Estudo de Caso 03: Comparison of Rising Drilling Configurations

Equipe 04

24 de Junho de 2017

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

1. Descrição do Problema

Tubos de perfuração *(drilling risers)* consistem em um tipo de conduíte ou tubo utilizado para transportar o petróleo extraído em plataformas oceânicas para a superfície. Um pesquisador precisa comparar o tempo médio até a falha *(mean time to failure ou MTTF)* de quatro configurações diferentes (níveis A, B, C, D) de equipamentos, de forma a escolher a que forneça a menor probabilidade de falha, considerando-se um período de 20 anos. Para isto será comparada a configuração padrão *(Riser1)* com as outras três, buscando encontrar qual delas proverá o maior $MTTF$. Assim, o experimento pode ser sumarizado na seguinte questão:

Algum dos risers alternativos é melhor que o padrão? Se sim, qual deles é o melhor?

Para realizar o experimento, serão feitos testes utilizando de um modelo de perfuradores em menor escala, sujeitos a um protocolo de testes de tempo de vida acelerado, em que há uma relação direta entre o tempo medido de cada observação (em minutos) e a configuração do sistema real. O custo de cada observação é, em média, $US\$10.000$ (dez mil dólares) por observação. Além dos dados obtidos com os testes, existem dadosdados históricos existentes para a configuração padrão *(Riser1)*.

Os parâmetros experimentais desejados são:

- Nível de significância: $\alpha = 0.05$;
- Tamanho de efeito de interesse prático: $\delta^{\star} = 0.25$;
- Potência desejada: $(1 - \beta) = \pi \geq 0.85$.

2. Planejamento Experimental

A priemira etapa do experimento é a verificação de diferenças estatísticas significativas entre as configurações testadas. Se diferenças forem verificadas e as premissas forem validadas, a etapa seguinte consiste consiste em investigar e quantificar essas diferenças através de testes de comparações múltiplas para as médias. Deve ser também efetuada uma análise exploratória qualitativa dos dados para complementar os testes.

A variância do processo será estimada utilizando-se dados históricos de operação disponíveis da configuração padrão. Ela será considerada uniforme para todas as configurações.

Os dados experimentais utilizados foram obtidos através de simulação, por meio de um aplicativo web. A data de nascimento do segundo membro mais jovem da equipe (15/10/1992) foi o parâmetro utilizado como semente para o gerador de números da simulação. Os dados gerados informam uma tabela com os níveis de cada fator de interesse em uma coluna e os tempos (tomados em escala logarítmica) em coluna respectiva. Por este motivo não necessitam de tranformações inversas.

2.1 Análise de Variância

Para a avaliar existência de diferenças significativas entre as configurações de perfuradores, será utilizado o teste estatístico ANOVA.

Essa técnica analisa médias e variâncias de observações de diferentes grupos para verificar se existe diferença estatístixa significativa entre as médias desses grupos. Assim, generaliza o teste t para mais de dois grupos, permitindo que sejam comparados simultâneamente [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 é 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.

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

Esta etapa foi considerada como uma primeira verificação ao questionamento deste estudo por apresentar um 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.

2.2 Comparações Múltiplas

Caso a análise ANOVA identifique a existência de diferenças entre os níveis, deve-se proceder com testes de comparação múltiplas. As comparações devem ser realizadas em relação à configuração padrão, de maneira a verificar se alguma das alternativas propostas é melhor. Desta forma, caso a análise de variância indique a existência de diferenças entre os níveis, será aplicado o teste de comparações múltiplas um-contra-todos (one-vs-all) de *Dunnett*, onde os *Risers* propostos serão confrontados com a configuração padrão *Riser1* para verificar se alguma proposta traria ganho de $MTTF$ frente à configuração já estabelecida.

Para realizar o teste de comparações múltiplas, deve-se manter o controle sobre os erros do tipo-I em cada comparação de maneira com que ele não se acumule a cada teste sucessivo. Assim, os valores de α são corrigidos para cada teste através do método de correção de Bonferroni (2):

$$\alpha_{adj} = \frac{\alpha_{família}}{K}\,, \tag{2}$$

no qual $K = 3$ consiste no número de comparações a serem feitas que no caso do teste um-contra-todos de *Dunnett*. O número de comparações nesse caso é dado pela Equação (3):

$$K = a - 1\,; \tag{3}$$

na qual $a = 4$ indica oo número de níveis. Assim, utiliza-se o nível de significância ajustado $\alpha_{adj} = 0.0166667$.

