

Projeto fantasma

Introdução

Este projeto tem como objetivo A fim de atender a demanda do cliente, foi feita uma análise estatística descritiva acerca dos atletas que participaram das olimpíadas dos anos de 2000 até 2016.

O banco de dados foi disponibilizado pelo cliente. Foi observado nome, sexo, idade, país, peso, altura, esporte, modalidade e medalha adquirida de 38366 atletas diferentes.

A manipulação e análise dos dados além da confecção das figuras foram feitas com auxílio do software estatístico R versão 4.3.3.

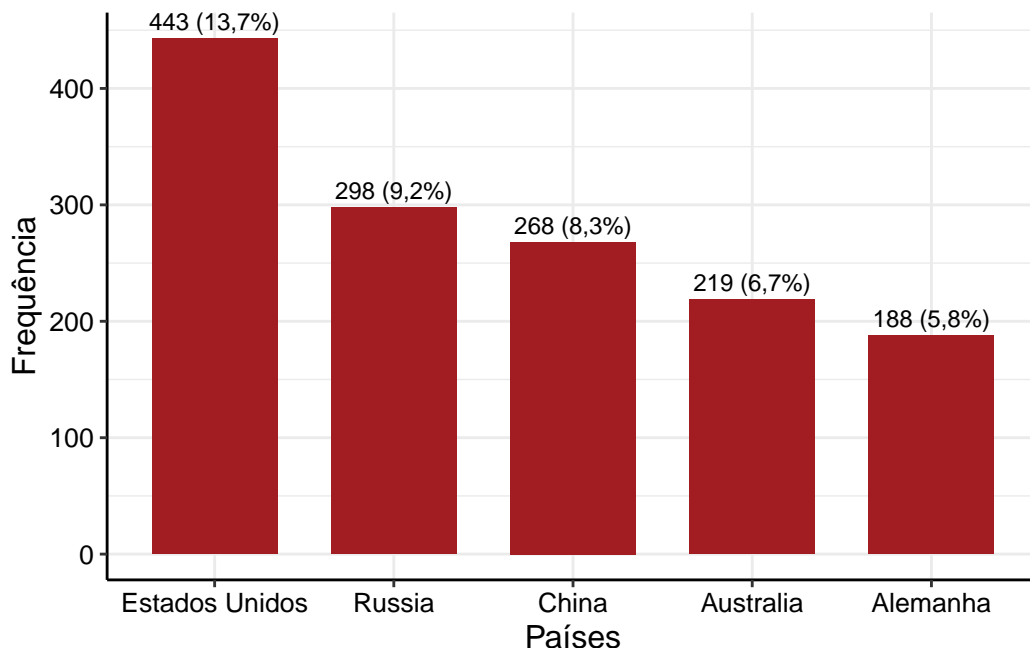
Referencial Teórico

Análises

1. Top 5 países com maior número de mulheres medalistas

Esta análise tem o intuito de identificar quais são os países com maior quantidade de mulheres medalhistas. Para isso foram utilizadas as variáveis sexo e medalhas, a primeira sendo qualitativa nominal e a segunda sendo qualitativa ordinal.

Figura 1: gráfico de colunas do número de mulheres medalistas



Como pode ser observado na figura 1, o ranque é formado por Estados Unidos, Rússia, China, Austrália e Alemanha respectivamente. Os Estados Unidos ocupa o primeiro lugar do ranque com mais de 100 medalhas de diferença da Rússia. Juntos esses 5 países sozinhos tem 43,7% das mulheres medalistas.

2. IMC por esportes

Essa análise tem como objetivo entender o comportamento do IMC nos esportes selecionados. Para tal, foram utilizadas as variáveis esporte e IMC, respectivamente qualitativa nominal e quantitativa contínua. A variável IMC foi obtida dividindo o peso do atleta pela altura ao quadrado. O valor do índice de massa corporal (IMC) é um importante indicador da saúde de uma pessoa. O número representa o quanto a pessoa tem de massa muscular + massa de gordura + massa óssea.

