



MICHELLY GUERRA COSTA

**ARREMESSOS DE BASQUETEBOL E SEQUÊNCIAS DE
BERNOULLI: UMA APLICAÇÃO DE MÉTODOS
ESTATÍSTICOS PARA ANÁLISE DE SÉRIES TEMPORAIS
BINÁRIAS**

CAMPINAS

2012



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA

MICHELLY GUERRA COSTA

ARREMESSOS DE BASQUETEBOL E SEQUÊNCIAS DE BERNOULLI:
UMA APLICAÇÃO DE MÉTODOS ESTATÍSTICOS PARA ANÁLISE DE
SÉRIES TEMPORAIS BINÁRIAS

Orientador: Prof. Dr. Cristiano Torezzan

Dissertação de mestrado apresentada ao Instituto de Matemática, Estatística e Computação Científica da Unicamp para obtenção do título de Mestra em Matemática.

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELA ALUNA MICHELLY GUERRA COSTA, E ORIENTADA PELO PROF. DR. CRISTIANO TOREZZAN

Assinatura do Orientador

A handwritten signature in black ink, appearing to be "Cristiano Torezzan", written over a horizontal line.

CAMPINAS

2012

FICHA CATALOGRÁFICA ELABORADA POR
MARIA FABIANA BEZERRA MULLER - CRB8/6162
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

Costa, Michelly Guerra, 1984-
C823a Arremessos de basquetebol e sequências de Bernoulli : uma aplicação de métodos estatísticos para análise de séries temporais binárias / Michelly Guerra Costa. – Campinas, SP : [s.n.], 2012.

Orientador: Cristiano Torezzan.
Dissertação (mestrado profissional) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Análise de séries temporais. 2. Algoritmos de computador. 3. Análise sequencial - Métodos estatísticos. 4. Estatística matemática. I. Torezzan, Cristiano, 1976-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em inglês: Basketball shots and sequences of Bernoulli : an application of statistical methods for analysis of binary time series

Palavras-chave em inglês:

Time-series analysis
Computer algorithms
Sequential analysis - Statistical methods
Mathematical statistics

Área de concentração: Matemática

Titulação: Mestra em Matemática

Banca examinadora:

Cristiano Torezzan [Orientador]
João Eloir Strapasson
Luciano Allegretti Mercadante

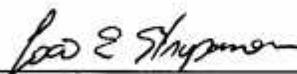
Data de defesa: 14-12-2012

Programa de Pós-Graduação: Mestrado Profissional em Matemática

Dissertação de Mestrado Profissional defendida em 14 de dezembro de 2012 e aprovada pela Banca Examinadora composta pelos Profs. Drs.



Prof. (a). Dr (a). CRISTIANO TOREZZAN



Prof. (a). Dr (a). JOÃO ELOIR STRAPASSON



Prof. (a). Dr (a). LUCIANO ALLEGRETTI MERCADANTE

*Às pessoas que me apoiaram e torceram para
a realização desta dissertação de mestrado.*

Agradecimentos

Agradeço em especial à Renata M. R. Miranda. Todo o meu agradecimento será sempre insuficiente diante da sua generosidade e disposição em me ajudar. Sempre será lembrada com muito carinho e gratidão.

À todos que me ajudaram no decorrer do desenvolvimento desta dissertação com incentivo, compreensão, afeto e respeito, contribuindo para esta minha conquista.

Ao Prof. Dr. Cristiano Torezzan. Não tenho palavras para expressar o quanto sou grata pela sua orientação, ensinamentos, paciência e incentivo para o desenvolvimento deste trabalho.

Aos professores que ministraram as disciplinas deste programa de mestrado.

Ao grupo do Laboratório de Biomecânica e Instrumentação (LABIN) da FCA - UNICAMP.

À toda a equipe do SAE, SAPPE, DAC e da Biblioteca do IMECC.

“Se quiser ir rápido, vá sozinho.

Se quiser ir longe, vá com alguém.”

Ditado Africano

Resumo

No presente trabalho investigamos as semelhanças entre registros de lançamentos extraídos de jogos reais de basquetebol e sequências aleatórias geradas por algoritmos computacionais. Nosso principal objetivo é comparar o comportamento das sequências de arremessos consecutivos de jogadores ao longo de uma temporada com sequências aleatórias de zeros e uns (sequências de Bernoulli) geradas computacionalmente. Para tanto, foram desenvolvidos algoritmos computacionais específicos para os testes e implementados na plataforma Scilab para análise dos dados. Os testes realizados neste trabalho indicam que, de maneira geral, não existem diferenças estatísticas entre os dois conjuntos de dados considerados. Além de uma breve revisão sobre testes estatísticos e sobre o problema da boa fase no jogo de basquetebol, apresentamos diversos exemplos visando tornar o texto acessíveis para alunos de graduação ou demais interessados em estatística aplicada, em especial ao esporte.

Palavras-chave: Métodos Estatísticos, Sequências de Bernoulli, Algoritmos Computacionais, Arremessos de Basquetebol.

Abstract

In this work we investigate the similarities between records of shots extracted from real basketball games and random sequences generated by computational algorithms. Our goal is to compare the behavior of sequences of consecutive shots of players throughout a season with random sequences of zeros and ones (Bernoulli sequences) generated computationally. For this purpose, computational algorithms have been developed for specific tests and implemented in the Scilab platform. The tests performed in this study indicate that, in general, there are no statistical differences between the two sets of data considered. Besides a brief review of statistical tests and the problem of good stage in the game of basketball, we present several examples in order to make the text accessible to undergraduates or others interested in applied statistics, especially the ones concerning this sport.

Keywords: Statistical Methods, Sequences of Bernoulli, Computational Algorithms, Throws Basketball.

Sumário

Resumo	ix
Abstract	x
Introdução	1
1 Introdução a Conceitos Estatísticos	3
1.1 Noções de Probabilidade e Estatística	3
1.1.1 Populações e Amostras	3
1.1.2 Variáveis Qualitativas e Quantitativas	4
1.1.3 Experimentos Aleatórios, Espaço Amostral e Eventos	4
1.1.4 Probabilidade	5
1.1.5 Probabilidade Condicional	6
1.1.6 Independência de Eventos	7
1.2 Variáveis Aleatórias e Distribuição de Probabilidade	7
1.2.1 Distribuição de Probabilidade	8
1.2.2 Função de Probabilidade	8
1.2.3 Função de Distribuição Acumulada	9
1.3 Alguns Modelos Discretos	9
1.3.1 Distribuição de Bernoulli	9
1.3.2 Distribuição Binomial	10
1.3.3 Distribuição de Poisson	11
1.4 Distribuição Normal: Um Modelo Contínuo	12
1.5 Conceitos de Séries Temporais	15
1.5.1 Noções Básicas	15
1.6 Testes Estatísticos	16
1.6.1 Teste de Aleatoriedade	17

1.6.2	Correlação Serial	19
1.6.2.1	Correlograma	20
1.6.3	Diferença entre Duas Proporções	23
2	Investigando a existência de boa fase no jogo de basquetebol: o polêmico artigo sobre “hot hand”	26
2.1	Estudo 1: pesquisa com torcedores de basquetebol	27
2.2	Estudo 2: registros de objetivo de campo do basquete profissional	27
2.3	Estudo 3: registros de lance livre do basquete profissional	32
2.4	Estudo 4: experimento controlado de lançamentos	33
2.5	Nova Análise	34
3	Análise estatística de arremessos de basquetebol	37
3.1	Preparação e processamento dos dados	38
3.2	Análise 1: Lançamentos de 3 Pontos	40
3.2.1	Teste da Diferença entre Duas Proporções	41
3.2.2	Correlação Serial de Ordem 1	42
3.2.3	Análise de Runs e o Teste de Aleatoriedade	44
3.3	Análise 2: Lançamentos Livres	45
3.3.1	Teste da Diferença entre Duas Proporções	47
3.3.2	Correlação Serial de Ordem 1	48
4	Análise estatística dos dados simulados computacionalmente	50
4.1	Simulação 1: Lançamentos de 3 Pontos	50
4.1.1	Teste da Diferença entre Duas Proporções	52
4.1.2	Correlação Serial de Ordem 1	53
4.1.3	Análise de Runs e o Teste de Aleatoriedade	54
4.2	Simulação 2: Lançamentos Livres	55
4.2.1	Teste da Diferença entre Duas Proporções	56
4.2.2	Correlação Serial de Ordem 1	56
	Conclusão	58
	Considerações Finais	60

SUMÁRIO

xiii

Referências Bibliográficas

62

Anexos

64

Introdução

No cenário esportivo mundial, o uso de ferramentas analíticas, tais como modelagem matemática, métodos estatísticos e algoritmos computacionais, tem sido cada vez mais frequente como forma de auxiliar os profissionais em processos de tomada de decisão. Nos esportes de alto rendimento, existe um particular interesse em estudos estatísticos sobre os registros individuais e coletivos dos jogadores ao longo do tempo, seja no decorrer de um jogo ou até mesmo no transcorrer de uma temporada.

Neste trabalho pesquisamos as semelhanças existentes entre registros de lançamentos extraídos de jogos de basquetebol e sequências aleatórias geradas por algoritmos computacionais. O estudo tem como ponto de partida o popular artigo "The hot hand in basketball: on the misperception of random sequences" [5], no qual os autores investigam a origem e a validade das crenças comuns sobre a existência de "hot hand" e "streak shooting", comumente traduzidos para o português como "mão quente" e "boa fase" dos jogadores, através de pesquisas e experimentos realizados com torcedores, atletas e treinadores de basquetebol. O artigo é centrado na seguinte questão: os jogadores acertam uma maior porcentagem de seus arremessos depois de acertos ou depois de erros recentes?

Para o desenvolvimento de nossos estudos, tivemos como base os registros de lançamentos da equipe de basquete profissional de Bauru, ocorridos na temporada 2009 – 2010. Os desempenhos dos jogadores foram obtidos através de tabulação feita por meio de vídeos dos jogos fornecidos pela Confederação Brasileira de Basketball (CBB). Os dados considerados para as análises foram as sequências de lances livres e lançamentos de três pontos dos dez integrantes da equipe estudada que obtiveram maior número de arremessos na temporada. As simulações computacionais, para produzir sequências (pseudo) aleatórias para cada um dos jogadores considerados, foram feitas de forma que cada sequência tivesse o mesmo número de lances e a mesma frequência relativa de acertos daquelas adquiridas dos jogos. Os mesmos estudos realizados sobre dados reais foram aplicados em dados simulados.

Este trabalho é apresentado em duas principais vertentes: a primeira trata-se de desenvolver algoritmos para exploração e manipulação de dados e aplicação de métodos estatísticos baseados

nas sequências reais de lançamentos (livres e de três pontos) consecutivos; a segunda trata-se de gerar sequências aleatórias que representem esses lances, para aplicar sobre estas os mesmos testes aplicados em dados reais. O trabalho aborda em seu primeiro capítulo uma breve revisão de Probabilidade e Estatística; o segundo capítulo trás as idéias principais do artigo [5]; os terceiro e quarto capítulos apresentam, respectivamente, as análises realizadas sobre registros reais e simulados de arremessos consecutivos.

Esperamos que este texto possa servir de incentivo à estudantes e profissionais interessados na aplicação de estatística e métodos quantitativos em esporte ou em outras situações cotidianas, cumprindo, assim, mais um dos objetivos deste programa de mestrado.

INTRODUÇÃO A CONCEITOS ESTATÍSTICOS

Neste primeiro capítulo apresentamos uma breve revisão sobre os principais conceitos de Probabilidade e Estatística que serão utilizados nesta dissertação. Uma referência mais completa sobre este assunto pode ser encontrada em [2, 8, 9, 10, 12], que constituem a base teórica deste capítulo.

1.1 Noções de Probabilidade e Estatística

1.1.1 Populações e Amostras

A disciplina Estatística fornece um conjunto de técnicas destinadas a organizar, analisar e interpretar dados e tirar conclusões a respeito de características de interesse presentes em um conjunto de informações quantificáveis.

Em uma investigação, o grande conjunto de dados que contém características de interesse denomina-se *população* [7]. Assim, dependendo do estudo, a população pode ser todas as barras de chocolates produzidas por uma fábrica durante um determinado período, ou todos os livros de uma biblioteca específica ou todos os lances de três pontos de um time de basquete ao longo de uma temporada – dentre outros exemplos. Um subconjunto da população é denominado *amostra*.

1.1.2 Variáveis Qualitativas e Quantitativas

É frequente o interesse em certas características dos objetos de uma população, como altura, peso, sexo e idade de um indivíduo, número de arremessos convertidos, por exemplo. Cada uma dessas características é denominada *variável*.

As variáveis podem ser classificadas em qualitativas ou quantitativas, como se verifica nas definições a seguir.

Definição 1.1 *Uma variável é denominada qualitativa quando os possíveis valores que assume representam atributos, qualidades.*

Definição 1.2 *Uma variável é denominada quantitativa quando possui natureza numérica.*

As variáveis qualitativas podem ser subdivididas em *ordinais* e *nominais*. Variáveis tais como tamanho (pequeno, médio ou grande) e classe social (baixa, média ou alta) são classificadas como *qualitativas ordinais*. Por outro lado, variáveis como livro (bom ou ruim) e sexo (masculino ou feminino) são classificadas como *qualitativas nominais*.

As variáveis quantitativas podem ser subdivididas em *discretas* e *contínuas*. Uma variável é *quantitativa discreta* se o seu conjunto de valores possíveis for finito ou enumerável. Variável *quantitativa contínua* assume valores no conjunto dos números reais. Variáveis tais como número de filhos (0, 1, 2,...), número de irmãos (0, 1, 2,...) e números de lances livres de uma equipe num jogo de basquetebol são classificadas como *quantitativas discretas*, enquanto que peso, altura e valor do imposto de renda pago por um jogador são *quantitativas contínuas*.

1.1.3 Experimentos Aleatórios, Espaço Amostral e Eventos

Definição 1.3 *Experimento Aleatório é qualquer processo de observação que ao ser repetido produz uma sequência de resultados.*

A seguir encontram-se alguns exemplos de experimentos aleatórios:

Exemplo 1.1 *Escolher um representante ao acaso em um grupo de 50 alunos.*

Exemplo 1.2 *Observar o tempo de vida útil de uma lâmpada no período de um ano.*

Exemplo 1.3 *Contar o número de arremessos de três pontos convertidos por um jogador ao longo de um quarto de partida de basquetebol.*

Definição 1.4 *O conjunto de todos os resultados possíveis de um experimento é chamado de espaço amostral, e será denotado por Ω .*

Os espaços amostrais associados aos experimentos exemplificados acima são, respectivamente:

$$\Omega_1 = \{A_1, A_2, \dots, A_{50}\}, \text{ onde } A_i \text{ representa o aluno } i$$

$$\Omega_2 = \{t \in \mathbb{R}, t \geq 0\}$$

$$\Omega_3 = \{a \in \mathbb{N}\}$$

Definição 1.5 *Seja Ω o espaço amostral associado a um experimento aleatório. Todo resultado ou subconjunto de resultados de um experimento é denominado evento.*

Exemplo 1.4 *Os eventos associados aos exemplos 1.1, 1.2 e 1.3 são, respectivamente:*

A_1 : o representante escolhido foi Pedro

A_2 : o tempo de vida útil da lâmpada foi maior que 180 horas

A_3 : o número de acertos do jogador Oscar no primeiro quarto da partida foi 5

A união de dois eventos A e B, representada por $A \cup B$, consiste na ocorrência de, pelo menos, um dos eventos A ou B. A interseção de dois eventos A e B, representada por $A \cap B$, consiste na ocorrência de ambos os eventos. Quando dois eventos não possuem elementos em comum, ou seja, $A \cap B = \emptyset$, diz-se que os eventos são *disjuntos* ou *mutualmente exclusivos*.

Os eventos A e B são *complementares* se sua união resulta no espaço amostral, $A \cup B = \Omega$, e sua interseção é vazia, $A \cap B = \emptyset$. Considere A^c o complementar de A, logo $A \cup A^c = \Omega$ e $A \cap A^c = \emptyset$.

1.1.4 Probabilidade

As pessoas possuem uma idéia intuitiva do que é probabilidade. Muitas vezes o fenômeno é previsível: ao soltar uma pedra várias vezes da mesma altura e sob as mesmas condições, a velocidade da queda será sempre a mesma, pois obedece a uma lei da física. Outras vezes, o fenômeno é imprevisível, porém tem um padrão a longo prazo. Por exemplo, não se sabe o sexo da próxima criança que vai nascer numa determinada cidade, mas se sabe que a longo prazo, um ano por exemplo, nascerão meninos e meninas quase na mesma proporção.

