

Analizando a diferença de altura e peso entre atletas de natação medalhistas e não medalhistas

Thays Ferreira Uchôa Albuquerque
Universidade Federal do Ceará
Campus Jardins de Anita
Curso de Ciência de Dados
Itapajé, CE-Brazil
thaysferreira@alu.ufc.br

Laura Lima Mendes
Universidade Federal do Ceará
Campus Jardins de Anita
Curso de Ciência de Dados
Itapajé, CE-Brazil
lauramendes@alu.ufc.br

Abstract—Este artigo investiga a diferença de altura e peso entre os atletas de natação medalhistas e não medalhistas, considerando a separação por sexo. Os dados foram obtidos da plataforma Kaggle, que forneceu observações abrangentes de todos os atletas que competiram nas Olimpíadas de 1896 a 2016. Foram selecionados apenas atletas de natação, a partir do ano de 1912, quando as mulheres começaram a participar do esporte. Esses atletas foram então categorizados em dois grupos: medalhistas e não-medalhistas, e ainda separados por gênero. De cada grupo, uma amostra de $n=100$ foi retirada aleatoriamente para análise. Para comparar a altura e o peso médios entre os ganhadores e não ganhadores de medalhas, foi realizado um teste t independente para cada gênero. Os resultados indicaram que não houve diferença significativa nessas características entre os grupos, independentemente do gênero.

Index Terms—Atletas medalhistas, Natação, Diferenças de gênero, Análise comparativa, Teste t independente.

I. INTRODUÇÃO

A altura e o peso são dois fatores fundamentais que foram amplamente estudados em vários esportes, inclusive na natação. Entender se há diferenças perceptíveis nessas características entre nadadores bem-sucedidos que ganharam medalhas e aqueles que não conseguiram medalhas pode fornecer informações valiosas sobre a função do físico na natação.

O conjunto de dados utilizado para esta análise está disponível na plataforma Kaggle, e abrange atletas que participaram dos Jogos Olímpicos de Verão nas edições entre 1896 e 2016.

A análise concentrou-se especificamente na natação, sendo selecionado atletas a partir do ano de 1912, quando as mulheres entraram pela primeira vez na competição. Também foi considerado as disparidades específicas de gênero, de modo que os dados foram estratificados considerando o sexo masculino e feminino.

A classificação dos atletas em dois grupos distintos - medalhistas e não medalhistas - permite uma comparação direta de seus atributos físicos. Ao categorizar os atletas em grupos com base em conquistas de medalhas e segregá-los ainda mais por gênero, o estudo procura explorar possíveis discrepâncias de altura e peso.

Trabalho Avaliativo da disciplina de Laboratório de Ciência de Dados

II. FUNDAMENTAÇÃO TEÓRICA

Neste estudo foram empregados vários testes estatísticos para garantir a validade e a confiabilidade da análise. Foram utilizados os seguintes testes:

A. Teste de Shapiro-Wilk

O teste de Shapiro-Wilk é um teste comumente usado para avaliar a normalidade de uma distribuição, e possui as seguintes hipóteses:

- H_0 : os dados não são normalmente distribuídos.
- H_1 : os dados são normalmente distribuídos.

O nível de significância escolhido foi de 0,05.

Este teste será aplicado para examinar se as variáveis de altura e peso do conjuntos de dados seguem uma distribuição normal. Ao confirmar a normalidade, uma das premissas para a realização do teste t é atendida.

B. Teste de Bartlett

O teste de Bartlett é um teste estatístico usado para avaliar a homogeneidade das variações em diferentes grupos.

- H_0 : as variâncias não são iguais.
- H_1 : as variâncias são iguais.

O nível de significância escolhido foi de 0,05.

O teste de Bartlett foi realizado para determinar se as variações de altura e peso eram semelhantes entre os atletas medalhistas e os não medalhistas. A suposição de homogeneidade das variações é essencial para a realização de um teste t independente.

C. Teste T independente

O teste t independente é um teste estatístico amplamente usado para comparar as médias de dois grupos independentes.

- H_0 : não há diferença significativa entre as médias dos dois grupos.
- H_1 : há diferença significativa entre as médias dos dois grupos

O nível de significância escolhido foi de 0,05.

Na análise, o teste t independente foi empregado para comparar a altura e o peso médios entre atletas medalhistas e não medalhistas, separadamente para homens e mulheres,

determinando se havia diferenças significativas nessas características entre os dois grupos.

