



UNIVERSIDADE DE BRASÍLIA - UnB
INSTITUTO DE EXATAS - IE
DEPARTAMENTO DE ESTATÍSTICA - EST

Análise estatística sobre as quedas de meteoritos em diferentes partes do globo

Júlia Lima Nosralla
Leonardo Farias Saraiva

Docente: Guilherme Souza Rodrigues

13 de dezembro de 2023

Introdução

O seguinte relatório tem como objetivo apresentar uma análise dos dados contidos no banco “Meteorite Landings”, obtido na plataforma Kaggle, que contém 45.716 observações em 10 colunas, com informações acerca de registros da NASA sobre quedas de meteoritos ao redor da superfície terrestre, tais como ano e local de queda, tipo de meteorito, nome, peso, entre outras.

A fim de obter informações relevantes, o presente relatório visa realizar uma análise descritiva das diversas variáveis contidas nessa base de dados, bem como identificar possíveis relações entre elas.

O software utilizado para as análises foi o R, em sua versão 4.3.1.

Metodologia

Medidas

Este relatório se serve de algumas técnicas estatísticas, descritas a seguir de acordo com o que foi utilizado para a realização das análises.

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

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:

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

Quartis

Os quartis são separatrizes que dividem o conjunto de dados em quatro partes iguais. O primeiro quartil (ou inferior) é o conjunto que delimita os 25% menores valores, o segundo representa a mediana e é o valor que ocupa a posição central (ou seja, metade dos dados estão abaixo dela e a outra metade está acima) e o terceiro delimita os 75% 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.

Variância

A variância é uma medida que avalia o quanto os dados estão dispersos em relação à média. Esta medida é calculada em uma escala diferente daquela dos dados, uma vez que considera os quadrados dos valores para obter um valor modularizado.

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

Desvio padrão

O desvio padrão é a raiz quadrada da variância. Avalia o quanto os dados estão dispersos em relação à média. A vantagem do uso do desvio padrão está na sua readequação à mesma escala de medida dos dados.

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

Limpeza dos dados

Como é de praxe, é importante ressaltar a necessidade de limpeza e correção de valores no banco de dados. Em particular, merece destaque o caso do meteorito “Northwest Africa 7701”, observação 30.683 do banco de dados, para o qual verificou-se após uma exploração inicial da base de dados que apresentava o valor 2101 referente ao ano em que foi encontrado. Evidentemente isso configura um erro, o qual provavelmente se deu no momento da tabulação dos dados. Fazendo uma breve pesquisa, foi possível verificar que o meteorito na verdade foi encontrado no ano de 2010. Após essa constatação, pôde-se tomar por certa a suposição inicial de que houve um erro de digitação no referido valor, de modo que optou-se por corrigi-lo para o seu valor devido ao invés de descartá-lo, como é o procedimento padrão para dados inconsistentes.

Visualização dos resultados

Após uma primeira exploração, a devida limpeza e correção dos dados contidos no banco, foram então construídos gráficos, tabelas e mapas, a fim de facilitar a visualização dos resultados obtidos. Assim, será possível interpretar as características das observações de forma mais clara.

Análises

Número de meteoritos achados ou observados por ano

Esta seção tem como objetivo comparar a quantidade de meteoritos documentados a cada ano. Cada meteorito é classificado como “observado”, caso tenha sido visto no momento da queda, ou “achado”, caso tenha sido encontrado posteriormente, já na superfície terrestre.

O registro mais antigo de um meteorito observado é do ano de 860, enquanto os meteoritos mais recentes documentados no banco de dados são do ano de 2013.

Ademais, vale salientar que, antes de ano de 1972, o ano com mais meteoritos documentados foi 1969, com 70 observações. No entanto, após 1971, o número de meteoritos achados por ano aumentou significativamente, de modo que optou-se por representar em um gráfico de linhas o número de meteoritos registrados por ano em uma série a partir de 1972. Dessa forma, ao restringir a representação gráfica aos dados de anos mais recentes, será otimizada a visualização desses dados.

