

# House of Excellence

**Consultores Responsáveis:**

Lucas Pessoa Ranieri

**Requerente:**

João Vitor Neves

Brasília, 7 de novembro de 2024.



## Sumário

	Página
1 Introdução . . . . .	3
2 Referencial Teórico . . . . .	4
2.1 Frequência Relativa . . . . .	4
2.2 Média . . . . .	4
2.3 Mediana . . . . .	4
2.4 Quartis . . . . .	5
2.5 Variância . . . . .	5
2.5.1 Variância Populacional . . . . .	5
2.5.2 Variância Amostral . . . . .	6
2.6 Desvio Padrão . . . . .	6
2.6.1 Desvio Padrão Populacional . . . . .	6
2.6.2 Desvio Padrão Amostral . . . . .	7
2.7 Boxplot . . . . .	7
2.8 Histograma . . . . .	8
2.9 Gráfico de Dispersão . . . . .	8
2.10 Tipos de Variáveis . . . . .	8
2.10.1 Qualitativas . . . . .	8
2.10.2 Quantitativas . . . . .	9
2.11 Coeficiente de Correlação de Pearson . . . . .	9
3 Análises . . . . .	11
3.1 Análise 1 . . . . .	11
3.2 Análise 2 . . . . .	12
3.3 Análise 3 . . . . .	13
3.4 Análise 4 . . . . .	14
4 Conclusões . . . . .	17
4.1 Análise 1: Desempenho por País . . . . .	17
4.2 Análise 2: Índice de Massa Corporal (IMC) por Modalidade Esportiva . . . . .	17
4.3 Análise 3: Perfil dos Principais Medalhistas . . . . .	17
4.4 Análise 4: Relação entre Altura e Peso . . . . .	18
4.5 Considerações Finais . . . . .	18

# 1 Introdução

João Neves, proprietário da House of Excellence, uma academia de alta performance, contratou nossa equipe para realizar análises estatísticas que visam otimizar o desempenho dos atletas de elite que representaram seu país nas Olimpíadas de 2000 a 2016. Ele solicitou um relatório estatístico abordando quatro análises principais e um dashboard interativo no Power BI para visualização dinâmica dos dados. As análises incluem identificar os cinco países com o maior número de mulheres medalhistas, em ordem decrescente; calcular e comparar o Índice de Massa Corporal (IMC) entre atletas de ginástica, futebol, judô, atletismo e badminton, observando diferenças e padrões específicos; determinar os três medalhistas com o maior número total de medalhas e analisar a distribuição de cada tipo de medalha (ouro, prata, bronze) entre eles; e explorar a correlação entre peso e altura dos atletas para verificar se há uma tendência proporcional entre essas variáveis. Essas análises serão apresentadas de forma detalhada no relatório e de maneira interativa no dashboard, permitindo que João obtenha insights valiosos para guiar estratégias de treinamento e aprimoramento de seus atletas.

Os dados utilizados foram coletados pelo cliente e incluem informações detalhadas sobre os atletas olímpicos, como nome, idade, altura, peso, país, esporte, modalidade, tipo de medalha e ano de conquista. A riqueza desses dados foi fundamental para a precisão e profundidade das análises solicitadas, possibilitando uma compreensão abrangente sobre o desempenho e características dos atletas entre 2000 e 2016.

Para conduzir as análises, utilizamos o RStudio versão 4.3.1(2023-06-16), o principal software empregado pela empresa devido à sua robustez e versatilidade. O RStudio oferece uma ampla gama de pacotes estatísticos e ferramentas de visualização de dados, permitindo a realização de análises complexas e a geração de gráficos detalhados. Este software não apenas facilita a análise estatística, mas também melhora a apresentação dos resultados, tornando-os mais acessíveis e compreensíveis.

## 2 Referencial Teórico

### 2.1 Frequência Relativa

A frequência relativa é utilizada para a comparação entre classes de uma variável categórica com  $c$  categorias, ou para comparar uma mesma categoria em diferentes estudos.

A frequência relativa da categoria  $j$  é dada por:

$$f_j = \frac{n_j}{n}$$

Com:

- $j = 1, \dots, c$
- $n_j$  = número de observações da categoria  $j$
- $n$  = número total de observações

Geralmente, a frequência relativa é utilizada em porcentagem, dada por:

$$100 \times f_j$$

### 2.2 Média

A média é a soma das observações dividida pelo 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

### 2.3 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 o valor que deixa metade das observações abaixo dela e metade dos dados acima.

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

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

## 2.4 Quartis

Os quartis são separatrizes que dividem o conjunto de dados em quatro partes iguais. O primeiro quartil (ou inferior) delimita os 25% menores valores, o segundo representa a mediana, e o terceiro delimita os 25% maiores 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 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.

