Projeto - Fantasma

Consultores Responsáveis:
Bruno Boaventura Xavier

Requerente:
House of Excellence

Brasília, 6 de novembro de 2024.

ESTAT
consultoria estatística

Sumário

1 Introdução

O seguinte projeto tem por objetivo a apresentação das análises estatísticas requisitadas por João Neves, proprietário da House of Excellence, academia de alta performance, que visam trazer informações para otimização do desempenho de seus atletas de elite, que participaram das olimpíadas dos anos de 2000 até 2016. Essas análises consistiram em observar quais eram os cinco países com mais medalhistas femininas, verificar padrões em relação ao IMC dos atletas de determinados esportes, entender se há relação nos três maiores medalhistas entre o atleta e cada tipo de medalha conquistada e, por último, estudar a relação entre peso e altura dos competidores. Portanto, para a produção desse projeto foram abordadas análises descritivas.

Ainda, a base de dados que foi utilizada nesse processo foi o banco Olímpiadas 2000 - 2016, em formato de planilha do Microsoft Excel disponibilizado pelo próprio cliente. Além disso, cabe ressaltar que os dados não são uma amostra probabilística, já que os dados são obtidos de toda a população de atletas, logo planos de amostragem não foram necessários. Para a produção do relatório foram considerados nas análises apenas os atletas que conquistaram medalhas olímpicas, bem como a única variável que não foi utilizada foi o evento em específico de cada atleta.

Assim, por meio destes dados e dos interesses do cliente, a plataforma escolhida para a confecção do projeto foi o *R Studio* na sua versão 4. 4. 1 e somente ela foi utilizada.

2 Referencial Teórico

2.1 Análise Descritiva Univariada

2.2 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:

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

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

$$100 \times f_j$$

2.3 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:

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

2.4 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:

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

2.5 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.

2.6 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.

2.6.1 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

2.6.2 Variância Amostral

Para uma amostra, a variância é dada por:

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

Com:

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

2.7 Desvio Padrão

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

2.7.1 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

2.7.2 Desvio Padrão Amostral

Para uma amostra, o desvio padrão é dado por:

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

Com:

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

2.8 Coeficiente de Variação

O coeficiente de variação fornece a dispersão dos dados em relação à média. Quanto menor for o seu valor, mais homogêneos serão os dados. O coeficiente de variação é considerado baixo (apontando um conjunto de dados homogêneo) quando for menor ou igual a 25%. Ele é dado pela fórmula:

$$C_V = \frac{S}{\bar{X}} \times 100$$

Com:

- $S =$ desvio padrão amostral
- $\bar{X} =$ média amostral

2.9 Coeficiente de Assimetria

O coeficiente de assimetria quantifica a simetria dos dados. Um valor positivo indica que os dados estão concentrados à esquerda em sua função de distribuição, enquanto um valor negativo indica maior concentração à direita. A fórmula é:

$$C_{Assimetria} = \frac{1}{n} \times \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{S}\right)^3$$

Com:

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

2.10 Curtose

O coeficiente de curtose quantifica o achatamento da função de distribuição em relação à distribuição Normal e é dado por:

$$Curtose = \frac{1}{n} \times \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{S} \right)^4 - 3$$

Com:

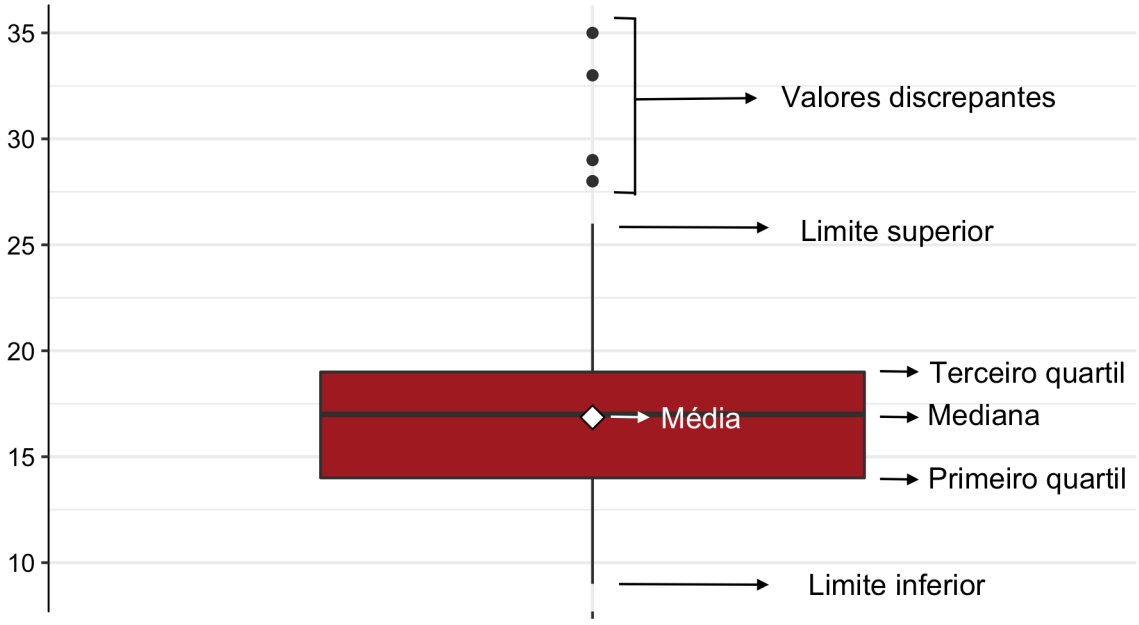
- $X_i =$ i-ésima observação da amostra
- $\bar{X} =$ média amostral
- $S =$ desvio padrão amostral
- $n =$ tamanho da amostra

Uma distribuição é dita mesocúrtica quando possui curtose nula. Quando a curtose é positiva, a distribuição é leptocúrtica (mais afunilada e com pico). Valores negativos indicam uma distribuição platicúrtica (mais achatada).