No contexto do basquete, não é possível afirmar com certeza se um determinado jogador acertará o próximo arremesso de lance livre que fizer, mas sabendo que nos treinos ele acerta 80% e supondo que esta taxa permanece igual no jogo, podemos intuir que a probabilidade dele acertar é de 80%.

Definição 1.6 *Considerando um experimento aleatório e um espaço amostral Ω , o valor $P(A)$, denominado probabilidade do evento A, fornece a chance de ocorrência do evento A, e satisfaz certas propriedades:*

a) Para todo evento $A \in \Omega$, $0 \leq P(A) \leq 1$

b) $P(\Omega) = 1$

Pode-se encontrar uma teoria bastante vasta sobre probabilidade na literatura. Neste trabalho usaremos apenas alguns conceitos e propriedades básicas como as listadas a seguir:

i) Se $(A_n)_{n \geq 1}$ é uma sequência finita de eventos mutuamente exclusivos, então:

$$P\left(\bigcup_{n=1}^k A_n\right) = \sum_{n=1}^k P(A_n)$$

ii) Para qualquer evento A , $P(A) = 1 - P(A^c)$, sendo $P(A^c)$ a probabilidade de A complementar.

iii) Sejam A e B eventos mutuamente exclusivos, então $P(A \cap B) = 0$.

iv) Para quaisquer dois eventos A e B ,

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

Quando A e B não são mutuamente exclusivos, a *regra de adição de probabilidades* é obtida do resultado iv).

1.1.5 Probabilidade Condicional

Podem haver situações nas quais queira saber a probabilidade de um evento A condicionada a ocorrência de um evento B , o que é dado pelo cálculo da probabilidade condicional.

Definição 1.7 Sejam A e B dois eventos quaisquer, com $P(B) > 0$. A probabilidade condicional de A dado que ocorreu B é definida por:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (1.1)$$

Multiplicando ambos os membros da equação 1.1 por $P(B)$, obtém-se a *Regra do Produto*:

$$P(A \cap B) = P(A | B) \cdot P(B) \quad (1.2)$$

Essa regra é muito útil quando o experimento envolve mais de duas etapas:

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3 | A_1 \cap A_2) \cdot P(A_1 \cap A_2) \\ &= P(A_3 | A_1 \cap A_2) \cdot P(A_2 | A_1) \cdot P(A_1) \end{aligned}$$

sendo que A_1 ocorre primeiro, A_2 em seguida e A_3 por último.

Exemplo 1.5 O lançamento de um dado produz o espaço amostral $\Omega = \{1, 2, 3, 4, 5, 6\}$. Ao lançar um dado, verifica-se que a face voltada para cima é de um número ímpar. Qual a probabilidade desse número ser igual a 3?

Solução: Representado por A o evento que corresponde a $x = 3$ e por B o evento “ x é um número ímpar”, segue que

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{(1/6)}{(1/2)} = \frac{1}{3}$$

1.1.6 Independência de Eventos

Em geral $P(A | B)$ é diferente da probabilidade não condicional $P(A)$. Quando isso acontece, significa que a ocorrência de B resultou em uma alteração da probabilidade de ocorrência de A . Mas, existem outras situações em que a informação da ocorrência ou não de B não altera a probabilidade de ocorrência de A , de forma que $P(A | B) = P(A)$, o que indica a independência de eventos.

Definição 1.8 Dois eventos A e B são independentes se a ocorrência ou não de B não alterar a probabilidade da ocorrência de A , isto é

$$P(A | B) = P(A)$$

1.2 Variáveis Aleatórias e Distribuição de Probabilidade

O espaço amostral de um experimento nem sempre produz resultado numérico. Por exemplo, ao descrever uma peça de carro, podemos classificá-la como *defeituosa* ou *não defeituosa*.

Definição 1.9 Uma Variável Aleatória é uma variável cujo valor é determinado pelo resultado de um experimento aleatório.

Em diversas situações, interessa a mensuração de algo e seu registro como um número. Podemos atribuir um valor real a todo elemento do espaço amostral.

Exemplo 1.6 O lançamento de duas moedas produz o espaço amostral $\Omega = \{CC, CR, RC, RR\}$, sendo $C = cara$ e $R = coroa$. Seja X uma função definida no espaço amostral que determina o número de caras nos dois lançamentos. Assim,

$X = 0$ corresponde ao evento (R, R)

$X = 1$ corresponde ao evento $(R, C), (C, R)$

$X = 2$ corresponde ao evento (C, C)

ou ainda $X(CC) = 2, X(RC) = 1$ e $X(RR) = 0$.

As variáveis aleatórias podem ser classificadas segundo o tipo de contradomínio.

Definição 1.10 Variáveis aleatórias que assumem valores num conjunto enumerável são denominadas discretas. Aquelas que assumem valores num intervalo dos números reais são denominadas contínuas.

1.2.1 Distribuição de Probabilidade

Uma vez definida a variável aleatória discreta, pode existir o interesse em estudar as probabilidades associadas. Dessa forma, o conjunto das variáveis e suas probabilidades é denominado *distribuição de probabilidade*, denotado por $(x_i, p(x_i)), i = 1, 2, \dots, n$.

1.2.2 Função de Probabilidade

Seja X uma variável aleatória discreta. Uma *função de probabilidade* é a função que associa a cada valor assumido pela variável aleatória a probabilidade correspondente, isto é, $P(X = x_i) = p(x_i) = p_i$, para $i = 1, 2, \dots$; ou ainda,

X	X ₁	X ₂	X ₃	...
p _i	p ₁	p ₂	p ₃	...

e satisfaz as seguintes asserções:

1. $0 \leq p_i \leq 1$
2. $\sum p_i = 1$

1.2.3 Função de Distribuição Acumulada

Definição 1.11 *Seja X uma variável aleatória discreta com função de probabilidade $p(x)$. A função de distribuição acumulada (FDA) de X denotada por $F(x)$ é definida como*

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y) \quad \text{ou} \quad \sum p(x_i)$$

com $x_i \leq x$.

Em outras palavras, a FDA é a probabilidade de a variável aleatória X assumir valores menores ou iguais a x .

Lema: *A função de distribuição acumulada de uma variável aleatória X satisfaz as seguintes condições:*

1. $0 \leq F(x) \leq 1$
2. $F(x)$ é não decrescente e contínua à direita
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$

1.3 Alguns Modelos Discretos

Ao estudar exemplos que envolvam variáveis aleatórias (discretas) é possível estudar o comportamento da variável em questão através de função de probabilidade.

Esta seção apresenta algumas das principais distribuições de probabilidade (ou funções de probabilidade) utilizadas para descrever experimentos discretos, que podem modelar situações encontradas na natureza.

1.3.1 Distribuição de Bernoulli

Existem variáveis que assumem apenas dois valores, por exemplo um objeto pode ser classificado como bom ou ruim; num questionário, uma pessoa pode responder se ingere bebidas alcoólicas regularmente ou não. No jogo de basquetebol, os possíveis resultados de um arremesso de lance livre, por exemplo, são acerto ou erro. Variáveis como estas, que assumem apenas dois valores, são denominadas variáveis de Bernoulli, ou variáveis binárias. O espaço amostral deste tipo de variável

pode ser restrita ao conjunto $\{S, F\}$, onde S representa sucesso (evento de interesse) e F representa fracasso (evento que não interessa).

Definição 1.12 *Uma variável X segue o modelo de Bernoulli quando $X(S) = 1$ ou $X(F) = 0$, à ocorrência de sucesso ou fracasso, respectivamente. Seja p a probabilidade de um sucesso, com $0 \leq p \leq 1$. Assim, sua função de probabilidade é dada por*

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

Em muitas aplicações tem-se o interesse em analisar uma sequência de resultados de uma variável de Bernoulli, ou uma sequência de ensaios de Bernoulli, que é definida através das seguintes condições:

- i)* Cada ensaio tem somente dois possíveis resultados: sucesso (S) ou falha / fracasso (F);
- ii)* Os ensaios são independentes;
- iii)* A probabilidade de sucesso (p) permanece a mesma em cada ensaio. Dessa forma, a probabilidade de fracasso é denotada por $(1 - p)$.

A média e a variância de uma variável aleatória com distribuição de Bernoulli são, respectivamente

$$E(X) = p \quad \text{e} \quad \text{Var}(X) = p(1 - p)$$

Neste trabalho as variáveis de Bernoulli desempenharão papel fundamental, pois fornecem um modelo natural para os resultados de arremessos num jogo de basquetebol. É importante observar que a modelagem de uma sequência de arremessos no basquetebol como um ensaio de Bernoulli implica em assumir que os ensaios são independentes, ou seja, que o resultado de um arremesso anterior não interfere no seguinte. Esta hipótese está diretamente relacionada à crença da boa fase (*hot hand*) e será discutida com mais detalhes no Capítulo 2.

1.3.2 Distribuição Binomial

Há situações em que o experimento consiste em n ensaios de Bernoulli. A repetição desses ensaios independentes dá origem a variável aleatória binomial.

Definição 1.13 *Seja a repetição de n ensaios de Bernoulli independentes e com probabilidade de sucesso p . A variável aleatória Binomial com parâmetro n e p , conta o número total de sucessos, e sua função de probabilidade é dada por*

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

O coeficiente binomial $\binom{n}{k}$ é calculado da seguinte forma:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Notação: $X \sim b(n, p)$ (lê-se: X tem distribuição binomial com parâmetros n e p).

A média e a variância de uma variável aleatória com distribuição Binomial são, respectivamente

$$E(X) = np \quad \text{e} \quad \text{Var}(X) = np(1-p)$$

Observação: *parâmetro é a medida usada para escrever uma característica numérica populacional. A média μ , e a variância σ^2 (ou $\text{Var}(X)$) são exemplos de parâmetros populacionais.*

1.3.3 Distribuição de Poisson

Definição 1.14 *Uma variável aleatória discreta X tem distribuição de Poisson com parâmetro $\lambda > 0$ se sua função de probabilidade for dada da seguinte forma:*

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \text{ para } K = 0, 1, 2, \dots$$

com $e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$, onde e representa a base dos logaritmos naturais.

O parâmetro λ é usado para designar a taxa de ocorrência por unidade de tempo ou unidade de suporte.

Notação: $X \sim P_o(\lambda)$ (lê-se: X tem distribuição de Poisson com parâmetro λ).

A média e a variância de uma variável aleatória com distribuição de Poisson são, respectivamente

$$E(X) = \lambda \quad \text{e} \quad \text{Var}(X) = \lambda$$

Exemplo 1.7 *Suponha que, ao longo de uma temporada, a média de arremessos convertidos, da linha de 3 pontos, pelo jogador J numa partida de basquete seja dada por $\lambda_J = 4,2$. Deseja-se determinar a probabilidade deste jogador fazer mais do que seis cestas de três pontos em uma partida.*

Solução: *Assumindo que o número de arremessos de três pontos tem uma distribuição de Poisson, com $\lambda_J = 4,2$, podemos calcular*

$$\begin{aligned} P(X > 6) &= 1 - P(X \leq 6) \\ &= 1 - \sum_{k=0}^6 \frac{e^{-\lambda} \lambda^k}{k!} \\ &= 1 - \sum_{k=0}^6 \frac{e^{-4,2} (4,2)^k}{k!} \\ &\cong 0,1325 \end{aligned}$$

Portanto, a probabilidade deste jogador fazer mais do que seis cestas de três pontos em uma partida é 0,1325.

1.4 Distribuição Normal: Um Modelo Contínuo

A atribuição de probabilidades para o caso contínuo é definida pela área abaixo de uma função $f(x)$ positiva denominada função de *densidade de probabilidade*. Assim, para uma variável aleatória contínua X , tal função satisfaz as seguintes condições:

- i) $f(x) \geq 0$, para todo $x \in (-\infty, \infty)$;
- ii) A área definida por $f(x)$ é igual a 1.

O item ii) pode ser caracterizado através de

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

Do mesmo modo, a probabilidade de uma variável assumir valores entre a e b é igual a área sob a curva da função densidade de probabilidade entre esses dois pontos, ou seja,

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

Dessa forma, tem-se $P(X = k) = 0$, para qualquer valor de k .

O valor esperado (ou média) e a variância são dados, respectivamente, por

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx \quad \text{e} \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

Ou ainda, uma expressão alternativa para a variância é

$$\sigma^2 = E(X^2) - \mu^2,$$

com $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$.

No caso de variáveis aleatórias contínuas, a *Distribuição Normal* é considerada a mais importante distribuição de probabilidade. Diversas populações numéricas possuem distribuições que podem ser ajustadas por uma curva normal. Existem situações nas quais a distribuição é discreta e mesmo assim a curva normal pode fornecer uma boa aproximação. As somas e as médias das variáveis terão uma distribuição aproximadamente normal sob condições adequadas ainda que as próprias variáveis individuais não sejam normalmente distribuídas.

Definição 1.15 Diz-se que uma variável aleatória contínua X possui uma distribuição normal com parâmetros μ e σ , $-\infty < \mu < \infty$ e $0 < \sigma^2 < \infty$, se a função densidade de probabilidade de X for dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-(x-\mu)^2/2\sigma^2}, \text{ para } -\infty < x < \infty \quad (1.3)$$

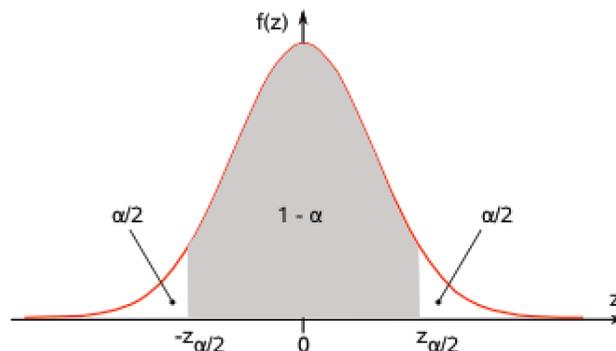


Figura 1.1: Gráfico da Distribuição Normal

Notação: $X \sim N(\mu, \sigma^2)$ (lê-se: X tem distribuição Normal com parâmetros μ (média) e $\sigma^2 = \text{Var}(X)$ (variância)).

Observando a figura, nota-se que a mesma é simétrica em relação a μ , disso decorre que $E(X) = \mu$.

Técnicas de integração analítica geralmente não são suficientes para calcular integrais envolvendo a função (1.3), por isso, as probabilidades envolvendo a distribuição normal são calculadas por métodos numéricos e tabelados para uso prático. Para facilitar estes cálculos, faz-se uma transformação que conduz ao cálculo de probabilidade com parâmetros $\mu = 0$ e $\sigma = 1$.

Considere $X \sim N(\mu, \sigma^2)$. Definindo $Z = \frac{X - \mu}{\sigma}$, através das propriedades do valor esperado e da variância, segue que

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}E(X - \mu) = \frac{1}{\sigma}[E(X) - \mu] = 0;$$

$$\text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2}\text{Var}(X - \mu) = \frac{1}{\sigma^2}\text{Var}(X) = 1.$$

Com essa transformação, é possível verificar que a variável aleatória Z tem distribuição $N(0, 1)$.

A distribuição normal com parâmetros $\mu = 0$ e $\sigma = 1$ é denominada *Distribuição Normal Padrão*, e portanto, Z é uma *variável aleatória normal padrão*. O cálculo da probabilidade de $X \in [a, b]$ é dado da seguinte forma:

$$\begin{aligned} P(a \leq X \leq b) &= P(a - \mu \leq X - \mu \leq b - \mu) \\ &= P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\ &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \end{aligned}$$

Assim, quaisquer que sejam μ e σ , utiliza-se a Normal Padrão para obter probabilidades com a Distribuição Normal. Os valores para $P(0 \leq Z \leq z)$, $z \geq 0$ podem ser encontrados em tabelas estatísticas que apresentam áreas sob a curva normal, ou seja, fornece a área abaixo do gráfico da função densidade de probabilidade normal padrão no intervalo $[0, z]$.

Apesar da maioria das variáveis de interesse deste trabalho serem discretas, a distribuição normal será bastante utilizada em alguns testes de hipóteses que faremos, como poderemos ver na seção 1.6.

1.5 Conceitos de Séries Temporais

Nesta seção apresentamos uma breve introdução de alguns conceitos de séries temporais que darão embasamento teórico ao estudo de *correlação serial* apresentado na seção 1.6.2. Existe uma literatura bastante vasta sobre séries temporais e que não reproduziremos aqui. Para uma referência mais completa o leitor pode consultar, por exemplo, [3, 11, 13].

No jogo de basquetebol, se considerarmos a ocorrência dos eventos no tempo, ao longo de um jogo ou de uma temporada, algumas técnicas de séries temporais podem ser aplicadas para modelar e analisar certas variáveis de interesse. Nos capítulos 3 e 4 faremos uso de alguns desses conceitos para investigar a correlação entre arremessos sucessivos de um mesmo jogador ao longo de uma temporada. Na sequência, introduzimos apenas alguns conceitos que serão utilizados.

