UNIVERSIDADE FEDERAL RURAL DO SEMIÁRIDO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

RAFAEL DO NASCIMENTO TOMÉ RIBEIRO

PROBABILIDADE E ESTATÍSTICA APLICADAS AO ESPORTE

Mossoró- RN
2016

Rafael do Nascimento Tomé Ribeiro

PROBABILIDADE E ESTATÍSTICA APLICADAS AO ESPORTE

Dissertação apresentada à Universidade Federal Rural do Semiárido – UFERSA, campus Mossoró para obtenção do título de mestre em Matemática.

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei n° 9.279/1996 e Direitos Autorais: Lei n° 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

R484p Ribeiro, Rafael do Nascimento Tomé Ribeiro.
Probabilidade e Estatística aplicadas ao Esporte / Rafael do Nascimento Tomé Ribeiro Ribeiro. - 2016.
67 f. : il.

Orientador: Maurício Zuluaga Martinez Martinez.
Dissertação (Mestrado) - Universidade Federal Rural do Semi-árido, Programa de Pós-graduação em Matemática, 2016.

1. Probabilidade. 2. Estatística. 3. Esporte. 4. Futebol. I. Martinez, Maurício Zuluaga Martinez, orient. II. Título.

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC´s) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

RAFAEL DO NASCIMENTO TOMÉ RIBEIRO

PROBABILIDADE E ESTATÍSTICA APLICADAS AO ESPORTE

Dissertação apresentada a Universidade Federal Rural do Semiárido – UFERSA, Campus Mossoró para obtenção do título de Mestre em Matemática.

APROVADO EM: 22 / 06 / 2016

BANCA EXAMINADORA

Mauricio Zuluaga M

Dr. Maurício Zuluaga Martinez - UFERSA

Presidente

Dr. Elmer Rolando Llanos Villarreal- UFERSA

Primeiro Membro

Aleksandre Saraiva Dantas

Dr. Aleksandre Saraiva Dantas- IFRN

Terceiro Membro

MOSSORÓ/RN, 2016.

DEDICATÓRIA

Dedico esse trabalho, principalmente, à minha filha, minha esposa, meus familiares e amigos que acompanharam o meu empenho para desenvolvê-lo com responsabilidade e aprimoramento.

.

AGRADECIMENTOS

A Deus, por ter me dado forças e proporcionar muitas graças em minha vida.

À minha esposa, Ana Karla Maranhão da Silva e minha filha, Sofia Maranhão Ribeiro, por me darem força nos momentos mais difíceis, pela compreensão e sempre estarem ao meu lado.

Aos meus familiares, principalmente ao meu pai, Israel Barbosa Ribeiro, professor de matemática, pelo exemplo e incentivo que sempre me deu.

A todos os funcionários da UFERSA, principalmente aos professores do programa, pela sabedoria na condução do processo de ensino-aprendizagem.

Aos meus colegas, pelo convívio e amizade.

"Este jogo não é matemática. É futebol, e no futebol dois mais dois raramente dá quatro. Às vezes é três. Com frequência é cinco"
Leo Beenhakker

Educar é impregnar de sentido o que fazemos a cada instante!”
Paulo Freire

“O futebol é a coisa mais importante, dentre as coisas menos importantes”
Nelson Gonçalves

“Alguns usam a estatística como os bêbados usam postes: mais para apoio do que para iluminação”
Andrew Lang

RESUMO

O presente trabalho tem como principal objetivo estudar a importância da aplicação da probabilidade e da estatística no esporte, especificamente, no futebol. A pesquisa pode ser utilizada para satisfazer a curiosidade dos interessados pelo esporte, como a probabilidade e a estatística são calculados e utilizados para o seu desenvolvimento, respaldado em uma pesquisa bibliográfica mostrando os conceitos básicos, definições, apresentação de um modelo para o cálculo das probabilidades no futebol, podem ser captados os dados estatísticos e como eles são obtidos. De certo, essa pesquisa pode contribuir para o estudo complementar de professores, alunos e demais interessados no assunto.

Palavras chaves: Probabilidade, Estatística, Esporte e Futebol.

ABSTRACT

The present work has as main purpose to study the importance of applying probability and statistics in sports, especially in football. It can be used to satisfy the curiosity of those interested in the sport, how the probability and statistics are calculated and used for the development of this sport ,based on a bibliographic research showing the basic concepts, definitions, presentation of a model for the calculation of probabilities in football, statistics data can be collected and showed how they are obtained. Indeed, this research may contribute to the further study of teachers, students and others interested in the subject.

Key words: Probability, Statistics, Sport and Football.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

INTRODUÇÃO

De um modo geral, a Estatística e a Probabilidade possibilitam, através da coleta de dados, que pesquisas sejam transformadas em informações úteis para a tomada de decisões, seja qual for a área de atuação.

Com a evolução tecnológica, nas últimas décadas, o esporte, sobretudo o futebol, cada vez mais utiliza estudos estatísticos e probabilísticos, por meio da coleta, organização e análise de dados para técnicos, dirigentes e até torcedores tomarem decisões coerentes e realizarem projeções adequadas para suas equipes.

Nesse contexto, essa pesquisa apresenta, inicialmente, uma pequena evolução histórica da Estatística e da Probabilidade. Em seguida, expõe os primeiros conceitos de Probabilidade e Estatística, assim como, as definições de população, amostra e variável; para consequentemente chegar à forma de desenvolvimento da coleta de dados, sua representação gráfica e análise, recorrendo ao estudo das medidas de posição e dispersão.

O objetivo é mostrar aplicações de Probabilidade e Estatística, com exemplos e problemas voltados ao esporte, especificamente ao futebol. É importante ressaltar a necessidade da utilização de recursos tecnológicos, por exemplo, softwares, planilhas eletrônicas e aplicativos.

Portanto, para uma melhor compreensão da pesquisa, o estudo foi dividido em cinco capítulos, contemplando os seguintes assuntos: o primeiro capítulo relata a parte histórica de estatística e probabilidade, também sua evolução no esporte. Já o segundo trata da fundamentação, definição intuitiva e formal de probabilidade, assim como sua aplicação no futebol. O terceiro capítulo aborda os principais conceitos de estatística usando exemplos voltados ao esporte. Enquanto o quarto capítulo mostra as principais medidas de posição ou tendência central e as de dispersão que servem para validar as de tendências centrais. O quinto capítulo apresenta aplicações com o uso de alguns softwares e sites, mostrando como eles captam os dados estatísticos no esporte para que leitor possa satisfazer suas curiosidades, utilizar e ensinar a outros interessados nos assunto.

Com o intuito de ressaltar o principal propósito desse trabalho, as considerações finais mostram a relevância do resultado dessa pesquisa e sua aplicabilidade nos estudos.

1. EVOLUÇÃO HISTÓRICA

Milênios antes de Cristo surge a estatística, concomitantemente com a organização da sociedade em Estados. Manifesta-se com o objetivo dos governantes conhecerem como os recursos estavam distribuídos entre a população e do que dispunha o Estado, os imperadores chineses, romanos, faraós egípcios, civilizações maias já utilizavam conhecimentos estatísticos em censos, com fins de coletas de impostos e militares.

Para Dante (2013) a palavra estatística significa análise de dados, sendo originada da palavra *status* em latim, que significa estado.

Segundo Lima (2009), foi no século XVIII que a Estatística recebeu essa nomenclatura, mas continuava a ser sistematizada com objetivo meramente descritivo pelos alemães e pelos matemáticos sociais, que buscavam enumeração, organização e fenômenos demográficos. A Estatística foi considerada como ramo da Matemática aplicada, ou seja, independente das demais, no inicio do século XX, sendo utilizada pela ciência para produzir pesquisas feitas com cautela e precisão. Entretanto, com a revolução tecnológica, a Estatística foi aplicada nas mais diversas áreas, incluindo o esporte, visto que o computador ajuda a fazer os cálculos e colher os dados para uma pesquisa estatística.

Já a Probabilidade nasceu com o estudo dos possíveis resultados de jogos de azar, como roleta e baralho. Ao passar do tempo a Teoria da Probabilidade com problemas genéricos evoluiu, sendo muito utilizada em estatística, genética, meteorologia, esporte, etc.

A estatística básica sempre esteve presente nos esportes, através das tabelas de classificação, número de finalizações, escanteios, faltas, cartões, posse de bola e índices para classificação para as Olimpíadas.

Um dos primeiros a utilizar os conceitos avançados de Estatística no esporte foi Billy Beane, dirigente de um time de basebol dos Estados Unidos da América chamado *Oakland Athletics* (1977), com auxilio do Economista Peter Brand.

O diferencial de Beane foi conseguir montar um time competitivo, com um pequeno orçamento, usando técnicas estatísticas sofisticadas na escolha dos jogadores, de comprar e vender em cada posição. Com o sucesso de Beane, outros times de beisebol, além de montar o time, passaram a usar a estatística também para escolher a forma de treinamento, posição tática e até a melhor escalação, analisando seu adversário, em determinados jogos. Em 2012, foi lançado um filme sobre esse pioneirismo de Beane, com o título de “Money Ball”, versão nacional intitulada “O homem que mudou o jogo”.

Atualmente, com a revolução tecnológica, a Estatística é aplicada às mais diversas modalidades esportivas como natação, voleibol, tênis, futebol, e outros. Tendo o profissional de estatística na comissão técnica de praticamente todos os atletas de ponta e equipes utilizou-se de câmeras e computadores nos treinamentos de atletas para colher os dados estatísticos e estudá-los.

Além disso, a probabilidade e estatística são muito utilizadas pela imprensa, por exemplo, mostrando aos torcedores os dados das partidas e evolução das equipes ao longo das competições, e as possibilidades de seus times serem campeões, conquistarem uma vaga para a Copa Libertadores da América ou de livrar-se do rebaixamento.

A Probabilidade também é muito usada em casas e sites de aposta, pois um resultado pouco provável pode trazer lucros ao apostador que um resultado com grandes chances de acontecer.

2. PROBABILIDADE

2.1 Introdução

Segundo Lipschutz (1993), Probabilidade é o ramo da matemática que estuda os eventos não determinísticos ou aleatórios. Por exemplo, se o árbitro da partida lança uma moeda ao ar é certo que ela cairá, mas não é certo que aparecerá como resultado coroa na face superior da moeda. Entretanto, suponha que repetimos esse evento de lançar a moeda n vezes; seja s o número de sucessos, ou seja, a quantidade de vezes que o resultado é coroa. Observa-se, empiricamente, que a razão entre s e n, $\frac{s}{n}$, chamada também de frequência relativa, tende a estabilizar-se, isto é, aproximar se de um limite. Essa é à base da teoria da probabilidade.

Na teoria da probabilidade, definimos um modelo matemático para um fenômeno e precisamos verificar a validade desse fenômeno, ou seja, ver se as probabilidades associadas a ele estão bem próximas dos limites das frequências relativas, o que origina testes de confiança, os quais constituem o assunto de probabilidade.

A probabilidade p de um evento A ocorrer, representada por $p(A)$, assim foi definida: Se o evento A pode ocorrer de s maneiras distintas dentre um total de n maneiras supostamente equiprováveis então:

$$p(A) = \frac{s}{n}$$

Exemplo 2.1: no lançamento de um dado, um resultado par pode ser obtido de três maneiras (2, 4 ou 6) dentre os seis resultados igualmente possíveis (1, 2, 3, 4, 5 ou 6). Pressupondo que o dado não é viciado. Logo

$$p(Par) = \frac{3}{6} \implies p(par) = \frac{1}{2} \text{ ou } p(par) = 50\%.$$

2.2 Espaço Amostral e Eventos

Espaço amostral é o conjunto de todos os resultados possíveis. No exemplo acima $S = (1, 2, 3, 4, 5, 6)$.

Ponto amostral é um resultado particular, ou seja, um elemento de $S.$ No exemplo acima o resultado 4 é um ponto amostral.

Um evento A é um conjunto de resultados favoráveis, ou seja, um subconjunto de $S.$ No exemplo acima o resultado par nos fornece o evento $A = (2, 4, 6)$.

2.3 Axiomas de Probabilidades

i. ∀ (para todo) evento *A*, $0 \leq p(A) \leq 1$.
ii. $p(S) = 1$
iii. Se A e B são eventos mutuamente exclusivos, ou seja, $A \cap B = \emptyset$, então $p(A \cup B) = p(A) + p(B)$.

Teorema 2.3.1: $p(\emptyset) = 0$.

Prova: $p(A) = p(A \cup \emptyset)$, como $A \cap \emptyset = \emptyset$.

$p(A) = p(A) + p(\emptyset)$

$p(A) - p(A) = p(\emptyset)$

$0 = p(\emptyset)$. ■

Exemplo 2.2: Qual a probabilidade de um time da região Norte ser campeão da série A do Campeonato Brasileiro de futebol de 2015?

Solução: Sendo *A* o evento time da região Norte ser campeão da série A do Campeonato Brasileiro de futebol de 2015. O conjunto de resultados favoráveis é vazio, pois não tinha nenhum time da região Norte na serie A do Campeonato Brasileiro de futebol de 2015, logo:

$p(A) = 0.$

Teorema 2.3.2: Se A^c é complemento de A, então $p(A^c) = 1 - p(A)$.

