Análise Estatística da Mega-Sena

LT

Abstract

Este artigo apresenta, de forma sistemática e detalhada, os fundamentos de probabilidade e estatística aplicados à Mega-Sena, incluindo cálculos de probabilidades para as faixas de premiação, análise de sequências específicas como $\{1, 2, 3, 4, 5, 6\}$ e a inclusão do princípio da casa dos pombos. Todos os resultados são obtidos através de combinações e fórmulas matemáticas em LaTeX.

1 Introdução

A Mega-Sena é a principal loteria oficial do Brasil, administrada pela Caixa Econômica Federal desde 1996. A cada sorteio, seis bolas numeradas de 1 a 60 são extraídas sem reposição, e os apostadores que acertam quatro, cinco ou seis números (quadra, quina ou sena) recebem prêmio em dinheiro. O objetivo deste artigo é:

- Definir formalmente o espaço amostral e os eventos de interesse.
- Calcular, usando combinações, as probabilidades de quadra, quina e sena.
- Demonstrar que qualquer combinação fixa de seis números (inclusive $\{1, 2, 3, 4, 5, 6\}$) tem probabilidade igual a $1/\binom{60}{6}$.
- Apresentar o *princípio da casa dos pombos* e discutir sua aplicação a conjuntos de números na Mega-Sena.
- Analisar o valor esperado de retorno para o jogador em uma aposta simples.

2 Regras Básicas da Mega-Sena

No volante oficial:

- O jogador escolhe k números, com $6 \leq k \leq 15$, dentre os inteiros de 1 a 60.
- Cada seleção de 6 números (jogo mínimo) custa um valor fixo (por exemplo, R$4,50). Marcar mais de 6 números aumenta o número de apostas implícitas, pois todas as combinações de 6 números dentre os k marcados são consideradas apostas independentes.
- São realizados dois sorteios semanais (quartas e sábados). Em cada sorteio, extraem-se 6 bolas sem reposição dentre 60. A ordem de extração é irrelevante para definir o conjunto sorteado.
- As faixas de premiação são:

Sena: acerto dos 6 números;

Quina: acerto de exatamente 5 dígitos dentre os 6 sorteados;

Quadra: acerto de exatamente 4 dígitos dentre os 6 sorteados.

3 Espaço Amostral e Eventos

Denote-se por Ω o espaço amostral de todas as combinações de 6 números dentre 60, sem considerar permutações. Assim,

$$\Omega \;=\; \big\{\{a_1, a_2, a_3, a_4, a_5, a_6\} : 1 \le a_1 < a_2 < \cdots < a_6 \le 60\big\}.$$

O número total de combinações é

$$|\Omega| \;=\; \binom{60}{6} \;=\; \frac{60!}{6!\,54!} \;=\; 50\,063\,860.$$

Sejam E_k o evento "acertar exatamente k números" em uma aposta simples (marcação de 6 números). Em particular:

$$E_6 = \{\text{sena}\}, \quad E_5 = \{\text{quina}\}, \quad E_4 = \{\text{quadra}\}.$$

Sob a hipótese de equiprobabilidade, a probabilidade de cada evento é

$$P(E_k) = \frac{|E_k|}{|\Omega|}.$$

4 Probabilidade de Acertar a Sena (E_6)

Quando se faz uma aposta simples de 6 números — digamos $A = \{a_1, \ldots, a_6\}$ — há *exatamente uma* combinação sorteada que coincide com A. Portanto,

$$|E_6| = 1 \quad\Longrightarrow\quad P(\text{sena}) = P(E_6) = \frac{1}{\binom{60}{6}} = \frac{1}{50\,063\,860} \approx 1{,}998 \times 10^{-8}.$$

5 Probabilidades de Quina (E_5) e Quadra (E_4)

5.1 Evento Quina (E_5)

Para acertar exatamente 5 dos 6 números apostados, deve-se:

1. Escolher quais 5 dos 6 números da aposta coincidirão com o sorteio: $\binom{6}{5}$ maneiras.
2. Escolher 1 número dentre os restantes ($60 - 6 = 54$), que estará no sorteio mas não na aposta: $\binom{54}{1}$ maneiras.

