

Trabalho Final - Avaliação de Algoritmos em um Problema de Otimização de Roteiros de Viagem

Equipe 04

03 de Julho de 2017

Coordenador: Gustavo Vieira
Relator: Bernardo Marques
Verificador: Danny Tonidandel
Monitor: Alessandro Cardoso

Sumário

...

1. Descrição do Problema

As ferramentas atuais para planejamento de viagens permitem que usuários busquem passagens e hotéis e construam seu próprio itinerário, mas não oferecem suporte para viagens com flexibilidade nas datas ou na ordenação de destinos. Para solucionar esse problema, foi proposto um novo sistema que busca as melhores combinações de passagens e hotéis para uma viagem, minimizando o custo e tempo de transporte. O sistema busca otimizar roteiros com datas flexíveis e não requer uma ordem fixa dos destinos, de forma que considere todas as possíveis configurações que atendem as especificações do usuário. Dessa forma, oferece boas opções de itinerários para viagens com esforço do usuário muito menor, representando uma melhoria significativa na experiência de planejamento de viagens.

O conjunto de restrições do problema considera datas inicial e final, datas específicas para cada destino, ordenação dos destinos e características das acomodações e transportes. Pode ser considerado uma aplicação multiobjetivo do problema do caixeiro viajante com janelas de tempo e dependência temporal.

Para realizar a otimização de roteiros, foi utilizado um algoritmo de colônia de formigas [?]. Uma melhoria para o algoritmo utilizando busca local foi proposta. O objetivo do trabalho é comparar os algoritmos e determinar se a melhoria proposta é melhor para solucionar o problema, avaliando tanto o tempo gasto na otimização e a qualidade do conjunto de soluções encontrado. Assim, as perguntas a serem respondidas são:

- TODO
- TODO

No entanto, o tamanho do roteiro a ser avaliado também influencia esses resultados. Assim, é necessário considerar o tamanho da instância como uma covariável do problema. Para lidar com isso, serão utilizadas técnicas de blocagem, e também exploradas técnicas estatísticas de análise de covariância.

Os parâmetros experimentais desejados são:

- Nível de significância: $\alpha = 0.05$;
- Tamanho de efeito de interesse prático: $d_{time}^* = d_{quality}^* = 0.5$;
- Potência desejada: $(1 - \beta) = \pi \geq 0.8$.

TODO: explicar melhor o problema

2. Planejamento Experimental

Os dados experimentais do problema foram obtidos através de simulações. Foram geradas instâncias do problema de diferentes tamanhos. Para cada uma delas, cada algoritmo foi aplicado 8 vezes, e os resultados registrados. A ordem das instâncias geradas e da aplicação dos algoritmos foi aleatorizada de maneira a bloquear possíveis efeitos de ordem.

Nos experimentos, foi medido o tempo gasto na execução de cada algoritmo. Para minimizar a influência de outros processos do sistema, a execução foi limitada a um único núcleo do processador, com prioridade de execução alta.

Cada algoritmo de otimização produz como resultado um conjunto de soluções. A qualidade desse conjunto é medida através do Hipervolume [?] do conjunto na fronteira de Pareto, considerando como referência o ponto da solução Nadir na fronteira de Pareto (isto é, a pior solução teórica). Para cada instância avaliada, um novo ponto de referência é calculado através de uma heurística.

O método de Hipervolume é bastante sensível ao ponto de referência definido [?]. Sendo assim, a qualidade medida entre diferentes instâncias não tem significado quando comparadas entre si. No entanto, a métrica é eficiente para comparar a qualidade de diferentes conjuntos de solução encontrados no mesmo ponto. Sendo assim, é necessário avaliar a qualidade relativa das soluções em cada instância, e não o seu valor absoluto.

2.1 Definição do Tamanho Amostral

O tamanho amostral do problema foi definido a partir de requisitos técnicos do projeto. Era desejável avaliar o efeito dos algoritmos para problemas de diferentes tamanhos, partindo do caso mais simples até problemas de tamanho real. No entanto, limitações no tempo de execução do trabalho impediram coleta de dados maiores.

As instâncias geradas utilizaram então combinações de 1 a 4 destinos e 1 a 6 opções de transportes e acomodação. Para essas características, foi gerado um conjunto de dados com 108 instâncias.

A função `power.t.test` foi utilizada para calcular o tamanho necessário para os requisitos do projeto, considerando comparações pareadas que podem ser efetuadas após o teste ANOVA. O resultado indica que, para a potência e tamanho de efeito desejados, 108 observações são mais que suficientes.

2.2 Análise de Variância com Blocagem

Para a avaliar existência de diferenças significativas entre os dois parâmetros, será utilizado o teste estatístico ANOVA.

Essa técnica analisa médias e variâncias de observações de diferentes grupos para verificar se existe diferença estatística significativa entre as médias desses grupos. Assim, generaliza o teste t para mais de dois grupos, permitindo que sejam comparados simultaneamente [Campelo, 2015], [Montgomery and Runger, 2011]. Em outras palavras, a análise de variância é utilizada quando se quer decidir se os níveis apresentam médias diferentes em relação a uma média global μ . As diferenças entre as médias dos níveis e a média global é $\tau_i \forall i$.

Vale ressaltar que a técnica é restrita apenas à indicação de existência ou não de diferenças entre os níveis avaliados, sem indicar quais níveis seriam diferentes. Além disso, quando a análise de variância tem como resultado um indicativo de refutação da hipótese nula (1) é que podem ser evidenciados os indícios de diferenças entre os níveis.

$$\begin{cases} H_0 : & \tau_i = 0, \quad \forall i \\ H_1 : & \exists \tau_i \neq 0 \end{cases} \quad (1)$$

Esta etapa foi considerada como uma primeira verificação ao questionamento deste estudo por apresentar uma abordagem que pode minimizar o custo de coleta de amostras menor do que o teste de comparações múltiplas discutido na Seção 2.2. Portanto, caso não seja detectada alguma diferenças entre os níveis, as comparações múltiplas são dispensáveis, o que reduz o custo do experimento.

2.3 Análise de Covariância (ANCOVA)

A análise de covariância é uma técnica – assim como a blocagem ou pareamento (em testes-t) – bastante útil para a melhora da precisão de um experimento [Montgomery, 1984]. Em diversos aspectos é similar à análise de variância (ANOVA), porém permite ter controle sobre a influência do covariante nas variáveis dependentes.

A covariável complementa o controle local e, obviamente, necessita estar correlacionada com a variável de resposta para que se possa fazer uso de tal análise. E quando a análise de variância é realizada com uma ou mais covariáveis, é usual chamar a análise de ANCOVA.

Na ANCOVA a variável dependente é contínua (e.g. tempo, velocidade, etc.), enquanto que a variável independente é normalmente categórica (e.g. “masculino/feminino”, “fumante/não-fumante” etc.). A análise de variância, por sua vez, poderá converter-se na ANCOVA quando for adicionado um covariante, que consiste em outra variável, que pode ser tanto categórica quanto contínua.

Existem aliás, duas razões principais para sair-se da ANOVA e passar a ser considerada a análise de covariância, i.e., razões técnicas para adicionar um covariante, quais sejam:

- Reduzir fatores de variabilidade inter-grupos, ou alcançar um nível maior de entendimento a partir da variância desconhecida;
- Isolar o efeito que ocorre quando o controle experimental não permite que o experimentador elimine, de maneira razoável, explicações alternativas para uma relação observada entre variáveis independentes e dependentes, o que é chamado de “confusão” *confounding*. Em outras palavras, esta variável é algo que pode estar influenciando o experimento mas que não está, a princípio, no modelo original, e é uma potencial fonte de viéses no experimento.