Prova: Como $S = A \cup A^c$, A e A^c são mutuamente exclusivos, então:

$p(S) = 1$

$p(A \cup A^c) = 1$

$p(A) + p(A^c) = 1$

$p(A^c) = 1 - p(A)$. ■

Exemplo 2.3: Se a probabilidade de um time A ser rebaixado é de 90%, então a probabilidade da equipe não ser rebaixada é de *10%.*

Teorema 2.3.3: Se $A \subset B$, então $p(A) \leq p(B)$.

Prova: $A \subset B \Longrightarrow B = A \cup (B \backslash A)$, com $A \cup (B \backslash A) = \emptyset$

$p(B) = p(A \cup (B \backslash A))$

$p(B) = p(A) + P(B \backslash A)$

Se $A = B$, então $(B \backslash A) = \emptyset$ e $p(A) = p(B)$.

Se $A \neq B$, então $(B \backslash A) \neq \emptyset \Rightarrow p(A \backslash B) > \emptyset \Rightarrow p(B) > p(A)$.

Portanto $p(A) \leq p(B)$. ■

Teorema 2.3.4: Se A e B são eventos quaisquer, então

$p(A \backslash B) = p(A) - p(A \cap B)$.

Prova: $A = (A \backslash B) \cup (A \cap B)$, como $(A \backslash B)$ e $(A \cap B)$ são eventos mutuamente exclusivos, então:

$p(A) = p[(A \backslash B) \cup (A \cap B)]$

$p(A) = p(A \backslash B) + p(A \cap B)$

$p(A \backslash B) = p(A) - p(A \cap B)$. ■

Teorema 2.3.5: Se A e B são eventos quaisquer, então:

$p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Prova: $A \cup B = (A \backslash B) \cup B$, como $(A \backslash B)\ e\ B$ são eventos mutuamente exclusivos, então:

$p(A \cup B) = p[(A \backslash B) \cup B]$

$p(A \cup B) = p(A \backslash B) + p(B)$, do teorema 4 temos que:

$p(A \cup B) = p(A) - p(A \cap B) + p(B)$

$p(A \cup B) = p(A) + p(B) - p(A \cap B)$ ■

2.4 Espaços de Probabilidades Finitos

Seja S um espaço amostral finito, $S = (a_1, a_2, \dots, a_n)$, um espaço de probabilidade é obtido associando-se a cada ponto $a_i \in S$, um número real p_i chamado de probabilidade de a_i.

i. Cada $p_i \geq 0$
ii. $\sum_{i=1}^{n} p_i = 1$, ou seja, $p_1 + p_2 + p_3 + \cdots + p_n = 1$

Exemplo 2.4: Um time que tem como probabilidades 45% de vitória, 35% de empate e 20% de derrota. Note que: $45\% + 35\% + 20\% = 1$.

2.5 Espaços Finitos Equiprováveis

Chamamos de espaço equiprovável ou uniforme, quando nesse espaço de probabilidade finito S, cada ponto tem probabilidades iguais. Em particular, quando S contém n pontos cada ponto tem como probabilidade $\frac{1}{n}$**.** Dessa forma, quando um evento A tem s pontos $p(A) = \frac{s}{n}$, ou seja,

$$p(A) = \frac{|A|}{|S|} = \frac{números\ de\ elementos\ de\ A}{número\ de\ elementos\ de\ S}.$$

Portanto, a probabilidade de A ocorrer pode ser calculada assim:

$$p(A) = \frac{números\ de\ eventos\ favoravéis}{número\ de\ eventos\ possivéis}$$

Exemplo 2.5: Selecione uma carta de um baralho comum com 52 cartas. Qual a probabilidade de:

i. Retirar uma carta de espadas?
ii. Retirar uma dama?
iii. Retirar uma dama de espadas?

Resolução: chamando de $A = \{retirar\ uma\ carta\ de\ espadas\}$, $B = \{retirar\ uma\ dama\}$ e $A \cap B = \{retirar\ uma\ dama\ de\ espadas\}$. Como temos um espaço equiprovável,

i. $p(A) = \frac{números\ de\ cartas\ de\ espadas}{número\ de\ cartas} = \frac{13}{52} = \frac{1}{4}$
ii. $p(B) = \frac{números\ de\ damas}{número\ de\ cartas} = \frac{4}{52} = \frac{1}{13}$
iii. $p(A \cap B) = \frac{números\ de\ damas\ de\ espadas}{número\ de\ cartas} = \frac{1}{52}$

2.6 Probabilidades no futebol

Como calcular a probabilidade de uma equipe vencer uma determinada partida de futebol?

Em uma partida de futebol temos como espaço amostral a vitória, o empate, e a derrota. É obvio que esse é um espaço amostral não equiprovável, pois há vários fatores que influenciam as probabilidades do resultado de um jogo de futebol, dentre eles temos o mando de campo, desfalques, campanhas das equipes no campeonato, histórico dos jogos entre as equipes, dentre outros. Tudo isso, pode tornar complexa a criação de um modelo matemático para calcular a probabilidade desse evento. Outra dificuldade é a grande frequência de zebras no futebol, mesmo assim, pode se afirmar que há uma lógica por trás das probabilidades.

Segundo Costa (2009), uma modelagem matemática bastante simples, que é usada no site “probabilidades no futebol” da UFMG, é levar em consideração apenas a história das equipes dentro do próprio campeonato, para simplificar e evitar preconceitos, tabus e o peso da camisa de cada time. Nesta modelagem o mando de campo é o fator mais importante, logo ele é utilizado para o cálculo das probabilidades de vitória (v), empate *(e)* e de derrota *(d)*. Dessa forma, vencer uma partida aumenta a probabilidade de vencer outras, perder aumenta a probabilidade de perder e empatar uma aumenta a probabilidade do empate.

Vamos ao modelo para o cálculo da probabilidade de uma partida. Temos dois trios de números, um para o mandante (v_A, e_A, d_A) e outro para visitante (v_B, e_B, d_B). Cada trio de números que descreve o perfil, frequência relativa dentro do campeonato, que pode ser pensado como a sua tendência a vencer, empatar ou perder jogos em casa ou fora. É bom ressaltar que a soma de cada trio é 1 e que cada um desses números é positivo. Quando o time A recebe o time B para jogar “em sua casa”, o número v que representa a probabilidade de vitória do mandante será inferido, fazendo a média aritmética, entre a tendência de vitória expressa pelo perfil de mandante de A e a tendência de derrota do perfil de visitante de B. De maneira análoga, obtém-se e e d fazendo a média das tendências de empate e derrota. Assim, calculamos probabilidades das três alternativas consideradas, com soma igual a 1.

Prova: como a soma de cada trio é igual a 1, ou seja, $v_A + e_A + d_A = 1$ e $v_B + e_B + d_B = 1$ e as probabilidades do mandante A de vitória v, derrota d empate e *são* obtidas por $v = \frac{v_A + d_B}{2}$, $d = \frac{v_B + d_A}{2}$ e $e = \frac{e_A + e_B}{2}$, então:

$$v + e + d = \frac{v_A + d_B}{2} + \frac{e_A + e_B}{2} + \frac{v_B + d_A}{2}$$

$$v + e + d = \frac{v_A + e_A + d_A + v_B + e_B + d_B}{2}$$

$$v + e + d = \frac{1 + 1}{2}$$

$$v + e + d = \frac{2}{2}$$

$$v + e + d = 1 \blacksquare$$

Exemplo 2.6: Se o time A vencesse 60% das partidas como mandante e o time B perdesse 80% como visitante, a probabilidade de vitória do time A será de 70%.

É comum, nas últimas rodadas do Campeonato Brasileiro de Futebol, vermos em noticia que a probabilidade de uma equipe A ser campeã, é x%, e ser classificada para a Copa Libertadores da América é y% e a probabilidade de uma equipe B ser rebaixada para a série B é de z%. A pergunta é, como esses cálculos são feitos?

O site “probabilidades no futebol” da UFMG realiza um grande número de simulações, a partir da situação atual do campeonato, usando o modelo matemático supracitado. Feitas as simulações, contam quantas terminaram com o resultado favorável (tal time campeão ou tal time rebaixado) e divide pelo número total de simulações. A simulação usa o sorteio de um número aleatório para definir o vencedor de cada partida. Esse número aleatório é sorteado como um número positivo menor que, ou igual a 1. Se considerar que o mandante de certo jogo (usando o modelo acima descrito) tem 50% de chance de vitória, 30% de chance de empate e 20% de chance de derrota, ele é vitorioso nessa simulação. Se o número aleatório sorteado for menor ou igual a 0,5, consideraremos o resultado como empate, se o número maior que 0,5 e menor ou igual 0,8 e consideraremos o resultado de vitória do visitante se aquele número aleatório estiver acima de 0,8. De modo mais geral, para cada partida, temos um número $v < 1$ que dá a probabilidade da vitória do mandante, outro número $e < 1 - d$ que dá a probabilidade do empate. Sorteamos um número aleatório x entre 0 e 1, se $x \leq v$ o mandante venceu este jogo nesta simulação; se $v < x \leq e + v$, o é considerado empate; se $x > v + e$, o visitante é que venceu. Cada rodada do Campeonato é simulada fazendo as 10 partidas desta forma. Depois de simular uma rodada, atualizam-se todas as probabilidades de todos os times, de acordo com os resultados da rodada, e simula-se a rodada seguinte.

Segundo Costa (2009) em seu artigo Probabilidades no Futebol página 4:

> *"São simuladas não só mil, mas milhares e até milhões delas se for necessário. O computador faz isto em segundos ou quando muito em minutos. Tudo o que temos de fazer é ensina-lo a sortear".*

Vale ressaltar que quando a simulação diz que uma equipe tem 99% de chances de vencer um título ou de ser rebaixada, não há garantia de que isso acontecerá com a equipe. Em 2009, o Fluminense tinha 99% de chances de ser rebaixado para a Série B, mas em uma reação espetacular nas últimas rodadas se salvou. A simulação errou? A resposta "é não, pois em momento algum, ela afirmou que o time já estava rebaixado". Era uma chance de se manter na série A e 99 de rebaixamento, ou seja, a cada cem simulações feitas pelo modelo acima descrito o fluminense tinha uma chance de continuar na sua posição.

Exemplo 2.7: Antes da última rodada do Campeonato Brasileiro de Futebol de 2015 a situação era a seguinte:

Tabela 1: Classificação do Campeonato Brasileiro de futebol de 2015, após a 37° rodada.

Classificação do Campeonato Brasileiro de futebol de 2015 após a 37° rodada.													
C	EQUIPE	**P**	J	V	E	D	GP	GC	SG	VM	VV	DM	DV
14º	CHAPECOENSE - SC	**47**	37	12	11	14	34	41	-7	9	3	5	9
15º	CORITIBA - PR	**43**	37	11	10	16	31	42	-11	6	5	5	11
16º	AVAÍ - SC	**41**	37	11	8	18	37	59	-22	8	3	6	12
17º	FIGUEIRENSE - SC	**40**	37	10	10	17	35	50	-15	6	4	6	11
18º	VASCO DA GAMA - RJ	**40**	37	10	10	17	28	54	-26	5	5	8	9
19º	GOIÁS - GO	**38**	37	10	8	19	39	48	-9	7	3	7	12
20º	JOINVILLE - SC	**31**	37	7	10	20	26	46	-20	6	1	5	15

Fonte: www.cbf.com.br/competicoes/brasileiro-serie-a/classificacao visualização em 30/11/2015

Na 37° rodada matematicamente, o Joinville foi rebaixado e para derradeira rodada as últimas três vagas no Z4 (grupo dos quatro times rebaixados) os times Coritiba, Avaí, Figueirense, Vasco e Goiás lutaram para não ficarem rebaixados. Enfrentavam-se na última rodada Figueirense x Fluminense em Santa Catarina, Corinthians x Avaí em São Paulo, Coritiba x Vasco em Curitiba e Goiás x São Paulo em Goiânia.

No programa Fantástico da Rede Globo, exibido em 29/11/2015, no quadro os “gols do fantástico” o apresentador Tadeu Schmidt informou que segundo o matemático Tristão Garcia, o Vasco e Goiás tinham 88%, Avaí 66%, Figueirense 57% e Coritiba 1% de chances de rebaixamento. Note que a soma das probabilidades é 300%, pois tinham três vagas. Não sabemos qual é o modelo matemático usado para o cálculo das probabilidades, por Tristão Garcia, em seu site infobola, mas para fazer uma comparação vamos calcular a probabilidade utilizando o modelo apresentado acima.

i) Para o Goiás não ser rebaixado ele deveria vencer sua partida e torcer para que Avaí perdesse e que Vasco e Figueirense não vencessem.

A probabilidade de o Goiás vencer era $\frac{VM+DV}{2} = \frac{\frac{7}{18}+\frac{5}{18}}{2} \cong 33{,}3\%$.

A probabilidade de o Vasco não vencer, pelo teorema 2.2, é igual a um menos a probabilidade de ele vencer $1 - \frac{VV+DM}{2} = \frac{\frac{5}{18}+\frac{5}{18}}{2} \cong 72{,}2\%$.

A probabilidade de o Avaí perder era de $\frac{VM+DV}{2} = \frac{\frac{16}{18}+\frac{12}{18}}{2} \cong 77{,}7\%$.

A Probabilidade de o Figueirense não vencer também pelo teorema 2.2 é igual a um menos a probabilidade de ele vencer $1 - \frac{VM+DV}{2} = 1 - \frac{\frac{6}{18}+\frac{12}{18}}{2} = 50\%$.