III. PRÉ-PROCESSAMENTO DOS DADOS

A. Categorização

Para garantir que a análise englobasse atletas masculinos e femininos, restringimos o conjunto de dados a registros a partir de 1912, quando as mulheres começaram a participar de eventos de natação. Esse critério de seleção permitiu um exame abrangente das diferenças específicas de gênero.

Para explorar o possível impacto da altura e do peso na conquista de medalhas, categorizamos os atletas em dois grupos distintos: medalhistas e não-medalhistas. Essa diferenciação nos permitiu comparar as características dos atletas que conquistaram medalhas com as dos que não conquistaram.

Além disso, reconhecendo a importância das diferenças específicas de gênero, dividimos ainda mais os dados em subgrupos com base no gênero dos atletas. Essa estratificação nos permitiu avaliar as possíveis disparidades de altura e peso em cada categoria de gênero.

B. Amostragem

De cada grupo, uma amostra aleatória de $n=100$ atletas foi selecionada para análise, visando garantir um subconjunto representativo de atletas de cada categoria. A seleção da amostra foi realizada usando técnicas de amostragem aleatória para garantir uma representação imparcial da população maior de atletas de natação, minimizando o potencial de viés de seleção.

IV. ANÁLISE DESCRITIVA

Antes da análise principal, foi realizado um exame descritivo dos conjuntos de dados.

Fig. 1: Proporção de atletas medalhistas e não medalhistas

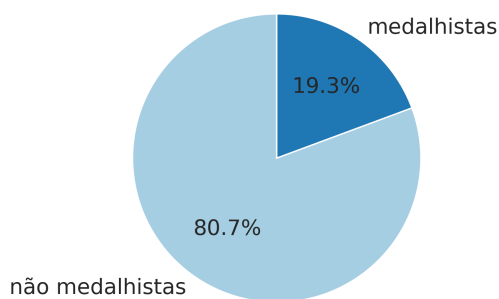


Fig. 2: Proporção de atletas por gênero

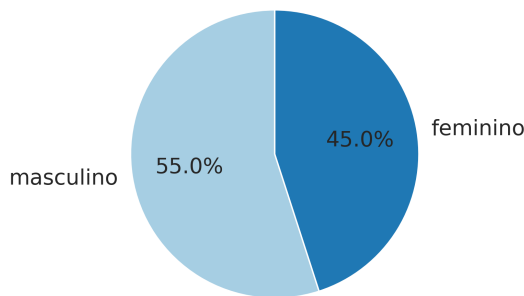


Fig. 3.1: Proporção de medalhistas e não medalhistas - sexo masculino

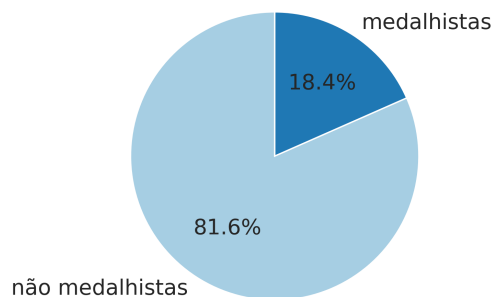
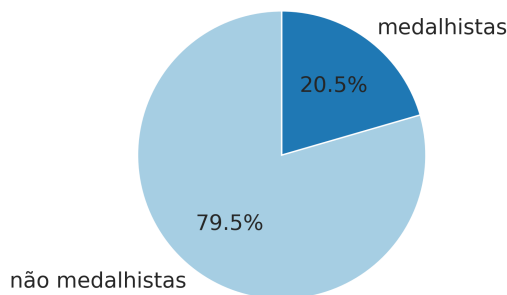


Fig. 3.2: Proporção de medalhistas e não medalhistas - sexo feminino



Através da figura 1 é possível observar que o número de medalhistas é consideravelmente menor do que o número de não medalhistas, alinhando-se com as expectativas, dada a natureza competitiva da natação. Além disso, a distribuição de atletas masculinos e femininos nos conjuntos de dados (figuras 2, 3.1 e 3.2) é razoavelmente proporcional, garantindo uma representação equilibrada de ambos os gêneros.

As estatísticas de tendência central e dispersão (table I e II) foram calculadas para as variáveis de altura e peso em cada conjunto de dados. As medidas de tendência central como média, mediana e moda, indicaram uma simetria aproximada na distribuição de peso e altura para atletas masculinos e femininos. Além disso, a medida de dispersão - desvio padrão - revelou uma variabilidade relativamente menor nos dados, embora algumas observações possam apresentar desvios notáveis da média.

