Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Instituto de Matemática e Estatística

Marcelo Tobias Magalhães de Carvalho

Mega Sena – Vitória Improvável

Rio de Janeiro

2018

Marcelo Tobias Magalhães de Carvalho

Mega Sena – Vitória improvável

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, do Instituto de Matemática e Estatística, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof.ª Dra. Patrícia Nunes da Silva

Coorientador: Prof. Dr. André Luiz cordeiro dos Santos

Rio de Janeiro

2018

CATALOGAÇÃO NA FONTE

UERJ / REDE SIRIUS / BIBLIOTECA CTC-A

C331 Carvalho, Marcelo Tobias Magalhães de.
Mega Sena: vitória improvável / Marcelo Tobias Magalhães de Carvalho. – 2018.
81f. : il.

Orientadora: Patrícia Nunes da Silva.
Coorientador: André Luiz Cordeiro dos Santos.
Dissertação (mestrado profissional em Matemática em rede nacional - PROFMAT) - Universidade do Estado do Rio de Janeiro, Instituto de Matemática e Estatísitca.

1. Combinações (Matemática) – Teses. 2. Probabilidade – Teses. 3. Loterias – Teses. I. Silva, Patrícia Nunes da. II. Santos, André Luiz Cordeiro dos. III. Universidade do Estado do Rio de Janeiro. Instituto de Matemática e Estatística. III. Título.

CDU 519.1

Marcelo Tobias Magalhães de Carvalho

Mega Sena – vitória improvável

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao programa de mestrado profissional em Matemática em rede nacional - PROFMAT, do instituto de Matemática e Estatística, da universidade do estado do Rio de Janeiro.

Aprovada em 03 de dezembro de 2018.

Banca Examinadora

Prof.ª Dra. Patrícia Nunes da Silva (Orientadora)

Instituto de Matemática e Estatística - UERJ

Prof. Dr. André Luiz Cordeiro dos Santos (Coorientador)

Instituto de Matemática e Estatística - UERJ

Prof. Dr. Helvécio Rubens Crippa

Instituto de Matemática e Estatística - UERJ

Prof.ª Dra. Brígida Alexandre Sartini

Universidade Federal Rural do Rio de Janeiro

Rio de Janeiro

2018

RESUMO

CARVALHO, Marcelo Tobias Magalhães de. Mega sena: vitória improvável. 2018. 81f. Dissertação (Mestrado profissional em Matemática em rede nacional - PROFMAT) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018.

Os jogos de apostas baseados em sorteios aleatórios surgiram no Brasil para aumentar a coleta de impostos para o governo. Ao mesmo tempo, trouxe para a população o desejo de ganhos financeiros rápidos e fáceis. Com este foco, algumas pessoas procuram meios de estudar os resultados anteriores de um sorteio para tentar adivinhar os possíveis resultados futuros. A mega-sena, hoje, é o jogo de sorteio aleatório mais utilizado pelos apostadores, daí a ideia de analisar seus resultados e os erros cometidos por aqueles apostadores que buscam prever os resultados futuros pela análise do passado. Neste contexto, a Lei dos grandes números ajudará a mostrar a aleatoriedade do sorteio e a lei dos pequenos números trará as explicações sobre os erros cometidos pelos jogadores ao tentar realizar essa análise profética.

Palavras-chave: Mega Sena. Lei dos grandes números. Lei dos pequenos números.

ABSTRACT

CARVALHO, Marcelo Tobias Magalhães de. *Mega Sena – unlikely victory.* 2018. 81f. Dissertação (Mestrado profissional em Matemática em rede nacional - PROFMAT) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018.

Gambling based on random lottery arose in Brazil to increase tax collection for the government. At the same time, it brought to population the desire for quick and easy financial gains. With this focus, some people search for ways to study the past results of a lottery to try to guess the possible future results. The mega-sena today is the random lottery game most used by bettors, therefor the idea of analyzing their results and the mistakes made by those bettors who seek to predict future results by analyzing the past. In this context, the Law of Large Numbers will help to show the randomness of the draw, and the law of small numbers will provide explanations for the mistakes made by players in attempting this prophetic analysis.

Keywords: Mega Sena. Law of Large Numbers. Law of Small Numbers.

LISTA DE TABELAS

SUMÁRIO

INTRODUÇAO

Por muitas vezes observei as casas lotéricas cheias, com filas enormes. Eram pessoas pagando pelo sonho de ficarem ricas. Pessoas que empregam seu último real que sobra do salário apertado do mês para sonhar com uma vida de milionário sem saber que a loteria federal surgiu, para o governo, como mais uma fonte de renda, mais um imposto a ser pago pelo cidadão brasileiro.

Com isso, surgiu a ideia de fazer um estudo detalhado com algumas das formas de cálculos probabilísticos tentando chegar a um resultado conclusivo sobre os jogos de sorte, ou azar, em que os apostadores escolhem números a serem sorteados.

Para basear este estudo escolhi o jogo da mega-sena por ser o mais conhecido, mais jogado, e o que paga maiores valores ao acertador das dezenas sorteadas. Durante os estudos estatísticos dos resultados já apresentados pelos sorteios da mega sena, pude perceber que os resultados tendem a atender a Lei dos grandes números. E que observando as dicas de jogadores que afirmam já ter ganho, encontrei os erros estatísticos enunciados na Lei dos pequenos números.

Neste trabalho, vou abordar os erros cometidos pelos jogadores, aproveitando para abordar, também, alguns erros que podemos cometer no dia a dia. Para isso, usarei alguns exemplos trazidos pelo livro *O andar do bêbado*, de Leonard Mlodinow.

Pensando em aprofundar a exploração do tema, vamos avaliar e refutar algumas das teorias criadas por apostadores, que tendem a acreditar que conseguem adivinhar os próximos eventos de ocorrência de um experimento aleatório analisando os eventos já ocorridos em tais experimentos. Neste contexto, estudaremos as heurísticas e vieses enunciados por Kahneman e Tversky, para tentar compreender os erros ocorridos nas análises dos jogos.

Por fim, vamos analisar a mega-sena com a intenção de caracterizar seu comportamento na lei dos grandes números. Farei uma análise estatística e gráfica dos resultados já ocorridos e assim poderemos concluir que os sorteios aleatórios não são passíveis de adivinhação.

Este estudo tem como objetivo anunciar aos incrédulos que tais jogos são de grande dificuldade de serem ganhos e que probabilisticamente podemos ter certeza que como todo jogo deste tipo a banca nunca perde. Mas não vamos entrar no mérito de a banca não perder e vamos nos concentrar na possibilidade que cada jogador tem de ganhar.

Na ânsia de ganhar muitos apostadores buscam estabelecer um padrão de sorteio dos resultados para conseguir adivinhar quais os próximos números a serem sorteados. Vamos aproveitar essas teses criadas por jogadores anônimos e que certamente já rondou a cabeça de muitos de nós para explicar onde se encontram os erros de interpretação dos resultados dos sorteios aleatórios.

1 **PROBABILIDADE**

1.1 **EXPERIMENTO ALEATÓRIO**

Definição 1. Chama-se **EXPERIMENTO ALEATÓRIO** àquele que, repetido em idênticas condições, produz resultados que não podem ser previstos com certeza, ou seja, cujo resultado é imprevisível. Entretanto, embora não saibamos qual o resultado irá ocorrer num experimento, em geral conseguimos descrever o conjunto de todos os resultados possíveis que podem ocorrer.

As variações de resultados, de experimento para experimento são devidas a uma multiplicidade de causas que não podemos controlar, as quais denominamos de acaso. (HAZZAN, 2004). Em oposição aos fenômenos aleatórios, existem os fenômenos determinísticos, que são aqueles cujos resultados são previsíveis, ou seja, temos certeza dos resultados a serem obtidos.

Definição 2. O conjunto de resultados possíveis de um determinado experimento é denominado **ESPAÇO AMOSTRAL**. Qualquer subconjunto desse espaço amostral é denominado **EVENTO**. Se este subconjunto possuir apenas um elemento, nós o denominamos **EVENTO ELEMENTAR**. Se não possuir nenhum elemento, nós o denominamos **EVENTO IMPOSSÍVEL**. Se coincidir com o espaço amostral, nós o denominamos **EVENTO CERTO.**

Para exemplificar os conceitos acima, imagine o lançamento de um dado e a observação da face superior. Neste exemplo o espaço amostral pode ser determinado, porém não há como ter certeza do resultado que irá acontecer. Por isso, este experimento é um experimento aleatório, os elementos de seu espaço amostral são os números de 1, 2, 3, 4, 5 e 6, e seus eventos possíveis como resultado é qualquer dos números de suas faces ou subconjuntos por eles formados. Podemos representar o espaço amostral como $U = \{1, 2, 3, 4, 5, 6\}$.

Exemplos de eventos no espaço amostral U:

A: sair número maior do que 4: $A = \{5, 6\}$

B: sair um número primo e par: $B = \{2\}$

C: sair um número ímpar: $C = \{1, 3, 5\}$.

No lançamento do dado acima, supõe-se que, sendo o dado perfeito, as chances de sair qualquer número de 1 a 6 são iguais. Estamos tratando então de um espaço equiprovável.

1.2 CONCEITO DE PROBABILIDADE

Para facilitar o entendimento dos resultados de determinado experimento podemos expressar as ocorrências de cada evento relativamente ao conjunto de todos os testes possíveis feitos em um dado experimento. Esse conceito faz com que tenhamos uma ideia de proporcionalidade dentre os diferentes eventos que um experimento pode apresentar.

Definição 3. Em um determinado experimento aleatório, chamamos de **FREQUÊNCIA RELATIVA** a razão entre a quantidade de ocorrências de um dado evento pelo total de repetições do experimento. (HAZZAN, 2004)

Definição 4. A **PROBABILIDADE** é definida como um número associado a cada evento de um experimento, de modo que ele tenha as mesmas características da frequência relativa. (HAZZAN, 2004).

Consideremos uma urna que contenha 49 bolas azuis e 1 bola branca. Para uma retirada, teremos duas possibilidades: bola azul ou bola branca. Percebemos, entretanto que será muito mais frequente obtermos numa retirada, uma bola azul, resultando daí, podermos afirmar que o evento “sair bola azul” tem maior probabilidade de ocorrer, do que o evento “sair bola branca”.

Definição 5: Dado um espaço amostral U, diremos que a distribuição de probabilidades em U é equiprovável se todos os eventos elementares de U tiverem a mesma probabilidade de ocorrer. Em geral, características do experimento é que nos levam a supor uma distribuição equiprovável. (HAZZAN, 2004).

Se um dado tem forma perfeita, ou seja, não é viciado, então a chance de ocorrer qualquer um dos seis eventos possíveis é igual. Sendo assim, dizemos que a probabilidade de acontecer qualquer dos eventos é igual, e daí, dizemos que o experimento de lançar um dado não viciado possui um espaço amostral **EQUIPROVÁVEL.**

Definição 6. Seja U um espaço amostral finito e equiprovável e A um determinado evento, ou seja, um subconjunto de U. A probabilidade $P(A)$ de ocorrência do evento A será calculada pela fórmula:

$$P(A) = \frac{n(A)}{n(U)}$$

com:

$n(A)$ = número de elementos de A; e

$n(U)$ = número de elementos do espaço amostral U.

Vamos utilizar a fórmula simples acima, para resolver os seguintes exercícios introdutórios:

Exemplo 1. Considere o lançamento de um dado. Calcule a probabilidade de:

a) sair o número 3:

Temos $U = \{1, 2, 3, 4, 5, 6\}$, logo, $n(U) = 6$; e

$A = \{3\}$, logo, $n(A) = 1$. Portanto, a probabilidade procurada será:

$$P(A) = \frac{1}{6}$$

b) sair um número par:

Agora, o evento A é o conjunto $A = \{2, 4, 6\}$ com 3 elementos; logo a probabilidade procurada será:

$P(A) = \frac{3}{6} = \frac{1}{2}$

c) sair um múltiplo de 3:

Dentre os elementos do conjunto universo, apenas os elementos 3 e 6 são múltiplos de 3, logo o conjunto evento A será $A = \{3, 6\}$ com 2 elementos; logo a probabilidade procurada será:

$P(A) = \frac{2}{6} = \frac{1}{3}$

d) sair um número menor do que 3:

O conjunto evento será $A = \{1, 2\}$ com dois elementos. Portanto:

$P(A) = \frac{2}{6} = \frac{1}{3}$

e) sair um quadrado perfeito:

O conjunto evento $A = \{1,4\}$ com dois elementos. Portanto:

$P(A) = \frac{2}{6} = \frac{1}{3}$

Exemplo 2. Considere o lançamento de dois dados. Calcule a probabilidade de:

a) sair a soma 8:

Observe que neste caso, o espaço amostral U é constituído pelos pares ordenados (i, j), onde i = número no dado 1 e j = número no dado 2.

É evidente que teremos 36 pares ordenados possíveis do tipo (i, j) onde $i = 1, 2, 3, 4, 5$ ou 6, o mesmo ocorrendo com j.

As somas iguais a 8, ocorrerão nos casos: $(2,6), (3,5), (4,4), (5,3)$ e $(6,2)$. Portanto, o evento "soma igual a 8" possui 5 elementos. Logo, a probabilidade procurada será igual a $P(A) = \frac{5}{36}$

b) sair a soma 12

Neste caso, a única possibilidade é o par $(6,6)$. Portanto, a probabilidade procurada será igual a $P(A) = \frac{1}{36}$.

Exemplo 3. Uma urna possui 6 bolas azuis, 10 bolas vermelhas e 4 bolas amarelas. Tirando-se uma bola com reposição, calcule as probabilidades seguintes: sair bola azul; sair bola vermelha e sair bola amarela.

Para iniciar esse problema, é interessante ressaltar a importância da informação dada no enunciado, em que o texto trata de um sorteio com reposição. Nesse sorteio, então, a bola retirada ao acaso em um dado sorteio é reposta, não alterando, assim, o espaço amostral para o próximo. Vamos aos casos.

a) sair bola azul

$P(A) = \frac{6}{20} = \frac{3}{10} = 0{,}30 = 30\%$

b) sair bola vermelha

$P(A) = \frac{10}{20} = \frac{1}{2} = 0{,}50 = 50\%$

c) sair bola amarela

$P(A) = \frac{4}{20} = \frac{1}{5} = 0{,}20 = 20\%$

Vemos no exemplo acima, que as probabilidades podem ser expressas como porcentagem. Esta forma é conveniente, pois permite a estimativa do número de ocorrências para um número elevado de experimentos. Por exemplo, se o experimento acima for repetido diversas vezes, podemos afirmar que em aproximadamente 30% dos casos, sairá bola azul, 50% dos casos sairá bola vermelha e 20% dos casos sairá bola amarela. Quanto maior a quantidade de experimentos, tanto mais a distribuição do número de ocorrências se aproximará dos percentuais indicados. Essa é uma consequência da Lei dos Grandes Números que será enunciada mais adiante neste trabalho.

Para melhor entender estes conceitos de probabilidade, vamos analisar alguns exemplos dados pelo autor do livro O andar do bêbado, Leonard Mlodinow. Este livro traz alguns exemplos do cotidiano em que a probabilidade pode nos ajudar. O exemplo abaixo traz uma boa noção de espaço amostral equiprovável e do conceito de probabilidade.

Problema 1: (PROBABILIDADE DOS GÊMEOS):

> "Uma amiga está grávida de gêmeos que sabe serem fraternos. Qual é a chance de que ao menos um dos bebês seja menina? " (MLODINOW, 2009).

Poderemos ter qualquer um dos resultados abaixo:

Menino e menino

Menina e menino

Menino e menina

Menina e menina

Consideramos então que o espaço amostral deste exemplo tem 4 elementos, em que em três deles ao menos um bebê é uma menina. Ou seja, o subconjunto evento tem 3 elementos.

Supomos que há a mesma probabilidade de ocorrer menina ou menino. E que os gêmeos fraternos são formados por dois zigotos diferentes. Ou seja, são formados independentemente um do outro. Sendo assim, a probabilidade de ao menos um dos bebês ser menina é de $3/4$. Ou seja, probabilidade de 75% de chance de ocorrer ao menos uma menina.

Outro problema que traz a noção de espaço amostral e como o desconhecimento do mesmo pode trazer erros de interpretação dos experimentos está no exemplo abaixo. O matemático D'Alembert autor de vários trabalhos sobre probabilidade cometeu um erro de avaliação natureza do espaço amostral. Isso mostra que não é difícil de cometermos esse erro, portanto, devemos ter muita atenção ao analisar os problemas em probabilidade.

Problema 2 (**O ERRO DE D'ALEMBERT**):

> Não muito tempo atrás, no século XVIII, o famoso matemático francês Jean Le Rond d'Alembert, autor de vários trabalhos sobre probabilidades, usou erroneamente o conceito ao analisar o lançamento de duas moedas. O número de caras obtido nos dois lançamentos pode ser 0, 1 ou 2. Como existem três resultados, raciocinou D'Alembert, a chance de cada um deve ser de $1/3$. Mas D'Alembert se enganou. (MLODINOW, 2009).

D'Alembert cometeu o erro de não perceber que associou ao evento um espaço amostral não equiprovável e ao mesmo tempo utilizou a Definição 6 para calcular a probabilidade do evento de interesse.

Ao analisar o espaço amostral como $\{0,1,2\}$ ele não considerou a frequência de ocorrência de cada elemento do espaço. Pois, analisando caso a caso, cada uma das moedas pode cair cara ou coroa o que faz com que existam dois casos com resultados iguais e dois casos com resultados diferentes.

Sendo assim, os resultados possíveis para o lançamento de duas moedas são {cara, cara}, {cara, coroa}, {coroa, cara} e {coroa, coroa}. Este é um espaço amostral equiprovável associado a esse experimento e, pela Definição 6, podemos concluir que existe apenas um resultado dos quatro possíveis no qual não se encontra cara, dois resultados com apenas uma cara e um resultado com duas caras.

Ou seja, ao contrário do que afirmou d'Alembert, a probabilidade de sair zero caras é de $1/4$, de sair uma cara é de $2/4$, ou $1/2$, e de sair duas caras é de $1/4$.

1.3 TEOREMAS SOBRE PROBABILIDADE EM ESPAÇO AMOSTRAL FINITO

Teorema 1. A probabilidade do evento impossível é nula.

Com efeito, sendo o evento impossível o conjunto vazio (Ø), teremos:

$P(\emptyset) = \frac{n(\emptyset)}{n(U)} = \frac{0}{n(U)} = 0$

Por exemplo, se numa urna só existem bolas brancas, a probabilidade de se retirar uma bola verde (evento impossível, neste caso) é nula.

Teorema 2. A probabilidade do evento certo é igual à 1.

Com efeito:

$P(\emptyset) = \frac{n(U)}{n(U)} = 1$

Por exemplo, se numa urna só existem bolas vermelhas, a probabilidade de se retirar uma bola vermelha (evento certo, neste caso) é igual a 1.

Teorema 3. Se um evento A está contido em um evento B, então $P(A) \leq P(B)$

Se A está contido em B.

Então, $n(A) \leq n(B)$.