2.11 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

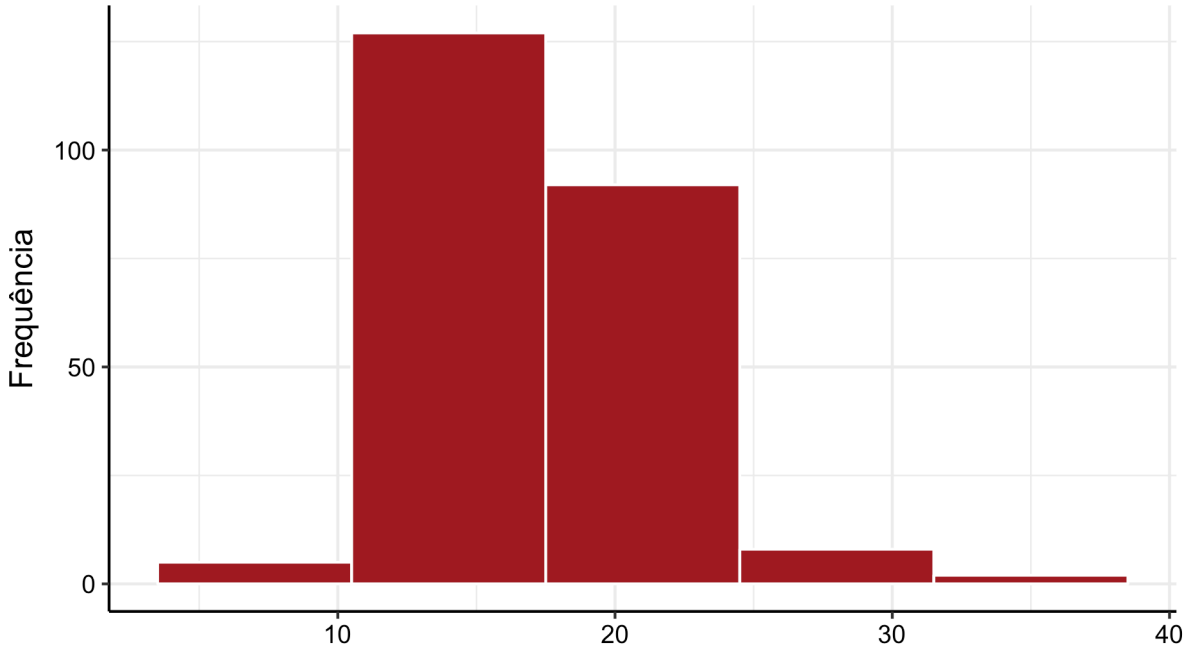


A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada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

2.12 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.

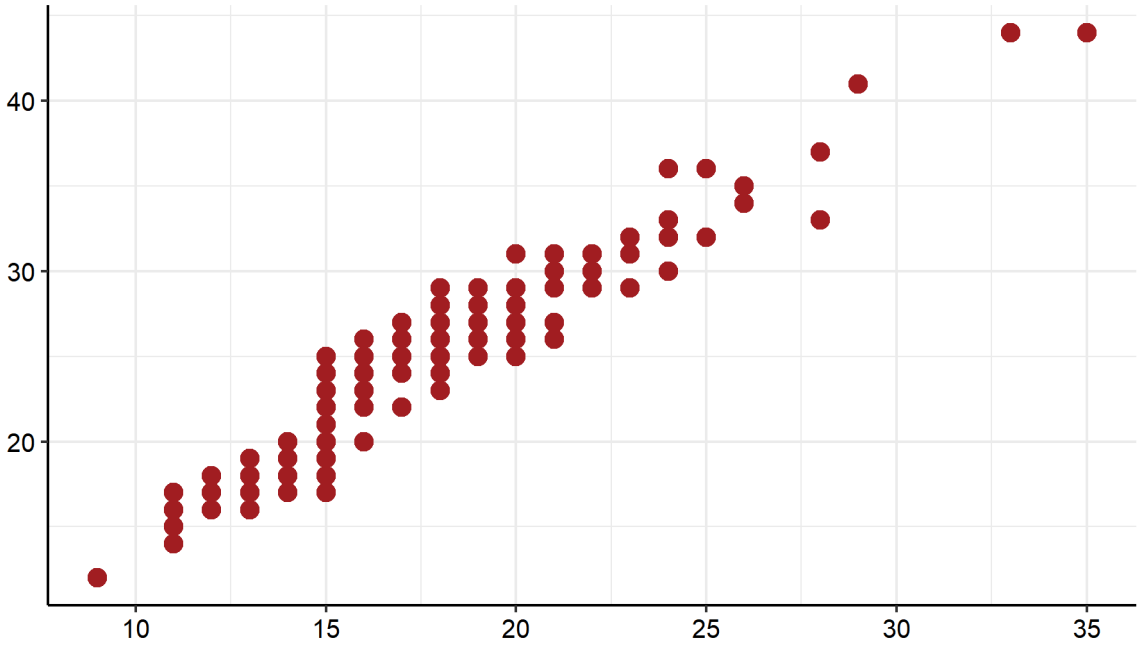
Figura 2: Exemplo de histograma



2.13 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



2.14 Tipos de Variáveis

2.14.1 Qualitativas

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

- **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)