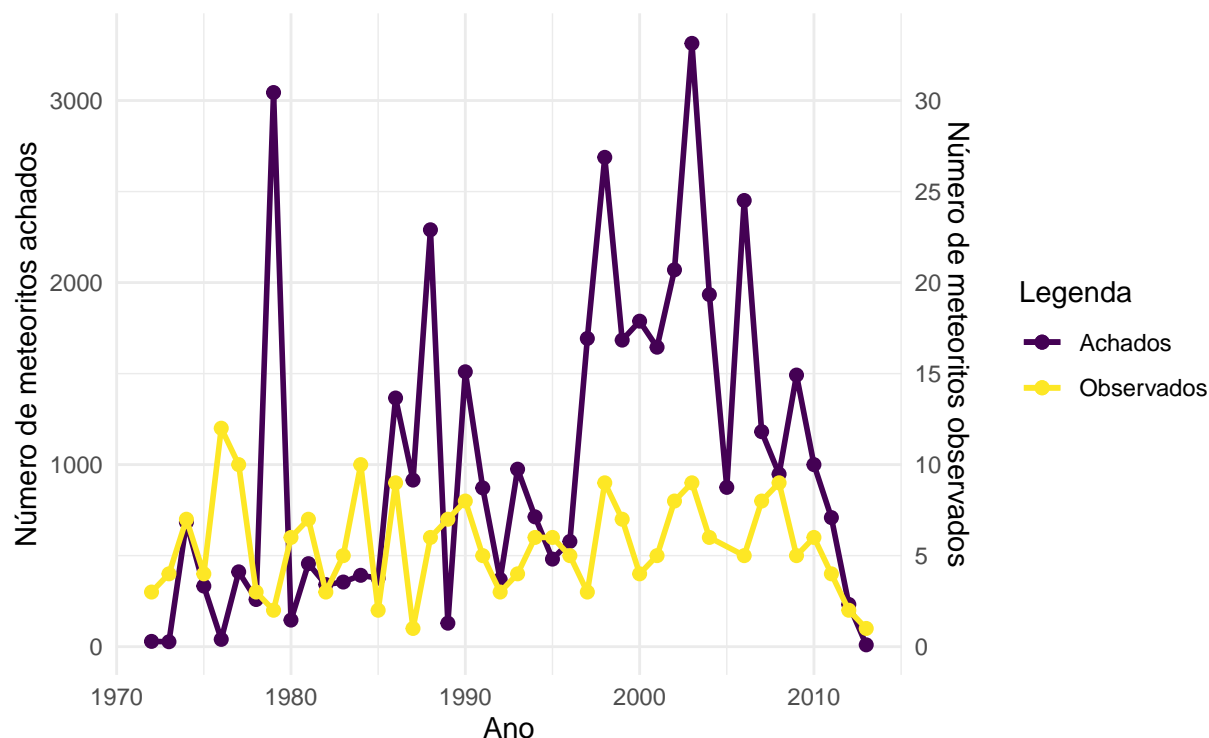


Figure 1: Gráfico de linhas do número de meteoritos registrados por ano

Com base no gráfico, observa-se que a quantidade de meteoritos achados foi de uma ordem significativamente maior do que o número de meteoritos observados. Ainda no intuito de melhorar a visualização, optou-se por representar graficamente esses valores em escalas distintas.

Durante o período analisado, a maior quantidade de meteoritos observados em um ano foi de 12 em 1976, enquanto a quantidade de meteoritos achados chegou a 3314 em 2003.

Por fim, é interessante notar que enquanto a série temporal para os meteoritos observados oscila dentro de uma faixa pequena, o gráfico de linhas para a quantidade de meteoritos achados apresenta alguns picos evidentes, que foram representados na tabela abaixo.

Table 1: Cinco anos com maior registro de meteoritos

Ano	Nº de meteoritos
2003	3323
1979	3046
1998	2697
2006	2456
1988	2296

Classificação dos meteoritos

Nesta seção, são identificadas as principais classificações dos meteoritos presentes no banco de dados. Os meteoritos costumam ser classificados de acordo com sua composição e estrutura, segundo uma miríade de propriedades e características. No entanto, não há padronização ou consenso científico quanto aos critérios ou à hierarquia dessas classificações, de modo que apenas nessa base de dados constam 466 tipos de meteoritos distintos, o que dificulta uma análise mais sistemática da distribuição dessas categorias. Dessa forma, foram considerados somente os dez tipos mais frequentes para a realização das análises.