Pelo princípio multiplicativo a probabilidade de o Goiás não ser rebaixado era dada por $33{,}3\% \cdot 72{,}2\% \cdot 77{,}7 \cdot 50\% = 9{,}3\%$. Logo o Goiás pelo nosso modelo tinha 90,7% de chances de ser rebaixado para a segunda divisão, quase três pontos percentuais a mais que o Modelo usado por Tristão.

ii) Para o Vasco não ser rebaixado, é necessário que ele ganhe a partida, como também os times Avaí e Figueirense não vençam.

A probabilidade de o Vasco vencer é o um menos a probabilidade dele não vencer $1 - 72{,}2\% = 27{,}8\%$.

A probabilidade de o Avaí não vencer pelo teorema 2.2 é igual a um menos a probabilidade de ele vencer $1-\frac{VV+DM}{2}=1-\frac{\frac{3}{18}+\frac{1}{18}}{2}\cong 88{,}9\%$.

A Probabilidade de o Figueirense não vencer já vimos que é 50%.

Pelo princípio multiplicativo a probabilidade de o Vasco não ser rebaixado era dada por $27{,}8\%\cdot 88{,}9\%\cdot 50\%=12{,}3\%$. Logo o Vasco da Gama tinha 87,7% de chances de ser rebaixado para a segunda divisão. Observe que as duas probabilidades são bem próximas e saiba que a divulgação na televisão usa geralmente o resultado inteiro mais próximo.

iii) Para o Figueirense não ser rebaixado ele deveria vencer sua partida e torcer para que Avaí não vencesse ou o Coritiba perdesse, e obtendo um saldo de quatro gols de diferença.

Sabe-se que a probabilidade de o Figueirense vencer era de 50%, a do Avaí não vencer era de 88,9% e a do Coritiba perder era mesma do vasco vencer $27{,}8\%$, no entanto, o Figueirense deveria ganhar por três gols de diferença como se fosse três vitórias.

Dessa forma, a probabilidade do Figueirense não ser rebaixado era dada por $50\%\cdot 88{,}9+50\%\cdot 50\%\cdot 50\%\cdot 27{,}8\%=48{,}0\%$, pois o time, pelo nosso modelo tinha 52% de chances de ser rebaixado para a segunda divisão.

iv) Para o Avaí não ser rebaixado, só precisava vencer sua partida; mas na condição de empate, teria que contar com a derrota do Vasco e Figueirense, e se não vencesse contaria com a derrota do Vasco, do Figueirense e do Goiás.

Portanto, a probabilidade de o Avaí não ser rebaixado era dada por $11{,}1\%+11{,}2\%\cdot 72{,}2\%\cdot 50\%+77{,}7\%\cdot 72{,}2\%\cdot 50\%\cdot 66{,}7\%=33{,}9\%$. Logo, o Avaí pelo nosso modelo tinha 66,1% de chances de ser rebaixado para a segunda divisão maior que a do Figueirense mesmo estando na frente na classificação.

v) Para o cálculo da probabilidade do Coritiba não ser rebaixado fica mais fácil subtrair de um a chances dele cair. Que seria derrota do Coritiba, vitória de Avaí e Figueirense, sendo que a vitória do Figueirense seja por três gols de diferença.

Portanto a probabilidade de o Coritiba ser rebaixado era dada por $27{,}8\% \cdot 11{,}1\% \cdot 50\% \cdot 50\% \cdot 50\% = 0{,}3\%$.

Para comparar os dois modelos, vamos arredondar as porcentagens para números inteiros, arredondando sendo que a porcentagem do Coritiba será um.

Tabela 2 - Comparação entre os modelos de probabilidade de ser rebaixado.

Comparação entre os modelos de probabilidade de ser rebaixado		
Time	Tristão	Costa
Goiás	88%	91%
Vasco	88%	88%
Figueirense	57%	52%
Avaí	66%	66%
Coritiba	1%	1%
Total	300%	298%

Note que a soma das probabilidades que fizemos aqui deu 298% por causa dos sucessivos arredondamentos, esse problema não existiria se fosse feito as simulações em um computador, pois o computador não está preocupado se o time A ou B vai ou não ser rebaixado ele não faz essas combinações que fizemos, o que ele faz é milhares de simulações da rodada final e mostra os resultados. Por exemplo, se ele diz que um time tem 88% de chances de cair, isso quer dizer que a cada 100 simulações o time é rebaixado em 88.

Os resultados dos jogos dos times que lutavam para não cair foram os seguintes: Figueirense 1 x 0 Fluminense, Corinthians 1 x 1 Avaí, Coritiba 0 x 0 Vasco e Goiás 0 x 1 São Paulo. Resultados que rebaixaram Goiás, Vasco e Avaí times que tinham maiores chances de cair para a série B. veja a classificação final abaixo.

Tabela 3 - Classificação Final do Campeonato Brasileiro de futebol de 2015.

Classificação Final do Campeonato Brasileiro de futebol de 2015																
	Classificação	**P**	**J**	**V**	**E**	**D**	**GP**	**GC**	**SG**	**VM**	**VV**	**DM**	**DV**	**CA**	**CV**	**%**
1º	CORINTHIANS - SP	**81**	38	24	9	5	71	31	40	16	8	1	4	64	2	71
2º	ATLÉTICO - MG	**69**	38	21	6	11	65	47	18	13	8	4	7	76	4	60
3º	GRÊMIO - RS	**68**	38	20	8	10	52	32	20	14	6	2	8	95	2	59
4º	SÃO PAULO - SP	**62**	38	18	8	12	53	47	6	12	6	1	11	73	5	54
5º	INTERNACIONAL - RS	**60**	38	17	9	12	39	38	1	14	3	2	10	101	6	52
6º	SPORT - PE	**59**	38	15	14	9	53	38	15	13	2	1	8	72	2	51
7º	SANTOS - SP	**58**	38	16	10	12	59	41	18	15	1	1	11	79	9	50
8º	CRUZEIRO - MG	**55**	38	15	10	13	44	35	9	10	5	3	10	89	5	48
9º	PALMEIRAS - SP	**53**	38	15	8	15	60	51	9	9	6	6	9	91	5	46
10º	ATLÉTICO - PR	**51**	38	14	9	15	43	48	-5	9	5	4	11	87	7	44
11º	PONTE PRETA - SP	**51**	38	13	12	13	41	40	1	9	4	6	7	99	5	44
12º	FLAMENGO - RJ	**49**	38	15	4	19	45	53	-8	8	7	8	11	82	6	42
13º	FLUMINENSE - RJ	**47**	38	14	5	19	40	49	-9	10	4	6	13	98	10	41
14º	CHAPECOENSE - SC	**47**	38	12	11	15	34	44	-10	9	3	5	10	84	3	41
15º	CORITIBA - PR	**44**	38	11	11	16	31	42	-11	6	5	5	11	113	5	38
16º	FIGUEIRENSE - SC	**43**	38	11	10	17	36	50	-14	7	4	6	11	111	5	37
17º	AVAÍ - SC	**42**	38	11	9	18	38	60	-22	8	3	6	12	119	3	36
18º	VASCO DA GAMA - RJ	**41**	38	10	11	17	28	54	-26	5	5	8	9	108	14	35
19º	GOIÁS - GO	**38**	38	10	8	20	39	49	-10	7	3	8	12	89	3	33
20º	JOINVILLE - SC	**31**	38	7	10	21	26	48	-22	6	1	6	15	94	8	27
P pontos - **J** jogos - **V** vitórias - **E** empates - **D** derrotas - **GP** gols pró - **GC** gols contra - **SG** saldo de gols - **VM** vitória mandante - **VV** vitória visitante - **DM** derrota mandante - **DV** derrota visitante - **CA** cartões amarelos - **CV** cartões vermelhos – **%** aproveitamento ■Libertadores ■Rebaixados																

Fonte: www.cbf.com.br/competicoes/brasileiro-serie-a visualização em 07/12/2015

3. ESTATÍSTICA

O objetivo deste capítulo é introduzir conceitos básicos em estatística. O aprendizado desses conceitos nos permitirá a compreensão da importância de se desenvolver um bom planejamento, a partir da definição da população, da amostra e das variáveis a serem utilizadas. Além disso, aprenderemos a organizar os dados em gráficos e tabelas.

3.1 Conceitos Básicos

De acordo com Memória (2004), a Estatística é o ramo da Matemática que nos permite de forma organizada recolher dados sobre uma população, analisá-los e tirar conclusões. A estatística é utilizada para a melhora do desempenho de empresas, órgãos governamentais e entidades esportivas.

3.1.1 População

É o conjunto de elementos a serem investigados, exemplos: fauna de um parque ecológico, jogadores inscritos no campeonato brasileiro, eleitores brasileiros, clientes de um restaurante.

3.1.2 Indivíduo

Todo elemento da população é denominado de indivíduo, exemplos: um tipo de arvore de um parque ecológico, um jogador escrito no campeonato brasileiro, um eleitor brasileiro, cliente de um restaurante.

3.1.3 Variáveis

O aspecto a ser estudado, comum a todos os indivíduos da pesquisa é chamado de variável. As variáveis são qualitativas ou quantitativas. É qualitativa quando ela expressa uma qualidade ou um atributo, exemplos: nome, profissão, cor, estado civil, preferência esportiva. É dita quantitativa quando ela expressa uma quantidade, ou seja, um número, exemplos: idade, remuneração, saldo de gols, pontos ganhos, peso, números de alunos, quantidade de jogadores relacionados, entre outros.

As variáveis qualitativas podem ser classificadas em discretas ou continuas. Quando a resposta para uma variável quantitativa apenas aceita números inteiros ela é dita variável quantitativa discreta, exemplo: quantidade de cartões, faltas, gols e saldo de gols. Quando a resposta para uma variável quantitativa aceita números reais, ou seja, pode ter valores não inteiros, ela é dita variável quantitativa continua exemplo remuneração dos jogadores, altura em metros, preço dos ingressos.

3.2 Tabelas

Para resumir, organizar e tabular os dados obtidos em uma pesquisa é muito utilizado quadros chamados de tabelas. Tabela é um arranjo sistemático de dados numéricos dispostos em forma de colunas e linhas, para fins de comparação. A apresentação em formas de tabela deve expor os dados de modo fácil e que deixe a leitura mais rápida.

Exemplo 3.1: Podemos analisar o Campeonato Brasileiro de Futebol de 2012 através da seguinte tabela:

Tabela 4: Classificação Final do Campeonato Brasileiro de Futebol de 2012.

CAMPEONATO BRASILEIRO DE FUTEBOL DE 2012 – SERIE A																	
Classifica-ção		**EQUIPE**	**P**	**J**	**V**	**E**	**D**	**GP**	**GC**	**SG**	**VM**	**VV**	**DM**	**DV**	**CA**	**CV**	**%**
1º		FLUMINENSE - RJ	**77**	38	22	11	5	61	33	28	11	11	3	2	85	3	67
2º		ATLÉTICO - MG	**72**	38	20	12	6	64	37	27	14	6	0	6	104	10	63
3º		GRÊMIO - RS	**71**	38	20	11	7	56	33	23	13	7	2	5	102	9	62
4º		SÃO PAULO - SP	**66**	38	20	6	12	59	37	22	15	5	2	10	93	6	57
5º		VASCO DA GAMA - RJ	**58**	38	16	10	12	45	44	1	9	7	6	6	91	3	50
6º		CORINTHIANS - SP	**57**	38	15	12	11	51	39	12	10	5	3	8	87	3	50
7º		BOTAFOGO - RJ	**55**	38	15	10	13	60	50	10	9	6	6	7	75	3	48
8º		SANTOS - SP	**53**	38	13	14	11	50	44	6	8	5	2	9	82	4	46
9º		CRUZEIRO - MG	**52**	38	15	7	16	47	51	-4	9	6	5	11	121	9	45
10º		INTERNACIONAL - RS	**52**	38	13	13	12	44	40	4	9	4	6	6	106	6	45
11º		FLAMENGO - RJ	**50**	38	12	14	12	39	46	-7	8	4	3	9	81	6	43
12º		NÁUTICO - PE	**49**	38	14	7	17	44	51	-7	13	1	3	14	86	6	43
13º		CORITIBA - PR	**48**	38	14	6	18	53	60	-7	11	3	6	12	99	7	42
14º		PONTE PRETA - SP	**48**	38	12	12	14	37	44	-7	9	3	3	11	94	2	42
15º		BAHIA - BA	**47**	38	11	14	13	37	41	-4	5	6	5	8	112	2	41
16º		PORTUGUESA - SP	**45**	38	10	15	13	39	41	-2	7	3	4	9	87	4	39
17º		SPORT - PE	**41**	38	10	11	17	39	56	-17	7	3	6	11	85	8	36
18º		PALMEIRAS - SP	**34**	38	9	7	22	39	54	-15	6	3	8	14	106	6	29
19º		ATLÉTICO - GO	**30**	38	7	9	22	37	67	-30	5	2	11	11	102	6	26
20º		FIGUEIRENSE - SC	**30**	38	7	9	22	39	72	-33	5	2	8	14	95	9	

P pontos - **J** jogos - **V** vitórias - **E** empates - **D** derrotas - **GP** gols pró - **GC** gols contra - **SG** saldo de gols - **VM** vitória mandante - **VV** vitória visitante - **DM** derrota mandante - **DV** derrota visitante - **CA** cartões amarelos - **CV** cartões vermelhos – **%** aproveitamento ■Libertadores■Rebaixados

Fonte: www.cbf.com.br visualizada em 28/11/2015

É fácil perceber que o Fluminense foi o campeão brasileiro de 2012; e que além do Fluminense, também se classificaram para a Taça Libertadores da América os times Atlético Mineiro, o Grêmio e o São Paulo sendo rebaixados para a série B o Sport, o Palmeiras, o Atlético Goianiense e o Figueirense.

