Estudo de Caso 4 - Comparação de Configurações de Dutos de Exploração de Petróleo

Áthila Rocha Trindade

Sumário

Dutos de perfuração (Risers), são componentes importantes dos sistemas de exploração de petróleo submarina. Os *Risers* são responsáveis pelo transporte do petróleo do fundo do oceano à superfície (plataformas). Um pesquisador deseja comparar o tempo médio até a falha de 4 diferentes configurações dos *Risers*, para que possa ser escolhida aquela com a menor probabilidade de falha dentro de 20 anos (ou seja, com maior tempo para falha). Para comparação prática dos *Risers*, o pesquisador adota uma escala de tempo até falha em minutos, usando modelos "acelerados" de teste de vida dos *Risers*. A configuração padrão de *Risers* é denominada como *Riser 1*, e o pesquisador está interessado em saber se qualquer uma das outras 3 configurações irão proporcionar um maior tempo médio até falha.

Com o objetivo de reduzir os custos do experimento, serão utilizados dados históricos disponíveis para o *Riser 1* (10 amostras), dados estes que foram obtidos utilizando a mesma modelagem "acelerada" prevista para as outras 3 configurações (*Riser 2*, *Riser 3* e *Riser 4*). As 10 amostras iniciais sobre o *Riser 1* não tem custo, enquanto que cada nova amostra obtida para o experimento tem o custo de US$ 25000.

Como a distribuição dos dados é do tipo lognormal, os dados estão disponíveis na escala logarítmica. Deve atentar-se para possibilidades de minimização do custo total do experimento.

Projeto Experimental

A partir do *sumário*, temos que se objetiva comparar 3 configurações de *Risers* com relação a uma já utilizada (padrão). Para cada configuração de *Riser*, deverão ser obtidas amostras para a realização do experimento. Como o que quer ser investigado é se existe algum *Riser* melhor que o padrão (*Riser 1*), a questão de interesse é a seguinte:

Existe alguma configuração alternativa dos Risers melhor que a configuração padrao (Riser 1)?

Para responder a questão acima, deverão ser analisadas as diferenças entre as médias de tempo até falhas dos *Risers* 2, 3 e 4 com relação à média do *Riser 1* (chamado de grupo de "controle"); ou seja, o teste estatístico deve ser implementado na forma de um teste estatistico de múltiplas comparações, do tipo "1 contra todos"; tendo a execução prévia do teste ANOVA apontado diferenças entre as médias dos *Raisers*.

Sendo τ_i (i=1,... 4) os efeitos das configurações dos *Risers* 1 a 4 no tempo médio geral até falha, o teste de hipótese é o seguinte:

$$\begin{cases} H_0 : \tau_i = 0, \forall_i \in 1,..4 \\ H_1 : \exists \tau_i > \tau_1, i \in 2,..4 \end{cases}$$

os parâmetros para o teste de hipótese foram definidos previamente pela equipe do experimento como se segue:

- Nível de significância: $\alpha = 0.1$;
- Mínimo Efeito de Interesse: $\delta^* = 0.25$;
- Potência do teste: $(1 - \beta) \geq 0.8$;

Levando em consideração o tipo do teste estatístico e o custo de cada amostra que deve ser coletada, deve-se pensar em alternativas a serem adotadas para o experimento e então escolher a mais barata, sem prejuízos na precisão do teste (índice de tipos de Erro I e II).

Foram neste experimento elencadas e analisadas as seguintes alternativas de teste:

Alternativa 1

Utilizar as 10 amostras sem custo do *Riser 1* para estimar o número total de amostras requeridas para o teste estatistico *ANOVA* e então calcular o valor das amostras necessárias para o teste de Dunnetts (1 contra todos) não balanceado.