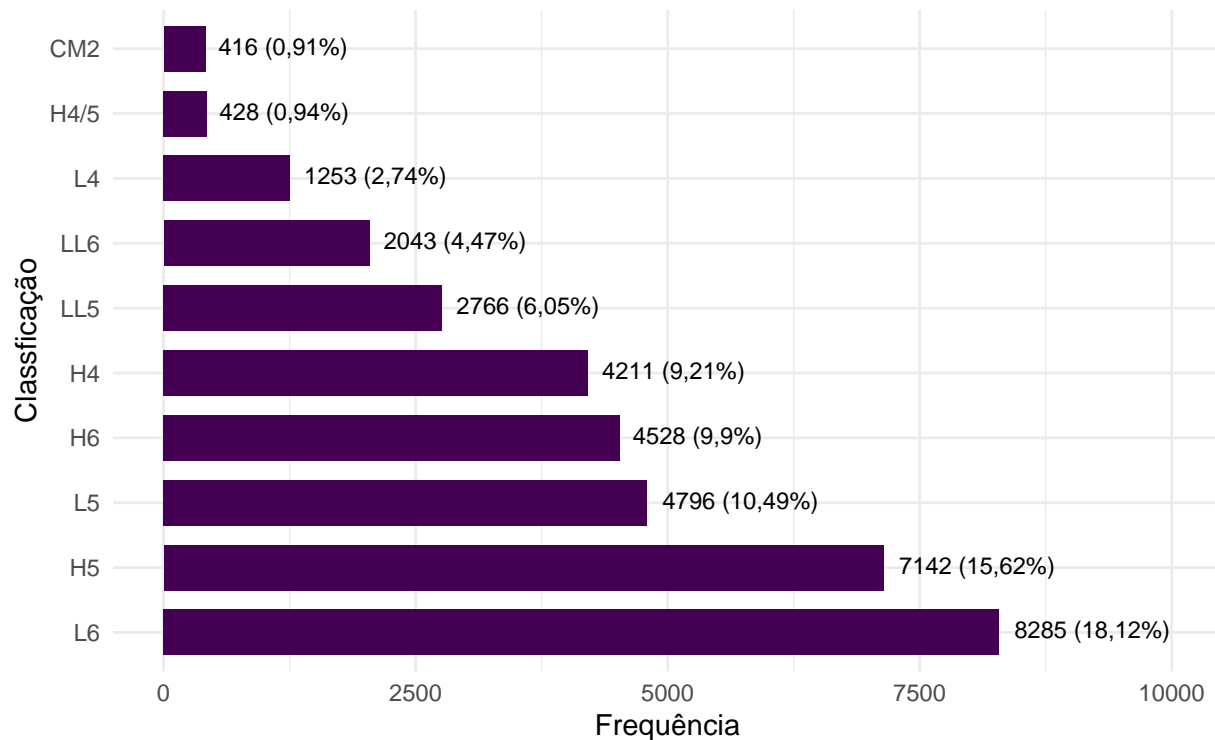


Figure 2: Gráfico de barras das 10 principais classificações de meteorito

Pelo gráfico acima, é possível perceber que as observações dos dez tipos de meteorito mais frequentes representam 78,45% do total de observações, o que é um valor absolutamente majoritário, justificando assim a opção por suprimir da análise os demais tipos, menos significativos. Além disso, nota-se que o tipo de meteorito com maior ocorrência é o L6, compondo sozinho 18,12% dos dados.

O menor valor apresentado no gráfico foi 0,91%, que representa a porcentagem de meteoritos do tipo CM2. Com isso, pode-se concluir que a porcentagem individual de todos os demais tipos de meteoritos não mostrados na figura é inferior a 1%.

Peso dos meteoritos segundo sua classificação

Esta seção tem como objetivo analisar o peso dos meteoritos de acordo com as dez classificações principais. O peso é uma variável quantitativa contínua que, nesta base de dados, possui valores que variam de 0,00001 a 4000 gramas.

Seguem o gráfico e a tabela com a distribuição das variáveis:

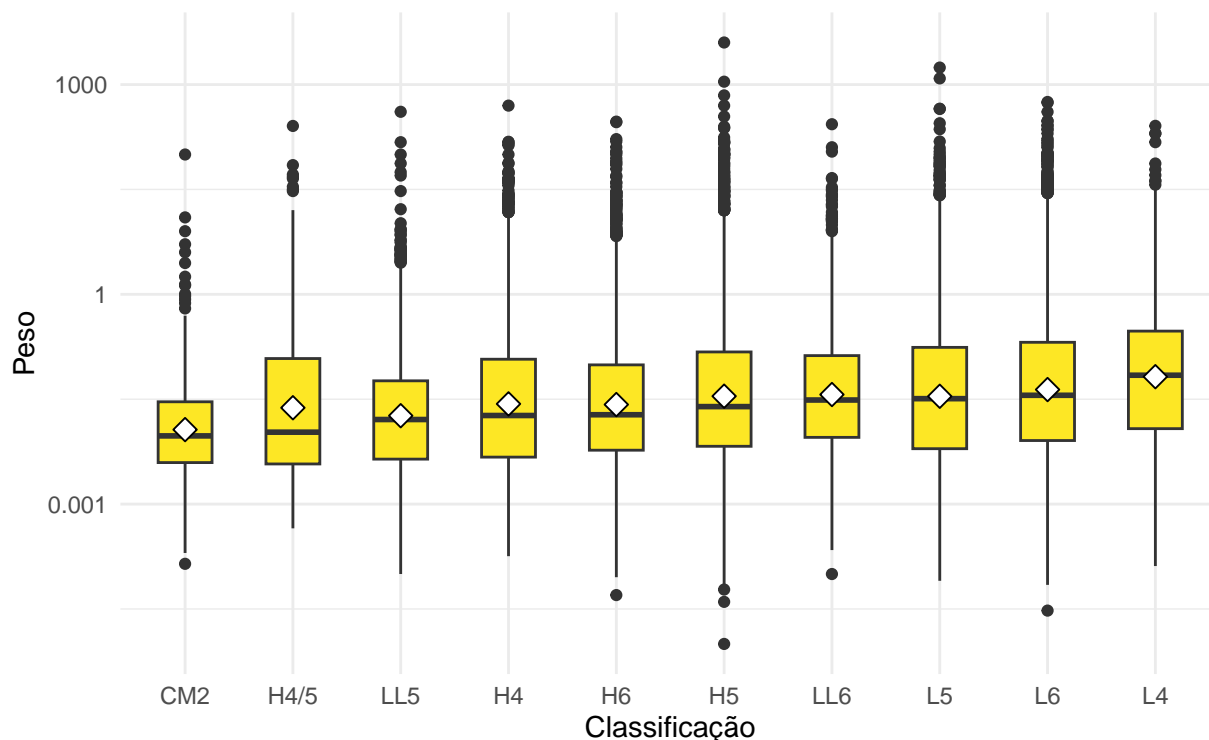


Figure 3: Boxplot do peso dos meteoritos nas 10 principais classificações

Table 2: Medidas resumo

	CM2	H4	H4/5	H5	H6	L4	L5	L6	LL5	LL6
Média	0,37	1,00	1,60	2,17	0,86	1,64	1,80	1,45	0,46	0,69
Desvio Padrão	4,98	9,86	13,70	51,17	8,52	12,43	33,13	14,01	8,75	7,48
Variância	24,77	97,14	187,80	2618,04	72,51	154,56	1097,65	196,29	76,59	55,91
Mínimo	0,00014	0,00018	0,00045	0,00001	0,00005	0,00013	0,00008	0,00003	0,00010	0,00010
1º Quartil	0,0039	0,0047	0,0037	0,0067	0,0059	0,0120	0,0062	0,0081	0,0044	0,0090
Mediana	0,01	0,02	0,01	0,02	0,02	0,07	0,03	0,04	0,02	0,03
3º Quartil	0,03	0,12	0,12	0,15	0,10	0,30	0,17	0,21	0,06	0,13
Máximo	100	500	256	4000	295	257	1750	564	408	271

A partir das informações obtidas, percebe-se que existem algumas variações nos pesos de cada categoria de meteorito. Primeiramente, é interessante destacar o comportamento dos meteoritos tipo H5: essa categoria apresentou a maior média e o maior desvio padrão, com valores de 2,17 e 51,17, respectivamente. Além disso, essa classe possui o menor e o maior peso observados na tabela, e é também a classe com mais valores discrepantes (*outliers*), tanto inferiores quanto superiores.

Em relação aos demais grupos, vale notar que tanto a menor média quanto o menor desvio padrão estão na classe CM2, com valores de 0,37 e 4,98, respectivamente. Pelo gráfico, é evidente que a categoria CM2

apresenta a menor caixa do boxplot. A maior mediana é de 0,07 na categoria L4, enquanto o menor valor de mediana, de 0,01, se aplica tanto aos meteoritos H4/5 quanto aos CM2.

Além disso, observa-se uma grande quantidade de valores discrepantes acima do limite superior em todos os grupos, enquanto poucos outliers inferiores aparecem apenas em CM2, H6, H5, LL6 e L6.

Localização da queda dos meteoritos

O mapa a seguir pode ser utilizado para melhor compreender a distribuição dos locais onde meteoritos foram encontrados ao redor do mundo.