Elementos característicos de uma tabela.

Titulo: Indica o assunto da tabela.

Cabeçalho: Explica o conteúdo das linhas e colunas da tabela.

Corpo: são os dados da tabela.

Colunas: Indicadoras dos conteúdos das linhas.

Fonte: Mostra de onde foram recolhidos os dados da pesquisa.

3.3 Distribuição de Frequência

Para conhecer o comportamento de uma variável é preciso analisar com que frequência os dados estão distribuídos dentro dessa variável.

Frequência absoluta, representada por f, de uma variável é o número de vezes em que ele é observado.

Frequência relativa, representada por fr, de uma variável é a razão entre a frequência absoluta e o número de elementos da população. Ou seja, $fr = \frac{f}{n}$. A frequência relativa pode ser expressa em fração, porcentagem ou número decimal.

Como já vimos, no capitulo anterior, à frequência relativa dos jogos como mandante ou visitante pode ser usada para o cálculo das probabilidades dos resultados de uma partida.

Exemplo 3.2: Observando a Tabela 4 podemos calcular as frequências relativas dos jogos do Fluminense como mandante e como visitante no campeonato brasileiro de 2012.

Frequência relativa como mandante de vitória $fr = \frac{11}{19} \cong 0{,}579$, empate $fr = \frac{5}{19} \cong 0{,}263$ e derrota $fr = \frac{3}{19} \cong 0{,}158$.

Frequência relativa como visitante de vitória $fr = \frac{11}{19} \cong 0{,}579$, empate $fr = \frac{6}{19} \cong 0{,}316$ e derrota $fr = \frac{2}{19} \cong 0{,}105$.

A tabela de frequência abaixo mostra a variável e suas frequências absolutas (f) e relativas (fr) em porcentagem.

Tabela 5: Frequência dos resultados dos jogos como mandante e visitante do Fluminense o campeão brasileiro de 2012.

Resultados dos jogos do campeão brasileiro de 2012				
Resultado	Frequência absoluta (f) como mandante	Frequência relativa (fr), como mandante.	Frequência absoluta (f) como visitante	Frequência relativa (fr) como visitante
Vitórias	11	57,9%	11	57,9%
Empates	5	26,3%	6	31,6%
Derrotas	3	15,8%	2	10,5%
Total	19	100 %	19	100 %

Fonte: site da CBF Última atualização efetuada em 04/02/2013.

Quando aparecem muitos valores diferentes agrupamos esses valores em intervalos de classes. Numa distribuição de frequência por intervalos (ou classe), é necessário determinar os limites das classes, calculando a diferença entre o maior e o menor resultado obtido, e dividindo essa diferença pelo número de intervalos desejado. Usamos o símbolo ⊢ que indica intervalo fechado à direita e aberto à esquerda.

Exemplo 3.3: A seleção brasileira de voleibol masculino que disputou a Liga Mundial em 2015 utilizou 24 atletas, veja a seguir as suas alturas em metros (m): 1,90; 2,05; 2,04; 1,98; 2,03; 2,12; 1,85; 1,90; 1,90; 1,84; 1,88; 1,96; 2,09; 2,07; 2,09; 1,94; 1,95; 1,92; 2,17; 1,88; 1,99; 2,04; 1,83 e 1,95.

Em virtude de aparecerem vários valores diferentes para a variável (altura), torna-se inviável colocar na tabela uma linha para cada valor. Vamos então agrupar os valores em intervalos (ou classes), da seguinte forma segundo Doane (2014):

I. Colocamos os dados em ordem crescente: 1,83; 1,84; 1,85; 1,88; 1,88; 1,90; 1,90; 1,90; 1,92; 1,94; 1,95; 1,95; 1,96; 1,98; 1,99; 2,03; 2,04; 2,04; 2,05; 2,07; 2,09; 2,09; 2,12 e 2,17.

II. Escolhemos o número de classes: utilizando a regra de Sturges $k = 1 + 3{,}3\log n$, onde k é o número de classes e n é o tamanho da amostra.

   Logo $k = 1 + 3{,}3\log 24 \Rightarrow k \cong 1 + 3{,}3 \cdot 1{,}38 \Rightarrow k \cong 1 + 4{,}5 \Rightarrow k \cong 5{,}5$
   Portanto, podemos utilizar 5 ou 6 intervalos. Usaremos neste caso 5 intervalos.

III. Estabelecemos os limites dos intervalos de classe, calculando a diferença entre o maior e o menor resultado adquirido, obtendo a amplitude total, dividimos pelo número de classes obtido em II $\frac{(2{,}17\ m - 1{,}83\ m)}{5} = \frac{0{,}34\ m}{5} = 0{,}068\ m.$ Arredondaremos para $0{,}07\ m$

IV. Construímos a tabela de frequências.

Tabela 6 - Altura dos jogadores da seleção brasileira de voleibol masculino que disputaram a Liga Mundial em 2015.

Altura dos jogadores da seleção brasileira de voleibol masculino que disputaram a Liga Mundial em 2015.	
Altura	Frequência absoluta (f)
$1{,}83 \vdash 1{,}90m$	5
$1{,}90 \vdash 1{,}97m$	8
$1{,}97 \vdash 2{,}04m$	3
$2{,}04 \vdash 2{,}11$	6
$2{,}11 \vdash 2{,}17$	2

Fonte: www.cbv.com.br visualizada em 28/12/2015.

Analisando a tabela 6, podemos observar que o intervalo com o maior número de jogadores, com oito atletas, é entre 1,90 e 1,97 *m* e o intervalo com menos jogadores com dois é o com os jogadores mais altos da seleção.

Podemos também usar a Frequência absoluta acumulada, representada por fa, e frequência relativa acumulada a partir de uma frequência absoluta, adicionamos os valores das frequências anteriores.

Veja como ficaria o exemplo anterior.

Tabela 7 Fr e Fra da altura dos jogadores da seleção brasileira de voleibol masculino que disputaram a Liga Mundial em 2015

Fr e Fra da altura dos jogadores da seleção brasileira de voleibol masculino que disputaram a Liga Mundial em 2015				
Altura	Frequência absoluta (f)	Frequência acumulada (fa)	Frequência relativa (fr),	Frequência relativa acumulada (fra),
$1,83 \vdash 1,90m$	5	5	0,21	0,21
$1,90 \vdash 1,97m$	8	13	0,33	0,54
$1,97 \vdash 2,04m$	3	16	0,125	0,665
$2,04 \vdash 2,11$	6	22	0,25	0,915
$2,11 \vdash 2,17$	2	24	0,085	1

Fonte: : www.cbv.com.br visualizada em 28/12/2015

3.4 Representações Gráficas

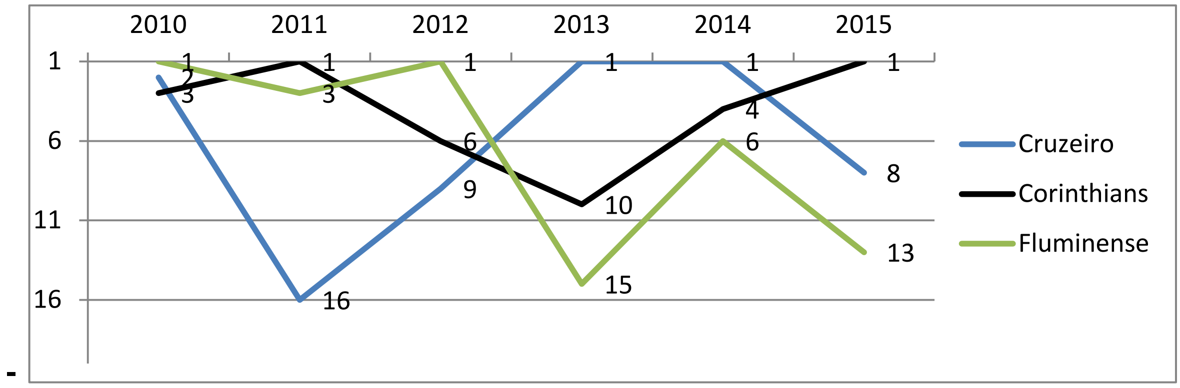
Outro modo de organizar os dados estatísticos é por meio de gráficos. Os gráficos estatísticos fornecem uma visão de um conjunto mais rápida, dinâmica e abrangente do estudo realizado. Motivo pelo qual os meios de comunicação com frequência oferecem informações por meio de gráficos. Os tipos de gráficos mais utilizados são os gráficos de segmentos, barras, setores e pictogramas.

3. 4. 1 Gráficos de Segmentos

O gráfico de segmento, também conhecido como gráfico de linha ou curva, utiliza uma linha poligonal para representar os dados estatísticos. Sendo utilizados principalmente para mostrar a evolução das frequências dos valores em certo período.

Exemplo 3.4: Análise da colocação no Campeonato Brasileiro de Futebol na segunda década do século XXI dos Campeões.

Figura 1 – Posição dos últimos campeões no campeonato brasileiro de 2010 a 2015



Fonte*: www.cbf.com.br* visualizado em 29/12/2015

Pelo gráfico acima, podemos observar que:

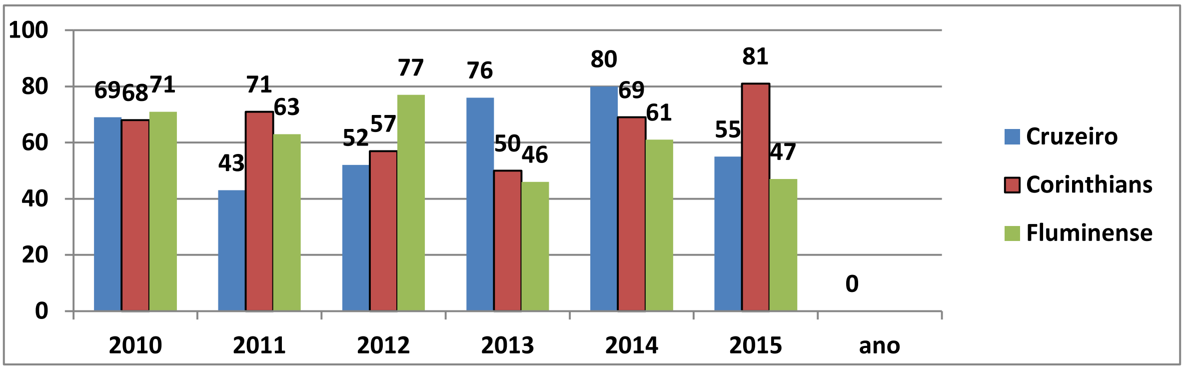
- Nos seis anos tiveram apenas três campeões com dois títulos cada.
- O Corinthians foi a equipe que menos oscilou estando sempre entre os dez primeiros e quatro vezes no G4 (grupo das quatro equipes que se classificam para a Taça Libertadores da América).
- Cruzeiro quase foi rebaixado em 2011, pois é rebaixado do 17° ao 20° colocado, e foi bicampeão em 2013 e 2014.

3. 4. 2 Gráficos de Barras

Gráfico de barras é uma forma de resumir um conjunto de dados categóricos. Nele, os dados estatísticos são representados por retângulos (barras) horizontais ou verticais. Sendo utilizados principalmente para a visualização do estudo através da comparação entre a variável e sua frequência absoluta ou relativa.

Exemplo 3.5: Agora veremos as pontuações de Corinthians, Cruzeiro e Fluminense nos anos de 2010 a 2015.

Figura 2 – Pontuação dos últimos campeões no campeonato brasileiro de 2010 a 2015.



Fonte: *www.cbf.com.br* visualizado em 29/12/2015

Pelo gráfico acima, podemos observar que:

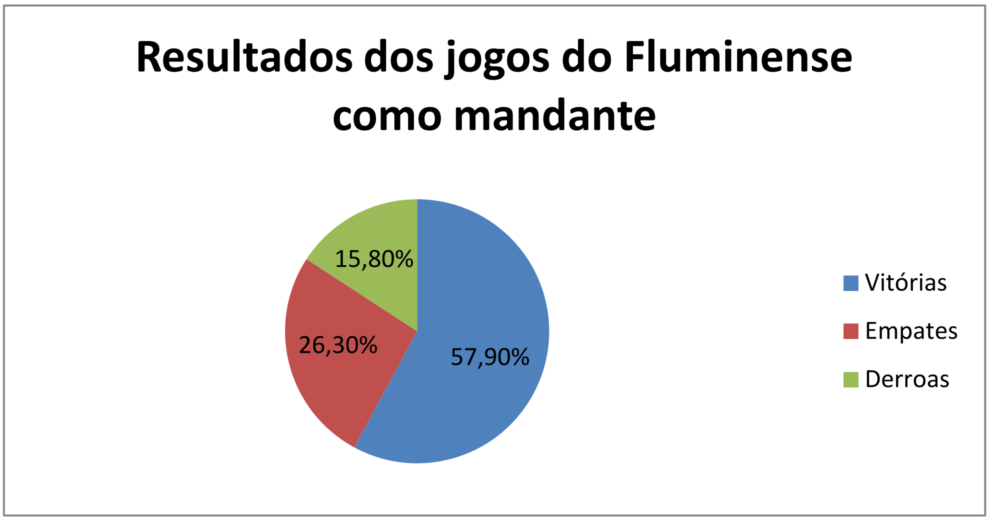
- Em 2010 e 2011 foram os anos em que os campeões tiveram a menor pontuação.
- O Corinthians foi à equipe que mais pontuou de 2010 a 2015.
- Dos campeões desse período de tempo o Cruzeiro foi o que menos fez pontos em um ano com 43 pontos em 2011.

