

# House of Excellence

**Consultores Responsáveis:**

João Gabriel Bertoli Rola

**Requerente:**

João Victor Neves

Brasília, 7 de novembro de 2024.



## Sumário

	Página
1 Introdução . . . . .	3
2 Referencial Teórico . . . . .	4
2.1 Média . . . . .	4
2.2 Mediana . . . . .	4
2.3 Quartis . . . . .	4
2.4 Variância . . . . .	5
2.4.1 Variância Populacional . . . . .	5
2.5 Desvio Padrão . . . . .	5
2.5.1 Desvio Padrão Populacional . . . . .	6
2.6 Boxplot . . . . .	6
2.7 Histograma . . . . .	6
2.8 Gráfico de Dispersão . . . . .	7
2.9 Coeficiente de Correlação de Pearson . . . . .	7
3 Análises . . . . .	8
3.1 Analise 1 . . . . .	8
3.1.1 TOP 5 PAÍSES COM MAIS MEDALHISTAS DAS OLIM- PIADAS DE SYDNEY(2000) - RIO DE JANEIRO (2016)	8
3.2 Analise 2 . . . . .	9
3.2.1 RELAÇÃO ENTRE IMC EM DIFERENTES ESPORTES	9
3.3 Analise 3 . . . . .	10
3.3.1 TOP 3 MEDALHISTAS DAS OLIMPIADAS 2000-2016 .	10
3.4 Analise 4 . . . . .	11
3.4.1 Analise da relação de peso por altura . . . . .	11
4 Conclusões . . . . .	13

# 1 Introdução

Este relatório apresenta uma análise estatística detalhada dos dados coletado para entender e otimizar o desempenho dos atletas da House of Excellence durante as olimpíadas de 2000 até 2016. Este relatório possui análises descritivas como média, mediana, variância e análise de correlação também com o uso do coeficiente de Pearson.

A coleta de dados foi feita pelo próprio cliente e é uma amostra não probabilística. Esse banco de dados contém variáveis sobre : Esporte, peso, idade, modalidade que participou do esporte, altura, medalha além do nome dos atletas .

As análises deste relatório foram feitas exclusivamente pelo R studio na versão R version 4.3.3 (2024-02-29 ucrt) e o relatório foi gerado a partir do Quarto.

## 2 Referencial Teórico

### 2.1 Média

A média se constitui como a razão entre a soma do valor das observações e o número total delas, dada pela fórmula:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Com:

- $i = 1, 2, \dots, n$
- $n =$  número total de observações

### 2.2 Mediana

Sejam as  $n$  observações de um conjunto de dados  $X = X_{(1)}, X_{(2)}, \dots, X_{(n)}$  de determinada variável ordenadas de forma crescente. A mediana do conjunto de dados  $X$  é a medida que divide o valor das observações ao meio, de modo que metade delas tenham valor menor que a mediana e a outra metade, maior.

Com isso, pode-se calcular a mediana da seguinte forma:

$$med(X) = \begin{cases} X_{\frac{n+1}{2}}, & \text{para } n \text{ ímpar;} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}, & \text{para } n \text{ par.} \end{cases}$$

### 2.3 Quartis

Os quartis são separatrizes que dividem o conjunto de dados em quatro partes iguais. O primeiro quartil, também chamado de inferior, é a medida que delimita os 25% menores valores. Já o segundo quartil representa a mediana e é o valor que ocupa a posição central, isto é, metade dos dados estão abaixo dela e a outra metade está acima. Por último, o terceiro quartil delimita os 75% menores valores. Inicialmente deve-se calcular a posição do quartil:

- Posição do primeiro quartil  $P_1$ :

$$P_1 = \frac{n + 1}{4}$$

- Posição da mediana (segundo quartil)  $P_2$ :

$$P_2 = \frac{n + 1}{2}$$

- Posição do terceiro quartil  $P_3$ :

$$P_3 = \frac{3 \times (n + 1)}{4}$$

Com  $n$  sendo o tamanho da amostra. Dessa forma,  $X_{(P_i)}$  é a o valor do  $i$ -ésimo quartil, onde  $X_{(j)}$  representa a  $j$ -ésima observação dos dados ordenados.