2.3 Definição do Tamanho Amostral

O experimento realizado possui duas etapas: ANOVA para verificar diferença entre as médias dos algoritmos, seguida de comparções múltiplas one-vs-all (se necessário). Cada uma dessas etapas requer diferentes tamanhos amostrais.

O cálculo do tamanho amostral para a técnica ANOVA pode ser feito iterativamente até encontrar o número n tal que:

$$F_{(1-\alpha)} = F_{\beta;\phi}\,, \tag{4}$$

em que ambas distribuições F têm $(a-1)$ graus de liberdade no numerador e $a(n-1)$ no denomiador. O parâmetro de não-centralidade ϕ é dado por:

$$\phi = \frac{\left(n\sum_{i=1}^{a}\tau_i^2\right)}{\hat{\sigma}^2}\,. \tag{5}$$

O valor de τ para comparações "todos contra um" pode ser obtido a partir da relação (6)

$$\tau = \left(-\frac{(a-1)\delta^\star}{a}, \frac{\delta^\star}{a}, \frac{\delta^\star}{a}, \frac{\delta^\star}{a}\right)\,. \tag{6}$$

O resultado desse teste indica um tamanho amostral $n = 60$ observações em cada grupo.

Para calcular o tamanho amostral das comparações múltiplas one-vs-all, serão utilizadas as mesmas relações para a comparação de duas amostras independentes emparelhadas, alterando-se apenas os valores de α para os valores corrigidos α_{adj}, considerando-se as múltiplas hipóteses e $a-1$ comparações. Em comparações desse tipo, a potência é maximizada utilizando um número maior de observações para o grupo de controle, dado pela Equação (7) [Campelo, 2015]:

$$n_0 = n_i\sqrt{K}\,, \tag{7}$$

onde $K =3 $ é o número de comparações, n_0 é o tamanho amostral do grupo controle e n_i o número de observações dos demais grupos. Esse valor é calculado conforme a Equação (8):

$$n_i = \left(1+\frac{1}{K}\right)\left(\frac{\hat{\sigma}}{\delta^\star}\right)^2 (t_{\alpha_{adj}} + t_\beta)^2\,, \tag{8}$$

em que $t_{\alpha_{adj}}$ e t_β são dependentes de n.

Para solucionar o problema da dependência de n, os termos $t_{\alpha_{adj}}$ e t_β foram substituídos por $z_{\alpha_{adj}}$ e z_β e a equação foi testada iterativamente até a convergência (implementação em anexo no arquivo *calcN.R*).

O resultado obtido indica um tamanho amostral $n_0 = 101$ para o grupo controle e $n_i = 58$ para os demais. Considerando as 10 observações históricas já disponíveis para o *Riser1*, tem-se:

- Número de observações experimentais necessárias para o ANOVA: $60 \times 3 + 50 = 230$
- Número de observações experimentais necessárias a comparação todos contra um: $58 * 3 + 91 = 265$

O primeiro caso apresenta um custo experimental de $2.300.000, versus $2.650.000 para o segundo.

Serão necessárias 60 coletas para a realização do ANOVA e 58 para as comparações (excetuando-se o grupo controle), o que resultaria em uma sobreamostragem de 2 unidades por grupo. Entretanto, o segundo experimento só é necessário caso o ANOVA indique diferença entre as configurações dos *Risers*. Caso contrário, apenas o primeiro experimento já seria o suficiente, o que acarretaria em uma economiacerca de $350.000 fazendo o ANOVA antecipadamente. Tais informações sumarizadas na Tabela 1.

Assim, serão realizados inicialmente apenas experimentos necessários para o teste ANOVA, em virtude da possível economia e do possível menor custo com a amostragem. Caso sejam identificadas diferenças entre as configurações possíveis de *risers*, mais experimentos serão realizados para completar o tamanho amostral necessário.