3. 4. 3 Gráficos de Setores

O gráfico de setores também conhecido como gráfico de pizza utiliza um circulo dividido em setores circulares para representar os dados estatísticos. Sendo utilizados principalmente para comparar as frequências dos valores com o todo.

Exemplo 3.6: Frequência relativa dos resultados dos jogos como mandante do Fluminense no Campeonato Brasileiro de 2012.

Figura 3 - Resultados dos jogos do Fluminense como mandante.



Fonte: site da CBF visualizado em 29/12/2015

Atualmente, usamos computadores para fazer gráficos. Para fazer manualmente um gráfico de pizza devemos calcular o ângulo de cada setor utilizando a regra de três, utilizar um compasso e um transferidor.

No exemplo acima o ângulo x da frequência relativa de vitórias é dado por:

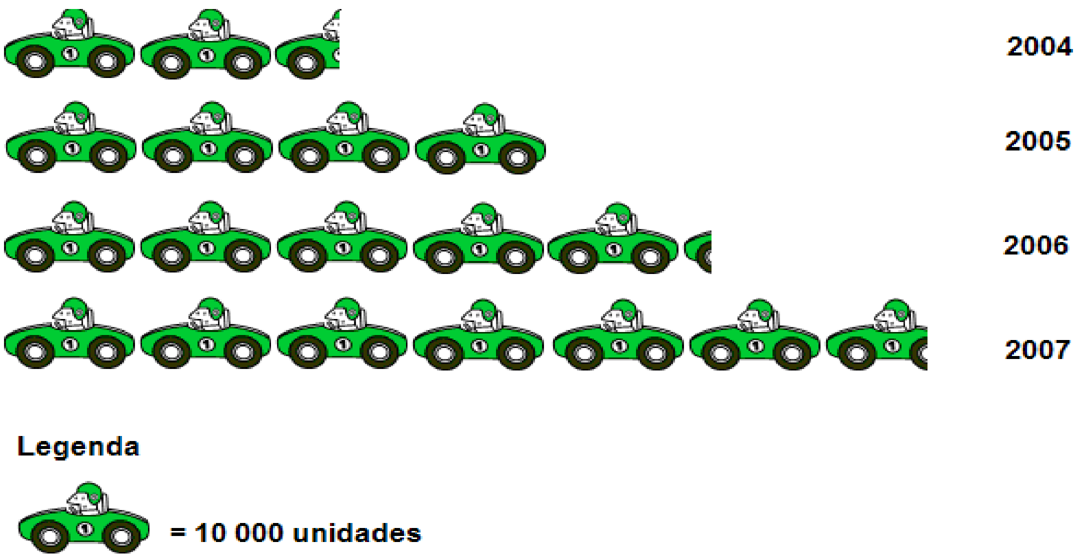
$$\frac{360°}{x} = \frac{100\%}{57{,}9\%} \Rightarrow 100x = 208{,}44° \Rightarrow x = 208{,}44°$$

3. 4. 4 Gráficos de pictóricos

O gráfico pictórico ou pictograma utiliza figuras relacionadas ao assunto para representar os dados estatísticos, tornando-os mais atrativos, sendo utilizados em publicações em revistas, jornais e sites.

Exemplo 3.7: Produção de Veículos da Empresa X, no período de 2004 a 2007.

Figura 4 - Produção de Veículos da Empresa X, no período de 2004 a 2007.



Fonte: dados fictícios

Pelo gráfico acima, podemos observar que:

- Em 2004 a produção de carros foi de aproximadamente 25.000 unidades.
- Em 2007 a produção de carros foi de aproximadamente 67.500 unidades.
- A produção de automóveis só cresceu de 2004 e 2007.

4 MEDIDAS DE TENDÊNCIA CENTAL E MEDIDAS DE DISPERSÃO

Como vimos no capítulo anterior, as tabelas de frequência e os gráficos nos fornecem informações para analisar o comportamento de uma variável estudada. Para iniciar suas análises preliminares, estabelecemos seus valores significativos como as médias, mediana e moda que são medidas de tendência central. Estudaremos também as medidas de dispersão, que são o desvio médio, a variância e o desvio padrão, pois essas medidas auxiliam na validação das medidas de tendência central.

4.1 Medidas de tendência central

Considerando o público pagante dos jogos em casa de um time A, podemos determinar um número que nos fornece uma ideia aproximada da frequência de público pagante em todos os jogos. Avaliando as notas bimestrais de um aluno, podemos registrar uma nota anual para saber se ele será promovido ou não para o próximo nível escolar. O número obtido em cada caso é denominado de medida de posição ou tendência central.

Existem situações em que exige uma redução de forma coerente dos dados coletados, para isso usamos os valores das medidas de posição.

As médias, a mediana e a moda são as principais medidas de posição.

4.1.1 Médias

Um dos conceitos estatístico mais utilizado no esporte para analisar desempenho de equipes e atletas por treinadores, pela imprensa e por torcedores é a média.

Além dos esportes as médias são aplicadas em diferentes contextos: escolas, empresas, governos, entre outros. As principais médias são a média aritmética, ponderada geométrica e harmônica.

4.1.1.1 Média aritmética

Dentre as médias, a mais utilizada é a média aritmética, representada por $\bar{x}$. Ela é obtida por meio da divisão entre a soma dos de todos os valores pelo número total dos valores, ou seja, $\bar{x} = \frac{x_1+x_{2+}x_{3+\cdots+}x_n}{n}$ ou $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Exemplo 4.1: Ainda usando os dados da figura 3, obtemos as médias de pontos obtidos por Corinthians, Cruzeiro e Fluminense de 2010 a 2015.

Média de pontos do Corinthians:

$$\bar{x} = \frac{68+71+57+50+69+81}{6}$$

$$\bar{x} = \frac{396}{6}$$

$$\bar{x} = 66$$

Média de pontos do Cruzeiro:

$$\bar{x} = \frac{69+43+52+76+80+55}{6}$$

$$\bar{x} = \frac{375}{6}$$

$$\bar{x} = 62{,}5$$

Média de pontos do Fluminense:

$$\bar{x} = \frac{71+63+77+46+61+47}{6}$$

$$\bar{x} = \frac{365}{6}$$

$\bar{x} = 60{,}8333 \ldots.$

4.1.1.2 Média ponderada

Quando os valores estão agrupados ou quando se encontram ponderados (tem pesos diferentes). Para o calculo da média levamos em conta a frequência absoluta ou peso. Assim, a fórmula para o cálculo da média é $\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_n f_n}{f_1 + f_2 + f_3 + \cdots + f_n}$,

Em que $\bar{x}$ é a média aritmética ponderada; x_i é o valor da série numérica utilizada e f_i é a frequência absoluta ou peso do valor x_i.

Exemplo 4.2: Em uma escola a média final é obtida da seguinte forma, para evitar que o aluno Pedrinho perca o interesse pelos estudos no quarto bimestre, o primeiro bimestre tem peso um, o segundo peso dois, o terceiro peso três e o quarto peso quatro. Sabendo que a média para a promoção do aluno no ano seguinte é seis, vamos verificar se o aluno foi promovido ou ficou em recuperação sendo que ele obteve nota nove no primeiro bimestre, seis no segundo, cinco no terceiro e quatro no quarto bimestre.

Utilizada a média ponderada, obtivemos o resultado que Pedrinho foi para recuperação, pois $\bar{x} = \frac{9 \cdot 1 + 6 \cdot 2 + 5 \cdot 3 + 4 \cdot 4}{1+2+3+4} \Rightarrow \bar{x} = \frac{9+12+15+16}{10} \Rightarrow \bar{x} = \frac{52}{10} \Rightarrow \bar{x} = 5{,}2.$

Note que se a média adotada pela escola fosse aritmética ele seria promovido para o nível escolar seguinte:

$$\bar{x} = \frac{9+6+5+4}{4} \Rightarrow \bar{x} = \frac{24}{4} \Rightarrow \bar{x} = 6.$$

Para dados apresentados em intervalos de classes e cálculo da média usamos o ponto médio de cada intervalo.

Exemplo 4.3: Observando a Tabela 6 podemos calcular a média de altura dos jogadores da seleção brasileira de voleibol masculino que disputaram a Liga Mundial de 2015.

$$\bar{x} = \frac{1{,}865 \cdot 5 + 1{,}935 \cdot 8 + 2{,}005 \cdot 3 + 2{,}075 \cdot 6 + 2{,}145 \cdot 2}{5+8+3+6+2} \Rightarrow \bar{x} = \frac{9{,}325 + 15{,}48 + 6{,}015 + 12{,}45 + 4{,}29}{24}$$

$$\Rightarrow \bar{x} = \frac{47{,}56}{24} \Rightarrow \bar{x} \cong 1{,}98.$$

Portanto, a média de altura dos jogadores da seleção brasileira de voleibol masculino que disputaram a Liga Mundial de 2015 é de 1,98 metros.

4.1.1.3 Média Geométrica

Quando os valores estão caracterizados segundo uma progressão geométrica, utilizamos a média geométrica, representada por $\overline{x_g}$. Para o cálculo dessa média extraímos a raiz enésima do produto das frequências absolutas, ou seja, $\overline{x_g} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$.

Exemplo 4.4: Para contratar um jogador de basquete o técnico analisou em três jogos os seus percentuais de acertos em seus arremessos, sendo 80% de acertos no primeiro jogo, 88% no segundo e 93% após três meses. Qual o percentual médio de acertos?

Como não sabemos a quantidade de arremessos em cada jogo a média aritmética não é indicada, logo usaremos a média geométrica $\overline{x_g} = \sqrt[3]{0{,}8 \cdot 0{,}88 \cdot 0{,}93}$ $\Rightarrow$ $\overline{x_g} = \sqrt[3]{0{,}65472} \Longrightarrow \overline{x_g} \cong 0{,}868$, ou seja, média de 86,8% de acertos por partida.

4.1.1.4 Média Harmônica

Quando os valores estão caracterizados de forma inversamente proporcional, utilizamos a média harmônica, representada por $\overline{x_h}$. A média harmônica é obtida por meio da divisão do número total dos valores n e a soma dos inversos de cada valor, ou seja,

$$\overline{x_h} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}}.$$

Exemplo 4.5: Um ultramaratonista participou de uma prova que realizou o trajeto de ida e volta entre as cidades *A* e *B* pelo mesmo caminho. Na ida, ele desenvolveu uma velocidade média de 15 km/h, na volta, a velocidade média desenvolvida foi de 10 km/h. Qual a velocidade média para realizar todo o percurso da prova?

É fácil perceber que a média aritmética das velocidades seria de 12,5 km/h, pois $\bar{x} = \frac{15+10}{2} = 12{,}5.$ No entanto, a pergunta não foi qual a média das velocidades, mas sim qual a velocidade média para realizar todo o percurso.

A resposta para esta pergunta seria a média harmônica $\overline{x_h} = \frac{2}{\frac{1}{15}+\frac{1}{10}} \Rightarrow$ $\overline{x_h} = \frac{2}{\frac{2+3}{30}} \Rightarrow \overline{x_h} = \frac{60}{5} \Rightarrow \overline{x_h} = 12$ km/h:

Por que 12 km/h? Em que se baseia este resultado?

Observemos: independentemente da distância entre as cidades as velocidades médias foram de 15 km/h na ida e de 10 km/h na volta, imaginemos que a distância entre as cidades *A* e *B* seja de 30 quilômetros.

Como base nessas informações, podemos concluir que, o tempo gasto na ida seria de duas horas, na volta seria de três horas e o percurso total da prova 60 quilômetros em cinco horas. Concluímos que a velocidade média da prova foi de $v = \frac{60}{5} = 12km/h.$

4.1.1.5 Desigualdades entre as Médias

Em uma sequência numérica, a média aritmética é sempre maior ou igual à média geométrica, que é sempre maior ou igual à média harmônica. Ou seja, $\bar{x} \geq \overline{x_g} \geq \overline{x_h}$ ou $\frac{x_1+x_{2+}x_{3+\cdots+}x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} \geq \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}}$, e as médias só são iguais quando todos os valores da sequência são iguais.

Demonstração feita por Caminha (1999): Demonstraremos primeiro que $\bar{x} \geq \overline{x_g}$, em seguida que $\overline{x_g} \geq \overline{x_h} \Rightarrow \bar{x} \geq \overline{x_g} \geq \overline{x_h}$.

Prova que $\bar{x} \geq \overline{x_g}$: Façamos a prova em dois passos:

i. A desigualdade é verdadeira quando *n* for uma potência de 2, só ocorrerá a igualdade, se e somente todos os números forem iguais.

ii. A desigualdade é verdadeira em geral, e a igualdade ocorre apenas, se os números forem todos iguais.

i. Façamos por indução finita, sendo $n = 2^m$: Para *m* = 1, temos:

$\frac{x_1+x_2}{2} \geq \sqrt{x_1 x_2} \Leftrightarrow x_1 + x_2 \geq 2\sqrt{x_1 x_2} \Leftrightarrow x_1 - 2\sqrt{x_1 x_2} + x_2 \geq 0 \Leftrightarrow (\sqrt{x_1} - \sqrt{x_2})^2 \geq 0,$ o que é verdade. Há igualdade $(\sqrt{x_1} - \sqrt{x_2})^2 = 0$ se e só se $x_1 = x_2$.