A ANCOVA, permite, portanto, um controle do erro experimental, aumentando sua precisão. Vale ressaltar que a técnica não é restrita apenas à indicação de existência ou não de diferenças entre os níveis avaliados.

Considerando um experimento com um fator e uma covariável, o modelo estatístico da análise de covariância pode ser estabelecido como [Montgomery and Runger, 2011]:

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \epsilon_{ij}, \quad (2)$$

em que: μ é uma constante, τ_i é o efeito do i -ésimo “tratamento”, X_{ij} é o valor observado da covariável e $\bar{x}_{..}$ a média dos valores x_{ij} (da covariável). Além disso, β consiste no coeficiente de regressão linear entre a covariável X e a variável de resposta Y , com $\beta \neq 0$ e ϵ_{ij} é uma componente aleatória de erro, com $i = 1, \dots, a$ e $j = 1, \dots, n$. Neste caso a relação deve ser linear. Vale ressaltar que este modelo pressupõe que a variável de resposta e a covariável estão relacionadas linearmente. Para efeito de simplificação, contudo, a análise de covariância pode ser vista como uma análise “ajustada” de variância, e pode ser sumarizada na tabela (??):

em que:

Table 1: ANCOVA como um “modelo ajustado” da ANOVA.

Fontes de variação	\sum de quadrados	Graus de Lib	média quadrática	F_0
regressão	(S_{xy}/S_{xx})	1		
Tratamentos	$\frac{SS'_E - SS_E}{(S_{xy})^2/S_{xx} - [E_{yy} - (E_{xy})^2/E_{xx}]}$	$a - 1$	$\frac{SS'_E - SS_E}{a-1}$	$\frac{1}{MSE} \frac{SS'_E - SS_E}{a-1}$
Erro	$SS_E = E_{yy} - (E_{xy})^2/E_{xx}$	$a(n-1) - 1$	$\frac{SS_E}{a(n-1)-1}$	
Total	S_{yy}	$an - 1$		

$$S_{yy} = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{an}, \quad (3)$$

$$S_{xx} = \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - \frac{x_{..}^2}{an}, \quad (4)$$

$$S_{xy} = \sum_{i=1}^a \sum_{j=1}^n x_{ij} y_{ij} - \frac{(x_{..})(y_{..})}{an}, \quad (5)$$

$$T_{yy} = \frac{1}{n} \sum_{i=1}^a y_{i.}^2 - \frac{y_{..}^2}{an}, \quad (6)$$

$$T_{xx} = \frac{1}{n} \sum_{i=1}^a x_{i.}^2 - \frac{x_{..}^2}{an}, \quad (7)$$

$$T_{xy} = \frac{1}{n} \sum_{i=1}^a x_{i.} y_{i.} - \frac{(x_{..})(y_{..})}{an}, \quad (8)$$

$$E_{yy} = S_{yy} - T_{yy}, \quad (9)$$

$$E_{xx} = S_{xx} - T_{xx}, \quad (10)$$

$$E_{xy} = S_{xy} - T_{xy}. \quad (11)$$

Assim, rejeita-se a hipótese nula $H_0 : \beta = 0$ se $F_0 > F_{\alpha, 1, a(n-1)-1}$.

3. Resultados Análise Estatística

3.1 Tratamento e Validação dos Dados

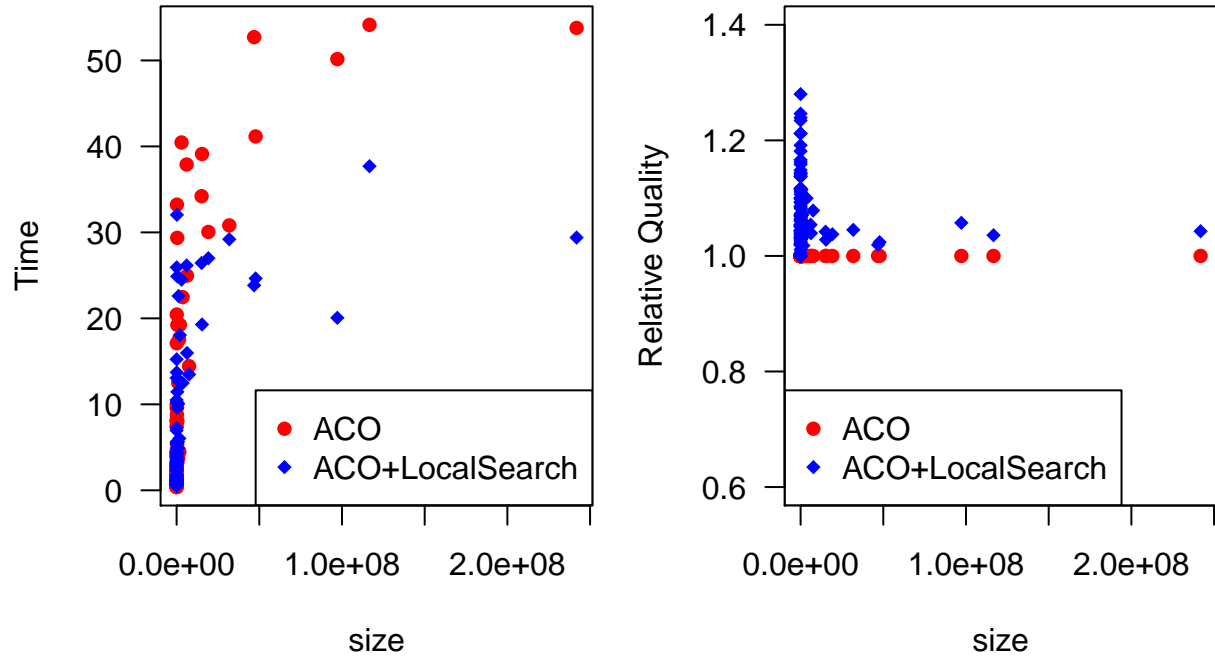
Após a obtenção dos dados, foi feita uma rotina de validação para identificar possíveis erros. Para isso, as seguintes condições foram avaliadas em cada observação:

1. Tempo > 0
2. 0 $>$ Qualidade > 1

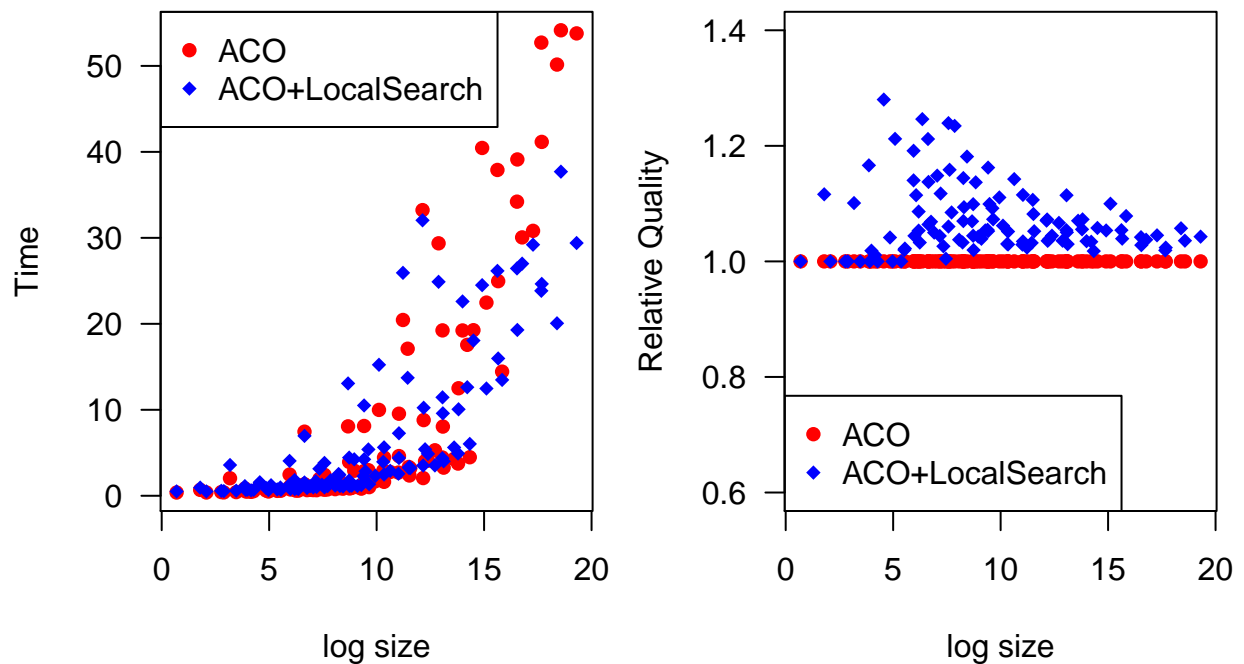
Em seguida, as diferentes execuções realizadas para cada instância foram agrupadas através da média. Dessa forma, cada bloco considerado é resumido por única estatística.

3.2 Análise Exploratória dos Dados

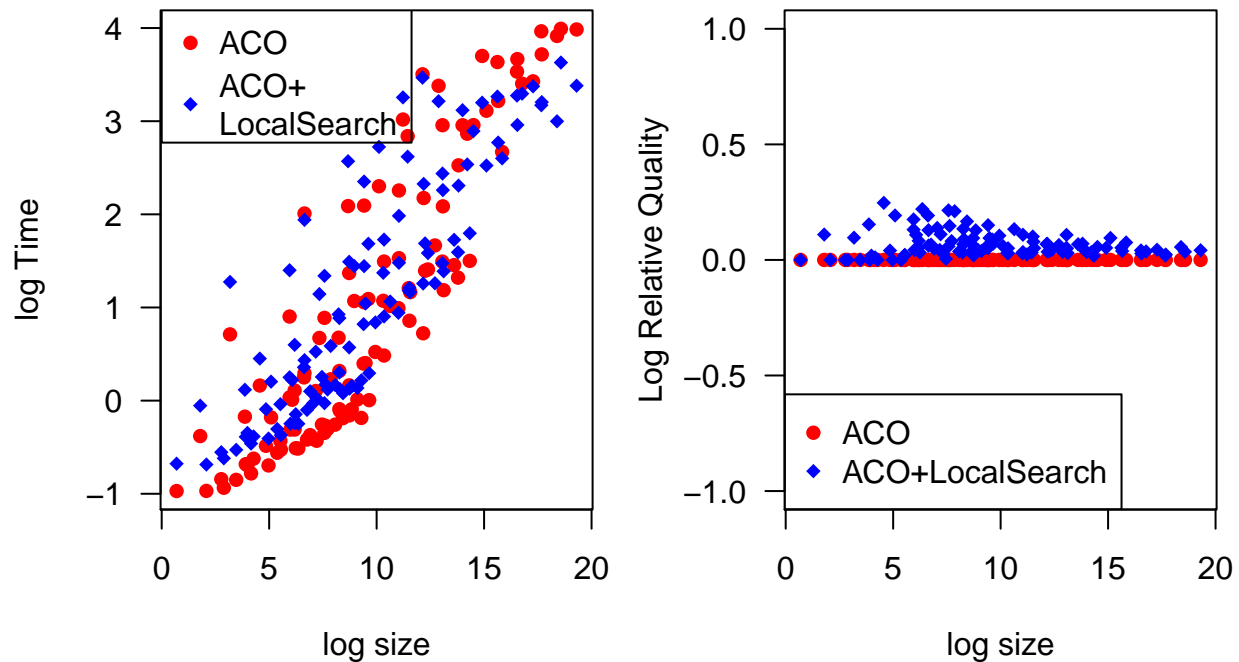
Foi então realizada uma análise qualitativa dos dados. Para cada variável de resposta, foram plotados seus gráficos em função do tempo, identificando cada algoritmo.



A escala não permite ter uma visualização adequada dos dados. Assim, é aplicada uma transformação logarítmica na variável Size para solucionar esse problema.



Para normalizar a distribuição dos dados, foi aplicada uma transformação logarítmica também nas variáveis de resposta. Essa transformação também permite obter uma relação mais linear entre a covariável tamanho e o tempo,



Os resultados parecem indicar uma relação linear entre tempo e tamanho de problema (ambos em escala logarítmica). Além disso, a qualidade relativa das soluções parece não depender do tamanho. Finalmente, há indícios de que o algoritmo proposto é melhor que a alternativa tanto em quesitos de tempo quanto qualidade da solução encontrada.

3.2 Análise de Variância com Blocagem

É então aplicado o método ANOVA com blocagem para obter a significância estatística das variáveis consideradas. Para ser possível aplicar o modelo ANOVA, a variável contínua de tamanho deve ser considerada como um fator. Assim, o tamanho é avaliado como uma variável com 99 níveis (número de tamanhos diferentes nas 108 instâncias avaliadas). Os testes são realizados para as duas variáveis de resposta.

Tempo de Execução

Inicialmente, a análise é realizada sem blocagem. Essa análise ignora os efeitos de tamanho do problema e tenta avaliar a diferença de tempo entre os métodos considerando apenas os algoritmos:

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## algorithm    1    1.3    1.308    0.706  0.402
## Residuals  214   396.6    1.853
```

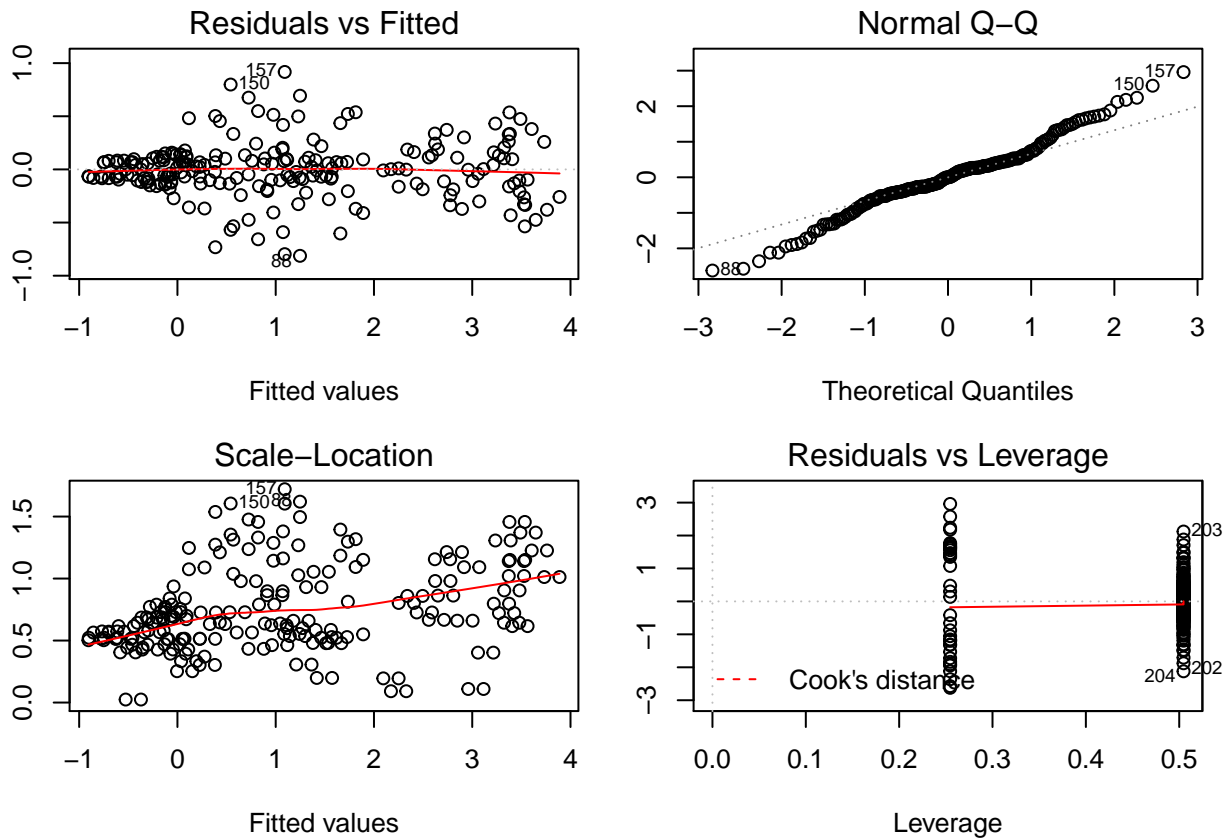
O resultado não indica nenhum efeito significativo dos algoritmos. Isso condiz com o esperado, uma vez que a análise visual indica que quaisquer possíveis efeitos dos algoritmos são muito menores que o efeito devido à variações no tamanho.

Incluindo o efeito do tamanho da instância no modelo, tem-se:

```
##           Df Sum Sq Mean Sq F value  Pr(>F)
## algorithm      1    1.3    1.308   10.15 0.00185 **
## as.factor(log(size)) 98  381.7   3.895  30.23 < 2e-16 ***
## Residuals     116   14.9    0.129
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



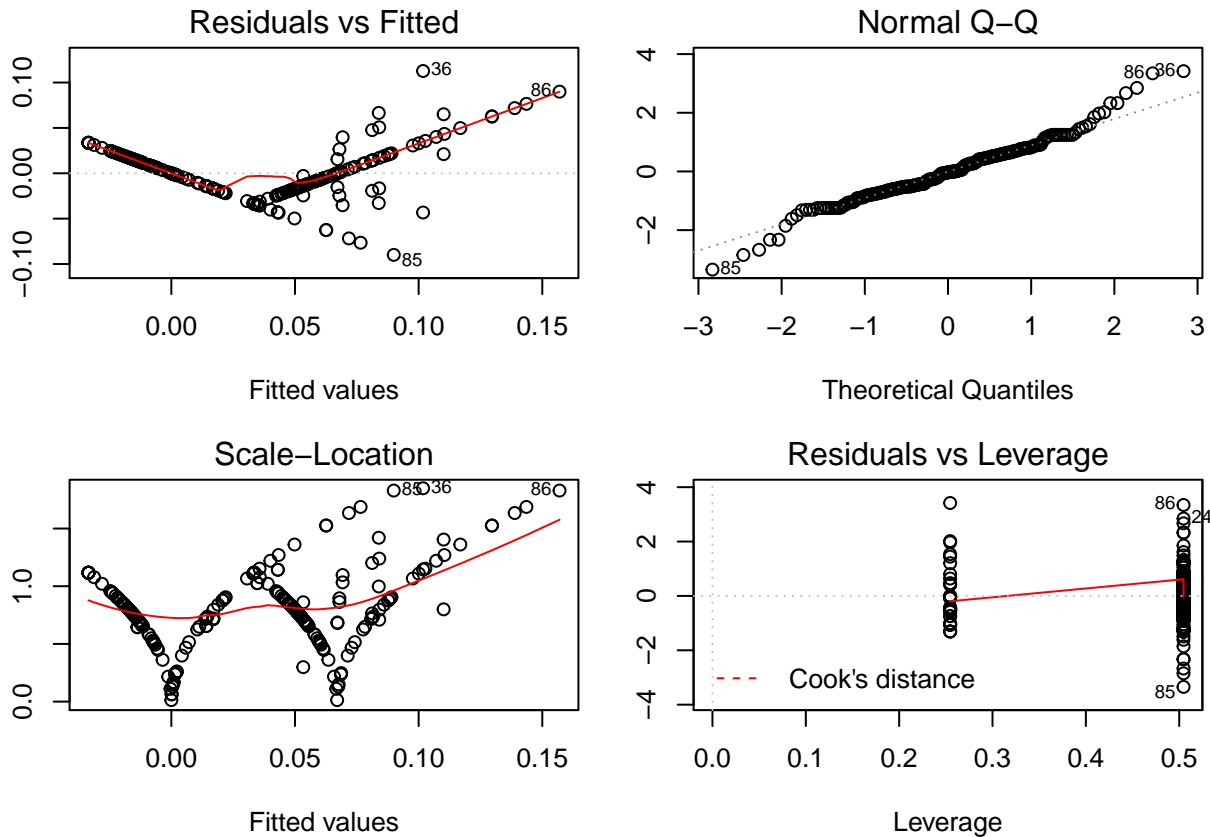
Nesse caso, é possível verificar que, uma vez compensado o efeito do tamanho, é possível verificar um efeito do algoritmo utilizado no tempo total de execução. Ainda assim, esse efeito é bastante menor que o efeito do tempo, o que pode ser concluído observando a relação da soma quadrática dos fatores e dos resíduos. A soma quadrática do tamanho é significativamente maior, o que indica uma maior contribuição desse fator para explicar o efeito observado.

O gráfico de distribuição dos resíduos (e resíduos padronizados) em relação ao valor ajustado indica homocedasticidade. O segundo gráfico quantil-quantil indica que os dados transformados possuem uma distribuição praticamente normal, o que valida a premissa de normalidade. Esses resultados são confirmados pelos testes Fligner-Killeen e Shapiro-Wilk, respectivamente. O design aleatório da ordenação das instâncias nos testes realizados permite assumir independência entre as observações. Assim, as premissas do teste ANOVA são bem atendidas. Finalmente, o gráfico Residuals vs. Leverage não indica nenhuma observação com Cook's distance muito diferente dos demais, o que poderia sugerir um possível outlier.