Se o cálculo da posição resultar em uma fração deve-se fazer a média entre o valor que está na posição do inteiro anterior e do seguinte ao da posição.

## 2.4 Variância

A variância é uma medida que avalia o quanto que os dados estão dispersos em relação à média, em uma escala ao quadrado da escala dos dados que dificulta a interpretação dessa medida.

### 2.4.1 Variância Populacional

Para uma população, a variância é dada por:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Com:

- $X_i$  =  $i$ -ésima observação da população
- $\mu$  = média populacional
- $N$  = tamanho da população

## 2.5 Desvio Padrão

O desvio padrão é a raiz quadrada da variância, haja vista a ideia de retirar a escala ao quadrado da variância para que se tenha uma medida mais facilmente interpretável. Avalia o quanto os dados estão dispersos em relação à média.

### 2.5.1 Desvio Padrão Populacional

Para uma população, o desvio padrão é dado por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

Com:

- $X_i$  = i-ésima observação da população
- $\mu$  = média populacional
- $N$  = tamanho da população

## 2.6 Boxplot

O boxplot é uma representação gráfica na qual se pode perceber de forma mais clara como uma variável quantitativa está distribuída. A figura abaixo ilustra um exemplo de boxplot.

Figura 1: Exemplo de boxplot



A parte inferior do retângulo corresponde ao primeiro quartil, enquanto a parte superior representa o terceiro quartil. O traço dentro do retângulo indica a mediana, que divide o conjunto de dados em duas partes de tamanhos iguais. A média é ilustrada por um losango branco, e os pontos representam os *outliers*. *Outliers* são valores discrepantes da série de dados, ou seja, valores que não refletem a realidade do conjunto.

## 2.7 Histograma

O histograma é uma representação gráfica utilizada para a visualização da distribuição de variáveis quantitativas e pode ser construído por valores absolutos, frequência relativa ou densidade. A figura abaixo ilustra um exemplo de histograma.

Figura 2: Exemplo de histograma



## 2.8 Gráfico de Dispersão

O gráfico de dispersão é uma representação gráfica utilizada para ilustrar o comportamento conjunto de duas variáveis quantitativas. A figura abaixo ilustra um exemplo de gráfico de dispersão, onde cada ponto representa uma observação do banco de dados.

Figura 3: Exemplo de Gráfico de Dispersão



## 2.9 Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson é uma medida que verifica o grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor zero significa que não há relação linear entre as variáveis. Quando o valor do coeficiente  $r$  é negativo, diz-se existir uma relação de grandeza inversamente proporcional entre as variáveis. Analogamente, quando  $r$  é positivo, diz-se que as duas variáveis são diretamente proporcionais.

O coeficiente de correlação de Pearson é normalmente representado pela letra  $r$  e a sua fórmula de cálculo é:

$$r_{Pearson} = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \times \sqrt{\sum_{i=1}^n y_i^2 - n\bar{y}^2}}$$

Onde:

- $x_i$  = i-ésimo valor da variável  $X$
- $y_i$  = i-ésimo valor da variável  $Y$
- $\bar{x}$  = média dos valores da variável  $X$
- $\bar{y}$  = média dos valores da variável  $Y$

Vale ressaltar que o coeficiente de Pearson é paramétrico e, portanto, sensível quanto à normalidade (simetria) dos dados.