Agora admitindo que para $m = k$ seja verdade, ou seja: $\frac{x_1+x_{2+}x_3+\cdots+x_{2^k}}{2^k} \geq \sqrt[2^k]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_{2^k}}$ e que há igualdade se e só se todos os números forem iguais.

Queremos demonstrar que para $m = k + 1$ também é verdade, logo:

$$\frac{x_1 + x_2 + \cdots + x_{2^k} + x_{2^k+1} + \cdots + x_{2^{k+1}}}{2^{k+1}} = \frac{(x_1 + x_2 + \cdots + x_{2^k}) + \left(x_{2^k+1} + \cdots + x_{2.2^k}\right)}{2.2^k}$$

$$\frac{x_1 + x_2 + \cdots + x_{2^k} + x_{2^k+1} + \cdots + x_{2^{k+1}}}{2^{k+1}} = \frac{1}{2}\left(\frac{(x_1 + x_2 + \cdots + x_{2^k})}{2^k} + \frac{(x_{2^k+1} + \cdots + x_{2^{k+1}})}{2^k}\right)$$

$$\frac{x_1 + x_2 + \cdots + x_{2^k} + x_{2^k+1} + \cdots + x_{2^{k+1}}}{2^{k+1}} \overset{hipótese\ \ de\ indução}{\geq} \frac{\sqrt[2^k]{x_1 x_2 \ldots x_{2^k}} + \sqrt[2^k]{x_{2^k+1} x_{2^K+2} \ldots x_{2^{k+1}}}}{2}$$

$$\frac{x_1+x_2+\cdots+x_{2^k}+x_{2^k+1}+\cdots+x_{2^{k+1}}}{2^{k+1}} \overset{provado\ por\ m=1}{\geq} \sqrt{\sqrt[2^k]{x_1x_2\ldots x_{2^k}}\sqrt[2^k]{x_{2^k+1}x_{2^K+2}\ldots x_{2^{k+1}}}}$$

$$\frac{x_1+x_2+\cdots+x_{2^k}+x_{2^k+1}+\cdots+x_{2^{k+1}}}{2^{k+1}} \geq \sqrt{\sqrt[2^k]{x_1x_2\ldots x_{2^k}x_{2^k+1}\ldots x_{2^{k+1}}}}$$

$$\frac{x_1+x_2+\cdots+x_{2^k}+x_{2^k+1}+\cdots+x_{2^{k+1}}}{2^{k+1}} \geq \sqrt[2^{k+1}]{x_1x_2\ldots x_{2^k}x_{2^k+1}\ldots x_{2^{k+1}}}$$

Para haver igualdade, devemos tê-la em todas as passagens. Então, deve ser $\frac{x_1+x_2+\cdots+x_{2^k}}{2^k}=\sqrt[2^k]{x_1x_2\ldots x_{2^k}}$, $\frac{(x_{2^k+1}+\cdots+x_{2.2^k})}{2^k}=\sqrt[2^k]{x_{2^k+1}x_{2^K+2}\ldots x_{2^{k+1}}}$ e $\frac{\sqrt[2^k]{x_1x_2\ldots x_{2^k}}+\sqrt[2^k]{x_{2^k+1}x_{2^K+2}\ldots x_{2^{k+1}}}}{2}=\sqrt[2^k]{x_1x_2\ldots x_{2^k}x_{2^k+1}\ldots x_{2^{k+1}}}$. Já vimos que para as duas primeiras igualdades, segue da hipótese de indução que deve ser $x_1=x_2=\cdots=x_{2^k}$ e $x_{2^k+1}=x_{2^K+2}=\cdots=x_{2^{k+1}}$ A última igualdade ocorre se e só se $\sqrt[2^k]{x_1x_2\ldots x_{2^k}}=\sqrt[2^k]{x_{2^k+1}x_{2^K+2}\ldots x_{2^{k+1}}}$. Estas duas condições juntas implicam que devemos ter $x_1=x_2=\cdots=x_{2^k}=x_{2^k+1}=x_{2^K+2}=\cdots=x_{2^{k+1}}$. É também evidente que se os números forem todos iguais a igualdade ocorre.

ii. Seja agora $n>1$ um natural qualquer e $x_1,x_2,\ldots x_n$ reais positivos. Tome *k* natural tal que $2^k>n$. Usando a desigualdade entre as médias para os 2^k números $x_1,x_2,\ldots x_n$ e 2^k-n cópias de $\overline{x_g}=\sqrt[n]{x_1\cdot x_2\cdot x_3\cdot\ldots\cdot x_n}$, obtemos:

$$\frac{x_1+x_2+\ldots x_n+x+x+\cdots+x}{2^k} \geq \sqrt[2^k]{x_1x_2\ldots x_n x^{2^k-n}} = \sqrt[2^k]{x^n x^{2^k-n}} = \sqrt[2^k]{x^{2^k}} = x$$ e daí

$x_1+x_2+\ldots+x_n+(2^k-n)x\geq 2^kx$, ou ainda $\frac{x_1+x_2+\ldots x_n}{n}\geq x=\sqrt[n]{x_1x_2\ldots x_n}$, que era a desigualdade desejada.

Para haver igualdade, segue do item *i* que deve ser $x_1=x_2=\cdots=x_n=x=\cdots=x$. Em particular, todos os números $x_1,x_2,\ldots x_n$ devem ser iguais. É fácil ver que se esses números forem todos iguais então há igualdade, logo $\bar{x}\geq\overline{x_g}$. □

Agora provaremos que $\overline{x_g}\geq\overline{x_h}$.

Já sabemos que $\bar{x}\geq\overline{x_g}$, ou seja,

$$\frac{x_1+x_2+\ldots x_n}{n}\geq\sqrt[n]{x_1x_2\ldots x_n}$$

$$x_1 + x_2 + \cdots x_n \geq n\sqrt[n]{x_1 x_2 \ldots x_n}$$

$$(x_1 + x_2 + \cdots x_n)^n \geq n^n x_1 x_2 \ldots x_n$$

$$(x_1 x_2 \ldots x_n)^{n-1}(x_1 + x_2 + \ldots x_n)^n \geq n^n x_1 x_2 \ldots x_n (x_1 x_2 \ldots x_n)^{n-1}$$

$$(x_1 x_2 \ldots x_n)^{n-1} \geq \frac{n^n (x_1 x_2 \ldots x_n)^n}{(x_1 + x_2 + \ldots x_n)^n}$$

$$\sqrt[n]{(x_1 x_2 \ldots x_n)^{n-1}} \geq \frac{n x_1 x_2 \ldots x_n}{x_1 + x_2 + \ldots x_n}$$

$$\sqrt[n]{(x_1 x_2 \ldots x_n)^{n-1}} \geq \frac{n}{\frac{x_1 + x_2 + \ldots x_n}{x_1 x_2 \ldots x_n}}$$

$$\sqrt[n]{(x_1 x_2 \ldots x_n)^{n-1}} \geq \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

Mas como $\sqrt[n]{(x_1 x_2 \ldots x_n)^n} \geq \sqrt[n]{(x_1 x_2 \ldots x_n)^{n-1}}$, pois $x_1, x_2, \ldots x_n$ reais positivos,

$\sqrt[n]{(x_1 x_2 \ldots x_n)^n} \geq \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$, portanto $\overline{x_g} \geq \overline{x_h} \Rightarrow \bar{x} \geq \overline{x_g} \geq \overline{x_h}$. □

4.1.2 Mediana

Para Tiboni (2003), mediana é uma medida de posição chamada de separatriz, pois ela divide um conjunto de dados, em duas partes com número iguais de elementos. A mediana é mais indicada quando as variáveis são qualitativas, por não poderem ser calculada as médias, e quando nas variáveis quantitativas há presença de valores que distorcem dos demais, por serem muito pequenos ou muito grandes, fazendo com que a média não caracterize de forma eficiente o conjunto de valores.

Existem três formas de calcular a mediana: uma para dados não agrupados, uma para dados agrupados sem intervalos de classe e uma para dados agrupados com intervalos de classe.

4.1.2.1 Mediana para Dados não Agrupados

No cálculo da mediana para dados não agrupados primeiramente colocamos os dados em rol (ordem crescente ou decrescente) quando possível em seguida:

- Se tivermos um número impar de dados, a mediana é determinada pelo termo de ordem $\frac{n+1}{2}$·, onde n é o número de termos da série.

Exemplo 4.6: Altura dos jogadores, da seleção brasileira de futebol campeã do mundo em 1994, Taffarel (1,83m); Jorginho (1,75m), Aldair (1,83m), Márcio Santos (1,87m) e Branco (1,80m); Mauro Silva (1,77m), Dunga (1,77m), Mazinho (1,76m) e Zinho (1,72m); Bebeto (1,76m) e Romário (1,69m)

Colocando as suas alturas em metros (m) em rol: 1,69; 1,72; 1,75; 1,76; 1,76; 1,77; 1,77; 1,80; 1,83; 1,83; e 1,87.

Como temos 11 valores à mediana é determinada pelo termo de ordem $\frac{11+1}{2} = 6$, logo a mediana é 1,77.

- Se tivermos um número par de dados, a mediana é determinada pela média aritmética entre os termos de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$.

Exemplo 4.7: A mediana da altura dos jogadores da seleção brasileira de voleibol masculino que disputou a Liga Mundial em 2015:

Colocamos as suas alturas em metros (m) em rol: 1,83; 1,84; 1,85; 1,88; 1,88; 1,90; 1,90; 1,90; 1,92; 1,94; 1,95; 1,95; 1,96; 1,98; 1,99; 2,03; 2,04; 2,04; 2,05; 2,07; 2,09; 2,09; 2,12 e 2,17.

Como temos 24 valores à mediana é determinada pela média aritmética entre o 12° e 13° termo, $Me = \frac{1{,}95+1{,}96}{2} \Rightarrow Me = 1{,}955$ metros.

4.1.2.2 Mediana para Dados Agrupados sem intervalos de classe

Para dados agrupados sem intervalos de classe calculamos a mediana da seguinte forma:

- Calculamos a localização da mediana na série numérica, dividindo o número total de elementos por dois, ou seja, $\frac{n}{2}$.
- Encontramos a mediana da série usando a frequência acumulada.

Exemplo 4.8: Análise do número de gols por partida em um torneio de futebol.

Tabela 8 - Número de gols por partida em um torneio de futebol.

Número de gols por partida em um torneio de futebol		
Número de gols	Frequência absoluta (f)	Frequência acumulada (fa)
0	6	6
1	12	18
2	15	33
3	5	38
4	2	40
Total	40	40

Fonte: Dados fictícios

Como o total de frequências é 40 (número par), os valores centrais são o 20° e o 21°, que são ambos 2 gols, logo $Me = \frac{2+2}{2} = 2$.

4.1.2.3 Mediana para Dados Agrupados com intervalos de classe

Para dados agrupados com intervalos de classe calculamos a mediana da seguinte forma:

- Localizamos da mediana na série numérica, dividindo o número total de elementos por dois, ou seja, $\frac{n}{2}$.
- Calculamos a mediana através da fórmula $Me = L + \frac{\left|\frac{n}{2} - \sum f\right| h}{f}$, onde: L é o limite inferior da classe que contem a mediana; n é o número total de elementos; f é a frequência da classe que contém a mediana; h é a amplitude da classe da mediana e $\sum f$ é a frequência acumulada da classe anterior da mediana.

Exemplo 4.9: A pesquisa sobre o “peso” (massa em quilogramas) das atletas de ginástica olímpica de um clube pode ajudar a definir parâmetros sobre alimentação, treinamento, cálculo do IMC (índice de massa corpórea) dentre outras aplicações.

Os dados de em pesquisa sobre a massa corpórea são apresentados com agrupamentos de intervalos de classe.

Tabela 9 - Peso das atletas de ginástica olímpica de um clube.

Peso das atletas de ginástica olímpica de um clube		
Peso	Frequência absoluta (f)	Frequência acumulada (fa)
$40 \vdash 44$	1	1
$44 \vdash 48$	3	4
$48 \vdash 52$	7	12
$52 \vdash 56$	6	18
$56 \vdash 60$	3	20
Total	20	20

Fonte: fictícia

Localizando a mediana na classe numérica, dividindo o número total de elementos por dois, obtendo assim o 10° termo.

Calculamos a mediana através da fórmula $Me = L + \frac{\left|\frac{n}{2} - \sum f\right| h}{f}$, onde:

$L = 48$ é o limite inferior da classe que contem a mediana;

$n = 20$ é o número total de elementos;

$f = 7$ é a frequência da classe que contém a mediana;

$h = 4$ é a amplitude da classe da mediana

$\sum f = 4$ é a frequência acumulada da classe anterior da mediana.

$Me = 48 + \frac{|10-4|4}{7}$,

$Me \cong 51{,}43$ kg.

4.1.3 Moda

Assim como a mediana, a Moda é mais indicada quando as variáveis são qualitativas, por não poderem ser calculadas as médias das variáveis, nem serem ordenadas.

Em Estatística, o valor mais frequente é denominado de valor modal. Quando os dados estão agrupados em intervalos de classe, a moda será representada pelo ponto médio do intervalo de classe, do valor mais frequente.