1.5.1 Noções Básicas

Entende-se por *série temporal* um conjunto de observações feitas de forma sequencial ao longo do tempo. Em determinadas aplicações, a variável *tempo* pode ser substituída por *espaço*, *profundidade*, ou outra variável de interesse.

Duas características importantes em uma série temporal são a **correlação** entre seus termos e a **estacionariedade** da série.

Em muitas ocasiões tem-se o interesse de modelar a dependência entre novos termos da série e termos já observados. Métodos de regressão linear tem sido aplicados nestes casos. Há situações onde se deseja investigar se uma determinada série é *aleatória*, ou seja, se seus termos são *não correlacionados*. Para isto, recorre-se a testes de hipóteses, como os apresentados na seção a seguir.

Dizemos que uma série é *correlacionada* quando a observação de um determinado evento no tempo depende, em algum grau, das observações anteriores. Em séries aleatórias suas observações são não correlacionadas.

Em alguns casos tem-se o interesse em investigar se a média de uma determinada série varia indefinidamente com o tempo, ou estaciona em algum valor a longo prazo. Isto dá origem ao conceito de *estacionariedade*.

O seguinte exemplo ilustra esses conceitos e ajuda a compreender a maneira com que eles serão utilizados para analisar arremessos de basquetebol.

Exemplo 1.8 Considere S_1 e S_2 duas sequências de zeros e uns com 20 termos. A sequência S_1 é gerada iniciando em zero e alternando entre um e zero em cada novo termo:

$$S_1 = 0101010101 \dots 01$$

A sequência S_2 é gerada através de um experimento de Bernoulli com $p = 0.5$ (como por exemplo, um lançamento de moeda no qual atribui-se 0 caso o resultado seja cara e 1 caso o resultado seja coroa). Um exemplo deste tipo de série é o seguinte:

$$S_2 = 00100101110010001111$$

Em ambas as séries, o número de zeros e uns é o mesmo, 10 números de cada. Portanto, a média de S_1 e S_2 é igual a $\frac{1}{2}$, e este valor permanecerá praticamente o mesmo independente de quão grande for as séries. Dizemos assim que ambas as séries são estacionárias em relação à média. No entanto, na série S_1 , os valores são totalmente correlacionados, pois todo valor 0 é sucedido de um valor 1, em outras palavras, podemos prever todos os elementos da série. Já na série S_2 , os valores são aleatórios e não é possível a mesma previsão.

No contexto do basquetebol, podemos considerar uma série S contendo os resultados de todos os arremessos de 3 pontos de um determinado jogador ao longo de uma temporada. Uma pergunta central deste trabalho é: esta série é aleatória? A resposta deste tipo de questão envolve o uso de testes estatísticos.

1.6 Testes Estatísticos

Será de suma importância apresentar alguns *testes estatísticos* que embasarão os estudos realizados no terceiro e quarto capítulos. O leitor pode encontrar um conteúdo com mais detalhes em qualquer livro de Probabilidade e Estatística.

Um *teste estatístico* é uma regra de decisão usada para avaliar as hipóteses estatísticas. A *hipótese estatística* é uma suposição ou afirmação relativa a uma ou mais populações.

Um dos principais objetos de estudo da estatística moderna é a inferência estatística, a qual é dividida em dois grandes tópicos: a estimação de *parâmetros* e os *testes de hipóteses*.

De um modo geral, os *testes de hipóteses* têm por objetivo verificar se as informações sobre os parâmetros de uma população são verdadeiras. A formulação de um teste de hipótese consiste, basicamente, em:

- determinar a população e o parâmetro de interesse;

- considerar duas hipóteses: a *hipótese nula*, denotada por H_0 , representa a conjectura a ser testada e a *hipótese alternativa*, denotada por H_1 , representa o que se deseja comprovar;
- fixar o *valor de significância* α do teste (geralmente usa-se 1%, 5% e 10%) para que, assim, possam ser definidos os pontos de cortes (ponto crítico) das regiões de rejeição e aceitação das hipóteses;
- retirar amostra significativa da população a avaliar;
- calcular o valor do teste estatístico adequado;
- verificar se o valor do teste estatístico está na região de rejeição ou não para poder decidir estatisticamente se rejeita ou não a hipótese nula.

1.6.1 Teste de Aleatoriedade

Existem diferentes testes estatísticos que podem ser aplicados para testar a aleatoriedade de uma série. Nesta seção apresentamos um teste que emprega o número de *runs* de uma sequência. Este teste também foi utilizado para analisar arremessos de basquetebol em [5].

O Teste de Aleatoriedade, também conhecido como *teste de sequências*, é usado para testar a hipótese de que uma série de observações do tipo *sim* ou *não* (chamada de *runs*) é aleatória, ou seja, testa se as sequências ocorrem ou não ao acaso [4, 14].

Considerando u como sendo o número de *runs* de uma sequência, a aplicação do teste procede da seguinte forma:

Para $n_1 \geq 10$ e $n_2 \geq 10$, a distribuição de probabilidade de u é aproximadamente normal [4], com

$$\text{Média} = \mu(u) = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad (1.4)$$

e

$$\text{Desvio Padrão} = \sigma(u) = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}. \quad (1.5)$$

Assim, a expressão dada por

$$z = \frac{u - \mu(u)}{\sigma(u)} \quad (1.6)$$

testa a hipótese de aleatoriedade das sequências rejeitando-se H_0 se $|z| > z_{\alpha/2}$.

Exemplo 1.9 A sequência de observações $S_2 = 00100101110010001111$, do exemplo 1.8, contém 10 runs, pois cada “bloco” de 0’s e de 1’s é contado como um run.

Exemplo 1.10 Num jogo de perguntas e respostas, designamos “0” para acerto da pergunta e “X” para erro, obtendo a seguinte série de observações:

XOXXXOXO0XXOXO0X000XXOXX0000XXXOXX0000XOXOX

Neste caso, há a ocorrência de $n_1 = 22$ acertos e $n_2 = 21$ erros, e o número de runs observado é $u = 25$. Queremos testar a hipótese de que a série de observações acima é aleatória.

Se a hipótese alternativa H_A é que a disposição dos X’s e 0’s **não** é aleatória, rejeitamos a hipótese nula H_0 se $|z| > z_{\alpha/2}$.

Para $\alpha = 0,05$, obtém-se um valor crítico de 1,96, encontrado na Tabela da Distribuição Normal. Assim, rejeitar a hipótese nula se $z \leq -1,96$ ou $z \geq 1,96$.

Aplicando (1.4), (1.5) e (1.6) aos dados que temos, segue que

$$\begin{aligned}\mu(u) &= \frac{2 \cdot 22 \cdot 21}{22 + 21} + 1 \cong 22,49 \\ \sigma(u) &= \sqrt{\frac{2 \cdot 22 \cdot 21 (2 \cdot 22 \cdot 21 - 22 - 21)}{(22 + 21)^2 (22 + 21 - 1)}} \cong 3,24 \\ \therefore z &= \frac{u - \mu(u)}{\sigma(u)} = \frac{25 - 22,49}{3,24} \cong 0,77\end{aligned}$$

Como $z \cong 0,77$ está no intervalo de -1,96 a 1,96, a hipótese nula não pode ser rejeitada e temos que aceitar o fato de que, segundo este teste, não existem indícios contrários ao fato de que esta série seja aleatória.

Exemplo 1.11 Considere a sequência $S_1 = 0101010101 \dots 01$ do exemplo 1.8, com 0 representando acerto e 1 representando erro. Queremos testar a hipótese de que a série de observações S_1 é aleatória.

Aplicando o teste de aleatoriedade em S_1 segue que $n_1 = 10$, $n_2 = 10$ e $u = 20$. Assim,

$$\begin{aligned}\mu(u) &= \frac{2 \cdot 10 \cdot 10}{10 + 10} + 1 = 11 \\ \sigma(u) &= \sqrt{\frac{2 \cdot 10 \cdot 10 (2 \cdot 10 \cdot 10 - 10 - 10)}{(10 + 10)^2 (10 + 10 - 1)}} \cong 2,18\end{aligned}$$

$$\therefore z = \frac{u - \mu(u)}{\sigma(u)} = \frac{20 - 11}{2,18} \cong 4,13$$

Assumindo $\alpha = 0,05$, obtém-se um valor crítico de 1,96. Como $z > 1,96$, a hipótese nula deve ser rejeitada, o que é uma evidência estatística de que a série não é aleatória.

Outra forma de investigar a aleatoriedade de uma série consiste em analisar o grau de independência entre seus termos. Na próxima seção, mostramos como isto pode ser feito utilizando o conceito de correlação serial.

1.6.2 Correlação Serial

Em muitas situações se tem interesse em verificar se há relação entre variáveis. Por exemplo, será que o aumento do número de crimes está relacionado com o aumento da taxa de desemprego? Ou ainda, a temperatura influencia o rendimento dos alunos dentro da sala de aula?

A análise de *correlação* entre duas ou mais variáveis tende a mostrar se e como essas variáveis estão relacionadas, fornecendo um valor que sintetiza, ou quantifica, o grau de relacionamento entre as duas variáveis. Isso pode ser feito utilizando um teste de correlação linear. Em outros casos, interessa investigar se existe relação entre os resultados de uma variável aleatória e sua própria série histórica, ou seja, se o resultado corrente apresenta algum grau de dependência com o(s) resultado(s) anterior(es). Para responder este tipo de questão podemos utilizar uma análise chamada de *correlação serial*, conforme apresentamos a seguir.

Dadas n observações x_1, x_2, \dots, x_n é possível formar os pares $(x_1, x_2), \dots, (x_{n-1}, x_n)$. Ao considerar x_1, \dots, x_{n-1} e x_2, \dots, x_n como duas variáveis, o coeficiente de correlação entre elas é dado da seguinte forma:

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)(x_{t+1} - \bar{x}_2)}{\sqrt{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)^2 \sum_{t=1}^{n-1} (x_{t+1} - \bar{x}_2)^2}} \quad (1.7)$$

sendo as médias amostrais iguais a

$$\bar{x}_1 = \sum_{t=1}^{n-1} x_t / (n-1) \quad \text{e} \quad \bar{x}_2 = \sum_{t=2}^n x_t / (n-1)$$

O coeficiente r_1 mede as correlações entre observações sucessivas, o que explica o fato de ser chamado de *coeficiente de autocorrelação* ou *coeficiente de correlação serial (de ordem 1)*.

O cálculo de r_1 pode ser simplificado usando-se a média de todas as observações, ou seja, $\bar{x} = \sum_{t=1}^n x_t/n$, já que $\bar{x}_1 \approx \bar{x}_2$, e assumindo variância constante. Logo, a equação simplificada de (1.7) é dada por

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(n-1) \sum_{t=1}^n (x_t - \bar{x})^2 / n} \quad (1.8)$$

Em alguns livros, verifica-se que autores eliminam o termo $n/(n-1)$, que é próximo de 1 para n não muito pequeno, e generalizam a equação acima para calcular a correlação entre observações defasadas de k períodos de tempo, originando a seguinte equação:

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad \text{com } -1 < r_k < 1. \quad (1.9)$$

Existe também outra forma de se calcular os coeficientes de autocorrelação de ordem k . Geralmente, calcula-se primeiro os coeficientes de autocovariância c_k , definidos por analogia com a forma usual de covariância. Assim, c_k e r_k são dados, respectivamente, por

$$c_k = \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})/n, \quad k = 0, 1, 2, \dots \quad \text{e} \quad r_k = c_k/c_0$$

1.6.2.1 Correlograma

Quando se consegue resumir os dados de um determinado problema em forma de gráfico, muitas vezes, a interpretação dos resultados de cálculos passa a ser uma tarefa mais simples. Nesse sentido, o *correlograma* se torna uma ferramenta muito importante para melhor interpretar os k primeiros coeficientes de autocorrelação, e pode identificar características da série temporal em estudo.

Para uma série aleatória, os valores defasados são não correlacionados e, dessa forma, espera-se $r_k \approx 0$ para $k = 1, 2, \dots$. Agora, suponha que x_1, x_2, \dots, x_n sejam variáveis aleatórias independentes e identicamente distribuída com médias arbitrárias. Então, pode-se mostrar que o coeficiente de autocorrelação amostral r_k é assintoticamente normal distribuído, com média e variância dados, respectivamente, por

$$E(r_k) \approx -1/n \quad \text{e} \quad \text{Var}(r_k) \approx 1/n$$

(ver Kendall, Stuart, & Ord 1983, Capítulo 48). Portanto, limites de confiança aproximados de 95% são dados por $\pm 1.96/\sqrt{n}$, conforme especificado em [3].

Exemplo 1.12 *Calcular o coeficiente de autocorrelação de ordem $k = 1$ e verificar se existe correlação entre as observações sucessivas de cada variável abaixo.*

$$a) X_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Solução: *Dado que $x_1 = 1, x_2 = 2, x_3 = 3, \dots, x_{10} = 10$ e $n = 10$, é preciso calcular \bar{x} para depois fazer uso da equação (1.8). Assim,*

$$\begin{aligned} \bar{x} &= \sum_{t=1}^n x_t / n \\ &= \sum_{t=1}^{10} x_t / 10 \\ &= (1/10)(1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10) \\ &= 5,5 \end{aligned}$$

Portanto,

$$\begin{aligned} r_1 &= \frac{\sum_{t=1}^{10-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(10-1) \sum_{t=1}^{10} (x_t - \bar{x})^2 / 10} \\ &= (10/9) \left(\frac{(1-5,5)(2-5,5)}{(1-5,5)^2} + \frac{(2-5,5)(3-5,5)}{(2-5,5)^2} + \dots + \frac{(9-5,5)(10-5,5)}{(10-5,5)^2} \right) \\ &\cong 0,78 \end{aligned}$$

O valor encontrado para r_1 pode ser melhor interpretado através do correlograma abaixo:

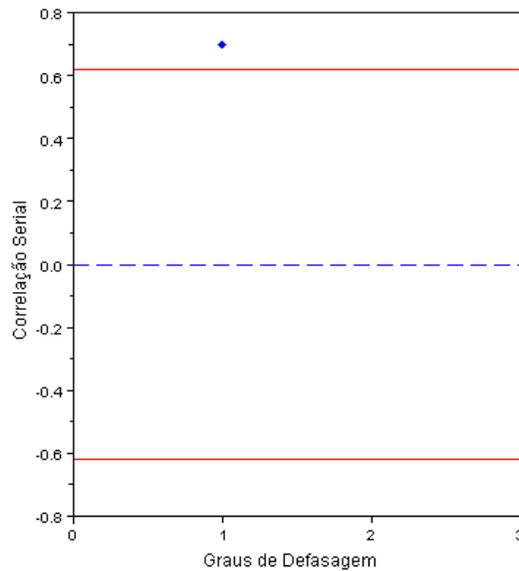


Figura 1.2: Correlograma com coeficiente de autocorrelação de ordem 1

Note que em $k = 1$ grau de defasagem (eixo horizontal), a correlação serial $r_1 \cong 0,78$ (eixo vertical) está fora dos limites de confiança dados, no caso, por $1.96/\sqrt{10} \cong 0.6198$ e $-1.96/\sqrt{10} \cong -0.6198$. Portanto, com nível de significância de 5%, pode-se afirmar que a série X_1 é correlacionada.

$$b) X_2 = \{1, 10, 2, 3, 8, 9, 7, 5, 6, 4\}$$

Solução: Neste caso, $x_1 = 1$, $x_2 = 10$, $x_3 = 2$, $x_4 = 3$, e assim por diante, com $n = 10$. Assim,

$$\begin{aligned} \bar{x} &= \sum_{t=1}^n x_t / n \\ &= \sum_{t=1}^{10} x_t / 10 \\ &= (1/10)(1 + 10 + 2 + 3 + 8 + 9 + 7 + 5 + 6 + 4) \\ &= 5,5 \end{aligned}$$

Logo,

$$\begin{aligned}
 r_1 &= \frac{\sum_{t=1}^{10-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(10-1) \sum_{t=1}^{10} (x_t - \bar{x})^2 / 10} \\
 &= (10/9) \left(\frac{(1-5,5)(10-5,5)}{(1-5,5)^2} + \frac{(10-5,5)(2-5,5)}{(10-5,5)^2} + \dots + \frac{(6-5,5)(4-5,5)}{(4-5,5)^2} \right) \\
 &\cong -0,29
 \end{aligned}$$

gera o seguinte correlograma

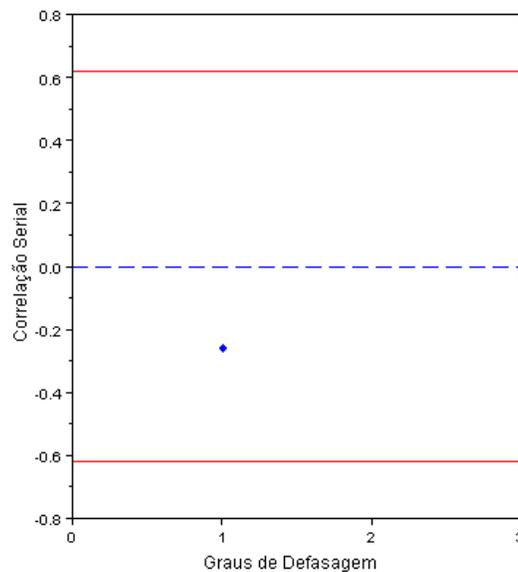


Figura 1.3: Correlograma com coeficiente de autocorrelação de ordem 1

Repare que o coeficiente $r_1 \cong -0,29$ está dentro dos limites de confiança dados por $1.96/\sqrt{10} \cong 0.6198$ e $-1.96/\sqrt{10} \cong -0.6198$. Portanto, com nível de significância de 5%, pode-se afirmar que a série X_2 é não correlacionada.

