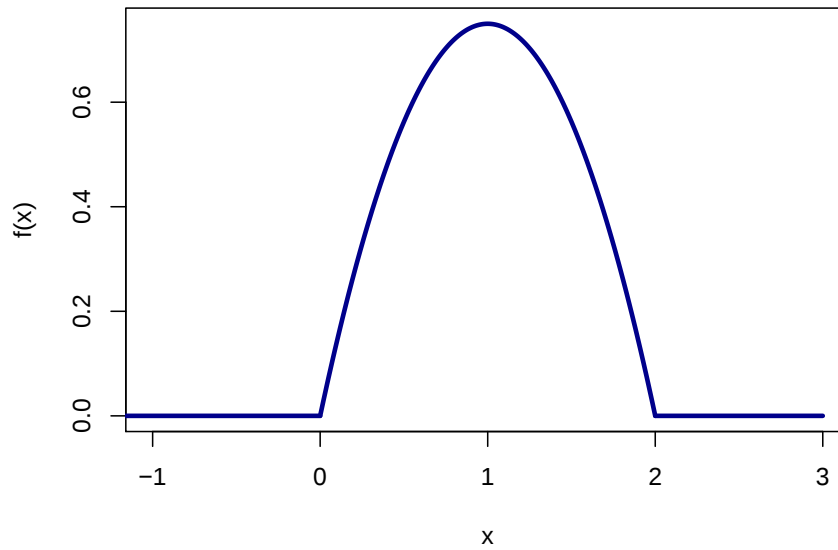
Vamos determinar o valor de c e $P(X < 1)$. Note que $f(x) \geq 0$ em $(0, 2)$. Como f é uma fdp, devemos ter que $\int_{-\infty}^{+\infty} f(x) dx = 1$, o que implica em

$$\begin{aligned} c \int_0^2 (4x - 2x^2) dx &= 1 \\ c \left(2x^2 - \frac{2x^3}{3} \right) \Big|_0^2 &= 1 \\ c &= \frac{3}{8}. \end{aligned}$$

$$\text{Portanto, } P(X < 1) = \int_0^1 f(x) dx = \frac{3}{8} \int_0^1 (4x - 2x^2) dx = \frac{1}{2}.$$

Podemos observar que $P(X < 1)$, assim como o valor esperado da v.a. X do Exemplo 2.7, pode ser determinado pela observação do gráfico da fdp de X , vide Figura 4. Ao se verificar que $f(x)$ é simétrica em relação à reta $x = 1$, de imediato se conclui que $E(X) = 1$.

Figura 4 – fdp da v.a. X do Exemplo 2.7 .



Fonte: A autora.

A formalização do conceito de valor esperado de uma v.a contínua é dada a seguir.

2.2.2 Média, valor esperado ou esperança matemática de uma v.a. contínua

Definição 2.9. O valor esperado de uma (v.a.) X contínua sendo $f(x)$ sua (fdp), é dado por

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx. \quad (2.13)$$

É importante observar que a integral em (2.13) pode não convergir. Assim, $E(X)$ existe se, e somente se,

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

for finita.

Exemplo 2.8. (ARA; MUSETTI; SCHNEIDERMAN, 2003) o valor esperado da variável aleatória contínua X , com função densidade de probabilidade $f(x) = 3x^2, 0 \leq x \leq 1$, é

dado por

$$\mu = E(X) = \int_{-\infty}^0 x \cdot 0 dx + \int_0^1 x \cdot 3x^2 dx + \int_1^{+\infty} x \cdot 0 dx = \int_0^1 3x^3 dx = \frac{3}{4}. \quad (2.14)$$

2.2.3 Variância

Definição 2.10. A variância de uma v.a. contínua X , denotada por $\text{Var}(X) = \sigma^2$, é dada por

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx. \quad (2.15)$$

De acordo com a Equação (2.8), podemos reescrever (2.15) na forma

$$\text{Var}(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2.$$

Exemplo 2.9. Vamos calcular a variância do Exemplo 2.8.

Observe que já temos que $E(X) = \frac{3}{4}$ e que $E(X^2) = \int_0^1 x^2 \cdot 3x^2 dx = \int_0^1 3x^4 dx = \frac{3}{5}$. Logo, $\text{Var}(X) = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{80}$.

2.2.4 Função distribuição acumulada

De acordo com [Morettin e Bussab \(2017\)](#) a função distribuição acumulada (fda) de uma v.a X pode ser definida do seguinte modo.

Definição 2.11. Dada a variável aleatória X , chamamos de função distribuição acumulada (fda), ou simplificada de função distribuição (fd) à função

$$F(x) = P(X \leq x). \quad (2.16)$$

Segue-se imediatamente desta definição que,

$$F(x) = P(X \leq x) = \sum_{x_j \leq x} p(x_j). \quad (2.17)$$

A função distribuição acumulada da v.a X do Exemplo 2.3 é determinada pelo exemplo seguinte.

Exemplo 2.10. Continuação do Exemplo 2.3. Primeiro calculamos as probabilidades:

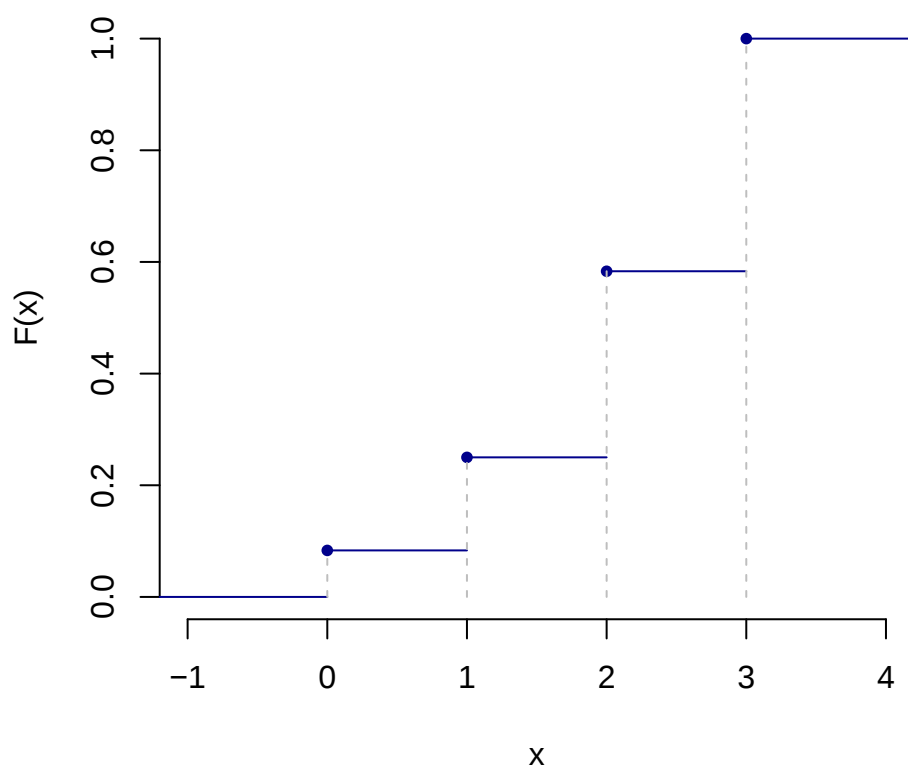
$$\begin{aligned} P(0 \leq x < 1) &= P(X = 0) = \frac{1}{12} \\ P(1 \leq x < 2) &= P(0 \leq x < 1) + P(X = 1) = \frac{3}{12} \\ P(2 \leq x < 3) &= P(1 \leq x < 2) + P(X = 2) = \frac{7}{12} \\ P(x > 3) &= P(0 \leq x < 1) + P(1 \leq x < 2) + P(2 \leq x < 3) = 1. \end{aligned}$$

Assim, temos que

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ \frac{1}{12} & \text{se } 0 \leq x < 1, \\ \frac{3}{12} & \text{se } 1 \leq x < 2, \\ \frac{7}{12} & \text{se } 2 \leq x < 3, \\ 1 & \text{se } x \geq 3. \end{cases}$$

Uma representação gráfica da $F(x)$ é dada pela Figura 5.

Figura 5 – Distribuição acumulada do Exemplo 2.10.



Fonte: A autora.

3 Algumas Distribuições de Probabilidade Importantes

3.1 Distribuição de Bernoulli

Definição 3.1. *Provas de Bernoulli são experimentos aleatórios com apenas dois resultados possíveis chamados por convenção um de "sucesso" e o outro de "fracasso".*

Definição 3.2. *Uma variável aleatória discreta X , que assume apenas os valores 1 para {sucesso} e 0 para {fracasso}, com função de probabilidade $(x, p(x))$ tal que*

$$P(X = x) = p^x(1 - p)^{1-x}, \quad x = 0, 1 \quad (3.1)$$

é chamada de variável aleatória de Bernoulli.

Notação. $X \sim \text{Bernoulli}(p)$, indica que a v.a. X tem distribuição de Bernoulli com parâmetro p para sucesso.

Exemplo 3.1. *São exemplos de aplicações da distribuição de Bernoulli:*

- *O resultado de um teste de gravidez é positivo ou negativo.*
- *A aprovação ou reprovação de um estudante no Exame Nacional do Ensino Médio (ENEM);*
- *Um candidato é eleito ou não em uma eleição presidencial;*
- *Ocorrência de cara ou coroa no lançamento de uma moeda.*

Seja $P(X = 1) = p$ e $P(X = 0) = 1 - p$. A sua função de distribuição acumulada é definida como

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1 - p & \text{se } 0 \leq x < 1, \\ 1 & \text{se } x \geq 1. \end{cases}$$

Notação. A esperança é representada por $E(X) = p$ e a variância por $\text{Var}(X) = p(1 - p)$.

Seguindo Meyer (1983) vamos enunciar a distribuição binomial a seguir.

3.2 Distribuição Binomial

Teorema 3.1. *Seja X uma variável binomial, baseada em n repetições. Então,*

$$P(X = x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, x = 0, 1, 2, \dots, n. \quad (3.2)$$

Demonstração: *Considere-se um particular elemento do espaço amostral de S satisfazendo à condição $X = k$. Um resultado como esse poderia surgir, por exemplo, se nas primeiras k repetições de S ocorresse A , enquanto nas últimas $n - k$ repetições ocorresse $\bar{A}$, isto é,*

$$\underbrace{AAA \dots A}_n \underbrace{\bar{A}\bar{A}\bar{A} \dots \bar{A}}_{n-k}$$

Como todas as repetições são independentes, a probabilidade desta sequência particular seria $p^x(1-p)^{n-x}$ mas exatamente essa mesma probabilidade seria associada a qualquer outro resultado para o qual $X = k$. O número total de tais resultados é igual a $\binom{n}{k}$, porque deveremos escolher exatamente k posições (dentre n) para o evento A . Ora, isso dá o resultado acima, porque esses $\binom{n}{k}$ resultados são todos mutuamente excludentes. ■

Para mostrar que $\sum_{x=0}^n P(X = x) = 1$, basta observar que

$$\sum_{x=0}^n P(X = x) = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = (p + 1 - p)^n = 1^n = 1,$$

pela aplicação direta do Teorema Binomial.

Notação. *Utiliza-se a notação $E(X) = np$ e $Var(X) = np(1-p)$ para representar a esperança e a variância de variável aleatória discreta com distribuição binomial.*

3.3 Distribuição Geométrica

Quando estamos interessados em calcular a probabilidade de ocorrer um sucesso em determinada quantidade de tentativas, podemos utilizar a distribuição geométrica que será enunciada a seguir pelo [Meyer \(1983\)](#).