O valor da variância e desvio padrão para todas as amostras foi estimado com base nas 10 amostras iniciais e de acordo com o vaor médio do intervalo de confiança calculado à partir da distribuição X^2 (qui-quadrado). O intervalo de confiança para a variância é calculado da seguinte maneira:

$$IC = \frac{(n-1)\sigma^2}{X^2_{(1-\alpha)/2,(n-1)}}; \frac{(n-1)\sigma^2}{X^2_{(2-\alpha)/2,(n-1)}}$$

onde:

σ^2: valor da variância para as 10 observações gratuitas disponíveis.

n: número de observações gratuitas

Para cálculo do intervalo de confiança para o desvio padrão, basta obter a raiz quadrada dos limites do intervalo de confiança da variância. Então, os valores médios determinados para a variância e desvio padrão foram aproximadamente 0.42333 e 0.63469, respectivamente.

Com os valores estimado para variância e desvio padrão, o tamanho amostral para o ANOVA foi feito como mostrado a seguir:

```r
deltaestrela <- 0.25
alpha <- 0.1
beta <- 0.2
a <- 4

tau <- c(-((deltaestrela*(a - 1))/a), rep(deltaestrela/a, a-1))
vartau <- var(tau)
n = ceiling(power.anova.test(groups = a,
                             between.var = vartau,
                             within.var = v,
                             sig.level = alpha,
                             power = 1-beta)$n)

print(n)
```

```
## [1] 81
```

De acordo com os cálculos acima seriam necessários 81 x 4 = 324 amostras para execução do teste ANOVA. Supondo que o teste ANOVA apontasse diferenças entre as médias dos *Raisers*, então o teste de Dunnetts teria que ser feito. Desta forma, deve ser calculado o número de amostras necessárias para o teste de Dunnetts desbalanceado. As fórmulas utilizadas são as seguintes:

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

$$n_i = (1 + \frac{1}{\sqrt{K}})(\frac{(t_{\alpha*/2,K} + t_{\beta,K})\hat{\sigma}}{\delta^*})^2$$

onde:

n_0: é o tamanho amostral do grupo de controle

n_i: tamanho amostral dos demais grupos ou Raisers

K: número de comparações para o teste de Dunnets. Neste caso K=3 (*Raiser 1* X *Raiser 2*, *Raiser 1* X *Raiser 3*, *Raiser 1* X *Raiser 4*)

$\alpha*$: representa o valor do nivel de significância (α) ajustado para K testes comparativos. $\alpha*= \alpha$/K

$\hat{\sigma}$: representa o desvio padrão estimado à partir das 10 amostras iniciais

β e δ^* são parâmetros do teste estatístico definidos previamente

Os valores determinados pelos cálculos para n_0 e n_i foram 394 e 227, respectivamente. Caso o teste ANOVA apontasse diferenças nas médias dos *Raisers*, então seria necessário um número de amostras igual a n_0+3*n_i = 394 + 3 x (227) = 1075 amostras. Com relação ao número de amostras para o ANOVA, temos 751 amostras adicionais.

Portanto, temos os seguintes custos para a *Alternativa 1*, considerando o pior e o melhor caso:

Melhor Caso - Teste ANOVA não apontar diferenças: o custo seria 324-10 = 314 x 25000 = US$7850000

Pior Caso - Teste ANOVA apontar diferenças e então ser realizado o teste de Dunnetts: custo seria = custo do teste ANOVA (US$7850000) + custo amostras adicionais. Então teriamos 7850000+(751*25000) = US$26625000.

Alternativa 2

Uma outra possível maneira proposta para o experimento foi baseada no trabalho apresentado em [1] e na estimativa da variância e desvio padrão das amostras, com base nas 10 amostras obtidas gratuitamente.

Em [1], Bechhofer e Tamhane determinam os tamanhos ótimos do grupo de controle n_0 e dos demais grupos n_i's para um experimento de multicomparações não balanceado com base nos valores de δ^*, α e σ_i^2 (variância dos i grupos). De acordo com [1], para determinação de n_0 e n_i, primeiramente calcula-se o parâmetro θ, dado por:

$$\theta = \frac{\sum_{i=1}^{p} \sigma_i^2}{\sigma_0^2}$$

onde:

p: número de grupos ou fatores, exceto o grupo de controle

σ_0^2: variância do grupo de controle

σ_i^2: variância do *i-ésimo* grupo

Com o valor de θ, α e p encontram-se dois valores tabelados $\hat{y_0}$ e $\hat{\lambda}$. O menor tamanho amostral total requerido $\hat{N}$ é dado pelo menor inteiro maior que:

$$(\frac{\hat{\lambda}\sigma_0}{\delta^*})^2$$

O menor tamanho amostral para o grupo de controle $\hat{N}_0$ é dado por:

$$\hat{N}_0 = \hat{y_0}\hat{N}$$

O menor tamanho amostral para cada um dos demais grupos é dado por:

$$\hat{N}_i = \frac{(\hat{N} - \hat{N}_0)\sigma_i^2}{\theta\sigma_0^2}$$

Para os cálculos acima, os valores das variâncias para os grupos são iguais, como é suposto no sumário. Os valores tabelados de $\hat{y_0}$ e $\hat{\lambda}$ são aqueles considerados para o testes do tipo "one-sided", devido à questão de interesse do problema.

Os cálculos da alocação ótima de acordo com [1] são mostrados a seguir:

```
theta<-v*3/v
yzerohat<-0.335
lambdahat<-4.819
Nhat<-ceiling(((lambdahat*sd)/deltaestrela)^2)
N0hat<-ceiling(yzerohat*Nhat)
Nihat<-ceiling((Nhat-N0hat)/3)
print(Nhat)
```

```
## [1] 150
```

```
print(N0hat)
```

```
## [1] 51
```

```
print(Nihat)
```

```
## [1] 33
```

Os valores encontrados para $\hat{N}$, $\hat{N}_0$ e $\hat{N}_i$ são iguais a 150, 51 e 33. Assim, seriam obtidas 51 amostras do *Raiser 1* e 33 amostras dos *Raisers* 2, 3 e 4. Entretanto, deve-se verificar se com este tamanho amostral garante-se a potência definida inicialmente (0.8).

Como em [1] os autores não apresentam a relação da potência do teste com a metodologia, para determinar de forma aproximada o tamanho amostral dos grupos para a potência desejada, foi realizado o cálculo do tamanho amostral requerido para o teste T "one-sided", com o valor de α devidamente corrigido (considerando que seriam executados 3 teste T pareados).

O cálculo é o seguinte:

```
ss<-power.t.test(delta=.25,sd=sd,sig.level=0.03,power=.8,
                      type="two.sample",alternative="one.sided")

print(ss)
```

```
##
##      Two-sample t test power calculation
##
##               n = 96.43327
```

```
##           delta = 0.25
##              sd = 0.6346925
##       sig.level = 0.03
##           power = 0.8
##     alternative = one.sided
##
## NOTE: n is number in *each* group
```

O valor de n é igual a 97 para cada grupo. Considerando 4 *Raisers*, os valores previamente encontrados para $\hat{N}$, $\hat{N}_0$ e $\hat{N}_i$ não são suficientes. Mas podemos recalcular estes valores. Fazendo N = 97*3=291 = $\hat{N}_i$, temos que: $\hat{N}$-$\hat{N}_0$ = 291 e $\hat{N}_0 = \hat{y_0}\hat{N} = 0.335\hat{N}$. Então, $\hat{N}$-0.335$\hat{N}$=291, o que significa que $\hat{N}$=438 e $\hat{N}_0$=147. Então teríamos, o grupo de controle ou *Raiser 1* com 147 amostras e os demais grupos ou *Raisers* com 97 amostras. Como o número total de amostras para o ANOVA já foi calculado previamente e é igual a 324 amostras, para o teste desbalanceado apresentado nesta alternativa teríamos 114 amostras adicionais. Pode-se então propor o custo da *Alternativa 2* baseado no melhor caso e no pior caso:

Melhor Caso - Teste ANOVA não apontar diferenças: o custo seria 324-10 = 314 x 25000 = US$7850000

Pior Caso - Teste ANOVA apontar diferenças e então ser realizado o teste de Dunnetts: o custo seria igual ao custo do ANOVA = 7850000+ custo das amostras adicionais = 7850000+114*25000 = US$10700000.

Alternativa Escolhida

Como a alternativa 2 apresenta menor custo no pior caso, ela foi utilizada. Primeiramente, será feito o teste ANOVA com as 324 amostras. De acordo com os resultados do ANOVA, será analisada a necessidade ou não de execução do teste de Dunnetts.