1.6.3 Diferença entre Duas Proporções

Existem testes de hipóteses que, com base em dados amostrais, permitem decidir se o verdadeiro valor de uma proporção, porcentagem ou probabilidade, difere (ou não) estatisticamente de um determinado valor alvo. Por exemplo, esses testes possibilitam verificar se a proporção de alunos do sétimo ano aprovados na disciplina de matemática difere estatisticamente do valor alvo de 0.79, ou se é verdade que 25% dos arremessos de dois pontos de uma grande equipe de basquete

ocorridos nos treinamentos não são convertidos. Geralmente, problemas como esses são resolvidos considerando o número observado de sucessos em n provas (ou a proporção observada de sucessos).

Em diversas situações, é preciso verificar se a diferença observada entre duas proporções amostrais pode ser atribuída ao acaso ou se indica que as proporções populacionais correspondentes não são iguais.

O teste da diferença entre duas proporções testa se uma diferença observada entre duas proporções amostrais é significativa ou se pode ser atribuída ao acaso.

Conforme [4], considere x_1 e x_2 números de sucessos obtidos em n_1 de um tipo e n_2 provas de outro tipo, independentes umas das outras, com p_1 e p_2 sendo, respectivamente, as probabilidades de sucesso correspondentes. Assim, a distribuição amostral de $\frac{x_1}{n_1} - \frac{x_2}{n_2}$ tem média $p_1 - p_2$ e desvio padrão (ou erro padrão entre duas proporções) dado por

$$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Ao testar a hipótese nula $p_1 = p_2 (= p)$ contra uma hipótese alternativa conveniente, temos $p_1 - p_2 = 0$ (média da distribuição da diferença entre duas proporções amostrais). Dessa forma, o erro padrão pode ser escrito como

$$\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

sendo p geralmente estimada combinando-se os dados e substituindo p pela proporção amostral combinada

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Assim, a estatística para o teste, em unidades padronizadas é dada por

$$z = \frac{\frac{x_1}{n_1} - \frac{x_2}{n_2}}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{com } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Para grandes amostras, essa estatística é um valor de uma variável aleatória que poderá ser aproximada pela distribuição normal padronizada. Rejeitar a hipótese nula H_0 se $|z| > z_{\alpha/2}$.

Exemplo 1.13 *Em um jogo de basquetebol, um determinado jogador possui probabilidade de acertar um arremesso de três pontos, dado que acertou os dois arremessos anteriores, igual a $15/37$. A probabilidade de acertar um arremesso, dado que errou os dois anteriores, é igual a $25/49$. Testar a hipótese nula de que a diferença entre essas probabilidades são iguais.*

Solução: *Seja $x_1 = 15$ o número de sucessos obtidos em $n_1 = 37$ provas e $x_2 = 25$ o número de sucessos obtidos em $n_2 = 49$ provas. Nestas condições, $\hat{p} = 0.4651$ e $z = -0.9647$.*

Com nível de significância de 5%, rejeitar a hipótese nula se a estatística do teste for um valor $|z| > z_{\alpha/2}$, ou seja, $z \leq -1,96$ ou $z \geq 1,96$. Como $z = -0.9647$ é um valor entre $-1,96$ e $1,96$, não podemos rejeitar a hipótese nula, e portanto, consideramos que a diferença entre as probabilidades é não significativa.

INVESTIGANDO A EXISTÊNCIA DE BOA FASE NO JOGO DE BASQUETEBOL: O POLÊMICO ARTIGO SOBRE “HOT HAND”

Neste capítulo apresentamos um resumo dos principais estudos realizados no artigo *The Hot Hand in Basketball: On the Misperception of Random Sequences* [5] – artigo polêmico que deu início ao desenvolvimento de vários outros trabalhos e que constitui-se na principal referência para esta dissertação.

O artigo mencionado busca investigar a origem e a validade das crenças comuns sobre “*hot hand*” e “*streak shooting*”, comumente traduzidos para o português como *mão quente* ou *boa fase* dos jogadores, através de pesquisas e experimentos realizados com torcedores, atletas e treinadores de basquete; mais precisamente, as equipes do Boston Celtics e Philadelphia 76ers do basquete profissional e alunos das Universidades de Cornell e Stanford estavam envolvidos nas pesquisas. O estudo desenvolvido em [5] está primariamente interessado em saber se os jogadores das equipes estudadas acertam uma maior porcentagem de seus arremessos depois de acertos ou depois de erros recentes e qual a percepção do mesmo fenômeno por parte dos expectadores.

As conclusões extraídas dos testes aplicados geraram muitas críticas [6, 15], pois segundo outros pesquisadores, os testes são insuficientes para concluir se um jogador está ou não passando por uma *boa fase*.

Na sequência apresentamos um resumo de cada um dos quatro principais estudos realizados no artigo. Nossa intenção com isto é fornecer subsídios básicos para o desenvolvimento do Capítulo 3 desta dissertação, que fará uma investigação similar utilizando dados da Liga Nacional de Basquete.

2.1 Estudo 1: pesquisa com torcedores de basquetebol

Para a realização do Estudo 1, foram convocados 100 alunos das Universidades de Cornell e Stanford, dos quais 50 eram capitães de times internos, para responder um questionário.

Todos os entrevistados tinham certa familiaridade com o basquete: 65% jogavam regularmente; todos assistiam pelo menos cinco jogos por ano e 75% assistiam mais de quinze jogos por ano.

O estudo tinha como objetivo apurar a opinião dos torcedores referente à dependência entre lances sucessivos num jogo de basquete. As respostas mostraram que 91% dos entrevistados acreditam que um jogador tem uma maior chance de acertar uma jogada depois de 2 ou 3 acertos do que depois de 2 ou 3 erros; 68% possuem opinião semelhante para lances livres; 96% acreditam que os jogadores tendem a fazer mais lances após uma série de acertos seguidos e 84% acreditam que é importante passar a bola para alguém que acabou de fazer vários lances seguidos com sucesso.

Com base nestes resultados, os autores do artigo concluem que os torcedores acreditam em boa fase (“*streak shooting*”).

2.2 Estudo 2: registros de objetivo de campo do basquete profissional

Este estudo envolveu os dados do basquete profissional. Os registros individuais de lançamentos dos jogadores do Philadelphia 76ers foram obtidos em 48 jogos (em casa) durante a temporada 1980 – 1981. As análises desses dados foram divididas da seguinte forma:

I) Análise de Probabilidade Condicional

A partir dos registros de lançamentos, pôde ser calculada a probabilidade de acerto condicionada a acertos e erros recentes dos jogadores, com o intuito de investigar se estes acertam uma maior porcentagem de seus lances depois de acertar ou depois de errar seu(s) último(s) lance(s).

A Tabela 2.1 mostra as probabilidades condicionais (calculadas como a frequências relativas condicional) dos nove principais jogadores do Philadelphia 76ers. A coluna 5 apresenta o percentual total de acerto de cada jogador que varia de 46% (para Hollins e Toney) a 62% (para Dawkins). As colunas 6, 7 e 8 apresentam a probabilidade de cada jogador acertar um lance dado que acertou seu último lance, últimos 2 lances e últimos 3 lances, respectivamente. Por outro lado, as colunas 2, 3 e 4 apresentam a porcentagem de acerto dos jogadores condicionada a terem errado seus últimos 3 lances, últimos 2 lances e último lance, respectivamente. A coluna 9 mostra a correlação serial entre os resultados de lances sucessivos.

Tabela 2.1: Probabilidade de acertar um lance condicionado ao resultado de lances anteriores para nove membros do Philadelphia 76ers

Jogador	P(A 3E)	P(A 2E)	P(A 1E)	P(A)	P(A 1A)	P(A 2A)	P(A 3A)	Correlação Serial r
Clint Richardson	0.50 (12)	0.47 (32)	0.56 (101)	0.50 (248)	0.49 (105)	0.50 (46)	0.48 (21)	- 0.020
Julius Erving	0.52 (90)	0.51 (191)	0.51 (408)	0.52 (884)	0.53 (428)	0.52 (211)	0.48 (97)	0.016
Lionel Hollins	0.50 (40)	0.49 (92)	0.46 (200)	0.46 (419)	0.46 (171)	0.46 (65)	0.32 (25)	- 0.004
Maurice Cheeks	0.77 (13)	0.60 (38)	0.60 (126)	0.56 (339)	0.55 (166)	0.54 (76)	0.59 (32)	- 0.038
Caldwell Jones	0.50 (20)	0.48 (48)	0.47 (117)	0.47 (272)	0.45 (108)	0.43 (37)	0.27 (11)	- 0.016
Andrew Toney	0.52 (33)	0.53 (90)	0.51 (216)	0.46 (451)	0.43 (190)	0.40 (77)	0.34 (29)	- 0.083
Bobby Jones	0.61 (23)	0.58 (66)	0.58 (179)	0.54 (433)	0.53 (207)	0.47 (96)	0.53 (36)	- 0.049
Steve Mix	0.70 (20)	0.56 (54)	0.52 (147)	0.52 (351)	0.51 (163)	0.48 (77)	0.36 (33)	- 0.015
Daryl Dawkins	0.88 (8)	0.73 (33)	0.71 (136)	0.62 (403)	0.57 (222)	0.58 (111)	0.51 (55)	- 0.142**
Média ponderada	0.56	0.53	0.54	0.52	0.51	0.50	0.46	- 0.039

Nota: O primeiro lançamento de cada jogo não pode ser condicionado. Os valores entre parênteses nas colunas 4 e 6 somados não resultam no valor entre parênteses na coluna 5. O número de lances em que cada probabilidade está baseada é dado entre parênteses.

Fonte: Extraída de [5]

Ao comparar as colunas 4 e 6, os autores observaram que para 8 dos 9 jogadores a probabilidade de um acerto é menor depois de um acerto que depois de um erro, o que contradiz a hipótese da mão quente. Comparações entre as colunas 3 e 7, e entre as colunas 2 e 8, fornecem resultados similares. Conseqüentemente, as correlações seriais são negativas para 8 dos 9 jogadores, verificadas na coluna 9.

II) Análises de Runs

Nesta etapa da pesquisa, foi feito o teste Run Wald-Wolfowitz nos registros de cada jogador. Para esse teste, cada seqüência de acertos ou erros é contada como um *run*. Por exemplo, uma seqüência de acertos (representados por 0) e erros (representados por X) dada por XX000X0 contém quatro *runs*. Quanto **maior** a quantidade de acertos (e erros) agrupados de um jogador, **menos** *runs* existem nos seus registros.

Na Tabela 2.2, a coluna 4 apresenta o número observado de *runs* através de todos os 48 jogos e a coluna 5 apresenta o número esperado de *runs* se os lances fossem independentes um do outro. Para 5 dos 9 jogadores, o número observado é **maior** que o número esperado de *runs*, contrariando a hipótese do “streak shooting”. A coluna 6 evidencia a diferença entre o observado e o esperado número de *runs*. Repare que o jogador Dawkins apresenta uma diferença significativa (segundo os autores do artigo) entre esses valores, pois o seu registro inclui mais *runs* que o esperado, sujeito a independência, novamente, contrariando a chamada boa fase.

Tabela 2.2: Teste Run - Philadelphia 76ers

Jogador	Acertos	Erros	N ^o _{obs} de Runs	N ^o _{esp} de Runs	Z
Clint Richardson	124	124	128	125.0	-0.38
Julius Erving	459	425	431	442.4	0.76
Lionel Hollins	194	225	203	209.4	0.62
Maurice Cheeks	189	150	172	168.3	-0.41
Caldwell Jones	129	143	134	136.6	0.32
Andrew Toney	208	243	245	225.1	-1.88
Bobby Jones	233	200	227	216.2	-1.04
Steve Mix	181	170	176	176.3	0.04
Daryl Dawkins	250	153	220	190.8	-3.09**
Média	218.6	203.7	215.1	210.0	-0.56

Fonte: Extraída de [5]

III) Teste de Estacionariedade

Os registros de arremessos dos jogadores foram divididos em conjuntos disjuntos de quatro lances consecutivos. Conjuntos contendo três ou quatro acertos significam que o jogador teve um alto desempenho; dois acertos, médio desempenho e zero ou um acerto, baixo desempenho.

Um jogador é considerado “quente” se o mesmo possuir mais conjuntos de alto desempenho que o esperado pela probabilidade. O número de grupos de alto, médio e baixo desempenho de cada um dos nove jogadores foi comparado aos valores esperados pela probabilidade, assumindo uma taxa constante de acerto (obtida da coluna 5 da Tabela 2.1) e lances independentes.

Contudo, segundo os autores de [5], o Teste de Estacionariedade também não forneceu evidências de que exista boa fase.

As pessoas tendem a acreditar que os jogadores de basquete acertam uma maior porcentagem de seus arremessos depois de terem acertado seu(s) último(s) lance(s) que depois de terem errado seu(s) últimos(s) lance(s). A crença na boa fase é evidenciada também na opinião dos jogadores e do treinador do Philadelphia 76ers.

Numa seção de treinos da equipe do Philadelphia 76ers, sete jogadores foram entrevistados e o treinador foi quem perguntou e também respondeu questões similares as que foram feitas no Estudo 1. A maioria dos jogadores (seis entre os oito envolvidos na pesquisa) revelou que depois de ter acertado alguns arremessos seguidos, sente que acertará o próximo. Cinco jogadores acreditam ter uma maior chance de acertar uma jogada depois de ter a pouco acertado seus últimos 2 ou 3 lances que depois de ter perdido seus últimos 2 ou 3 lances (dois jogadores não apoiaram essa afirmação e um não respondeu a questão). Sete dos oito participantes informaram que depois de terem feito uma série de lances seguidos, tendem a fazer mais lances do que normalmente fariam. Todos eles acreditam que, em um jogo, é importante passar a bola para alguém que tenha feito vários lances seguidos. Alguns jogadores fazem estimativas numéricas calculando suas porcentagens de acerto depois de um acerto (média: 62%) e depois de um erro (média: 49,5%).

Segundo os autores de [5], a pesquisa evidencia que os jogadores entrevistados acreditavam na chamada boa fase.

2.3 Estudo 3: registros de lance livre do basquete profissional

O terceiro estudo aborda os registros de lançamento livre dos jogadores do Boston Celtics, do basquete profissional. Este estudo tinha como objetivo analisar a dependência entre lances sucessivos que, neste caso, seriam isentos da influência de nocivos fatores externos tais como seleção de lances e pressão da defesa adversária.

Para existir uma correlação positiva entre lances sucessivos, espera-se que os jogadores acertem uma maior porcentagem de seu segundo lance livre depois de terem acertado que depois de terem errado o primeiro lance (ver Tabela 2.3).

Tabela 2.3: Probabilidade de acertar o segundo lance livre condicionado ao resultado do primeiro lance livre para nove membros do Boston Celtics durante as temporadas 1980 - 1981 e 1981-1982

Jogador	$P(H_2 M_1)$	$P(H_2 H_1)$	Correlação Serial r
Larry Bird	0.91 (53)	0.88 (285)	- 0.032
Cedric Maxwell	0.76 (128)	0.81 (302)	0.061
Robert Parish	0.72 (105)	0.77 (213)	0.056
Nate Archibald	0.82 (76)	0.83 (245)	0.014
Chris Ford	0.77 (22)	0.71 (51)	- 0.069
Kevin McHale	0.59 (49)	0.73 (128)	0.130
M. L. Carr	0.81 (26)	0.68 (57)	- 0.128
Rick Robey	0.61 (80)	0.59 (91)	- 0.019
Gerald Henderson	0.78 (37)	0.76 (101)	- 0.022

Nota: O número de lances em que cada probabilidade está baseada é dado entre parênteses

Fonte: Extraída de [5]

A coluna 2 indica a probabilidade de um acerto no segundo lance livre dado que houve um erro no primeiro lance e a coluna 3 indica a probabilidade de um acerto no segundo lance livre dado que ocorreu um acerto no primeiro lance. A coluna 4 apresenta a correlação entre o primeiro e o segundo lance livre. As correlações são positivas para 4 jogadores e negativas para os outros 5, porém nenhuma é significativamente diferente de zero.