Como $P(A) = \frac{n(A)}{n(U)}$ e $P(B) = \frac{n(B)}{n(U)}$

Conclui-se que $P(A) \leq P(B)$

Teorema 4. Se A é um evento, então $0 \leq P(A) \leq 1$

Decorre diretamente da definição e do teorema anterior, pois se A é evento de um experimento, então o conjunto vazio é seu subconjunto e A é subconjunto do espaço amostral, ou seja, A contém o evento nulo e está contido no evento certo.

Portanto, $P(0) \leq P(A) \leq P(U)$, isto é, $0 \leq P(A) \leq 1$

Teorema 5. A soma das probabilidades de um evento e do seu evento complementar é igual à unidade.

Seja o evento A e o seu complementar A'. Sabemos que $A \, U \, A' = U$. Logo, $n(A \, U \, A') = n(U)$ e, como a união disjunta, temos $n(A) + n(A') = n(U)$.

Dividindo ambos os membros por $n(U)$, vem:

$\frac{n(A)}{n(U)} + \frac{n(A')}{n(U)} = \frac{n(U)}{n(U)}$, de onde conclui-se: $P(A) + P(A') = 1$.

Este teorema é muito importante pois facilita a solução de muitos problemas aparentemente complicados. Em muitos casos, é mais fácil calcular a probabilidade do evento complementar e, pela propriedade acima, fica fácil determinar a probabilidade do evento.

Conforme mostrado nos exemplos anteriores, é possível aplicar esses teoremas em problemas reais que tratam do acaso. Outro problema interessante exposto no livro é o chamado problema de Monty Hall. Vamos a ele:

Problema 3 (O PROBLEMA DE MONTY HALL):

> Suponha que os participantes de um programa de auditório recebam a opção de escolher uma dentre três portas: atrás de uma delas há um carro; atrás das outras, há cabras. Depois que um dos participantes escolhe uma porta, o apresentador, que sabe o que há atrás de cada porta, abre uma das portas não escolhidas, revelando uma cabra. Ele diz então ao participante: “Você gostaria de mudar sua escolha para a outra porta fechada?” Para o participante é vantajoso trocar sua escolha inicial? A pergunta foi inspirada nas regras do programa de televisão *Let's Make a Deal*, transmitido de 1963 a 1976 e relançado diversas vezes entre 1980 e 1991. O principal atrativo do programa era o belo e simpático apresentador, Monty Hall, e sua assistente de roupas provocativas, Carol Merril, Miss Azusa (Califórnia) de 1957.” (MLODINOW, 2009).

Apesar de à primeira vista se supor que a troca ou não de escolha após a revelação de uma das cabras não trará nenhum benefício, uma análise do problema mais a fundo pode revelar o contrário.

Quando o participante faz sua primeira escolha ele tem que escolher 1 porta de 3. Evidentemente, o seu objetivo é acertar a porta em que está escondido o carro. Nesse caso, ele possui aproximadamente 33% (1 porta em 3) de chance de acertar a porta com o carro e 66% (2 portas em 3) de chance de errar.

Ao revelar uma cabra em uma das duas portas não escolhidas pelo participante, o apresentador traz algumas certezas para o jogo que muda suas probabilidades. Pois, se o participante havia escolhido a porta correta (33%) e muda de porta, ele perde. Porém, se o participante havia escolhido uma das portas erradas (66%) e muda de porta, ele ganha, já que a única porta não escolhida é a porta certa.

Veja bem, mudando de porta ele muda sua probabilidade de vitória para 66%, enquanto que ficando com a porta escolhida anteriormente ele manteria seus 33% de chance de acertar a porta com o carro.

Conclui-se, então, que para aumentar suas probabilidades de vitória, o participante deveria mudar de porta após a revelação de uma das cabras pelo apresentador.

Durante a troca das portas, o jogador poderia utilizar o teorema apresentado acima para facilitar suas escolhas. Obviamente, se temos três portas para escolher uma podemos dividir as portas entre escolhas com cabras e com carro. Sendo assim, o conjunto de portas com cabras é o complementar de portas com carros, visto que não há outra possibilidade no espaço amostral de portas apresentadas.

Esse recurso pode ser usado da seguinte maneira. Se eu tenho 1 porta em 3 com carro, terei 33% de chance de acertar a porta do carro. Com essa probabilidade, e mesmo sem calcular a chance de errar saberemos que temos mais chance de errar pois o complementar de 33% será $100\% - 33\% = 66\%$, aproximadamente.

Logo, com a abertura da porta e conforme mostrado acima, temos maior chance de acertar o carro trocando de porta.

Evidentemente, esse exemplo é simples para tal aplicação, mas pode ser usado em outros exemplos maiores. Imagine um jogo parecido com novecentas portas e após a escolha as portas são reveladas até sobrarem apenas duas. Dá para usar a mesma ideia, sem problemas e com toda certeza o leitor chegará a conclusão de que ainda assim vale mais trocar de porta.

Teorema 6. Sendo $A\ e\ B$ dois eventos, então $P(AUB) = P(A) + P(B) - P(A \cap B)$

Observe que se $A \cap B = \emptyset$ (ou seja, a interseção entre os conjuntos A e B é o conjunto vazio), então $P(A\ U\ B) = P(A) + P(B)$.

Com efeito, já sabemos da Teoria dos Conjuntos que $n(A\ U\ B) = n(A) + n(B) - n(A \cap B)$

Dividindo ambos os membros por $n(U)$ e aplicando a definição de probabilidade, concluímos rapidamente a veracidade da fórmula acima.

Para melhor ilustrar como pode ser usado o teorema apresentado acima, suponha que em uma certa comunidade existem dois jornais J e P. Sabe-se que

5.000 pessoas são assinantes do jornal J, 4.000 são assinantes de P, 1.200 são assinantes de ambos e 800 não leem jornal. Qual a probabilidade de que uma pessoa escolhida ao acaso seja assinante de ambos os jornais?

Precisamos calcular o número de pessoas do conjunto universo, ou seja, o espaço amostral.

Teremos:

$n(U) \ = \ n(J \ U \ P) \ + \ N.$º de pessoas que não leem jornais.

$n(U) = n(J) + N(P) – N(J \cap P) + 800$

$n(U) = 5.000 + 4.000 – 1.200 + 800$

$n(U) = 8.600$

Portanto, a probabilidade procurada será igual a:

$$P = \frac{1.200}{8.600} = \frac{6}{43} = 0{,}1395$$

Logo, $P \ = \ 0{,}1395 \ = \ 13{,}95\%$.

A interpretação do resultado é a seguinte: escolhendo-se ao acaso uma pessoa da comunidade, a probabilidade de que ela seja assinante de ambos os jornais é de aproximadamente 14% (contra cerca de 86% de probabilidade de não ser).

1.4 PROBABILIDADE CONDICIONAL

Considere que desejamos calcular a probabilidade da ocorrência de um evento A, sabendo-se de antemão que ocorreu um evento B.

Seja U um espaço amostral e consideremos dois eventos, A e B. Com o símbolo $P(A|B)$ indicamos a probabilidade do evento A, dado que o evento B ocorreu, isto é, $P(A|B)$ é a probabilidade condicional do evento A, uma vez que B tenha ocorrido. Quando calculamos $P(A|B)$, tudo se passa como se B fosse o novo espaço amostral "reduzido" dentro do qual queremos calcular a probabilidade de A. (HAZZAN, 2004)

Teremos então:

$P(A|B) = \frac{n(A \cap B)}{n(B)}$, onde $A \cap B$ = interseção dos conjuntos A e B e desde que $n(B) \neq 0$.

Esta fórmula é importante, mas pode ser melhorada.

Vejamos:

Ora, a expressão acima, pode ser escrita sem nenhum prejuízo da elegância, nem do rigor, como:

$$P(A|B) = \frac{n(A \cap B)}{n(U)} . \frac{n(U)}{n(B)} = \frac{P(A \cap B)}{P(B)}$$

Vem, então: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, de onde concluímos finalmente:

$$P(A \cap B) = P(A|B).P(B) \quad (1)$$

Esta fórmula é denominada **Lei das Probabilidades Compostas**. Esta importante fórmula, permite calcular a probabilidade da ocorrência simultânea dos eventos A e B, sabendo-se que já ocorreu o evento B.

Definição 7: São ditos **EVENTOS INDEPENDENTES** aqueles em que a ocorrência do evento B, não muda a probabilidade da ocorrência do evento A. Sendo assim, $P(A|B) = P(A)$ e, neste caso a fórmula (1) pode ser reescrita como:

$$P(A \cap B) = P(A) . P(B)$$

Podemos então afirmar, que a probabilidade de ocorrência simultânea de eventos independentes, é igual ao produto das probabilidades dos eventos considerados.

Para exemplificar a propriedade enunciada acima tomemos uma urna que possui cinco bolas vermelhas e duas bolas brancas. Nesta condição, vamos calcular as probabilidades de:

a) em duas retiradas, sem reposição da primeira bola retirada, sair uma bola vermelha (V) e depois uma bola branca (B).

$P(V \cap B) = P(V).P(B|V)$

$P(V) = \frac{5}{7}$ (5 bolas vermelhas de um total de 7).

Supondo que saiu bola vermelha na primeira retirada, ficaram 6 bolas na urna.

Logo:

$P(B|V) = 2/6 = 1/3$

Da lei das probabilidades compostas, vem finalmente que:

$$P(V \cap B) = \frac{5}{7}.\frac{1}{3} = \frac{5}{21} = 0{,}2380 = 23{,}8\%$$

b) em duas retiradas, com reposição da primeira bola retirada, sair uma bola vermelha e depois uma bola branca.

Com a reposição da primeira bola retirada, os eventos ficam independentes. Neste caso, a probabilidade buscada poderá ser calculada como:

$$P(V \cap B) = P(V).P(B) = \frac{5}{7}.\frac{2}{7} = \frac{10}{49} = 0{,}2041 = 20{,}41\% .$$

Mlodinow trouxe um exemplo claro do uso da multiplicação de probabilidades em seu livro *O Andar do Bêbado*. Vamos a ele:

PROBLEMA 4 (SINCERIDADE E BELEZA)

> Suponha que você faça um baralho com as fotografias dos 100 sujeitos que conheceu até agora num serviço de relacionamentos na internet – homens que, pelas fotos apresentadas nas páginas, se parecem com Tom Cruise, mas que pessoalmente lembram mais Danny DeVito. Suponha também que, no verso de cada carta, você liste algumas informações sobre cada homem, como sincero (sim ou não) e bonito (sim ou não). Por fim, suponha que 1 de cada 10 de suas possíveis almas gêmeas tenha um “sim” para cada uma das categorias. Quantos homens dentre os 100 do seu baralho passarão no teste em ambas as categorias? (MLODINOW, 2009).

Considerando, primeiramente, a condição sinceridade, em que um de cada dez dos cem homens tem um sim no verso do cartão, podemos afirmar que dez deles terão sim no verso. Da mesma forma, um a cada dez homens responderão sim para o critério beleza. Avaliando os critérios de escolha apresentados, podemos dizer que esses são critérios que não dependem um do outro para acontecer. Desta forma, podemos dizer que a ocorrência dos eventos beleza e sinceridade são independentes. Pelo exposto anteriormente, quando nos deparamos com o cálculo de probabilidades simultâneas de eventos independentes podemos usar a multiplicação de probabilidades.

Se um de cada dez homens são sinceros, então a probabilidade de encontrar um homem sincero é de $\frac{1}{10}$. O mesmo acontece para beleza, se um a cada dez é bonito, então a probabilidade de um homem escolhido ser bonito é de $\frac{1}{10}$. Como o objetivo é escolher um homem sincero e bonito, podemos calcular através da multiplicação das probabilidades:

$$P = \frac{1}{10} \cdot \frac{1}{10} = \frac{1}{100} = 1\%$$

Ou seja, existe a probabilidade de um a cada cem de um dos homens ser sincero e bonito.

Neste exemplo hipotético de Mlodinow podemos ver a aplicação direta do teorema das multiplicações de probabilidades que vai nos ajudar a calcular, mais a frente, algumas das probabilidades dos sorteios aleatórios estudados neste trabalho.

Da mesma forma, quando tratamos de um experimento em que os eventos não são independentes, não podemos usar a multiplicação direta das probabilidades. Para o caso das probabilidades condicionais está claro que não há que se fazer a multiplicação. Mlodinow trouxe um exemplo bem característico para esse tema. Nele, ao invés de multiplicar as probabilidades de dois eventos, devemos considera-los como um só. Vamos a ele:

PROBLEMA 5 (PASSAGEIROS DE UMA CIA AÉREA):

> Suponha que uma companhia aérea tenha 1 lugar restante num voo, e ainda restem 2 passageiros por chegar. Suponha que, a partir da experiência, a companhia saiba que existe uma chance de $2/3$ de que um passageiro que reservou um voo se apresente para viajar. Utilizando a regra da multiplicação, a funcionária da companhia poderá concluir que existe uma chance de $2/3 \times 2/3$, ou cerca de 44%, de que ela tenha que lidar com um cliente insatisfeito. A chance de que nenhum cliente apareça e de que o avião tenha que voar com um lugar vago, por outro lado, é de $1/3 \times 1/3$, ou apenas

> 11%, aproximadamente. Mas isso só ocorre se presumirmos que os passageiros são independentes. Se, por exemplo, eles estiverem viajando juntos, a análise acima estará errada. A chance de que ambos queiram viajar é de $2/3$, igual à de um passageiro independente. É importante lembrar que só calculamos a probabilidade combinada a partir das probabilidades simples por meio da multiplicação quando os eventos não têm nenhuma relação entre si. (MLODINOW, 2009).

Conforme visto anteriormente, a multiplicação da probabilidade de dois eventos só é possível quando eles são independentes. Sendo assim, não cabe, neste caso, pensarmos em multiplicação das probabilidades. Como o exemplo acima trata de pessoas viajando juntas, então, podemos considerar eventos que se relacionam entre si da seguinte forma: $A \subseteq B$ e $B \subseteq A$. Ou seja, uma pessoa não viaja sem a outra. Sendo assim, o não comparecimento de qualquer das pessoas implica também no não comparecimento da outra. Sendo assim, cabe analisarmos este evento de duas pessoas como sendo o evento de apenas uma pessoa.

Durante o texto do livro, o autor, Mlodinow, trouxe mais um exemplo para a utilização dos teoremas apresentados neste capítulo. O teorema que diz que a soma de todas as probabilidades de um determinado experimento é igual a 1. Vamos a ele:

PROBLEMA 6 (SOMA DAS PROBABILIDADES):

> De volta à companhia aérea: em que momento a funcionária deveria somar as probabilidades em vez de multiplicá-las? Suponha que ela quisesse descobrir a probabilidade de que os dois passageiros se apresentem para viajar ou de que nenhum deles o faça. Neste caso, ela deve somar as probabilidades individuais dos dois eventos, o que, segundo o que calculamos acima, geraria uma probabilidade de 55%. (MLODINOW, 2009).

Para comprovar o que foi apresentado, podemos começar calculando a ocorrência da negativa do que se pretende. Portanto, calcularemos inicialmente a probabilidade de que apenas uma das pessoas apareça para o voo. Ou seja, nomeando os passageiros como A e B, teremos a ocorrência de A aparecendo para o vôo (probabilidade de $2/3$) e B não aparecendo para o vôo (probabilidade de $1/3$). Sendo então esta probabilidade de:

$$\frac{2}{3} \times \frac{1}{3} = \frac{2}{9}$$

O mesmo ocorrendo para o caso em que A não aparece para o vôo e B aparece para o vôo (probabilidade de $2/9$).

Utilizar a negativa é subtrair do evento certo o que não se pede. Sendo assim, temos que calcular a probabilidade de que os dois passageiros apareçam para o voo ou que nenhum deles apareça e simplesmente realizar a subtração do evento certo. Ou seja,

$$P = 1 - \frac{2}{9} - \frac{2}{9}$$

Portanto, a probabilidade de os dois passageiros aparecerem ou nenhum deles aparecer é de $5/9$, ou melhor, aproximadamente 55%.

Repare que todos os eventos possíveis desse experimento são:

1- Os dois passageiros aparecem para o voo: $P = 4/9$

2- Os dois passageiros não aparecem para o voo: $P = 1/9$

3- Apenas o passageiro A aparece para o voo: $P = 2/9$

4- Apenas o passageiro B aparece para o voo: $P = 2/9$

Sendo esses todos os eventos possíveis, sua soma deve ser de 1, ou seja, o evento certo.

$$\frac{4}{9} + \frac{1}{9} + \frac{2}{9} + \frac{2}{9} = \frac{9}{9} = 1$$

Comprovando o apresentado.

Utilizando o mesmo método de pensamento do exemplo acima, e ainda a multiplicação de probabilidades, podemos ajudar os especialistas em exames de DNA. Veja no próximo problema apresentado por Mlodinow (2009) que nem todo teste pode ser trazido à certeza extrema e sempre é importante realizar um exame comprobatório.

PROBLEMA 7 (TAXA DE ERRO NO EXAME DE DNA):

> Não é raro que especialistas em análise de DNA testemunhem em julgamentos nos quais uma amostra de DNA retirada da cena de um crime corresponde ao DNA extraído de um suspeito. Quanta certeza podemos ter quanto ao resultado de tais testes? Quando os exames de DNA passaram a ser usados em tribunais, diversos especialistas afirmaram que estes jamais

> poderiam gerar resultados falsos positivos. Hoje, tais especialistas afirmam regularmente que a chance de que uma pessoa aleatória tenha DNA semelhante ao de uma amostra retirada da cena de um crime é menor que **1 em 1 milhão ou 1 em 1 bilhão**. Com essa probabilidade, seria difícil culparmos um jurado por pensar: “Prendam e joguem a chave fora.” Mas existe outra estatística, pouco apresentada ao júri, ligada ao fato de que os laboratórios cometem erros, por exemplo, ao obterem e manusearem amostras, misturando-as ou trocando-as acidentalmente, ou então interpretando e relatando os resultados de maneira incorreta. Todos esses erros são raros, mas não tão raros quanto o caso da pessoa aleatória que possui DNA semelhante ao da cena do crime. O Laboratório de Criminologia da Cidade da Filadélfia, por exemplo, admitiu ter trocado a amostra de referência do réu e da vítima num caso de estupro, e uma companhia de exames chamada Cellmark Diagnostics confessou ter cometido erro semelhante. Infelizmente, a força das estatísticas ligadas ao DNA apresentado na corte é tal que um tribunal de Oklahoma condenou um homem chamado Timothy Durham a mais de 3.100 anos de prisão, muito embora 11 testemunhas afirmassem que ele estava em outro estado no momento do crime. No fim das contas, verificou-se que a análise inicial do laboratório não fora capaz de separar completamente o DNA do estuprador e o da vítima no líquido que testaram, e a combinação do DNA da vítima e do estuprador gerou um resultado positivo quando comparado com o de Durham. Um novo teste realizado posteriormente revelou o erro, e Durham foi libertado depois de passar quatro anos na prisão. As estimativas da taxa de erros por falhas humanas variam, mas muitos especialistas consideram que seja algo em torno de 1%. No entanto, como as taxas de erro de muitos laboratórios nunca foram medidas, as cortes muitas vezes não permitem o testemunho quanto a essa estatística geral. E mesmo que as cortes permitissem o testemunho relacionado aos falsos positivos, de que maneira os jurados o avaliariam? A maioria dos jurados presume que, dados dois tipos de erro – a semelhança acidental de 1 em 1 bilhão e o erro humano no laboratório, de 1 em 100 –, a taxa de erro geral deve se situar em algum lugar no meio do caminho, digamos, 1 em 500 milhões, o que, para a maior parte dos jurados, ainda não constitui uma dúvida razoável. Porém, empregando as leis da probabilidade, encontramos uma resposta muito diferente. (MLODINOW, 2009).