Preparação e Exploração dos Dados

A figura 1 mostra a variação nas médias dos tempos de vida até falha dos 4 *Risers*. O gráfico sugere variações muito próximas, indicando a princípio a não existência significativa na diferença de tempo médio até falha entre os *Risers*.

```r
sample<-read.table("1978-06-12_81_81_81_81.csv",header=TRUE,sep=",")

summary(sample)
```

```
##     Riser        LogTTF
##  Riser1:81   Min.   : 7.501
##  Riser2:81   1st Qu.: 8.664
##  Riser3:81   Median : 9.036
##  Riser4:81   Mean   : 9.021
##              3rd Qu.: 9.371
##              Max.   :10.479
```

Teste estatístico ANOVA

```r
model <- aov(LogTTF~Riser,
          data = sample)
summary.aov(model)
```

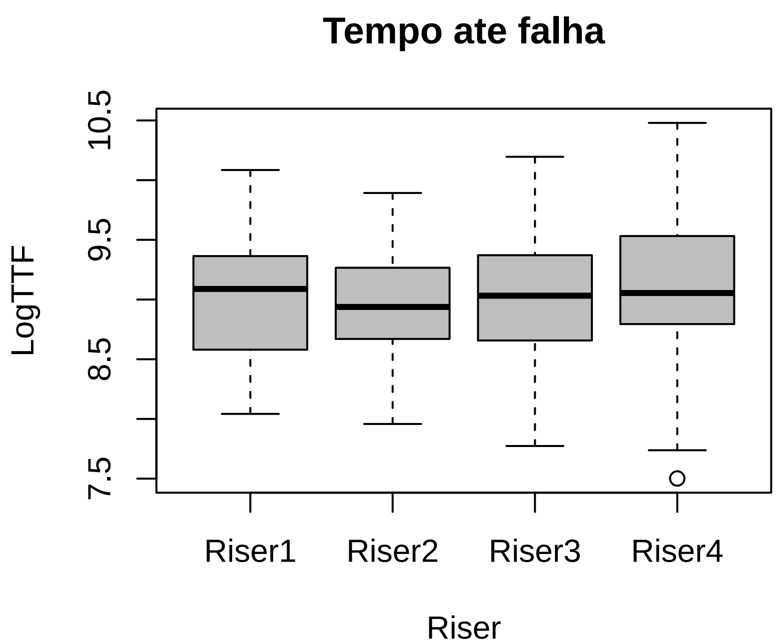

Figure 1: variação do tempo de vida dos Raisers

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Riser         3   1.16  0.3872   1.462  0.225
## Residuals   320  84.76  0.2649
```

De acordo com o resultado do teste estatístico ANOVA, o valor determinado (F Value = 1.462) é menor que o valor crítico da Distribuição F com 3 graus de liberdade no denominador, 320 graus de liberdade no denominador e α=0.1 (F critico = 2.08). Além disso, o valor de P encontrado (P-*Value* = 0.225) > α, nos dão evidências suficientes para a não rejeição da hipótese nula. Como não rejeitamos a hipótese nula, não serão executados os testes de Dunnetts.

Checando as premissas do modelo

Normalidade dos dados

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.99575, p-value = 0.5287
```

