

Análise Estatística da Relação Massa–Raio de Exoplanetas

Enzo Rodrigues Teixeira de Andrade

Brasília – DF, 2025

Resumo

Este relatório apresenta uma análise estatística da relação entre massa e raio de exoplanetas utilizando Regressão Linear Múltipla (RLM) em escala logarítmica. Foram incluídas como variáveis explicativas a massa planetária, a temperatura de equilíbrio e a metalicidade estelar. O estudo avalia o ajuste do modelo, a significância estatística dos parâmetros e a validade dos pressupostos clássicos de regressão (normalidade, homocedasticidade, independência e multicolinearidade).

Sumário

1	Introdução	2
1.1	Hipótese	2
2	Metodologia	2
3	Análise Exploratória	2
3.1	Estatísticas Descritivas	2
3.2	Correlação e Distribuições	4
3.3	Transformação Logarítmica	4
4	Modelagem Estatística	4
4.1	Modelos Ajustados	4
4.2	Comparação OLS vs Erros Robustos (HC1)	5
5	Diagnóstico dos Resíduos	5
6	Validação Preditiva	5
7	Discussão	8
8	Conclusão	8

1 Introdução

A relação massa-raio é um dos pilares da caracterização de exoplanetas, pois permite inferir sua composição interna e estrutura atmosférica. Contudo, essa relação não é linear: planetas gasosos, por exemplo, apresentam variações do raio que saturam com o aumento da massa. A utilização da escala logarítmica lineariza parcialmente esse comportamento, permitindo aplicar modelos lineares para estudar a dependência física entre essas grandezas.

1.1 Hipótese

A hipótese testada é que o raio planetário (em log) depende linearmente da massa (em log), com correções associadas à temperatura de equilíbrio (`pl_eqt`) e à metalicidade da estrela hospedeira (`st_met`). Espera-se que planetas mais massivos e estrelas mais metálicas resultem em raios maiores.

2 Metodologia

- Base utilizada: `exoplanets.csv`, contendo 6028 registros e 682 colunas.
- Após limpeza e remoção de valores nulos: 5980 observações válidas.
- Para o modelo múltiplo (com `pl_eqt` e `st_met`): 4343 observações.
- Transformações aplicadas:

$$\log_{10}(\text{massa}), \quad \log_{10}(\text{raio})$$

- Modelos testados:

$$\text{Modelo 1: } \log R = \beta_0 + \beta_1 \log M + \varepsilon$$

$$\text{Modelo 2: } \log R = \beta_0 + \beta_1 \log M + \beta_2 \text{pl_eqt} + \beta_3 \text{st_met} + \varepsilon$$

3 Análise Exploratória

A análise exploratória teve como objetivo compreender o comportamento das variáveis antes da modelagem. Foram observadas medidas de tendência central, dispersão e possíveis relações entre os parâmetros físicos.

3.1 Estatísticas Descritivas

A amostra revela forte dispersão das massas e raios, variando de planetas menores que a Terra até gigantes gasosos. A metalicidade média das estrelas é próxima de zero, indicando valores semelhantes à composição solar.

Tabela 1: Resumo estatístico das variáveis principais.

Variável	N	Média	Desvio-padrão	Mínimo	Mediana	Máximo
Massa (M_{\oplus})	5735	385.22	1105.75	0.02	8.96	9344.16
Raio (R_{\oplus})	5735	5.73	5.31	0.31	2.80	77.34
Densidade	5676	4.93	35.33	0.01	2.58	2000.00
Temperatura (K)	4322	916.70	467.01	34.00	821.50	4050.00
Fluxo Estelar	4061	427.07	1344.54	0.00	99.24	44900.00
Metallicidade	5267	0.0145	0.188	-1.00	0.02	0.60