Para que se possa afirmar que alguém é culpado é necessário que os dois erros sejam descartados. Pois, então, se a possibilidade de erro é de 1 em 1 bilhão, teríamos $\frac{999999999}{1000000000}$ chance de acerto. E o caso de erro humano sendo de 1 em 100, teríamos $\frac{99}{100}$ de acerto. Sendo necessário que aconteçam os dois acertos mutuamente, podemos multiplicar essas probabilidades para chegar na probabilidade de acerto.

$$\frac{999999999}{1000000000} \times \frac{99}{100} \approx \frac{99}{100} = 99\%$$

Se a probabilidade de acerto é de aproximadamente 99% então a probabilidade de erro é de aproximadamente 1%. Portanto, um valor muito mais expressivo a ser considerado do que aquele em que se acredita.

Outro exemplo trazido por Mlodinow que vai exemplificar a importância de saber usar cálculos de probabilidades e o quanto o desconhecimento desses cálculos pode nos fazer cometer erros de escolhas durante a vida é o exemplo da loteria canadense. Segundo o texto, os organizadores cometeram um erro grande ao desconsiderar as probabilidades dos sorteios aleatórios. Vamos ver no problema abaixo que os teoremas acima são muito importantes e resolveriam esse problema para os canadenses.

PROBLEMA 8 (**A LOTERIA CANADENSE)**:

> Alguns anos atrás, os administradores da loteria canadense aprenderam, da pior maneira possível, a importância de se fazer uma contagem cuidadosa, quando tiveram que devolver um prêmio em dinheiro não reclamado que ficara acumulado. Compraram 500 automóveis como prêmios especiais e programaram um computador para determinar os vencedores, selecionando aleatoriamente 500 números de uma lista de 2,4 milhões de participantes. A loteria publicou a lista dos 500 números vencedores, prometendo um automóvel para cada número listado. Para seu embaraço, uma pessoa alegou (corretamente) que havia ganhado dois carros. Os administradores da loteria ficaram embasbacados – sorteando números de uma lista de mais de 2 milhões de participantes, como o computador poderia ter sorteado duas vezes o mesmo número? Haveria alguma falha no programa? (MLODINOW, 2009).

A pergunta deixada pelo autor relativa a uma possível falha no programa não pode ser respondida facilmente. Afinal, não podemos ter certeza, pois não está explícito no texto, quais eram as reais intenções dos realizadores do sorteio. O fato é que um sorteio totalmente aleatório pode trazer resultados iguais. Sendo assim, um resultado possível, apesar da baixíssima probabilidade de acontecer nesse sorteio, seria que uma mesma pessoa seja sorteada para ganhar os 500 carros.

Assim, a discussão de o programa estar certo ou errado passa por descobrir qual era a real intenção dos organizadores. Com esse pensamento, podemos estabelecer duas linhas de pensamento dos organizadores. A primeira, voltada para o sorteio totalmente aleatório no qual qualquer resultado é possível e aceitável. E uma segunda na qual os organizadores queriam sortear um carro por pessoa.

Caso a intenção seja o sorteio aleatório, sem restrições de resultados, então o programa cumpriu seu papel. Poderia até ocorrer a possibilidade de serem sorteados 500 pessoas diferentes para receber os 500 carros, entretanto, um dos sorteados recebeu dois carros. Nesse caso, o programa cumpriu bem o seu papel e não o que se falar em mudanças ou erros.

Porém, se a intenção dos organizadores era que cada carro sorteado fosse para uma pessoa diferente, o programa apresentado possui um erro.

Para que cada pessoa possa receber apenas um carro, seria necessário que o programador retirasse do sorteio seguinte os números sorteados anteriormente. Só assim, seria possível garantir que o fato relatado não ocorresse.

Neste caso, podemos intuir que os organizadores do sorteio cometeram um erro comum ao acreditar que um evento de baixa probabilidade para uma repetição do experimento mantém suas probabilidades para muitas repetições. Afinal, sorteando 500 participantes num universo de 2,4 milhões, qual seria a probabilidade de ocorrerem ao menos dois resultados iguais?

Para calcular essa probabilidade é necessário retirarmos do evento certo a probabilidade de ocorrerem 500 resultados diferentes, ou seja:

$$P=1-\frac{(2{,}4*10^6)(2{,}4*10^6-1)(2{,}4*10^6-2)...(2{,}4*10^6-499)}{(2{,}4*10^6)^{500}}$$

Para garantir que sejam diferentes, no primeiro fator qualquer resultado possível satisfaz. Portanto, a probabilidade de ocorrer um evento satisfatório é o evento certo, ou seja 1, que pode ser escrito como: $\frac{(2{,}4*10^6)}{(2{,}4*10^6)}$

Para o segundo fator, retiraremos uma unidade do numerador da fração que representa o inteiro, já que para os resultados satisfatórios não poderemos contar o evento que ocorreu no primeiro sorteio, portanto, o segundo fator será: $\frac{(2{,}4*10^6)-1}{(2{,}4*10^6)}$

Para o terceiro fator, retiraremos duas unidades e assim sucessivamente até o quingentésimo sorteio que retiraremos 499 unidades do numerador, relativos aos 499 sorteios já realizados.

Multiplicando todos esses fatores chegamos na probabilidade a ser calculada P acima.

Mas esse cálculo não precisa ser realizado com precisão, já que claramente, tende ao evento certo com o aumento das repetições. Porém, para melhorar a intuição do evento vamos calcular tal probabilidade para dois, três e quatro sorteios.

Em caso de apenas dois sorteios, qual a probabilidade de ocorrer ao menos dois resultados iguais?

$$P=1-\frac{(2{,}4*10^6)(2{,}4*10^6-1)}{(2{,}4*10^6)^2}=1-\frac{(2{,}4*10^6-1)}{(2{,}4*10^6)^2}=\frac{1}{2{,}4*10^6}\approx 0{,}000000417$$

Em caso de apenas três sorteios, qual a probabilidade de ocorrer ao menos dois resultados iguais?

$$P=1-\frac{(2{,}4*10^6)(2{,}4*10^6-1)(2{,}4*10^6-2)}{(2{,}4*10^6)^3}=0{,}0000012499$$

Em caso de apenas quatro sorteios, qual a probabilidade de ocorrer ao menos dois resultados iguais?

$$P=1-\frac{(2{,}4*10^6)(2{,}4*10^6-1)(2{,}4*10^6-2)(2{,}4*10^6-3)}{(2{,}4*10^6)^4}\approx 0{,}0000025$$

Claramente as probabilidades de ocorrência de ao menos 2 resultados iguais estão aumentando à medida que aumentamos a quantidade de sorteios, portanto, haverá maior chance de ocorrer resultados iguais repetindo esse sorteio 500 vezes.

Podemos fazer uma aproximação para o evento acima da seguinte maneira:

$$P=1-\frac{(2{,}4*10^6)(2{,}4*10^6-1)(2{,}4*10^6-2)...(2{,}4*10^6-500)}{(2{,}4*10^6)^{500}}$$

$$P=1-\frac{(2{,}4*10^6)}{(2{,}4*10^6)}*\frac{(2{,}4*10^6-1)}{(2{,}4*10^6)}*\frac{(2{,}4*10^6-2)}{(2{,}4*10^6)}...\frac{(2{,}4*10^6-500)}{(2{,}4*10^6)}$$

Sendo $\frac{(2{,}4*10^6-500)}{(2{,}4*10^6)}$ o menor dos fatores do produto apresentado, então:

$$P>1-\frac{(2{,}4*10^6-500)}{(2{,}4*10^6)}*\frac{(2{,}4*10^6-500)}{(2{,}4*10^6)}*\frac{(2{,}4*10^6-500)}{(2{,}4*10^6)}...\frac{(2{,}4*10^6-500)}{(2{,}4*10^6)}$$

$$P>1-\left(\frac{(2{,}4*10^6-500)}{(2{,}4*10^6)}\right)^{500}\approx 0{,}099$$

Concluímos, então que a probabilidade de ocorrer sorteio repetido é maior que $9{,}9\%$. Um percentual que chamaria a atenção dos organizadores do sorteio, caso esses tivessem calculado seus resultados.

Neste exemplo podemos ver o quão importante para cálculo de probabilidades pode ser a quantidade de vezes que um experimento acontece. Mesmo que a probabilidade de algo acontecer seja muito pequena em um experimento, a probabilidade de este evento raro ocorrer ao menos uma vez aumenta de acordo com o aumento das repetições deste mesmo experimento.

Durante os estudos vamos reparar que a quantidade de repetições de um determinado experimento é muito importante para tirarmos conclusões acerca de um determinado evento.

Jogando um dado não viciado apenas uma vez, temos a probabilidade de $1/6$ de sair o número 1. Jogando o dado duas vezes a probabilidade de sair o número 1 pelo menos uma vez aumenta para $11/36$. Essa chance de ocorrer o número 1 ao menos uma vez aumenta a cada vez que aumentamos a quantidade de repetições desse experimento.

Não podemos garantir a frequência, ou em que jogada acontecerá o número 1 pela primeira vez. As probabilidades calculam a chance de ocorrência, mas não a garante. Entretanto, podemos garantir que em uma quantidade muito grande de repetições desse experimento a frequência relativa de sair cada número tenderá para 1/6, sua probabilidade calculada. Essa garantia é dada pela Lei dos Grandes Números.

1.4 A LEI DOS GRANDES NÚMEROS

A lei dos grandes números (LGN) foi definida por Jacob Bernoulli na obra Arte da Conjectura (1713) da seguinte maneira: Se um evento de probabilidade P é observado repetidamente durante independentes repetições, a razão da frequência observada deste evento para o número total de repetições converge em direção a P conforme o número de repetições se torna arbitrariamente grande.

Durante o estudo da LGN podemos ver várias formas de defini-la. Em seu trabalho, Ismael Araújo Silva citou-a com a escrita de Boyer:

> Se p é a probabilidade de um evento, se m é o número de ocorrências do evento em n experiências, se ε é um número positivo arbitrariamente pequeno, e se P é a probabilidade de que a desigualdade $\left|\frac{m}{n} - p\right| < \varepsilon$ esteja satisfeita, então lim P = 1. (BOYER, 1996, p. 308)

Essa forma de expressar a definição da LGN é exatamente a mesma à enunciada por Bernoulli escrita de uma forma mais moderna, como dita pelo próprio.

De acordo com o texto de Carlos Fiolhais, a LGN pode, também, ser chamada de LEI FUNDAMENTAL DAS PROBABILIDADES ou TEOREMA DE BERNOULLI e ela permite ligar, com precisão, frequências relativas e probabilidades. (SILVA, 2002)

De fato, ao se tratar as probabilidades por suas frequências relativas em uma grande quantidade de repetições de um experimento apresentado estamos atuando com a visão frequentista da probabilidade.

Apesar de perfeitas, as definições apresentadas acima para a LGN podem ser difíceis de entender. Para melhor explicar e entender a Lei dos grandes números, podemos utilizar a definição trazida por Murilo Sampaio em seu trabalho “A lei dos grandes números na percolação multifuncional” . Nele, Sampaio a define como o princípio geral das ciências de observação, segundo a qual a frequência de determinados acontecimentos, observados em um grande número de casos análogos, tende a se estabilizar cada vez mais, à medida que aumenta o número de casos observados, aproximando-se do valor previsto pela teoria das probabilidades. (SAMPAIO, 2007)

Exemplificando o apresentado acima, pense em um experimento simples, no qual uma pessoa lança uma moeda para cima diversas vezes e anota os resultados: cara ou coroa. A Lei dos grandes números afirma que quanto mais vezes a pessoa repetir o experimento, mais próximo da probabilidade calculada estará a frequência relativa das suas anotações.

Ou seja, nas primeiras 10 jogadas, podem ocorrer 10 caras. Essa é uma das possibilidades do experimento. Dessa forma, em 10 jogadas a frequência relativa de caras estará em 100%. Entretanto, se nos próximos 10 sorteios ocorrerem números iguais de caras e coroas, a frequência relativa de caras cai para 75%. Se nos próximos 10 sorteios ocorrerem novamente números iguais de caras e coroas, a frequência relativa de caras cai para 66%. Repare que mesmo sem ocorrer número maior de coroas, a frequência relativa de caras vai caindo e tendendo a se estabilizar próximo a 50% que é a probabilidade calculada para este experimento.

Kahneman e Tversky afirmam que amostras muito grandes são altamente representativas da população da qual elas são retiradas. A estatística, no entanto, não afirma algo semelhante para amostras pequenas, ou seja, nem todos os segmentos de sequências obtidos do lançamento de uma moeda serão altamente representativos da confiabilidade da mesma. Aparentemente, afirmaram os autores, as pessoas esperam que algum desvio em uma direção será brevemente cancelado pelo correspondente desvio na direção oposta, mas os erros não são cancelados nas amostras e sim simplesmente diluídos. (KAHNEMAN E TVERSKY,1971)

Esta diluição será evidenciada mais à frente neste trabalho, quando analisarmos estatisticamente os números da mega-sena já sorteados. Vamos conseguir verificar a lei dos grandes números atuando neste sorteio aleatório.

2 PROBLEMAS DO LIVRO: FALÁCIAS DE INTERPRETAÇÃO

O conhecimento das teorias probabilísticas pode, de fato, nos trazer benefícios para entender melhor alguns casos da vida real. Ter a exata noção do que buscar e a probabilidade de cada evento acontecer pode ser determinante para ajudar a revelar todas as possibilidades a que estamos sujeitos. O autor do livro O andar do bêbado, Leonard Mlodinow, trouxe alguns problemas nos quais a falta de conhecimento das teorias probabilísticas fez com que as pessoas cometessem erros que poderiam ter consequências graves.

Um exame de HIV com uma certeza falsa, um homem preso injustamente, outro inocentado usando argumentos falsos, taxas de desemprego manipuladas da forma que mais convir e notas em vinhos que pode salvar ou falir uma marca. Todas essas situações foram expostas por Mlodinow para mostrar alguns dos possíveis erros de interpretação das probabilidades.

Neste capítulo, vamos analisar esses erros cometidos nos casos apresentados e interpretar o que poderia ser feito para evitar os erros a fim de aprender melhor e não os cometer mais à frente no curso de nossas vidas.

PROBLEMA 9 (EXAME DE HIV)

> Fui impulsivamente ao médico e fiz um exame de HIV. O resultado foi positivo. Embora, a princípio, eu estivesse chocado demais para interrogar meu médico sobre a probabilidade que citara, soube mais tarde que ele havia derivado a minha chance de 1/1 mil de estar saudável a partir da seguinte estatística: o exame de HIV gera um resultado positivo quando o sangue não está infectado pelo vírus da Aids em apenas 1 de cada 1 mil amostras de sangue. Isso pode soar exatamente como a informação que ele me passou, mas não é. Meu médico confundira a probabilidade de que o exame desse positivo se eu não fosse HIV-positivo com a probabilidade de que eu não fosse HIV-positivo se o exame desse positivo.)
>
> Para entender o erro do meu médico, vamos empregar o método de Bayes. O primeiro passo consiste em definirmos o espaço amostral. Poderíamos incluir todas as pessoas que já fizeram um exame de HIV, mas teremos maior precisão se utilizarmos mais algumas informações relevantes ao meu respeito, considerando apenas os homens brancos americanos, heterossexuais e não usuários de drogas intravenosas que já realizaram o exame (veremos mais adiante que diferença isso faz).

> Agora que sabemos a quem incluir no espaço amostral, vamos classificar os integrantes do espaço. Em vez de menino e menina, as classes relevantes neste caso são pessoas infectadas que tiveram um exame positivo (verdadeiros positivos), pessoas não infectadas que tiveram um exame positivo (falsos positivos), pessoas não infectadas que tiveram um exame negativo (verdadeiros negativos) e pessoas infectadas que tiveram um exame negativo (falsos negativos).
>
> Por fim, perguntamos: quantas pessoas existem em cada uma das classes? Consideremos uma população inicial de 10 mil. Podemos estimar, a partir das estatísticas dos Centros de Controle e Prevenção de Doenças dos Estados Unidos, que, em 1989, aproximadamente 1 de cada 10 mil homens brancos americanos, heterossexuais e não usuários de drogas testados estava infectado pelo HIV. Presumindo que a taxa de falsos negativos fosse próxima de 0, isso significa que aproximadamente 1 pessoa de cada 10 mil apresentaria um exame positivo em virtude da presença da infecção. Além disso, como a taxa de falsos positivos é de, como citou meu médico, $1/1\,mil$, haverá aproximadamente 10 outros casos de exames positivos em pessoas que não estão infectadas pelo HIV. Os outros 9.989 dos 10 mil homens do espaço amostral apresentarão exames negativos.
>
> Agora vamos podar o espaço amostral, de modo a incluir apenas os que tiveram exames positivos. Acabamos com 10 pessoas cujos exames foram falsos positivos e 1 cujo exame foi um verdadeiro positivo. Em outras palavras, apenas 1 de cada 11 pessoas que apresentam exames positivos está realmente infectada pelo HIV. Meu médico me informou que a probabilidade de que o exame estivesse errado – e eu, na verdade, estivesse saudável – era de $1/1\,mil$. Ele deveria ter dito: “Não se preocupe, existe uma chance de mais de $10/11$ de que você não esteja infectado. (MLODINOW, 2009)

Nessa explicação o autor aprofundou bastante. Entretanto para melhor entender o que foi apresentado faremos uma tabela representada conforme abaixo:

	POSITIVO	**NEGATIVO**
INFECTADO	Verdadeiro positivo	Falso negativo
NÃO INFECTADO	Falso positivo	Verdadeiro negativo

Tendo sido dito que 1 a cada 10 mil está infectado, que a taxa de falsos positivos é de 1 a cada mil e que a taxa de falsos negativos é próxima de zero, tomando uma população de 10 mil, podemos concluir que 1 estará infectado, os falsos negativos podemos considerar sendo zero e sabendo que a taxa de falso positivo é de um a cada mil, podemos concluir que nessa população 10 em 10 mil terão um resultado falso positivo. Em resumo:

Verdadeiro positivo: 1

Falso positivo: 10

Falso negativo: 0

Verdadeiro negativo: 9989

Total: 10.000

	POSITIVO	**NEGATIVO**
INFECTADO	1	0
NÃO INFECTADO	10	9.989
Total	**11**	**9.989**

Tendo em vista que o exame do autor deu positivo, então é necessário utilizar como espaço amostral apenas os exames positivos, ou seja, 11. Então a probabilidade de ele não estar infectado tendo sido diagnosticado positivo será de 10/11, ou seja, aproximadamente 90,9%.

