

EC03 — Análise de Desempenho (DE) na Classe Rosenbrock

Relatório de Estudo de Caso (EC03)

04/11/2025

Bernardo Bacha

bernardobr@ufmg.br

Gustavo Reis

augustogustavo94@gmail.com

Marília Melo

mariliamacedomelo@gmail.com

1 Formulação das Hipóteses de Teste

Para analisar os resultados, definimos a variável de interesse como a diferença pareada (Δ) para cada dimensão i :

$$\Delta_i = \text{Fbest}_{\text{cfg2},i} - \text{Fbest}_{\text{cfg1},i}$$

Se $\Delta_i > 0$, a Configuração 1 (cfg1) foi melhor (encontrou um valor menor). Se $\Delta_i < 0$, a Configuração 2 (cfg2) foi melhor (encontrou um valor menor).

As hipóteses estatísticas bicaudais sobre a mediana (ou média) das diferenças são: $H_0 : \mu_{\Delta} = 0$ (Hipótese Nula: Não há diferença significativa no desempenho médio/mediano entre as duas configurações). $H_1 : \mu_{\Delta} \neq 0$ (Hipótese Alternativa: Existe uma diferença significativa no desempenho).

2 Cálculo dos Tamanhos Amostrais

O planejamento amostral foi definido para detectar uma diferença considerada “média”.

- **Parâmetros (do Enunciado):**

- Efeito mínimo detectável (padronizado): $d^* = 0.5$
- Nível de significância: $\alpha = 0.05$
- Poder estatístico desejado: $1 - \beta = 0.8$

- **Cálculo (para Teste-t pareado):**

- O cálculo realizado (usando `pwr.t.test`) resultou em $n = 33.37$.

- **Tamanho Amostral Definido:**

- O valor foi arredondado para cima, exigindo 34 pares (ou seja, 34 dimensões de teste).

3 Coleta e Tabulação dos Dados

Os dados foram lidos do arquivo `resultados_EC03.csv`. A análise utilizou $n = 34$ pares, conforme planejado.

Table 1: Resumo Descritivo das Métricas

Variavel	Media	Mediana	DP
Fbest_cfg1	1398355	1085956	1379838
Fbest_cfg2	4114822	3164911	3760523
Delta (cfg2 - cfg1)	2716467	1871029	2466475

Table 2: Placar de Vitórias (por Dimensão)

N_pares	Vitorias_cfg2	Vitorias_cfg1	Empates
34	0	34	0

*Observação: A Configuração 1 obteve um **Fbest** menor (melhor) em todas as 34 dimensões testadas.*

