

Relatório Olimpíadas 2000 - 2016

Consultores Responsáveis:

Felipe Bretas

Requerente:

João Neves

Brasília, 2 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 Variância | 5 |
| 2.4.1 Variância Populacional | 5 |
| 2.4.2 Variância Amostral | 6 |
| 2.5 Desvio Padrão | 6 |
| 2.5.1 Desvio Padrão Populacional | 6 |
| 2.5.2 Desvio Padrão Amostral | 7 |
| 2.6 Boxplot | 7 |
| 2.7 Gráfico de Dispersão | 8 |
| 2.8 Tipos de Variáveis | 8 |
| 2.8.1 Qualitativas | 8 |
| 2.8.2 Quantitativas | 9 |
| 2.9 Coeficiente de Correlação de Spearman | 9 |
| 3 Análises | 10 |
| 3.1 Análise 1 | 10 |
| 3.2 Análise 2 | 11 |
| 3.3 Análise 3 | 12 |
| 3.4 Análise 4 | 13 |
| 4 Conclusões | 15 |

| | total_med4\$Medalha | | |
|-------------------------------|---------------------|------|--------|
| total_med4\$Nomes | Bronze | Gold | Silver |
| Michael Fred Phelps, II | 2 | 23 | 3 |
| Natalie Anne Coughlin (-Hall) | 5 | 3 | 4 |
| Ryan Steven Lochte | 3 | 6 | 3 |

[1] 0,8408162

1 Introdução

Este relatório tem como objetivo um estudo mais aprofundado, por meio de análises descritivas, sendo também analisadas suas correlações, sobre a performance de atletas nas Olimpíadas dos anos 2000 até 2016, a fim de entender melhor o desempenho dos atletas de elite da academia de alta performance “House of Excellence”.

O banco de dados foi coletado pelo próprio cliente, possuindo uma boa amostra para fins de estudo contendo um total de 9 variáveis para a confecção da análise, possuindo variáveis tanto qualitativas quanto quantitativas.

O software utilizado para a realização desta análise é o R, sendo este um software livre em sua versão 4.3.3.

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:

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

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.

2.4 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.

2.4.1 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
- μ = média populacional
- N = tamanho da população

2.4.2 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

2.5 Desvio Padrão

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

2.5.1 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
- μ = média populacional
- N = tamanho da população

2.5.2 Desvio Padrão Amostral

Para uma amostra, o desvio padrão é dado por:

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

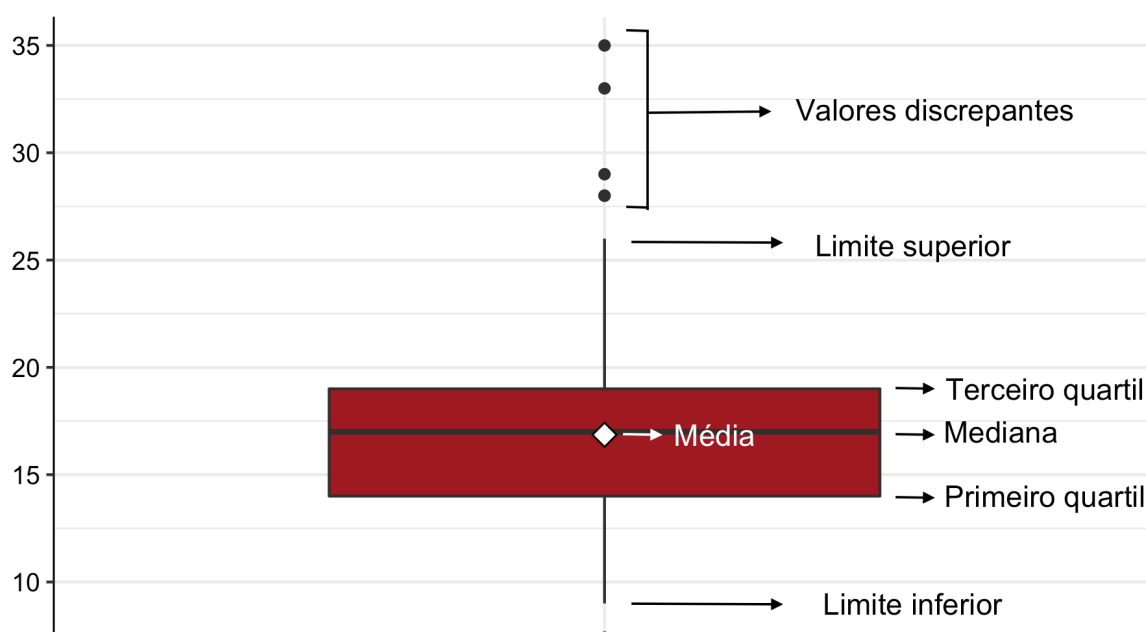
Com:

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

2.6 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



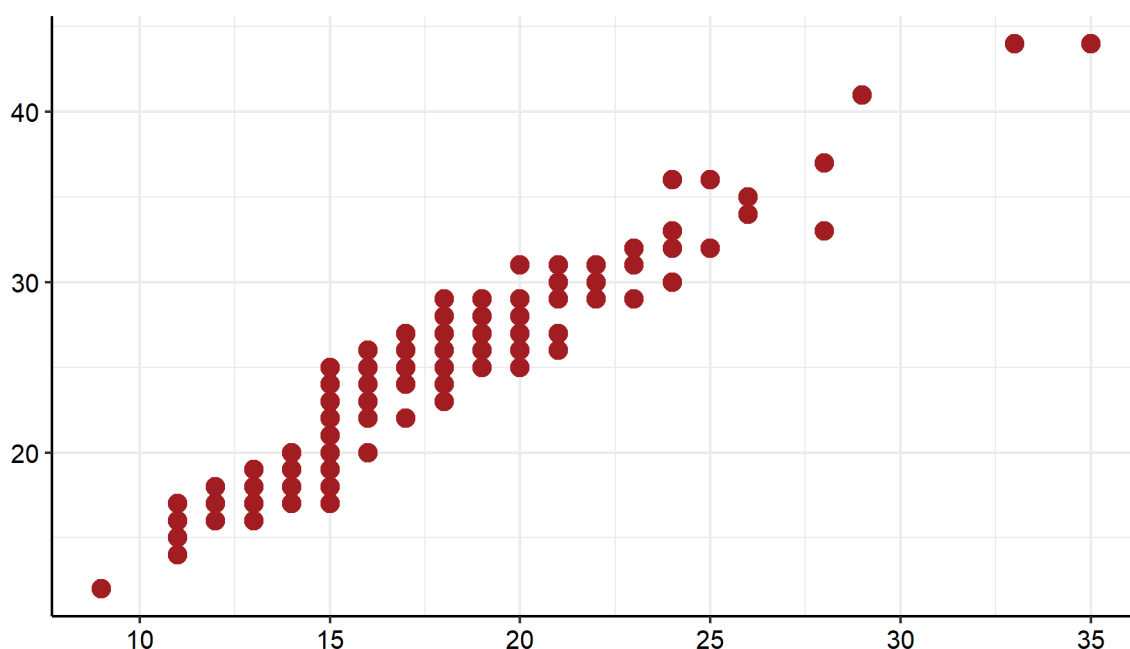
A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada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

2.7 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 2: Exemplo de Gráfico de Dispersão



2.8 Tipos de Variáveis

2.8.1 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)

2.8.2 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)

2.9 Coeficiente de Correlação de Spearman

O coeficiente de correlação de Spearman é uma medida não paramétrica que verifica, através de postos de variáveis quantitativas ou qualitativas ordinais, o grau de relação linear entre duas variáveis. 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 ρ é negativo, diz-se existir uma relação de grandeza inversamente proporcional entre as variáveis. Analogamente, quando ρ é positivo, diz-se que as duas variáveis são diretamente proporcionais.

O coeficiente é calculado da seguinte maneira:

$$\rho_{Spearman} = \frac{\sum_{i=1}^n \left[\left(R(x_i) - \frac{n+1}{2} \right) \left(R(y_i) - \frac{n+1}{2} \right) \right]}{\sqrt{\sum_{i=1}^n (R(x_i)^2) - n \left(\frac{n+1}{2} \right)^2} \times \sqrt{\sum_{i=1}^n (R(y_i)^2) - n \left(\frac{n+1}{2} \right)^2}}$$

Onde:

- x_i = i-ésimo valor da variável X
- y_i = i-ésimo valor da variável Y
- $R(x_i)$ = posto relativo à observação i de X
- $R(y_i)$ = posto relativo à observação i de Y
- n = número total de observações na amostra

