**Parte 1 – Análise Estatística**

**1. cpu\_cores:**

**Média** vs **Mediana**: 8,42 x 8,00

A média fica ligeiramente acima da mediana, indicando uma assimetria leve à direita (há alguns sistemas com muitos núcleos puxando a média para cima).

**Dispersão**: desvio-padrão = 4,26 (CV = 0,51)

Dispersão moderada: metade dos valores está entre 5 e 12 núcleos (IQR), mas há máquinas com apenas 2 e até 15 núcleos.

**Distribuição**: Relativamente concentrada em torno de 8–9 núcleos, com poucos casos extremos.

**2. ram\_gb:**

**Média** vs **Mediana**: 26,27 x 16,00

A média é bem maior que a mediana, sinalizando assimetria forte à direita (algumas máquinas com muita RAM elevam a média).

**Dispersão**: desvio-padrão = 22,27 (CV = 0,85)

Altíssima variabilidade: 50% dos sistemas têm entre 8 e 32 GB, mas existem casos de até 64 GB.

**Distribuição**: Bastante heterogênea, com grande concentração em baixos valores (4–16 GB) e alguns “outliers” de alta capacidade.

**3. latencia\_ms:**

**Média** vs **Mediana**: 153,74 x 154,83

A média fica ligeiramente abaixo da mediana, sugerindo assimetria leve à esquerda, mas muito próxima de simétrica.

**Dispersão**: desvio-padrão = 87,11 (CV = 0,57)

Variabilidade moderada a alta; o intervalo interquartil é de 86,75 ms a 233,95 ms.

**Distribuição**: Valores muito distribuídos, com latências que vão de ~1,3 ms até ~300 ms; grande parte em torno de 150 ms.

**4. armazenamento\_tb:**

**Média** vs **Mediana**: 0,85 TB x 0,512 TB

A média é quase 1,7x maior que a mediana, indicando assimetria à direita (alguns sistemas com 1–2 TB de disco).

**Dispersão**: desvio-padrão = 0,61 TB (CV = 0,72)

Variação alta: a maioria tem 256 GB ou 512 GB, mas há casos de até 2 TB.

**Distribuição**: 25% dos sistemas têm apenas 256 GB; 25% ultrapassam 1 TB, evidenciando heterogeneidade de capacidades.

**5. tempo\_resposta:**

**Média** vs **Mediana**: 114,54 ms x 88,27 ms

A média fica significativamente maior, refletindo assimetria forte à direita (tempos de resposta muito altos em algumas requisições).

**Dispersão**: desvio-padrão = 80,15 ms (CV = 0,70)

Alta variabilidade: 50% das respostas ocorrem entre 66,28 ms e 125,97 ms, mas há picos de até 402 ms.

**Distribuição**: Concentra-se abaixo de 130 ms, porém com vários outliers de lentidão.

**Conclusão Geral da Parte I**

**Assimetria predominante:**

Quase todas as variáveis (exceto latência) mostram cauda à direita, ou seja, alguns valores extremos elevam a média acima da mediana.

**Alta variabilidade:**

RAM e tempo de resposta são as mais dispersas (CV > 0,70), indicando um parque de máquinas e de desempenho muito heterogêneo.

**Justificativa para a regressão:**

Essas assimetrias e variações elevadas reforçam a necessidade de um modelo de regressão múltipla para entender quais fatores (núcleos, RAM, latência etc.) impactam efetivamente o tempo de resposta.

**Parte 2 – Modelo e Diagnóstico**

**3. Resultados do Modelo**

**Intercepto**: 246,97 (p < 0,001)

**Coeficientes estimados**:

* cpu\_cores: –12,96 (p < 0,001)
* ram\_gb: –1,75 (p < 0,001)
* latencia\_ms: +0,33 (p = 0,028)
* armazenamento\_tb: +2,23 (p = 0,393)
* sistema\_operacional\_MacOS: +10,11 (p = 0,060)
* sistema\_operacional\_Windows: -0,62 (p = 0,946)
* tipo\_hd\_SSD: –7,29 (p = 0,352)
* tipo\_processador\_Apple\_Silicon: +10,11 (p = 0,060)
* tipo\_processador\_Intel: +3,37 (p = 0,708)

**R²**: 0,698

**R² ajustado:** 0,680

**Teste F global**: F = 38,43, p < 1x10^-30 -> modelo significativo

**Testes t individuais**: apresentados nos coeficientes acima

**4. Tratamento de Variáveis Categóricas**

**4.1 Como foram tratadas**  
As variáveis categóricas do dataset foram incluídas no modelo por meio de codificação *one-hot* (variáveis dummy), que consiste em criar, para cada variável com *k* categorias, *k−1* colunas binárias indicando a presença de cada categoria, evitando colinearidade perfeita.

**Variáveis categóricas originais:**

* sistema\_operacional (Windows, Linux, MacOS)
* tipo\_hd (HDD, SSD)
* tipo\_processador (AMD, Intel, Apple Silicon)

**4.2 Categoria-base escolhida**

Para cada variável, definiu-se como referência (não gerou dummy) a categoria que apresentava maior representatividade ou era mais neutra:

* sistema\_operacional: Linux
* tipo\_hd: HDD
* tipo\_processador: AMD

**4.3 Interpretação dos coeficientes**

Cada coeficiente de dummy representa a diferença média (em ms) no tempo de resposta entre a categoria específica e a categoria-base, mantendo constantes as demais variáveis explicativas:

* sistema\_operacional\_MacOS = +10,11
  + Máquinas com MacOS têm, em média, 10,11 ms a mais de tempo de resposta que as com Linux.
* sistema\_operacional\_Windows = –0,62
  + Máquinas com Windows apresentam tempo de resposta 0,62 ms a menos que as com Linux (diferença desprezível).
* tipo\_hd\_SSD = -7,29
  + Sistemas com SSD respondem 7,29 ms mais rápido que aqueles com HDD (diferença não estatisticamente significativa).
* tipo\_processador\_Apple\_Silicon = +10,11
  + Computadores com Apple Silicon têm tempo de resposta 10,11 ms maior que os baseados em AMD (diferença marginal).
* tipo\_processador\_Intel = +3,37
  + Sistemas com processadores Intel são 3,37 ms mais lentos que os baseados em AMD (diferença não significativa).