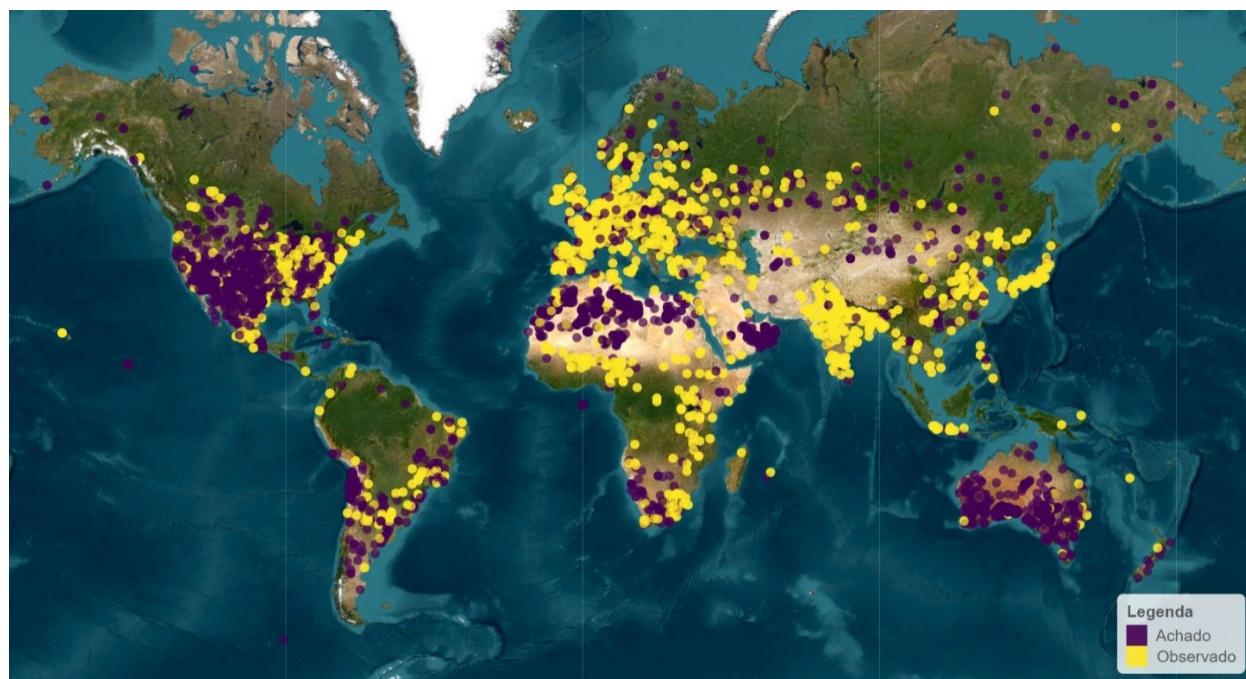


Figure 4: Mapa dos locais de queda de meteoritos

[Link de uma versão interativa](#)

Com base no mapa acima, é possível visualizar um comportamento interessante na distribuição dos locais de queda dos meteoritos documentados. Primeiramente, observa-se que há uma concentração de meteoritos achados na América do Norte, na Oceania, no norte da África e na península arábica, sobretudo na região onde fica o Omã. Já os meteoritos que foram observados em queda se encontram concentrados na Europa, na África Subsaariana e na região onde se encontra Índia, leste da China, Japão e países do sudeste asiático.

Pode-se ainda notar que existem regiões onde poucos ou nenhum meteorito foi encontrado. A partir disso, é possível levantar algumas hipóteses sobre esse fenômeno. As regiões norte do Canadá e da Rússia e a Groenlândia, por exemplo, apresentam climas frios extremos e são, portanto, pouco habitadas, de modo que meteoritos que caem nessas regiões dificilmente são encontrados.

Outras regiões em que se observa uma ausência do registro de meteoritos são o centro da América do Sul e da África. Uma das similaridades entre as duas regiões é que ambas apresentam florestas tropicais densas, a Floresta Amazônica e a Floresta do Congo. Assim, as características da vegetação, do clima e do relevo dessas regiões poderia explicar a dificuldade em descobrir meteoritos caídos nesses locais.

Além disso, em contrapartida, o mapa induz a especular de forma análoga que exista uma maior facilidade para encontrar meteoritos em regiões áridas ou desérticas, tais como o Atacama, o Saara, a península arábica, grande parte do território australiano, o sudoeste dos Estados Unidos e o norte do México.

Conclusão

Pode-se concluir, após a finalização deste trabalho, que a análise exploratória e a construção de gráficos, tabelas e mapas com o auxílio de ferramentas estatísticas propiciam uma melhor compreensão das hipóteses investigativas extraídas e/ou informações adicionais de bases de dados, ainda que estas sejam bastante limitadas. Nesse caso, o uso do software R e de diversos pacotes possibilitou contornar a escassez de variáveis, inconsistências de registro, diferenças de escala e a existência de categorias pouco significativas na base de dados para expor características interessantes acerca do fenômeno da queda de meteoritos que não eram óbvias ou não estavam evidentes até então.

Participações

Júlia Lima Nosralla (232014825)

Leonardo Farias Saraiva (232014834)