## 2.5 Variância

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

### 2.5.1 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

### 2.5.2 Variância Amostral

Para uma amostra, a variância é dada por:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Com:

- $X_i$  =  $i$ -ésima observação da amostra
- $\bar{X}$  = média amostral
- $n$  = tamanho da amostra

## 2.6 Desvio Padrão

O desvio padrão é a raiz quadrada da variância. Ele avalia o quanto os dados estão dispersos em relação à média.

### 2.6.1 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

### 2.6.2 Desvio Padrão Amostral

Para uma amostra, o desvio padrão é dado por:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

Com:

- $X_i$  = i-ésima observação da amostra
- $\bar{X}$  = média amostral
- $n$  = tamanho da amostra

## 2.7 Boxplot

O boxplot é uma representação gráfica na qual se pode perceber de forma mais clara como os dados estão distribuídos. A figura abaixo ilustra um exemplo de boxplot.

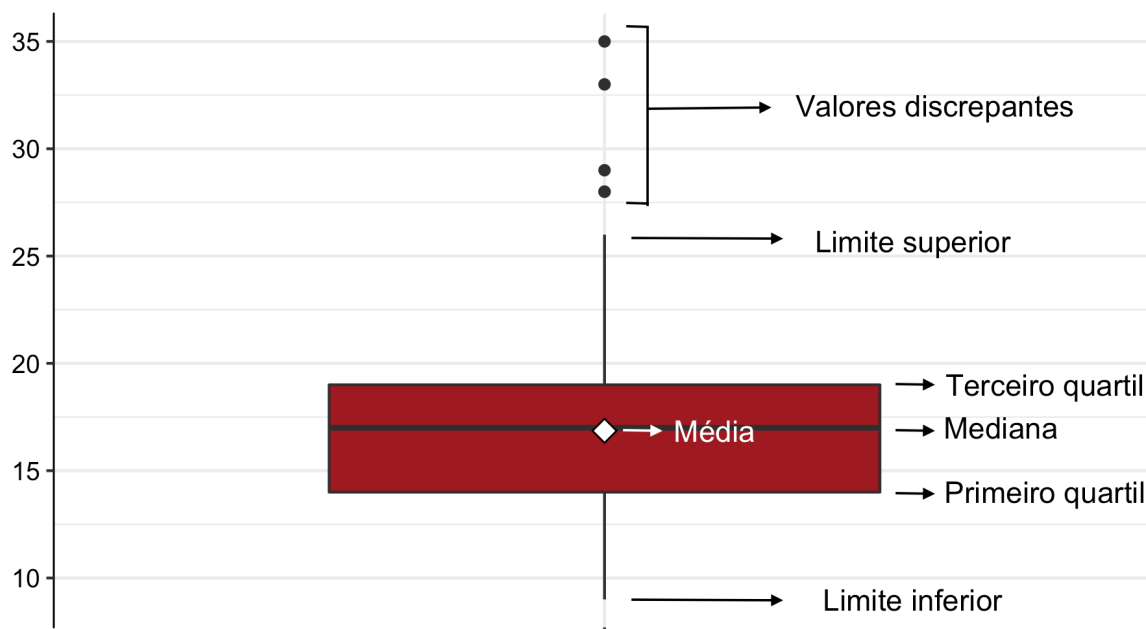


Figura 1: Exemplo de boxplot

A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada pelo losango branco e os

pontos são *outliers*. Os *outliers* são valores discrepantes da série de dados, ou seja, valores que não demonstram a realidade de um conjunto de dados.