Definição 3.3. *Seja uma variável aleatória X definida como o número de repetições necessárias para obter a primeira ocorrência de um evento A (nele se incluindo a última). Como X toma os valores possíveis $(k-1)$ repetições de A darem o resultado $\bar{A}$, temos:*

$$P(X = k) = q^{k-1}p, k = 1, 2, \dots \quad (3.3)$$

A equação 3.3 é obtida porque para que X seja igual a k é necessário que as $(k-1)$ repetições sejam fracassos e que a n -ésima repetição seja um sucesso. Como se supõe que os resultados sejam independentes e $P(X = k) \geq 0$, temos que:

$$\sum_{k=1}^{\infty} P(X = k) = p(1 + q + q^2 + \dots) = p \left[p \frac{1}{1-q} \right] = 1$$

dessa forma, com a probabilidade igual 1 o sucesso acabará ocorrendo. Assim, qualquer variável aleatória definida pela equação 3.3 é definida como uma variável aleatória geométrica com parâmetro p .

Podemos obter o valor esperado de X , da seguinte maneira:

$$E(X) = \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} \frac{d}{dq} q^k = p \frac{d}{dq} \sum_{k=1}^{\infty} q^k = p \frac{d}{dq} \left[\frac{q}{1-q} \right] = \frac{1}{p}.$$

a troca da derivação e do somatório é válida porque a série converge quando $|q| < 1$. De forma análoga a $V(X) = \frac{q}{p^2}$.

Uma generalização da distribuição geométrica será definida a seguir pelo [Meyer \(1983\)](#).

3.4 Distribuição Binomial Negativa

Suponha que um experimento seja realizado até que um evento A ocorra na r – *énésima* vez. Se $P(A) = p$, $P(A) = q = 1 - p$ em cada repetição, definimos a variável aleatória Y como sendo o número de repetições necessárias a fim de que A possa ocorrer exatamente r vezes .

Ora $Y = k$ se, e somente se, A ocorrer na k – *ésima* repetição e A tiver ocorrido exatamente $(r - 1)$ vezes nas $(k - 1)$ repetições anteriores. A probabilidade desse evento é $p \binom{k-1}{r-1} p^{r-1} q^{k-r}$, desde que o que acontece nas $(k - 1)$ repetições é independente daquilo que acontece na k – *ésima* repetição. Portanto,

$$P(Y = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad k = r, r+1, \dots \quad (3.4)$$

Observe que para $k = 1$ a equação 3.4 acima se reduz a 3.3. Assim, uma variável aleatória que tenha a distribuição de probabilidade dada pela expressão 3.4 é chamada de distribuição binomial negativa ou de distribuição de Pascal com parâmetros (r, p) .

Para verificar a condição

$$\sum_{k=1}^{\infty} P(X = k) = 1,$$

obtemos

$$\sum_{k=1}^{\infty} \binom{k-1}{r-1} p^r q^{k-r} = p^r \sum_{k=1}^{\infty} \binom{k-1}{r-1} q^{k-r} = p^r (1-q)^{-r},$$

que é obviamente igual a 1. Isso porque a última igualdade provém do desenvolvimento em série de

$$(1-q)^{-r} = \sum_{n=0}^{\infty} \binom{-r}{n} (-q)^n,$$

que é igual a

$$\sum_{k=1}^{\infty} \binom{k-1}{r-1} q^{k-r},$$

depois de algumas simplificações algébricas e com o uso do coeficiente binomial generalização. Por conta do expoente negativo $(-r)$, a distribuição é conhecida por binomial negativa.

Notação. Utiliza-se $E(X) = \frac{r}{p}$ e $Var(X) = \frac{rq}{p^2}$ para representar respectivamente a esperança e a variância de uma variável aleatória com distribuição binomial negativa.

Seguindo Meyer (1983) vamos enunciar a distribuição hipergeométrica.

3.5 Distribuição Hipergeométrica

Seja um lote com N peças, r das quais sejam defeituosas e $(N - r)$ sejam não-defeituosas. Suponha que escolhamos, ao acaso, n peças desse lote ($n \leq N$), sem reposição. Seja X o número de peças defeituosas encontradas. Desde que $X = k$ se, e somente se, obtivermos exatamente k peças defeituosas (dentre as r defeituosas do lote) e exatamente $(n - k)$ não-defeituosas [dentre as $(N - r)$ não-defeituosas do lote], teremos:

$$P(X = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots \quad (3.5)$$

Notação. Denotamos por $X \sim \text{hiper}(N, R, n)$ a distribuição hipergeométrica.

Teorema 3.2. X tem distribuição hipergeométrica como indica a equação 3.5. Seja $p = \frac{r}{N}$, $q = 1 - p$. Nessa caso, temos:

- $E(X) = np$;
- $V(X) = npq \frac{N - n}{N - 1}$;
- $P(X = k) \simeq \binom{n}{k} p^k (1 - p)^{n-k}$, para N grande.

Vamos ilustrar o último item com um exemplo a seguir. Suponha que se deseje calcular $P(X = 0)$.

Para $n = 1$, obtemos para a distribuição hipergeométrica, $P(X = 0) = \frac{(N-r)}{N} = 1 - \frac{r}{N} = q$. Para a distribuição binomial, obtemos que $P(X = 0) = q$. Portanto, essas respostas são iguais, quando se trata de $n = 1$.

Para $n = 2$, obtemos para a distribuição hipergeométrica

$$P(X = 0) = \frac{N-r}{N} \frac{N-r-1}{N-1} = \left(1 - \frac{r}{N}\right) = \left(1 - \frac{r}{N-1}\right)$$

Para a distribuição binomial, obtemos $P(X = 0) = q^2$. Observa-se que $1 - \frac{r}{N} = q$, enquanto $\left[1 - \frac{r}{N-1}\right]$ é quase igual a q . Em geral, a aproximação da distribuição hipergeométrica pela binomial é bastante boa, se $\frac{n}{N} \geq 0,1$.

Notação. Denotamos por $E(X) = n \frac{r}{N}$ e $Var(X) = n \frac{r}{N} \frac{N-r}{N} \frac{N-n}{N-1}$.

3.6 Distribuição de Poisson

Definição 3.4. Dada uma variável aleatória X discreta, assumindo os valores $0, 1, 2, \dots$, se

$$P(X = k) = \frac{e^{-\alpha} \alpha^k}{k!}, \quad k = 0, 1, 2, \dots, \quad (3.6)$$

então, X tem uma distribuição de Poisson com parâmetros $\alpha > 0$.