3.1 Análise Gráfica Exploratória

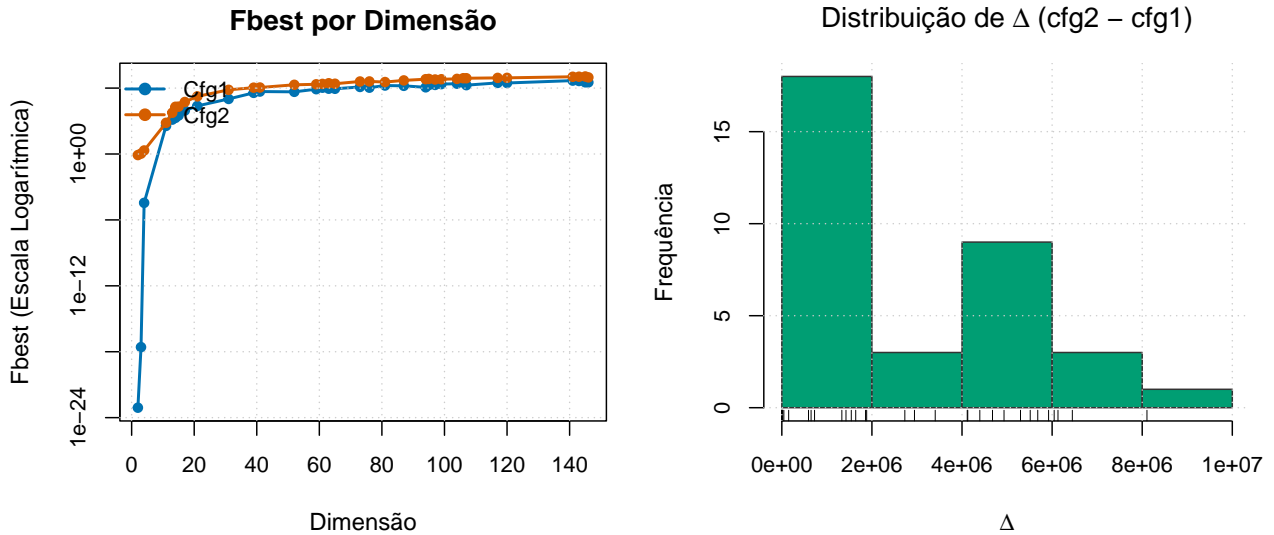


Figure 1: Gráficos exploratórios: (Esquerda) Fbest por dimensão (escala log-y); (Direita) Histograma das diferenças.

4 Teste das Hipóteses

O procedimento de teste seguiu a lógica de verificar a premissa de normalidade antes de aplicar o teste principal.

- **Checagem da Premissa (Normalidade):**
 - O Teste de Shapiro-Wilk foi aplicado sobre o vetor de diferenças (Δ).
 - O resultado foi $p = 2.86e-03$.
- **Decisão sobre o Teste:**
 - Como $p < 0.05$, a hipótese de normalidade dos dados foi **rejeitada**.
 - Portanto, o teste não-paramétrico (que não exige normalidade) foi utilizado: **Teste de Wilcoxon pareado**.
- **Resultado do Teste Principal (Wilcoxon):**
 - O teste retornou um $p\text{-valor} = 3.82e-07$.
- **Decisão Estatística:**
 - Como $p < 0.05$ (o nível de significância α), **rejeitamos a hipótese nula (H_0)**.

5 Estimação da Magnitude da Diferença

Como o teste de Wilcoxon foi usado, estimamos a mediana das diferenças (Δ):

- **Estimativa da Mediana:** $1.87e+06$
- **Intervalo de Confiança (IC 95% para Δ):** $[1.66e+06, 3.53e+06]$
- **Tamanho de Efeito (r aprox.):** 0.871 (considerado um efeito de grande magnitude)

O IC 95% não contém o zero, reforçando a rejeição de H_0 .

6 Verificação das Premissas dos Testes

Conforme detalhado no item 4, a premissa de normalidade das diferenças, necessária para o Teste-t pareado, foi verificada e **rejeitada** (Shapiro-Wilk $p = 2.86e-03$). A análise foi ajustada corretamente ao utilizar o Teste de Wilcoxon pareado, que é robusto a essa violação.

O gráfico QQ-plot abaixo confirma visualmente a falta de normalidade, com os pontos se desviando significativamente da linha teórica.

7 Derivação de Conclusões

Com base nos resultados, podemos responder às perguntas do enunciado:

- **Há diferença no desempenho?**
 - **Sim.** O p -valor ($3.82e-07$) foi extremamente baixo, indicando que a diferença observada não se deve ao acaso.
- **Qual a melhor configuração e qual a magnitude?**