O problema acima trazido por Mlodinow traz um erro simples de interpretação do espaço amostral, por parte do médico que causou um grande incômodo na vida e história do autor. Afinal, ser diagnosticado com um vírus incurável com a chance de 0,1% de erro é desesperador para a maioria das pessoas. Entretanto, como mostrado acima, caso o médico tivesse um pouco mais de experiência e habilidade matemática e estatística poderia ter evitado o desespero e pedido um exame de confirmação de pronto, assim que o resultado deu positivo.

Outro erro médico grave trazido por Mlodinow em seu livro O andar do bêbado está no problema abaixo. Os médicos alemães estimaram que 90% das mulheres sem sintomas estariam com câncer de mama após apresentar uma mamografia positiva. Vamos conferir mais esse absurdo estatístico.

PROBLEMA 10 (ERROS MÉDICOS)

> ***A TEORIA DE BAYES nos mostra que a probabilidade de que A ocorra se B ocorrer geralmente difere da probabilidade de que B ocorra se A ocorrer. Não levar esse fato em consideração é um erro comum na profissão médica.*** Por exemplo, em estudos feitos na Alemanha e nos Estados Unidos, pesquisadores pediram a médicos que estimassem a probabilidade de que uma mulher assintomática com idade entre 40 e 50 anos que apresentasse uma mamografia positiva realmente tivesse câncer, sabendo que 7% das mamografias mostram câncer quando ele não existe.
>
> Além disso, os médicos receberam a informação de que a incidência real da doença era de aproximadamente 0,8%, e que a taxa de falsos negativos era de aproximadamente 10%. Juntando todas as informações, podemos usar o método de Bayes para determinar que uma mamografia positiva representa a presença de câncer em apenas cerca de 9% dos casos. No grupo alemão, no entanto, um terço dos médicos concluiu que a probabilidade seria de 90%, e a mediana das estimativas foi de 70%. No grupo americano, 95 de cada 100 médicos estimaram que a probabilidade seria próxima de 75%.

Para solucionar o problema vamos utilizar uma tabela parecida com a que foi utilizada no problema anterior em que teremos:

Verdadeiro positivo (mamografia positiva em pessoas com câncer): 0,8%

Falso positivo (mamografia positiva em pessoas sem câncer): 7%

Verdadeiro negativo (mamografia negativa em pessoas sem câncer): 82,2%

Falso negativo (mamografia negativa em pessoas com câncer): 10%

	POSITIVO	**NEGATIVO**
COM CÂNCER	0,8%	10%
SEM CÂNCER	7%	82,2%
Total	7,8%	92,2%

Sendo assim, para calcular a probabilidade de que uma mulher assintomática estivesse realmente com câncer (evento A) tendo sido diagnosticada positivo (evento B) será de:

$$\frac{0{,}8}{7{,}8} = \frac{8}{78} \approx 0{,}102 = 10{,}2\%$$

Veja bem, o cálculo acima levou em consideração 7,8% como a frequência relativa do espaço amostral porque no problema temos a garantia de que o paciente foi diagnosticado positivo, sendo assim, não faria sentido manter no espaço amostral qualquer paciente que tenha sido diagnosticado negativo.

Como podemos concluir, na verdade, a chance de uma mulher assintomática, com a mamografia positiva ter câncer é de cerca de 10%. Quando os médicos concluem que a chance de a mamografia estar certa é de 90% eles podem trazer desespero às pacientes, afinal, isso significa a quase certeza de o exame corresponder com a realidade. Entretanto, com o cálculo correto, vimos que para garantir que a paciente tem realmente câncer de mama, o médico, certamente, precisará realizar um exame mais preciso, pois a mamografia só lhe dá certeza de cerca de 10% em casos de positivo.

Não é só nos hospitais que o desconhecimento das teorias de probabilidades pode causar transtornos. Nos tribunais, essas falácias podem gerar erros catastróficos para a sociedade. Afinal, prender um inocente ou soltar um assassino são falhas gravíssimas no processo judicial. Vamos ver no que a probabilidade ajudaria no caso abaixo.

PROBLEMA 11 (FALÁCIA DA ACUSAÇÃO)

> Nos círculos jurídicos americanos, o erro da inversão costuma ser chamado de falácia da acusação, porque os advogados de acusação frequentemente utilizam esse tipo de argumento falacioso para levar o júri a condenar suspeitos com base em provas frágeis. Consideremos, por exemplo, o caso de Sally Clark, na Grã-Bretanha. O primeiro filho de Clark morreu com 11 semanas de idade. A causa da morte foi declarada como síndrome da morte súbita infantil (SMSI), um diagnóstico feito quando um bebê morre inesperadamente e a autópsia não revela nenhuma causa de morte. Clark ficou grávida novamente, e o novo bebê morreu às 8 semanas de vida, novamente por SMSI. Quando isso aconteceu, a mãe foi presa e acusada de sufocar seus dois filhos. Durante o julgamento, a acusação apresentou um pediatra especialista, sir Roy Meadow, para testemunhar que, com base na raridade da SMSI, a probabilidade de que essa houvesse sido a causa da morte das duas crianças era de

> 1/73 milhões. A acusação não apresentou nenhuma outra prova substancial contra ela. Isso deveria ser suficiente para condená-la? O júri achou que sim e, em novembro de 1999, a sra. Clark foi mandada para a prisão.
>
> Sir Meadow estimara que a probabilidade de que uma criança morresse de SMSI era de 1/8.543. Ele chegou à cifra de 1/73 milhões multiplicando esse fator por si mesmo, uma vez para cada bebê. No entanto, esse cálculo presume que as mortes sejam independentes – isto é, que não haja nenhum efeito ambiental ou genético envolvido, de modo a aumentar o risco do segundo bebê uma vez que seu irmão mais velho tenha morrido de SMSI. Na verdade, em um editorial publicado no *British Medical Journal* algumas semanas após o julgamento, estimou-se que a probabilidade de que dois irmãos morressem de SMSI seria de 1/2,75 milhões. Ainda era uma possibilidade remota.
>
> Para entendermos por que Sally Clark foi condenada injustamente é fundamental considerar novamente o erro da inversão: o que buscamos não é a probabilidade de que duas crianças morram de SMSI, e sim a probabilidade de que as duas crianças que morreram tenham morrido de SMSI. Dois anos após a condenação de Clark, a Royal Statistical Society entrou na briga lançando um comunicado de imprensa no qual declarava que a decisão do júri se baseara em "um grave erro de lógica conhecido como a falácia da acusação. O júri precisa considerar duas explicações concorrentes para as mortes dos bebês: SMSI ou assassinato. Duas mortes por SMSI ou duas mortes por assassinato são ambas bastante improváveis, mas, neste caso, uma delas aparentemente aconteceu. O que importa é a probabilidade relativa das mortes … não só o quanto é improvável. Um matemático estimou posteriormente a probabilidade relativa de que uma família perdesse dois bebês por SMSI ou por assassinato. Ele concluiu, com base nos dados disponíveis, que a chance de que dois bebês morressem de SMSI era 9 vezes maior do que a de que houvessem sido assassinados. Os Clark recorreram da sentença e, na apelação, contrataram seus próprios estatísticos como testemunhas. A sentença foi mantida, mas eles continuaram a buscar explicações médicas para as mortes e, no processo, descobriram que o patologista que trabalhava para a acusação havia omitido a informação de que o segundo bebê tinha uma infecção bacteriana no momento da morte, uma infecção que poderia tê-la provocado. Com base nessa descoberta, um juiz revogou a sentença, e após quase três anos e meio, Sally Clark foi libertada da prisão.

Para a acusação, o cálculo da probabilidade de que dois bebês morram de SMSI foi simplesmente observar a quantidade de casos de morte por SMSI em um total de crianças nascidas. Foi verificado que a cada 8.543 crianças nascidas 1 morre de SMSI. Porém, essa probabilidade é falha já que estamos comparando nascimento e morte. O correto deveria ser a comparação entre mortes, como já especificado pelo autor.

Como o texto não nos diz o total de mortes ocorridas dentro daquele universo de crianças apresentado, fica impossível, com os dados apresentados, calcular a probabilidade de que uma criança morra de SMSI. No entanto, para melhor ilustrar o que já foi dito, poderemos utilizar uma referência do Brasil.

Segundo o site www.brasil.gov.br entre os anos 1990 e 2012 o Brasil obteve uma redução nos índices de mortalidade infantil de cerca de 70%. E em 2013, obteve a taxa de mortalidade de 15 crianças a cada 1.000, ou seja, $1{,}5\%$.

Utilizando essa estatística, e considerando os dados proporcionais apresentados pelo problema, poderemos perceber que se 15 a cada 1.000 morrem, então fazendo uma proporção simples temos a informação de que a cada 8.543 crianças nascidas, cerca de 128 ($1{,}5\%\ \text{de}\ 8.543$) morrem. E conforme foi estabelecido pelo caso, a cada 8.543 crianças nascidas uma morre de SMSI. Ou seja, podemos considerar que a cada 8.543 crianças nascidas, 128 morrem e 1 morre de SMSI. Consideraremos, então, que do total de 128 mortes 1 foi por SMSI. Portanto, a taxa de mortalidade por SMSI será de $1/128$. E assim, a probabilidade de que duas crianças morram de SMSI é de $\frac{1}{128} x \frac{1}{128} = \frac{1}{16.384}$. Ainda remoto, porém muito mais provável que o cálculo apresentado pela acusação.

Neste contexto, o próprio autor já mencionou que se trata da falácia da acusação. Ou seja, um erro que pode causar a condenação de pessoas com provas frágeis. No caso apresentado, mais uma vez os dados usados para calcular a probabilidade de ocorrerem duas mortes na mesma família foram falsos, visto que foram comparadas mortes com nascimentos.

Além de condenar inocentes esses erros podem inocentar culpados conforme mostrado em mais esse exemplo trazido por Mlodinow.

PROBLEMA 12 (MULHERES ESPANCADAS)

> Alan Dershowitz, renomado advogado e professor da Faculdade de Direito de Harvard, também empregou com sucesso a falácia da acusação – para ajudar a defender O.J. Simpson em seu julgamento pelo assassinato da ex-mulher, Nicole Brown Simpson, e seu namorado. O julgamento de Simpson, antigo astro do futebol americano, foi um dos maiores acontecimentos na mídia entre 1994 e 1995. A polícia tinha uma enorme quantidade de provas contra ele. Encontraram em sua fazenda uma luva manchada de sangue que parecia corresponder ao sangue encontrado na cena do crime. Manchas de sangue que correspondiam ao de Nicole foram encontradas nas luvas de Simpson, em seu Ford Bronco branco,

> num par de meias em seu quarto, na casa e na entrada da garagem. Além disso, o DNA retirado das manchas de sangue na cena do crime correspondia ao de O.J. Simpson. A defesa não teve muito a fazer, além de acusar o Departamento de Polícia de Los Angeles de racismo – Simpson é negro – e criticar a integridade da polícia e a autenticidade das provas.
>
> A acusação decidiu concentrar a abertura do caso na propensão de Simpson a se tornar violento com Nicole. Os advogados de acusação passaram os primeiros dez dias do julgamento apresentando provas de que ele havia sido violento com ela, alegando que isso já era um bom motivo para acreditarmos que seria suspeito do homicídio. Em suas palavras: “Um tapa é um prelúdio de um homicídio.” Os advogados de defesa usaram essa estratégia como base para suas acusações de falsidade, afirmando que a acusação havia passado duas semanas tentando enganar o júri e que as provas de que Simpson havia batido em Nicole em ocasiões prévias não tinham nenhum significado. Eis o raciocínio de Dershowitz: 4 milhões de mulheres são espancadas anualmente por maridos e namorados nos Estados Unidos; ainda assim, em 1992, segundo as estatísticas do FBI sobre crimes, um total de 1.432 dessas mulheres, ou $1/2.500$, foram assassinadas por seus maridos ou namorados. Portanto, replicou a defesa, poucos homens que dão tapas ou espancam suas parceiras domésticas acabam assassinando-as. Verdade? Sim. Convincente? Sim. Relevante? Não. O número relevante não é a probabilidade de que um homem que bate na mulher acabe matando-a ($1/2.500$), e sim a probabilidade de que uma mulher espancada que foi assassinada tenha sido assassinada pelo espancador. Segundo as mesmas estatísticas do FBI em 1993, a probabilidade que Dershowitz (ou a acusação) deveria haver relatado era esta: de todas as mulheres que apanhavam dos maridos e morreram nos Estados Unidos em 1993, cerca de 90% foram mortas pelo espancador. Essa estatística não foi mencionada no julgamento.
>
> À medida que se aproximava a hora do veredicto, o volume de ligações telefônicas de longa distância caiu pela metade, o volume de transações na Bolsa de Valores de Nova York caiu em 40% e cerca de 100 milhões de pessoas ligaram suas televisões e rádios para ouvir a sentença: inocente. Dershowitz pode ter sentido que sua decisão de enganar o júri foi justificada, pois, em suas palavras: “O juramento feito no fórum – ‘dizer a verdade, toda a verdade e nada mais que a verdade’ – só se aplica às testemunhas. Advogados de defesa, de acusação e juízes não assumem esse compromisso… De fato, podemos dizer que uma das fundações sobre as quais se apoia o sistema de justiça americano é *não* dizer toda a verdade. (MLODINOW, 2009)

Claramente o autor resolveu o caso. A defesa de O.J. Simpson utilizou como espaço amostral todas as mulheres espancadas por seus parceiros, quando na verdade precisaria ter considerado todas as mulheres mortas que eram espancadas pelo marido (como era o caso).

Se 1.432 mulheres espancadas pelo marido morreram, e 90% delas foram mortas pelo próprio espancador, então podemos concluir que $1.288\ (90\%\ \text{de}\ 1.432\ =\ 1.288)$ mulheres que eram espancadas pelo marido foram mortas pelo próprio.

Como mostrado aqui, não precisamos excluir as estatísticas e cálculos probabilísticos dos processos jurídicos, mas é preciso utilizá-los de forma correta. No caso apresentado, a probabilidade bem usada poderia condenar o assassino sem maiores problemas.

Em outros casos, apesar de não ter uma consequência tão grave quanto as anteriores, o uso das probabilidades e estatísticas podem ser responsáveis pelo sucesso ou fracasso de uma marca ou empresa, conforme mostrado nos dois próximos problemas. No primeiro uma taxa de aumento ou queda tão pequena que não deveria ter sido considerado variação, enquanto que no segundo, as notas subjetivas dadas pelos críticos podem trazer sucesso ou falência da marca.

PROBLEMA 13 (TAXA DE DESEMPREGO)

> Num agosto recente, a Agência de Estatísticas do Trabalho dos Estados Unidos relatou que a taxa de desemprego era de $4{,}7\%$. Em julho, a agência relatara uma taxa de $4{,}8\%$. A variação desencadeou notícias como está no *New York Times*: “Empregos e salários aumentaram modestamente no último mês.” Porém, nas palavras de Gene Epstein, editor de economia da revista *Barron’s*: “Só porque o número se alterou, não devemos concluir necessariamente que a coisa em si tenha mudado. Por exemplo, a qualquer momento a taxa de desemprego se altera em $1/10$ de ponto percentual … tal variação é tão pequena que não temos como saber se realmente ocorreu uma mudança.” Em outras palavras, se a Agência de Estatísticas do Trabalho medir a taxa de desemprego em agosto e repetir a medição uma hora depois, há uma boa chance de que, em virtude unicamente do erro, a segunda medição seja diferente da primeira em ao menos $1/10$ de ponto percentual. Será que o *New York Times* traria então a manchete “Empregos e salários aumentaram modestamente às 14h”? (MLODINOW, 2009)

PROBLEMA 14 (NOTAS DOS VINHOS)

> Outra medição subjetiva que recebe mais atenção do que merece é a classificação de vinhos. Nos anos 1970, a indústria dos vinhos era um empreendimento adormecido; crescia, mas principalmente nas vendas de vinhos de baixa qualidade. Então, em 1978, ocorreu um evento muitas vezes creditado pelo rápido crescimento da indústria: um advogado autoproclamado crítico de vinhos, Robert M. Parker Jr., decidiu que, além de escrever artigos críticos, daria notas numéricas aos vinhos, numa escala de 100

> pontos. Ao longo dos anos, a maior parte das publicações especializadas passou a fazer o mesmo. Atualmente, as vendas anuais de vinhos nos Estados Unidos passam de US$ 20 bilhões, e milhões de consumidores fanáticos evitam colocar a mão no bolso sem antes observar a classificação de um vinho para sustentar sua escolha. Assim, quando a revista *Wine Spectator* deu, digamos, nota 90 em vez de 89 ao cabernet sauvignon Valentín Bianchi de 2004, um vinho argentino, esse ponto a mais representou uma enorme diferença nas vendas do produto. De fato, se observarmos uma loja de vinhos, veremos que as bebidas em promoção, menos procuradas, muitas vezes possuem notas um pouco abaixo de 90. Mas qual é a probabilidade de que o Valentín Bianchi de 2004, que recebeu nota 90, tivesse recebido nota 89 se a avaliação houvesse sido repetida, digamos, uma hora depois? (MLODINOW, 2009)

Os dois exemplos acima trazem erros de interpretação de dados. O próprio autor do livro já questiona os dados apresentados por tentar quantificar caso subjetivos. Qual a probabilidade de que a taxa de desemprego mude de uma hora para a outra? Qual a probabilidade de que uma nota dada a um vinho mude se a avaliação fosse repetida uma hora depois, ou que seja feita por outro crítico? Essas respostas são tão subjetivas quanto afirmar que uma proposição sobre a lei da física ainda não provada esteja correta. Podemos ter ideias sobre elas, mas não pode ser quantificada.

Esses casos de erros de interpretação, as falácias probabilísticas, causam impressões erradas sobre o que podemos esperar de certos experimentos aleatórios. Para tentar evitá-los, é importante o conhecimento dos conceitos probabilísticos, mas não só eles garantem que não vamos errar, afinal, alguns erros são induzidos pelo sentimento enquanto outros por processos para simplificar a tomada de decisão. Esses processos são chamados de heurísticas.

No próximo capítulo, vamos estudar como as pessoas costumam tomar suas decisões acerca de experimentos aleatórios e como são formadas essas heurísticas e vamos conhecer alguns erros muito comuns cometidos pelas pessoas que tentam usar métodos estatísticos para tentar prever ocorrências futuras.

3 HEURÍSTICAS

Vimos nos capítulos anteriores alguns problemas probabilísticos que podem afetar diretamente a vida das pessoas. Casos como um exame de HIV com uma falsa certeza, um homem preso injustamente, outro inocentado usando argumentos falsos, taxas de desemprego manipuladas da forma que mais convir e notas em vinhos que pode salvar ou falir uma marca. Todos esses casos, podem ser melhor interpretados com os conhecimentos de probabilidade, conforme vimos. Entretanto, poderíamos ter mais atenção a cada caso conhecendo os erros mais comuns cometidos pelas pessoas nesses casos.