Qualidade da Solução

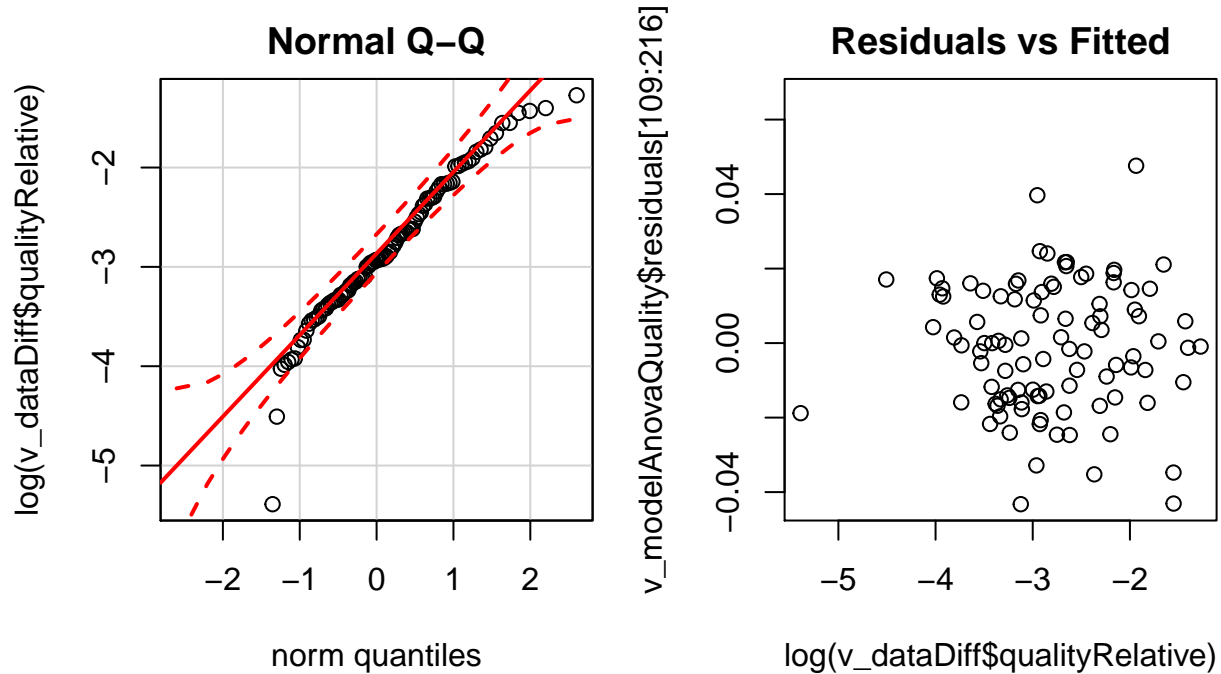
A mesma análise realizada para o tempo de execução é aplicada para a qualidade da solução:

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## algorithm      1  0.2427  0.24272  166.369 <2e-16 ***
## as.factor(log(size)) 98  0.1419  0.00145   0.992  0.514
## Residuals     116  0.1692  0.00146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



O resultado do teste ANOVA condiz com o esperado após a análise exploratória. O algoritmo possui um efeito significativo na qualidade relativa, enquanto o tamanho do problema não possui significância para explicar as diferenças de qualidade entre os algoritmos.

Os gráficos de distribuição dos resíduos e qq-plot possuem uma configuração que desperta dúvidas e pode indicar violações nas premissas. No entanto, essa configuração é consequência da transformação utilizada para avaliar qualidade relativa das soluções, que concentra a qualidade de todas as soluções do algoritmo padrão no valor 1. Para contornar esse problema e avaliar a normalidade da distribuição e homocedasticidade dos resíduos, é então considerada a diferença entre qualidade relativa das soluções



O gráfico indica que a distribuição das diferenças é bem próxima da distribuição normal. Não há também indícios de heterocedasticidade, o que valida as premissas do teste.

Uma vez que foi verificado que o tamanho do problema não é significativo para a qualidade relativa da solução, é possível realizar um teste-t pareado para testar a hipótese de que o algoritmo proposto apresenta qualidade maior que o original, considerando um tamanho de efeito de 5%.

```
##
## One Sample t-test
##
## data: (v_dataDiff$qualityRelative)
## t = 12.335, df = 107, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  0.06137889      Inf
## sample estimates:
## mean of x
## 0.07091809
```

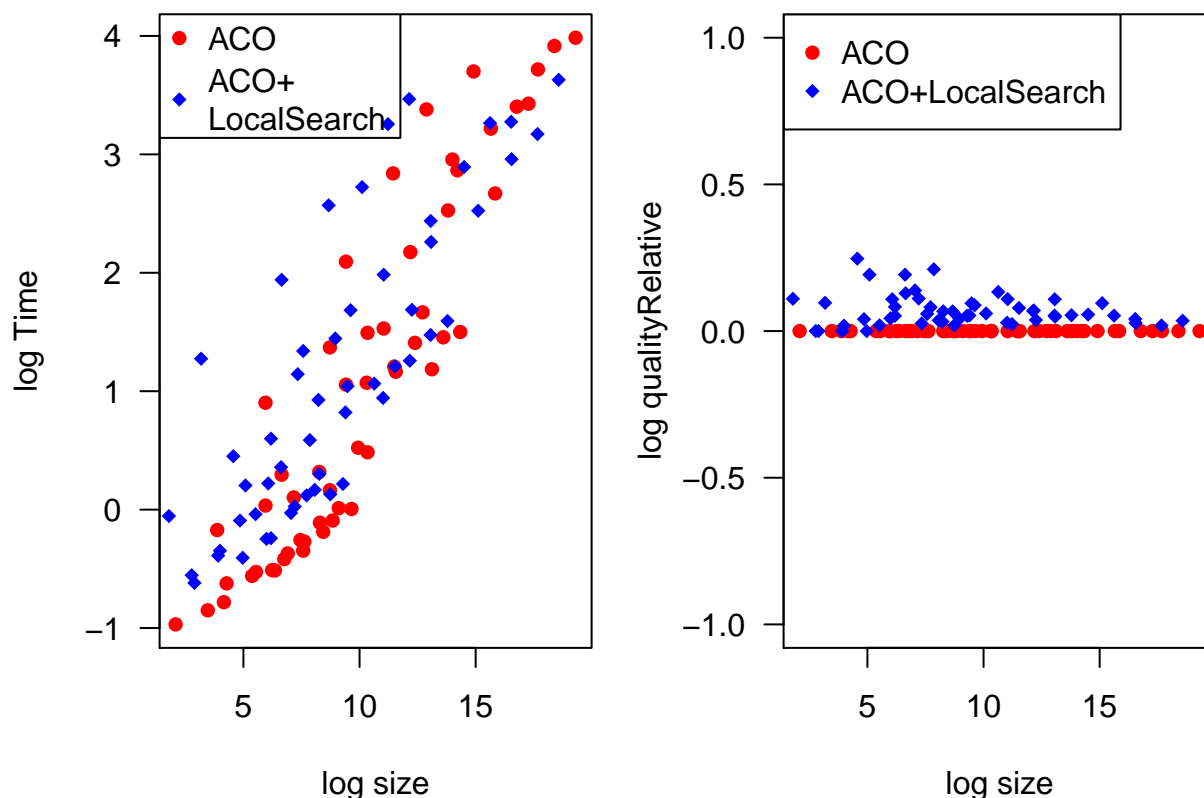
O p-valor encontrado rejeita a hipótese de que a qualidade dos algoritmos são equivalentes. O intervalo de confiança de 95% obtido indica que a média da diferença relativa de qualidade é de pelo menos 6.137%, em favor do algoritmo proposto.