2.14.2 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)

2.15 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:

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

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

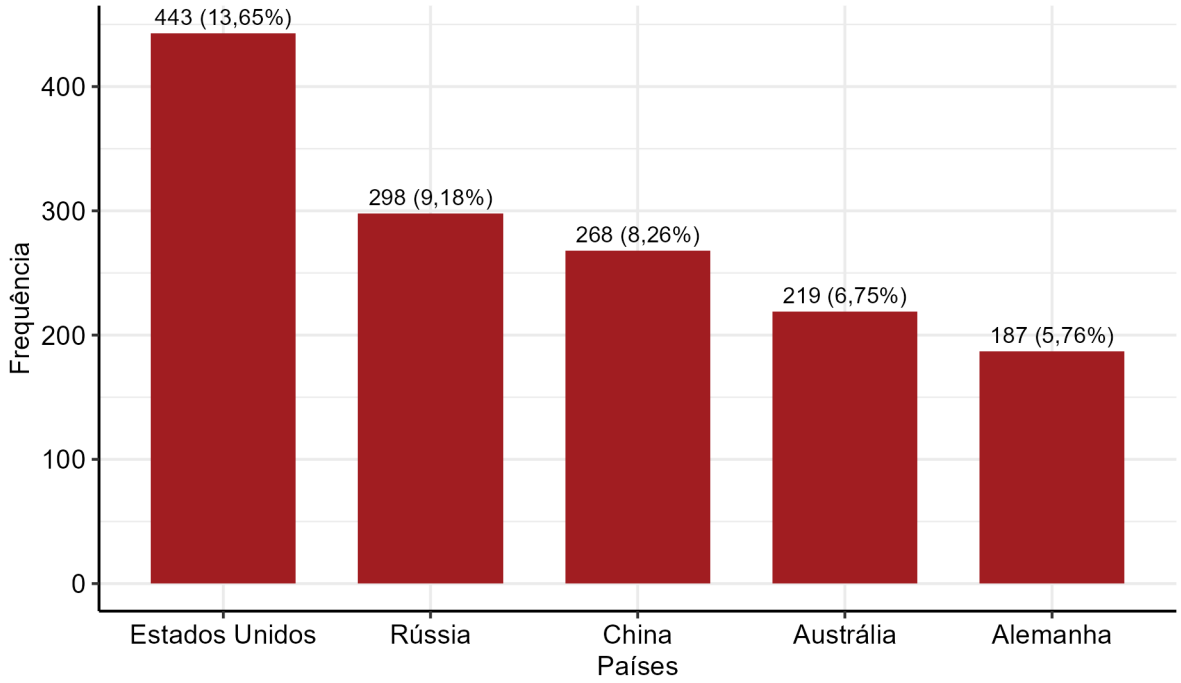
Vale ressaltar que o coeficiente de Pearson é paramétrico e, portanto, sensível quanto à normalidade (simetria) dos dados.

3 Análises

3.1 Top 5 países com mais medalhistas femininas

A partir do acesso ao banco de dados disponibilizado pela “House of Excellence”, tivemos acesso aos dados de todos os atletas que participaram das olimpíadas de 2000 até 2016. Nesse cenário, com o intuito de produzir um ranqueamento das cinco delegações que obtiveram mais conquistas nesse período nas modalidades femininas, do banco de dados original, foram utilizadas as variáveis de sexo, ainda foi utilizada a variável que descrevia qual país era a origem do atleta, sendo essa a base para agrupar as conquistas e, por último, a que continha os nomes das medalhistas, todas elas classificadas como qualitativas nominais, já que não existe ordem entre as categorias.

Figura 4: Gráfico de colunas do total de conquistas de cada delegação do Top 5



Tendo a **Figura 4** como referência podemos analisar que os Estados Unidos lideram a lista com 443 medalhistas, seguidos da delegação russa com 298- o que representa uma diferença de 145 entre o primeiro e o segundo colocado - em terceiro colocado está a China com 268, a Austrália na quarta posição com 219 e fechando o ranqueamento a Alemanha com 187, distanciando-se do penúltimo colocado por 32 atletas e em relação ao primeiro são 256- valor maior que as medalhistas da Austrália que ocupa o quarto lugar.