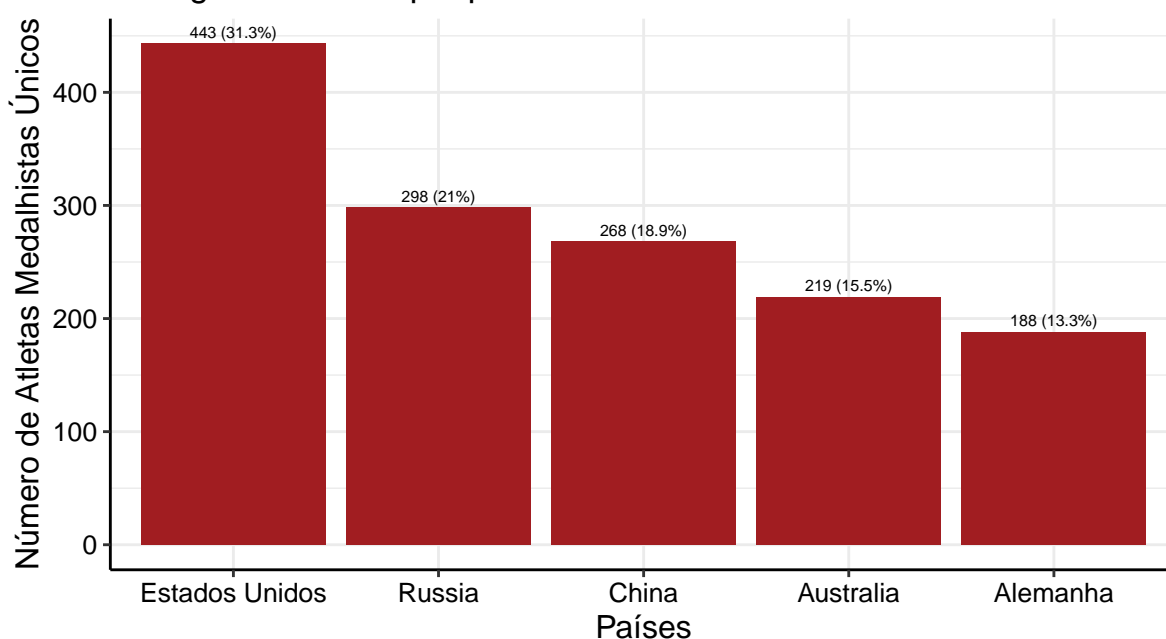
## 3 Análises

### 3.1 Analise 1

#### 3.1.1 TOP 5 PAÍSES COM MAIS MEDALHISTAS DAS OLIMPIADAS DE SYDNEY(2000) - RIO DE JANEIRO (2016)

Nesta primeira análise, consideramos como variável dependente (eixo y) o número de atletas femininas medalhistas ao longo de cinco ciclos olímpicos, caracterizada como uma variável quantitativa discreta. No eixo x, temos os países participantes, uma variável qualitativa nominal, pois representa os nomes dos países. O objetivo desta análise é identificar quais países possuem o maior número de mulheres medalhistas nesse período de cinco Olimpíadas, oferecendo insights sobre os países mais competitivos em termos de desempenho feminino olímpico.