## 2.8 Histograma

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

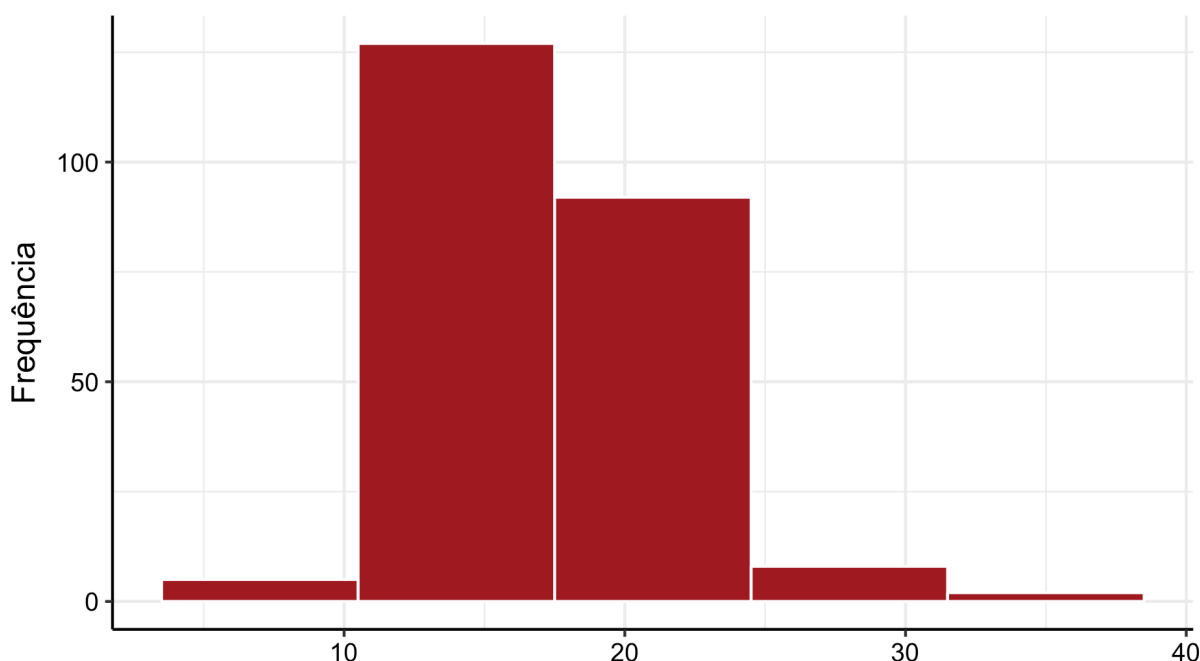


Figura 2: Exemplo de histograma

## 2.9 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.

## 2.10 Tipos de Variáveis

### 2.10.1 Qualitativas

As variáveis qualitativas são as variáveis não numéricas, que representam categorias ou características da população. Estas subdividem-se em:



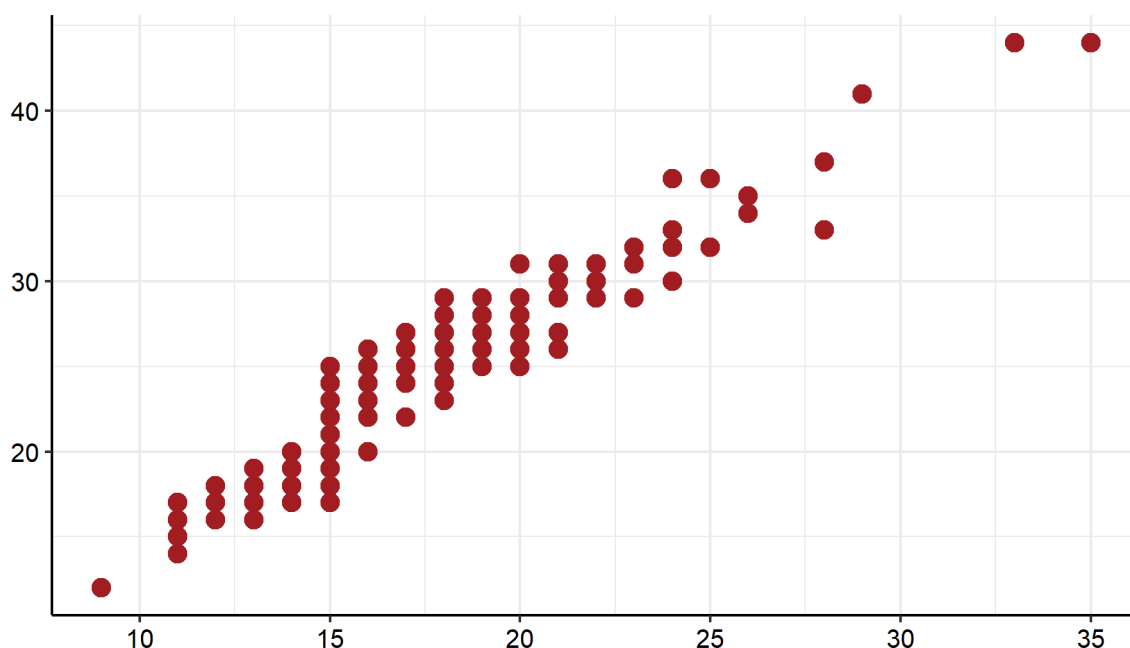


Figura 3: Exemplo de Gráfico de Dispersão

- **Nominais:** quando não existe uma ordem entre as categorias da variável (exemplos: sexo, cor dos olhos, fumante ou não, etc)
- **Ordinais:** quando existe uma ordem entre as categorias da variável (exemplos: nível de escolaridade, mês, estágio de doença, etc)

### 2.10.2 Quantitativas

As variáveis quantitativas são as variáveis numéricas, que representam características numéricas da população, ou seja, quantidades. Estas subdividem-se em:

- **Discretas:** quando os possíveis valores são enumeráveis (exemplos: número de filhos, número de cigarros fumados, etc)
- **Contínuas:** quando os possíveis valores são resultado de medições (exemplos: massa, altura, tempo, etc)

## 2.11 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:

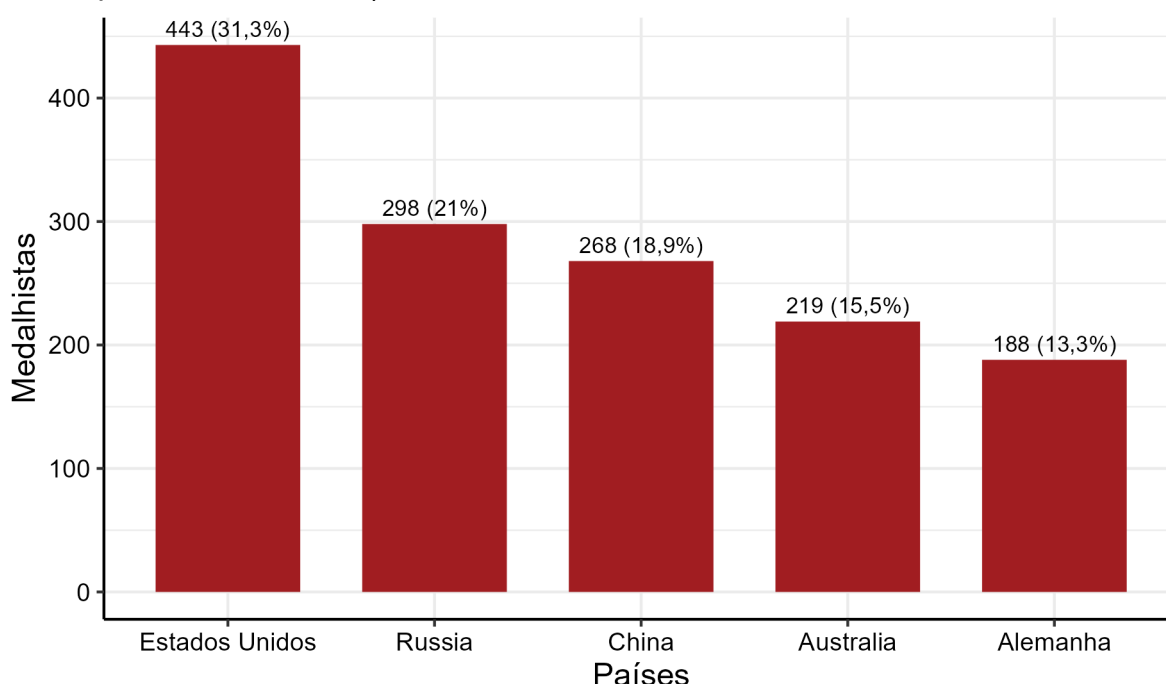
- $x_i$  = i-ésimo valor da variável  $X$
- $y_i$  = i-ésimo valor da variável  $Y$
- $\bar{x}$  = média dos valores da variável  $X$
- $\bar{y}$  = média dos valores da variável  $Y$

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

## 3 Análises

### 3.1 Análise 1

A análise a seguir tem como foco comparar a distribuição de medalhas entre diferentes países. O objetivo é identificar as variações no desempenho esportivo de cada nação, avaliando as proporções de medalhas conquistadas. As variáveis analisadas incluem os países (variável qualitativa nominal) e o número de medalhas (variável quantitativa discreta).



O gráfico evidencia a **liderança expressiva dos Estados Unidos**, com um total de **443 medalhistas, representando 31,3% do total**. Este valor reflete a **dominância histórica do país em competições internacionais**, especialmente nas Olimpíadas, com uma ampla base de **atletas altamente competitivos** em várias modalidades esportivas.

A **Rússia aparece em segundo lugar, com 298 medalhistas (21%)**. Apesar das mudanças geopolíticas e algumas restrições recentes em competições, o país continua a **demonstrar sua importância no cenário esportivo mundial**.

A **China ocupa a terceira posição, com 268 medalhistas (18,9%)**, evidenciando o **crescimento esportivo do país nas últimas décadas**. O **forte investimento em modalidades como ginástica, levantamento de peso e natação** reflete-se nos números expressivos de medalhas.