Através da análise dos dados evidencia-se que dentre as modalidades femininas, durante os anos de 2000 a 2016 considerando todas as atletas que medalharam sem distinção entre ouro, prata e bronze, totalizam-se 3245 medalhistas. Dado o interesse

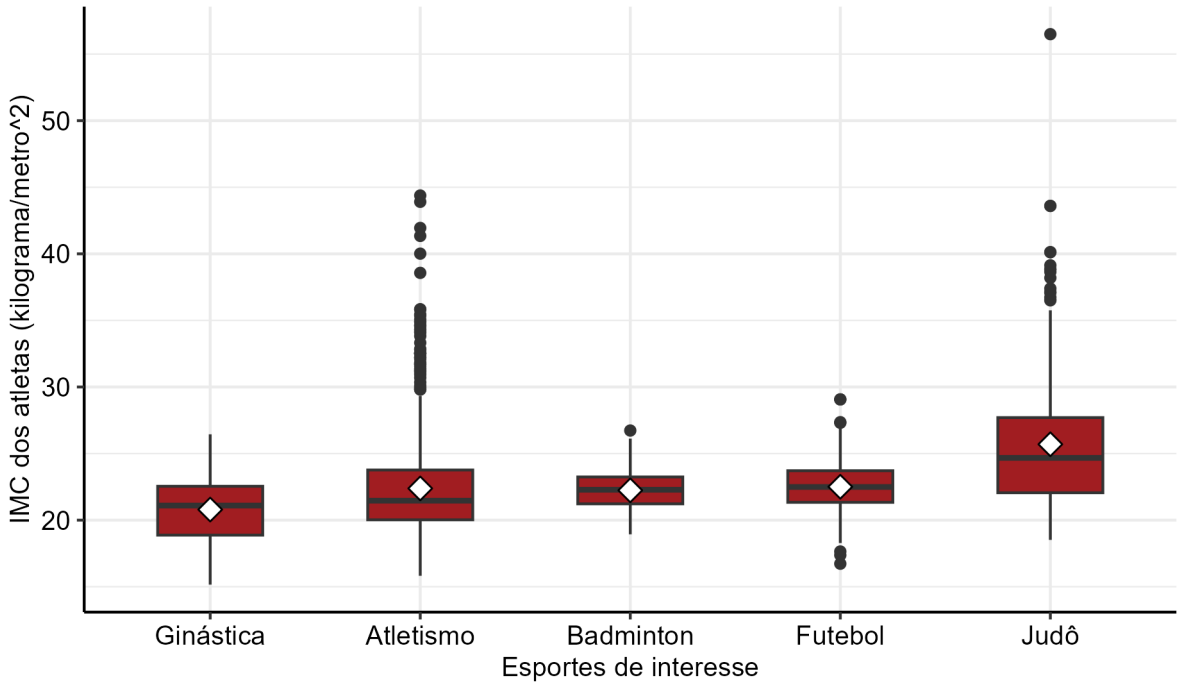
da “House of Excelence” em entender o cenário das conquistas olímpicas femininas, foi contruída uma análise da frequência dessas atletas do *Top* 5 em relação ao total de atletas femininas. Assim por meio da **Figura 4**, no *Top* 5, percebe-se que os Estados Unidos detém o topo do quadro de medalhas com 13,65%, seguido da Rússia que contém 9,18% , a China em terceiro com 8„26%, a Austrália na quarta posição com 6,75% e fechando o ranqueamento a Alemanha com 5,76%.

Para mais, a fim de esclarecer como os melhores países se comparam aos demais, diante da **Figura 4**, nota-se que os cinco países com melhor perfomance nos jogos possuem 1415 atletas, o que condiz a 43,61% de todas as medalhistas, Nesse cenário, analisa-se que as delagações fora do “Top 5”, totalizam 1830, o que representa 56,39% do todo.

3.2 IMC por Esporte

Esta análise tem por objetivo comparar os valores do Índice de Massa Corporal (IMC) entre atletas de diferentes esportes, especificamente ginástica, judô, futebol, atletismo e badminton, para entender a variação entre eles e identificar se algum esporte tende a ter IMCs geralmente menores, maiores ou se não há diferença significativa. Para isso, foi utilizada como base de cálculo do IMC as variáveis quantitativas contínuas que descrevem o peso dos atletas em libras (lbs) e suas respectivas alturas em centímetros (cm), tendo isso, o IMC de cada um dos atletas foi calculado, realizando as conversões de unidade para estabelecer o padrão do índice que é kg/m².

Figura 5: Boxplot do IMC dos atletas pelos esportes de interesse



Quadro 1: Medidas resumo do IMC por esportes

Estatística	Atletismo	Badminton	Futebol	Ginástica	Judô
Média	22,38	22,24	22,51	20,79	25,70
Desvio Padrão	3,97	1,52	1,73	2,40	5,12
Variância	15,75	2,32	2,99	5,75	26,23
Mínimo	15,82	18,94	16,73	15,16	18,52
1º Quartil	20,03	21,22	21,34	18,88	22,06
Mediana	21,46	22,28	22,49	21,10	24,68
3º Quartil	23,77	23,24	23,71	22,54	27,70
Máximo	44,38	26,73	29,07	26,45	56,50
Coeficiente de Variação	17.74%	6.83%	7.69%	11.54%	19.92%

A partir do **??** e da **Figura 5** pode-se observar que o Judô tende a ter valores de IMC maiores que os demais por ter seus valores de centralidade - média e mediana - e de primeiro quartil maiores, o que significa que a maior parte dos dados possui valores acima dos outros esportes. Além disso, o Judô é o esporte que tem maior dispersão dos índices, percebido pelo maior coeficiente de variação e pelo intervalo interquartílico ser o maior, intervalo esse que é a distância entre o primeiro e o terceiro quartil - limite inferior e superior da caixa - bem como uma leve assimetria positiva, aquela que ocorre quando ocorre maior frequência de valores entre o primeiro quartil e a median. Contudo, ao observar os valores extremos, conclui-se que a ocorrência deles gera deslocamento da média para esses valores maiores.