Podemos observar que a expressão (3.6) é uma verdadeira distribuição de probabilidade, visto que

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{e^{-\alpha} \alpha^k}{k!} = e^{-\alpha} e^{\alpha} = 1.$$

A determinação do valor esperado e da variância de uma v.a. X com distribuição de Poisson com parâmetro λ é dada a seguir.

Teorema 3.3. Se X tem distribuição de Poisson com parâmetro α , então $E(X) = Var(X) = \alpha$.

Demonstração: Temos que,

$$E(X) = \sum_{k=0}^{\infty} k \frac{e^{-\alpha} \alpha^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{-\alpha} \alpha^k}{(k-1)!}.$$

Fazendo $s = k - 1$ na última soma, segue-se que

$$E(X) = \sum_{s=0}^{\infty} \frac{e^{-\alpha} \alpha^{s+1}}{s!} = \alpha \sum_{s=0}^{\infty} \frac{e^{-\alpha} \alpha^s}{s!} = \alpha.$$

Para determinar a variância basta calcular $E(X^2)$ e aplicar ao resultado $Var(X) = E(X^2) - E^2(X)$.

O cálculo de $E(X^2)$ é dado por

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \frac{e^{-\alpha} \alpha^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\alpha} \alpha^k}{(k-1)!}.$$

Fazendo novamente $s = k - 1$ na última soma, obtemos

$$E(X^2) = \sum_{s=0}^{\infty} (s+1) \frac{e^{-\alpha} \alpha^{s+1}}{s!} = \alpha \sum_{s=0}^{\infty} \frac{e^{-\alpha} \alpha^s}{s!} + \alpha \sum_{s=0}^{\infty} \frac{e^{-\alpha} \alpha^s}{s!} = \alpha^2 + \alpha.$$

Portanto,

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \alpha^2 + \alpha - \alpha^2 = \alpha.$$

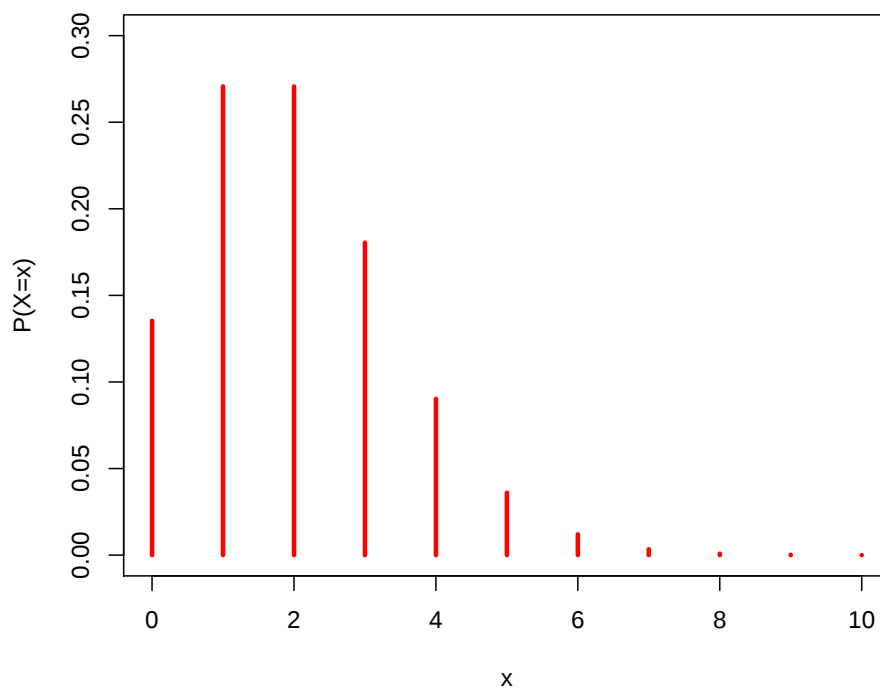


Exemplo 3.2. Suponha que a quantidade de erros de finalização em um jogo de futebol tenha distribuição de Poisson com o parâmetro $\alpha = 2$. Vamos calcular a probabilidade de se ter ao menos três erros de finalização no jogo. Denotamos $X =$ número de erros de finalização de um jogo. Então,

$$P(X \geq 3) = 1 - P(X < 3) = 1 - \left(\frac{e^{-2}(2)^0}{0!} \right) - \left(\frac{e^{-2}(2)^1}{1!} \right) - \left(\frac{e^{-2}(2)^2}{2!} \right) = 0,3233.$$

A Figura (6) mostra o gráfico da função de probabilidade da v.a. de Poisson associada ao Exemplo 3.2.

Figura 6 – Função de probabilidade da uma v.a. Poisson com parâmetro $\alpha = 2$ do Exemplo 3.2.



Fonte: A autora.

3.7 Distribuição Normal

Em estatística e probabilidade a distribuição normal é uma das mais importantes. É também conhecida como distribuição gaussiana, distribuição de Gauss ou distribuição

de Laplace–Gauss, em referência aos matemáticos, francês Pierre–Simon Laplace (1749 – 1827) e alemão Johann Carl Friedrich Gauss (1777 – 1855).

Definição 3.5. Uma v.a. X tem distribuição normal com média μ e variância σ^2 se a função densidade de probabilidade é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty, \quad (3.7)$$

sendo que, $-\infty < \mu < +\infty$ e $\sigma > 0$. Usamos a notação $X \sim N(\mu, \sigma^2)$ para afirmar que X tem distribuição normal com média μ e variância σ^2 .