3 Análises

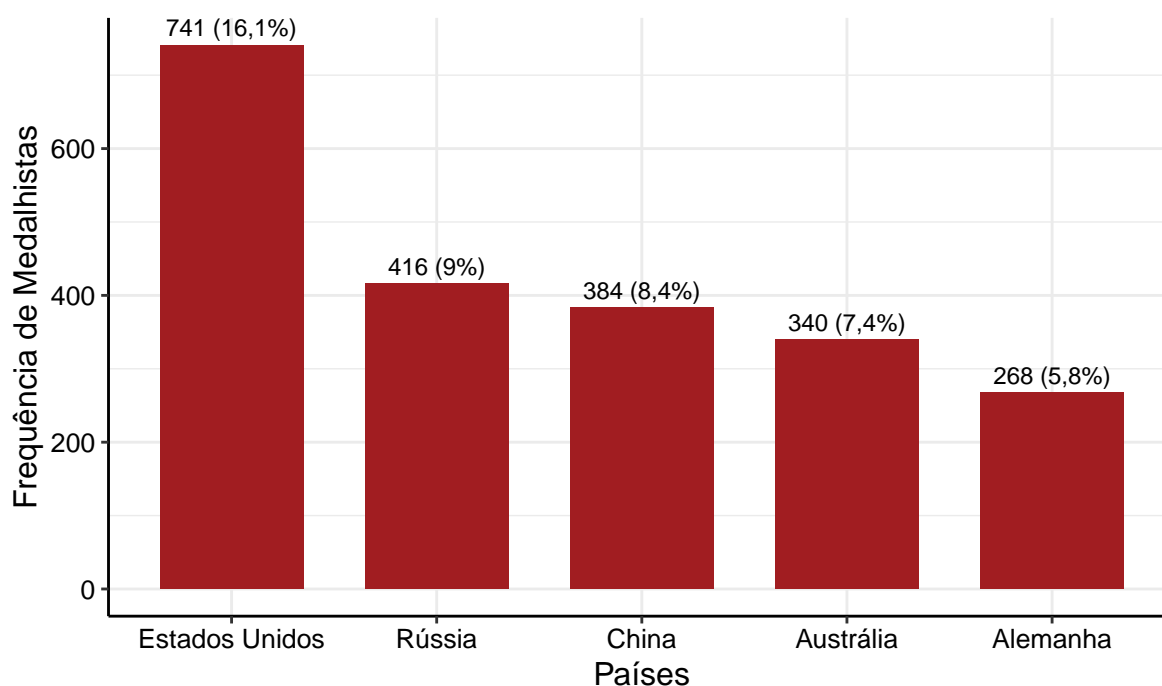
3.1 Análise 1

Desde o século XX, as mulheres vêm ganhando espaço em diversas áreas de atuação que antes eram consideradas exclusivamente dos homens. Uma conquista muito grande do público feminino foi o ingresso nas Olimpíadas e, com a virada do século, o investimento se tornou ainda maior.

Para a confecção desta análise, foram utilizadas as variáveis: “Sexo” (qualitativa nominal), “Medalha” (qualitativa ordinal) e “País” (qualitativa nominal) que significam respectivamente, o gênero dos atletas participantes; se o atleta ganhou medalha e seu tipo, sendo estes ouro, prata ou bronze; e o país que o atleta representou.

Este estudo tem como finalidade compreender quais foram os países que tiveram o maior número de medalhistas femininas nas Olimpíadas de 2000 até 2016, analisando os 5 que se saíram mais vencedores nessas edições, como é mostrado na figura a seguir:

Figura 3: Gráfico de colunas do total de medalhas femininas por país



Como pode ser observado pelo @figura-1, os Estados Unidos possui o maior número de medalhas quanto se trata de atletas femininas, com um total de 146, representando 16,1% no total de medalhistas mulheres das Olimpíadas de 2000 até 2016, havendo uma diferença de 7,2% para o segundo colocado Rússia.

Algo que também pode ser observado é a grande variedade de continentes presentes dentro do top 5 medalhistas femininas, sendo estes: América, Europa, Ásia e

Oceania, mostrando que existe uma grande variedade muito grande entre as atletas ganhadoras.