Histogram sobre Top 5 países com mais atletas medalhistas únicos em



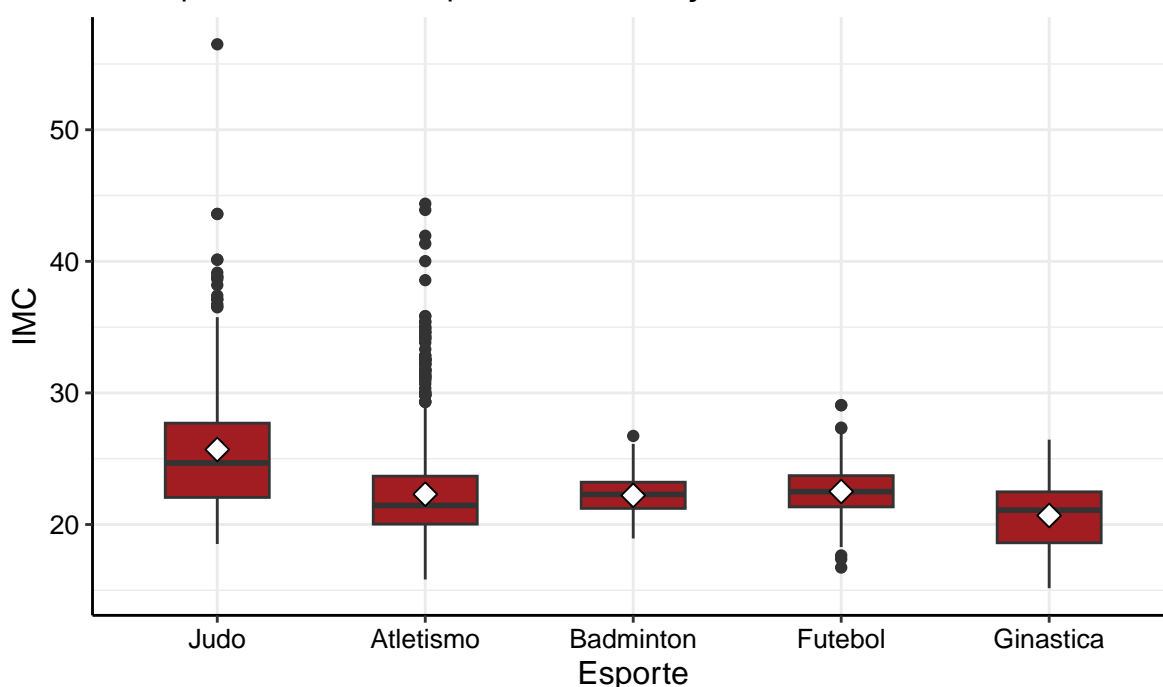
Os Estados Unidos lideram o número de mulheres medalhistas, com 443 medalhas, o que representa quase um terço do total de medalhistas do top 5. Em segundo lugar, a Rússia acumula 298 medalhistas, refletindo seu alto investimento em esportes. A China, com 268 medalhistas, teve um aumento significativo desde as Olimpíadas de Pequim 2008 e continuou a se destacar nas edições seguintes, o que demonstra uma estratégia bem-sucedida de investimento, especialmente durante os Jogos realizados no país. A Austrália, com 219 medalhistas (15,5% do total do top 5), se destaca principalmente em esportes aquáticos, como natação e vela. Em quinto lugar, a Alemanha apresenta 188 medalhistas nas últimas cinco edições, ficando atrás dos outros quatro países.



## 3.2 Análise 2

### 3.2.1 RELAÇÃO ENTRE IMC EM DIFERENTES ESPORTES

Nesta análise, trabalhamos com duas variáveis: o IMC (Índice de Massa Corporal), uma variável quantitativa contínua, e o esporte praticado, uma variável qualitativa nominal. A análise apresenta um boxplot que compara cinco esportes: Judô, Badminton, Futebol, Ginástica e Atletismo, com o objetivo de ilustrar a distribuição do IMC entre os diferentes esportes e identificar possíveis diferenças nos índices de IMC entre eles.



Quadro 1: Medidas resumo dos esportes

Estatística	Atletismo	Badminton	Futebol	Ginastica	Judo
Média	22,30	22,21	22,51	20,68	25,70
Desvio Padrão	3,86	1,50	1,73	2,38	5,12
Variância	14,92	2,26	2,99	5,67	26,23
Mínimo	15,82	18,94	16,73	15,16	18,52
1º Quartil	20,03	21,22	21,34	18,61	22,06
Mediana	21,45	22,28	22,49	21,09	24,68
3º Quartil	23,67	23,21	23,71	22,48	27,70
Máximo	44,38	26,73	29,07	26,45	56,50

A comparação dos valores de IMC entre os esportes analisados indica tendências interessantes. O Judô apresenta as maiores médias de IMC (25,70), desvio padrão (5,12) e variância (26,23), refletindo a divisão das competições por categorias de peso, que gera uma variação maior. No Atletismo, o desvio padrão (3,86) e a variância (14,92) também são elevados, devido às distintas modalidades, como corrida e arre-

nesso de peso, que atraem atletas com IMCs variados.