No jogo da mega-sena, as interpretações estatísticas giram em torno de tentar prever as próximas ocorrências dos números a serem sorteados. Assim, os apostadores tentam determinar os próximos seis números afim de ganhar o prêmio máximo.

Neste sentido, **Daniel Kanneman** e **Amos Tversky** fizeram um estudo psicológico acerca do comportamento humano quando pretendem usar conhecimentos estatísticos para prever futuras ocorrências.

Segundo o que foi apresentado por Adriana Sbicca, Daniel Kanneman começou seu estudo sobre as decisões humanas em 1955 quando pretendia traçar perfis de recrutas para as forças de defesa de Israel. Deste trabalho, escreveu um questionário para tentar gerar uma previsão quanto a performance dos recrutas. Este trabalho deu início às pesquisas de Tversky sobre a psicologia da previsão intuitiva.

Em 1974, escreveram um artigo na revista ***Science*** intitulado "*Judgment under Uncertainty: heuristics and biases"* e, em 1982, um livro com este mesmo título que se tornaram marcos do programa *heuristics and biases* (heurísticas e vieses). O foco destas publicações eram princípios heurísticos que criavam atalhos para julgamentos de probabilidade.

HEURÍSTICA, pelo dicionário, diz-se do processo pedagógico de encaminhar o aluno a descobrir por si mesmo o que se quer ensinar, geralmente através de perguntas. Enquanto **VIÉS**, também pelo dicionário, é distorção ou tortuosidade na maneira de observar, de julgar ou de agir. Em resumo, o que Kanneman e Tversky pretendiam era estudar os erros cometidos por todos aqueles que buscavam adivinhar o futuro através de frequências de eventos ocorridos no passado.

Segundo eles, ao se depararem com situações em que precisariam de uma análise probabilística muito complexa, as pessoas desenvolvem processos heurísticos mais simples para facilitar a avaliação e tomada de decisão. Para eles, “Em geral, estas heurísticas são totalmente úteis, mas, algumas vezes, elas levam a erros graves e sistemáticos.” (TVERSKY & KAHNEMAN, 1974, p. 1124).

Este resultado apresentado por Kahneman e Tversky, foi concluído empiricamente, ou seja, através de testes realizados. Um dos seus primeiros testes, inclui uma loteria simples.

Eles apresentaram um jogo em que o apostador teria duas opções de escolha. Em uma delas ele terá 20% de chance de ganhar $\$4.000,00$ ou 25% de chance de ganhar $\$3.000,00$. Neste teste, a maioria dos consultados escolheu a primeira opção.

Entretanto, numa segunda possibilidade de jogo, os apostadores deveriam escolher entre 80% de chance de ganhar $\$4.000,00$ e 100% de chance de ganhar $\$3.000,00$, a escolha preferida foi a segunda.

Através do cálculo do valor esperado de cada uma das situações, segundo Tversky e Kahneman, as escolhas lógicas seriam as duas primeiras. No primeiro caso, temos $0{,}2 \times \$4.000 = \$800$ e $0{,}25 \times \$3.000 = \$750$. E a maior possibilidade de ganho, torna a primeira escolha logicamente mais vantajosa. No segundo cenário, temos $0{,}8 \times \$4.000 = \$3.200$ e $1 \times \$3.000 = \$3.000$. E, novamente, a maior possibilidade de ganho, torna a primeira escolha logicamente mais vantajosa. Entretanto neste jogo as pessoas colocaram mais uma variável, a certeza de ganhar. Este comportamento se dá pelo padrão de comportamento de simplificação, este padrão mostra a tendência das pessoas de superestimarem resultados que são considerados certos em relação a resultados considerados prováveis. Esta maior atração pelos resultados mais próximos do certo foi denominada de **efeito certeza.**

O trabalho de Adriana Sbicca trouxe mais um exemplo importante para mostrar o que Kahneman e Tversky buscaram em seu trabalho. Neste exemplo eles conseguiram mostrar que mesmo que as pessoas tenham conhecimento suficiente para tirar conclusões sobre o assunto, quando se trata de eventos novos, elas tendem a criar heurísticas para simplificar seus entendimentos de tomada de decisão, e esses procedimentos podem leva-las a erros.

> Outro teste ilustra a estrutura dos estudos feitos pelos dois autores. Foram dados sessenta segundos a estudantes da University British Columbia (UBC) para listar palavras em inglês com sete letras sendo que receberam a instrução de listar palavras cuja sexta letra fosse "n" e palavras que terminassem em "ing". Como resultado, os estudantes listaram muito mais palavras terminando com "_ing" do que com "_n_" (médias de 6,4 e 2,9, respectivamente) (TVERSKY & KAHNEMAN, 1983, p. 295). Os autores chamaram a atenção para o fato de que a última forma (_n_) inclui a primeira (_ing); palavras como writ**ing**, work**ing**, fill**ing** e bias**ing** poderiam estar nas duas listas, mas outras como airli**n**e, accou**n**t e repri**n**t, só poderiam constar da segunda lista. Seria esperado, portanto, que a lista contendo a forma "_n_" fosse maior, ou, no mínimo, igual à primeira, o que se contrapõe ao resultado obtido no experimento. K&T explicaram esse resultado, pelo fato das palavras terminadas com "ing" estarem mais disponíveis em nossa memória, o que faz com que as pessoas se lembrem mais de palavras com essa forma. Segundo Kahneman, esse é um exemplo de uso de *heurística da disponibilidade* (*availability heuristics*), que se refere à maior influência nas decisões de eventos que são mais fáceis de imaginar ou de se lembrar. O uso desta regra pode explicar o fato das pessoas tenderem a superestimar a probabilidade de eventos que ocorreram recentemente em relação àqueles que aconteceram há mais tempo (como o medo de viajar de avião devido à ocorrência de um acidente recente ou o aumento da procura por seguro logo após um terremoto). (SBICCA, 2011)

Este exemplo mostra que mesmo exemplos simples podem trazer resultados inesperados. Repare que a falha de interpretação do problema por parte das pessoas não está pautada na falta de conhecimento. Afinal, as pessoas conhecem as palavras, entretanto, neste teste, foram utilizadas as palavras que vieram mais rapidamente a cabeça. A simplificação do problema pelo simples resultado que chegar mais rápido trouxe o erro de não utilizar as palavras nos dois grupos.

Os autores trouxeram um outro teste interessante que traz um resultado parecido. O teste consistia em perguntar às pessoas qual das opções apresentadas é mais provável sobre a personagem Linda.

> Linda tem 31 anos, é solteira, sincera e muito brilhante. Ela é formada em filosofia. Quando era estudante, preocupava-

> se com temas como discriminação e justiça social, e também participou de manifestações anti-nucleares:
> a. Linda é professora do ensino fundamental.
> b. Linda trabalha numa livraria e faz yoga.
> c. Linda é ativista do movimento feminista.
> d. Linda trabalha como assistente social.
> e. Linda é um membro da Liga das mulheres eleitoras.
> f. Linda é uma caixa de banco.
> g. Linda é uma corretora de seguros.
> h. Linda é uma caixa de banco e é ativista do movimento feminista.

O teste trouxe como resposta que 80 a 90% das pessoas responderam a letra h. Kahneman e Tversky analisaram o resultado como a **falácia da Razão**. Afinal, como poderia a letra h ser mais provável que a resposta f. Veja que a resposta f traz que Linda é caixa de banco, enquanto que a resposta h diz que além de caixa de banco, Linda é ativista do movimento feminista. E, como dito no trabalho de Sbicca (2011), este julgamento viola a regra da conjunção $(P(A \cap B) \leq P(B))$, uma das mais simples e básicas leis da probabilidade, qualificaram os autores.

Estatisticamente, podemos garantir que a probabilidade de Linda ser apenas caixa de banco é maior que a probabilidade de ela ser caixa de banco e ativista do movimento feminino. Por isso, a resposta da maioria está errada.

Mais um exemplo explorado por Kahneman e Tversky trata do lançamento de moedas. As pessoas esperam, em uma moeda não viciada, que os resultados saiam próximos de 50% para cara e 50% para coroa, mesmo em uma sequência pequena. (TVERSKY & KAHNEMAN, 1974). Sendo assim, lançando uma moeda 4 vezes e ocorrendo 2 caras e 1 coroa nos três primeiros lançamentos, as pessoas tendem a acreditar que o ocorrerá coroa no quarto lançamento. Essa expectativa dos resultados representarem a probabilidade calculada, mesmo em sequências pequenas foi chamada de **heurística da representatividade**. Herbert Simon (1979) chamou situações como a descrita acima de **FALÁCIA DO JOGADOR**.

Representatividade é um atributo do nível de correspondência entre uma amostra e uma população, um exemplo e uma categoria, um ato e um ator ou, de maneira mais geral, um resultado e um modelo. (TVERSKY & KAHNEMAN, 1983).

Um caso interessante de falácia do jogador ocorreu em um jogo de roleta no Cassino de Monte-Carlo. Segundo, Lehrer (2009), em 1913 a bola da roleta caiu 26 vezes seguidas em uma casa preta. A roleta de apostas possui o mesmo número de casas pretas e vermelhas, por isso, é fácil intuir que a probabilidade de uma bola cair em casas pretas e vermelhas é de 50% para cada. Como a bola caiu, em sequência, em casas pretas, muitos apostadores passaram a apostar mais em casas vermelhas, acreditando, incorretamente, que o “desbalanço” na aleatoriedade do jogo traria uma sequência de casas vermelhas em seguida.

Aparentemente os jogadores esperam que algum desvio em uma direção será brevemente cancelado pelo correspondente desvio na direção oposta, mas os desvios não são cancelados nas amostras e sim simplesmente diluídos (TVERSKY & KAHNEMAN, 1971).

Durante o desenvolvimento de seu trabalho Kahneman e Tversky enunciaram outros tipos de heurísticas. Além da heurística de representatividade, se destaca em seu trabalho as de ancoragem e de disponibilidade.

A **heurística de disponibilidade** é uma referência usada para processos que utilizam fatores que vem mais rápido a mente. São elementos que estão mais disponíveis durante a tomada de decisão. Enquanto que a **heurística da ancoragem** é o processo decisório que leva em consideração determinadas informações usadas como referência (TVERSKY & KAHNEMAN , 1983).

Em seu trabalho, O papel das heurísticas no julgamento e na tomada de decisão sob incerteza, Leandro Miletto **TONETTO** (2006), descreve a heurística da ancoragem como um processo que influencia na tomada de decisão mesmo que a âncora em questão não tenha relação lógica com a decisão a ser tomada.

A heurística da ancoragem é iniciada em diversos trabalhos, segundo Tonetto, com a solicitação explícita para que as pessoas comparem o valor da âncora com o valor alvo. Em um estudo de TVERSKY & KAHNEMAN (1974), foi solicitado que as pessoas estimassem a porcentagem de países africanos nas Nações Unidas e o grupo que recebeu o número 10 como âncora inicial (obtido por meio de uma roda da fortuna) estimou em 25% em média, enquanto que o grupo que recebeu o número 65 como valor inicial teve uma estimativa média de 45%.

Claramente, no exemplo trazido acima, houve influência da âncora na decisão dos grupos de pessoas, apesar de não haver qualquer relação lógica entre o número sorteado e a resposta a ser dada.

Podemos ver casos de heurística de ancoragem ao determinar preços de coisas que não temos certeza. É claro que buscaremos os preços de itens parecidos para determinar o valor de algum bem, entretanto, se esses itens não existem no mercado temos uma dúvida muito maior ao tentar estabelecer tal preço.

A heurística de Ancoragem mostra a necessidade de aproximação que as pessoas buscam ao se deparar com situações em que a tomada de decisão não está dentro de qualquer atitude já vista por elas anteriormente. Neste caso, qualquer ponto de apoio traz uma sensação de estar próximo da realidade. É mais um caso de pensamento simplificado que pode levar as pessoas a tomar decisões erradas conforme trazido por Tonetto.

> A influência das âncoras persiste inclusive quando elas são claramente não informativas para o julgamento (Tversky & Kahneman, 1974; Wilson et al., 1996). Mesmo a advertência às pessoas sobre a natureza não informativa da âncora não impede que elas sejam influenciadas por este valor (Chapman & Johnson, 1999; George et al., 2000). Em julgamentos e decisões efetuados por pessoas com auxílio de programas de computador, o efeito de ancoragem também foi observado (George et al., 2000), o que demonstra efetivamente a robustez do fenômeno. (TONETTO, 2006)

Pelo trabalho de Tverky e Kahneman (1974), a heurística da disponibilidade mostra que a facilidade com que um determinado fato é lembrado ou imaginado por um indivíduo pode determinar uma hiper ou subestimação da probabilidade ou frequência desse evento ocorrer. Significa dizer que as pessoas julgam a probabilidade pela facilidade de evocar exemplos em suas memórias.

Um dos testes sobre heurística da disponibilidade foi feito por Tversky e Kahneman (1974). Nele, os autores pediram que os participantes julgassem a seguinte questão: “Se uma palavra de três letras é mostrada aleatoriamente de um texto em inglês, é mais frequente que a palavra comece com ‘r’ ou que tenha ‘r como sua terceira letra?” A maioria dos participantes disse ser mais provável que comece com “r”. Entretanto, a língua inglesa possui mais palavras de três letras que terminam em “r” do que as que começam em “r”. Isso acontece porque os participantes têm em sua memória mais palavras iniciando do que terminando com “r”. Segundo Tonetto (2006), O uso da heurística da disponibilidade leva-nos cotidianamente ao erro no que tange à influência da memória de longo prazo. Este ainda traz a conclusão de que a heurística da disponibilidade é um procedimento eficaz e rápido, utilizado com frequência no julgamento e tomada de decisão em condições de incerteza. Evidentemente, julgar e tomar decisões baseado apenas nas memórias pode nos levar a erros de avaliação.

Em seu trabalho, “*Misconceptions of probability: An Experiment with a small-group, Activity-Based, Model building approach to introductory probability at the college level”*, J. Mihael Shaughnessy, afirma que as falhas ocasionadas por heurísticas de avaliação e representatividade são muito difíceis de serem superadas (Shaughnessy,1977) e que a exposição à teoria da probabilidade e estatística não é necessariamente suficiente para superar os vieses sistemáticos que são introduzidos pela disponibilidade heurística e representatividade.

Para comprovar o apresentado o autor buscou realizar atividades com uma turma experimental de nível superior.

Nestas atividades práticas acerca de probabilidade, ele pediu para os estudantes realizarem experimentos aleatórios. Jogar seis moedas, jogar três percevejos de papel, jogar três dados dentre outros. Nas atividades os alunos deveriam realizar o experimento um determinado números de vezes e anotar os resultados possíveis. Depois deveriam tentar estimar a probabilidade de ocorrer cada um dos resultados possíveis do experimento. Em seguida, eles calculavam a probabilidade através de um modelo matemático e comparavam os resultados. Não foi surpreendente que em todas as atividades a diferença entre a probabilidade estimada pelos alunos não coincidiam com as probabilidades calculadas através dos modelos matemáticos.

Entretanto, após o curso, houve uma melhora na avaliação de probabilidades. Em seu texto, Shaughnessy cita que fez testes antes e depois do curso com todos os alunos comparando os resultados. Os instrumentos testaram o conhecimento de alguns conceitos de probabilidade e a confiança na representatividade e disponibilidade na estimativa de probabilidade de eventos.

Os problemas apresentados para teste foram os seguintes:

R1: A probabilidade de ter um menino é cerca de $1/2$. Qual das sequências seguintes é mais provável tendo seis filhos?

(A) H M H H M H

(B) H H H H M H

(C) a mesma chance de ocorrer cada sequência.

R2: (Assumindo o mesmo que em R1) Qual das sequências é mais provável de ocorrer tendo seis filhos?

(A) H M M H M H

(B) H H H M M M

(C) a mesma chance de ocorrer cada sequência.

R3: Qual a probabilidade de que em seis crianças, três sejam meninas?

R4: O que é mais provável de ocorrer?

(A) Retirar uma bola vermelha de um jarro com 10 bolas vermelhas e 90 brancas; ou

(B) Retirar quatro bolas vermelhas em sequencia de um jarro com 50 bolas vermelhas e 50 brancas.

Em todos os problemas apresentados acima houve uma melhora considerável na interpretação das probabilidades. Entretanto, o resultado do problema seguinte não apresentou uma grande melhora após os alunos terem participado do curso. Vamos analisá-lo.

R5: A chance de um bebê nascer menino é de cerca de ½. Ao longo de um ano inteiro, haveriam mais dias em que pelo menos 60% dos bebês nascidos seriam meninos:

(A) Em um hospital grande;

(B) Em um hospital pequeno;

(C) Não faz diferença.

A grande maioria dos alunos testados marcou a opção (C) não faz diferença. Este resultado apoia a afirmação de Kahneman e Tverskky de que as pessoas tendem a acreditar que ambos os hospitais deveriam ser igualmente representativos para a proporção da população de meninos e meninas. Portanto, para a maioria o tamanho da amostra não faz diferença.

Para exemplificar e melhorar o entendimento, vamos analisar por que a resposta certa é o hospital pequeno.

Em um pequeno hospital existe a maior chance de ocorrer poucos nascimentos, e sendo assim, existe maior chance de ocorrer uma frequência relativa maior para um dos sexos em alguns dias.

Repare que se nascem 3 bebês em um hospital 2 meninos e 1 menina, podemos dizer que a frequência relativa de nascimento foi de aproximadamente 66% de meninos, o que atende a pergunta.

Entretanto, em um grande hospital, por ser maior, a chance de acontecerem poucos nascimentos é remota e, portanto, podemos concluir que em um hospital pequeno é mais provável que a frequência relativa de nascimento de meninos em um dia ao longo do ano seja maior que 60%.

Este exemplo ajuda a concluir que o tamanho da amostra faz toda a diferença quando tratamos da representatividade de um grupo em relação a uma população.

3.1 A lei dos pequenos números

O termo Lei dos pequenos números foi cunhado por Tversky e Kahneman para descrever como pessoas exageram o grau em que a distribuição de probabilidade em um pequeno grupo se assemelhará à distribuição de probabilidade na população média. (Rabin, 2002)

Da heurística da representatividade já mostrada anteriormente, podemos inferir a tendência que as pessoas têm de aumentar a representação de uma pequena amostra proporcionalmente a toda uma população.

No exemplo já citado anteriormente que trata de quatro lançamentos de uma moeda em que os três primeiros resultados são duas caras e uma coroa, as pessoas que se utilizam da heurística da representatividade acreditam que no quarto lançamento ocorrerá uma coroa, pois assim a frequência relativa da sequência de experimentos se aproximará da probabilidade de cada evento.

Da mesma forma, no caso do cassino de Monte-Carlo os apostadores acreditaram que uma grande sequência de resultados pretos seria seguida por uma outra grande sequência de resultados vermelhos para que a frequência relativa do experimento se aproxime da probabilidade de cada evento.

Tanto no caso da moeda, quanto no caso da roleta, a tendência de as pessoas acreditarem que pequenas sequências de resultados de um dado experimento podem representar a probabilidade calculada de cada evento foi chamada por Kahneman e Tversky de **LEI DOS PEQUENOS NÚMEROS**.