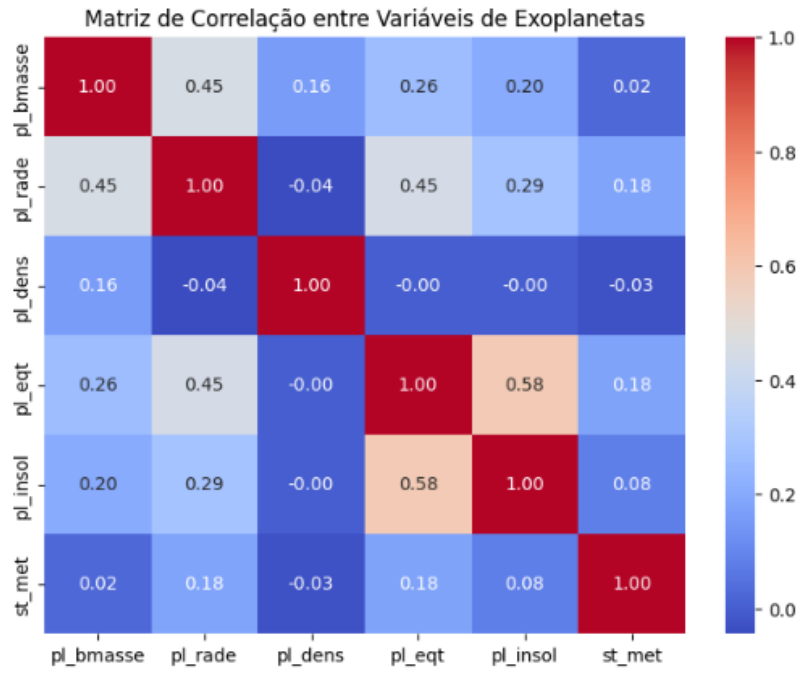


Figura 1: Matriz de correlação entre variáveis. Observa-se correlação positiva entre massa e raio.

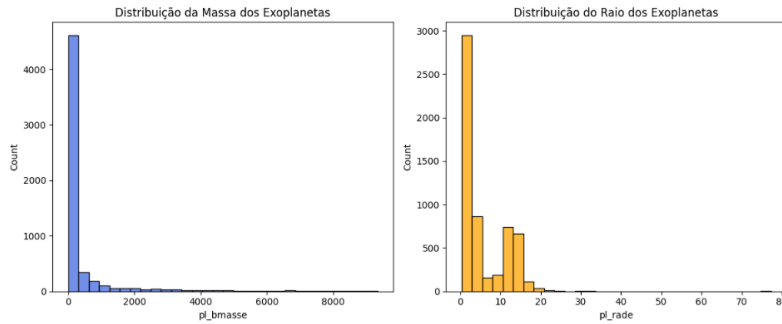


Figura 2: Distribuições de massa e raio (sem transformação logarítmica).

3.2 Correlação e Distribuições

3.3 Transformação Logarítmica

Devido à forte assimetria das distribuições, aplicou-se a transformação logarítmica para linearizar a relação e reduzir o impacto de valores extremos:

$$\log_massa = \log_{10}(pl_bmasse), \quad \log_raio = \log_{10}(pl_rade)$$

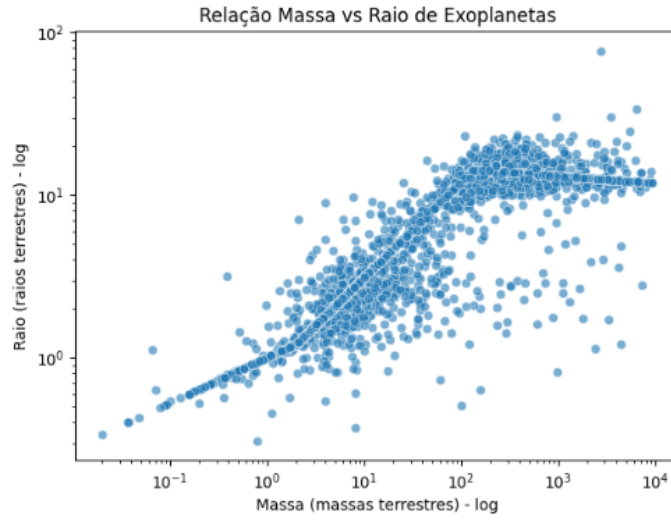


Figura 3: Relação entre massa e raio em escala logarítmica. O padrão linear justifica o uso de regressão.