Logo,

$$|E_5| = \binom{6}{5}\binom{54}{1} = 6 \times 54 = 324,$$

e

$$P(\text{quina}) = \frac{324}{\binom{60}{6}} = \frac{324}{50\,063\,860} \approx 6{,}470 \times 10^{-6}.$$

5.2 Evento Quadra (E_4)

Para acertar exatamente 4 dos 6 números apostados, tem-se:

1. Escolher quais 4 dos 6 números da aposta estarão no sorteio: $\binom{6}{4}$ maneiras.
2. Escolher 2 números dentre os 54 restantes (que não fazem parte da aposta) para completar os 6 sorteados: $\binom{54}{2}$ maneiras.

Portanto,

$$|E_4| = \binom{6}{4}\binom{54}{2} = 15 \times \binom{54}{2} = 15 \times \frac{54 \times 53}{2} = 15 \times 1\,431 = 21\,465,$$

e

$$P(\text{quadra}) = \frac{21\,465}{\binom{60}{6}} = \frac{21\,465}{50\,063\,860} \approx 4{,}287 \times 10^{-4}.$$

6 Probabilidade de Combinações Específicas (ex.: $\{1, 2, 3, 4, 5, 6$

Não existe qualquer distinção probabilística entre as $\binom{60}{6}$ combinações possíveis. Assim, para a sequência aritmética exata

$$\{1, 2, 3, 4, 5, 6\},$$

tem-se

$$P\big(\{1, 2, 3, 4, 5, 6\}\big) = \frac{1}{\binom{60}{6}} = \frac{1}{50\,063\,860}.$$

Qualquer outra sextina fixa, seja ela “aleatória” ou “padrão”, possui a mesma probabilidade. Em geral, se quisermos contar quantas sequências aritméticas de 6 termos existem em $\{1, 2, \dots, 60\}$, observamos:

Seja uma progressão aritmética de 6 termos definida por

$$a_1,\ a_1 + r,\ a_1 + 2r,\ \dots,\ a_1 + 5r,$$

com $r \geq 1$. Para que $a_1 + 5r \leq 60$, precisamos de

$$1 \leq a_1 \leq 60 - 5r.$$

Como r pode variar em $\{1, 2, \dots, \lfloor (60-1)/5 \rfloor\} = \{1, 2, \dots, 11\}$, o número de progressões aritméticas de 6 termos é

$$\sum_{r=1}^{11}(60 - 5r) = 11 \times 60 - 5\sum_{r=1}^{11} r = 660 - 5 \times 66 = 660 - 330 = 330.$$

Portanto, há exatamente 330 progressões aritméticas de 6 termos em $\{1, \dots, 60\}$, cada uma delas com probabilidade $1/\binom{60}{6}$. Se quisermos a probabilidade de que *alguma* dessas 330 seja sorteada, obtemos

$$P\big(\text{progressão aritmética de 6}\big) = \frac{330}{\binom{60}{6}} \approx 6{,}59 \times 10^{-6}.$$

7 Princípio da Casa dos Pombos

[Princípio da Casa dos Pombos] Sejam n "pombos" e m "gaiolas". Se $n > m$, então, ao menos uma gaiola conterá pelo menos dois pombos.

Demonstração (esquemática). Suponha, por contradição, que cada uma das m gaiolas contenha no máximo um pombo. Então, no máximo, podemos colocar m pombos, um em cada gaiola. Mas como se deseja colocar $n > m$ pombos, haverá pelo menos um pombo que não encontrará "gaiola livre" e, portanto, alguma gaiola acabará contendo dois ou mais pombos. □ □

7.1 Aplicação ao Espaço de Números da Mega-Sena

Embora o princípio da casa dos pombos seja frequentemente apresentado em contextos discretos gerais, podemos ilustrar sua aplicação em problemas relacionados a números sorteados:

Considere os 60 inteiros de 1 a 60, particionados em 6 *classes de restos* na divisão por 6:

$$C_j = \{\, x \in \{1, \dots, 60\} : x \equiv j \pmod 6\}, \quad j = 1, 2, 3, 4, 5, 0.$$

Cada classe C_j tem exatamente 10 elementos (por exemplo, $C_1 = \{1, 7, 13, \dots, 55\}$). Se um apostador escolhe 7 números distintos dentre 1 a 60, aplicando o princípio da casa dos pombos (com 7 pombos e 6 gaiolas), obtemos:

$$7 > 6 \quad \Longrightarrow \quad \exists\, j \text{ tal que pelo menos dois dos 7 números estão na mesma classe } C_j.$$

Em particular, ao marcar 7 números, garantimos que existem ao menos dois que têm o mesmo resto na divisão por 6.