Essa diferença pode ser explicada pelo alto investimento que os Estados Unidos faz em suas atletas desde a infância, as incentivando a praticarem esportes de alto rendimento em toda sua fase de educação.

Algo que também deve ser observado é a frequência relativa acumulada entre esses 5 países mais vencedores os quais equivalem à 46,7% do total de medalhas ganhas por atletas femininas em todas as Olimpíadas em análise.

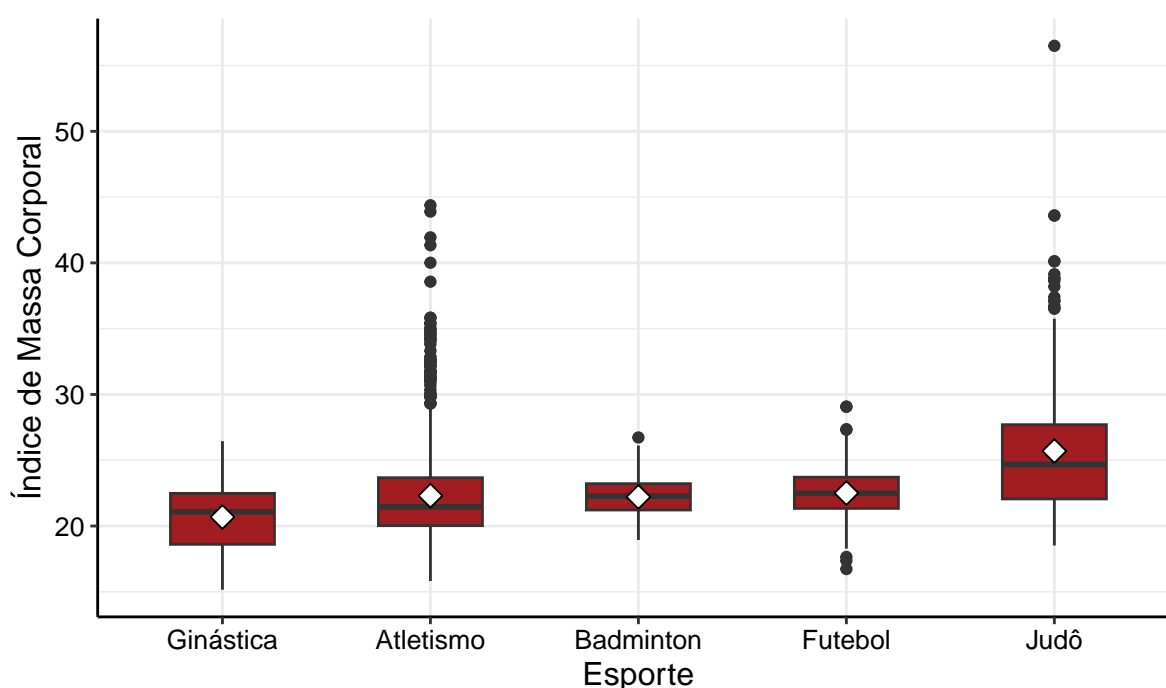
3.2 Análise 2

O Índice de Massa Corporal, mais conhecido como IMC, é uma medida internacional adotada para verificar se uma pessoa estaria numa certa faixa de “peso ideal”. Esse índice varia por diversos aspectos e, entre um deles, está se a pessoa pratica ou não algum esporte.

Para esta análise, foram utilizadas as variáveis: “Altura (cm)” (quantitativa contínua), “Peso (kg)” (quantitativa contínua) e “Esporte” (qualitativa nominal), significando respectivamente: altura de cada atleta participante em centímetros; peso de cada atleta em quilogramas; esporte que o competidor participou.

A fim de tentar compreender melhor essa relação, foi feita uma análise descritiva entre alguns dos esportes olímpicos e seus atletas, demonstrado pela figura abaixo:

Figura 4: Boxplot do IMC por esporte



Como pode ser observado pelo @figura-2, os atletas de ginástica são os que apre-

sentam menor mediana em relação aos outros esportes. Uma explicação para esse motivo se dá pelo fato dos atletas de ginástica precisarem muito de sua mobilidade e agilidade para executarem suas acrobacias, necessitando de uma menor taxa de IMC para facilitar os movimentos.