Para os outros esportes, analisar existem tendências de IMCs menores fica menos nítido, como é no caso do Judô, uma vez que a Ginástica mesmo que com menor média possui assimetria negativa, quando os valores se tornam mais frequentes entre a mediana e o terceiro quartil - limite superior da caixa - e uma média pouco afetada por valores extremos, bem como ter uma amplitude interquartílica maior se comparada ao Badminton e Futebol. Já para o Atletismo, seus valores de quartis, de mediana são maiores, ao mesmo tempo que possui assimetria positiva, logo maior concentração entre primeiro quartil e mediana, se aproximando da Ginástica que tinha mais concentração para cima da caixa e, ainda, contém significativa amplitude como é a Ginástica e seu valor médio é afetado pelos diversos *outliers* acima do máximo estipulado no gráfico. Dessa forma, embora a Ginástica conter o menor valor de média e mediana, dificulta-se a análise de que tende a ter os menores índices por ambos esportes terem uma notória amplitude interquartílica, a Ginástica ter maior frequência para valores próximos do terceiro quartil e o Atletismo ter maior ocorrência dos valores próximos ao primeiro quartil, nota-se que é possível observar proximidade dos valores dos dois esportes.

Por outro lado, ao observar a dispersão e assimetria, percebe-se que o badminton é o esporte mais concentrado em relação aos demais, já que sua caixa é a mais

achatada, o que indica que a maioria dos valores gira em torno da média - não afetada por valores extremos - e da mediana, que para esse caso tem todos os seus dados praticamente simétricos, já que seus valores de centralidade são signifcativamente próximos e a mediana próxima da metade do boxplot. Poranto, evidencia-se que os índices para o Badminton são concentrados, pouco assimétrico e, por isso, é plausível compreender uma tendência do esporte conter majoritariamente valores em torno da média. Caso que também pode ser percebido para o futebol, entretanto com menor intensidade, posto que possui maior amplitude interquartílica e mais influência de *outliers.*

3.3 Os 3 maiores medalhistas

Para a análise buscou-se entender quais são os 3 medalhistas com maior número de medalhas no total, e dentre eles, observar a quantidade de cada tipo de medalha que cada um destes atletas conquistou, sendo ouro, prata ou bronze. Nesse caso, utilizou-se duas variáveis, o atleta em si, identificado como qualitativa nominal e a quantidade de medalhas de cada tipo (Bronze, Prata e Ouro), dada por uma variável qualitativa ordinal, uma vez que suas categorias possuem ordens entre si.

Figura 6: Gráfico de colunas da frequência de cada tipo de medalha dos 3 maiores medalhistas

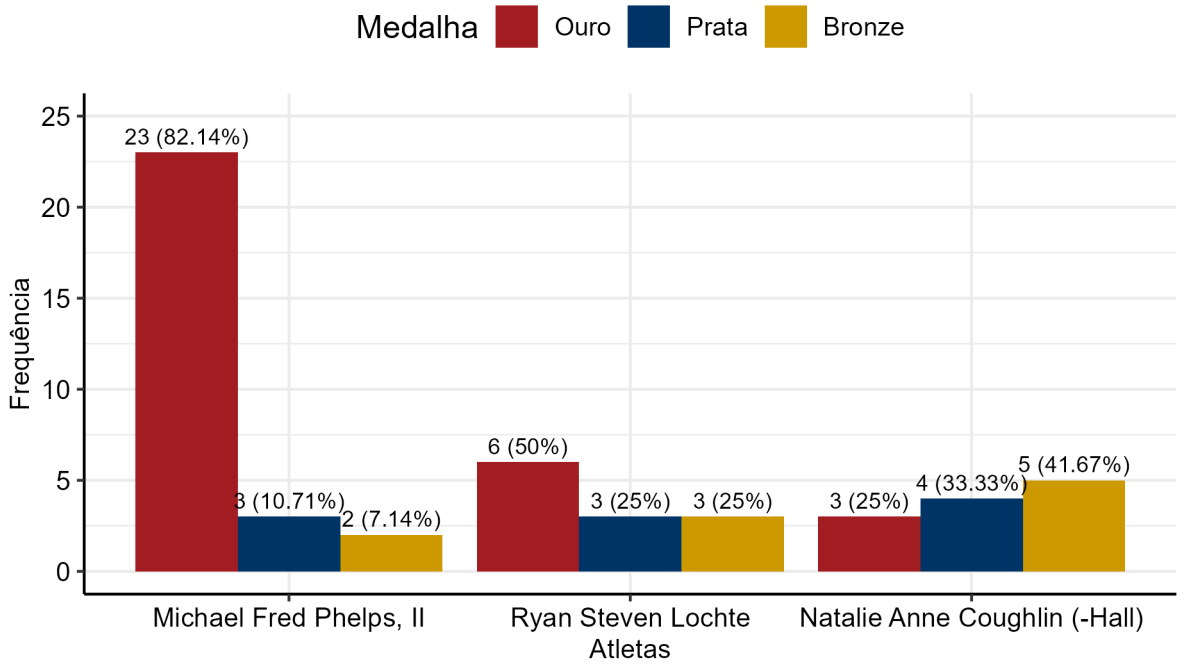


Tabela 1: Valores que representam a quantidade de cada um tipo de medalha para os 3 maiores medalhistas

