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

O objetivo do problema é realizar uma comparação entre quatro tipos de tubos de perfuração *drilling risers*, que consiste em uma espécie de conduíte ou tubo utilizado para servir como passagem temporária para o petróleo extraído em plataformas oceânicas. Mais especificamente, pretende-se 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 serão comparadas a configuração padrão *(Riser 1)* com as outras três, buscando encontrar qual delas proverá o maior *MTTF*. Ou seja, a questão a se responder é:

Algum dos risers alternativos é melhor que o padrão?

A plataforma de testes escolhida pela equipe de engenharia consiste na utilização de um modelo em escala para os *risers* com um protocolo de tempo acelerado, em que há uma relação direta entre o tempo medido de cada observação (em minutos) e a configuração do sistema real. Todavia, o custo de cada observação é significativamente custoso, sendo cerca de $US\$10000$ (dez mil dólares). Afim de minimizar os custos com amostras coletadas é possível utilizar os dados históricos existentes para a primeira configuração *(Riser1)*.

Os parâmetros experimentais desejados são:

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

2. Planejamento Experimental

Esta etapa do presente estudo de caso consiste em investigar o comportamento dos níveis de fator a partir de uma análise exploratória, e posteriormente, aplicar um teste de comparações múltiplas para as médias. A variância do processo é desconhecida e será estimada utilizando-se os dados históricos de operação fornecidos em <https://git.io/vHDG3. A variância será considerada como sendo uniforme para todas as configurações neste estudo.

Os dados experimentais utilizados foram obtidos através de simulação, por meio de um [aplicativo web] <http://orcslab.cpdee.ufmg.br:3838/riserdata/. 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.

Os dados estão na escala de *log* e serão analisados nesta escala, pois não necessitam de tranformação para tal.

2.1 Análise de Variância

Para a avaliar o questionamento acima, será utilizada a ferramenta estatística para análise de variância denominada ANOVA.

A técnica ANOVA compara médias de diferentes amostras para verificar se estas possuem médias iguais ou não. Assim, essa técnica permite que vários grupos sejam comparados [1], [2].

Em outras palavras, a análise de variância é utilizada quando se quer decidir se os níveis apresentam médias diferentes em relação à uma média global μ. As diferenças entre as médias dos níveis e a média global é τ_i $\forall\ i$. Desta forma a ANOVA se basea no teste de a hipóteses abaixo, onde quando refutado a hipótese nula, evidencia-se indícios de diferenças entre os níveis.

Esta técnica é restrita apenas à indicação da existência ou não de diferenças entre os níveis avaliados sem indicar quais níveis seriam diferentes.

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

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 abaixo. Portanto, caso não seja detectada alguma diferenças entre os níveis o estudo será concluído por meio do teste ANOVA.

2.2 Comparações Múltiplas

Caso a análise ANOVA identifique a existência de diferenças entre os níveis, deve-se proceder com testes de comparação múltipla, no intuito de identificar qual ou quais níveis apresentam tal diferença.

Desta forma, somente, 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 que se proceda com o teste de comparações múltiplas, deve-se manter o controle sobre os erros do tipo-I em cada comparação. É preciso corrigir os valores de α para cada teste. A escolha para este caso é o método de correção de Bonferroni:

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

no qual $K = 3$ é o número de comparações a serem feitas que no caso do teste um-contra-todos de *Dunnett* é confome abaixo:

$$K = a - 1 \, ,$$

onde a é o número de níveis, que neste do estudo é 4 e o α experimental desejado é o $\alpha_{família} = 0.05$.

2.3 Definição do Tamanho Amostral

Para calcular o tamanho amostral neste caso, utilizaremos, portanto, as mesmas relações utilizadas na comparação de duas amostras independentes emparelhadas “todos contra um”, alterando-se apenas os valores de α para os valores corrigidos α_{adj} para as múltiplas hipóteses e $a - 1$ comparações. O tamanho amostral para os níveis que não são o grupo é calculado como abaixo:

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

em que $t_{\alpha_{adj}}$ e t_β são dependentes de n. Para solucionar esse problema, eles 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*). Dessa forma, foi encontrado o valor $n_i = 60$.