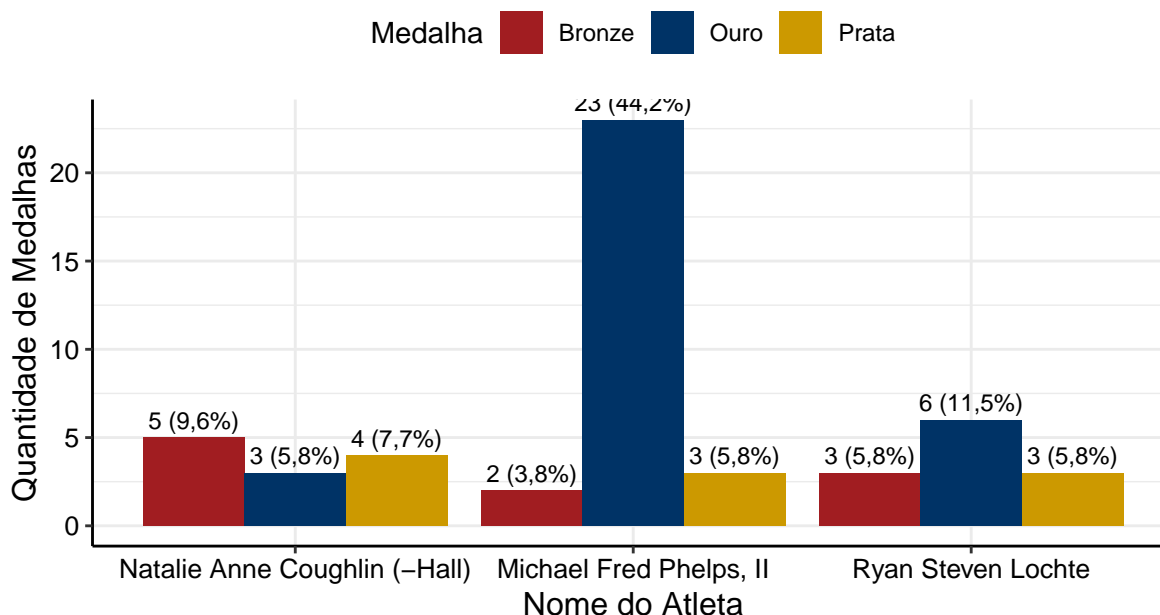
No Badminton, o desvio padrão (1,50) e a variância (2,26) são baixos, sugerindo uma uniformidade no IMC dos atletas, com pequena diferença entre mínimo (18,94) e máximo (26,73). O Futebol também apresenta baixa variação (desvio padrão de 1,73 e variância de 2,99), indicando IMCs semelhantes entre os atletas. Já a Ginástica mostra a menor média (20,68), com uma variância moderada (5,67) inferior à do Judô e do Atletismo, mas ainda maior que a do Futebol e do Badminton.

### 3.3 Análise 3

#### 3.3.1 TOP 3 MEDALHISTAS DAS OLIMPIADAS 2000-2016

Esta análise examina o perfil dos três atletas que mais conquistaram medalhas durante o ciclo olímpico de Sydney 2000 a Rio 2016, detalhando as quantidades e tipos de medalhas obtidas por cada um. Trata-se de uma variável qualitativa ordinal, dado que reflete as posições ocupadas em conquistas olímpicas. O objetivo desta análise é identificar padrões entre os principais medalhistas, explorando quais cores de medalhas prevalecem entre os atletas mais vitoriosos, e fornecer insights sobre o tipo de conquista que é mais frequente entre os atletas de elite.