Atletas	Ouro	Prata	Bronze	Total
Natalie Coughlin	3	4	5	12
Michael Phelps	23	3	2	28
Ryan Lotche	6	3	3	12

Através da **Figura 6**, nota-se que por haver diferenças marcantes da quantidade de medalhas de ouro do atleta dos Estados Unidos, Michael Phelps, em relação aos demais, pode-se sugerir que há uma aparente correlação visual entre as variáveis. Caso contrário, se as quantidades de medalhas fossem semelhantes entre os atletas, poderia-se supor que não há uma forte associação entre os atletas e os tipos de medalhas, porém não é esse o caso dos *Top* 3 medalhistas. Essa mesma percepção de correlação pode ser percebida nas frequências, se um atleta conquistar uma proporção muito maior de um tipo de medalhas em específico - muito mais ouros a exemplo - pode-se sugerir um caso de associação, esse é o caso da **Figura 6**, a qual Phelps e Ryan Lochte se destacam por uma proporção diferenciada de medalhas de ouro, evidenciando uma plausível correlação entre esses atletas e o tipo de medalha.

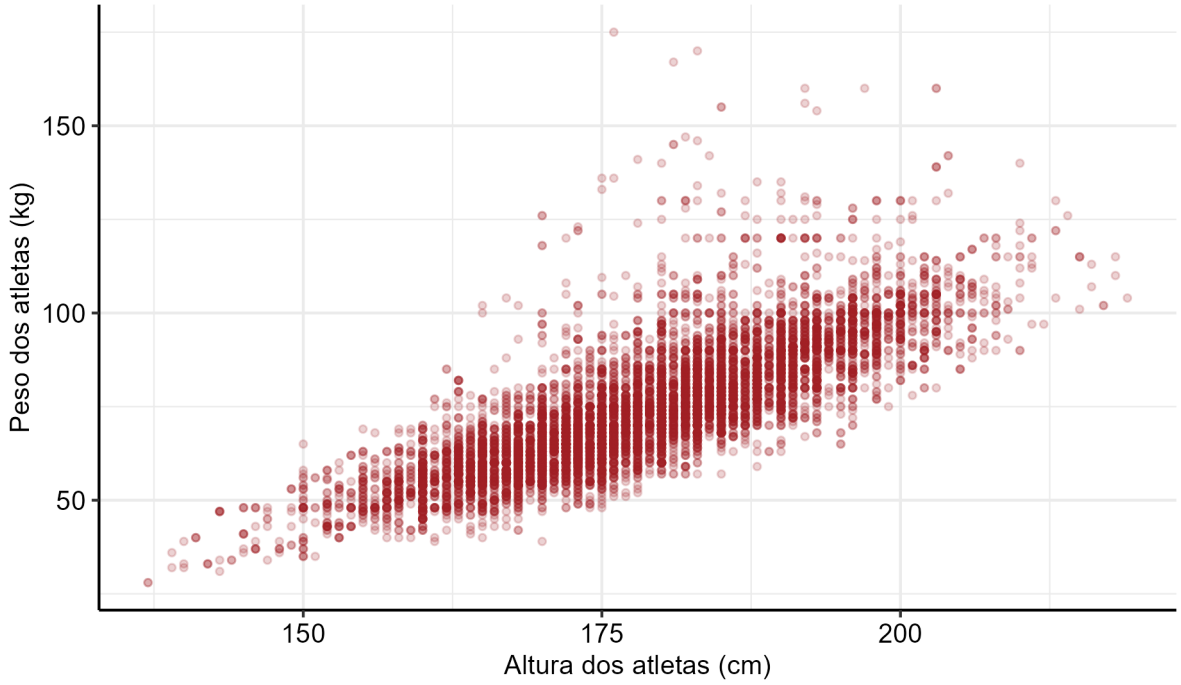
Por meio da **Tabela 1** nota-se que Michael Phelps tem um valor altamente discrepante em relação aos demais, uma vez que seu total ultrapassa os outros dois juntos, somente de ouro ele possui 23, o que significa que caso somadas todas as conquistas dos outros dois maiores medalhistas, totalizando 24 medalhas, Phelps perderia apenas por 1 medalha, suas conquistas de ouro representam 82,14% de todos os seus pódios, evidenciando uma alta efetividade em suas conquistas. Ainda, nota-se que Ryan Lotche, mesmo que não contendo tantas medalhas comparadas às de Phelps, destaca-se por 50% das duas 12 medalhas serem de ouro, e suas demais conquistas representarem 25% cada, mostrando uma alta representatividade dos ouros. Já para Natalie Coughlin sua maior frequência de medalhas são os bronzes com 5 conquistas (41,67%) e com 3 de ouro (25%) de seus pódios olímpicos.

3.4 Variação Peso por Altura

O intuito neste momento é entender a relação entre o peso e altura dos medalhistas olímpicos, isso quer dizer que se à medida que o peso aumenta, a altura também aumenta? Ou o contrário, ou não tem diferença, ou não dá pra inferir nada? Dessa forma, a fim de responder essas perguntas foram utilizadas a variável que armazena os pesos dos atletas em kilogramas e a que representa as alturas em centímetros, ambas as variáveis são classificadas como quantitativas contínuas, para que então seren realizadas as análises estatisticas. Diante desses dados, visando atingir este

interesse, foram construídos os gráficos e o quadro a seguir.