Contudo, para maximizar o poder do procedimento de múltiplas comparações, o tamanho amostral do grupo de controle deve ser calculado como [1]:

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

```
## Parsed with column specification:
## cols(
##   Riser = col_character(),
##   LogTTF = col_double()
## )
```

Já temos então, neste ponto, subsídios para o cálculo do tamanho amostral do ANOVA: (Discutir essa ideia ainda…)

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

Número necessário pra anova: 60*3 + 50

Número necessário pra análises subsequentes: 58*3 + (101 - 10)

Mais barato fazer ANOVA antes e pegar mais amostras do grupo 1 apenas se necessário. N das comparações multiplas não é tão grande por ser unilateral.

CUSTO TOTAL!

APRESENTAR OS VALORES DOS PARÂMETROS EXPERIMENTAIS

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.

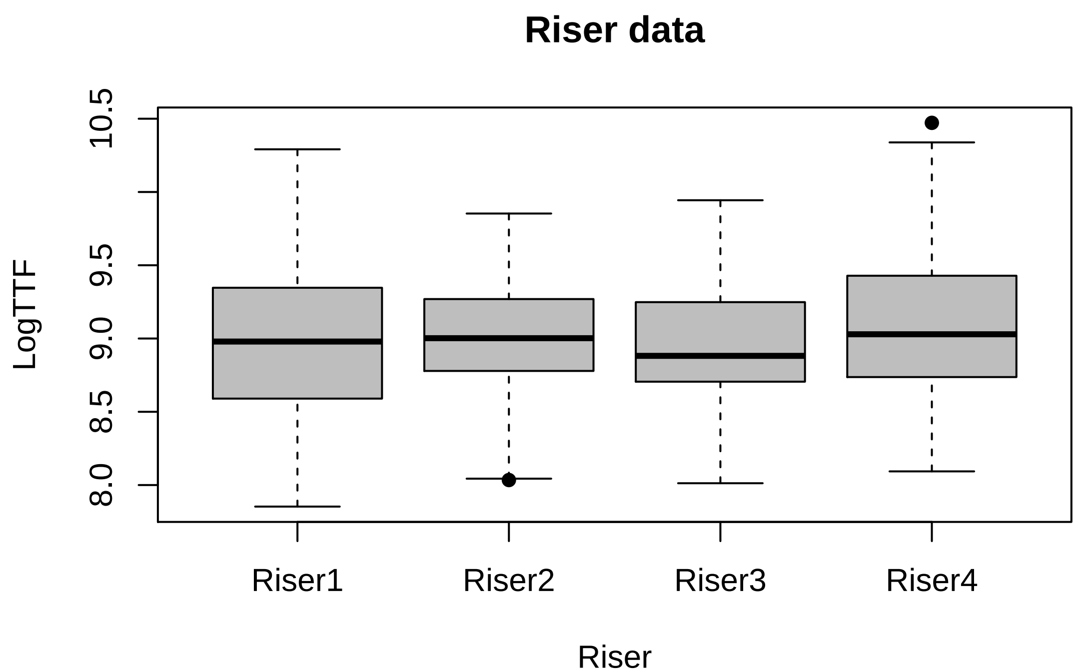
1. LogTTF > 0

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

3. Análise Estatística

3.1 Análise de Variância

```
## Parsed with column specification:
## cols(
##   Riser = col_character(),
##   LogTTF = col_double()
## )
```