Este tipo de argumento pode ser útil quando se analisa padrões de distribuição em classes modulares. Embora não altere a equiprobabilidade das combinações sorteadas, fornece garantias combinatórias sobre qualquer seleção de números.

8 Valor Esperado para o Jogador

Em cada concurso, a premiação total (fundo de prêmios) corresponde a 46% da arrecadação bruta. Desse montante de 46%:

- 35% são destinados à *sena.*
- 19% são destinados à *quina.*
- 19% são destinados à *quadra.*
- 22% vão para o fundo de acumulação (rolagem/mega acumulado).
- 5% são repassados a programas sociais da União.

Suponha que, em um concurso hipotético:

$$\text{Arrecadação total} = R\$100\,000\,000.$$

Então:

$$\text{Prêmio bruto} = 0{,}46 \times 100\,000\,000 = 46\,000\,000 \text{ (em R\$)}.$$

Destinações:

$$\begin{aligned}
\text{Sena: } & 0{,}35 \times 46\,000\,000 = 16\,100\,000,\\
\text{Quina: } & 0{,}19 \times 46\,000\,000 = 8\,740\,000,\\
\text{Quadra: } & 0{,}19 \times 46\,000\,000 = 8\,740\,000.
\end{aligned}$$

Suponha ainda que haja:

2 ganhadores da sena, 50 ganhadores da quina e 2000 ganhadores da quadra.

Então, os valores unitários médios para cada faixa são:

$$v_6 = \frac{16\,100\,000}{2} = 8\,050\,000, \qquad v_5 = \frac{8\,740\,000}{50} = 174\,800, \qquad v_4 = \frac{8\,740\,000}{2000} = 4\,370.$$

Seja $a = 4{,}50$ (custo de uma aposta simples de 6 números). A probabilidade de cada faixa, em uma aposta simples, é:

$$P(\text{sena}) = \frac{1}{50\,063\,860}, \quad P(\text{quina}) = \frac{324}{50\,063\,860}, \quad P(\text{quadra}) = \frac{21\,465}{50\,063\,860}.$$

O valor esperado $E[V]$ (antes de descontar o custo da aposta) é:

$$E[V] = v_6\,P(\text{sena}) + v_5\,P(\text{quina}) + v_4\,P(\text{quadra}).$$

Substituindo valores:

$$\begin{aligned}
v_6\,P(\text{sena}) &= 8\,050\,000 \times \frac{1}{50\,063\,860} \approx 0{,}1608,\\
v_5\,P(\text{quina}) &= 174\,800 \times \frac{324}{50\,063\,860} = 174\,800 \times 6{,}470 \times 10^{-6} \approx 1{,}1311,\\
v_4\,P(\text{quadra}) &= 4\,370 \times \frac{21\,465}{50\,063\,860} = 4\,370 \times 4{,}287 \times 10^{-4} \approx 1{,}8730.
\end{aligned}$$

Logo,

$$E[V] \approx 0{,}1608 + 1{,}1311 + 1{,}8730 \approx 3{,}1649 \text{ (R\$)}.$$

Subtraindo o custo $a = 4{,}50$,

$$\text{Retorno} = E[V] - a \approx 3{,}1649 - 4{,}50 = -1{,}3351 \text{ R\$},$$

isto é, expectativa negativa de aproximadamente R$1,34 por aposta simples.