	Melhor caso	Pior caso
Com ANOVA	$2.300.000	$2.710.000
Sem ANOVA	$2.650.000	$2.650.000
Diferença	$350.000	-$60,000

Table 1: Análise de custo das opções.

2.4 Tratamento e Validação dos Dados

Considerando o experimento realizado, foi criada uma rotina para validação dos dados obtidos e identificação de erros, onde o tempo de $MTTF$ na escala logarítma deve ser maior que 0.

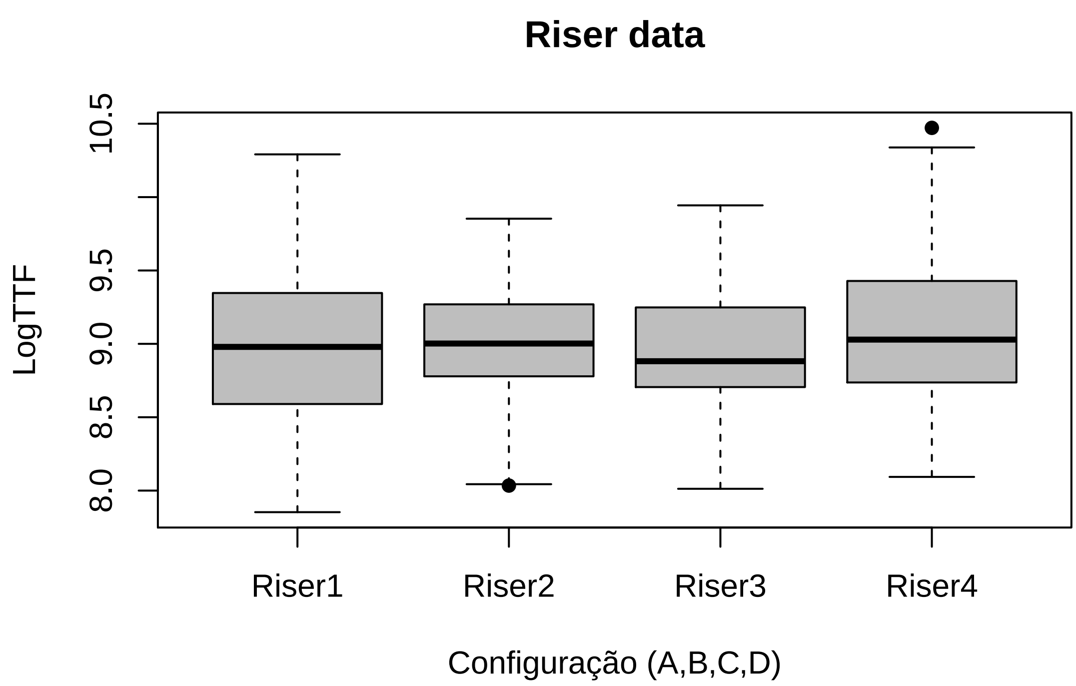
1. LogTTF > 0

Caso os valores de uma execução não atendam essas condições, ela seria descartada. No entanto, nenhuma das amostras apresentou tal problema.

3. Análise Estatística

3.1 Análise de Variância

O teste ANOVA realizado apresenta *F-valor = 0.608*, o que indica que a hipótese nula de que não há diferenças entre os grupos não foi refutada. O boxplot dos resíduos corrobora o resultado encontrado.