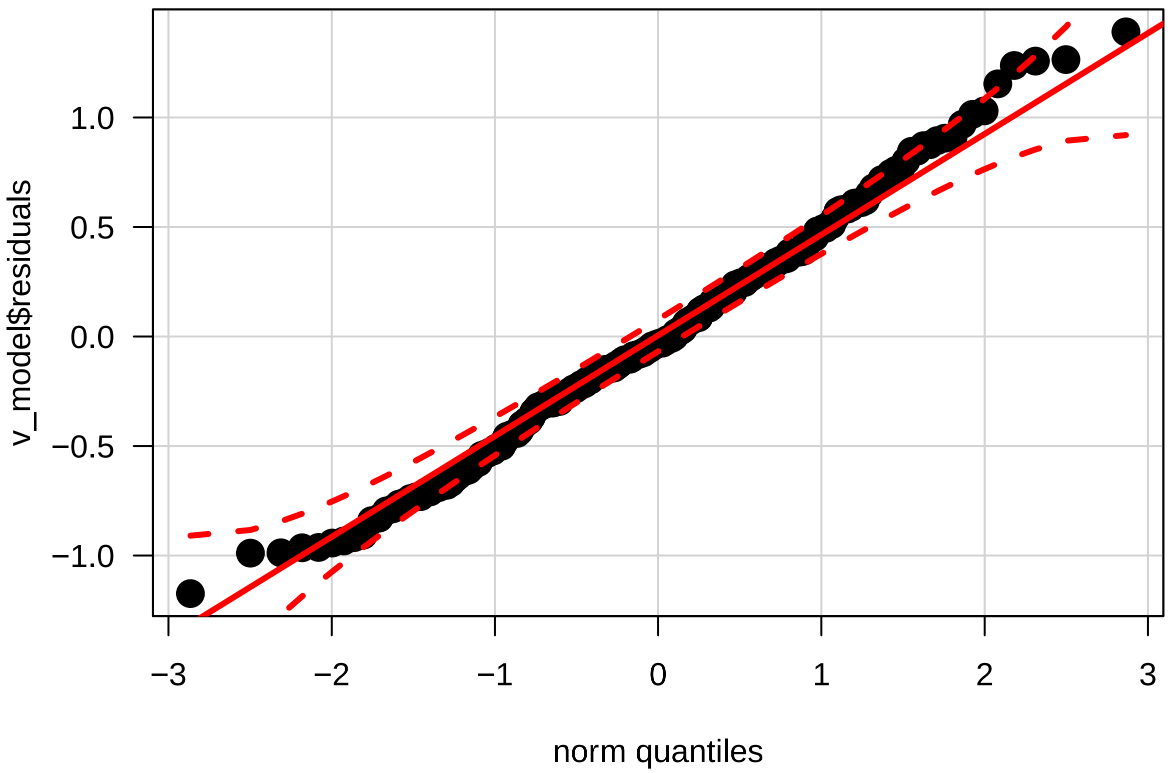


Como pode ser observado, o gráfico acima não indica visualmente diferença significativa entre os níveis. Corrobora com isto, o *F-valor = 0.608* do teste ANOVA indicando pela não rejeição da hipótese de que não há diferença entre os grupos.

Este resultado também indica que não há necessidade de testes de comparação múltiplas que verificariam quais níveis teriam se apresentado diferentes do grupo de controle *Riser1*.

