

Modelo Projeto - Quarto

Consultores Responsáveis:

Estatiano 1

Estatiano 2

...

Estatiano n

Requerente:

ESTAT

Brasília, 27 de outubro de 2024.



Sumário

	Página
1 Introdução	3
2 Referencial Teórico	4
3 Análises	5
3.0.1 Top 5 países com maior número de mulheres medalhistas	5
3.0.2 Valor IMC por esporte, estes sendo, ginástica, futebol, judô, atletismo e badminton	6
3.0.3 Top 3 medalhistas gerais por quantidade de cada tipo de medalha	10
4 Conclusões	13

1 Introdução

O seguinte projeto tem como objetivo realizar uma análise estatística do desempenho dos atletas de elite da House of Excellence que participaram das Olimpíadas entre 2000 e 2016. As análises incluem estatísticas descritivas e comparações entre grupos, visando entender fatores como a performance das medalhistas mulheres, a variação do Índice de Massa Corporal (IMC) por esporte, o número de medalhas conquistadas por atletas destacados e a relação entre peso e altura.

O banco de dados utilizado foi coletado pelo cliente, contendo informações sobre atletas e suas performances nas Olimpíadas, incluindo as variáveis nome, gênero, idade, altura, peso, país, esporte, evento, tipo de medalha e ano de conquista da medalha. A amostra é boa e inclui dados de diferentes países e esportes, proporcionando uma visão abrangente e probabilística do desempenho atlético. Essas informações são cruciais para as análises, permitindo uma compreensão mais profunda dos fatores que impactam o desempenho.

Para a realização das análises, foi utilizado o software R, versão 4.2.0. Este software é amplamente reconhecido na comunidade estatística por suas capacidades de manipulação e visualização de dados, além de fornecer uma ampla gama de pacotes para análises estatísticas avançadas. A utilização do R garantiu a precisão e a eficácia na execução das análises propostas, assim como na geração de gráficos e relatórios visuais que complementam os resultados obtidos.

2 Referencial Teórico

3 Análises

3.0.1 Top 5 países com maior número de mulheres medalhistas

O objetivo dessa análise é ver quais países possuem o maior número de mulheres que conquistaram medalhas, em ordem do país com maior número de mulheres medalhistas para o país com menor número.

Para isso, foi criado um gráfico de barras verticais, ideal para analisar uma variável quantitativa discreta, que são valores enumeráveis como a quantidade de medalhas e uma variável qualitativa nominal que são os nomes dos países. No gráfico, a quantidade de medalhas está representada no eixo X e os países no eixo Y em ordem do maior número de mulheres medalhistas para o menor, quantificados por frequência relativa que é utilizada para a comparação entre classes de categorias, como os países.

Para melhor visualização, acompanhe o gráfico **Figura 1** abaixo:

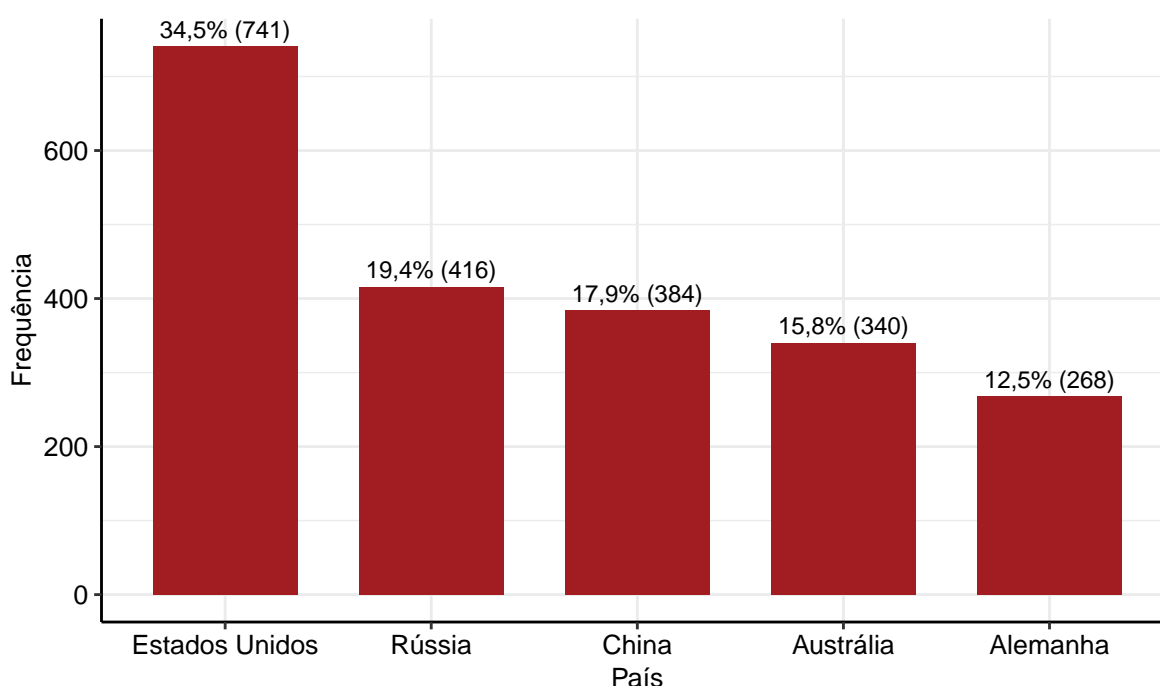


Figura 1: Gráfico de colunas do total de medalhas

É possível visualizar que em primeiro lugar está os Estados Unidos, com o total de 741 medalhistas, seguido da Rússia com 416 medalhistas, seguido da China com 384 medalhistas, seguida da Austrália com 340 medalhistas, seguida da Alemanha com 268 medalhistas.

É possível ver uma diferença significativa entre os 5 primeiros países, sendo os Estados Unidos com 325 medalhas a mais que o segundo colocado e 473 medalhas a mais que o quinto colocado, diferença essa maior que qualquer total de medalhas

de outro país. Assim, os Estados Unidos tem mais medalhas que a soma do total de medalhas que a Rússia e Alemanha juntas. Dessa forma, conclui-se que os Estados Unidos é um país que prepara com excelência as atletas para as Olimpíadas.

Além disso, é possível ver que não há diferença significativa entre a Rússia, China e Austrália quanto o total de mulheres medalhistas, visto que sua diferença entre as medalhas é baixa. Dessa forma, esses países se equivalem na preparação de atletas medalhistas.

Essa análise não apenas destaca a importância do investimento em esportes femininos, como enfatiza países com estratégias eficazes que podem ser implementadas para fomentar o sucesso de mulheres atletas em nível internacional. Portanto, enquanto os Estados Unidos estabelecem um padrão elevado, a competitividade entre Rússia, China e Austrália também ressalta a crescente importância do esporte feminino globalmente.

3.0.2 Valor IMC por esporte, estes sendo, ginástica, futebol, judô, atletismo e badminton

O objetivo dessa análise é observar o IMC dos esportes selecionados, estes sendo, ginástica, futebol, judô, atletismo e badminton. Para isso, foi preciso calcular esse índice para cada atleta usando a fórmula $\text{peso (kg)} / \text{altura (m)}^2$, agrupando o valor do IMC do atleta por esporte.

O IMC é uma variável quantitativa contínua, ou seja apresenta valores resultados de medições, permitindo construir um **Figura 2** para perceber de forma mais clara como os dados estão distribuídos, sendo a porção inferior do retângulo o 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, valores discrepantes da série de dados, ou seja, valores que não demonstram a realidade de um conjunto de dados.

Por outro lado, o IMC é classificado de acordo com a seguinte **Tabela 1**:

Assim, a variável categoria do IMC é qualitativa ordinal, permitindo construir um **Figura 3** com a quantidade de atletas por categoria do IMC em cada esporte.