Figura 2: Boxplot do IMC pelo Esporte

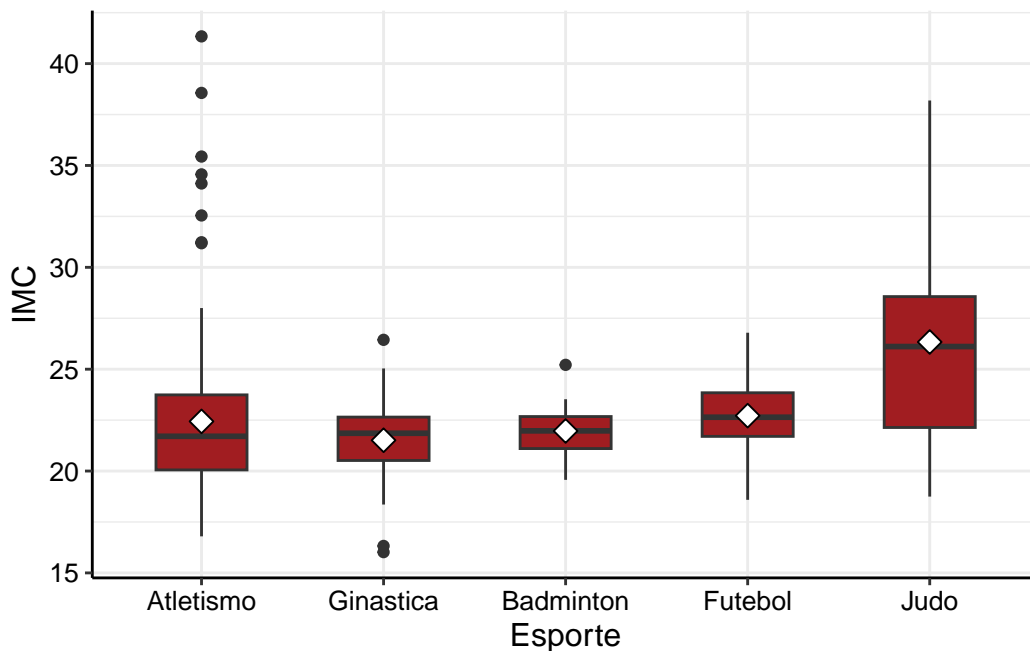


Figura 3

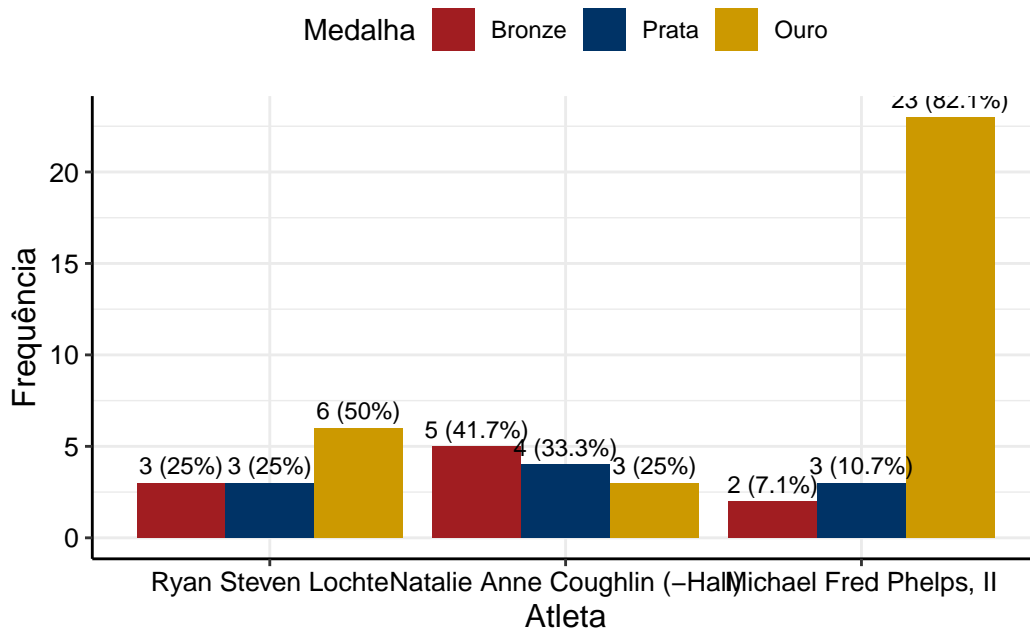
NA

Pode se observar que o IMC segue um comportamento diferente para cada esporte. No judô há diversas categorias para pessoas com pesos diferentes e é um esporte que exige maior massa muscular, o que explica a maior dispersão dos dados (Desvio padrão=6.24) e uma média elevada(26.78) se comparada com as dos outros esportes. A medida de centralidade média no Badminton, na Ginastica, no futebol e no Atletismo são semelhantes: 22.17, 20.76, 22.72 e 22.12 respectivamente, entretanto cada um tem uma configuração única. No atletismo há uma assimetria positiva, na ginastica há uma assimetria negativa e no badminton e no futebol a distribuição é simétrica e centralizada(desvio padrão=1.61). Em geral, quanto mais um esporte ou modalidade exige massa muscular, maior o IMC.

3. Top 3 medalhistas gerais

Esta análise tem como objetivo observar quais são os 3 maiores medalhistas e verificar se há relação entre o medalhista e o tipo de medalha conquistada. Os três atletas que conquistaram mais medalhas nessas 5 edições dos jogos olímpicos foram: Michael Fred Phelps com 28 medalhas, Natalie Anne Coughlin e Ryan Steven Lochte ambos com 12 medalhas. Os três são estadunidenses e tem como esporte a natação.

Figura 4: Gráfico de colunas da quantidade de medalhas pelo tipo da medalha



Observa-se pelo gráfico a relação entre o atleta e o tipo de medalha. Conforme o valor da medalha aumentou, Natalie conquistou menos medalhas, enquanto para Michael e Ryan o comportamento foi o contrário. Também é possível observar que a grande maioria das medalhas que Michael Fred Phelps conquistou foram medalhas de ouro.

4. Relação peso x altura

Esta análise tem o intuito de compreender a relação entre o peso e a altura dos atletas. Para isso, foi utilizado as variáveis peso(Kg) e altura(m), ambas são quantitativas contínuas. O comportamento conjunto das variáveis está ilustrado pelo gráfico de dispersão a seguir:

Figura 5: Gráfico de dispersão do peso pela altura do atleta