3.3 Análise de Covariância

A Análise de Covariância é uma maneira mais adequada de lidar com variáveis contínuas. Além disso, ela permite realizar a análise mesmo quando uma covariável contínua não pode ser controlada e as observações possuem diferentes valores para a mesma, de forma que não podem ser pareadas.

Para ilustrar esse cenário, o conjunto de dados foi truncado para que, para cada instância, apenas resultados de um algoritmo esteja disponível. O objetivo principal é demonstrar que, mesmo na situação onde não é possível realizar o pareamento, ainda é possível fazer a análise do efeito da variável de interesse. O novo conjunto de dados possui portanto metade das observações do conjunto anterior.

O gráfico abaixo ilustra o novo conjunto de dados, já transformado:



O modelo ANCOVA é primeiro ajustado em relação à uma das variáveis. O erro restante é então ajustado em relação as variáveis seguintes. Sendo assim, há uma relação de dependência da segunda variável em relação à primeira. Por isso, a ordem das variáveis utilizadas no modelo é importante.

A seguir, é realizada a análise dos dados para cada variável de resposta.

Tempo

Inicialmente, é ajustado um modelo do tempo tomando o algoritmo como variável independente:

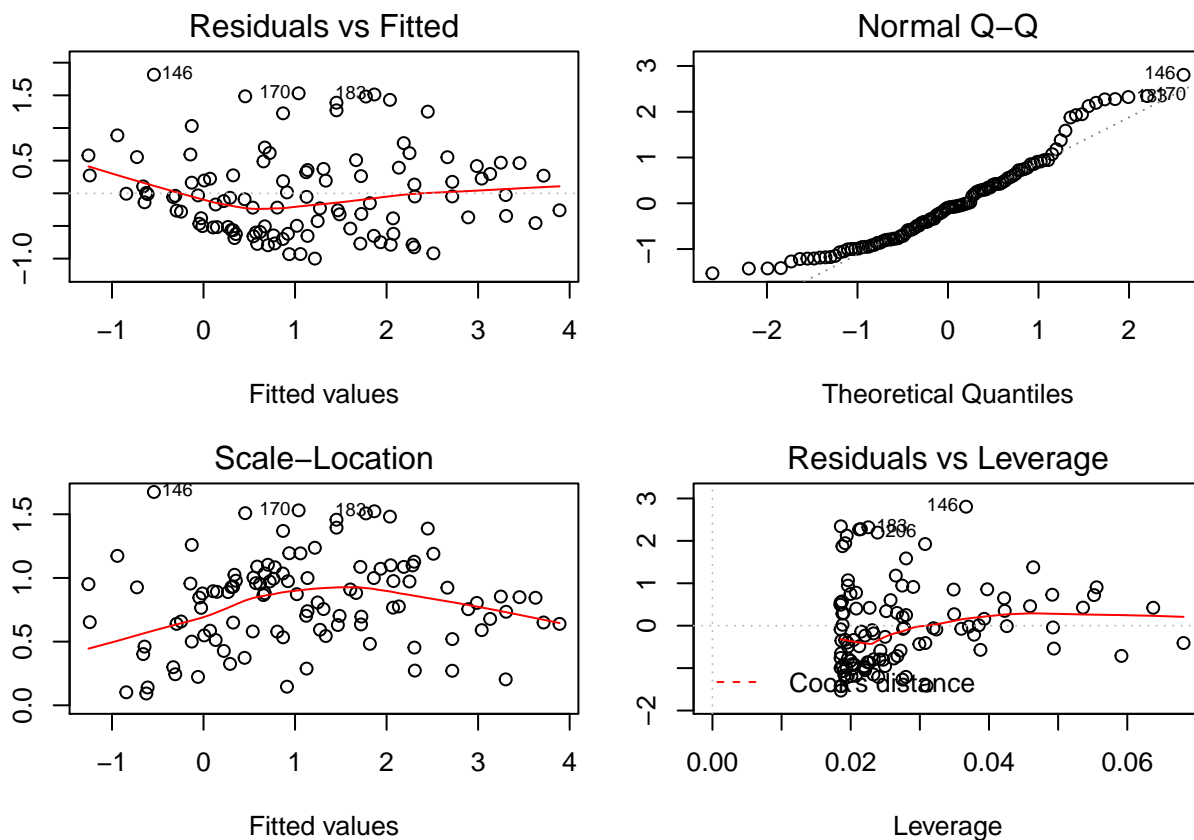
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## algorithm    1   0.07    0.07   0.154  0.695
## log(size)    1 153.64  153.64 353.305 <2e-16 ***
## Residuals  105  45.66    0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nesse modelo, não foi identificada nenhuma significância do algoritmo para explicar o tempo de execução, e apenas o tamanho tem significância. Isso condiz com o resultado esperado, uma vez que o efeito do algoritmo observado no gráfico é ordens de grandeza menor que o efeito do tamanho.

Em seguida, ajusta-se um modelo tomando o tamanho como variável independente.

```
##           Df Sum Sq Mean Sq F value  Pr(>F)
## log(size)    1 149.79  149.79 344.452 < 2e-16 ***
```