9 Acúmulo de Prêmios e Probabilidade de Não Haver Ganhador da Sena

Sejam:

$$p = P(\text{sena}) = \frac{1}{\binom{60}{6}}, \quad n = \text{número de apostas simples de 6 números.}$$

A probabilidade de que *nenhuma* aposta acerte a sena em um concurso é

$$P(\text{nenhum acerta a sena}) = (1-p)^n.$$

Por exemplo, se $n = 30\,000\,000$, então

$$(1-p)^n \approx \left(1 - \frac{1}{50\,063\,860}\right)^{30\,000\,000} \approx \exp\Big(-30\,000\,000/50\,063\,860\Big) \approx e^{-0{,}5996} \approx 0{,}5493.$$

Ou seja, mesmo com 30 milhões de apostas, a chance de não haver acertador da sena é de cerca de 55%. Quando isso ocorre, o valor destinado à sena (16,1% da arrecadação) "rola" para o próximo concurso, criando o famoso "mega acumulado".

10 Simulações Monte Carlo

Para ilustrar numericamente a concordância teoria–prática, podemos realizar simulações de Monte Carlo, repetindo N experimentos de sorteio:

1. Para cada experimento $i = 1, 2, \ldots, N$:
   (a) Gerar, aleatoriamente, um conjunto $S_i = \{s_{i1}, \ldots, s_{i6}\}$ de 6 números distintos em $\{1, \ldots, 60\}$.
   (b) Para cada aposta (ou para um conjunto de m apostas fixas), contar quantos acertos ocorrem (sena, quina, quadra).
2. Após N experimentos, calcular as frequências relativas:

$$\hat{P}(\text{sena}) = \frac{\#\{\text{sena simuladas}\}}{N}, \quad \hat{P}(\text{quina}) = \frac{\#\{\text{quina simuladas}\}}{N}, \quad \hat{P}(\text{quadra}) = \frac{\#\{\text{quadra si}}{N}$$

Para N suficientemente grande, observaremos que $\hat{P}$ aproxima-se muito bem dos valores teóricos de Seções 4 e 5.

11 Teste de Qui-Quadrado para Uniformidade de Números

Considere um histórico de n concursos, e seja O_i a frequência absoluta de ocorrência do número $i \in \{1, \ldots, 60\}$ nesses n sorteios. Pelo modelo de sorteios idêntica e independentemente distribuídos, cada número i deveria aparecer cerca de $n \times (6/60) = n/10$ vezes. Defina-se

$$\mu_i = \frac{n}{10}, \quad \sigma_i^2 = n \times \frac{6}{60} \times \left(1 - \frac{6}{60}\right) = \frac{9\,n}{100}.$$

O estatístico de Qui-Quadrado é

$$X^2 = \sum_{i=1}^{60} \frac{(O_i - \mu_i)^2}{\mu_i}.$$

Sob a hipótese nula de uniformidade, X^2 aproxima-se de uma distribuição χ^2 com 59 graus de liberdade. Se $X^2 > \chi^2_{0{,}95}(59) \approx 78{,}11$, rejeita-se a hipótese de equiprobabilidade.

12 Conclusão

Neste artigo em LaTeX, abordamos:

- A definição do espaço amostral Ω para a Mega-Sena e o cálculo exato de $|\Omega| = \binom{60}{6}$.
- As probabilidades de sena, quina e quadra em uma aposta simples de 6 números:

$$P(\text{sena}) = \frac{1}{50\,063\,860}, \quad P(\text{quina}) \approx 6{,}470 \times 10^{-6}, \quad P(\text{quadra}) \approx 4{,}287 \times 10^{-4}.$$

- A demonstração de que qualquer combinação fixa de 6 números (incluindo $\{1, 2, 3, 4, 5, 6\}$) possui probabilidade igual a $1/\binom{60}{6}$.

- O enunciado e uma breve aplicação do princípio da casa dos pombos, mostrando que, ao selecionar 7 números dentre 60, pelo menos dois compartilham o mesmo resto em $\bmod 6$.
- O cálculo do valor esperado do prêmio para o jogador, demonstrando a expectativa negativa em uma aposta simples.
- A logística de acumulação de prêmios (rolagem) quando não há acertador da sena, com probabilidade aproximada de 55% de "nenhum seis" quando existem 30 milhões de apostas simples.
- Metodologia de simulação Monte Carlo e teste de Qui-Quadrado para avaliar uniformidade de sorteios históricos.

Portanto, o leitor detém agora as ferramentas matemáticas em LaTeX para compreender, calcular e criticar, com precisão, as probabilidades e os riscos relacionados à Mega-Sena, bem como a aplicabilidade do princípio da casa dos pombos em contextos de combinação de números.