4 Modelagem Estatística

4.1 Modelos Ajustados

O modelo simples apresentou:

$$\log(R) = 0.0503 + 0.3907 \log(M), \quad R^2 = 0.832$$

O modelo múltiplo apresentou:

$$\log(R) = 0.0379 + 0.3846 \log(M) + 1.987 \times 10^{-5} pl_eqt + 0.0591 st_met, \quad R^2 = 0.834$$

Todos os coeficientes mostraram significância estatística ($p < 0.05$).

Tabela 2: Coeficientes estimados por OLS e com erros robustos HC1.

Variável	OLS	Std.Err (OLS)	HC1	Std.Err (HC1)
Constante	0.0379	0.005	0.0379	0.005
$\log M$	0.3846	0.003	0.3846	0.005
pl_eqt	1.987e-05	5.23e-06	1.987e-05	5.75e-06
st_met	0.0591	0.014	0.0591	0.015

4.2 Comparação OLS vs Erros Robustos (HC1)

Interpretação física:

- $\beta_1 \approx 0.38$: relação sublinear — um aumento de $10\times$ na massa implica aumento de $\sim 2.4\times$ no raio;
- $\beta_2 > 0$: temperatura de equilíbrio eleva o raio, coerente com dilatação atmosférica;
- $\beta_3 > 0$: maior metalicidade estelar está associada a planetas ligeiramente maiores.

5 Diagnóstico dos Resíduos

Os testes indicaram:

- **Shapiro–Wilk:** $W = 0.845$, $p < 0.001$ — rejeita normalidade;
- **Breusch–Pagan:** $p \approx 2.36 \times 10^{-82}$ — evidencia heterocedasticidade;
- **Durbin–Watson:** ≈ 1.95 — ausência de autocorrelação;
- **VIF:** entre 1.08 e 1.14 — sem multicolinearidade relevante.

Esses resultados justificam o uso de erros robustos (HC1) para garantir estimativas consistentes.

6 Validação Preditiva

- $R^2_{teste} = 0.811$
- $\text{RMSE}_{teste} = 0.159$

Os resultados mostram estabilidade e baixa variabilidade entre as partições de treino e teste, reforçando a consistência do modelo.

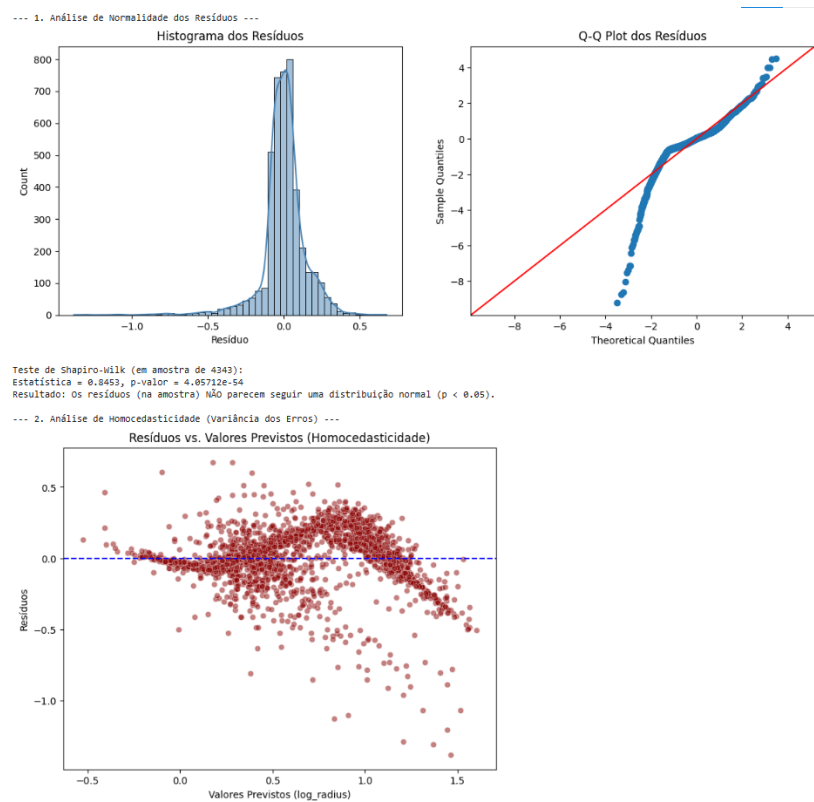


Figura 4: Análise de normalidade dos resíduos: histograma, Q-Q plot e resíduos vs valores previstos.