No exemplo da moeda, caso no quarto lançamento saia cara em vez de coroa, as pessoas começam a duvidar da honestidade e confiabilidade da moeda. Enquanto que na roleta, os apostadores trouxeram o pensamento de que deveriam acontecer outros 26 sorteios que sairão casas vermelhas para que o a frequência do jogo apresente a tendência da probabilidade.

Portanto, a lei dos pequenos números apresentada por Kahneman e Tversky trata do erro que as pessoas cometem ao acreditar que uma pequena amostra de um experimento deve representar a população total deste experimento.

Durante seu trabalho *Believers in the law of small numbers*, Matthew Rabin diz que muitas pessoas exageram em como uma probabilidade de uma pequena amostra se assemelha a população da qual é desenhada. (Rabin, 2002)

Nele, Rabin traz um modelo para capturar a crença na lei dos pequenos números. Uma pessoa observa uma sequência de sinais binários. Cada valor do sinal é gerado aleatoriamente a partir de uma probabilidade estacionária referida como “taxa”. O observador possui antecedentes probabilísticos corretos sobre essa taxa e os sinais são gerados por sorteios aleatórios sem substituição de uma “urna” de N sinais, onde a urna contém a proporção dos dois valores de sinal correspondente à taxa.

Segundo o autor, o modelo captura a crença na lei dos pequenos números, pois significa que a pessoa acredita que a proporção de sinais deve se equilibrar com a taxa da população antes que os N sinais sejam observados. Quanto menor for N mais a pessoa acredita na lei dos pequenos números.

Rabin investigou o que as pessoas pensam quando observam muitos resultados de mesmo sinal em uma sequência. Para isso ele estabeleceu um procedimento de uma urna com reposição a cada dois períodos. Segundo ele, isso serve para capturar de um modo atrativo a evidencia de que as pessoas esperam em pequenas e grandes sequências uma produção de sinais aproximadamente proporcional a toda a sequência. O trabalho mostra, mais uma vez, que as pessoas não esperam que as composições de pequenas sequências difiram dramaticamente da proporção de toda a população, eles também não esperam ver “estrias” de sinais que não são representativos da frequência de toda a sequências.

Esses modelos levam diretamente a **falácia do jogador**: as pessoas esperam que o segundo empate de um sinal seja negativamente correlacionado com o primeiro sorteio. Porque nós exageramos em como a probabilidade de uma pequena sequência de jogadas de moeda pode trazer metade de resultados caras e metade de coroas. (Rabin, 2002)

4 A Mega-sena

A Mega-Sena é o jogo que paga milhões para o acertador dos 6 números sorteados. Mas quem acerta 4 ou 5 números também ganha. Para realizar o sonho de ser o próximo milionário, você deve marcar de 6 a 15 números, entre os 60 disponíveis no volante. Você pode deixar que o sistema escolha os números para você (Surpresinha) e/ou concorrer com o mesmo jogo por 2, 4 ou 8 concursos consecutivos (Teimosinha).

Os sorteios são realizados duas vezes por semana, às quartas e aos sábados. A aposta mínima, de 6 números, custa R$ 3,50. Quanto mais números marcar, maior o preço da aposta e maiores as chances de faturar o prêmio mais cobiçado do país.

O prêmio bruto corresponde a $\mathbf{46\%}$ da arrecadação, já computado o adicional destinado ao Ministério do Esporte. Dessa porcentagem:

- 35% são distribuídos entre os acertadores dos 6 números sorteados (Sena);
- 19% entre os acertadores de 5 números (Quina);
- 19% entre os acertadores de 4 números (Quadra);
- 22% ficam acumulados para serem distribuídos aos acertadores dos 6 números nos concursos de final 0 ou 5.
- 5% ficam acumulado para a primeira faixa - sena - do último concurso do ano de final zero ou 5.

Não havendo acertador em qualquer faixa, o valor acumula para o concurso seguinte na respectiva faixa de premiação. (CAIXA.GOV.BR)

4.1 Os valores das apostas

Os valores a serem apostados são calculados de acordo com a quantidade de números a serem sorteados, logicamente, quanto mais números apostados maior será a probabilidade de ganhar e sendo assim, maior deverá ser o valor apostado.

Esses valores aumentam proporcionalmente a probabilidade de vitória, para isso, basta saber quantas combinações de 6 números estão contempladas na quantidade de números apostada. Por exemplo, uma aposta de 7 números pode gerar 7 apostas de seis números, visto que $C_{7,6} = 7$. Sendo assim, o apostador que marcar sete números na cartela de apostas deverá pagar proporcionalmente sete vezes mais do que aquele que marcar apenas seis números. Como cada aposta de seis números custa $\text{R\$}3{,}50$, então o apostador de sete números deve pagar $7 \times \text{R\$}3{,}50 = \text{R\$}24{,}50$.

Evidentemente, a caixa econômica limita a quantidade de números apostados em no máximo 15 números por cartela. Até porque um apostador que marque todos os números da cartela teria 100% de chance de acertar. E daí bastaria que ele verificasse se o prêmio a ser oferecido é maior do que o valor apostado. Desse jeito a loteria deixaria de contar com a sorte do apostador e passaria a ser um investimento.

Os preços e probabilidades das apostas estão descritas no próprio site da caixa conforme tabela abaixo:

Tabela 1 - Valor das apostas e probabilidade de ganhos de acordo com os números jogados.

Quantidade de n° jogados	**Valor da aposta**	**Probabilidade de acerto (1 em)**		
		Sena	**Quina**	**Quadra**
6	R$ 3,50	50.063.860	154.518	2.332
7	R$ 24,50	7.151.980	44.981	1.038
8	R$ 98,00	1.787.995	17.192	539
9	R$ 294,00	595.998	7.791	312
10	R$ 735,00	238.399	3.973	195
11	R$ 1.617,00	108.363	2.211	129
12	R$ 3.234,00	54.182	1.317	90
13	R$ 6.006,00	29.175	828	65
14	R$ 10.510,50	16.671	544	48
15	R$ 17.517,50	10.003	370	37

Fonte: tabela retira do site www.caixa.gov.br

Não faz sentido fazer as contas de cada valor de aposta aqui, mas como mais um exemplo vamos calcular o valor de uma aposta de 15 números usando a análise combinatória.

Primeiramente, vamos calcular quantas combinações de 6 números são possíveis utilizando os 15 números apostados.

$$C_{15,6} = \frac{15!}{6!\,(15-6)!} = \frac{15.14.13.12.11.10}{6.5.4.3.2.1} = 5.005$$

Isso quer dizer que o apostador que fizer uma aposta com 15 números deverá pagar 5.005 vezes mais do que aquele que apostar apenas seis números. Sendo assim, $\text{R\$}3{,}50 \times 5.005 = \text{R\$}17.517{,}50$.

4.2 As probabilidades do jogo

O jogo consiste de um sorteio de 6 números dentre os 60 possíveis. O apostador que conseguir acertar todos os 6 números sorteados ganha o maior prêmio. Também são premiados aqueles que acertam 5 e 4 números, conforme disposto acima.

O sorteio dos números é realizado através de um globo giratório com bolas numeradas. E sendo assim, podemos concluir que se trata de um sorteio totalmente aleatório.

Porém, analisando a tabela de sorteios da mega-sena mais a fundo pode-se perceber uma variação a ser considerada. Essa variação pode causar dúvidas sobre a aleatoriedade dos sorteios. Vamos agora usar o exemplo da mega-sena, calcular suas probabilidades e aplicar a Lei dos Grandes Números para obter alguma conclusão acerca da aleatoriedade dos sorteios.

Esses valores probabilísticos estão expostos na tabela mostrada acima que também se encontram no site da Caixa Econômica Federal.

4.2.1 Probabilidade de ganhar jogando 6 números.

Podemos dizer que todos os números possuem a mesma possibilidade de serem sorteados em um sorteio qualquer. Assim, temos que o espaço amostral será $U = \{1, 2, 3, \ldots, 60\}$. Considerando o evento de acertarmos os seis números.

Dessa forma, temos:

No evento A (acerto do primeiro número sorteado) temos:

$$P(A) = \frac{6}{60}$$

No evento B (acerto do segundo número sorteado) temos:

$$P(B) = \frac{5}{59}$$

Assim, por diante,

$$P(C) = \frac{4}{58}$$

$$P(D) = \frac{3}{57}$$

$$P(E) = \frac{2}{56}$$

$$P(F) = \frac{1}{55}$$

Portanto,

$$P(A \cap \mathrm{B} \cap \mathrm{C} \cap \mathrm{D} \cap \mathrm{E} \cap \mathrm{F}) = P(A).P(B).P(C).P(D).P(E).P(F)$$

$$P(A \cap \mathrm{B} \cap \mathrm{C} \cap \mathrm{D} \cap \mathrm{E} \cap \mathrm{F}) = \frac{6}{60}.\frac{5}{59}.\frac{4}{58}.\frac{3}{57}.\frac{2}{56}.\frac{1}{55} = \frac{720}{36.045.979.200} = \frac{1}{50.063.860}$$

Portanto, a probabilidade de conseguirmos acertar os seis números sorteados em uma aposta de seis números é de $1 \text{ em } 50.063.860$ (cinquenta milhões sessenta e três mil oitocentos e sessenta).

Mas, veja bem, se um apostador jogar cinquenta milhões sessenta e três mil oitocentos e sessenta a $R\$3{,}50$ por jogo ele gastará $R\$175.223.510{,}00$ (cento e setenta e cinco milhões duzentos e vinte e três mil quinhentos e dez reais). Porém, historicamente, a mega-sena pagou valor superior a esse apenas uma vez, e a média do prêmio pago é de $R\$14.059.948{,}21$. Valor bem abaixo dos cem milhões apostados.

Há ainda a possibilidade de apostar em mais de seis números. Para simplificar o trabalho, vamos calcular a probabilidade e os valores a serem pagos pela aposta com maiores quantidades de números aceita pela loteria federal que é uma aposta de 15 números com valor de $R\$17.517{,}50$ por aposta.

4.2.2 Probabilidade de ganhar jogando 15 números

Obviamente a probabilidade de acertar seis números em 15 é muito maior. Esta probabilidade pode ser calculada da seguinte forma:

P(A) = 15/60

P(B) = 14/59

P(C) = 13/58

P(D) = 12/57

P(E) = 11/56

P(F) = 10/55

$$P(A \cap B \cap C \cap D \cap E \cap F) = \frac{15}{60} \times \frac{14}{59} \times \frac{13}{58} \times \frac{12}{57} \times \frac{11}{56} \times \frac{10}{55} = \frac{3.603.600}{36.045.979.200}$$

$$= \frac{13}{130.036} \approx \frac{1}{10.002}$$

Portanto, a probabilidade de conseguirmos acertar os seis números sorteados em uma aposta de quinze números é de aproximadamente $1 \text{ em } 10.002$ (dez mil e dois).

Mas, veja bem, se um apostador jogar dez mil e duas vezes a R\$17.517,50 por jogo ele gastará R\$175.210.035,00 (cento e setenta e cinco milhões duzentos e dez mil e trinta e cinco reais). Valor próximo do calculado para jogos de seis números. O que nos leva a mesma conclusão sobre as apostas.

Mostrando ainda que como a média de pagamento é de cerca de quatorze milhões, podemos dizer que o governo federal terá um gordo lucro de cerca de 86 milhões a cada pagamento de prêmio que se faça.

Em resumo, não importa quantos números o jogador aposta, a mega-sena foi feita para lucro do governo e a relação entre preço da aposta e a probabilidade de ganhar sempre será proporcional em uma relação fixa. Quanto mais números apostados, mais caras são as apostas e esses preços são estabelecidos pela probabilidade de ocorrer vitória naquela aposta.

4.3 **Métodos de apostas**

Mesmo com as demonstrações matemáticas mostradas, alguns apostadores insistem em tentar métodos “milagrosos” para tentar aumentar as chances de acertar os seis números sorteados. Surpreendentemente, quando mostrei em uma tabela a análise dos números já sorteados a um grupo de mestrandos em matemática, a grande maioria anotou os números mais sorteados ao longo dos anos.

Em pesquisa pela internet podemos encontrar vários métodos e dicas que supostamente poderiam aumentar as probabilidades de vitória, neste capítulo vamos discutir alguns métodos expostos e tentar desmistifica-los.

Por exemplo, no site www.numerosmegasena.com existe um artigo não assinado dizendo que é possível prever os números a serem sorteados e que os sorteios seguem um padrão previsto. Segundo o artigo as loterias não devem mais estar sendo vistas como uma forma de jogo, mas uma verdadeira representação da teoria probabilística e da lei dos grandes números. No entanto, o mesmo artigo não menciona a fonte, quem foi o matemático e nem sequer qual seriam os padrões seguidos pelos sorteios. Por isso, cabe aos leitores apenas desconfiar da veracidade das informações.

Outro artigo interessante e que chama a atenção de muitos está no portal noticias.r7.com. A chamada do artigo diz: “Especialista dá dicas de como se dar bem e faturar prêmio da mega sena”.

Neste artigo o “especialista”, matemático e dono de casa lotérica Munir W. Niss, conhecido como Munir Pé Quente ensina:

1 – Evite jogar números seguidos.

2 - Caia fora das apostas com todas as dezenas terminadas com o mesmo número

3 - O cartão da Mega-Sena tem cinco números dobrados: 11, 22, 33, 44 e 55. Pé Quente aconselha que se escolham, no máximo, dois por aposta.

4 - Procure equilibrar a escolha entre números pares e ímpares.

5 - Pegue o cartão com 60 números e faça uma cruz imaginária dividindo-o em quatro partes. Dê um risco horizontal, até embaixo, entre a linha do 05 e a do 06. Depois, faça outro risco, agora horizontal, separando a linha do 21 ao 30 da seguinte, que vai de 31 a 40. Você terá um cartão dividido em quatro partes, cada uma com 15 números. No caso de apostas com seis dezenas, marque um número em cada um desses quadros e, em seguida, coloque o segundo número em dois deles, à sua escolha. Se o cartão tiver sete ou mais apostas (o máximo são 15, que custam R$17.517,80 – isso mesmo: um bom carrinho usado), distribua de forma equilibrada por esses quatro quadros. Um exemplo: uma aposta de oito dezenas permite colocar duas delas em cada “quadrado” de 15 números.

6 - Para quem quer (e pode) gastar mais um pouco na fezinha, Pé Quente aconselha o cartão com oito apostas. O com oito é a melhor opção se pensarmos na relação entre custo e benefício. Não exige um dinheiro absurdo, aumenta

consideravelmente as chances de ganho e dá a oportunidade de o apostador ou o grupo fazer mais de uma tentativa.

O método utilizado para sorteio das dezenas é um globo com 60 bolas, sendo as seis sorteadas uma a uma. Evidentemente, se todas as bolas têm a mesma forma e mesmo peso, o que faria com que números em sequência não sejam sorteados? Sendo assim, e admitindo que a probabilidade de qualquer das bolas ser sorteada é a mesma, não faz sentido pensar que a probabilidade de ocorrer o número 51 é menor que a de ocorrer qualquer outro número logo depois de sorteado o 50.

Este pensamento refuta todas as 5 primeiras dicas, porém a dica 6 é uma boa dica. Visto que aumentando o valor das apostas e aumentando a quantidade de números jogados, a probabilidade de acertar os seis sorteados aumenta. Sendo assim, a única dica realmente útil deste artigo é a sexta que em outras palavras menciona o que a probabilidade nos ensina. Quanto mais números apostados maior a chance de acertar os números sorteados.

Outro artigo falando sobre a mega-sena foi encontrado no site www.g1.globo.com com o título: “O que os matemáticos pensam sobre as ditas fórmulas e dicas infalíveis para se ganhar na loteria. ”

Neste artigo as regras giram em torno de apostar nos números mais sorteados, não apostar em números sequenciais e não apostar em números que estão na mesma coluna. Nele, o autor menciona também, a divisão do cartão em quatro quadrantes, assim como o artigo acima.

Entretanto, este artigo veio com outra intenção, ao invés de tentar conquistar apostadores, ele vem com a intenção de alertá-los sobre os métodos de apostas. Sendo assim, o texto traz um alerta do físico e doutor em matemática Augusto Quadros Teixeira, do instituto de matemática pura e aplicada (IMPA): “Infelizmente, essas fórmulas não podem ajudar quem aposta na Mega-Sena.”

Além do Matemático Dr. Augusto Quadros Teixeira, o artigo cita ainda Ricardo Miranda Martins, do Instituto de Matemática, Estatística e Computação Científica (IMECC) da Universidade Estadual de Campinas (Unicamp) e Hubert Marie Lacoin, do IMPA. Todos eles usando diferentes argumentos a fim de mostrar que o sorteio é equiprovável e que as dicas e métodos ditos como infalíveis não causam nenhum efeito nas probabilidades de ocorrer acerto dos seis números sorteados.

5 A LEI DOS GRANDES NÚMEROS APLICADA À MEGA SENA

Os cálculos mostrados no capítulo anterior foram feitos tomando-se por base que o sorteio da mega-sena se trata de um espaço amostral equiprovável. Porém, analisando os dados dos sorteios, podemos perceber que, atualmente, alguns números foram sorteados cerca de 50% mais vezes que outros. Por exemplo, na atualização de 11/04/2011, o número 5 havia sido sorteado 155 vezes enquanto o número 26 havia sido sorteado em apenas 102 oportunidades.

Esta disparidade de frequência nos chama atenção para a Lei dos Grandes Números. Afinal, será que nesses 1273 sorteios analisados podemos considerar que estamos tratando de um espaço amostral grande suficiente para que a razão da frequência observada de cada evento para o número total de repetições convirja em direção a sua probabilidade real?

Vamos avaliar essa possibilidade considerando que esses muitos sorteios sejam suficientes para que as frequências convirjam para a probabilidade real de cada número ser sorteado, conforme determina a LGN.

O algoritmo que será apresentado para que se possa calcular a probabilidade real levando em consideração a Lei dos Grandes Números é de simples aplicação e entendimento.

Ora, se ao levarmos em consideração que um espaço amostral é equiprovável consideramos cada evento como um subconjunto unitário do espaço amostral, logo de tamanho um, e o espaço amostral terá tamanho igual à soma dos tamanhos de seus subconjuntos unitários, ou seja, seus eventos. Por exemplo, a probabilidade de sair o número 6 num dado não viciado é $1/6$, onde o 1 no numerador é a unidade relativa ao tamanho do subconjunto do evento “sair o número 6” e o 6 a soma dos tamanhos de todos os subconjuntos deste espaço amostral.

Supondo ainda que este dado seja viciado no número 1, de forma que este tenha a probabilidade dobrada de sair. Podemos dizer que o subconjunto do evento “sair o número 1” terá tamanho 2. Sendo assim, a soma dos tamanhos dos subconjuntos do espaço amostral passa a ser 7 e o tamanho do subconjunto do evento “sair o número 6” continua sendo 1. Dessa forma, a probabilidade de sair o número 6 neste dado viciado é de $1/7$. Da mesma forma, que neste dado viciado a probabilidade de sair o número 1 é de $2/7$.