A **Austrália, com 219 medalhistas (15,5%)**, se destaca, especialmente considerando sua **menor população** em comparação aos demais países no gráfico. O país é tradicionalmente forte em **esportes aquáticos, como natação e**

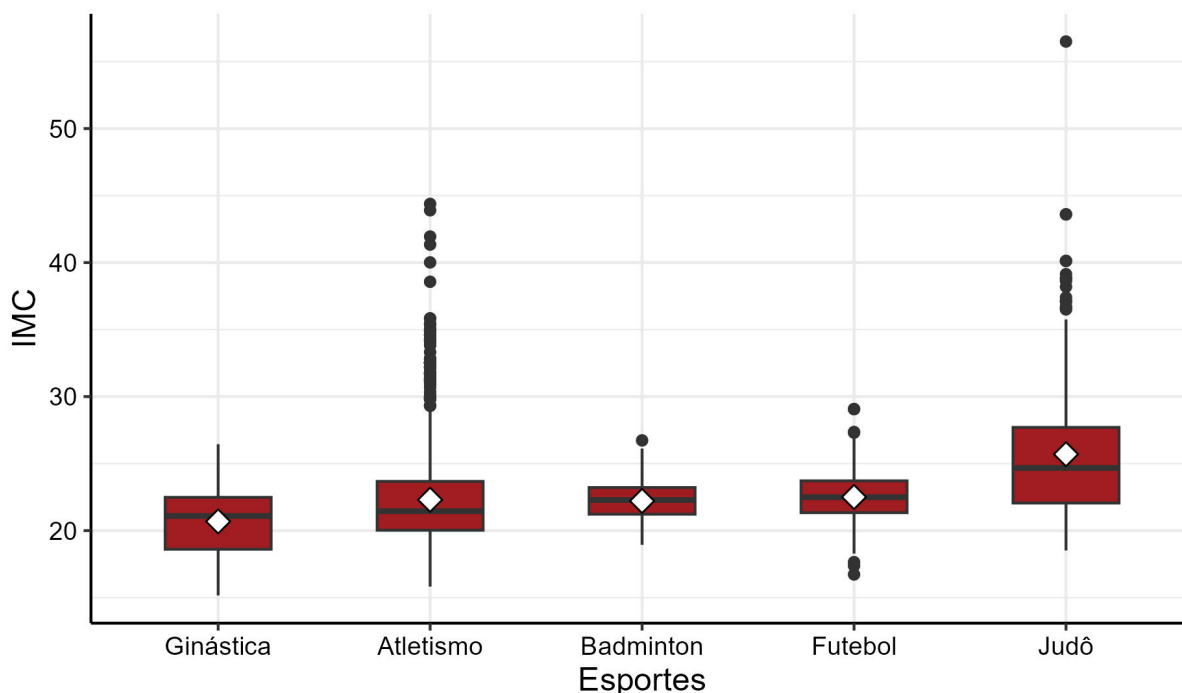
**vela**, contribuindo para seu desempenho robusto.

Por fim, a **Alemanha, com 188 medalhistas (13,3%)**, completa o grupo dos cinco principais países. Sua **longa tradição de sucesso em esportes como atletismo, ciclismo e esportes de inverno** assegura sua posição entre as potências esportivas globais.

Este gráfico oferece uma **visão clara das principais potências esportivas**, evidenciando a **dominância dos Estados Unidos, seguidos pela Rússia e China**. A análise dos **números de medalhistas permite insights** sobre os **investimentos e políticas esportivas** de cada país, bem como a **importância dada ao desenvolvimento de atletas de alta performance**.

## 3.2 Análise 2

A análise a seguir tem como foco comparar o índice de massa corporal (IMC) de atletas de diferentes modalidades esportivas. O objetivo é identificar as variações no IMC entre os esportes, avaliando as características físicas predominantes em cada um. As variáveis analisadas incluem o esporte (variável qualitativa nominal) e o IMC dos atletas (variável quantitativa contínua).



Quadro 1: Medidas resumo da nota de IMC por esporte

Estatística	Atletismo	Badminton	Futebol	Ginástica	Judô
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

Os atletas de **Judô** têm o **maior IMC médio (25,70)**, com **grande dispersão** (desvio padrão de 5,12) e vários **outliers**. Isso reflete a **diversidade de perfis corporais** devido às diferentes categorias de peso, que exigem variações na **composição corporal** para maximizar o desempenho.

Esportes como **Ginástica, Atletismo, Badminton e Futebol** apresentam **IMCs mais baixos e menor variação**. Ginástica, com o **menor IMC médio (20,68)**, destaca-se pela **uniformidade física** entre os atletas, refletindo a necessidade de **corpos leves e flexíveis**. O Atletismo tem um IMC médio de **22,30**, com maior variação devido à diversidade das disciplinas, onde **corredores têm IMCs mais baixos** e competidores de força, mais altos.

**Badminton** exibe um **IMC médio baixo (22,21)** e a **menor variação** (desvio padrão de 1,50), refletindo a necessidade de **agilidade e resistência**. Já o **Futebol** tem um **IMC médio de 22,51**, com leve dispersão e sem outliers, indicando uma **uniformidade física** entre os jogadores, que equilibram **força e leveza**.