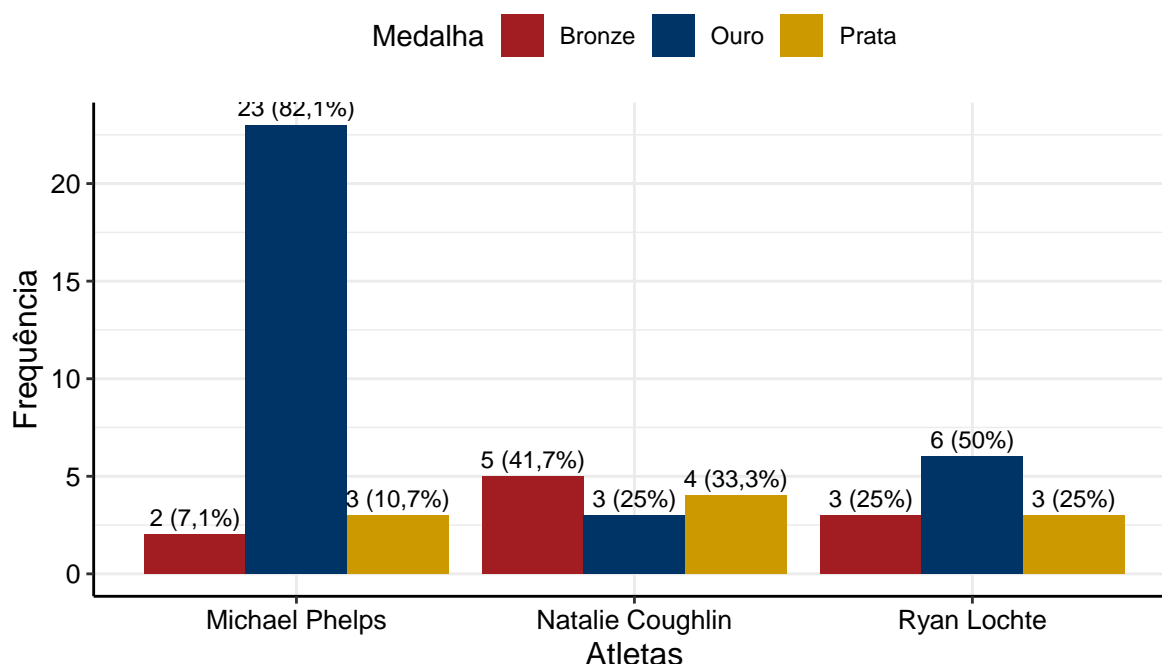
Algo que também vale a pena ser observado é o grande número de outliers nos competidores de atletismo, evidenciando a diversidade possível para os atletas entrarem em suas competições. Porém, mesmo com essa grande presença desse outliers, ainda apresentam a segunda menor mediana de IMC por atletas por esporte.

3.3 Análise 3

O sonho de cada atleta que adentra uma Olimpíadas é conquistar alguma medalha para seu país. Para isso é necessário anos de dedicação, treinos intensos e muito talento para ter uma chance de conquistar a tão sonhada medalha. Graças à combinação de todos esses fatores, existem competidores que conseguem atingir um maior sucesso que os demais, adquirindo um grande total de medalhas ao longo de vários anos.

A fim de tentar entender melhor quais são os atletas que possuem mais medalhas entre as Olimpíadas de 2000 até 2016, foi feita uma análise descritiva utilizando as variáveis: “Nomes” (qualitativa nominal), “Medalha” (qualitativa ordinal), significando respectivamente: o nome dos atletas participantes; medalha conquistada e seu tipo, sendo estes, ouro, prata e bronze. Além dessas variáveis, foi utilizado a figura a seguir:

Figura 5: Gráfico de colunas de medalhas por atleta



Como pode ser observado pela @figura-3, o atleta Michael Phelps é quem detem o maior número de medalhas olímpicas entre 2000 até 2016, totalizando 28 medalhas, dentre elas 23 sendo de ouro, representando 82,1%. Para fechar o top 3 tem a atleta Natalie Coughlin e Ryan Lochte, ambos empatados no total de medalhas com 12 cada.

Nesta análise também foi estudado se há ou não uma associação entre o tipo da medalha e o atleta que a conquistou e para isso, foi utilizado o Coeficiente de Contingência Modificado, $C^*=0,2719$, evidenciando uma associação moderada.