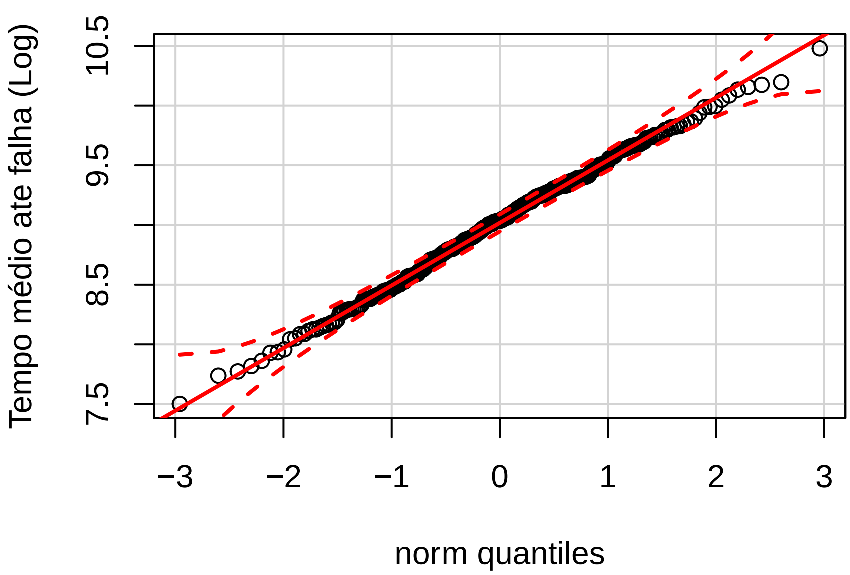

Figure 2: Distribuição dos dados

Analisando a figura 2 percebe-se que a distribuição dos dados, embora seja do tipo Lognormal, apresenta-se como uma distribuição normal. Quanto maior o tamanho amostral, mais a distribuição Lognormal se aproxima da distribuição normal. Os valores do teste de Shapiro-Wilk (W=0.99575 e *P-Value*=0.5278) confirmam a normalidade representada graficamente.

Homocedasticidade

```
##
##  Fligner-Killeen test of homogeneity of variances
##
## data:  LogTTF by Riser
## Fligner-Killeen:med chi-squared = 6.7176, df = 3, p-value =
## 0.08147
```

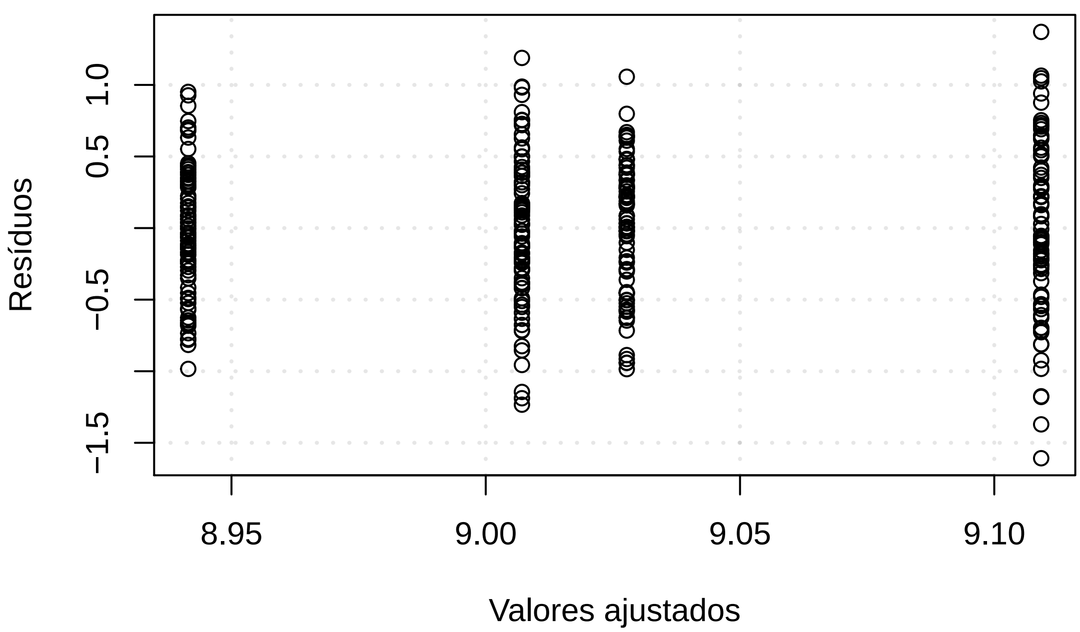

Figure 3: Análise das variâncias

Analisando a figura 3, não se apresenta diferenças significativas na variância dos resíduos, para todos os *Raisers*. O teste de Fligner-Killen foi executado e seu valor (*P-value*=0.08147) confirmam a homocedasticidade (igualdade de variâncias).