Como prova que $f(x)$ na expressão (3.7) é uma função densidade de probabilidade, vamos mostrar de acordo com (ROSS, 2009), que

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$$

Demonstração: Fazendo a substituição $y = (x - \mu)/\sigma$, temos que

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma e^{-y^2/2} dy.$$

Agora, precisamos mostrar que

$$\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi}.$$

Com esse objetivo, considere $I = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$. Então,

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{y^2+x^2}{2}} dy dx.$$

Avaliamos agora a integral dupla por meio de uma mudança de variáveis para coordenadas polares (isto é, $x = r\cos\theta$, $y = r\sin\theta$, e $dydx = r d\theta dr$). Assim,

$$I^2 = \int_0^{\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r \theta dr = 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr = 2\pi.$$

Com isso, $I = \sqrt{2\pi}$ e o resultado está demonstrado. ■

Podemos observar que $f(x; \mu, \sigma^2) \rightarrow 0$, quando $x \rightarrow \infty$, $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão de $f(x; \mu, \sigma^2)$, $x = \mu$ é ponto máximo de $f(x; \mu, \sigma^2)$. A densidade $f(x; \mu, \sigma^2)$ é simétrica em relação à reta $x = \mu$, isto é,

$$f(\mu + x; \mu, \sigma^2) = f(\mu - x; \mu, \sigma^2)$$

para todo x real.

De acordo com Ross (2009) se uma variável aleatória normal X for normalmente distribuída com parâmetros μ e σ^2 , então $Y = aX + b$ é normalmente distribuída com

parâmetros $a\mu + b$ e $a^2\sigma^2$. Assim $Z = \frac{X - \mu}{\sigma}$ é normalmente distribuída com parâmetros $\mu = 0$ e $\sigma^2 = 1$. Tal variável aleatória é chamada variável aleatória normal padronizada ou padrão. A função de distribuição acumulada de uma variável aleatória normal padrão é dada por

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{y^2}{2}} dy.$$

Não existe uma forma analítica simples para representar a função Φ . Por isso, ela é avaliada numericamente por meio de tabelas nos livros que tratam do assunto e por funções de bibliotecas de softwares estatísticos e de matemática, tais como *R* e *Matlab*, e *Maple* e *Geogebra*, respectivamente. Em geral, há tabelas para calcular $P(Z \leq z)$, $P(Z > z)$ e $P(0 \leq Z \leq z)$. Neste último tipo, os valores negativos de z , o valor $\Phi(z)$ pode ser obtido a partir da relação $\Phi(-z) = 1 - \Phi(z)$, $-\infty < z < \infty$. Essa equação diz que se Z é uma variável normal padrão, então

$$P(Z \leq -z) = P(Z > z), -\infty < z < \infty$$

Sendo X uma v.a. com função acumulada F , então podemos verificar que, a v.a. $Z = \frac{X - \mu}{\sigma}$ tem $E(Z) = 0$ e $Var(Z) = \sigma^2 = 1$. Podemos ainda afirmar que se $X \sim N(\mu, \sigma^2)$, então a v.a. $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$. Este resultado pode ser verificado diretamente pela obtenção da função acumulada

$$\Phi(z) = P(Z \leq z) = P(X \leq \mu + \sigma z) = F(\mu + \sigma z).$$

De onde, segue-se que $\phi(z) = \Phi(z)'$,

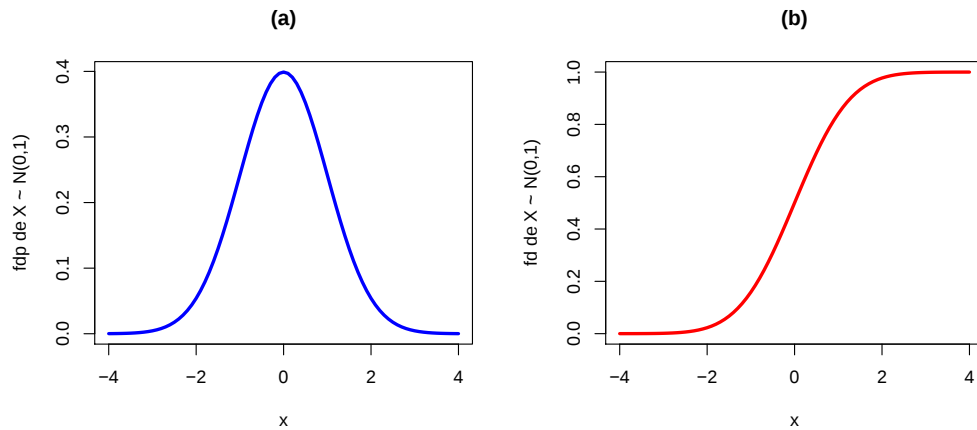
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty. \quad (3.8)$$

Para calcular $P(X \leq a)$ com uso de uma tabela para a v.a. $Z \sim N(0, 1)$ com os valores de $P(Z \leq z)$, procedemos do seguinte modo.

$$P(X \leq a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Os gráficos da fdp e da fda de uma v.a. normal padronizada estão representados na Figura 7.

Figura 7 – Distribuição normal padronizada. (a) Função densidade de probabilidade. (b) função acumulada.



Fonte: A autora.

O teorema limite de DeMoivre Laplace, diz que que n é grande, uma variável aleatória binomial com parâmetros n e p tem aproximadamente a mesma distribuição. Ao subtrair desta distribuição sua média np e então dividirmos o resultado pelo seu desvio padrão $\sqrt{np(1-p)}$, então a função distribuição dessa *v.a.* padronizada (média 0 e variância 1) convergirá para a função distribuição normal que $n \rightarrow \infty$ (ROSS, 2009).

4 Aplicações

4.1 Introdução

O futebol é o esporte mais praticado no Brasil e se tornou parte insubstituível de sua cultura. Das comunidades mais pobres até as mais ricas, desperta a alegria, a paixão e principalmente o sentimento de pertencimento a um grupo; a maior parte dos brasileiros já nascem com um time para torcer. Dentro e fora dos campos, ele movimenta a economia com altos salários e receitas e, torna a ida aos estádios ou o acompanhamento à distância um autêntico evento brasileiro. Formado por 22 jogadores, sendo 11 titulares e 11 reservas, tem como objetivo principal marcar mais gols que o time adversário. O tempo regulamentar é de 90 minutos podendo ter acréscimos ou em caso de competições estilo mata mata, pode haver prorrogações. Caso persista o empate, segue para a disputa por pênaltis que consiste em 5 rodadas para cada time, de confronto direto entre o jogador de um time A contra o goleiro de um time B , e vice-versa.