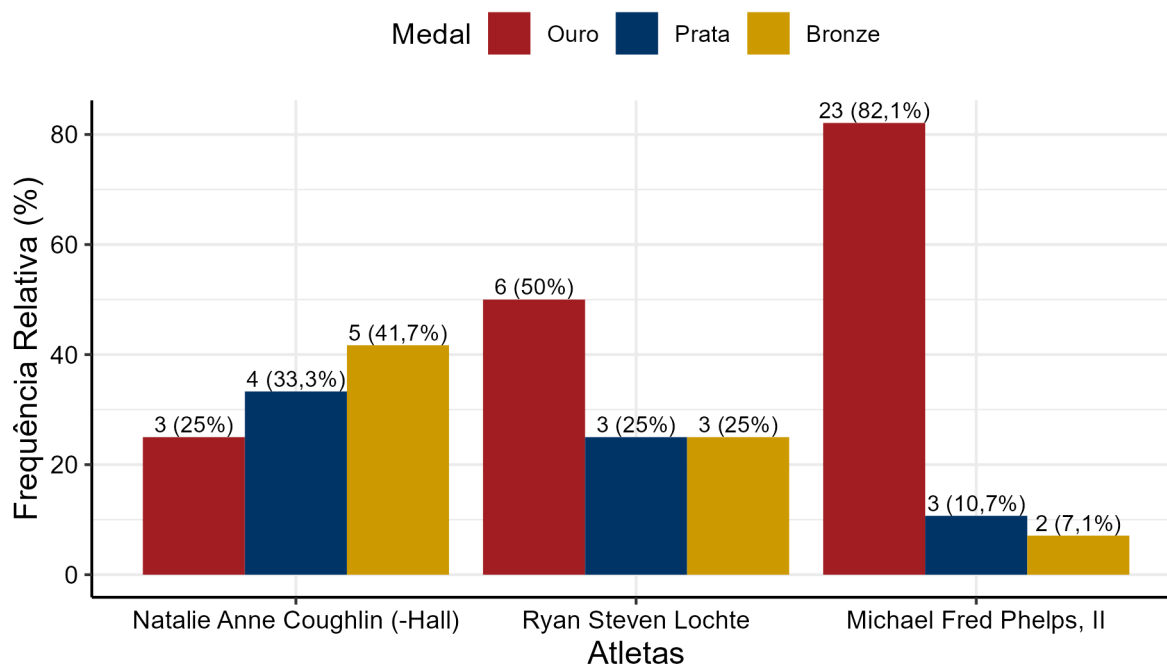
Os **box-plots** confirmam essas conclusões: Judô apresenta **maior variação**, enquanto os outros esportes mostram distribuições mais concentradas, refletindo a necessidade de **biotipos mais homogêneos** para desempenho ideal.

Em resumo, as **diferenças de IMC** entre as modalidades esportivas destacam como cada esporte impõe **exigências físicas únicas** aos seus atletas. **Judô**, com suas categorias de peso, apresenta uma **maior variação** no perfil corporal, enquanto esportes como **Futebol, Badminton, Atletismo e Ginástica** tendem a promover um **biotipo mais uniforme**, ajustado às demandas específicas de cada modalidade.

### 3.3 Análise 3

A análise a seguir tem como foco identificar os três principais medalhistas em termos de quantidade total de medalhas, destacando a distribuição entre os tipos de medalha (ouro, prata e bronze). O objetivo é compreender a relação entre o desempenho desses atletas e a proporção de medalhas conquistadas, oferecendo uma visão sobre

a regularidade e consistência nos pódios.



O gráfico destaca **Michael Fred Phelps, II** como o maior medalhista, com **82,1% de suas medalhas sendo de ouro** (23 medalhas), evidenciando sua **dominância nas competições**. Ele possui apenas 3 medalhas de prata (10,7%) e 2 de bronze (7,1%), o que indica que sua presença no pódio está majoritariamente associada a **vitórias absolutas**.

Em contraste, **Ryan Steven Lochte e Natalie Anne Coughlin (-Hall)** apresentam uma **distribuição mais equilibrada** entre os tipos de medalhas. Lochte conquistou 50% de medalhas de ouro (6), com o restante igualmente distribuído entre prata e bronze (25% cada, 3 de cada tipo), refletindo uma **consistência entre os primeiros colocados**, mas sem a dominância de Phelps.