TABLE I
ESTATÍSTICAS DE PESO RESUMIDAS POR GÊNERO E STATUS DE MEDALHISTA

| Sexo | Medalhista | Média | Mediana | Moda | Máximo | Mínimo | Variância | Desvio padrão |
|------|------------|--------|---------|------|--------|--------|-----------|---------------|
| F | Não | 60.041 | 60.0 | 60.0 | 85.0 | 39.0 | 41.434 | 6.437 |
| F | Sim | 63.410 | 63.0 | 63.0 | 85.0 | 45.0 | 39.239 | 6.264 |
| M | Não | 76.411 | 76.0 | 75.0 | 112.0 | 45.0 | 71.352 | 8.447 |
| M | Sim | 80.188 | 80.0 | 80.0 | 114.0 | 54.0 | 68.360 | 8.268 |

TABLE II
ESTATÍSTICAS DE ALTURA RESUMIDAS POR GÊNERO E STATUS DE MEDALHISTA

| Sexo | Medalhista | Média | Mediana | Moda | Máximo | Mínimo | Variância | Desvio padrão |
|------|------------|---------|---------|-------|--------|--------|-----------|---------------|
| F | Não | 169.768 | 170.0 | 170.0 | 191.0 | 131.0 | 48.058 | 6.932 |
| F | Sim | 173.404 | 173.0 | 173.0 | 193.0 | 150.0 | 44.177 | 6.647 |
| M | Não | 182.870 | 183.0 | 180.0 | 209.0 | 147.0 | 61.173 | 7.821 |
| M | Sim | 186.442 | 187.0 | 188.0 | 208.0 | 152.0 | 57.365 | 7.574 |

Fig 4.1: Atletas do sexo feminino

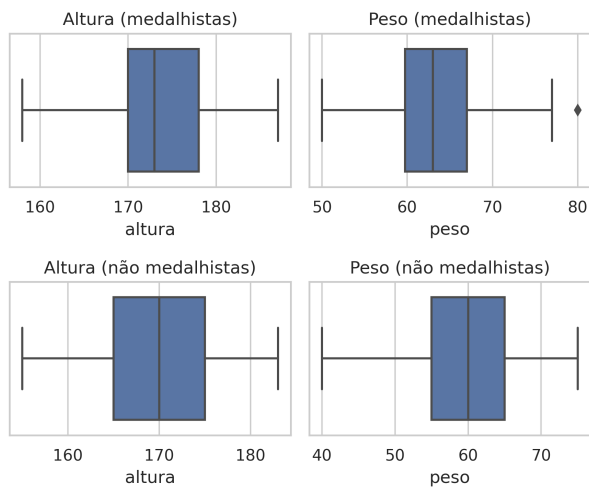
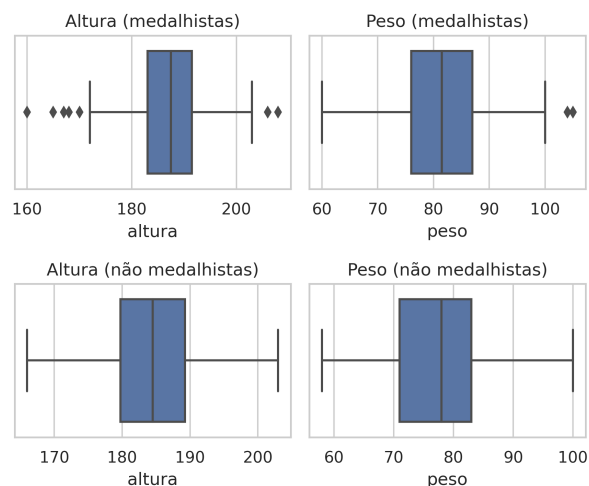


Fig 4.2: Atletas do sexo masculino



Para entender melhor as características dos dados, construímos gráficos de caixa para cada uma das quatro amostras (figuras 4.1 e 4.2). Os gráficos de caixa revelaram a presença de alguns valores discrepantes; no entanto, como pesquisadores, reconhecemos que a remoção desses valores discrepantes poderia introduzir um viés na análise. Como os valores não causaram assimetria significativa nas distribuições, optamos por mantê-los no conjunto de dados.