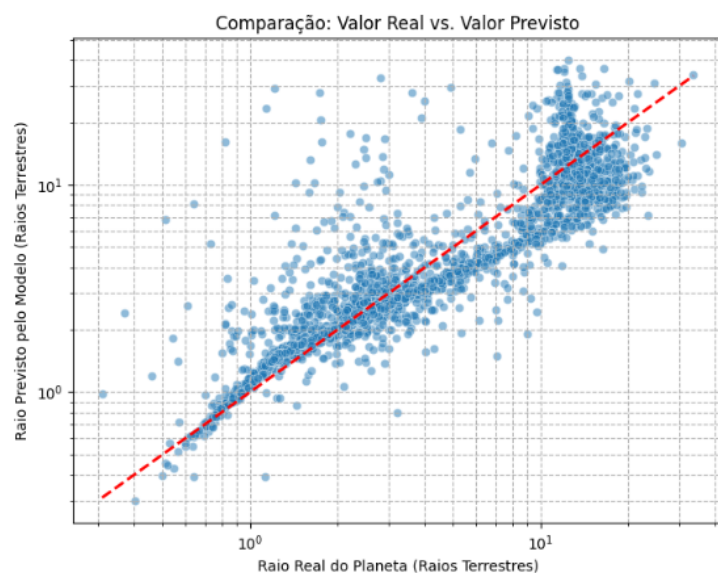


Figura 5: Comparação entre raios reais e previstos (escala linear, em raios terrestres).

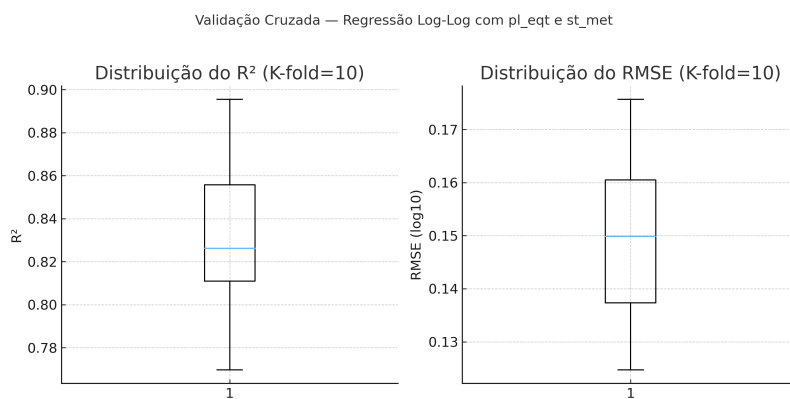


Figura 6: Validação cruzada ($K=10$): distribuição dos valores de R^2 e RMSE.

7 Discussão

Os coeficientes e ajustes obtidos são física e estatisticamente coerentes:

- O expoente ~ 0.38 está em linha com observações empíricas para amostras mistas de planetas rochosos e gasosos;
- A heterocedasticidade sugere diferentes regimes de formação e estrutura planetária;
- O modelo, embora simples, captura bem o comportamento geral e oferece boa capacidade preditiva ($R^2 \approx 0.83$).

8 Conclusão

A hipótese inicial foi confirmada: existe uma relação positiva e estatisticamente significativa entre massa, temperatura de equilíbrio, metalicidade e raio planetário. O modelo log-log ajustado apresentou $R_{aj}^2 \approx 0.834$ e RMSE=0.159, com coeficientes consistentes com teorias de estrutura planetária.

As análises de resíduos indicam leve não-normalidade e heterocedasticidade, justificando o uso de erros robustos. Em síntese, o modelo proposto é estatisticamente sólido e fisicamente interpretável, fornecendo subsídios para futuras investigações sobre regimes composicionais de exoplanetas.