Além disso, criou-se um quadro de medidas resumo, tabela que apresenta estatísticas descritivas de um conjunto de dados, como média, mediana, moda, desvio padrão, quartis e extremos. Ele serve para fornecer uma visão rápida e concisa das características principais da distribuição dos dados, permitindo entender sua tendência central, variabilidade e forma, facilitando a comparação entre diferentes grupos, nesse caso, diferentes países. Segue o @quad-quadro_resumo1 abaixo:

Tabela 1: Tabela do IMC

(a)

Categoria	IMC Mínimo	IMC Máximo
Abaixo do peso	0	18.4
Peso normal	18.5	24.9
Sobrepeso	25	29.9
Obesidade grau I	30	34.9
Obesidade grau II	35	39.9
Obesidade grau III	40	>

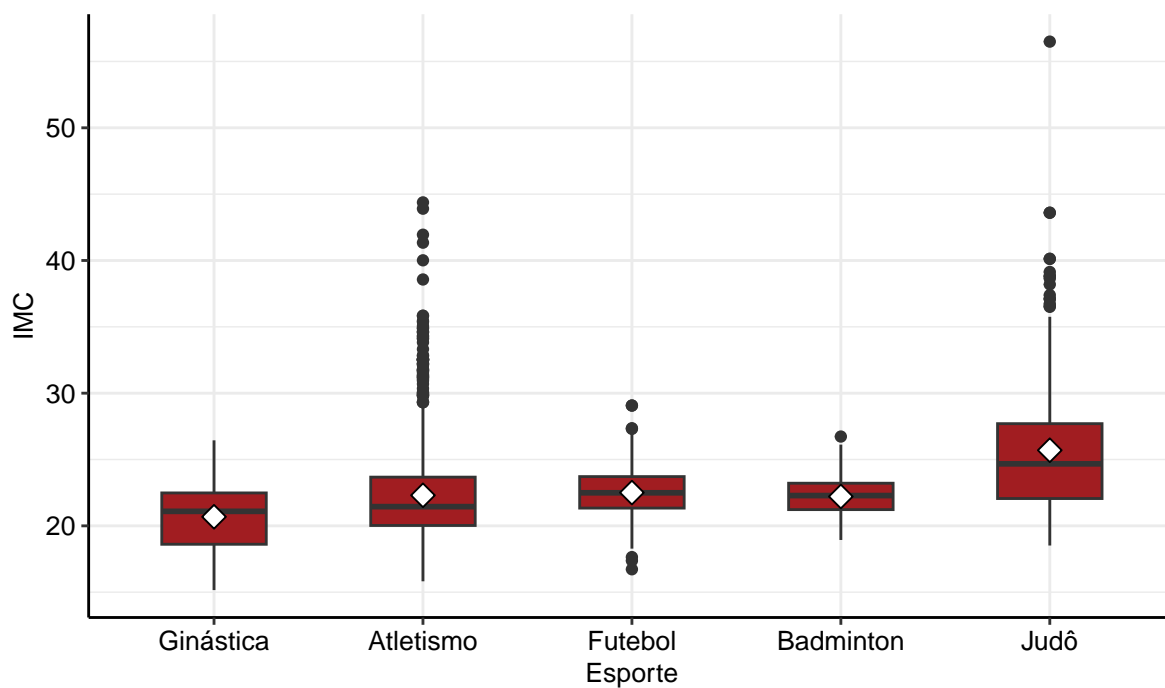


Figura 2: Boxplot do IMC

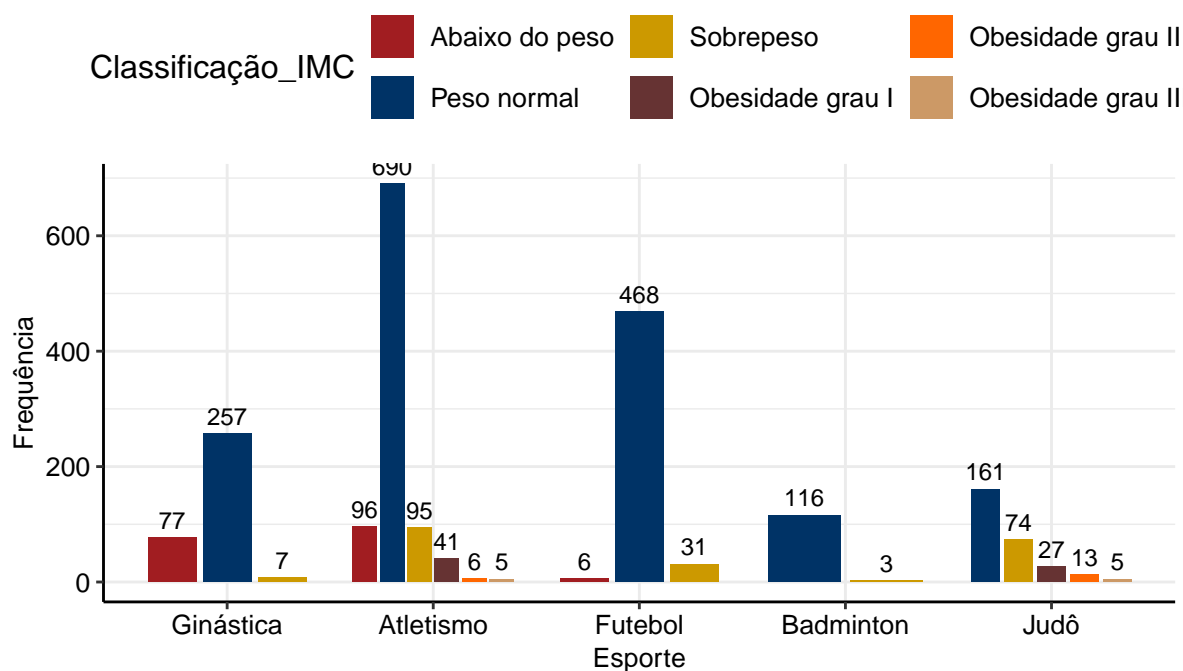


Figura 3: Gráfico de colunas da classificação do IMC

Quadro 1: Medidas resumo do IMC

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

Medidas resumo do IMC

Para ver se os dados do IMC seguem uma distribuição normal, ou seja, se a média, a mediana e a moda são iguais e estão localizadas no centro da distribuição, foi usado o @teste_hipotese1, com as seguintes hipóteses:

$$\begin{cases} H_0 : \text{Os dados seguem uma distribuição normal} \\ H_1 : \text{Os dados não seguem uma distribuição normal} \end{cases}$$

Shapiro-Wilk normality test


```
data: banco_esportes_selecionados$IMC