Neste capítulo, estamos interessados em aplicar as distribuições de probabilidade importantes em dados do futebol com o objetivo de incentivar o uso da Estatística em análises de desempenho de equipes de futebol. Neste trabalho, usaremos dados de jogos visando suas aplicações em fundamentos do futebol com uso do R ([R Core Team, 2020](#)).

Os dados da Copa do Mundo foram retirados de [FootStats \(2023\)](#), os do Brasileirão do [OGol \(2023\)](#) e do Clássico do Povo de [Timão \(2023\)](#).

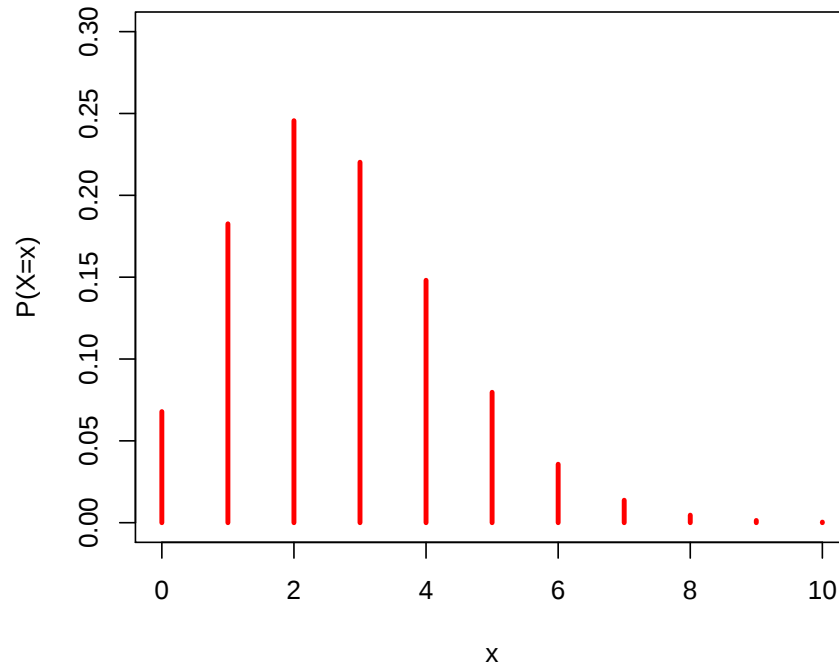
4.2 Copa do Mundo 2022

O Campeonato Mundial de Futebol (FIFA) é um torneio internacional de futebol masculino que acontece de quatro em quatro anos e, no ano de 2022 ocorreu no Catar sendo a primeira edição realizada no Oriente Médio. Formado por 32 equipes o evento é referência de audiência contando com bilhões de espectadores em suas edições.

Em 64 jogos foram 172 gols marcados com uma média de 2,69 gols por jogo. A probabilidade de sair dois gols em um jogo (independente da equipe) pode ser calculada pela distribuição de Poisson:

$$P(X = 2) = \frac{e^{-2,69}(2,69)^2}{2!} = 0,25$$

ou seja, em cada jogo disputado tinha-se a chance de 25% de ver dois gols serem marcados.

Figura 8 – Distribuição de probabilidade de Poisson com parâmetro $\lambda = 2,69$.

Fonte: Autora.

Ao analisar o gráfico acima podemos perceber que a medida que o número de gols cresce as probabilidades de ocorrência se tornam cada vez menores.

Considerando que 23% dessas partidas resultaram em um empate, vamos calcular com a distribuição de probabilidade geométrica, a probabilidade de em 5 partidas se obter um empate. Veja que a

$$P(X = 5) = 0.23(0.77)^{5-1} = 0.23(0.77)^4 \simeq 0.08$$

ou seja, a probabilidade de se obter um empate em 5 partidas é de 8%.

Agora vamos fazer a análise de alguns fundamentos futebolísticos da última partida disputada pelo Brasil na competição.

4.3 Análise de dados da partida Brasil x Croácia

O confronto foi pelas quartas de final entre a Croácia que é um país do sul da Europa e o Brasil, que é um país do sul da América Latina. O resultado foi marcado pela vitória da seleção croata nas penalidades máximas que eliminou a seleção brasileira da competição.

4.3.1 Número de Finalizações

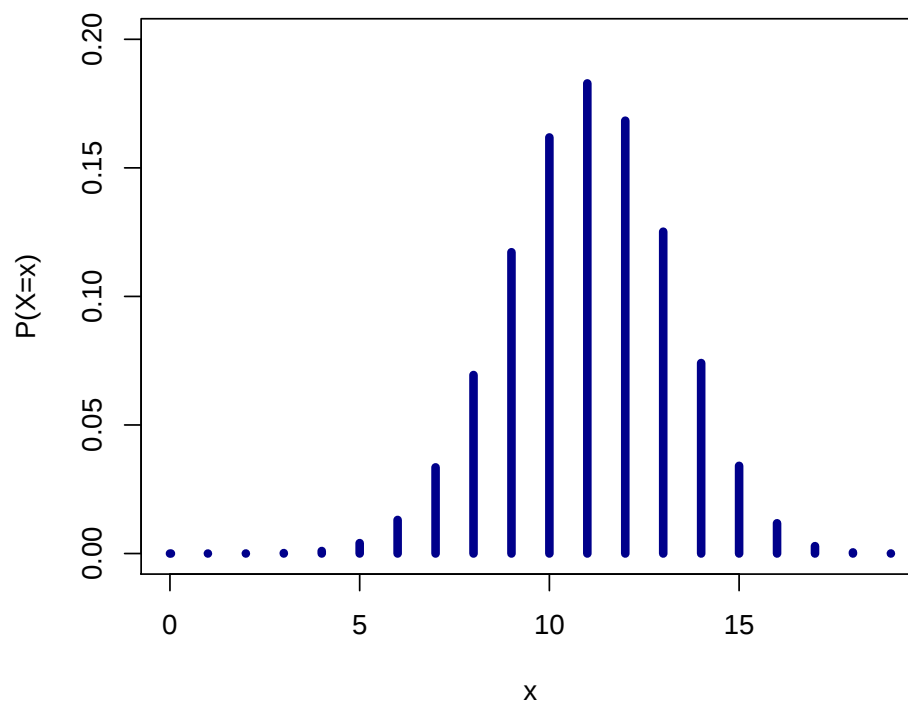
A finalização é quando o jogador lança a bola em direção ao gol, considerando possíveis desvios ou mesmo o erro do alvo. As finalizações dessa quarta de final está ilustrada na Tabela 4.3.1.