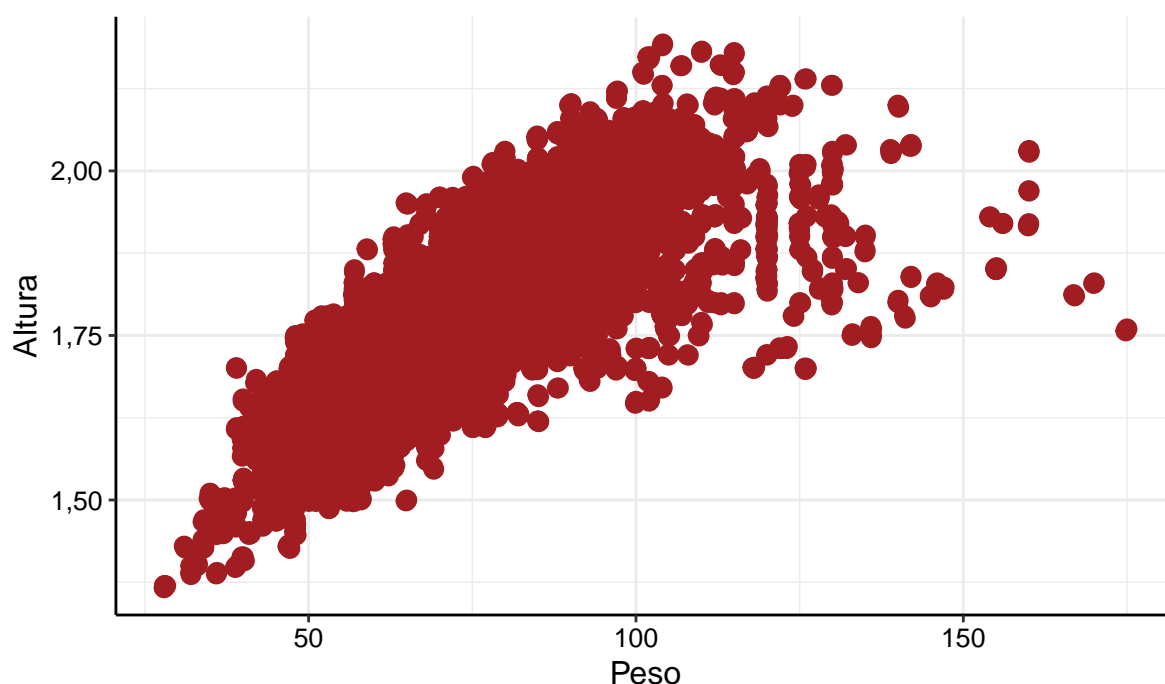
Graças ao Coeficiente estudado, podemos dizer que uma certa relação entre o tipo da medalha conquistada e atleta que a conquistou, mostrando que o talento individual, além de anos de treino e esforço, pode ser um fator determinante quanto o assunto se trata de uma conquista grande.

3.4 Análise 4

Algo bastante comum na fase de crescimento de todos os seres vivos é o aumento tanto de peso quanto de altura, podendo variar bastante os dois dependendo de cada indivíduo. Porém, isso nem sempre acontece, havendo uma grande variedade possível de alturas e pesos diferentes para todos.

A fim de tentar compreender melhor essa possível relação, foi feita uma análise descrita, estudando sua correlação, utilizando das variáveis: “Peso (kg)” (quantitativa discreta) e “Altura (m)” (quantitativa contínua), possuindo seus respectivos significados: peso em quilogramas de cada atleta; altura em metros de cada competidor. Para está compreender melhor está análise foi feita a seguinte figura:

Figura 6: Gráfico de dispersão do peso pela altura



Como pode ser observado pela @figura-4, uma grande parte da amostra analisada possui entre 1,50 e 2,00 metros, e de forma semelhante, os atletas se encontram na faixa de peso entre 50 a 100 quilos.

Para analisar sua correlação, foi utilizado o coeficiente de correlação de Spearman, o qual obteve um valor de $\rho = 0,841$, demonstrando que existe uma relação forte entre o peso e a altura das pessoas.

Essa grande correlação pode ser explicada pelo fato de quanto mais alto é o indivíduo, os músculos e a estrutura óssea dessa pessoa também tendem a ser maiores, gerando assim um maior valor de peso para tal.

4 Conclusões