Exemplo 4.10: Em uma pesquisa com um grupo de adolescentes cariocas, foi perguntado qual era o time de futebol preferido. O resultado foi o seguinte: 14 torcem pelo Flamengo, 9 pelo Vasco, 7 pelo Fluminense, 6 pelo Botafogo e 5 por clubes de outros estados. Explicitado anteriormente, não seria conveniente querer identificar a média ou a mediana, pois a variável é qualitativa, por isso usaremos como medida de tendência central a Moda. Dessa forma a Moda dessa situação é “Flamengo”, com 14 ocorrências.

As séries numéricas, que apresentam mais de uma moda são denominadas de séries multimodais, e as que não têm repetição de números são séries amodais.

4.2 Medidas de dispersão

Nem sempre a média, a mediana e a moda são valores que representam uma série numérica. Para comprovar, utilizamos as medidas de dispersão. As principais medidas de dispersão são o desvio médio, a variância e o desvio padrão.

Primeiramente, veremos como calcular as medidas de dispersão para dados não agrupados, em seguida, para dados agrupados sem e com intervalos de classe.

4.2.1 Medidas de dispersão para dados não agrupados

Exemplo 4.11: O técnico de uma equipe de basquetebol está precisando de um jogador. Dois jogadores disputam a vaga, foi analisado o desempenho dos dois candidatos, Cristiano e Vicente, na última temporada. A tabela a seguir mostra o desempenho dos dois candidatos nos arremessos das seguintes posições:

Tabela 10: Frequência relativa dos acertos dos arremessos dos candidatos.

Frequência relativa dos acertos dos arremessos dos candidatos.		
Posição do arremesso	Cristiano	Vicente
Dentro do garrafão	95%	90%
Fora do garrafão	85%	95%
Lance livre	80%	85%
Depois da linha dos três pontos	70%	80%
Da zona morta	70%	50%

Fonte: fictícia

Note que as médias de acertos dos dois jogadores são iguais:

Cristiano: $\overline{x_c} = \frac{95\%+85\%+80\%+70\%+70\%}{5} = 80\%$

Vicente: $\overline{x_v} = \frac{90\%+95\%+85\%+80\%+50\%}{5} = 80\%$

Já que os dois jogadores obtiveram a mesma média, como proceder cientificamente, para determinar qual dos dois teve o melhor desempenho na avaliação?

Será contratado o jogador que menos oscila, logo é necessário recorrer as Medidas de Dispersão ou de Variabilidade, pois elas qualificam os valores de uma variável, ressaltando a maior ou menor dispersão (ou variabilidade), entre esses valores e uma respectiva medida de posição.

4.2.1.1 Desvio Médio

O desvio médio mede a dispersão média entre os valores da distribuição e a média dos dados coletados, através da média da diferença entre os valores da distribuição e a média dos dados coletados em módulo. O desvio médio é dado pela fórmula $Dm = \frac{|x_1-\bar{x}|+|x_2-\bar{x}|+|x_3-\bar{x}|+\cdots+|x_n-\bar{x}|}{n}$

A comparação entre os dois desempenhos pode ser feita através do desvio médio, pois ele determina o quanto cada frequência está afastada da média.

Cristiano: $Dm = \frac{|95-80|+|85-80|+|80-80|+|70-80|+|70-80|}{5} = 8.$

Vicente: $Dm = \frac{|90-80|+|95-80|+|85-80|+|80-80|+|50-80|}{5} = 12.$

Desse modo, os arremessos de Cristiano estão, em média, 0,8 acima ou abaixo da média, enquanto os arremessos de Vicente estão, em média, 1,2 acima ou abaixo da média aritmética (8,0). Isso mostra que os arremessos de Cristiano são menos dispersos que os de Vicente. Então, Cristiano merece a vaga no time.

4.2.1.2 Variância

Outra forma de evitar o uso de números negativos é, ao invés de utilizar o módulo da diferença dos valores da distribuição e a média dos dados coletados, podemos elevá-los ao quadrado gerando, assim a variância, representada por Var.

Sendo assim a variância é dada pela fórmula

$Var = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}.$

Vamos analisar o exemplo 4.11 através da variância:

Cristiano: $Var = \frac{(95-80)^2+(85-80)^2+(80-80)^2+(70-80)^2+(70-80)^2}{5} = 90.$

Vicente: $Var = \frac{(90-80)^2+(95-80)^2+(85-80)^2+(80-80)^2+(50-80)^2}{5} = 250.$

Logo, por esse processo, Cristiano também teve um desempenho mais regular.

4.2.1.3 Desvio Padrão

No cálculo da variância, tivemos que elevar os valores ao quadrado para eliminar os resultados negativos, o que torna difícil a análise dos dados. Para facilitar a interpretação desses dados, extraímos a raiz quadrada da variância, gerando o desvio padrão, representado por DP, logo: $DP = \sqrt{Var}$

Ainda analisando o desempenho de Cristiano e Vicente, temos que:

Cristiano: $DP = \sqrt{90} \cong 9{,}48$

Vicente: $DP = \sqrt{250} \cong 15{,}81$

Logo, por esse processo, as notas de Cristiano são menos dispersas que as notas de Vicente, ou seja, Cristiano é sempre melhor que Vicente, e deve ser contratado.

4.2.2 Medidas de dispersão com dados agrupados

No cálculo das medidas de dispersão para dados agrupados, usamos a frequência f_i como o peso para multiplicar cada módulo ou quadrado da diferença dos valores da distribuição e a média dos dados coletados. Sendo assim, a formulação matemática para o cálculo das medidas de dispersão em dados agrupados, com intervalos ou sem intervalos, é a mesma, e no caso de dados agrupados com intervalos, x_i representa o ponto médio da classe.

4.2.2.1 Desvio Médio com dados agrupados

O desvio médio para dados agrupados é dado pela fórmula

$$Dm = \frac{|x_1 - \bar{x}|f_1 + |x_2 - \bar{x}|f_2 + |x_3 - \bar{x}|f_3 + \cdots + |x_n - \bar{x}|f_n}{n}$$

Vamos retomar o exemplo 4.8. Observando a tabela 8 da variável discreta número de gols por partida em um torneio de futebol, dada sem intervalos de classe, para o cálculo do desvio médio.

Primeiro, calculamos a média de gols por partida

$$\bar{x} = \frac{6 \cdot 0 + 12 \cdot 1 + 15 \cdot 2 + 5 \cdot 3 + 4 \cdot 2}{6 + 12 + 15 + 5 + 2} \Rightarrow \bar{x} = \frac{0 + 12 + 30 + 15 + 8}{40} \Rightarrow \bar{x} = \frac{65}{40} \Rightarrow$$

$$\bar{x} = 1{,}625.$$

Em seguida, calculamos o desvio médio

$$Dm = \frac{|0 - 1{,}625|6 + |1 - 1{,}625|12 + |2 - 1{,}625|15 + |3 - 1{,}625|3 + |4 - 1{,}625|2}{40}$$

$$Dm = \frac{1{,}625 \cdot 6 + 0{,}625 \cdot 12 + 0{,}375 \cdot 15 + 1{,}375 \cdot 3 + 2{,}375 \cdot 2}{40}$$

$$Dm = \frac{9{,}75 + 7{,}5 + 5{,}625 + 4{,}125 + 4{,}75}{40}$$

$$Dm = \frac{31{,}74}{40}$$

$$Dm = 0{,}7935.$$

Analisemos agora o exemplo 4.9. Observando a Tabela 9 da variável contínua, peso das atletas de ginástica olímpica de um clube, dada por intervalos de classe.

Primeiro calculamos a média

$$\bar{x} = \frac{42 \cdot 1 + 46 \cdot 3 + 50 \cdot 7 + 54 \cdot 6 + 58 \cdot 3}{1 + 3 + 7 + 6 + 3} \Rightarrow \bar{x} = \frac{84 + 138 + 350 + 324 + 174}{20} \Rightarrow \bar{x} = \frac{1070}{20} \Rightarrow$$

$\bar{x} = 53{,}5$

Em seguida calculamos o desvio médio

$$Dm = \frac{|42 - 53{,}5|1 + |46 - 53{,}5|3 + |50 - 53{,}5|7 + |54 - 53{,}5|6 + |58 - 53{,}5|3}{20}$$

$$Dm = \frac{11{,}5 \cdot 1 + 7{,}5 \cdot 3 + 3{,}5 \cdot 7 + 0{,}5 \cdot 6 + 4{,}5 \cdot 3}{20}$$

$$Dm = \frac{11{,}5 + 22{,}5 + 24{,}5 + 3 + 13{,}5}{20}$$

$$Dm = \frac{75}{20}$$

$$Dm = 3{,}75.$$

4.2.2.2 Variância com dados agrupados

A variância para dados agrupados é dada pela fórmula

$$Var = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + (x_3 - \bar{x})^2 f_3 + \cdots + (x_n - \bar{x})^2 f_n}{n}$$

Vamos para o cálculo da variância do exemplo 4.8

$$Var = \frac{(0 - 1{,}625)^2 6 + (1 - 1{,}625)^2 12 + (2 - 1{,}625)^2 15 + (3 - 1{,}625)^2 3 + (4 - 1{,}625)^2 2}{40}$$

$$Var = \frac{(1{,}625)^2 6 + (0{,}625)^2 12 + (0{,}375)^2 15 + (1{,}375)^2 3 + (2{,}375)^2 2}{40}$$

$$Var = \frac{2{,}640625 \cdot 6 + 0{,}390625 \cdot 12 + 0{,}140625 \cdot 15 + 1{,}890625 \cdot 3 + 5{,}640625 \cdot 2}{40}$$

$$Var = \frac{15{,}84375 + 4{,}6875 + 3{,}109375 + 5{,}671875 + 11{,}28125}{40}$$

$$Var = \frac{40{,}59375}{40}$$

$$Var = 1{,}01484375.$$

Dando continuidade, iremos para o cálculo da variância do exemplo 4.9

$$Var = \frac{(42-53{,}5)^2 1 + (46-53{,}5)^2 3 + (50-53{,}5)^2 7 + (54-53{,}5)^2 6 + (58-53{,}5)^2 3}{20}$$

$$Var = \frac{(11{,}5)^2 \cdot 1 + (7{,}5)^2 \cdot 3 + (3{,}5)^2 \cdot 7 + (0{,}5)^2 \cdot 6 + (4{,}5)^2 \cdot 3}{20}$$

$$Var = \frac{11{,}5 \cdot 1 + 56{,}25 \cdot 3 + 12{,}25 \cdot 7 + 0{,}25 \cdot 6 + 20{,}25 \cdot 3}{20}$$

$$Var = \frac{11{,}5 + 168{,}75 + 85{,}75 + 3 + 60{,}75}{20}$$

$$Var = \frac{329{,}75}{20}$$

$$Var = 16{,}4875$$

4.2.2.3 Desvio padrão com dados agrupados

O Desvio Padrão em todos os casos é sempre definido como a raiz quadrada da Variância. Assim temos: $DP = \sqrt{Var}$

No exemplo 4,8 $DP = \sqrt{1{,}01484375} \Rightarrow DP \cong 1{,}0074$

No exemplo 4,9 $DP = \sqrt{16{,}4875} \Rightarrow DP \cong 4{,}06$

5. APLICAÇÕES AO ESPORTE COM O USO DE SOFWARES

Já vimos que, a grande maioria dos atletas de ponta e das equipes esportivas tem um profissional de estatística em sua comissão técnica. Agora que conhecemos os principais conceitos estatísticos veremos como as equipes, técnicos, jornalistas e torcedores podem obter os dados estatísticos para analisá-los. Também veremos como esses dados são obtidos.

Antigamente, quando um competidor de judô ia disputar pela primeira vez uma competição fora do país, ele tinha uma enorme dificuldade, pois não conhecia a maioria dos seus oponentes; mas hoje, esse problema não ocorre, já que existem aplicativos, em que ao digitar o nome do atleta, aparecem diversos dados estatísticos tais como: número de lutas disputadas, quantidade de vitórias, de derrotas, quais golpes ele mais utiliza, quais seus pontos fracos, vídeos, entre outros.

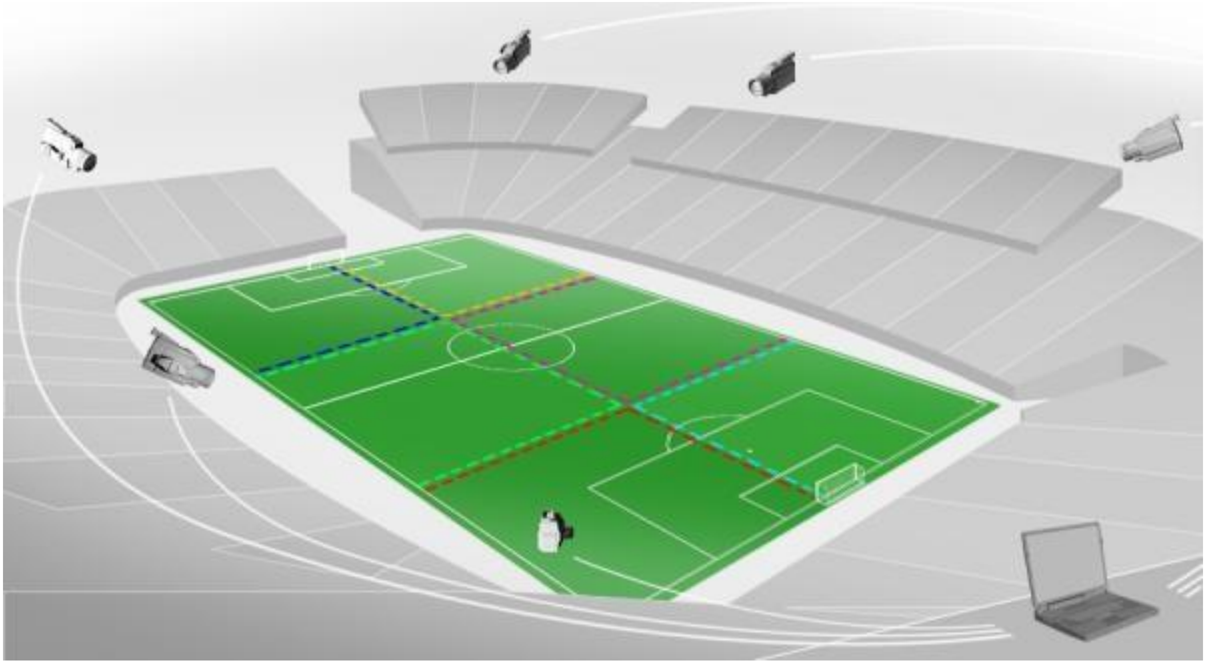
O uso desse aplicativo, já pode ser feito em vários outros esportes, com o uso de programas (softwares) como o Winnerfoot, Scoolt, Volleyplus e o HANA (High-Performance Analytic Appliance) desenvolvidos pela empresa alemã SAP, este, que foi usado pela seleção da Alemanha campeã da Copa Do Mundo de Futebol de 2014.