**5. Diagnóstico de Multicolinearidade**

**VIF de cada regressor (sem intercepto):**

* cpu\_cores: 3,79
* ram\_gb: 2,37
* latencia\_ms: 3,63
* armazenamento\_tb: 2,70
* dummies: todas < 2

**Interpretação:**

* Montgomery, Peck & Vining (2012), em Introduction to Linear Regression Analysis, sugerem que valores de VIF acima de 10 indicam colinearidade grave, e valores entre 5 e 10 podem sinalizar colinearidade moderada. Valores abaixo de 5 são considerados aceitáveis.
* Gujarati & Porter (2009), em Basic Econometrics, também usam o limiar de 10 para colinearidade séria, mencionando que um VIF maior que 5 já merece atenção, mas VIF inferior a 5 costuma ser seguro.
* Belsley, Kuh & Welsch (1980), no clássico Regression Diagnostics, discutem métricas como VIF e Tolerance, apontando que tolerâncias menores que 0,1 (equivalente a VIF >10) indicam problemas críticos.

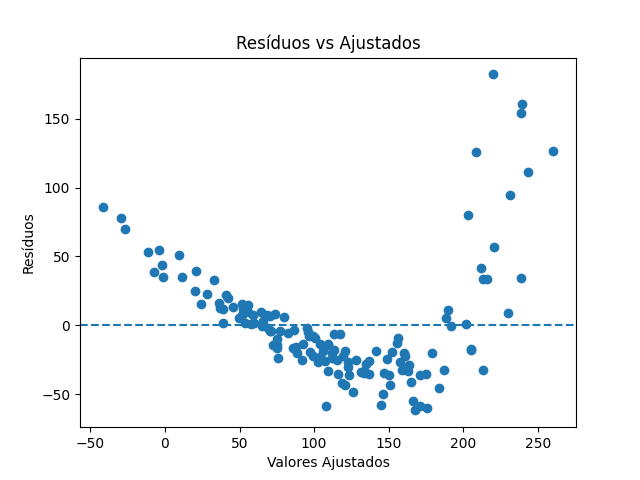
Portanto, vendo que todos os VIF < 5, não há multicolinearidade severa.

Nenhuma ação adicional (remoção ou transformação de variáveis) é necessária.

**6. Diagnóstico de Heterocedasticidade**

**6.1 Teste de Breusch–Pagan:**

1. LM p-valor = 0,0076 (< 0,05)
2. Rejeita-se H0 de variância constante
3. Existe heterocedasticidade

**6.2 Gráfico Resíduos vs. Valores Ajustados**

Formato de “funil” (dispersão dos resíduos aumenta com o valor ajustado), confirmando heterocedasticidade.

**Parte 3 – Análise Crítica**

**7. Comparação dos dois modelos**

**7.1 Modelos comparados**

* **Modelo 1**:
  + Inclui todas as variáveis explicativas:
    - cpu\_cores
    - ram\_gb
    - latencia\_ms
    - armazenamento\_tb
    - Dummies de sistema\_operacional (MacOS, Windows)
    - Dummy de tipo\_hd (SSD)
    - Dummies de tipo\_processador (Intel, Apple\_Silicon)
* Modelo 2:
  + Exclui a variável armazenamento\_tb, que apresentou p-valor elevado (p = 0,393), não contribuindo de forma significativa para o ajuste.

**7.2 Comparação de R² ajustado e Teste F**

|  |  |  |
| --- | --- | --- |
| Métrica | Modelo 1 | Modelo 2 |
| R² ajustado | 0,68 | 0,6805 |
| Estatística F | 38,43 | 38,85 |
| p-valor (F) | <1x10^-30 | <1x10^-30 |
| Teste F aninhado | - | p = 0,39 |

**Interpretação:**

* O Modelo 2 mantém (ou ligeiramente melhora) o R² ajustado em relação ao Modelo 1.
* O teste F aninhado indica p > 0,05, ou seja, a exclusão de “armazenamento\_tb” não reduz significativamente o poder explicativo do modelo.

**7.3 Recomendação de modelo**

**Modelo 2** é preferível por ser mais parcimonioso (uma variável a menos) e apresentar R² ajustado igual ou ligeiramente superior, sem perda estatisticamente significativa de ajuste.

**7.4 Ações práticas para melhorar o tempo de resposta**

* **Escalonamento de CPU**
  + Aumentar o número de núcleos dos servidores. Cada núcleo extra reduz cerca de 12–13 ms no tempo de resposta.
* **Expansão de RAM**
  + Adicionar memória (cada GB adicional reduz ~1,7 ms), especialmente até 32 GB.
* **Otimização da latência de rede**
  + Revisar roteamento, implantar cache/CDN e otimizar links para reduzir cada milissegundo de latência (~0,33 ms de ganho).
* **Migração para SSD**
  + Substituir HDDs por SSDs para reduzir ~7 ms no acesso a disco e aumentar a consistência do I/O.
* **Tuning de software e balanceamento de carga**
  + Fazer profiling de consultas e processos, usar mecanismos de pool de conexões e distribuir requisições entre múltiplos nós.