Com base na Tabela 2.3, os pesquisadores do artigo verificaram que não há evidências que o resultado do segundo lance livre seja influenciado pelo resultado do primeiro.

2.4 Estudo 4: experimento controlado de lançamentos

Com a pretensão de eliminar os efeitos que a pressão defensiva e a seleção de lances podiam causar sobre os jogadores, foi realizado um experimento controlado considerado como um método alternativo. Além disso, o experimento pôde investigar a habilidade dos jogadores em prever seus desempenhos. O grupo que participou deste teste era composto por 14 homens e 12 mulheres da equipe principal de basquete dos times do Colégio de Cornell.

Para cada jogador foi determinada uma distância até a cesta, onde sua porcentagem de lançamento era de aproximadamente 50%. A essa distância, desenhou-se dois arcos no chão, sobre os quais cada pessoa deveria fazer 100 arremessos (50 de cada arco) e, entre lances, se mover ao longo do arco de forma que lances consecutivos não fossem feitos do mesmo lugar.

Os jogadores foram pagos por suas participações e a quantia recebida era baseada na precisão com que faziam seus lances e como eles previam acertos e erros.

Com os dados coletados nesse experimento pôde ser feita algumas análises, como as que seguem:

I) Análise de Probabilidade Condicional

Novamente, buscou-se saber se os jogadores acertam uma maior porcentagem de seus lances depois de terem acertado seu último lance (ou últimos lances) que depois de terem errado seu último lance (ou últimos lances).

Através das probabilidades condicionais, perceberam que para a maioria dos jogadores (14 de um total de 26) a probabilidade de acerto dado que ocorreu um erro ($P(A | 1E)$) é maior que a probabilidade de acerto dado que ocorreu um acerto ($P(A | 1A)$), o que contraria a expectativa da mão quente. Conseqüentemente, suas correlações seriais são negativas.

Comparações feitas entre $P(A | 2E)$ e $P(A | 2A)$, e entre $P(A | 3E)$ e $P(A | 3A)$ fornecem resultados semelhantes (maiores detalhes, ver [5], Tabela 4, página 307).

II) Análises de Runs

Com base nos registros de lançamentos realizados no experimento controlado, descrito anteriormente, pôde ser feito o teste *run*. Notaram que 14 dos 26 jogadores apresentaram um pouco mais de *runs* que o esperado, 12 apresentam um pouco menos que o esperado (o que contraria a expectativa da chamada boa fase) e que somente uma pessoa do grupo masculino conteve significativamente menos *runs* que o esperado (ver [5], Tabela 5, página 308).

III) Teste de Estacionariedade

Os registros dos 100 lances feitos por cada jogador foram divididos em conjuntos disjuntos de 4 lances consecutivos, semelhante ao que foi feito no Estudo 2. A partir dos conjuntos, contou-se o número de conjuntos com alto, médio e baixo desempenho. Caso o número de conjuntos de alto desempenho exceder o número esperado pela probabilidade, o jogador é considerado “quente”. Para melhor comparação entre o número observado e o número esperado de alto, médio e baixo desempenho, aplicou-se um teste Qui-Quadrado. A análise foi repetida por quatro vezes, cada qual iniciando do primeiro, segundo, terceiro e quarto lance, respectivamente.

Os resultados proveram evidências de estacionariedade apenas para um dos 26 jogadores.

IV) Teste de Previsibilidade

Acredita-se, muitas vezes, que “estar quente” envolve previsibilidade no lugar de dependência sequencial. Se um jogador pode prever um acerto antes de fazer um arremesso, o mesmo pode ter um senso justificável de ser “quente”.

O Teste de Previsibilidade foi realizado com o objetivo de, como o próprio nome sugere, analisar a habilidade dos jogadores em prever seus desempenhos. Para tanto, antes de realizar um lance, cada jogador fazia sua aposta e previsão em relação ao resultado do lançamento. Os jogadores foram aconselhados a apostar alto quando sentissem confiança em suas habilidades de lançamento e apostar baixo quando não.

Para uma segunda etapa do teste, os participantes ficavam em duplas alternando entre os papéis de arremessador e observador de modo que o observador apostava alto ou baixo no resultado do lance a ser apresentado pelo seu parceiro de dupla.

Uma idéia intuitiva faz pensar que se os jogadores podem prever seus acertos e erros, suas apostas deveriam correlatar com seus desempenhos. No entanto, os pesquisadores de [5] constataram através do teste de previsibilidade que os jogadores eram malsucedidos em suas previsões.

2.5 Nova Análise

Embora as análises apresentadas até agora não evidenciassem dependência entre lances sucessivos, as pessoas que participaram das pesquisas mostraram-se crentes na existência da boa fase.

Isso talvez se deva, por exemplo, a uma confusão que se faz ao pensar em seqüências longas e alternadas de acertos e erros. Se seqüências longas de acertos ou erros são mais notáveis, e é possível que o observador superestime a correlação entre lances sucessivos.

Para realizar uma nova investigação, 100 torcedores de basquete foram solicitados para classificar seis diferentes seqüências como:

- chance shooting
- streak shooting
- alternative shooting

sendo que cada uma delas continha 11 acertos, indicados por X's, e 10 erros, indicados por 0's.

Seqüências de acertos e erros iguais as de caras e coroas resultantes dos arremessos de moedas foram definidas de *chance shooting*. As seqüências com agrupamentos de acertos e erros maiores que os de caras e coroas foram denominadas *streak shooting*. Por fim, seqüências com agrupamentos menores do que os obtidos num lançamento de moedas foram definidas como *alternative shooting*.

As seis seqüências se diferem umas das outras no número de *runs* (9, 11, 13, 15, 17 e 19) e, conseqüentemente, na probabilidade de alternância (P_{AI}) obtida dividindo o número de *runs* pelo número total de acertos e erros, que neste caso é 21.

A probabilidade de alternância gerada ao lançar uma moeda é 0,5 – o resultado de um lançamento é independente do resultado anterior – caracterizando a chamada *chance shooting*. Já $P_{AI} < 0,5$ e $P_{AI} > 0,5$ caracterizam *streak shooting* e *alternative shooting*, nesta ordem. Por exemplo, a seqüência dada por

$$X0X0X000XX0X0X00XXX0X$$

consiste em 15 *runs* com $P_{AI} \cong 0,7$. Quanto **maior** o agrupamento de acertos e erros, **menor** será o número de *runs* ocasionando, eventualmente, probabilidade de alternância **menor** que 0,5. No caso em que essa probabilidade for **maior** que 0,5, verifica-se a presença de **mais** *runs* devido ao **menor** agrupamento de acertos e erros.

Na pesquisa, somente 32% das pessoas classificaram corretamente como *chance shooting* a seqüência com $P_{AI} = 0,5$ e 62% a classificaram, de forma errada, como *streak shooting*. As seqüências com $P_{AI} = 0,7$ e $P_{AI} = 0,8$ foram identificadas como o melhor exemplo de *chance shooting*, mas que deveriam ser classificadas como *alternative shooting*.

A tendência de perceber uma sequência como *streak shooting* decresce conforme a probabilidade de alternância aumenta. As pessoas tendem a reconhecer *chance shooting* como *streak shooting* e elas esperam sequências exemplificando *chance shooting* que contenha mais alternâncias que realmente seria produzido por um processo aleatório (*chance shooting*). Assim, pessoas veem uma correlação serial positiva em sequências independentes e elas não conseguem detectar uma correlação serial negativa em sequências alternadas. Por esta razão, pessoas consideram sequências aleatórias como correlacionadas positivamente e também consideram sequências relacionadas negativamente como aleatórias.

Existe uma idéia errada das leis de probabilidade associada com a crença de que as pequenas, assim como as grandes sequências, são representativas em seu processo de geração. Isso induz uma expectativa de que sequências aleatórias deveriam ser muito mais equilibradas do que elas são, e a percepção errada de uma correlação positiva entre lances sucessivos. Se sequências aleatórias são percebidas como *streak shooting*, então nenhuma quantidade de exposição para tais sequências convencerá o jogador, treinador ou torcedor que as sequências são de fato aleatórias.

A independência entre lances sucessivos, claro que, não significa que basquete seja um jogo de probabilidade em lugar de habilidade, nem menos emocionante de jogar, assistir ou analisar. Meramente indica que a probabilidade de um acerto é em grande medida independente de resultados de lances anteriores, embora certamente dependa de outros fatores tais como habilidade, pressão do adversário, distância à cesta; essa situação é análoga a jogar uma moeda utilizando um mecanismo controlado, de forma que os resultados de jogadas sucessivas sejam independentes, mas probabilidade de caras depende de fatores mensuráveis tais como a posição inicial da moeda, e seu momento angular e vertical (ver [5], páginas 312 e 313).

ANÁLISE ESTATÍSTICA DE ARREMESSOS DE BASQUETEBOL

Neste capítulo apresentamos os resultados de alguns testes estatísticos realizados com base nos registros de lançamentos da equipe de basquete profissional de Bauru, ocorridos na temporada 2009 – 2010. Alguns destes testes são baseados em estudos realizados em [5] e outros foram trazidos de outros contextos para o objeto de estudo desta dissertação.

As informações relativas ao desempenho dos jogadores foram adquiridas através de tabulação feita em colaboração com o grupo do Laboratório de Biomecânica e Instrumentação (LABIN) da FCA - UNICAMP, com base em vídeos dos jogos fornecidos pela Liga Nacional de Basquetebol (LNB). Dessa forma, puderam ser coletados todos os dados de interesse.

Objetivamos analisar estatisticamente o comportamento das sequências de arremessos consecutivos de cada integrante da equipe estudada diante dos testes estatísticos aplicados, para que assim possamos, no capítulo subsequente, verificar se existem semelhanças entre as sequências advindas dos jogos de basquetebol e sequências aleatórias geradas computacionalmente.

Para a análise dos dados foi fundamental desenvolvermos algoritmos computacionais. Todos os algoritmos utilizados nesta dissertação (desde a separação e manipulação dos registros de arremessos de cada jogador, aplicação de métodos estatísticos até simulação de dados) foram implementados no software livre Scilab e os principais estão disponíveis no apêndice deste trabalho.

3.1 Preparação e processamento dos dados

A princípio, fez-se necessário separar os dados de cada membro do time considerado, pois possuíamos uma planilha com mais de 2.600 registros representando o desempenho dos jogadores no decorrer da temporada. Para as nossas análises, nos interessam lances livres e cestas de três pontos dos dez integrantes que obtiveram maior número de arremessos em toda temporada.

Ao fazer uso do software Scilab, a sequência de lances juntamente com seus respectivos arremessadores pode ser considerada como uma matriz de n linhas e duas colunas, onde a primeira coluna representa as ações de cada pessoa no jogo e a segunda, o respectivo autor do arremesso, representado pelo número da camisa.

Aplicando os algoritmos que desenvolvemos, pudemos isolar os registros de cada atleta. Cada sequência de lances consecutivos pode ser interpretada como um vetor \vec{v}_i de n entradas, onde n é o número de lançamentos efetuados pelo jogador i . A tabela subsequente é parte do que tínhamos acesso:

Tabela 3.1: Parte da planilha de arremessos dos jogadores da equipe de Bauru

Ação	Jogador
1	10
1	10
-2	8
-3	14
2	15
-3	33
-2	8
-3	10
2	10
2	8
1	8
-2	14
3	8
-2	8
2	7
2	8
-3	10
-2	19

Na coluna *Ação*, 1, 2 e 3 representam acerto em arremessos de lance livre, dois e três pontos, respectivamente; os valores negativos representam erros nesses arremessos. A coluna *Jogador*

corresponde ao número da camisa do jogador que promoveu a ação. Logo, podemos considerar a seguinte matriz (transposta):

$$A' = \begin{pmatrix} 1 & 1 & -2 & -3 & 2 & -3 & -2 & -3 & 2 & 2 & 1 & -2 & 3 & -2 & 2 & 2 & -3 & -2 \\ 10 & 10 & 8 & 14 & 15 & 33 & 8 & 10 & 10 & 8 & 8 & 14 & 8 & 8 & 7 & 8 & 10 & 19 \end{pmatrix}$$

Portanto, ao separar os dados de cada um dos jogadores na matriz A (com o auxílio de algoritmos), obtemos vetores da seguinte forma:

$$v_7 = (2)$$

$$v_8 = (-2 -2 2 1 3 -2 2)$$

$$v_{10} = (1 1 -3 2 -3)$$

$$v_{14} = (-3 -2)$$

$$v_{15} = (2)$$

$$v_{19} = (-2)$$

$$v_{33} = (-3)$$

onde v_7 mostra a sequência de lances consecutivos do jogador 7, v_8 mostra a sequência de lances consecutivos do jogador 8, e assim por diante.

Caso se queira obter somente as sequências de lances de dois pontos, por exemplo, é possível fazê-lo. Novamente, aplicando sobre os vetores um determinado algoritmo para esse fim, os mesmos passam a ter as seguintes sequências:

$$v_7 = (2)$$

$$v_8 = (-2 -2 2 -2 2)$$

$$v_{10} = (2)$$

$$v_{14} = (-2)$$

$$v_{15} = (2)$$

$$v_{19} = (-2)$$

$$v_{33} = []$$

sendo v_{33} um vetor vazio.

Dessa forma, é possível aplicar testes estatísticos sobre as sequências desejadas, dadas pelos vetores.

3.2 Análise 1: Lançamentos de 3 Pontos

Inicialmente, faremos um estudo acerca das sequências de lançamentos de 3 pontos dos dez integrantes do Bauru que tentaram um maior número de arremessos, como outrora mencionado. Com os registros de arremessos devidamente separados e organizados por jogador, numa primeira análise obtemos uma tabela com o rendimento dos mesmos.

A Tabela 3.2 exibe o desempenho de cada jogador, identificado pelo número da camisa na coluna 1. O número total de tentativas em cestas de 3 pontos é apresentado na coluna 2. A terceira e quarta colunas da tabela apresentam o número de acertos e de erros, respectivamente. As colunas 5 e 6 apresentam a frequência relativa de acertos e erros, usualmente interpretadas como a probabilidade de acerto e de erro. Por fim, a última coluna mostra a probabilidade do jogador i efetuar um arremesso de 3 pontos em seu time durante a temporada. Para esta amostra de jogadores, a probabilidade de erro, $P(E)$, é maior que a probabilidade de acerto, $P(A)$, já que todos erraram mais que acertaram.

Como um exemplo prático, $P(J = 14) = 0,269$ significa que o *jogador 14* efetuou 26,9% dos arremessos de 3 pontos da sua equipe na temporada analisada. Os valores registrados na Tabela 3.2 auxiliarão em várias análises que faremos.

Tabela 3.2: Desempenho dos jogadores em lançamentos de 3 pontos

Jogador	Tentativas	Acertos	Erros	P(A)	P(E)	P(J=i)
14	203	95	108	0.468	0.532	0.269
8	136	61	75	0.448	0.551	0.180
33	85	29	56	0.341	0.658	0.112
26	78	30	48	0.384	0.615	0.103
15	71	26	45	0.366	0.633	0.094
13	70	28	42	0.40	0.60	0.093
10	51	16	35	0.313	0.686	0.067
12	35	15	20	0.428	0.571	0.046
16	19	8	11	0.421	0.578	0.025
11	5	1	4	0.20	0.80	0.006
Total	753	309	444			1.0

Podemos também avaliar as sequências de arremessos com relação ao estudo da probabilidade condicional (calculada a partir do algoritmo que desenvolvemos e apresentamos no Anexo C).

O cálculo da probabilidade condicional sobre os registros de arremessos de três pontos é de grande utilidade para termos uma idéia do que acontece com as probabilidades de acerto depois de um, dois e três erros (indicadas na Tabela 3.3 pelas colunas 4, 3 e 2, respectivamente) e de acerto depois de um, dois e três acertos (colunas 6, 7 e 8, respectivamente). Assim, conseguimos notar se a probabilidade de um acerto em um lançamento é **maior** depois de acerto(s) ou depois de erro(s) recentes.