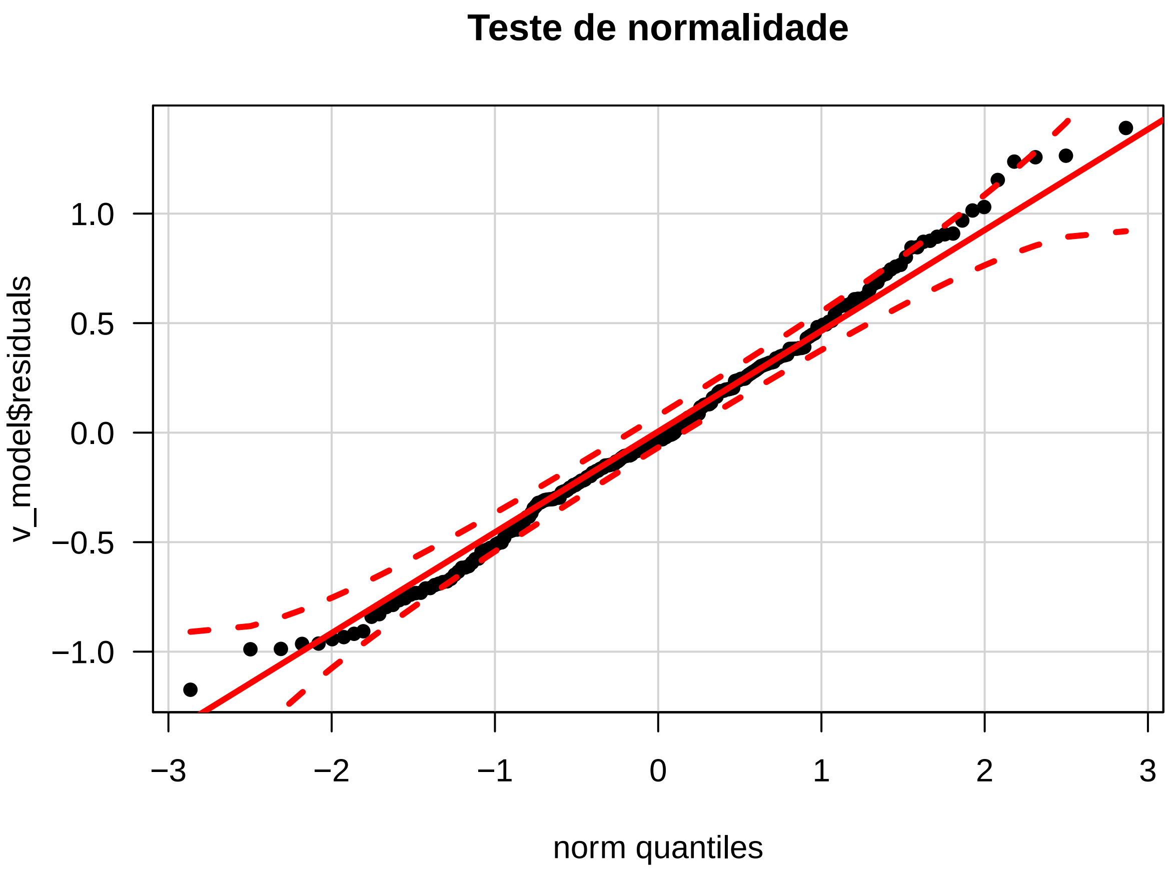


Este resultado indica que não há necessidade de testes de comparação múltiplas, que poderiam identificar quais os níveis diferentes do grupo de controle *(Riser1)*.

3.2 Validação das Premissas

Normalidade

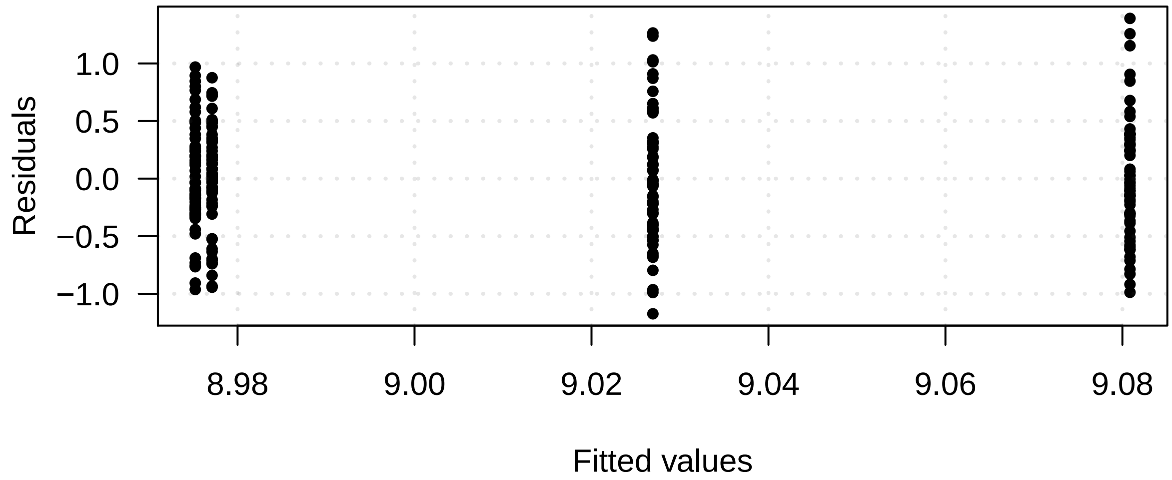
O *p-valor = 0.299* encontrado no teste de *Shapiro-Wilk* não indica rejeição hipótese de normalidade das amostras. A análise visual do qq plot confirma que não há violações de normalidade significativas. Além disso, o test ANOVA é robusto a pequenas variações de normalidade [Campelo, 2015]. Desta forma considerou-se a premissa de normalidade atendida.