Histograma sobre a Quantidade de Medalhas por Atleta e Tipo de Medalha



No top 3 de medalhistas olímpicos do ciclo Sydney 2000 até Rio 2016, Michael Phelps ocupa a primeira posição com 28 medalhas, seguido por dois atletas empatados com 12 medalhas cada. Todos os três são nadadores, destacando a natação como um esporte com maior possibilidade de acumulação de medalhas. Essa característica se deve ao fato de a natação contar com diversas modalidades baseadas em estilos de nado, ao invés de divisão por peso, permitindo que atletas competem em múltiplas

provas.

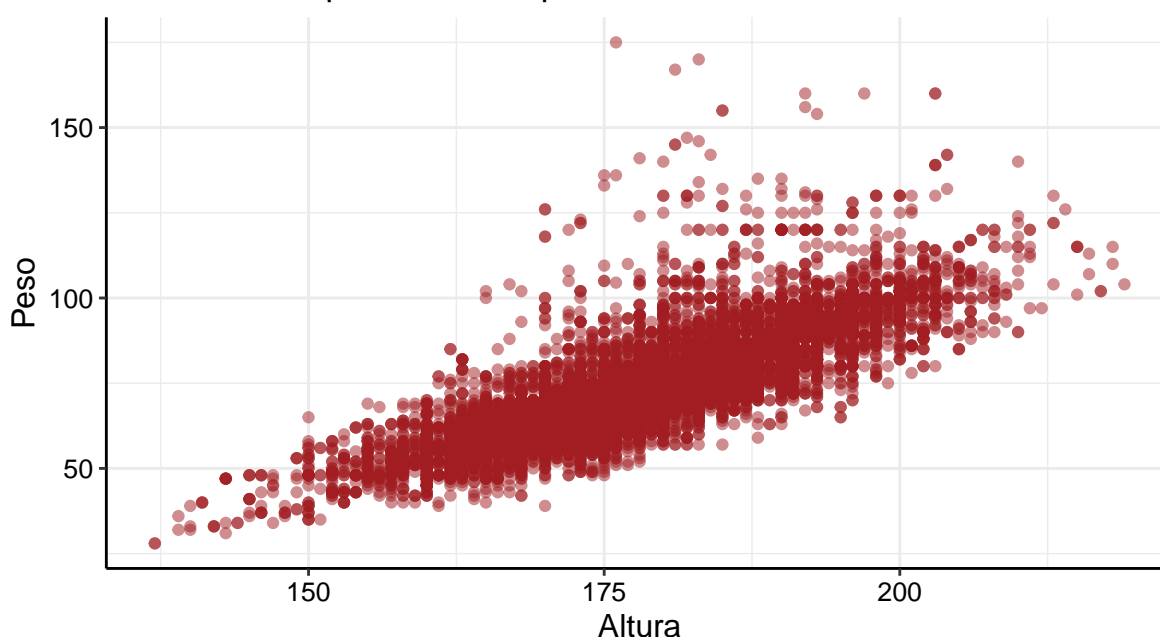
A análise detalhada dos três principais medalhistas revela que a maioria de suas conquistas é composta por medalhas de ouro, totalizando 32 medalhas, ou 61,53% do total. Além disso, esses atletas acumularam 10 medalhas de prata e 10 de bronze. Os dados sugerem uma tendência clara entre os principais medalhistas a conquistarem predominantemente medalhas de ouro, o que ressalta o nível de excelência competitivo destes atletas. Dessa forma, observa-se que os atletas de maior sucesso não apenas acumulam um número elevado de medalhas, mas também mantêm um alto desempenho, refletido na predominância de medalhas de ouro.