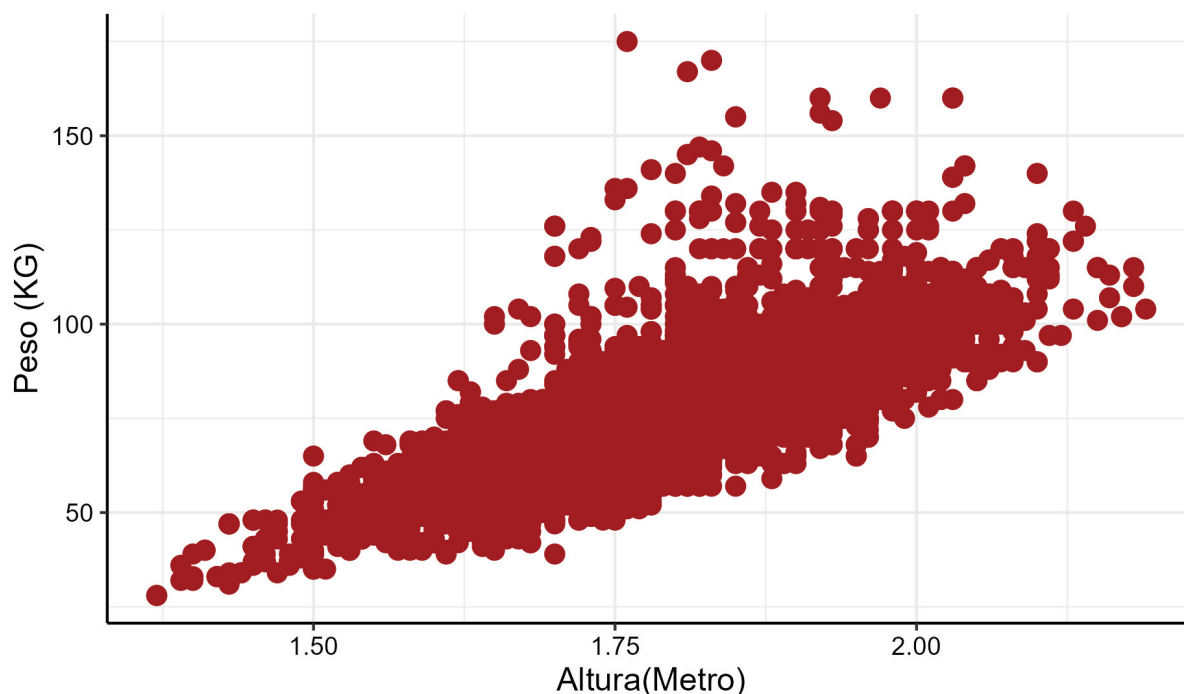
Natalie Coughlin tem uma **maior concentração em medalhas de bronze** (41,7%, 5 medalhas), seguida de 33,3% de ouro (4) e 25% de prata (3). Esse perfil indica que, apesar de frequente nos pódios, ela terminou mais vezes em terceiro lugar, mas ainda assim é uma atleta **altamente competitiva**.

Em resumo, enquanto **Phelps se destaca pela predominância de ouro**, Lochte e Coughlin mostram um **equilíbrio em suas conquistas**, revelando a **alta competitividade** entre esses nadadores de elite, cada um com uma trajetória distinta de sucesso.

### 3.4 Análise 4

A análise a seguir investiga a relação entre a altura e o peso dos atletas olímpicos, abrangendo diversas modalidades esportivas. Ambas as variáveis – altura (em metros)

e peso (em quilogramas) – são quantitativas contínuas, e o objetivo é verificar se há uma correlação entre essas medidas físicas. Para avaliar essa relação, foi utilizado o coeficiente de correlação de Pearson.



O cálculo do coeficiente de Pearson resultou em um valor de aproximadamente **0,8053**, indicando uma **correlação positiva forte** entre altura e peso dos atletas olímpicos. Esse coeficiente sugere que, em média, conforme a altura dos atletas aumenta, o peso também tende a crescer proporcionalmente. Essa relação é coerente com o fato de que **atletas mais altos frequentemente possuem maior massa corporal**, uma característica que pode ser benéfica em muitos esportes de alto rendimento.

Observando a distribuição dos dados, a maioria dos atletas olímpicos se concentra em uma faixa de **altura entre 1,5 e 2,0 metros** e **peso entre 50 e 100 kg**. Essa faixa representa o **perfil físico predominante** nas Olimpíadas, embora existam variações significativas entre as modalidades. Por exemplo, esportes como **ginástica** apresentam atletas menores e mais leves, enquanto modalidades como **basquete, vôlei e natação** costumam exigir maior estatura e peso, o que favorece o desempenho nesses esportes. Já nos esportes de combate, como **judô e boxe**, observamos uma ampla gama de perfis físicos para a mesma altura, devido às **categorias de peso** que permitem maior variabilidade entre os atletas.