Homocedasticidade

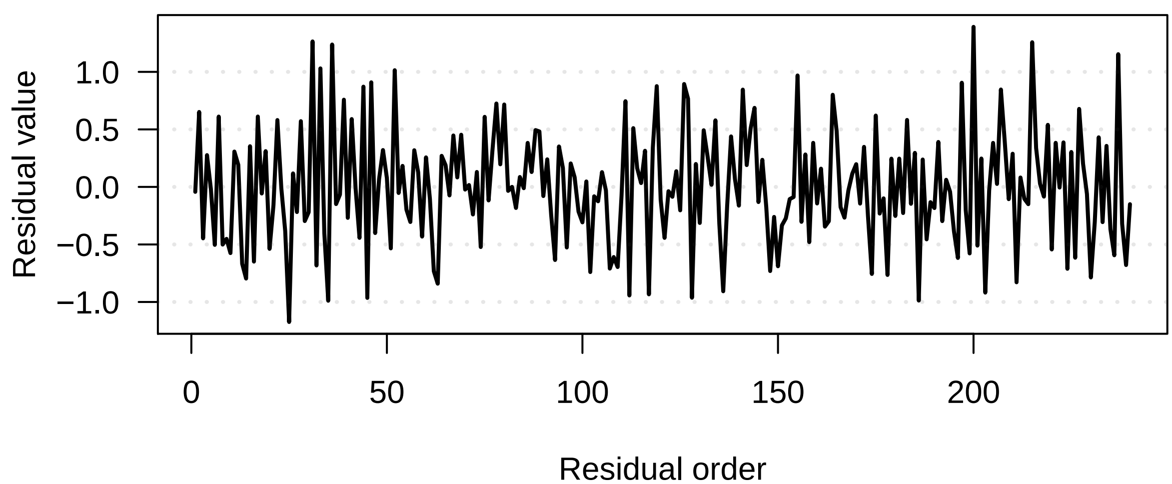
O teste de igualdade de variância dos resíduos de *Fligner-Killeen* apresentou um *p-valor = 0.1416*, o que indica falha em rejeitar a hipótese de homocedasticidade das amostras.

Além disso, podemos observar no gráfico que não há indíveios de diferença de variância significativa para entre os resíduos de diferentes grupos. O teste ANOVA também é robusto a pequenas violações na homocedasticidade como indicado por [Campelo, 2015]. Assim, essa premissa também é atendida.



Independência

O plot dos valores ordenados de diferenças de tempo entre os algoritmos não apresenta nenhum indício de dependência temporal dos valores. O teste de autocorrelação serial Durbin-Wastson apresenta $p = 0$, o que reforça a hipótese de que não há autocorrelação serial entre as amostras.



4. Discussão e Conclusões

Os testes realizados indicam que não há diferença significativa entre o tempo de vida das diferentes configurações de *Risers*. Todas as premissas do teste ANOVA realizado foram confirmadas, o que reforça o resultado alcançado. Dessa forma, recomenda-se manter a configuração padrão *Riser1*, uma vez que não há vantagem observada nas alternativas consideradas.

O custo do experimento realizado é significativo ($2.300.000), mas previniu um custo potencialmente maior de trocar a configuração de *Riser*.

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 and G. C. Runger. *Applied Statistics and Probability for Engineers*, volume 5. John Wiley and Sons, 2011.