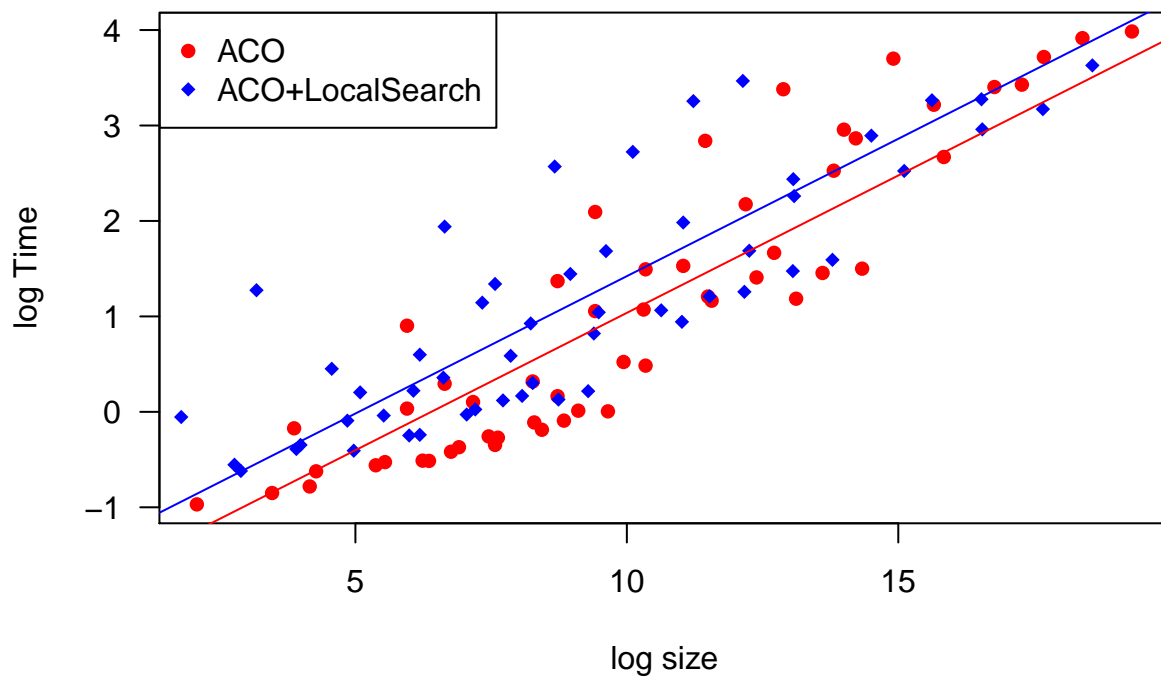
```
## algorithm      1    3.92    3.92    9.007 0.00336 **
## Residuals     105  45.66    0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Nesse modelo, é possível notar que, uma vez compensado o efeito do tamanho, o algoritmo possui sim um efeito significativo no tempo de execução. Assim, a análise realizada leva às mesmas conclusões que o teste ANOVA aplicado na seção anterior.

Os gráficos e testes realizados permitem avaliar a normalidade dos dados e homocedasticidade dos resíduos. Há indícios de pequenas violações na normalidade, mas o método ANCOVA é robusto a isso.

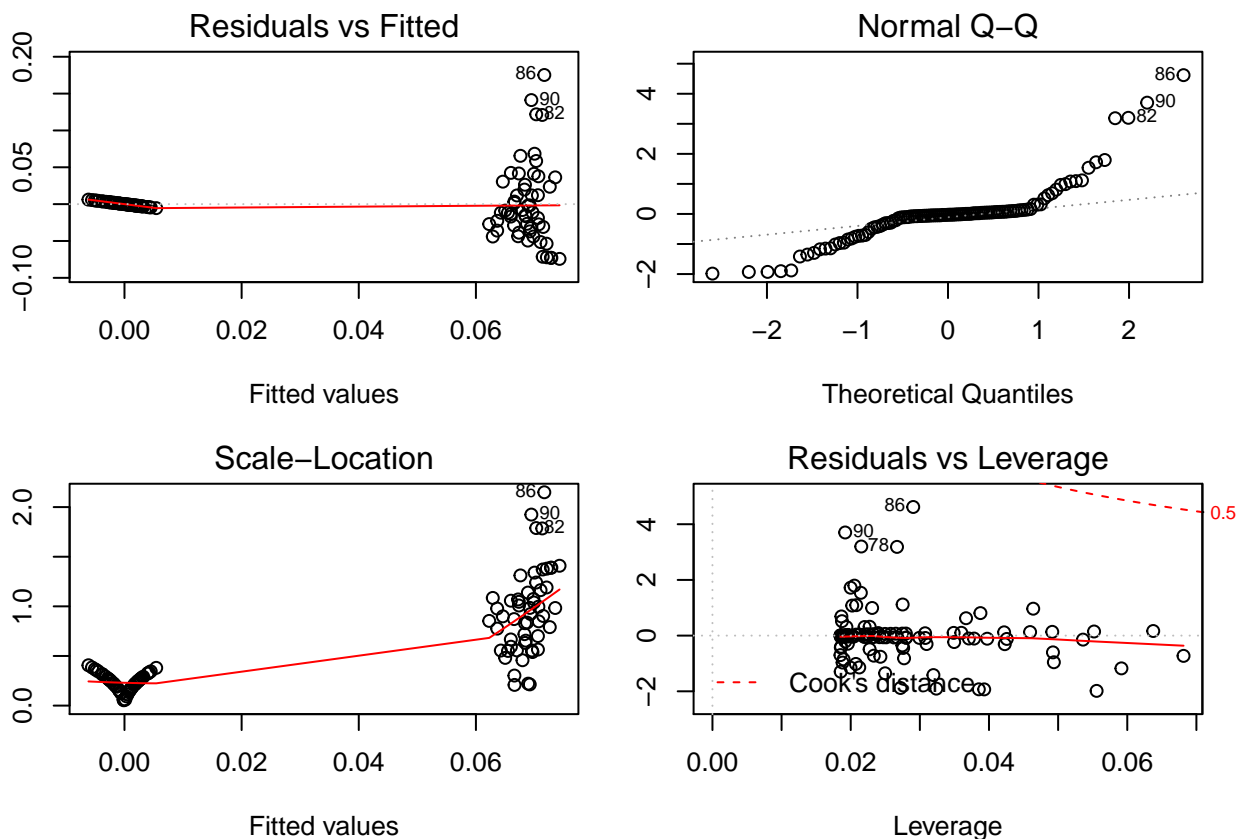
Abaixo, é feita uma regressão linear dos resultados, que parece indicar que o algoritmo original possui tempos um pouco menores que o proposto.



Qualidade

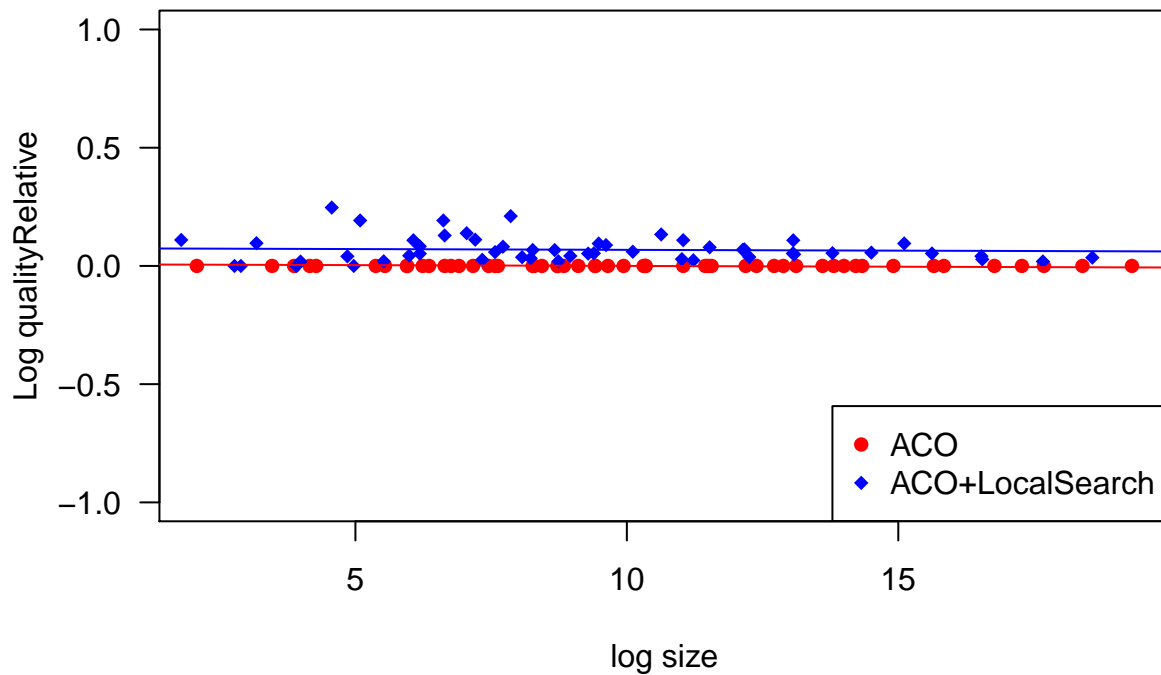
A mesma análise é feita para a qualidade dos algoritmos:

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## algorithm    1  0.12746   0.12746   85.986 2.62e-15 ***
## log(size)    1  0.00084   0.00084    0.564   0.454
## Residuals  105  0.15565   0.00148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



O resultado novamente condiz com o resultado obtido no teste ANOVA. O algoritmo tem um efeito significativo sobre a qualidade relativa, enquanto o tamanho do problema não. As premissas de normalidade e homocedasticidade são validadas de maneira análoga ao teste ANOVA.

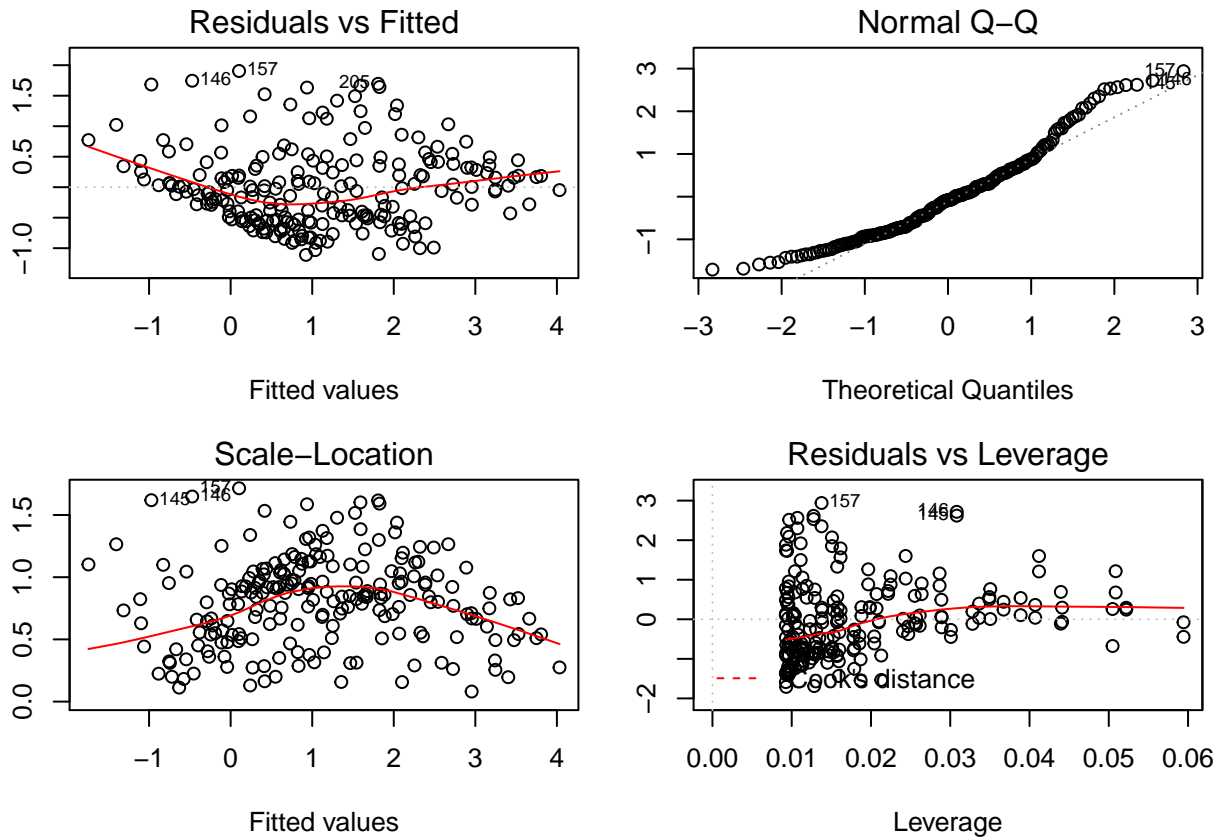