Tabela 3.3: Probabilidade Condicional

Jogador	$P(A 3E)$	$P(A 2E)$	$P(A 1E)$	$P(A)$	$P(A 1A)$	$P(A 2A)$	$P(A 3A)$
14	0.416	0.510	0.542	0.468	0.389	0.405	0.40
8	0.370	0.372	0.405	0.448	0.491	0.466	0.357
33	0.304	0.394	0.321	0.341	0.392	0.30	0.333
26	0.363	0.312	0.297	0.384	0.50	0.666	0.70
15	0.291	0.272	0.266	0.366	0.560	0.538	0.50
13	0.250	0.291	0.390	0.40	0.392	0.363	0.50
10	0.40	0.347	0.342	0.313	0.266	0.250	0
12	0.750	0.555	0.473	0.428	0.333	0.20	0
16	0.333	0.40	0.40	0.421	0.375	0.666	0.50
11	—	0	0.333	0.20	0	—	—

Repare que cinco dos dez jogadores listados na Tabela 3.3 possuem probabilidade de acerto após um acerto **menor** que após um erro ($P(A | 1A) < P(A | 1E)$), o que contraria a expectativa da *mão quente*. Comparando as colunas 3 e 7, e as colunas 2 e 8, evidenciam resultados semelhantes pois, basicamente 4 dentre os 10 jogadores apresentam $P(A | 2E) > P(A | 2A)$ e $P(A | 3E) > P(A | 3A)$. No entanto, buscamos identificar o quão significativas são essas diferenças entre as probabilidades. Faremos isto através do *teste da diferença entre duas proporções*.

3.2.1 Teste da Diferença entre Duas Proporções

Em lançamentos de três pontos, calculamos para cada jogador as probabilidades de acerto após um, dois e três acertos, bem como as probabilidades de acerto após um, dois e três erros. Como é usual, consideramos que esta probabilidade é aproximada pela frequência relativa com que cada evento ocorreu. De posse destes dados, aplicamos um teste de diferença entre proporções (Capítulo 1, Seção 1.6.3).

Para a aplicação do *teste da diferença entre duas proporções* faremos uso da Tabela 3.3, com base no conteúdo apresentado na Seção 1.6.3. Nos interessa saber se a diferença entre as probabilidades condicionais, de cada jogador, analisadas duas a duas são ou não significativas.

Com nível de significância de 5%, rejeitamos a hipótese nula de que as proporções são iguais se a estatística do teste resultar em um valor $|z| > z_{\alpha/2}$, ou seja, $z \leq -1,96$ ou $z \geq 1,96$; isto quer dizer que a diferença entre as probabilidades condicionais analisadas no *teste de diferenças entre duas proporções* são significativas (S). Caso o resultado do teste seja um valor $-1,96 < z < 1,96$, consideramos que as diferenças são não significativas (N).

Tabela 3.4: Diferença entre duas proporções para o jogador 14

J=14	P(A 3E)	P(A 2E)	P(A 1E)	P(A)	P(A 1A)	P(A 2A)	P(A 3A)
P(A 3E)	0	N	N	N	N	N	N
P(A 2E)	N	0	N	N	N	N	N
P(A 1E)	N	N	0	N	S	N	N
P(A)	N	N	N	0	N	N	N
P(A 1A)	N	N	S	N	0	N	N
P(A 2A)	N	N	N	N	N	0	N
P(A 3A)	N	N	N	N	N	N	0

S = diferença significativa

N = diferença não significativa

A Tabela 3.4 exemplifica o resultado do teste para o jogador 14. Verifica-se que a diferença entre $P(A | 1A)$ e $P(A | 1E)$ resultou num valor significativo. Em valores absolutos, a tabela é simétrica em relação a diagonal, e assim o resultado do teste entre $P(A | 1E)$ e $P(A | 1A)$ é um valor significativo.

Através de nossas análises, verificamos que as diferenças entre proporções são não significativas para os jogadores objetos de estudo, a exemplo do que ocorre para o jogador 14 analisado na Tabela 3.4, salvo quatro atletas que não possuíam dados suficientes para tal.

A interpretação desses resultados nos revela que não existem diferenças estatísticas entre as probabilidades de acerto, dado que ocorreram acertos, e as probabilidades de acerto, dado que ocorreram erros, fato que contraria a expectativa da boa fase.

3.2.2 Correlação Serial de Ordem 1

Nesta seção vamos utilizar os conceitos introduzidos nas Seções 1.6.2 e 1.6.2.1 para investigar se existe alguma correlação entre os arremessos sucessivos de 3 pontos de um jogador .

Na Tabela 3.5 é apresentada de forma sintética a análise dos coeficientes de correlação serial de ordem 1 das seqüências de arremessos de 3 pontos de cada jogador ao longo da temporada (ver algoritmo no Anexo B). A última coluna da tabela exhibe o resultado do teste de hipótese realizado com nível de significância de 5%, considerando como hipótese nula H_0 que a seqüência não é autocorrelacionada e como hipótese alternativa H_1 que a seqüência é autocorrelacionada. Se o valor calculado de r_1 estiver fora dos limites de aceitação de H_0 concluímos que o teste é significativo (S), e rejeitamos H_0 . Caso o valor calculado de r_1 esteja dentro dos limites de aceitação de H_0 , concluímos que o teste é não significativo (N) e aceitamos H_0 .

Dos dez jogadores listados na Tabela 3.5, oito possuem o coeficiente r_1 dentro dos limites de confiança dados por $\pm 1.96/\sqrt{n}$, onde n é o número de tentativas à cesta, conforme visto em 1.6.2.1. Portanto, com nível de significância de 5%, podemos afirmar que as seqüências de lançamentos de 3 pontos desses oito jogadores são *não correlacionadas*.

Tabela 3.5: Teste de hipóteses sobre o coeficiente de correlação serial de ordem 1 das seqüências de lançamentos de 3 pontos para cada jogador ao longo da temporada.

Jogador	Valor de n	Limites de Confiança	Correlação Serial r_1	Localização de r_1	Resultado do Teste
14	203	± 0.137	- 0.152	fora dos limites	S
8	136	± 0.168	0.086	dentro dos limites	N
33	85	± 0.212	0.070	dentro dos limites	N
26	78	± 0.221	0.202	dentro dos limites	N
15	71	± 0.232	0.289	fora dos limites	S
13	70	± 0.234	0.002	dentro dos limites	N
10	51	± 0.274	- 0.074	dentro dos limites	N
12	35	± 0.331	- 0.142	dentro dos limites	N
16	19	± 0.449	- 0.028	dentro dos limites	N
11	5	± 0.876	- 0.375	dentro dos limites	N

S = teste significativo

N = teste não significativo

3.2.3 Análise de Runs e o Teste de Aleatoriedade

Nesta seção, apresentamos uma análise acerca do número de *runs* juntamente com o resultado da aplicação do *teste de aleatoriedade*, realizados sobre os registros de lançamentos de 3 pontos estudados até o momento.

Para a análise de *runs*, foi desenvolvido um algoritmo que computa o número de runs existente na sequência de lançamentos de 3 pontos de cada jogador (Anexo A). Os cálculos para a obtenção do número de *runs esperados* e a aplicação do *teste de aleatoriedade* estão descritos com detalhes na Seção 1.6.1.

A Tabela 3.6 mostra os resultados do *teste de aleatoriedade* para cada jogador. A coluna 4 exhibe o número de *runs observado* nos registros de arremessos e a coluna 5 exhibe o número de *runs esperado* caso os resultados de todas as tentativas fossem independentes uns dos outros. A coluna 6 evidencia a diferença entre o *observado* e o *esperado* número de *runs*. Observe que os valores da estatística Z são positivos se o número de *runs observado* for maior que o número de *runs esperado*.

Tabela 3.6: Análise de Runs e o Teste de Aleatoriedade sobre os registros de lançamentos de 3 pontos

Jogador	Acertos	Erros	N ^o <i>obs</i> de Runs	N ^o <i>esp</i> de Runs	Z
14	95	108	117	102.08	2.107
8	61	75	62	68.27	- 1.092
33	29	56	36	39.21	- 0.780
26	30	48	30	37.92	- 1.908
15	26	45	24	33.95	- 2.567
13	28	42	34	34.60	- 0.150
10	16	35	24	22.96	0.342
12	15	20	20	18.14	0.650
16	8	11	10	10.26	- 0.127
11	1	4	—	—	0.816
Média	30.9	44.4	36	36.74	

Para o *teste de aleatoriedade*, consideramos como hipótese nula H_0 que as sequências de lançamentos de 3 pontos são aleatórias. Logo, a hipótese alternativa H_A é que tais sequências não são

aleatórias. Dessa forma, com $\alpha = 0.05$, rejeitamos a hipótese nula se $|z| > z_{\alpha/2}$, isto é, se $z \leq -1.96$ ou $z \geq 1.96$.

Analisando os oito primeiros jogadores da Tabela 3.6, verificamos que seis deles possuem o valor da estatística Z dentro do intervalo de -1.96 a 1.96 e assim, a hipótese nula não pode ser rejeitada. Isso quer dizer que não há evidências reais de que as sequências de lançamentos de 3 pontos desses jogadores não sejam aleatórias. Para os jogadores 14 e 15, a estatística Z está fora do intervalo citado, e portanto, a hipótese nula deve ser rejeitada.

No caso do *jogador 16*, o teste procede de uma outra forma, pois o número de acertos n_1 é menor que 10. Como $n_1 = 8$, $n_2 = 11$ e $u = 10$, rejeitamos a hipótese nula de aleatoriedade se $u \leq u'_{\alpha/2}$ ou $u \geq u_{\alpha/2}$, onde esses valores são encontrados nas Tabelas I e II do Anexo G para valores de n_1 e n_2 até 15 e $\alpha = 0,05$. Sendo assim, rejeitar a hipótese nula se $u \leq 5$ e $u \geq 15$, onde 5 e 15 são os valores de $u'_{0,025}$ e $u_{0,025}$, respectivamente, para $n_1 = 8$ e $n_2 = 11$. Como $u = 10$ é maior que 5 e menor que 15, segue que a hipótese nula não deve ser rejeitada, ou seja, não há indícios de que a sequência de lançamentos de 3 pontos desse jogador não seja aleatória.

O último jogador listado possui dados insuficientes para a análise.

3.3 Análise 2: Lançamentos Livres

Nesta seção vamos analisar os lançamentos livres – ou cestas de um ponto ou ainda cobrança de falta, como muitas pessoas os chamam. Os testes utilizados e a apresentação dos resultados seguirão de forma análoga ao que foi feito na Seção 3.2.

Em um jogo de basquetebol, normalmente os lances livres são feitos aos pares à uma determinada distância da cesta. No momento de cobrar a falta, o arremessador está isento da marcação do time adversário. Perceba que a situação em que este jogador está inserido é bastante diferente daquele que tenta efetuar um lance de 3 pontos, por exemplo, já que este último está sujeito a qualquer momento perder a bola antes mesmo da finalização à cesta, devido a pressão e tática de jogo dos seus oponentes.

Para se ter uma idéia dos valores gerais, a coluna 2 da Tabela 3.7 exhibe, para cada jogador, o valor total de tentativas à cestas de 1 ponto, bem como o número de acertos e erros nesses lances (colunas 3 e 4). As probabilidades globais de acerto, erro e do jogador i efetuar uma jogada foram calculadas nas colunas 5, 6 e 7, respectivamente. Note que quase todos fizeram mais acertos que

erros, exceto o *jogador 10*. Portanto, para os nove restantes, a porcentagem de acerto é maior que a de erro.

Tabela 3.7: Desempenho global dos jogadores em lançamentos livres

Jogador	Tentativas	Acertos	Erros	P(A)	P(E)	P(J=i)
14	129	112	17	0.868	0.131	0.188
8	115	89	26	0.773	0.226	0.168
33	60	44	16	0.733	0.266	0.087
26	77	48	29	0.623	0.376	0.112
15	29	22	7	0.758	0.241	0.042
13	65	41	24	0.630	0.369	0.095
10	72	35	37	0.486	0.513	0.105
12	26	18	8	0.692	0.307	0.038
16	3	2	1	0.666	0.333	0.004
11	107	85	22	0.794	0.205	0.156
Total	683	496	187			1.0

Procuramos também calcular a probabilidade de acerto em arremessos livres condicionada ao último acerto e ao último erro. Para esta análise, iremos considerar, de cada jogador, os lançamentos livres que foram arremessados aos pares. Isso implica numa pequena diminuição nos valores totais de tentativas, acertos e erros (comparar as Tabelas 3.7 e 3.8).

Tabela 3.8: Desempenho dos jogadores e probabilidades condicionais em lançamentos livres arremessados aos pares

Jogador	Tentativas	Acertos	Erros	$P(A_2 E_1)$	$P(A_2 A_1)$
14	106	92	14	0.923	0.858
8	94	73	21	0.70	0.808
33	48	35	13	0.769	0.705
26	66	42	24	0.478	0.714
15	26	21	5	0.60	0.90
13	56	32	24	0.695	0.50
10	68	33	35	0.411	0.545
12	22	15	7	0.50	0.733
16	2	1	1	—	0
11	88	68	20	0.70	0.805
Total	576	412	164		

Desenvolvemos no software Scilab um algoritmo para localizar os lances feitos aos pares (Anexo F), levando em consideração a sequência inteira (string) de lançamentos de cada jogador, para posteriormente calcular a probabilidade condicional. Será que ao acertar o primeiro lance livre existe uma maior probabilidade de acertar o segundo? Ou ainda, o erro no primeiro lance implica uma maior chance de acertar o segundo?

A Tabela 3.8, além de mostrar o desempenho dos jogadores em lances livres feitos aos pares, apresenta a probabilidade de acerto no segundo lance livre dado que ocorreu um erro no primeiro (coluna 5) e a probabilidade de acerto no segundo lance livre dado que ocorreu um acerto no primeiro (coluna 6). Excluindo o *jogador 16* da análise de probabilidade condicional pelo mesmo não ter um número suficiente para tal, dos nove restantes, seis possuem $P(A_2 | A_1) > P(A_2 | E_1)$. Desejamos determinar quão significativas são essas probabilidades e isso obtemos por meio do *teste da diferença entre duas proporções*.

3.3.1 Teste da Diferença entre Duas Proporções

As probabilidades condicionais em lances livres efetuados aos pares tornam-se elementos para o cálculo da diferença entre proporções. Considerando os dados das colunas 5 e 6 da Tabela 3.8, iremos proceder de forma similar ao que foi feito em 3.2.1.

A Tabela 3.9 ilustra o resultado do teste para o *jogador 14*. As diferenças entre as probabilidades condicionais são não significativas para este jogador.

Tabela 3.9: Diferença entre duas proporções para o *jogador 14*

J=14	P(A₂ E₁)	P(A₂ A₁)
P(A₂ E₁)	0	N
P(A₂ A₁)	N	0

S = diferença significativa

N = diferença não significativa

O mesmo teste aplicado a todos os jogadores mostrou que 7, dentre os 10 jogadores analisados, não possuem alteração significativa entre o percentual de acerto após um acerto e após um erro.

3.3.2 Correlação Serial de Ordem 1

Nesta seção, investigamos se existe alguma correlação entre os arremessos sucessivos de lances livres de um jogador, com base nos conceitos apresentados nas Seções 1.6.2 e 1.6.2.1.

A correlação serial de ordem 1 foi calculada sobre a sequência completa (*string*) de arremessos de cada atleta. Resta saber o quão significativas são essas correlações e para isto, podemos fazer uso do conceito de correlograma. A Tabela 3.10 resume as informações referentes ao coeficiente de correlação serial de ordem 1.

De forma análoga a desenvolvida na Seção 3.2.2, verificamos que todos os jogadores da Tabela 3.10 possuem seus valores de r_1 dentro dos limites de confiança, dados por $\pm 1.96/\sqrt{n}$. Portanto, com nível de significância de 5%, podemos afirmar que as sequências de lançamentos livres desses jogadores são *não correlacionadas*.

Tabela 3.10: Teste de hipóteses sobre o coeficiente de correlação serial de ordem 1 das sequências de lançamentos livres de cada jogador ao longo da temporada.

Jogador	Valor de n	Limites de Confiança	Correlação Serial r_1	Localização de r_1	Resultado do Teste
14	106	± 0.190	- 0.061	dentro dos limites	N
8	94	± 0.202	0.105	dentro dos limites	N
33	48	± 0.282	- 0.064	dentro dos limites	N
26	66	± 0.241	0.232	dentro dos limites	N
15	26	± 0.384	0.307	dentro dos limites	N
13	56	± 0.261	- 0.193	dentro dos limites	N
10	68	± 0.237	0.133	dentro dos limites	N
12	22	± 0.417	0.217	dentro dos limites	N
16	2	± 1.385	-1	dentro dos limites	N
11	88	± 0.208	0.106	dentro dos limites	N

S = teste significativo

N = teste não significativo

ANÁLISE ESTATÍSTICA DOS DADOS SIMULADOS COMPUTACIONALMENTE

Este capítulo é destinado a apresentação das análises realizadas com dados simulados de forma computacional. Todo o estudo apresentado aqui é abordado de modo análogo ao Capítulo 3.