Figura 7: Gráfico de dispersão da altura pelo peso dos atletas



Quadro 2: Medidas resumo da Altura (m) e do Peso (kg) dos atletas

Estatística	Altura	Peso
Média	178,24	74,00
Desvio Padrão	11,80	16,26
Variância	139,23	264,26
Mínimo	137,00	28,00
1º Quartil	170,00	62,00
Mediana	178,00	72,00
3º Quartil	186,00	84,00
Máximo	219,00	175,00
Coeficiente de Variação	7%	22%

Quadro 3: Coeficiente de correlação de Pearson para Altura (m) e Peso (kg) dos atletas

Estatística	Valor
Coeficiente de Correlação	0,805

Ao observar a **Figura 7**, é perceptível que a maioria dos atletas, os quais são representados pelos pontos no gráfico estão agrupados abaixo das cem kilogramas, tal análise também pode ser feita percebendo que a maioria possui altura inferior aos a

Figura 8: Boxplot da altura dos atletas

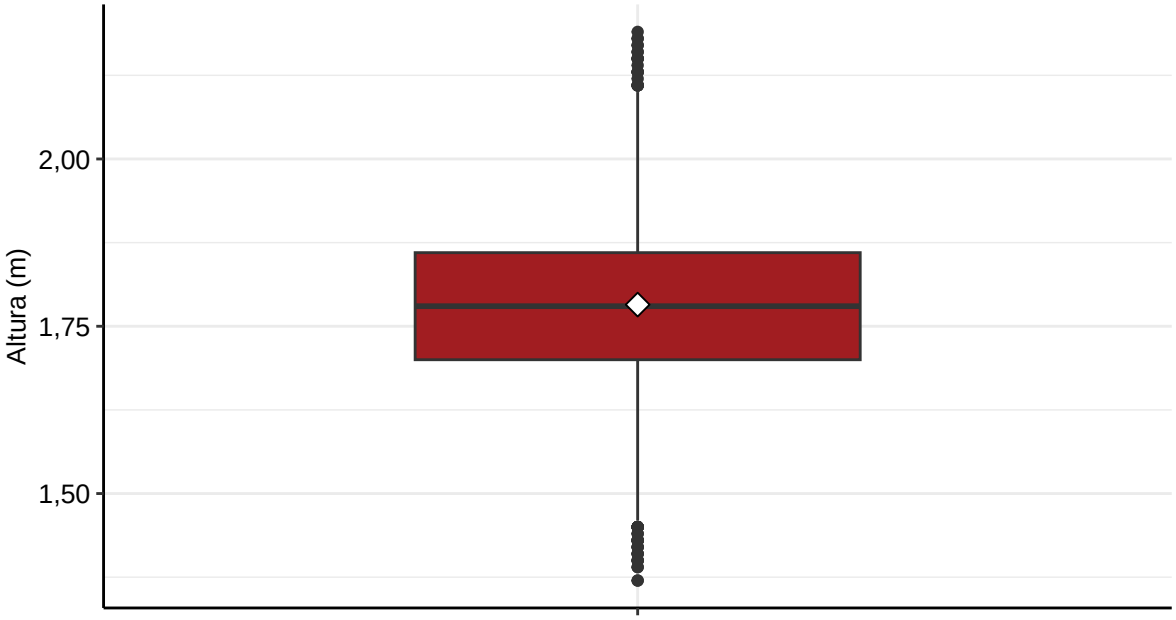
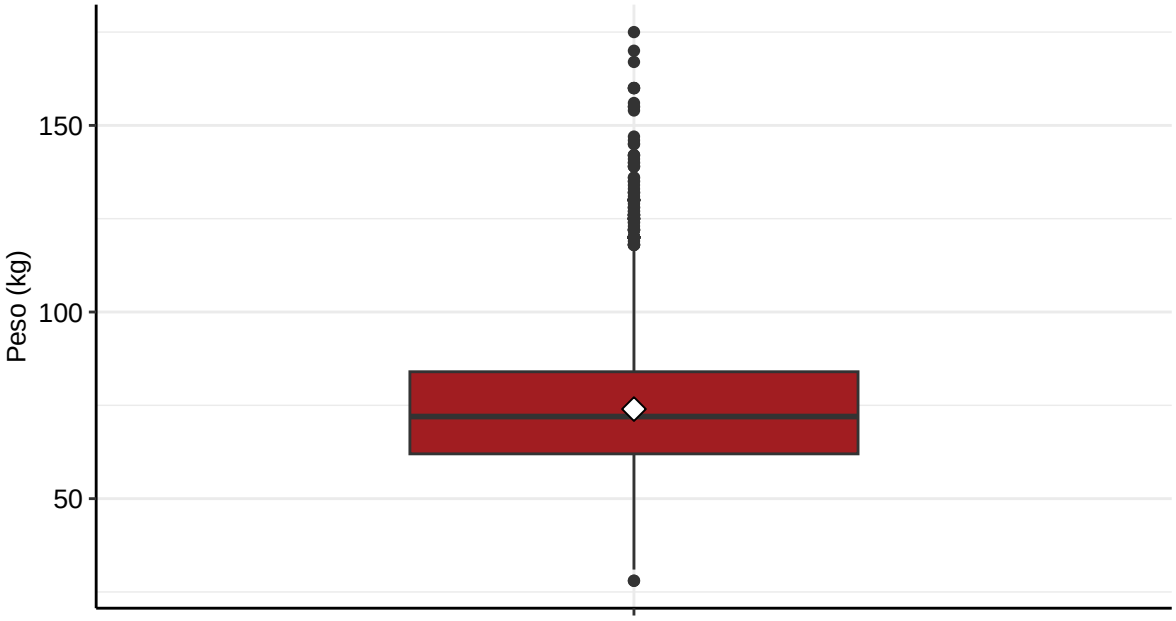