W = 0,8571, p-value < 0,00000000000000022
```

Tabela 2: P-valor do Teste de Shapiro-Wilk do IMC

Variável	P-valor	Decisão do teste
IMC	0,050	Não rejeita H_0

P-valor do teste de normalidade Shapiro-Wilk

O resultado do teste com 5% de nível de significância revelou que os dados não estão em conformidade com a distribuição normal. Portanto, os dados são não paramétricos. Assim, para verificar se as médias do IMC dos esportes são iguais, foi realizado o @tab-teste_hipotese2, comparando os 5 esportes sem supor nenhuma distribuição.

$$\begin{cases} H_0 : \text{Não há diferença significativa nos valores de IMC entre os diferentes esportes} \\ H_1 : \text{Há diferença significativa nos valores de IMC entre os diferentes esportes} \end{cases}$$

Kruskal-Wallis rank sum test

```
data: IMC by Esporte
Kruskal-Wallis chi-squared = 280,87, df = 4, p-value <
0,00000000000000022
```

Tabela 3: P-valor do Teste de Kruskal-Wallis da média do IMC

Variável	P-valor	Decisão do teste
IMC	0,050	Não rejeita H_0

P-valor do Teste de Kruskal-Wallis

Os resultados do teste de Kruskal-Wallis indicam que há uma diferença significativa nos valores de IMC entre os diferentes esportes, com um valor de p extremamente baixo, o que leva a rejeitar a hipótese nula. A estatística de Kruskal-Wallis foi de 280.87, com 4 graus de liberdade, evidenciando uma variação significativa nas distribuições do IMC conforme o esporte praticado.