Para garantir **consistência e clareza** nos gráficos, todas as unidades de peso foram padronizadas em **quilogramas (kg)**, evitando distorções e facilitando uma leitura precisa da **dispersão dos dados**. Além disso, a análise da dispersão revela uma concentração dos atletas na faixa central de altura e peso, com maior densidade entre **1,5-2,0 metros e 50-100 kg**, enquanto nos extremos, observamos dispersão

especialmente em esportes que exigem perfis físicos específicos. Essa descrição complementa a análise textual e gráfica, proporcionando uma visão mais clara do comportamento dos dados.

Em conclusão, a análise sugere que características físicas como **altura e peso** estão fortemente relacionadas, influenciando o tipo de esporte em que os atletas se destacam nas Olimpíadas. Modalidades que demandam **força e estabilidade**, como levantamento de peso, ou que favorecem maior estatura, como **basquete e vôlei**, apresentam perfis físicos específicos. Essa **forte correlação (0,8053)** entre altura e peso oferece **insights valiosos para treinadores e profissionais esportivos**, auxiliando na identificação de perfis físicos ideais para diferentes modalidades e orientando o desenvolvimento de talentos.



## 4 Conclusões

Este relatório reúne análises abrangentes e detalhadas sobre o desempenho e características físicas de atletas olímpicos entre 2000 e 2016, proporcionando insights para o aprimoramento de estratégias de treinamento na House of Excellence. A seguir, são apresentadas as principais conclusões das análises:

### 4.1 Análise 1: Desempenho por País

A análise dos cinco países com maior número de medalhas revela o domínio expressivo dos **Estados Unidos**, que representam 34,5% do total de medalhas, refletindo sua tradição e investimento constante no esporte de elite. **Rússia** e **China** também apresentam desempenhos significativos, indicando fortes programas de desenvolvimento de atletas. **Austrália** e **Alemanha** completam a lista, com o primeiro se destacando em esportes aquáticos e o segundo em modalidades variadas, reforçando a importância de focar em especializações estratégicas. Essa visão fornece uma base para que a House of Excellence avalie modelos de sucesso e desenvolva talentos de forma direcionada.

### 4.2 Análise 2: Índice de Massa Corporal (IMC) por Modalidade Esportiva

A comparação do IMC entre diferentes esportes destaca variações significativas que refletem as exigências físicas de cada modalidade: - **Judô** apresenta o maior IMC médio e a maior variabilidade, indicando a importância da diversidade corporal nas diferentes categorias de peso. - **Ginástica** possui o menor IMC médio, evidenciando a necessidade de um corpo leve e flexível. - **Atletismo**, **Badminton** e **Futebol** mostram IMCs relativamente baixos e homogêneos, reforçando a uniformidade física para otimizar a performance.

Esses resultados sugerem que a House of Excellence pode adaptar programas de treinamento considerando as características físicas ideais para cada modalidade, otimizando o rendimento dos atletas.

### 4.3 Análise 3: Perfil dos Principais Medalhistas

A análise dos três maiores medalhistas olímpicos destaca **Michael Phelps** pela dominância absoluta, com 82,1% de suas medalhas sendo de ouro, evidenciando sua consistência no topo do pódio. **Ryan Lochte** e **Natalie Coughlin** apresentam uma distribuição mais equilibrada de medalhas, demonstrando alto nível competitivo,

embora sem a mesma supremacia de Phelps. Esses perfis indicam diferentes abordagens de treinamento e especialização que podem ser exploradas, permitindo à House of Excellence identificar e aplicar estratégias que fomentem a consistência e a excelência entre os atletas.

#### 4.4 Análise 4: Relação entre Altura e Peso

A forte correlação (0,8053) entre altura e peso mostra que, em geral, à medida que a altura aumenta, o peso tende a crescer proporcionalmente. Este padrão é particularmente relevante em esportes como **basquete**, **vôlei** e **natação**, onde altura e massa corporal são frequentemente associadas a melhor desempenho. Nos esportes de combate, como o **judô**, a variabilidade entre altura e peso é mais acentuada devido às categorias de peso. Esses insights fornecem uma base científica para orientar treinamentos específicos, ajudando a moldar atletas com características físicas adequadas para maximizar o desempenho em suas modalidades.

#### 4.5 Considerações Finais

Este relatório, em conjunto com o dashboard interativo no Power BI, oferece uma visão detalhada e flexível para João Neves e a equipe da House of Excellence. A análise das potências esportivas, a comparação de IMCs, o perfil dos principais medalhistas e a correlação entre altura e peso fornecem uma base sólida para estratégias de desenvolvimento e aprimoramento de atletas. Utilizando essas informações, a House of Excellence poderá direcionar esforços para adaptar programas de treinamento, promover especializações e identificar perfis físicos ideais para diferentes modalidades, maximizando as chances de sucesso de seus atletas em competições de alto nível.