Fig 5: População geral

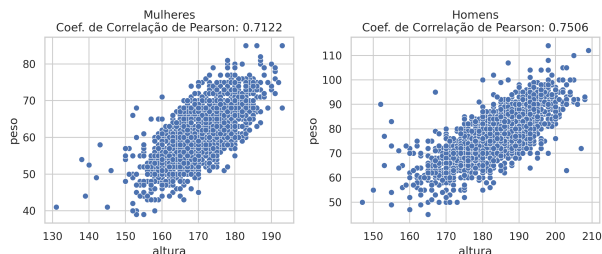
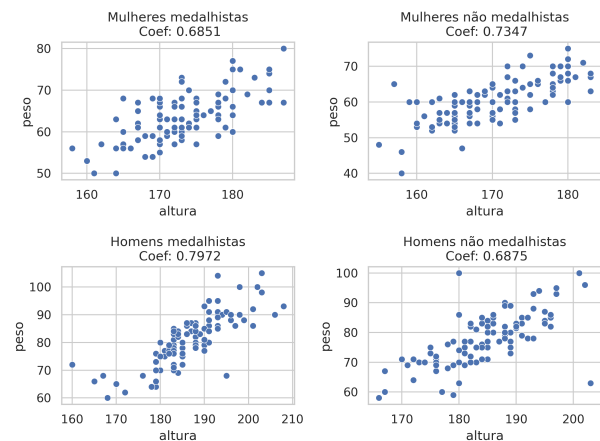


Fig 6: Amostras



Adiante, examinamos a correlação entre peso e altura tanto para a população geral de nadadores quanto para cada amostra específica. A análise revelou uma forte correlação positiva entre peso e altura em todos os casos, sugerindo que, à medida que a altura aumenta, o peso tende a aumentar também.

As figuras 7.1 e 7.2 ilustram a distribuição das variáveis por grupos:

Fig 7.1: Atletas do sexo feminino

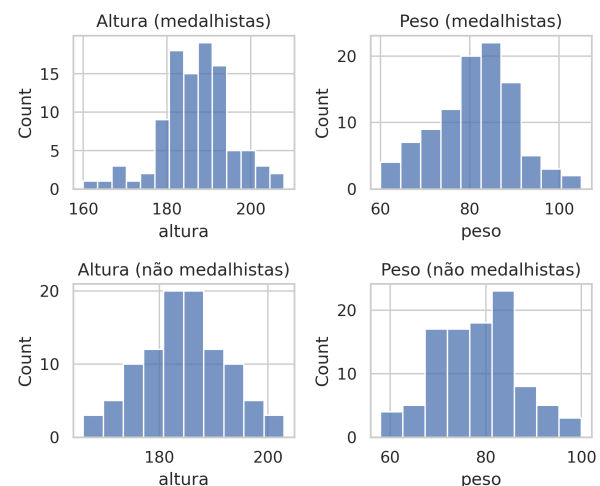
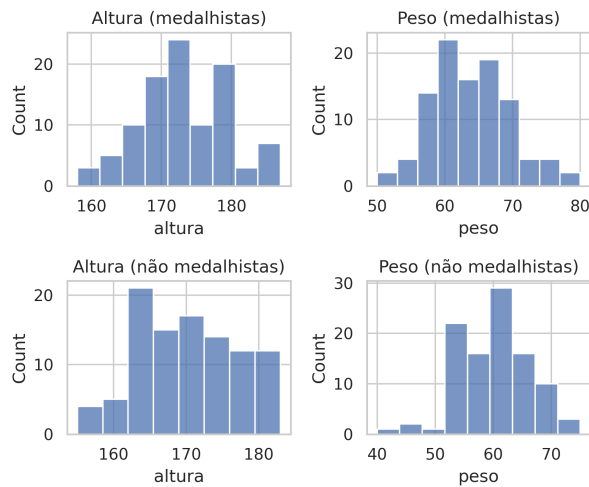


Fig 7.2: Atletas do sexo masculino



Para avaliar a normalidade das variáveis de altura e peso em cada amostra, realizamos o teste de Shapiro-Wilk:

TABLE III
TESTE DE NORMALIDADE SHAPIRO WILK

| Sexo | Medalhista | Altura | Peso |
|------|------------|---------------------|--------------------|
| F | Não | 0.0635407567024231 | 0.2560797333717346 |
| F | Sim | 0.36466866731643677 | 0.5825064778327942 |
| M | Não | 0.9035086035728455 | 0.52390056848526 |
| M | Sim | 0.05040418729186058 | 0.6188229322433472 |

Os resultados indicaram que as variáveis de peso e altura em todas as amostras seguiram uma distribuição normal. Esse foi um resultado favorável, pois a normalidade é um requisito para a realização do teste t com precisão.