Verifica-se a partir do **Figura 2** que as medianas de Ginástica e Atletismo são próximas, assim como as de Futebol e Badminton, já a de Judô está acima dos

terceiros quartis dos outros esportes, mostrando que o valor que está em 50% para o IMC de Judô está acima de 75% dos outros esportes, revelando que os dados de IMC para os judocas estão muito elevados. Além disso, a amplitude interquartil, que indica a dispersão dos dados, é pequena para Futebol e Badminton, intermediária para Ginástica e Atletismo e alta para judô, revelando a variabilidade do índice IMC nesses esportes. Quanto aos outliers, valores atípicos de uma categoria, estão evidentes no Atletismo e Judô, revelando que há muitos atletas nesses esportes que apresentam alto IMC, além do comum. Quanto à simetria dos dados, tem-se que é simétrica para o Futebol, Badminton e Judô, revelando que os dados estão simetricamente distribuídos, ou seja, não tendam para os extremos e assimétrica à direita para o Atletismo, dados próximos de um IMC mais baixo e assimétrica à esquerda para a Ginástica, revelando que os dados tendem para um IMC maior entre essa categoria. Comparando os dados entre os esportes, vê-se claramente que os IMC menores pertencem aos atletas da Ginástica, intermediários para os atletas de Atletismo, Futebol e Badminton e IMC maior para os judocas.

A partir do **Figura 3** vemos que há 22,58% ginastas abaixo do peso, 75,37% no peso normal e 2,05% em sobrepeso. Para o atletismo, há 10,29% de atletas abaixo do peso, 73,95% no peso normal, 10,18% em sobrepeso e 5,57% em obesidade. Já no futebol, há 1,19% de atletas abaixo do peso, 92,67% com peso normal e 6,14% em sobrepeso. No badminton 97,48% estão com peso normal e 2,152% em sobrepeso. E no judô temos apenas 57,5% de atletas no peso normal, 26,43% em sobrepeso e 16,07% em obesidade. Em conclusão, a análise do IMC dos atletas dos esportes selecionados revela padrões significativos em relação à composição corporal de cada grupo. Os dados indicam que os atletas de ginástica apresentam um IMC médio inferior ao peso normal, enquanto os atletas de judô estão, em média, acima desse padrão, com uma diferença notável entre os dois. Por outro lado, os atletas de atletismo, futebol e badminton apresentam IMCs dentro da faixa normal, evidenciando uma distribuição saudável.

Esses resultados sugerem a necessidade de uma avaliação cuidadosa das exigências físicas de cada esporte e a importância de estratégias nutricionais e de treinamento adequadas para promover a saúde e o desempenho atlético. A análise do IMC não apenas fornece insights sobre a condição física dos atletas, mas também pode informar decisões sobre intervenções que visem melhorar a saúde e o bem-estar dentro dessas modalidades esportivas.

3.0.3 Top 3 medalhistas gerais por quantidade de cada tipo de medalha

O objetivo dessa análise é entender quais são os 3 medalhistas com maior número de medalhas no total, observar entre eles a quantidade de cada tipo de medalha que

cada um conquistou e se existe relação entre o medalhista e cada tipo de medalha conquistada. Para isso, foi criado um **Figura 4** para visualizar visualmente a proporção do tipo de medalhas no total de medalhas de cada atleta, sendo essa uma variável quantitativa discreta.

Para avaliar se há relação entre o medalhista e cada tipo de medalha conquistada, ou seja, se o medalhista e o tipo de medalha são variáveis independentes, usou-se o teste de hipótese de Fisher, utilizado quando se quer testar a independência de variáveis com poucos dados de observação.