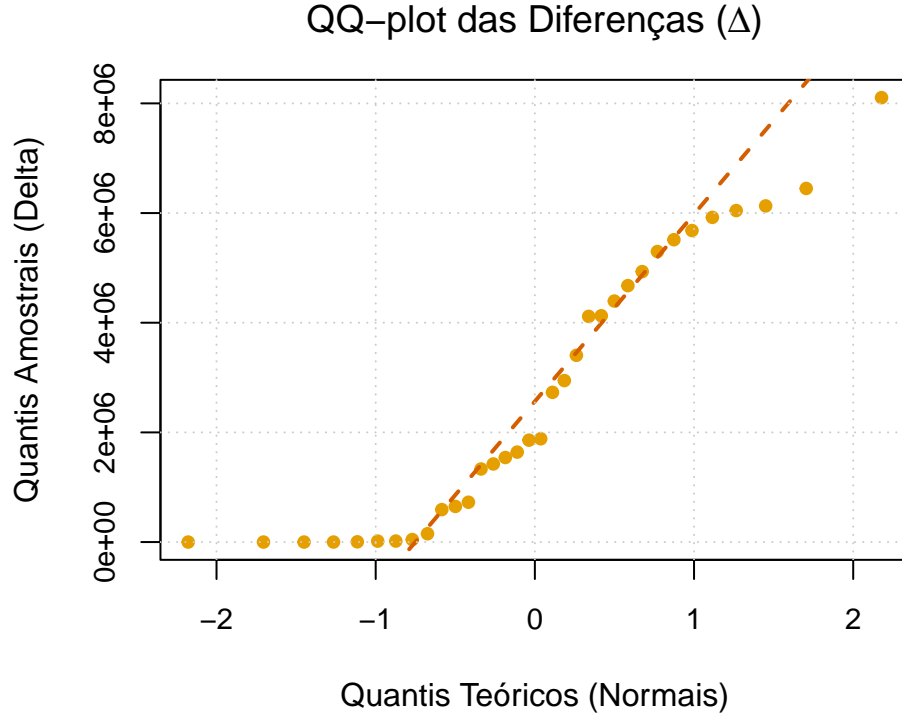


Figure 2: Gráfico QQ-plot para verificar a normalidade das diferenças (Delta).

- O IC 95% para a diferença ($\Delta = \text{cfg2} - \text{cfg1}$) é $[1.66\text{e}+06, 3.53\text{e}+06]$.
- Como o intervalo é inteiramente positivo, temos alta confiança de que $Fbest_{\text{cfg2}} > Fbest_{\text{cfg1}}$.
- Lembrando que **menor = melhor**, a **Configuração 1 (cfg1)** apresentou um **desempenho estatisticamente superior**.
- Qual configuração deve ser recomendada?
 - Para a classe de problemas Rosenbrock, a **Configuração 1** deve ser recomendada.

8 Discussão sobre Possíveis Limitações do Estudo

O estudo identifica as seguintes limitações e sugestões de melhoria:

- **Limitações:**
 - **Estocasticidade:** O DE é um algoritmo estocástico (aleatório). Este estudo usou apenas uma execução (réplica) por dimensão. O desempenho observado pode ter sido influenciado pelo ruído.
 - **Generalização:** Os resultados são válidos apenas para a classe de funções Rosenbrock.
 - **Orçamento Fixo:** O estudo usou um orçamento de avaliação fixo.
- **Sugestões de Melhoria (Extensões):**
 - Incluir múltiplas réplicas por dimensão para obter uma estimativa mais robusta e reduzir o ruído.

- Testar as configurações em outras classes de funções de benchmark.
- Analisar a sensibilidade a outros hiperparâmetros (como F , CR e tamanho da população).

9 Papéis desempenhados

Marília Melo: Metodologia, Pesquisa, Design da apresentação de dados, Desenvolvimento, Redação original; **Gustavo Reis:** Metodologia, Supervisão, Validação de dados e experimentos, Redação - revisão; **Bernardo Bacha:** Análise de dados, Metodologia, implementação e teste de software, Redação - revisão;

10 Referências

BESSANI, M. **Análise de Variância (ANOVA)**. Belo Horizonte: UFMG, 2025. Material da disciplina de Planejamento e Análise de Experimentos.

BESSANI, M. **Comparações Simples e Pareadas**. Belo Horizonte: UFMG, 2025. Material da disciplina de Planejamento e Análise de Experimentos.

BESSANI, M. **Inferência Estatística**. Belo Horizonte: UFMG, 2025. Material da disciplina de Planejamento e Análise de Experimentos.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics and Probability for Engineers**. 5. ed. Hoboken: John Wiley & Sons, 2010.

LEE, S.; LEE, D. K. What is the proper way to apply the multiple comparison test? **Korean Journal of Anesthesiology**, v. 71, n. 5, p. 353-360, 2018.