Além disso, realizamos o teste de Bartlett para avaliar a homogeneidade das variações entre os grupos:

TABLE IV
TESTE DE HOMOGENEIDADE DE VARIÂNCIA BARLETT

| sexo | variavel | pvalue |
|------|----------|---------------------|
| F | altura | 0.39602271536286415 |
| F | peso | 0.5246105689119875 |
| M | altura | 0.41752951324001075 |
| M | peso | 0.6663197216506267 |

Os resultados do teste indicaram que todos os grupos apresentaram homogeneidade de variâncias, sugerindo que a variabilidade de altura e peso foi semelhante entre os grupos de medalhistas e não medalhistas de ambos os sexos.

Com as condições de pré-requisito satisfeitas, passamos a realizar um teste t independente para comparar a altura e o peso médios entre atletas medalhistas e não medalhistas, separadamente para cada gênero. O teste t independente baseia-se nas seguintes considerações:

- Independência: As amostras são de dois grupos distintos, ou seja, medalhistas e não medalhistas, garantindo a independência entre os grupos.
- Normalidade: O teste de Shapiro-Wilk indicou que as variáveis de altura e peso em todas as amostras seguiram

uma distribuição normal, satisfazendo a suposição de normalidade para realizar o teste t com precisão.

- Homogeneidade da variância: O teste de Bartlett confirmou que as variações de altura e peso eram homogêneas entre os grupos de medalhistas e não-medalhistas para ambos os sexos.

TABLE V
TESTE T DE AMOSTRAS INDEPENDENTES

| sexo | variavel | pvalue |
|------|----------|------------------------|
| F | Altura | 0.0008357441183156177 |
| F | Peso | 0.00019499050689463155 |
| M | Altura | 0.030486501294745432 |
| M | Peso | 0.011380688267562432 |

Os resultados do teste t independente indicaram que não houve diferenças significativas na média de altura e peso entre as mulheres medalhistas e não medalhistas. Da mesma forma, não foram encontradas diferenças significativas na média de altura e peso entre homens medalhistas e não medalhistas. Essas descobertas sugerem que, com base nos dados analisados, a altura e o peso não servem como fatores de distinção que contribuem significativamente para a probabilidade de um atleta ganhar medalhas em competições de natação.

V. RESULTADOS E CONCLUSÃO

Em conclusão, a análise das amostras selecionadas do conjunto de dados abrangente não revelou diferenças significativas de altura e peso entre atletas de natação medalhistas e não medalhistas. Essas descobertas indicam que outros fatores, como técnica, treinamento, habilidade e até mesmo atributos psicológicos, podem desempenhar papéis mais influentes na determinação do sucesso de um atleta em competições de natação. O teste t independente confirmou a ausência de diferenças significativas na altura e no peso entre os grupos, satisfazendo as premissas necessárias de independência, normalidade e homogeneidade da variação. É importante observar que este estudo se limita ao conjunto de dados analisados, e pesquisas futuras devem explorar outras variáveis e fatores que contribuem para o desempenho de um atleta em competições de natação.

Os códigos desta análise podem ser encontrados em: <https://github.com/taferreiraau/LAB-CD/tree/main/P2-Artigo>

REFERENCES

- [1] MON, Daniel et al. Finger flexor force influences performance in senior male air pistol olympic shooting. PloS one, v. 10, n. 6, p. e0129862, 2015.
- [2] BURTON, Robert Weinberg Damon. of Olympic Athletes: An Exploratory Investigation. The Sport Psychologist, v. 14, p. 279-295, 2000.
- [3] SANDS, William A. et al. Historical trends in the size of US Olympic female artistic gymnasts. International Journal of Sports Physiology and Performance, v. 7, n. 4, p. 350-356, 2012.
- [4] PEDRO ALBERTO MORETTIN; WILTON OLIVEIRA BUSSAB. ESTATÍSTICA BÁSICA. [s.l.] Saraiva Educação S.A., 2017.
- [5] LAMBELET COLEMAN, D.; SHREVE, W. Comparing Athletic Performances The Best Elite Women to Boys and Men — Duke University School of Law. Disponível em: <https://law.duke.edu/sports/sex-sport/comparative-athletic-performance/>.