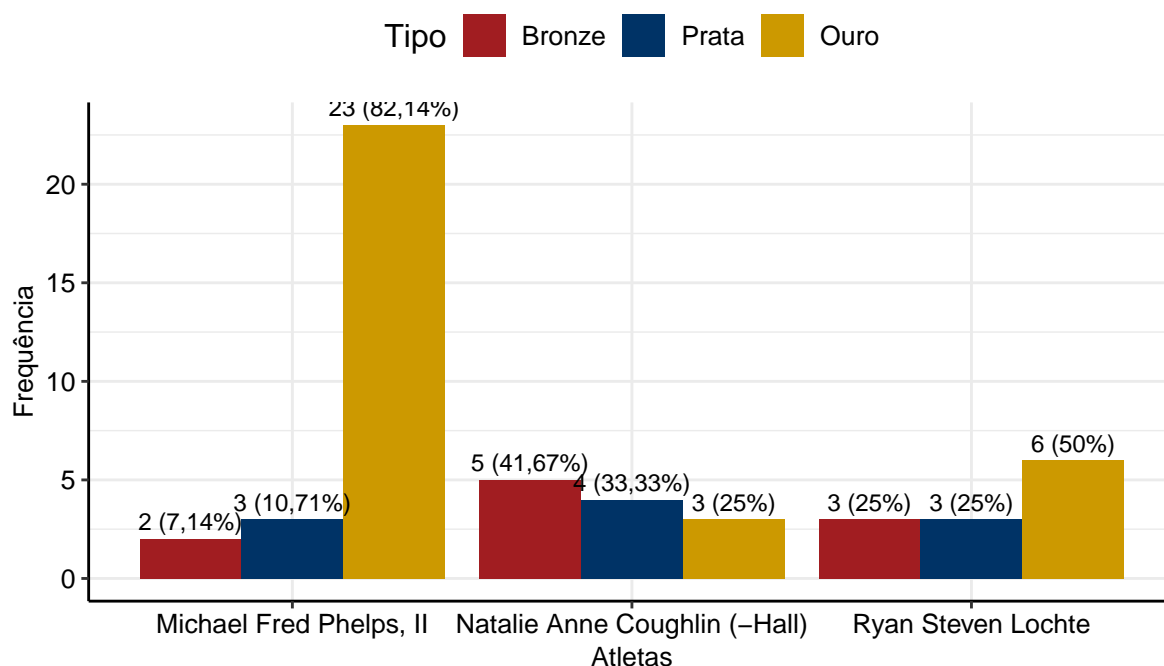


Figura 4: Gráfico de colunas do tipo de medalha

$$\begin{cases} H_0 : \text{Não há relação entre o medalhista e o tipo de medalha conquistada (são independentes)} \\ H_1 : \text{Há relação entre o medalhista e o tipo de medalha conquistada (não são independentes)} \end{cases}$$

Fisher's Exact Test for Count Data

```
data: tabela_fisher
p-value = 0,005638
alternative hypothesis: two.sided
```

Tabela 4: P-valor do teste exato de Fisher) de X

Variável	P-valor	Decisão do teste
Medalhas	0,050	Não rejeita H_0

P-valor do Teste de Fisher

A análise estatística revela informações significativas sobre a distribuição de medalhas entre os atletas. Phelps acumula um total de 28 medalhas, das quais 23 são de ouro, 3 de prata e 2 de bronze. Este desempenho não apenas o posiciona como o atleta com o maior número absoluto de medalhas, mas também evidencia uma predominância acentuada nas medalhas de ouro, que representam aproximadamente 82% de suas conquistas totais.

Em contraste, Natalie Coughlin e Ryan Lochte, ambos com 12 medalhas, apresentam perfis de medalhas diferentes. Coughlin possui uma distribuição de 5 medalhas de bronze, 4 de prata e 3 de ouro, enquanto Lochte tem 3 de bronze, 3 de prata e 6 de ouro. Essa variação na distribuição das medalhas sugere que, embora ambos os atletas tenham um número total equivalente de medalhas, a qualidade das conquistas, refletida pela quantidade de medalhas de ouro, é significativamente inferior em comparação com Phelps.

A discrepância observada nas medalhas de ouro de Phelps em relação aos outros atletas é notável, pois a soma das medalhas de ouro de Coughlin e Lochte (9) ainda é inferior ao total conquistado por Phelps. Essa diferença ressalta a importância do tipo de medalha na avaliação do desempenho dos atletas, indicando que a medalha de ouro não apenas representa a conquista máxima em competições, como é um indicador crítico de sucesso em comparação com medalhas de menor valor, como as de prata e bronze.

Para investigar a relação entre o tipo de medalha conquistada e o atleta, foi realizado um teste de hipótese utilizando o teste exato de Fisher. Com um nível de significância de 5%, os resultados indicaram que existe uma associação entre o medalhista e o tipo de medalha conquistada. Isso implica que as variáveis “atleta” e “tipo de medalha” não são independentes, sugerindo que o desempenho em termos de medalhas varia de maneira entre os atletas. Essa conclusão é relevante para a compreensão das dinâmicas competitivas no contexto do esporte, pois indica que fatores como habilidade, treinamento e experiência podem influenciar a capacidade de um atleta em conquistar medalhas de diferentes categorias.

4 Conclusões