É claro que, no mundo real, para que consigamos uma relação entre as probabilidades de acontecer cada número em um dado viciado deveremos lançar esse dado inúmeras vezes e observar seus resultados e fazendo uma tabela estatística deles. Assim, pela lei dos grandes números, a probabilidade de acontecer cada evento será dada pela frequência relativa desses eventos. Logo, no exemplo acima, a probabilidade de sair o número 1 será de $2/7$, ou seja, aproximadamente, 28,6% e a probabilidade de sair o número 6 será de $1/7$, ou seja, aproximadamente, 14,3%.

Então, pensando como da forma acima, bastaria conseguir uma relação entre as probabilidades não equiprováveis de cada número a ser sorteado.

Para essa relação poderíamos usar como a probabilidade real a própria frequência dos sorteios de cada número, considerando uma quantidade muito grande de repetições do experimento. O que nos traria uma aproximação da probabilidade não equiprovável de cada evento de acordo com a LGN.

A tabela abaixo conta com as frequências relativas de cada número já sorteado, atualizada em 01 de abril de 2018.

Tabela 2 - Frequência de sorteio de cada número da Mega Sena (continua)

NÚMERO	FREQUÊNCIA	FREQUÊNCIA RELATIVA	FREQUÊNCIA EQUIPROVÁVEL
5	233	1,92%	1,67%
53	230	1,89%	1,67%
10	228	1,87%	1,67%
23	225	1,85%	1,67%
4	224	1,84%	1,67%
51	221	1,82%	1,67%
54	221	1,82%	1,67%
24	220	1,81%	1,67%
33	220	1,81%	1,67%
17	218	1,79%	1,67%
52	218	1,79%	1,67%
28	217	1,78%	1,67%
42	216	1,78%	1,67%
32	215	1,77%	1,67%
30	213	1,75%	1,67%
43	212	1,74%	1,67%
16	211	1,73%	1,67%
34	210	1,73%	1,67%
41	210	1,73%	1,67%
2	209	1,72%	1,67%
13	209	1,72%	1,67%
50	209	1,72%	1,67%
29	208	1,71%	1,67%
36	208	1,71%	1,67%
37	208	1,71%	1,67%
56	207	1,70%	1,67%
27	206	1,69%	1,67%
44	206	1,69%	1,67%
6	204	1,68%	1,67%
1	202	1,66%	1,67%
8	202	1,66%	1,67%
18	201	1,65%	1,67%
45	201	1,65%	1,67%
59	201	1,65%	1,67%
35	200	1,64%	1,67%
47	200	1,64%	1,67%
49	200	1,64%	1,67%
12	199	1,64%	1,67%
38	197	1,62%	1,67%
11	196	1,61%	1,67%

20	195	1,60%	1,67%
31	195	1,60%	1,67%
58	195	1,60%	1,67%
46	193	1,59%	1,67%
3	191	1,57%	1,67%
7	191	1,57%	1,67%
40	191	1,57%	1,67%
48	190	1,56%	1,67%
60	190	1,56%	1,67%
14	189	1,55%	1,67%
39	189	1,55%	1,67%
57	188	1,55%	1,67%
19	186	1,53%	1,67%
15	184	1,51%	1,67%
25	184	1,51%	1,67%
9	183	1,50%	1,67%
21	178	1,46%	1,67%
22	177	1,46%	1,67%
55	174	1,43%	1,67%
26	164	1,35%	1,67%
TOTAL	12162	100,00%	100,00%

Nela, podemos notar que a diferença entre as frequências relativas dos números mais e menos sorteados é razoavelmente grande. Pelo pensamento acima, podemos intuir que eles teriam, então, probabilidades diferentes de acontecer. Considerando a probabilidade de cada evento como no caso do dado viciado mostrado acima podemos concluir que a “probabilidade não equiprovável” de o número 5 ser sorteado é de $1{,}92\%$.

Para mostrar que este pensamento, intuitivamente, poderia trazer algum benefício ao jogador, vamos avaliar individualmente cada jogo que pode ser feito.

Por exemplo, o indivíduo que joga nos números 5, 53, 10, 23, 4 e 51, que tem maior frequência sorteada, teria maior possibilidade de acerto do que o que jogar 26,22,46,9,21,39, que foram sorteados, até então, em menos oportunidades.

Assim, podemos calcular a probabilidade de alguns jogos escolhidos aleatoriamente como, por exemplo, qual seria a probabilidade de ocorrer os números 5, 53, 10, 23, 4 e 51.

Traremos da tabela de frequências a probabilidade de sair cada um dos números apostados, sendo $P(A)$ a probabilidade de sair o número A. Desta forma:

$P(5) = 1{,}92\%$

$P(53) = 1{,}89\%$

$P(10) = 1{,}87\%$

$P(23) = 1{,}85\%$

$P(4) = 1{,}84\%$

$P(51) = 1{,}82\%$

Assim, podemos calcular as interseções dessas probabilidades multiplicando-as:

$$P(5 \cap 53 \cap 10 \cap 23 \cap 4 \cap 51) = 1{,}92\% * 1{,}89\% * 1{,}87\% * 1{,}85\% * 1{,}84\% * 1{,}82\% \cong 4{,}2 * 10^{-11}$$

Esse resultado gira em torno de uma chance a cada 25 milhões. Ou seja, pensando em um sorteio não equiprovável, que a quantidade de repetições dos sorteios até 01 de abril de 2018 é grande suficiente para que as frequências relativas girem em torno da probabilidade de cada número sorteado como define a LGN, a chance de vitória dobra jogando os seis números que mais foram sorteados até hoje.

Lembre-se que a probabilidade calculada de vitória com qualquer jogo de seis números considerando o jogo da mega-sena como um espaço amostral equiprovável gira em torno de 52 milhões para 1.

É muito atrativo a qualquer pessoa pensar que considerando a Mega-sena como um espaço não equiprovável, as chances de vitória aumentam e com isso sua sorte pode aumentar. Esse tipo de pensamento é que causa os enganos probabilísticos.

As falácias probabilísticas são erros de interpretação de um evento ou espaço amostral que pode causar enganos nos resultados de um determinado estudo.

O exemplo mostrado acima foi determinado por Kahneman e Tversky como heurística da disponibilidade. As pessoas tendem a acreditar que os resultados que mais aconteceram em um experimento aleatório continuam acontecendo nas próximas repetições deste experimento.

No jogo da Mega Sena, ouvimos muitas coisas sobre a aleatoriedade do jogo quando acontece de serem sorteados uma sequência qualquer de números. A verdade é que, em um sorteio aleatório a probabilidade de sair o número 6 logo após ser sorteado o 5 é tão grande quanto a probabilidade de ser sorteado qualquer outro.

No caso das frequências dos sorteios apresentadas até hoje pode estar faltando quantidade de sorteios. Os 12162 números sorteados até aqui não são suficientes para descrever, segundo a Lei dos Grandes Números uma frequência relativa próxima da probabilidade real de acontecerem os eventos.

Segundo DANIEL KAHNEMAN, a Lei dos Pequenos Números é o pensamento dos pesquisadores que mesmo admitindo uma amostra pequena de certa população para testar uma determinada teoria, essa amostra seria suficiente para descrever os resultados esperados. (KAHNEMAN,2012)

É interessante ressaltar que mesmo com teses corretas, uma amostra pequena da população pode trazer resultados falhos. Veja um exemplo simples de como a Lei dos Pequenos Números pode atrapalhar nossas impressões e modificar sua forma de pensar.

Uma das crenças populares mais difundidas é que se passar por baixo de uma escada terá azar. Alguém que tenha esse dito em mente, tenha passado por baixo de uma escada e lhe tenha acontecido algo de ruim logo em seguida, certamente não o fará nunca mais. Na verdade, somente pelo fato de conhecermos o dito já é motivo para que não passemos por baixo da escada. Porém, a pessoa que o tenha feito e logo em seguida tenha um revés, seja qual for, aleatório ou não, evitará de todas as formas. Mesmo que o fato de passar por baixo da escada não tenha ligação direta com o acontecido.

Acontece que se essa mesma pessoa passasse por baixo da escada todos os dias e anotasse seus eventos bons e ruins do dia. Em uma determinada quantidade de dias, suas impressões sobre esses eventos se aproximariam da probabilidade real que ele tem de sofrer um revés ou de ter um acontecimento favorável no dia.

É claro que o fator psicológico age tendenciosamente nesse caso, mas a falta de testes faz com que a pessoa acredite em relações que não são verdadeiras.

No jogo da Mega-sena acontece o mesmo. Já encontrei muitos apostadores que fazem contas e traçam estratégias para jogar e aumentar suas probabilidades de acerto. Porém, a falta de sorteios suficientes faz com que tiremos a conclusão de que certos números acontecem mais vezes do que outros. Na verdade, quando apresentei minha proposta para meus colegas de mestrado, todos professores de matemática, apresentei a tabela de frequências dos jogos analisados até 2016, percebi que muitos deles anotaram os números que mais foram sorteados. Eles acreditaram que tendo essa informação poderiam levar alguma vantagem na previsão dos números dos próximos sorteios. Lembre-se que este comportamento foi descrito anteriormente como heurística da disponibilidade.

Em rápido e devagar, duas maneiras de pensar, o autor Daniel Kahneman define a Lei dos pequenos números como uma manifestação de viés geral que favorece a certeza sobre a dúvida.

Em um artigo publicado junto com Amos Tversky, Daniel Kahneman atribui à intuição alguns resultados acertados acerca de eventos aleatórios. Neste texto eles alertam sobre acreditar demais na validade de conclusões baseadas em pequenos exemplos.

No mesmo trajeto, Sérgio Rangel, atuário e Mestre em Economia pela UFRGS, especialista em Seguros de Vida pelo SITC/Zurique, acadêmico da Academia Nacional de Seguros e Previdência (ANSP) e consultor sênior da Mirador Atuarial também trata a lei dos pequenos números como meios psicológicos. Em um artigo para a revista apólice, Sérgio diz que os psicólogos e economistas comportamentais definem essa lei como a tendência das pessoas em desprezarem o tamanho das amostras para fins de julgamento e tomada de decisão. Esse viés conduz a conclusões precipitadas baseadas em dados, muitas vezes insuficientes.

Sendo assim, podemos concluir que a Lei dos Pequenos Números pode ser definida como uma tendência que as pessoas têm de acreditar em sequências ou na não aleatoriedade de eventos. Buscar explicação simples e lógicas para o caos e tentar dominar os eventos aleatórios da natureza. Essa tendência então, mostra algo mais voltado ao psicológico e não matemático.

Já foram mostrados no capítulo 2 desse trabalho que as falácias são erros que podem comprometer a saúde ou a liberdade no dia a dia das pessoas, como mostrados e analisados nos exemplos do livro O andar do bêbado.

Os erros de interpretação na Mega-sena não causam grandes resultados negativos para os apostadores enganados, mas trazem-lhes falsas esperanças e fazem com que estes percam tempo analisando resultados anteriores para prever o resultado de um sorteio futuro.

Para comprovar que os sorteios são equiprováveis e que aplicar os pensamentos mostrados anteriormente são erros, vamos seguir dois tópicos: A aplicação gráfica dos sorteios e a análise dos dados através do tempo.

Com essas análises, vamos visualizar de forma simples que quanto mais sorteios são realizados a frequência relativa dos números sorteados se aproximam da probabilidade equiprovável, conforme enunciado pela LGN.

5.1 **- Visualização gráfica da lei dos grandes números aplicada à Mega sena.**

Apesar do algoritmo acima apresentado para que se possa calcular a probabilidade não equiprovável de algum evento ser válido, na aplicação mega-sena, as análises dos gráficos mostram uma tendência de as frequências relativas dos eventos se aproximarem de um determinado valor. Esse valor seria, teoricamente, $1{,}6\%$, calculado como a probabilidade do espaço amostral mega-sena considerando-o como um espaço equiprovável.

Esta tendência nos mostra que apesar da grande quantidade de sorteios já realizados, a frequência relativa ainda não pode representar a probabilidade de ocorrerem os eventos deste experimento.

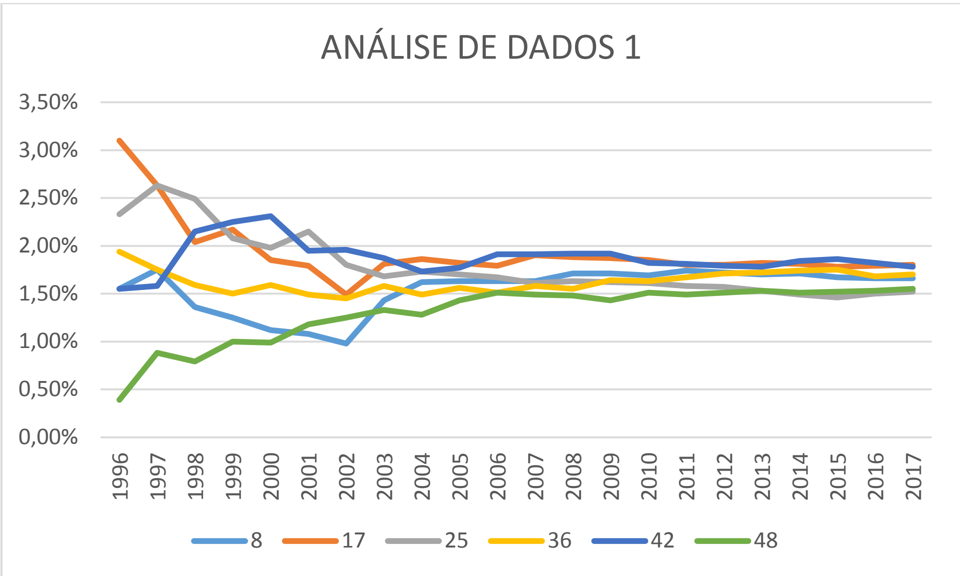
As análises foram feitas de forma aleatória. Utilizando seis números por vez, de forma a mostrar o avanço anual e a tendência das frequências relativas dos eventos se aproximarem ou não de certo valor.

Exemplo 1: Analisando os números: 8, 17, 25, 36, 42, 48.

Os números acima foram escolhidos aleatoriamente com a intenção de analisar o comportamento de suas frequências relativas divididas anualmente para mostrar visualmente a tendência que foi enunciada na lei dos grandes números.

Para fazer essa análise, vamos separar os números que pretendemos estudar e fazer sua contagem relativa anualmente até o último sorteio de 2017. Desta forma poderemos ter uma ideia de como se comporta o sorteio de cada número.

	8	17	25	36	42	48
1996	1,55%	3,10%	2,33%	1,94%	1,55%	0,39%
1997	1,75%	2,63%	2,63%	1,75%	1,58%	0,88%
1998	1,36%	2,04%	2,49%	1,59%	2,15%	0,79%
1999	1,25%	2,17%	2,08%	1,50%	2,25%	1,00%
2000	1,12%	1,85%	1,98%	1,59%	2,31%	0,99%
2001	1,08%	1,79%	2,15%	1,49%	1,95%	1,18%
2002	0,98%	1,49%	1,80%	1,45%	1,96%	1,25%
2003	1,43%	1,81%	1,68%	1,58%	1,87%	1,33%
2004	1,62%	1,86%	1,73%	1,49%	1,73%	1,28%
2005	1,63%	1,82%	1,70%	1,56%	1,77%	1,43%
2006	1,63%	1,79%	1,67%	1,51%	1,91%	1,51%
2007	1,63%	1,90%	1,61%	1,58%	1,91%	1,49%
2008	1,71%	1,88%	1,63%	1,55%	1,92%	1,48%
2009	1,71%	1,87%	1,62%	1,64%	1,92%	1,43%
2010	1,69%	1,85%	1,61%	1,63%	1,82%	1,51%
2011	1,74%	1,80%	1,58%	1,67%	1,81%	1,49%
2012	1,72%	1,80%	1,57%	1,71%	1,79%	1,51%
2013	1,70%	1,82%	1,53%	1,72%	1,78%	1,53%
2014	1,71%	1,81%	1,49%	1,74%	1,84%	1,51%
2015	1,67%	1,78%	1,46%	1,75%	1,86%	1,52%
2016	1,66%	1,79%	1,50%	1,68%	1,82%	1,53%
2017	1,66%	1,80%	1,52%	1,70%	1,78%	1,55%



Repare que, como todos os outros gráficos apresentados abaixo, estes valores representados pelo suposto sorteio dos números 8,17,25,36,42 e 48, mostra uma tendência de suas frequências relativas se aproximarem de algum valor entre $1{,}5\%\ e\ 2\%$.

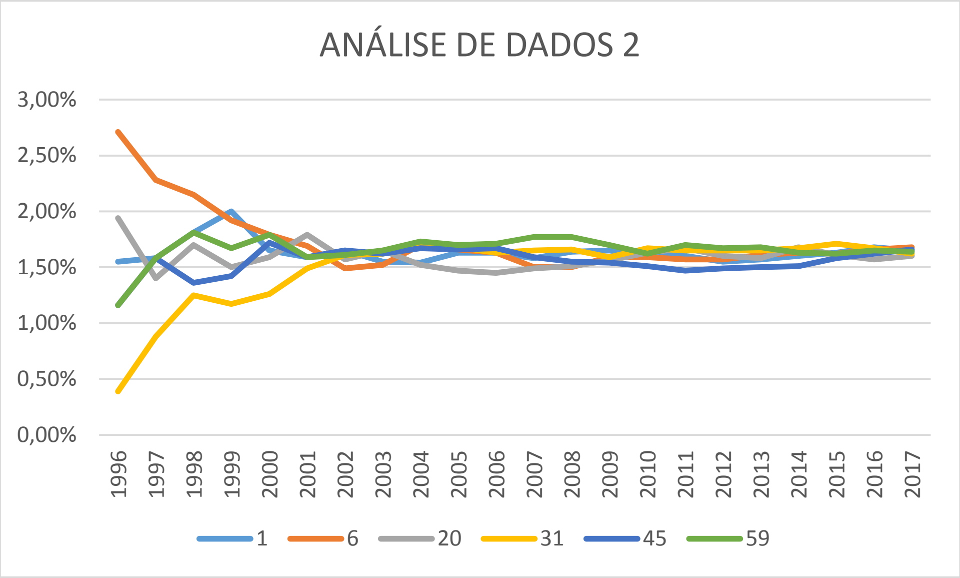
Esse gráfico mostra o que a Lei dos Grandes Números enuncia. Quanto mais vezes um experimento é repetido, mais próximo da probabilidade real se aproximará a frequência relativa dos eventos.

Para mostrar que a impressão de que os valores das frequências relativas se aproximam da probabilidade esperada, vamos fazer mais dois gráficos como exemplo do que está sendo proposto.

Exemplo 2: analisando os números: 1, 6, 20, 31, 45 e 59.