Finalizações	Certas	Erradas	Total
Brasil	11	8	19
Croácia	1	7	8

Das 19 finalizações o Brasil acertou 11 e errou 8, obtemos então a probabilidade de acerto por $p = \frac{11}{19} \cong 0,58$. Suponha que das 19 finalizações o Brasil obtenha 5 acertos, assumindo X como o número de finalizações certas e utilizando a distribuição de probabilidade binomial, temos:

$$P(X = 5) = \binom{19}{5} (0,58)^5 (0,42)^{11} = 0.004056347$$

Figura 9 – Distribuição de probabilidade de uma v.a. Binomial com parâmetros $n = 19$ e $p = 0,58$.

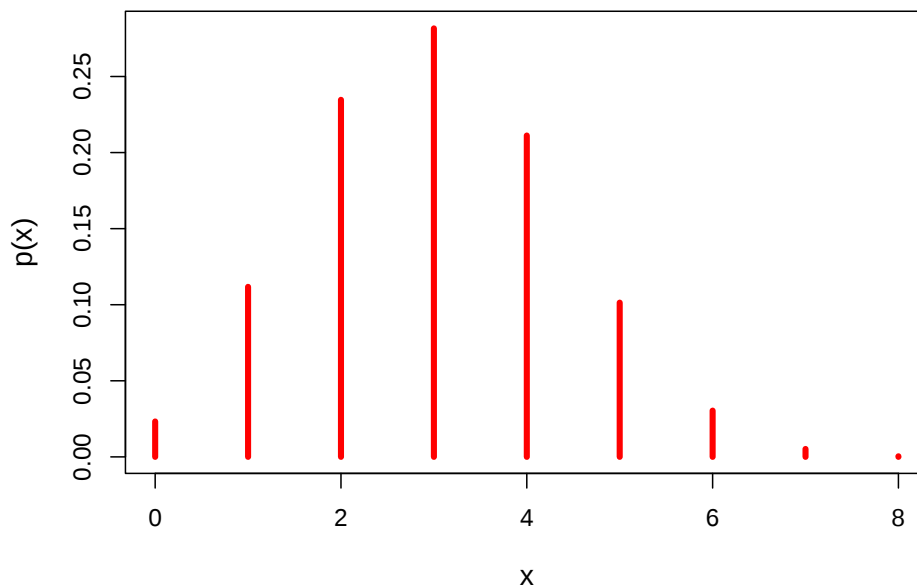


Fonte: Autora.

Por outro lado a Croácia, acertou 1 e errou 7 das suas 8 finalizações, com a sua $p = \frac{3}{8} = 0,375$. Suponha que ele acerte 5 finalizações de 8 com a mesma p .

$$P(X = 5) = \binom{8}{5} (0,375)^5 (0,625)^3 = 0.1013875$$

Figura 10 – Distribuição de probabilidade de uma v.a. Binomial com parâmetros $n = 8$ e $p = 0,375$.



Fonte: Autora.

4.3.1.1 Cruzamentos

O cruzamento é um passe longo que vem das linhas laterais, podendo ser rasteiro ou aéreo. Podemos aplicar a distribuição geométrica considerando calcular dentro de um número de 10 cruzamentos a possibilidade se obter o primeiro cruzamento certo. Para o Brasil consideramos $p = 0,2$ e $p = 0,8$ temos que a

$$P(X = 10) = (0.2)(0.8)^{10-1} \simeq 0.03$$

ou seja, era de 3% as chances da seleção acertar um cruzamento em número de 10 tentativas.

4.3.1.2 Tempo de jogo

A partida precisou ir para a prorrogação que consiste em mais 30 minutos de jogo após o tempo regulamentar, onde os dois times podem definir a partida. Vamos utilizar a

distribuição de Poisson para determinar a probabilidade de sair um gol em uma finalização certa da Croácia em um tempo de 4 minutos. Usaremos o $\lambda = 0.125$ que consiste na probabilidade de finalizações certas da Croácia de acordo com a Tabela 4.3.1. Dessa forma denotamos X como o número de minutos e calculamos $P(X = 4)$.

$$\begin{aligned} P(X \leq 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\ &= \frac{e^{-0.14}(0.14)^0}{0!} + \frac{e^{-0.14}(0.14)^1}{1!} + \frac{e^{-0.14}(0.14)^2}{2!} + \frac{e^{-0.14}(0.14)^3}{3!} + \frac{e^{-0.14}(0.14)^4}{4!} = 0.9999996 \end{aligned}$$

4.4 Brasileiro 2022

O campeonato brasileiro de futebol mais conhecido como "Brasileirão" foi disputado por 20 times e ocorreu do dia 9 de abril ao dia 13 de novembro de 2022. Essa foi a 66ª edição da principal divisão do futebol brasileiro conhecida por "Série A".

O campeão Palmeiras foi definido na 35ª rodada, isto porque matematicamente nenhum time alcançaria o seu total de pontos. Já o primeiro rebaixado à segunda divisão "Série B" foi o Juventude, visto que não era mais possível somar os pontos necessários para o tirar da zona de rebaixamento. Vamos aplicar algumas distribuições de probabilidade nos 4 melhores times que fizeram mais gols, passes certos, defesas, escanteios e que tiveram mais posse de bola.

4.4.1 Gols

A tabela abaixo indica os quatro times que mais fizeram gols. Eles estão distribuídos em gols marcados em casa e gols fora de casa.