A princípio, sequências de Bernoulli serão geradas por algoritmos computacionais com o intuito de representarem arremessos de 3 pontos e lances livres para dez jogadores de basquetebol. Objetivamos analisar e comparar o comportamento das sequências produzidas aleatoriamente com os registros reais de desempenho dos jogadores envolvidos nas pesquisas do terceiro capítulo. Queremos saber se as novas sequências de arremessos ao serem submetidas aos estudos aplicados no Capítulo 3 apresentarão resultados semelhantes aqueles obtidos quando se tratavam das verdadeiras sequências de arremessos.

Toda a base computacional existente nesta etapa do trabalho assim como os algoritmos computacionais que desenvolvemos para as simulações foram implementados no software Scilab.

4.1 Simulação 1: Lançamentos de 3 Pontos

Ao tentar produzir sequências aleatórias que representem arremessos de 3 pontos, precisamos primeiramente recorrer à Tabela 3.2, pois nela estão todas as informações que precisamos para a geração computacional de dados.

Para gerar as sequências, optamos por elaborar um algoritmo no Scilab para esta finalidade (Anexo D). Com o intuito de criar uma situação que se aproxime da realidade, tal programa inclui

as probabilidades de acertos de cada jogador bem como suas probabilidades de participação, além da soma total de lances realizados dentro do campeonato.

Através das sequências geradas, numa primeira exploração dos dados, obtemos a Tabela 4.1. Podemos observar que para todos os jogadores, exceto o *jogador 11*, a probabilidade de erro $P(E)$ é maior que a probabilidade de acerto $P(A)$, ou seja, para nove dos dez jogadores tem-se $P(E) > P(A)$. Observe ainda que, no geral, os valores da Tabela 4.1 se aproximam muito dos valores originais (comparar com a Tabela 3.2).

Tabela 4.1: Desempenho dos jogadores na simulação de lançamentos de 3 pontos

Jogador	Tentativas	Acertos	Erros	P(A)	P(E)	P(J=i)
14	207	93	114	0.449	0.550	0,274
8	133	56	77	0.421	0.578	0,176
33	94	31	63	0.329	0.670	0,124
26	70	28	42	0.40	0.60	0,093
15	83	32	51	0.385	0.614	0,110
13	69	29	40	0.420	0.579	0,091
10	45	17	28	0.377	0.622	0,059
12	35	15	20	0.428	0.571	0,046
16	12	4	8	0.333	0.666	0,015
11	5	4	1	0.80	0.20	0,006
Total	753	309	444			1.0

Podemos também avaliar as sequências com relação ao estudo de probabilidade condicional. Após gerar as sequências de lançamentos de 3 pontos para todos os integrantes do Bauru, fez-se necessário o cálculo das probabilidades condicionadas aos últimos acertos e erros (calculadas através do algoritmo computacional dado pelo Anexo C), o que deu origem a Tabela 4.2.

Tabela 4.2: Probabilidades Condicionais sobre as sequências de lançamentos de 3 pontos geradas pela simulação computacional

Jogador	$P(A 3E)$	$P(A 2E)$	$P(A 1E)$	$P(A)$	$P(A 1A)$	$P(A 2A)$	$P(A 3A)$
14	0.410	0.40	0.429	0,449	0.478	0.511	0.666
8	0.434	0.428	0.434	0,421	0.410	0.434	0.20
33	0.565	0.394	0.396	0,329	0.20	0.166	0
26	0.250	0.346	0.341	0,40	0.50	0.642	0.666
15	0.368	0.354	0.360	0,385	0.406	0.615	0.375
13	0.70	0.523	0.475	0,420	0.357	0.60	0.50
10	0.250	0.235	0.333	0,377	0.411	0.428	0.333
12	0.250	0.375	0.526	0,428	0.266	0.250	0
16	1.0	0.50	0.428	0,333	0.250	0	—
11	—	—	1.0	0,80	0.666	1.0	—

Note que seis jogadores listados na tabela possuem probabilidade de acerto após um acerto menor que após um erro ($P(A | 1A) < P(A | 1E)$). Comparações entre as colunas 3 e 7, e entre as colunas 2 e 8, evidenciam resultados semelhantes. Esses valores não se diferem muito dos encontrados na seção 3.2, pois nela cinco dentre os dez jogadores apresentaram $P(A | 1A) < P(A | 1E)$. Buscamos identificar o quão significativas são essas diferenças entre as probabilidades e conseguimos isto através do *teste da diferença entre duas proporções*.

4.1.1 Teste da Diferença entre Duas Proporções

A partir das sequências de arremessos de 3 pontos, obtidas na simulação computacional de dados, procedemos de forma análoga ao que foi desenvolvido na Seção 3.2.1.

Para a aplicação do *teste da diferença entre duas proporções* faremos uso da Tabela 4.2, com base no conteúdo apresentado na Seção 1.6.3 do Capítulo 1. Para cada jogador da simulação, nos interessa saber se a diferença entre as probabilidades condicionais, analisadas duas a duas, são ou não significativas.

Conforme o esperado, o teste nos revela que as diferenças entre proporções são não significativas para os jogadores objetos de estudo, ou seja, não existem diferenças estatísticas entre as probabilidades de acerto, dado que ocorreram acertos, e as probabilidades de acerto, dado que ocorreram erros.

4.1.2 Correlação Serial de Ordem 1

Nesta seção utilizaremos os conceitos desenvolvidos nas Seções 1.6.2 e 1.6.2.1, semelhantemente ao que foi feito na Seção 3.2.2, porém agora trata-se de arremessos gerados de forma computacional.

Para a análise do coeficiente de autocorrelação, calculamos a *correlação serial de ordem 1* (r_1) sobre as sequências simuladas de arremessos de 3 pontos, bem como identificamos os limites de confiança e localização de r_1 para formular o resultado do teste de hipótese, conforme mostra a Tabela 4.3. Consideramos como hipótese nula H_0 que as sequências de arremessos são *não correlacionadas*.

Tabela 4.3: Teste de hipóteses sobre o coeficiente de correlação serial de ordem 1 das sequências de lançamentos de 3 pontos geradas pela simulação computacional

Jogador	Valor de n	Limites de Confiança	Correlação Serial r_1	Localização de r_1	Resultado do Teste
14	207	± 0.136	0.048	dentro dos limites	N
8	133	± 0.169	- 0.023	dentro dos limites	N
33	94	± 0.202	- 0.194	dentro dos limites	N
26	70	± 0.234	0.159	dentro dos limites	N
15	83	± 0.215	0.046	dentro dos limites	N
13	69	± 0.235	- 0.117	dentro dos limites	N
10	45	± 0.292	0.078	dentro dos limites	N
12	35	± 0.331	- 0.262	dentro dos limites	N
16	12	± 0.565	- 0.181	dentro dos limites	N
11	5	± 0.876	- 0.375	dentro dos limites	N

S = teste significativo

N = teste não significativo

Todos os dez jogadores listados na Tabela 4.3 possuem seus coeficientes de *correlação serial de ordem 1* dentro dos limites de confiança, dados por $\pm 1.96/\sqrt{n}$, sendo n igual ao número de

tentativas à cestas de 3 pontos. Portanto, com nível de significância de 5%, podemos afirmar que as sequências simuladas de arremessos de 3 pontos desses jogadores são *não correlacionadas*.

4.1.3 Análise de Runs e o Teste de Aleatoriedade

Esta seção tem como correspondente a Seção 3.6. O procedimento usado aqui foi anteriormente explicado com mais detalhes na Seção 1.6.1, no Capítulo 1.

Para fazer uso do *teste de aleatoriedade*, consideramos como hipótese nula H_0 que as sequências de lançamentos de 3 pontos geradas pela simulação computacional são aleatórias e como hipótese alternativa H_A que tais sequências não são aleatórias. Para $\alpha = 0.05$, rejeitamos a hipótese nula se $|z| > z_{\alpha/2}$, ou seja, se $z \leq -1.96$ ou $z \geq 1.96$, conforme visto em 1.6.1.

Os cálculos obtidos na análise de *runs* e no *teste de aleatoriedade* sobre as sequências simuladas de lançamentos de 3 pontos estão exibidos na Tabela 4.4.

Tabela 4.4: Análise de Runs e o Teste de Aleatoriedade sobre as sequências de lançamentos de 3 pontos geradas pela simulação computacional

Jogador	Acertos	Erros	N ^o _{obs} de Runs	N ^o _{esp} de Runs	Z
14	93	114	98	103.43	- 0.765
8	56	77	67	65.84	0.206
33	31	63	50	42.55	1.749
26	28	42	29	34.60	- 1.405
15	32	51	38	40.32	- 0.542
13	29	40	38	34.62	0.840
10	17	28	20	22.15	- 0.692
12	15	20	22	18.14	1.351
16	4	8	7	6.33	0.459
11	4	1	3	2.60	0.816
Média	30.9	44.4	37.2		

Podemos verificar na tabela que os oito primeiros jogadores possuem os resultados da estatística Z dentro do intervalo de -1.96 a 1.96 , e desta forma, a hipótese nula não pode ser rejeitada. Portanto, não há evidências estatísticas de que as sequências de lançamentos de 3 pontos, geradas pela simulação computacional, não sejam aleatórias.

Os jogadores 11 e 16 não possuem dados suficientes para realizar a análise.

4.2 Simulação 2: Lançamentos Livres

Para produzir sequências aleatórias que representem lances livres arremessados aos pares, recorreremos a Tabela 3.8 para calcularmos as probabilidades de acerto e de participação de cada jogador no jogo e então, as inserimos no algoritmo computacional que desenvolvemos para a geração de dados (Anexo E). Com as sequências de arremessos completamente geradas, iremos desenvolver os mesmos estudos realizados na Seção 3.3 do Capítulo 3.

A Tabela 4.5 descreve o desempenho dos jogadores na simulação de lances livres arremessados aos pares e irá nos auxiliar nas análises subsequentes. De um modo geral, os valores encontrados na tabela se aproximam bastante dos valores encontrados na Tabela 3.8.

Após gerarmos sequências que representam lances livres para os dez jogadores em estudo, podemos calcular as probabilidades de acerto condicionadas ao último acerto e ao último erro. Isso nos ajuda a saber se um jogador tem uma maior probabilidade de acertar depois de um acerto ou depois de um erro.

Tabela 4.5: Desempenho dos jogadores na simulação de lançamentos livres arremessados aos pares

Jogador	Tentativas	Acertos	Erros	P(A)	P(J=i)
14	86	75	11	0.872	0.151
8	92	74	18	0.804	0.162
33	62	43	19	0.693	0.109
26	66	39	27	0.590	0.116
15	28	23	5	0.821	0.049
13	58	35	23	0.603	0.102
10	68	28	40	0.411	0.119
12	20	13	7	0.650	0.035
16	2	2	0	1.0	0.003
11	86	67	19	0.779	0.151
Total	568	399	169		1.0

A Tabela 4.6 mostra o cálculo das probabilidades de acerto no segundo lance livre condicionada ao erro no primeiro lance (coluna 2) e ao acerto no primeiro lance (coluna 3), na simulação de lançamentos livres arremessados aos pares.

Tabela 4.6: Probabilidades condicionais na simulação de lançamentos livres arremessados aos pares

Jogador	$P(A_2 E_1)$	$P(A_2 A_1)$
14	0.727	0.891
8	0.823	0.797
33	0.578	0.738
26	0.555	0.631
15	0.80	0.818
13	0.565	0.647
10	0.435	0.392
12	0.857	0.583
16	—	1.0
11	0.736	0.787

Excluindo o *jogador 16* da análise de probabilidade condicional por este não ter um número suficiente de dados para tal, dos nove restantes, seis possuem $P(A_2 | A_1) > P(A_2 | E_1)$. No entanto, queremos determinar quão significativas são essas diferenças de probabilidades e isto obtemos através do *teste da diferença entre duas proporções*.

4.2.1 Teste da Diferença entre Duas Proporções

A partir das sequências de arremessos de lance livre, obtidas na simulação computacional de dados, iremos proceder de acordo com o que foi desenvolvido na Seção 3.3.1 do terceiro capítulo.

Para a aplicação do *teste da diferença entre duas proporções*, consideremos os dados da Tabela 4.6. Nos interessa identificar se a diferença entre as probabilidades condicionais, analisadas duas a duas, são ou não significativas.

Para cada jogador da simulação, excluindo aqueles que possuem dados insuficientes, o teste nos revela que as diferenças entre as proporções são não significativas para os jogadores objetos de estudo, isto significa que não existem diferenças estatísticas entre as probabilidades de acerto, dado que ocorreu um acerto, e as probabilidades de acerto, dado que ocorreu um erro.

4.2.2 Correlação Serial de Ordem 1

Assim como nas outras seções que desenvolvemos o estudo de correlação serial, esta também segue os conceitos apresentados nas Seções 1.6.2 e 1.6.2.1 do Capítulo 1.

O cálculo da *correlação serial de ordem 1* foi feito com base na sequência completa de arremessos livres, simulada para cada jogador.

Pela análise da Tabela 4.7 verificamos que todos os jogadores possuem seus valores de r_1 dentro dos limites de confiança dados por $\pm 1.96/\sqrt{n}$. Portanto, com nível de significância de 5%, podemos afirmar que a sequência simulada de lances livres desses jogadores são *não correlacionadas*.

Tabela 4.7: Teste de hipóteses sobre o coeficiente de correlação serial de ordem 1 das sequências de lances livres geradas pela simulação computacional

Jogador	Valor de n	Limites de Confiança	Correlação Serial r_1	Localização de r_1	Resultado do Teste
14	86	± 0.211	0.166	dentro dos limites	N
8	92	± 0.204	- 0.025	dentro dos limites	N
33	62	± 0.248	0.160	dentro dos limites	N
26	66	± 0.241	0.076	dentro dos limites	N
15	28	± 0.370	0.019	dentro dos limites	N
13	58	± 0.257	0.082	dentro dos limites	N
10	68	± 0.237	- 0.043	dentro dos limites	N
12	20	± 0.438	- 0.282	dentro dos limites	N
16	2	± 1.385	divisão por zero	—	—
11	86	± 0.211	0.051	dentro dos limites	N

S = teste significativo

N = teste não significativo

Conclusão

Nesta dissertação, procuramos desenvolver um texto que pudesse ser acessível a alunos de graduação e/ou demais pessoas interessadas em estatística aplicada ao esporte, em particular aos jogos de basquetebol. Neste sentido, vários exemplos foram criados no Capítulo 1 para tornar mais claro o trabalho que iríamos fazer com os registros de arremessos de basquetebol.

Com base no que foi desenvolvido até o momento, podemos comparar os resultados dos testes aplicados sobre as sequências de arremessos simuladas de forma computacional com os resultados dos testes aplicados sobre as verdadeiras sequências de arremessos.

Após fazermos uma breve revisão sobre probabilidade e estatística e falarmos sobre o artigo que também nos motivou ao desenvolvimento desta dissertação, analisamos estatisticamente o desempenho de dez jogadores da equipe profissional de basquete de Bauru, nos jogos ocorridos na temporada 2009 – 2010.

De posse dos dados da equipe considerada para os estudos, tivemos todo um trabalho de separação dos registros de cada um dos atletas para darmos início as análises pretendidas. Para o processamento dos dados foi utilizado o software Scilab. Em seguida, aplicamos testes estatísticos sobre as sequências de arremessos de um e três pontos, isoladas por jogador.

Em uma outra etapa da dissertação, produzimos sequências de Bernoulli por algoritmos computacionais de tal forma que representassem arremessos de um e três pontos, simulando os registros reais da equipe de Bauru. O objetivo nesta fase do trabalho era aplicar sobre arremessos simulados computacionalmente os mesmos estudos realizados com sequências reais de arremessos de basquetebol, para que assim pudéssemos comparar os resultados dos testes. Logo abaixo, seguimos com uma discussão sobre os resultados obtidos ao longo do nosso trabalho.

• Resultados do Capítulo 3

No Capítulo 3, onde tratamos somente de registros reais de arremessos de basquetebol, aplicamos um teste estatístico denominado *teste da diferença entre duas proporções*, o qual nos permite identificar se a diferença entre probabilidades condicionadas a acertos e erros é ou não significativa. Por sua vez, o teste nos mostrou que, em lançamentos de três pontos, **seis** jogadores possuem essa

diferença de probabilidade *não significativa*; em lances livres, **sete** jogadores possuem a diferença de probabilidade *não significativa*. Isso significa que não existem diferenças estatísticas entre as probabilidades de acerto, condicionadas ao(s) último(s) acerto(s), e as probabilidades de acerto, condicionadas ao(s) último(s) erro(s). Os jogadores restantes não possuíam dados suficientes para a análise.

A análise de *correlação serial de ordem 1* mostra que **oito** jogadores possuem suas sequências de arremessos de três pontos *não correlacionadas* e **todos** os dez jogadores analisados possuem suas sequências de arremessos livres *não correlacionadas*, ou seja, o teste revela que para a sequência de desempenho desses atletas um lançamento não tem relação com o anterior. Em outras palavras, o resultado de um lançamento não é influenciado pelo resultado do anterior.