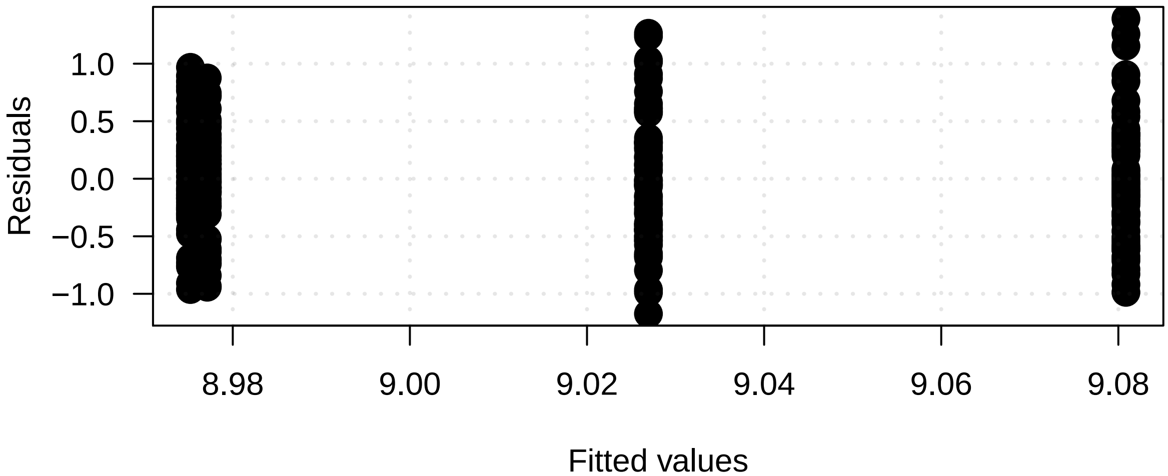
3.2 Validação das Premissas

Normalidade



O *p-valor = 0.299* encontrado no teste de *Shapiro-Wilk* indica pela rejeição normalidade das amostras No entanto, qq plot mostra que as violações de normalidade são relativamente pequenas. A análise ANOVA é robusto a pequenas variações de normalidade [1]. Desta forma considerou-se a premissa de normalidade atendida.

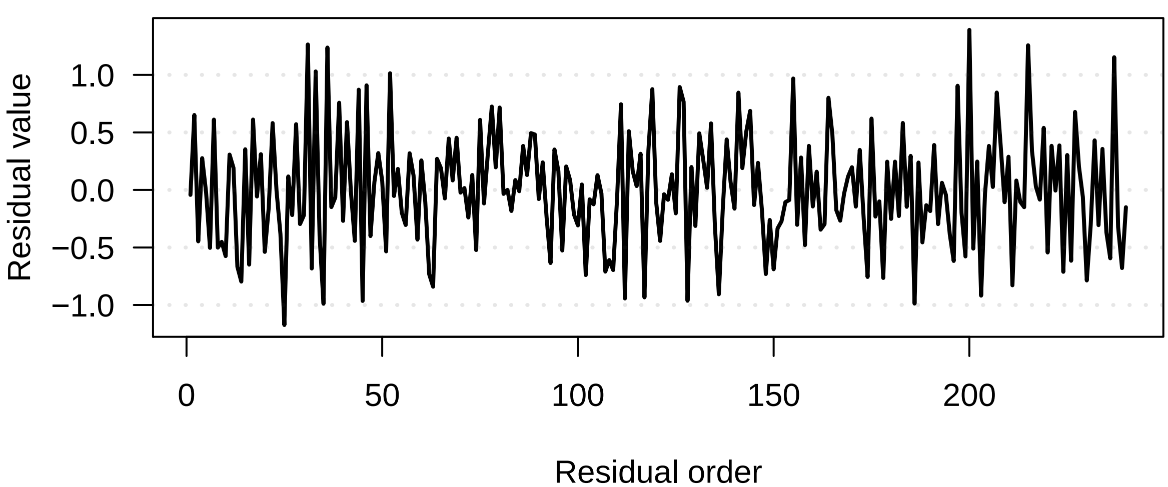
Homocedasticidade



O teste de igualdade de variância dos resíduos de *Fligner-Killeen* apresentou um *p-valor = 0.1416* o que também indica pela rejeição da homocedasticidade das amostras.

Contudo, podemos observar no gráfico acima, uma variância relativamente baixa entre os resíduos das amostras. Novamente com base de que a análise ANOVA é robusto a pequenas variações também de homocedasticidade como indicado por [1], considerou-se a premissa 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 levam às seguintes conclusões:

Variância entre grupos é explicada pela variância intra grupo. Não há indício de diferença significativa entre eles.

Recomenda-se manter riser 1. Custo do experimento é significativo, mas previniu um custo potencialmente maior de trocar o Riser.

Referências

[1] F. Campelo, "Lecture notes on design and analysis of experiments." https://github.com/fcampelo/Design-and-Analysis-of-Experiments, 2015.

[2] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, vol. 5. John Wiley; Sons, 2011.