Figura 9: Boxplot do peso dos atletas



dois metros. Agora, visualmente, analisando a dispersão dos pontos na imagem é fácil compreender que a maior concentração deles sugere que quanto maior a altura dos atletas maior será seu peso, uma vez que o conjunto de pontos mais próximos uns dos outros supõem uma correlão positiva forte dos dados, isso acontece quando os pontos se assemelham a uma reta crescente. Ainda, nota-se que essa relação é evidenciada pelo **??**,uma vez que para o coeficiente de correlação de Pearson, valores próximos de 1 e -1 indicam associação, neste caso por obter 0,805 as duas variáveis são correlacionadas no sentido de serem diretamente proporcionais, o que quer dizer que ao passo que a Altura aumenta o peso também aumenta, logo a associação linear é de 80,53%.

Através da **Figura 8** pode-se perceber que pelo valor de mediana e de média estarem bastante próximos uns dos outros - como foi observado na **??** e visualizado agora por meio da **Figura 8** - concluisse que existe simetria dos valores, que significa dizer que entre o primeiro e o terceiro quartil os valores estão distribuídos relativamente de forma simétrica em torno da média e da mediana. Ainda, nota-se que os valores de máximo e mínimo apresentam distancia semelhante em relação à caixa, além de possuir valores extremos que também se distribuem de maneira semelhante tanto para cima, quanto para baixo.

Diante da **Figura 9** nota-se, diferentemente da altura, haver maior assimetria dos dados. Em primeira análise, destaca-se do gráfico que sua mediana está distante da sua média, o que representa, neste caso, que há assimetria positiva, isso ocorre quando existe maior ocorrência de valores que númericamente são próximas no intervalo entre o primeiro quartil e a mediana. Para mais, observando os valores extremos, percebe-se que existe maior quantidade deles a acima do máximo estipulado na contrução do gráfico, tal fato corrobora à análise da **??** sobre a média ser influenciada por esses valores extremos e se distanciar da mediana.

Para mais, estudando a **??** destaca-se que a média de altura dos atletas é de 1,78 metros, assim como sua mediana - termo esse que divide em 50% porcento todas as observações em ordem cresecente, logo conclui-se que para este caso a média não tem seus valores afetados por valores extremos -ainda, verificam-se valores de mínimo em 1,37 e máximo em 2,19. Contudo, mesmo com uma disparidade alta entre máximo e mínimo, nota-se que os dados estão bastante concentrados, o que pode ser percebido pelo desvio padrão de 0,12 metro. Além disso, ao analisar a medida do coeficiente de variação - índice que determina quando o desvio padrão representa em relação à média - observa-se apenas um valor de 7%, corroborando ao entendimento de uma alta homogeneidade dos dados e uma baixa dispersão da altura.

Outrossim, por meio da **??** analisa-se um valor médio de 74 kilogramas e com sua mediana em 72 kg, evidenciando que a média é influencidada por valores extremos. Ainda, destacam-se a mínima ser de 28 kg e a máxima de 175 kg. Diferentemente da

altura, aqui percebe-se haver uma maior dispersão dos valores de peso, isso porquê o desvio padrão aqui passa a ser de 16,26 kg e analisando seu coeficiente de variação em 22%, pela teoria, valores menores que 25% são considerados homogêneos, dessa forma, a dispersão neste caso está relativamente próxima do limite para ser homogênea.

4 Conclusões

A partir do exposto, observa-se que Estados Unidos, Rússia, China, Austrália, Alemanha como os países com mais medalhistas olímpicas, nessa respectiva ordem, percebendo uma maior distância entre uma delegação e outra no caso do primeiro e segundo lugar. Ainda, nota-se que que juntos eles somam 43,61% das atletas com conquistas.

No caso da distribuição do IMC para aqueles esportes de interesse, observa-se maior tendência do Judô obter maiores valores, mesmo com maior variabilidade. Para a tentativa de identificar o que teria menor índice, a dispersão e assimetria da Ginástica e do Atletismo dificultam essa abordagem. Contudo, para o Badminton é possível perceber tendência de a maioria de seus valores ocorrerem próximos dos valores de tendência central.

Ao analisar os três maiores medalhistas e seus tipos de conquista, evidencia-se que por haver diferenças marcantes na quantidade e proporção de ouro de Ryan Lotche e, principalmente, de Michael Phelps, sugere-se uma correlação entre as variáveis. Ainda, destaca-se uma discrepante quantidade de medalhas de ouro entre Phelps e Lotche para a terceira colocada.

Por último, na quarta análise, destaca-se uma perceptível correlação positiva forte, ao fato do peso crescer ao passo que a altura cresce. Além disso, concluisse que a altura é mais concentrada e menos assimétrica se comparada aos valores de peso dos atletas.