No *teste de aleatoriedade*, não houve evidências estatísticas de que as sequências de arremessos de três pontos de **sete** jogadores, dentre os dez analisados, não ocorram de forma *aleatória*.

• Resultados do Capítulo 4

No Capítulo 4 discutimos os mesmos estudos realizados no Capítulo 3, porém levando em consideração as sequências de arremessos livres e de três pontos simuladas de forma computacional.

A análise de *correlação serial de ordem 1*, o *teste de aleatoriedade* e a aplicação do *teste da diferença entre duas proporções* resultaram em informações um tanto quanto esperadas: para os jogadores da simulação que possuíam número suficiente de dados para realizar os estudos, suas sequências de arremessos de três pontos se mostraram *não correlacionadas*, *aleatórias* e apresentaram diferença entre probabilidades condicionais *não significativas*; em arremessos livres, todos os jogadores possuíam suas sequências *não correlacionadas* e, excluindo aqueles que não possuíam dados suficientes para o teste, diferença entre probabilidades condicionais *não significativas*.

• Conclusão dos resultados

A comparação dos resultados dos testes nos permite concluir que as sequências de arremessos obtidas de jogos reais de basquetebol possuem comportamento semelhante às sequências de Bernoulli geradas de forma pseudoaleatórias por algoritmos computacionais.

Considerações Finais

A crença na chamada boa fase, de um modo geral, é algo identificado de forma natural por torcedores e pessoas envolvidas com o esporte. No basquetebol isso não é diferente. Percebemos em entrevistas a jornais e a outros meios de comunicação que até mesmo os atletas de alto rendimento falam em momentos de boa fase.

As pessoas tendem a acreditar que o acerto em um arremesso ocasiona acertos nos próximos (e vice-versa). Se um jogador acerta alguns arremessos consecutivos, este é denominado pela grande maioria como um jogador que está em boa fase. Geralmente essa crença é baseada em curtos intervalos de tempo, como por exemplo um jogo, um quarto de jogo ou uma pequena série de jogos.

Os trabalhos que tentam verificar evidências da boa fase, usando testes estatísticos, precisam utilizar um grande conjunto de dados, uma série histórica com muitos termos, uma sequência de vários arremessos, etc, tudo isto para que a *Lei dos Grandes Números* possa ser empregada. Isto quer dizer que só faz sentido afirmar algo sobre a proporção de sucessos a longo prazo, e que quase nada tem a dizer sobre o experimento isolado. Por exemplo, em uma simulação computacional da jogada de uma moeda equilibrada, nas 5 primeiras jogadas simuladas ocorreram 3 caras, nas 10 primeiras jogadas ocorreram 6 caras, nas 15 primeiras ocorreram 8 caras, nas 20 primeiras ocorreram 12 caras, nas 25 primeiras ocorreram 14 caras, ..., e nas 100 primeiras jogadas ocorreram 51 caras; as proporções correspondentes são 0.6, 0.6, 0.53, 0.6, 0.56, ..., 0.51. Embora a proporção de cara oscile, cada vez chega mais próxima de 0.5 a probabilidade de cara em cada jogada da moeda.

A partir de um grande número de dados, os testes estatísticos podem acabar concluindo coisas semelhantes às que concluímos neste trabalho. Mas e a crença na chamada boa fase? Será que as pessoas estão erradas?

Analisar a boa fase de forma fiel ao que a maioria das pessoas acreditam é uma tarefa difícil, pois neste caso, teríamos que tomar conclusões estatísticas com base em um conjunto muito pequeno de informações, tal como os últimos quatro ou cinco arremessos de um jogador. Isso poderia

não relatar o verdadeiro desempenho de um atleta, ou seja, concluir que um jogador está em boa ou má fase levando em consideração apenas uma pequena quantidade de arremessos efetuados num jogo, por exemplo, é algo pouco preciso.

Existem ainda outros fatores que podem ser favoráveis (ou não) e que poderiam produzir resultados positivos (ou negativos) dentro do jogo, ou em uma temporada de jogos: confiança, disposição, angústia, enfim, questões físicas e psicológicas e esses elementos geralmente não são levados em consideração nos estudos.

Neste trabalho analisamos longas séries de arremessos de basquetebol da equipe profissional de Bauru. Consideramos registros de arremessos consecutivos de lances livres e arremessos de três pontos dos dez integrantes da equipe estudada que obtiveram maior número de lançamentos em uma temporada inteira de jogos. Em seguida, produzimos computacionalmente sequências aleatórias e aplicamos sobre estas os mesmos estudos realizados com dados reais, a fim de comparar os resultados. Os resultados dos métodos estatísticos que aplicamos sobre os dois conjuntos de dados revelaram que as sequências de arremessos obtidas dos jogos de basquetebol possuem comportamento semelhante às sequências simuladas de forma aleatória por algoritmos computacionais.

Estudos como estes tem ajudado pesquisadores a desenvolver simuladores de jogos para auxiliar os profissionais em processos de tomada de decisão. Acreditamos que análises como as que desenvolvemos ao longo desta dissertação podem contribuir para uma maior compreensão do jogo e orientação dos atletas nos treinamentos para que assim possam conquistar, cada vez mais, um melhor desempenho nas competições.

Referências Bibliográficas

- [1] Albert, J., Bennett, J. and Cochran, J. J. Anthology of Statistics in Sports, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, 2005.
- [2] DeGroot, M. H. and Schervish, M. J. Probability and Statistics, 3rd ed. Boston: Addison-Wesley, c2002.
- [3] Ehlers, R. S. (2009) Análise de Séries Temporais:
<http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>
- [4] Freund, J. E. and Simon, G. A. Estatística Aplicada: economia, administração e contabilidade, 9ª ed., Porto Alegre: Bookman, 2000.
- [5] Gilovich, T., Vallone, R. and Tversky, A. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* 17, pp. 295-314, (1985).
- [6] Larkey, P., Smith, R. and Kadane, J. It's Okay to Believe in the "Hot Hand." *Chance*, 2(4), 22-30, (1989).
- [7] Magalhães, M. N. & Lima, A. C. P. Noções de Probabilidade e Estatística. 6ed rev. São Paulo: Editora da Universidade de São Paulo, 2005.
- [8] Mayer, P.L, Probabilidade: Aplicações à Estatística, 2ª ed., Rio de Janeiro, RJ: Livros Técnicos e Científicos, c1984.
- [9] Milone, G. Estatística Geral e Aplicada. São Paulo, SP: Thomson Learning, 2004.
- [10] Mood, A. M., Graybill, F. A. and Boes, D. C. Introduction to the Theory of Statistics. 3rd ed. New York, N.Y.: McGraw-Hill, 1974.
- [11] Morettin, P.A., Toloi, C.M.C. Análise de Séries Temporais, 2ª ed., São Paulo, SP: E. Blucher, 2006.

-
- [12] Ross, S. A first course in probability, 7th ed. Upper Saddle River, N. J.: Prentice-Hall, c2006.
- [13] Shumway, R. H. Applied Statistical Time Series Analysis, Englewood Cliffs, N. J.: Prentice Hall, c1988.
- [14] Siegel, S. Nonparametric Statistics for the Behavioral Sciences”, New York, N.Y.: McGraw-Hill, 1956.
- [15] Wardrop, R. L. Statistical Tests for the Hot-Hand in Basketball in a Controlled Setting, Technical report. Department of Statistics, University of Wisconsin-Madison -EUA.

Anexos

- **Anexo A - Algoritmo para computar o número de *runs* numa sequência de arremessos.**

```
function [n]=contarun(v)
m=length(v);
if length(v)>0
x=v(1);
n=1;
for i=2:m
y=v(i)
if y~=x
n=n+1;
x=y;
end
end
else
n=%nan;
end
endfunction
```

- **Anexo B - Algoritmo para o cálculo da correlação serial de ordem 1.**

```
function [r,Li,Ls]=correlserial(v)
n = length(v);
xbarra = sum(v)/n;
S1 = 0;
for i = 1:n-1
S1 = S1 + ( v(i) - xbarra)*( v(i+1) - xbarra);
end
S2 = 0;
for i = 1:n
S2 = S2 + (v(i) - xbarra)^2
end
r = S1/(((n-1)/n)*S2);
Li = (1/n) - (1.96/sqrt(n));
Ls = (1/n) + (1.96/sqrt(n));
endfunction
```

- **Anexo C - Algoritmo para o cálculo da probabilidade condicional.**

```
function [prob,ac,total]=contaerros(v,pp)
prob=[];
ac=[];
total=[];
if pp==0
total=0;
ac=0;
m=length(v);
total=m;
for i = 1:m
if v(i)>0
ac=ac+1;
end
end
if total>0
prob=ac/total;
else prob =%nan;
end
end
if pp == -1
total=0;
ac=0;
m=length(v); /
for i = 1:m-1
if v(i)<0
total=total+1;
if v(i+1)>0
ac=ac+1;
end
end
end
end
```

```
if total>0
prob=ac/total;
else
prob =%nan;
end
end
if pp==-2
total=0;
ac=0;
m=length(v);
for i = 1:m-2
if (v(i)<0) & (v(i+1)<0)
total=total+1;
if v(i+2)>0
ac=ac+1;
end
end
end
if total>0
prob=ac/total;
else
prob =%nan;
end
end
if pp==-3
total=0;
ac=0;
m=length(v);
for i = 1:m-3
if (v(i)<0) & (v(i+1)<0) & (v(i+2)<0)
total=total+1;
if v(i+3)>0
```

```
ac=ac+1;
end
end
end
if total>0
prob=ac/total;
else
prob =%nan;
end
end
if pp==1
total=0;
ac=0;
m=length(v);
for i = 1:m-1
if (v(i)>0)
total=total+1;
if v(i+1)>0
ac=ac+1;
end
end
end
if total>0
prob=ac/total;
else
prob =%nan;
end
end
if pp==2
total=0;
ac=0;
m=length(v);
```

```
for i = 1:m-2
if (v(i)>0) & (v(i+1)>0)
total=total+1;
if v(i+2)>0
ac=ac+1;
end
end
end
if total>0
prob=ac/total;
else
prob =%nan;
end
end
if pp==3
total=0;
ac=0;
m=length(v);
for i = 1:m-3
if (v(i)>0) & (v(i+1)>0) & (v(i+2)>0)
total=total+1;
if v(i+3)>0
ac=ac+1;
end
end
end
if total>0
prob=ac/total;
else prob =%nan;
end
end
endfunction
```

• Anexo D - Algoritmo para simulação de arremessos de 3 pontos.

```
Num = [14, 8, 33, 26, 15, 13, 10, 12, 16, 11];
J = [0.2696, 0.1806, 0.1129, 0.1036, 0.0943, 0.0930, 0.0677, 0.0465, 0.0252, 0.0066];
P=[0.4680, 0.4485, 0.3412, 0.3846, 0.3662, 0.4000, 0.3137, 0.4286, 0.4211, 0.2000];
A=zeros(753,2);
[m,n]=size(A);
for i=1:m
x = round(10000*rand());
if x <= 2696
jogador=14;
y = round(10000*rand());
if y <=4680
lance = 3;
else
lance=-3;
end
elseif (x>2696) & (x<=4502)
jogador=8;
y = round(10000*rand());
if y <=4485
lance = 3;
else
lance=-3;
end
elseif (x>4502) & (x<=5631)
jogador=33;
y = round(10000*rand());
if y <=3412
lance = 3;
else
lance=-3;
end
end
```

```
elseif (x>5631) & (x<=6667)
jogador=26;
y = round(10000*rand());
if y <=3846
lance = 3;
else
lance=-3;
end
elseif (x>6667) & (x<=7610)
jogador=15;
y = round(10000*rand());
if y <=3662
lance = 3;
else
lance=-3;
end
elseif (x>7610) & (x<=8540)
jogador=13;
y = round(10000*rand());
if y <=4000
lance = 3;
else
lance=-3;
end
elseif (x>8540) & (x<=9217)
jogador=10;
y = round(10000*rand());
if y <=3137
lance = 3;
else
lance=-3;
end
end
```

```
elseif (x>9217) & (x<=9682)
jogador=12;
y = round(10000*rand());
if y <=4286
lance = 3;
else
lance=-3;
end
elseif (x>9682) & (x<=9934)
jogador=16;
y = round(10000*rand());
if y <=4211
lance = 3;
else
lance=-3;
end
elseif (x>9934) & (x<=10000)
jogador=11;
y = round(10000*rand());
if y <=2000
lance = 3;
else
lance=-3;
end
end
end
A(i,2)=jogador;
A(i,1)=lance;
end
```

• Anexo E - Algoritmo para simulação de lançamentos livres feitos aos pares

```
Num = [14, 8, 33, 26, 15, 13, 10, 12, 16, 11];
J = [0.1840, 0.1632, 0.0833, 0.1146, 0.0451, 0.0972, 0.1181, 0.0382, 0.0035, 0.1528];
P=[0.8679, 0.7766, 0.7292, 0.6364, 0.8077, 0.5714, 0.4853, 0.6818, 0.5000, 0.7727];
A=zeros(576,2);
[m,n]=size(A);
for i=1:m /
x = round(10000*rand());
if x <= 1840
jogador=14;
y = round(10000*rand());
if y <=8679
lance = 1;
else
lance=-1;
end
elseif (x>1840) & (x<=3472)
jogador=8;
y = round(10000*rand());
if y <=7766
lance = 1;
else
lance=-1;
end
elseif (x>3472) & (x<=4305)
jogador=33;
y = round(10000*rand());
if y <=7292
lance = 1;
else
lance=-1;
end
end
```

```
elseif (x>4305) & (x<=5451)
jogador=26;
y = round(10000*rand());
if y <=6364
lance = 1;
else
lance=-1;
end
elseif (x>5451) & (x<=5902)
jogador=15;
y = round(10000*rand());
if y <=8077
lance = 1;
else
lance=-1;
end
elseif (x>5902) & (x<=6874)
jogador=13;
y = round(10000*rand());
if y <=5714
lance = 1;
else
lance=-1;
end
elseif (x>6874) & (x<=8055)
jogador=10;
y = round(10000*rand());
if y <=4853
lance = 1;
else
lance=-1;
end
```

```
elseif (x>8055) & (x<=8437)
jogador=12;
y = round(10000*rand());
if y <=6818
lance = 1;
else
lance=-1;
end
elseif (x>8437) & (x<=8472)
jogador=16;
y = round(10000*rand());
if y <=5000
lance = 1;
else
lance=-1;
end
elseif (x>8472) & (x<=10000)
jogador=11;
y = round(10000*rand());
if y <=7727
lance = 1;
else
lance=-1;
end
end
end
A(i,2)=jogador;
A(i,1)=lance;
end
```

- **Anexo F - Algoritmo para selecionar pares de lance livre dentro de uma sequência de arremessos com pontos variados.**

```
function [v]=seleciona(A,jog)
A(A(:,2)~=jog) = 0;
lances = A(:,1);
[m,n] = size(lances);
v = [];
flag = 0;
for i = 1:(m - 1)
if abs(lances(i)) == 1
if flag == 0
if abs(lances(i + 1)) == 1
v = [v;lances(i);lances(i + 1)]
flag = 1;
end
else
flag = 0;
end
end
end
endfunction
```

- Anexo G - Tabelas Estatísticas

Tabela I: Valores críticos de $u_{0,025}$

$n_2 \setminus n_1$	4	5	6	7	8	9	10	11	12	13	14	15
4		9	9									
5	9	10	10	11	11							
6	9	10	11	12	12	13	13	13	13			
7		11	12	13	13	14	14	14	14	15	15	15
8		11	12	13	14	14	15	15	16	16	16	16
9			13	14	14	15	16	16	16	17	17	18
10			13	14	15	16	16	17	17	18	18	18
11			13	14	15	16	17	17	18	19	19	19
12			13	14	16	16	17	18	19	19	20	20
13				15	16	17	18	19	19	20	20	21
14				15	16	17	18	19	20	20	21	22
15				15	16	18	18	19	20	21	22	22

Tabela II: Valores críticos de $u'_{0,025}$

$n_2 \setminus n_1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2											2	2	2	2
3					2	2	2	2	2	2	2	2	2	3
4				2	2	2	3	3	3	3	3	3	3	3
5			2	2	3	3	3	3	3	4	4	4	4	4
6		2	2	3	3	3	3	4	4	4	4	5	5	5
7		2	2	3	3	3	4	4	5	5	5	5	5	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6
9		2	3	3	4	4	5	5	5	6	6	6	7	7
10		2	3	3	4	5	5	5	6	6	7	7	7	7
11		2	3	4	4	5	5	6	6	7	7	7	8	8
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10