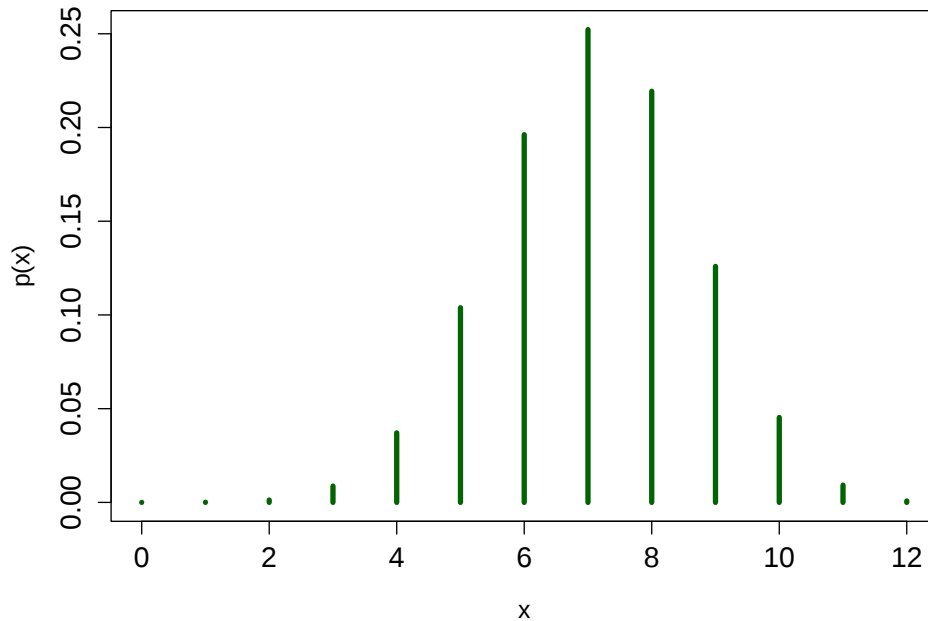
Gols	Casa	Fora	Total
Palmeiras	39	27	66
Fluminense	39	24	63
Flamengo	37	23	60
Internacional	40	18	58

Fonte: Autora.

O fator "jogar em casa" é considerado uma vantagem pela maioria dos torcedores. Pois com o apoio maior da torcida, o time consegue ter o foco que precisa em busca do melhor resultado. Assumindo X como o número de gols marcados em casa. Vamos considerar a extração de uma amostra $n = 12$ de gols do time campeão (Palmeiras). A probabilidade de 5 desses gols serem marcados dentro de casa pode ser calculada pela distribuição hipergeométrica a seguir.

- Palmeiras $P(X = 5) = \frac{\binom{39}{5} \binom{66-39}{12-5}}{\binom{66}{12}} = 0.1038598$

Figura 11 – Distribuição de probabilidade de uma v.a. Hipergeométrica com parâmetros $N = 66$, $n = 12$ e $p = 0.125$.



Fonte: Autora.

4.5 Clássico do Povo

O "clássico do povo" reúne as maiores torcidas do Brasil. Elas são representadas pelo time carioca (Flamengo) e pelo time paulista (Corinthians). As duas equipes batem recordes de público nos estádios com altas receitas em seus jogos.

4.5.1 Vitórias por estádio

No Maracanã, estádio considerado "casa" do Flamengo, as equipes somam 49 confrontos. Sendo 30 vitórias do Flamengo contra 11 vitórias do time adversário e 8 empates.

Já no Pacaembu, estádio onde o Corinthians atuou por 8 décadas, o confronto soma 47 jogos, 26 vitórias do time da casa, 8 empates e 13 vitórias do time carioca. Ambos os times indicam que jogar em casa traz na maioria das vezes, bons resultados.

Vamos agora aplicar a distribuição binomial.

Assumindo uma v.a X como sendo o número de vitórias fora de casa do Flamengo. Adotamos $p = 0.28$ do time carioca vencer. Vamos calcular a probabilidade de em 20 jogos fora de casa obtermos mais de 2 vitórias.

$$P(X > 2) = 1 - (P(X = 0) + P(X = 1) + P(X = 2)) = 0.05258043.$$

E a probabilidade ocorrer mais de 3 empates em 10 jogos no Pacaembu é calculada a seguir. (Considere $p = 0.17$).

$$P(X > 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)) = 0.9258528$$

Podemos observar, portanto, que em 20 jogos fora casa do Flamengo conta-se com uma probabilidade de 5% de vitória. Já o Corinthians em 10 jogos em casa, conta-se com a possibilidade de 90% de chance de empate. Assim, podemos concluir que time paulista conta com chances muito positivas quando joga em casa, já o time carioca tem chances negativas quando joga fora de casa.

5 Considerações Finais

As aplicações do Capítulo 4 são importantes para destacar a importância do estudo das distribuições de probabilidade no contexto do futebol. Ao analisar as estatísticas de cada jogo contribuimos para uma maior assertividade em apostas, análises táticas e no estudo das variáveis aleatórias.

Apostas em partidas de futebol tem atraído muitas pessoas, principalmente pelo amplo mercado na internet de casas de apostas. Atribuídos pelos apostadores, dados das probabilidades de vitória, empate ou derrota dos times podem ser incorporados nos modelos de previsão. Considere, por exemplo, o Campeonato Brasileiro da serie *A*, estatísticas dos jogos até uma determinada rodada podem ser usadas para a previsão de classificação, rebaixamento e título por equipe.

Referências

- ARA, A. B.; MUSETTI, A. V.; SCHNEIDERMAN, B. *Introdução à estatística*. [S.l.]: Editora Blucher, 2003.
- FOOTSTATS. *FootStats*. 2023. Disponível em: <<https://www.footstats.com.br/#/>>. Acesso em: 02 de março 2023.
- MEYER, P. L. *Probabilidade: aplicações à estatística*. [S.l.]: Livros Técnicos e Científicos Rio de Janeiro, 1983.
- MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. São Paulo: Saraiva Educação SA, 2017.
- OGOL. *O gol*. 2023. Disponível em: <<https://www.ogol.com.br/>>. Acesso em: 20 de maio 2023.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>.
- ROSS, S. *Probabilidade: um curso moderno com aplicações*. [S.l.]: Bookman Editora, 2009.
- SANTOS, R. d. Futebol e sua história: possibilidade de efetivação da proposta crítico superadora. 2014.
- TIMÃO. *Timão*. 2023. Disponível em: <<https://www.meutimao.com.br/>>. Acesso em: 20 de maio 2023.