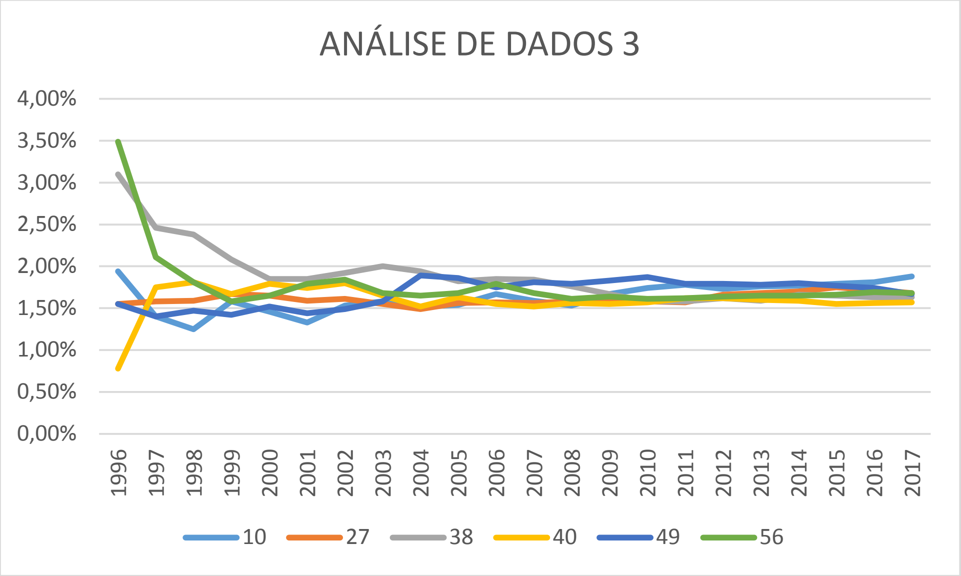
	1	**6**	**20**	**31**	**45**	**59**
1996	1,55%	2,71%	1,94%	0,39%	1,16 %	1,16 %
1997	1,58%	2,28%	1,40%	0,88%	1,58 %	1,58 %
1998	1,81%	2,15%	1,70%	1,25%	1,36 %	1,81 %
1999	2,00%	1,92%	1,50%	1,17%	1,42 %	1,67 %
2000	1,65%	1,79%	1,59%	1,26%	1,72 %	1,79 %
2001	1,59%	1,69%	1,79%	1,49%	1,59 %	1,59 %
2002	1,65%	1,49%	1,57%	1,61%	1,65 %	1,61 %
2003	1,55%	1,52%	1,65%	1,62%	1,62 %	1,65 %
2004	1,54%	1,70%	1,52%	1,70%	1,67 %	1,73 %
2005	1,63%	1,66%	1,47%	1,68%	1,66 %	1,70 %
2006	1,63%	1,63%	1,45%	1,63%	1,67 %	1,71 %
2007	1,58%	1,50%	1,49%	1,65%	1,59 %	1,77 %
2008	1,64%	1,50%	1,51%	1,66%	1,55 %	1,77 %
2009	1,65%	1,59%	1,56%	1,59%	1,54 %	1,70 %
2010	1,61%	1,59%	1,63%	1,67%	1,51 %	1,62 %
2011	1,60%	1,57%	1,67%	1,65%	1,47	1,70

					%	%
2012	1,55%	1,57%	1,60%	1,65%	1,49 %	1,67 %
2013	1,57%	1,61%	1,58%	1,65%	1,50 %	1,68 %
2014	1,60%	1,63%	1,68%	1,67%	1,51 %	1,63 %
2015	1,63%	1,62%	1,61%	1,71%	1,58 %	1,62 %
2016	1,68%	1,66%	1,57%	1,67%	1,62 %	1,65 %



Exemplo 3: Analisando os números: 10, 27, 38, 40, 49 e 56.

	10	27	38	40	49	56
1996	1,94%	1,55%	3,10%	0,78%	1,55 %	3,49 %
1997	1,40%	1,58%	2,46%	1,75%	1,40 %	2,11 %
1998	1,25%	1,59%	2,38%	1,81%	1,47 %	1,81 %
1999	1,58%	1,67%	2,08%	1,67%	1,42 %	1,58 %
2000	1,46%	1,65%	1,85%	1,79%	1,52 %	1,65 %
2001	1,33%	1,59%	1,85%	1,74%	1,44 %	1,79 %
2002	1,53%	1,61%	1,92%	1,80%	1,49 %	1,84 %
2003	1,58%	1,55%	2,00%	1,65%	1,58 %	1,68 %
2004	1,52%	1,49%	1,94%	1,52%	1,89 %	1,65 %
2005	1,54%	1,56%	1,82%	1,63%	1,86 %	1,68 %
2006	1,67%	1,57%	1,85%	1,55%	1,75 %	1,79 %
2007	1,59%	1,56%	1,84%	1,52%	1,81 %	1,68 %
2008	1,53%	1,58%	1,76%	1,56%	1,79 %	1,61 %
2009	1,67%	1,61%	1,67%	1,55%	1,83 %	1,64 %
2010	1,74%	1,58%	1,58%	1,57%	1,87 %	1,61 %
2011	1,78%	1,57%	1,58%	1,62%	1,79 %	1,62 %
2012	1,73%	1,66%	1,62%	1,62%	1,79 %	1,64 %
2013	1,76%	1,68%	1,59%	1,60%	1,78 %	1,65 %
2014	1,77%	1,70%	1,66%	1,59%	1,80 %	1,65 %
2015	1,79%	1,75%	1,65%	1,55%	1,77 %	1,66 %
2016	1,81%	1,72%	1,63%	1,56%	1,74 %	1,69 %
2017	1,88%	1,68%	1,63%	1,57%	1,67 %	1,68 %



Estes gráficos mostram o que foi enunciado acima. Eles mostram que as frequências relativas dos eventos a serem sorteados se aproximam de algum valor entre $1{,}5\%\ e\ 2{,}0\%$.

Mostrando assim, segundo a Lei dos Grandes Números, à ideia de que o sorteio é equiprovável.

Mas, se cada número a ser sorteado tem as mesmas chances de aparecer no sorteio, então, podemos dizer que eles tendem a se aproximar do mesmo valor, no caso, o $1{,}6\%$.

Essa afirmação poderia levar a pensar que os números sorteados que estão com frequência relativa menor que $1{,}6\%$ tendem a ser mais sorteados do que aqueles que estão com as frequências maiores que $1{,}6\%$.

Esse pensamento, nos remete a heurística de representatividade, e foi chamado por Kahneman e Tversky de falácia do jogador.

Portanto, do exposto acima, podemos concluir que neste momento, com essa quantidade de sorteios não podemos usar as frequências relativas como probabilidade. E nem poderia. Afinal, em cada sorteio essas frequências vão se alterando. Caso essa ideia fosse verdadeira em cada sorteio os números teriam probabilidades diferentes de ocorrer.

Mas o que explicaria uma variação entre as frequências dos sorteios?

A **Lei dos Pequenos Números** pode responder essa questão sem maiores problemas.

O erro neste pensamento está em acreditar que a quantidade de repetição já ocorrida neste experimento é suficientemente grande para que a Lei dos Grandes Números garanta a aproximação das frequências relativas às probabilidades reais de cada número ser sorteado.

Apesar dos mais de doze mil números sorteados desde o início dos sorteios da mega-sena, relativamente a quantidade de eventos associados a esse experimento, esse valor não é grande. Veja bem, no experimento de lançar uma moeda para o alto, o espaço amostral tem dois elementos apenas: cara ou coroa. Neste experimento, lançar a moeda doze mil vezes pode ser uma quantidade de repetições suficiente para aproximar a frequência relativa da probabilidade de ocorrer cada evento. Afinal, cada evento ocorrerá aproximadamente 6000 vezes.

Para a mega-sena, hoje, com cerca de doze mil sorteios e com 60 eventos diferentes dentro do espaço amostral, cada evento deve acontecer em torno de 200 vezes.

Perceba que é o que ocorre na avaliação das frequências relativas dos eventos da mega-sena. E que a quantidade de repetição de cada evento é muito maior no lançamento da moeda do que nos sorteios da mega-sena, o que pode explicar as divergências apreciadas nas tabelas.

Um outro erro de análise que pode ocorrer nesse pensamento é acreditar que os eventos que ocorreram em menor frequência tendem a acontecer com mais frequência a partir de um dado ponto analisado.

Este pensamento também não pode ser encarado como verdade. A Lei dos Grandes Números enuncia a aproximação da frequência relativa dos eventos a partir da grande quantidade de repetições do experimento, mas não necessariamente serão aproximadas as quantidades reais dos resultados. A cada repetição do experimento as probabilidades são as mesmas, o que acontece é que a cada repetição as diferenças relativas de acontecimentos anteriores tendem a diminuir.

Para exemplificar esse pensamento, vamos analisar as diferenças reais e relativas de dois números sorteados através do tempo.

Curiosamente, o número 5 é o mais sorteado há alguns anos e o número 26 o menos sorteado também há muitos anos. Sendo assim, vamos analisar a tabela abaixo que mostra esses eventos no tempo:

FREQUÊNCIA RELATIVA				FREQUÊNCIA REAL			
ANO	5	26	Diferença	ANO	5	26	Diferença
1996	2,71%	0,78%	1,93%	1996	7	2	5
1997	2,11%	1,05%	1,06%	1997	12	6	6
1998	2,04%	1,25%	0,79%	1998	18	11	7
1999	2,08%	1,33%	0,75%	1999	25	16	9
2000	1,79%	1,46%	0,33%	2000	27	22	5
2001	2,05%	1,59%	0,46%	2001	40	31	9
2002	2,20%	1,45%	0,75%	2002	56	37	19
2003	2,12%	1,39%	0,73%	2003	67	44	23
2004	1,99%	1,25%	0,74%	2004	75	47	28
2005	1,95%	1,29%	0,66%	2005	85	56	29
2006	1,93%	1,33%	0,60%	2006	96	66	30
2007	1,86%	1,31%	0,55%	2007	104	73	31
2008	1,96%	1,30%	0,66%	2008	122	81	41
2009	1,94%	1,23%	0,71%	2009	133	84	49
2010	2,01%	1,29%	0,72%	2010	150	96	54
2011	2,04%	1,33%	0,71%	2011	165	108	57
2012	2,03%	1,34%	0,69%	2012	177	117	60
2013	1,96%	1,37%	0,59%	2013	183	128	55
2014	1,97%	1,37%	0,60%	2014	197	137	60
2015	1,94%	1,35%	0,59%	2015	207	144	63
2016	1,96%	1,38%	0,58%	2016	222	156	66
2017	1,93%	1,37%	0,56%	2017	232	164	68

A análise que pode ser feita através dessa tabela é que a diferença entre as frequências relativas entre dois eventos foi diminuindo através do tempo, enquanto que a diferença real de ocorrência desses mesmos eventos aumentou no mesmo período.

Fica claro ao analisar os anos de 2016 e 2017. Em 2017, o evento 5 ocorreu 2 vezes a mais que o evento 26. Visto que a diferença entre eles aumentou em duas unidades. Mesmo assim, a diferença entre suas frequências relativas diminuiu de 0,58% para 0,56%.

É desse jeito que a ideia da Lei dos Grandes Números afirma que as frequências relativas tendem a se aproximar da probabilidade real. Aumentando as repetições de um experimento, as frequências se aproximam da probabilidade.

Kahneman, em seu livro Rápido e devagar, duas formas de pensar, trouxe uma frase que se encaixa perfeitamente ao jogo da mega-sena: *“Um evento aleatório, por definição não se presta a explicação, mas grupos de eventos aleatórios se comportam de um modo altamente regular.”* (Kahneman, 2012). Essa frase nos mostra que qualquer tipo de intenção de adivinhar resultados futuros de um experimento aleatório é mera perda de tempo, e que apesar das imperfeições dos resultados aleatórios passados, no contexto de eventos independentes, o próximo evento terá sempre as mesmas probabilidades de ocorrer.

CONCLUSÃO

O trabalho apresentado trouxe algumas visões dos erros que as pessoas cometem ao avaliar as probabilidades de alguns experimentos aleatórios. As heurísticas e vieses apresentados por Kahneman e Tversky podem explicar alguns dos maiores erros cometidos pelas pessoas ao tentar inferir a ocorrência de algum evento analisando os eventos passados. Essas heurísticas são as explicações dos erros cometidos quando tratamos da lei dos pequenos números.

Não é incomum que pessoas cometam erros por acreditar que as frequências relativas dos eventos em uma pequena quantidade de resultados de um determinado experimento possam representar as probabilidades desse experimento. Isso pode ser visto em inúmeros casos, e em algumas teorias dos apostadores.

Durante as várias horas de conversa sobre a mega-sena, encontrei muitos apostadores que fazem contas e traçam estratégias para jogar e aumentar suas probabilidades de acerto. Porém, a falta de sorteios suficientes e o desconhecimento faz com que tirem a conclusão de que certos números acontecem mais vezes do que outros ou até mesmo que certos números têm maior chance de acontecer em determinado sorteio. Na verdade, quando apresentei minha proposta para meus colegas de mestrado, todos professores de matemática, apresentei a tabela de frequências dos jogos analisados até 2016, percebi que muitos deles anotaram os números que mais foram sorteados (heurística da disponibilidade), enquanto outros anotaram os que menos foram sorteados (heurística da representatividade). Eles acreditaram que tendo essas informações poderiam levar alguma vantagem na previsão dos números dos próximos sorteios.

Parece impressionante que mentes informadas e formadas para cálculos probabilísticos sejam ludibriados pela possibilidade de vitória analisando os dados anteriores de um experimento aleatório e equiprovável. Essa ideia faz com que muitos apostadores gastem muito dinheiro com o sonho de ficar milionário.

São erros comuns de ocorrer. Como visto durante o trabalho esses tipos de erros já aconteceram em muitos outros casos semelhantes ou não. Pessoas foram condenadas por erros de interpretação da probabilidade, uma organização canadense cometeu o grave erro de ignorar uma probabilidade que parecia muito pequena e ainda um caso gravíssimo em que o médico deu um diagnóstico errado de uma doença incurável acreditando que a probabilidade de um exame estar errado é praticamente nula.

Estes erros normalmente são causados por falha na interpretação correta do espaço amostral a que o experimento está se referindo. De qualquer modo, erro no cálculo teórico de probabilidade, falha na interpretação do espaço amostral ou de interpretação de resultados estatísticos apresentados são falhas que podem trazer nos trazer problemas.

E em muitos casos não dependem necessariamente do conhecimento acadêmico, conforme apresentado no trabalho de Adriana Sbicca (2014). Afinal, se pessoas com um grau de instrução considerável e com posições importantes na sociedade cometem erros probabilísticos graves, é claro que o apostador comum também está sujeito a cometer tais erros.

Esse trabalho se prestou a explicar alguns dos possíveis erros que ocorrem quando um jogador procura adivinhar a ocorrência dos eventos dos próximos sorteios. E alertar aos que buscam métodos práticos e rápidos para acertar a ocorrência dos seis próximos números da mega-sena que não há método eficaz de adivinhar quais os próximos números sorteados.

Além disso, cabe ressaltar que os gráficos apresentados no capítulo 5 nos mostram claramente que as frequências de ocorrências dos números a serem sorteados neste experimento se aproximam em muito no que se espera de um experimento equiprovável. E que apesar de não possuirmos sorteios suficientes para que a Lei dos grandes números nos mostre as probabilidades reais de ocorrência de todos os números, podemos afirmar que elas estão bem próximas de serem iguais.

Portanto, é possível inferir que a mega-sena é um experimento aleatório e equiprovável, e que qualquer tipo de tentativa de adivinhar os próximos números a serem sorteados será perda de tempo.

É claro que existe um meio de ter mais chance de acertar os seis números, porém isso só será possível caso o apostador faça mais tentativas, e isso inclui investir mais no jogo. Quanto mais jogos, maior a chance de acertar. Cabe ao apostador decidir se vale a pena tamanho investimento.

REFERÊNCIAS

HAZZAN, Samuel. Fundamentos de Matemática Elementar. Atual Editora. 2004.

MLODNOW, Leoanard . O andar do bêbado. Editora Jorge Zahar. 2009.

LOBÃO, Júlio. Finanças comportamentais: quando a economia encontra a psicologia. Editora actual. 2014.

KAHNEMAN, Daniel. Rápido e devagar, duas formas de pensar. Editora objetiva. 2011.

RABIN, Matthew. Inference by Believers in law of Small Numbers. Oxford Uniersity Press,2002.

SBICCA, Adriana. A contribuição de Daniel Kahneman e Amos Tversky para o estudo das decisões econômicas. Departamento de Economia – UFPR.

TVERSKY, Amos. KAHNEMAN, Daniel. Availability: A heuristic for judging frequency and probability. Hebrew University of jerusalem and the oregon research institute. 1973.

SBICCA, Adriana. Heurísticas no estudo das decisões econômicas: Contribuições de Herbert Simon, Daniel Kahneman e Amos Tversky. Universidade Federal do Paraná – UFPR. 2013.

TONETTO, Miletto et all. O papel das heurísticas no julgamento e na tomada de decisão sob incerteza. Estudos de Psicologia. Pontifícia Universidade Católica de Campinas. 2006.

TVERSKY, Amos. KAHNEMAN, Daniel. On the psychology of prediction. Hebrew University of jerusalem and the oregon research institute. 1973.

SILVA, Ismael de Araújo. Probabilidades: a visão laplaciana e a visão frequentista na introdução do conceito. Dissertação apresentada à banca examinadora da Pontifícia Universidade Católica de São Paulo - PUC-SP. 2002.

FIOLHAIS, Carlos. A matemática e a física do azar. Revista portuguesa de psicanálise. Vol. 29. Pág, 121 a 127. 2009.

SAMPAIO, Murilo de Medeiros. Lei dos grandes números da percolação multi-dimensional. Dissertação submetida a banca a Universidade Federal de Pernambuco. 2007.

SANTOS, Jorian Pereira. A teoria da probabilidade e a teoria dos jogos em uma abordagem para o ensino médio. Universidade Federal do Rio Grande do Norte. 2016.

ROLLA, Leonardo T. Introdução a probabilidade. PUC-Rio. 2012.

LOPES, Eric. Lei dos grandes números em sistemas aparentemente aleatórios. Instituto de matemática, estatística e computação cientifica. Universidade estadual de Campinas.

RANGEL, Sérgio. Pequenos números, grandes erros. <https://www.revistaapolice.com.br/2012/05/pequenos-numeros-grandes-erros>. Acessado em: 17 de abril de 2018.

A probabilidade matemática pode prever números de loteria mais prováveis. <https://www.numerosmegasena.com/probabilidade-matematica-pode-prever-numeros-de-loteria-mais-provaveis/. Acessado em: 17 de abril de 2018.

<http://loterias.caixa.gov.br/wps/portal/loterias/landing/megasena>. Acessado em: 01 de abril de 2018.

MARINI, Eduardo. Especialista dá dicas para se dar bem e faturar o prêmio da mega-sena. Portal R7. <https://noticias.r7.com/brasil/especialista-da-dicas-para-se-dar-bem-e-faturar-premio-da-mega-sena-01032014> . Acessado em: 25 de maio de 2018.

BBC. O que dizem os matemáticos sobre as ditas fórmulas certeiras e dicas infalíveis para se ganhar na loteria. Portal G1. <https://g1.globo.com/loteria/noticia/o-que-dizem-os-matematicos-sobre-as-ditas-formulas-certeiras-e-dicas-infaliveis-para-se-ganhar-na-loteria.ghtml> . Acessado em: 25 de maio de 2018.