Independência dos dados

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.02266952      1.954075    0.54
##  Alternative hypothesis: rho != 0
```

Analisando a figura 4 e os resultados do teste de Durbin-Watson (DW=1.954075 e *P-Value*=0.57), podemos afirmar que os dados apresentam independência.

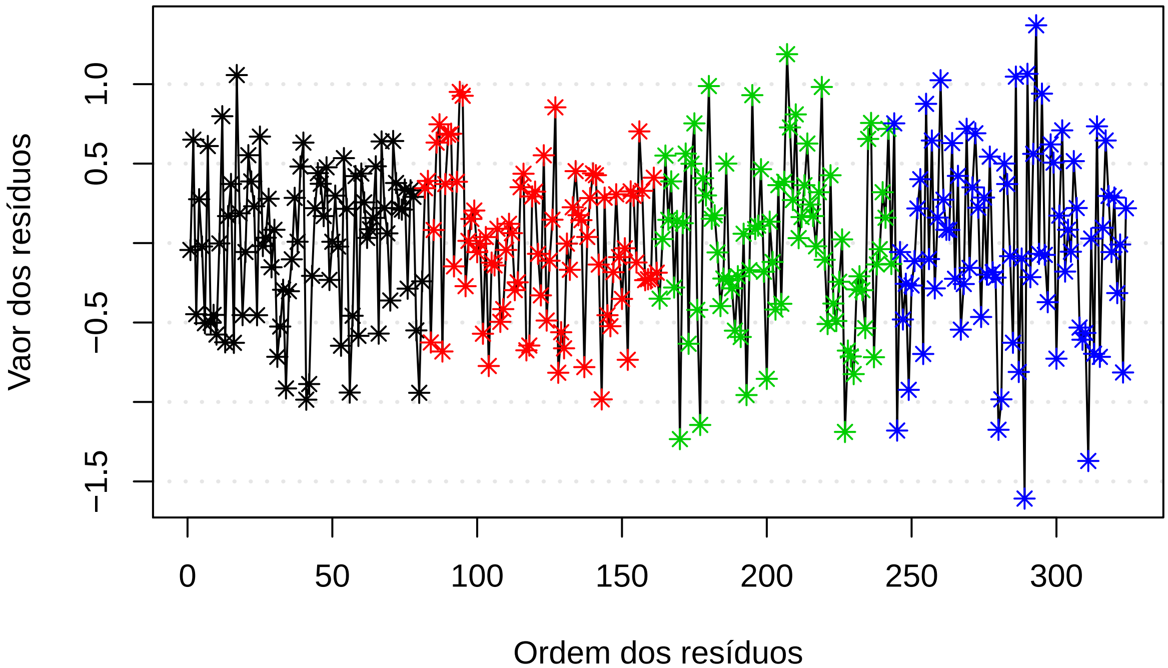

Figure 4: Análise de independêcia dos dados

Conclusões e Recomendações

De acordo com o experimento realizado, não há evidências que apontem alguma (pelo menos uma) configuração de *Riser* melhor (com tempo de vida até falha maior) que a configuração padrão (*Riser1*).

Diante de diferentes alternativas de experimentação, foi escolhida a mais barata, que não precisou ser executada por completo. Somente o teste estatístico ANOVA foi executado e foi suficiente para não rejeitar a hipótese nula (H_0). O custo do experimento foi de US$7850000 (Sete Milhões, oitocentos e cinquenta mil dólares).

Na elaboração das duas alternativas, optou-se por utilizar um valor médio estimado para a variância e desvio padrão, numa tentativa de não ser demasiadamente conservador nem demasiadamente otimista com relação à estes fatores.

Acredita-se que uma melhor investigação sobre o comportamento da potência estatística em testes multicomparativos aliada à alocação ótima apresentada em [1] pode apresentar resultados ainda mais satisfatórios (com necessidade de um menor tamanho amostral e que obedeça aos requisitos de potência).

Referências Bibliográficas

[1] Bechhofer, E. R; Tamhane, A.C. *Design of Experiments for Comparing Treatments With a Control: Tables*

of Optimal Allocations of Observations. TECHNOMETRICS, 25, nro. 1, 87-95.