### 3.4 Análise 4

#### 3.4.1 Análise da relação de peso por altura

Esta análise tem como objetivo verificar se existe uma relação entre a altura (em centímetros) e o peso (em quilos) dos atletas medalhistas das Olimpíadas de Sydney 2000 e Rio 2016, sendo ambas as variáveis quantitativas contínuas. Esse entendimento é relevante, pois a correlação entre essas variáveis pode revelar características físicas predominantes entre atletas de elite, auxiliando na identificação de perfis ideais para o alto desempenho em competições. Para isso, foi utilizado o coeficiente de correlação de Pearson, escolhido por ser uma medida precisa que permite avaliar a força e a direção dessa relação, oferecendo assim uma visão quantitativa fundamental para a análise.

Gráfico de Dispersão Altura pelo Peso



Quadro 2: Medidas resumo da variável peso

Estatística	Valor
Média	74,00
Desvio Padrão	16,26
Variância	264,26
Mínimo	28,00
1º Quartil	62,00
Mediana	72,00
3º Quartil	84,00
Máximo	175,00

Quadro 3: Medidas resumo da variável Altura

Estatística	Valor
Média	178,24
Desvio Padrão	11,80
Variância	139,23
Mínimo	137,00
1º Quartil	170,00
Mediana	178,00
3º Quartil	186,00
Máximo	219,00

O coeficiente de Pearson calculado foi de **0.8053352**, indicando uma forte correlação positiva entre peso e altura, o que confirma a tendência de que, à medida que a altura dos atletas aumenta, o peso também tende a aumentar. No entanto, é importante ressaltar que alguns dos pesos mais altos estão associados a alturas relativamente baixas. Isso sugere que, embora haja uma tendência de correlação positiva, uma altura maior não implica necessariamente um peso maior. Em resumo, o gráfico revela uma relação significativa entre peso e altura, mas outros fatores, como composição corporal e características individuais, também desempenham um papel importante na determinação do peso dos atletas.

## 4 Conclusões

Este estudo teve como objetivo analisar o desempenho dos atletas da House of Excellence nas Olimpíadas de 2000 a 2016, com o intuito de identificar padrões que possam contribuir para otimizar o treinamento e os resultados dos atletas.

Na primeira análise, que abordou os países com mais medalhistas entre as mulheres, ficou claro que os Estados Unidos, seguidos pela Rússia, são as maiores potências olímpicas, o que reflete o impacto do forte investimento no esporte. Juntos, esses dois países representam mais de 50% dos medalhistas no top 5, o que reforça a importância de investimentos significativos para alcançar resultados superiores.

A segunda análise, que comparou o IMC de atletas de cinco esportes diferentes (ginástica, judô, futebol, badminton e atletismo), revelou diferenças evidentes entre os esportes. O judô apresenta o maior IMC devido às suas categorias por peso, enquanto o futebol e o badminton têm índices semelhantes, refletindo uma maior uniformidade entre os atletas. A ginástica, por sua vez, exige menores índices de IMC devido à necessidade de flexibilidade. O atletismo também exibe variação nos dados devido às diferentes modalidades dentro do esporte. Isso destaca a necessidade de treinamentos e preparações específicas para cada esporte.

Na terceira análise, ao investigar os atletas com mais medalhas conquistadas de 2000 a 2016, observou-se que a natação é o esporte com maior probabilidade de gerar múltiplas medalhas devido à diversidade de categorias e estilos. Além disso, os atletas de natação destacam-se pela sua alta taxa de medalhas de ouro, representando 61,5% das conquistas.

Por fim, a quarta análise utilizou um gráfico de dispersão para examinar a relação entre altura e peso dos atletas, evidenciando a correlação direta entre essas variáveis, o que pode fornecer insights valiosos para otimização de treinamentos baseados nas características físicas dos atletas.

Em suma, este relatório oferece insights valiosos que podem ser aplicados para aprimorar o desempenho dos atletas da House of Excellence, direcionando estratégias específicas de treinamento e suporte com base nas características e necessidades de cada esporte.