Exemplificando, em uma disputa de mata-mata pode ser feito a análise das cobranças dos principais cobradores de pênalti, para que o goleiro possa ter a probabilidade das cobranças deles, em caso do jogo ir para a disputa de pênaltis. Além desse software, existem sites voltados a fornecer dados estatísticos, muito utilizados por jornalistas, como o infobola.com.br, footstats.net, statsport.com.br e www.sportytrader.

Por outro lado, ter acesso a vários dados do esporte não seria possível sem as tecnologias de câmeras de reconhecimento, algoritmos complexos e cálculos que levariam dias para serem feitos por pessoas, e que são feitos em poucos segundos com a ajuda de um computador e de câmeras.

Segundo Kleina (2012), o processo de captura e obtensão dos dados são bastante simples, configurá-lo, exige tanta paixão pela computação quanto pelo esporte. Primeiramente, são instaladas câmeras ao redor do local da competição ou do treinamento. Ao contrário daquelas usadas na televisão e que precisam sempre acompanhar a bola ou o atleta, essas permanecem estáticas e gravam tudo o que se passa por uma determinada zona do campo, tatame da pista, ou seja, do local de competição para depois essas imagens serem analisadas pelos softwares.

Figura 5 - Câmeras gravam setores do campo.



Fonte : Baixaki/Tecmundo

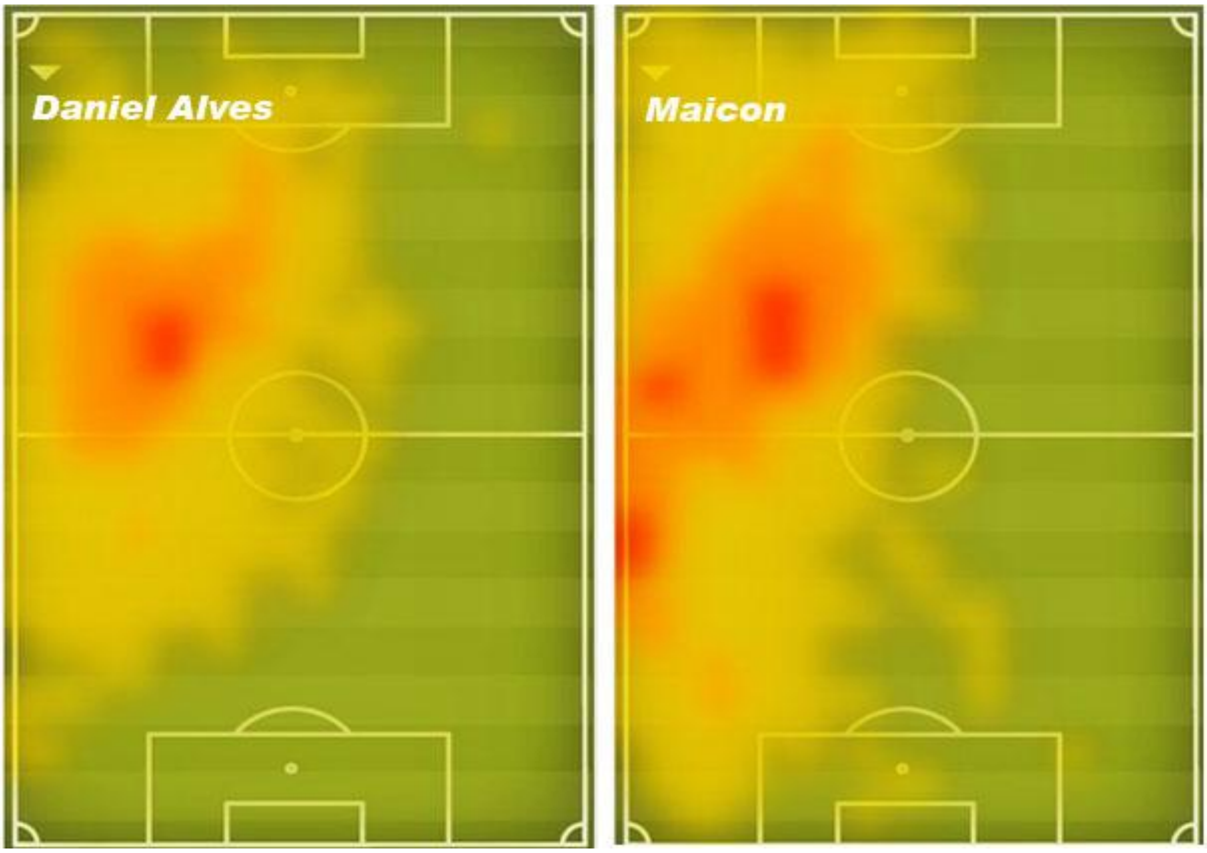
Por exemplo, no futebol durante a partida as câmeras gravam as zonas do campo, esse material é passado para um computador, que analisa a partida e começa a contagem de passes, desarmes, chutes a gol, distância percorrida, velocidade máxima atingida por um jogador, velocidade média e até definir quem foi o melhor jogador da partida. Tudo é inspecionado por especialistas em informática e no mundo da bola.

Com esse propósito, os softwares são programados para reconhecer a bola, mesmo que ela esteja em movimento. Em seguida, os atletas são marcados individualmente a partir do mesmo método, sendo que cada perna do jogador recebe algoritmo diferenciado.

Com tudo configurado, é só aguardar o computador “assistir” a tudo para fornecer os dados de quantos desarmes a zaga de um time realizou, o deslocamento de um determinado jogador ou de qual lado do campo saíram mais ataques e depois fornecer as frequências absolutas, relativas, às médias, medianas, desvio padrão, entre outras. Feito isso, é só analisar esses dados para inúmeras decisões para melhoria esportiva.

Segundo Dias (2013) um bom exemplo foi a disputa por posição de o lateral direito, titular da seleção brasileira de futebol, que participou da Copa das Confederações de 2013, entre os jogadores Daniel Alves e Maicon.

Figura 6 - Gráfico de comparação de movimentação do lateral direito da seleção brasileira de futebol, em jogos da Copa das Confederações de 2013.



Fonte: http://globoesporte.globo.com visualizada em 24/03/2016

O gráfico acima, mostra em dois jogos da Copa das Confederações de 2013, a movimentação do lateral direito da seleção, Daniel Alves, que atuou contra a seleção do Egito e o lateral Maicon, que atuou contra a seleção dos Estados Unidos.

A parte de cima do gráfico corresponde ao setor da defesa brasileira e a parte de baixo a do ataque. A zona amarela mostra por onde o jogador se movimentou e a vermelha indica a região em que o jogador mais se concentrou. Note que Daniel Alves e Maicon apoiaram bastante o ataque, principalmente Maicon, que ganhou a posição de titular da seleção, antes, era do colega de seleção, ainda que, ocupando a mesma área a maior parte do tempo, principalmente atrás da linha do meio-campo, ajudando a defesa e apoiando bastante o ataque.

Observe que Maicon se movimentou mais que Daniel Alves. Se não fossem utilizadas essas câmeras no esporte, dificilmente teríamos informações como essas. Obter esses dados requer de uma pessoa a responsabilidade de observar e acompanhar cada jogador durante todo o jogo. Perca um desarme e as estatísticas já saem erradas porque nem tudo acontece com a bola no pé, já que há fatores como posicionamento tático e marcação. Seria impossível saber o setor do campo que tal atleta mais atua, qual foi sua distância percorrida na partida, sua velocidade média e sua velocidade máxima atingida.

CONSIDERAÇÕES FINAIS

É comum o professor ouvir de seus alunos perguntas como estas: ”para que serve esse assunto?” ou “porque eu tenho que estudar isso”? Esse tipo de curiosidade é de extrema importância, pois alunos instigados são essenciais, tanto para o desempenho das aulas, quanto para a melhoria da aprendizagem. A Probabilidade e a Estatística aplicadas ao esporte podem servir para satisfazer essas curiosidades, uma vez que elas preparam o cérebro para a aprendizagem e torna-a gratificante.

Partindo dessa ideia, o objetivo principal deste trabalho foi apresentar e analisar a importância do uso da Estatística e da Probabilidade no esporte, com o intuito de servir como referência para futuras pesquisas e aplicações em salas de aula, além disso, chamar a atenção e motivar alunos, a partir de exemplos de matemática aplicada e satisfazendo as curiosidades de como são obtidas as probabilidades no futebol, para que serve a estatística nele, e como são obtidos os dados estatísticos.

Vimos como se deu a evolução de probabilidade e da estatística, principalmente com o desenvolvimento tecnológico. Expomos os conceitos básicos de probabilidade, apresentando um modelo para o cálculo das probabilidades no futebol, o que pode servir para satisfazer curiosidades de alunos que apreciam o esporte mais popular do mundo. Do mesmo modo foram apresentados os conceitos básicos e definições de estatística, formas de representá-los, as medidas de posição, as medidas de dispersão, assim como são obtidos esses dados no esporte e onde eles podem ser adquiridos.

Em síntese, é preciso resaltar que o esporte, sobretudo o futebol, não é uma ciência exata, mas os números sempre dizem alguma coisa.

REFERÊNCIAS BIBLIOGRÁFICAS

CAMINHA, Antonio. **Desigualdades Elementares**. Eureka! (Rio de Janeiro), Rio de Janeiro, v. 5, 1999.

CBF, Confederação Brasileira de Futebol. **Tabela do Campeonato Brasileiro de Futebol de 2012 da serie A**, Disponível em: <www.cbf.com.br/ brasileiro-serie-a/classificacao/2012>. Acesso em 28/12/2015.

CBF, Confederação Brasileira de Futebol. **Tabela do Campeonato Brasileiro de Futebol de 2015 da serie A após a 37° Rodada**, Disponível em: <www.cbf.com.br/ brasileiro-serie-a/classificacao/2015>. Acesso em 30/11/2015.

CBV, Confederação Brasileira de Voleibol. **Atletas da Seleção Brasileira de Voleibol que disputou a Liga Mundial em 2015**, Disponível em < http: //ligamundial.cbv.com.br/index.php/selecao-brasileira>. Acesso em 28/12/2015.

COSTA, Gilcione Nonato. **Probabilidades no futebol**. Bernardo Nunes Borges de Lima, Fábio Brochero, Gilcione Nonato Costa, Gustavo Zeferino, Marcelo Terra Cunha e Renato Vidal Martins, Minas Gerais: UFMG, 2009.

DANTE, Luiz Roberto. **Matemática: contexto & aplicações,** Luiz Roberto Dante, 2°. Edição, São Paulo, Ática, 2013.

DIAS, Thiago. **Sem aparecer no ataque, lateral-esquerdo torna-se fundamental para a seleção.** Disponível em < http://globoesporte.globo.com/Esportes/ Noticias/Times/Selecao_Brasileira/>. Acesso em 13/03/2016.

DOANE, David P. **Estatística Aplicada à Administração e Economia.** David P. Doane, Lori E. Seward, 4° Edição, Porto Alegre, AMGH Editora Ltda, 2014.

FANTASTICO, O Show da Vida. **Gols do Fantástico: Vasco vence e continua na briga para não cair.** Disponível em: <http://g1.globo.com/fantastico/noticia/2015/11/ gols-do-fantastico-vasco-vence-e-continua-na-briga-para-nao-cair>. Acesso em 01/12/2015.

KLEINA, Nilton. **Como Softwares Analisam o Futebol**. Disponível em: <http://www.tecmundo.com.br/futebol/26738-como-softwares-analisam-partidas-de-futebol>. Acesso em: 11/03/2016.

LIMA, Luciana de. **Estatística aplicada: semestre II.** Coordenação Cassandra Ribeiro de Oliveira e Silva. – Fortaleza: UAB/IFCE, 2009.

LIPSCHUTZ, SEYMOUR. **Probabilidade,** tradução Rutth Ribas Itacarabi; revisão técnica Helio Migon; revisão técnica 4. Ed. Dario Nery. São Paulo: Makron Books, 1993.

MEMÓRIA, José Maria Pompeu. **Breve história da Estatística.** Brasília: EMBRAPA, 2004.

TIBONI, Conceição Gentil Rabelo. **Estatística básica para o curso de turismo**. São Paulo, Atlas, 2003.