A regressão linear abaixo permite observar a superioridade da média da qualidade relativa do algoritmo proposto, bem como o efeito (praticamente nulo) do tamanho do problema.



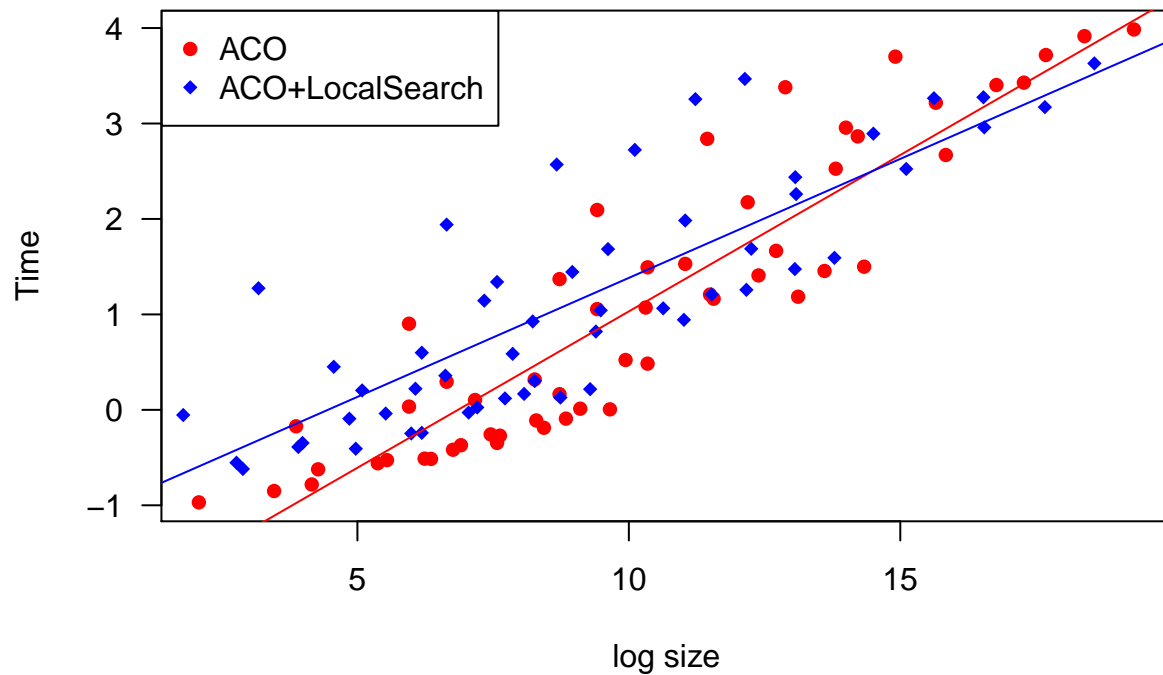
Validação das Premissas

Conforme discutido na seção 2.3, o teste ANCOVA depende fortemente da premissa de igualdade do coeficiente das retas de regressão de cada grupo avaliado em relação à covariável. Para verificar se essa premissa é atendida, é considerado um modelo com termos de interação entre tempo e tamanho do problema.

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## algorithm      1   1.31    1.31    3.072 0.0811 .
## log(size)      1 303.60  303.60 712.948 <2e-16 ***
## algorithm:log(size) 1   2.76    2.76    6.488 0.0116 *
## Residuals    212   90.28    0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



O resultado desse modelo indica que o termo de iteração é significativo para explicar o tempo de execução. Isso é um sinal de que a premissa pode ser violada. Realizando a regressão com um modelo com termos de iteração, o seguinte resultado é obtido:



4. Discussão e Conclusões

Referências

- Felipe Campelo. Lecture notes on design and analysis of experiments. <https://github.com/fcampelo/Design-and-Analysis-of-Experiments>, 2015. Version 2.11, Chapter 7; Creative Commons BY-NC-SA 4.0.
- D. C. Montgomery. *Design and analysis of experiments*, volume 7. Wiley New York, 1984.
- D. C. Montgomery and G. C. Runger. *Applied Statistics and Probability for Engineers*, volume 5. John Wiley and Sons, 2011.