

Estudo de caso: Grupo D 3

Gilmar and Maressa Nunes R. Tavares and Victor

3 de Setembro, 2019

1 Summary

O presente trabalho tem realizado o delineamento e executou os testes estatísticos para avaliar uma nova versão de um software, em relação aos resultados obtidos na versão anterior. Tendo em vista que a última versão possui uma distribuição de custos com média $\mu = 50$ e variância $\sigma = 100$, dados da população, objetiva-se verificar se a nova versão apresenta resultados melhores para tais características. Para tanto, utilizou-se o teste z com nível de significância $\alpha = 0,01$ e $\alpha = 0,05$, para os testes de média e variância, respectivamente. Após os testes verificou-se que....

2 Planejamento do experimento

Nesta seção são apresentados os detalhes do delineamento dos testes que foram executados para comparar o desempenho das duas versões do software em relação à média e à variância do custo de execução. Essa etapa é fundamental, pois trata-se de uma abordagem que fornece resultados importantes em análises de sistemas complexos, além disso, os testes servem para validar a teoria que está por trás dele ???.

2.1 Objetivo do experimento

Para a versão atual de um dado sistema, sabe-se que sua distribuição de custos de execução possui média populacional de $\mu = 50$ e variância $\sigma^2 = 100$. Uma nova versão desse software foi desenvolvida, portanto realizou-se uma análise estatística para investigar os ganhos de desempenho obtidos em relação à versão atual.

Inicialmente, o teste foi executado para as médias do custo, assim, para verificar se a nova versão é melhor que a anterior, formulou-se as seguintes hipóteses:

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu < 50 \end{cases}$$

Como a média da população para a primeira versão é $\mu = 50$, considerou-se como hipótese nula (H_0) a ausência de melhoria do software, isto é, a segunda versão apresenta a mesma performance da versão anterior, com média igual $\mu = 50$. Por outro lado, a hipótese alternativa considera que houve melhorias entre as versões, portanto, a média é menor que 50 (H_1).

Além disso, para o teste da média foram definidos os seguintes objetivos: nível de significância $\alpha = 0.01$; nível de confiança $1 - \alpha = 0.99$; o tamanho mínimo de efeito $\delta^* = 4$; e a potência desejada $\pi = 1 - \beta = 0.8$

Por outro lado, para a variância o experimento foi realizado com base nas seguintes hipóteses:

$$\begin{cases} H_0 : \sigma^2 = 100 \\ H_1 : \sigma^2 < 100 \end{cases}$$

Assim como no teste da média, neste caso adotou-se como hipótese nula (H_0) a ausência de melhoria do software, mantendo os resultados de variância da versão anterior ($\sigma^2 = 100$). Enquanto a hipótese alternativa considera que houve melhorias entre as versões, portanto, a variância é menor que 100 (H_1).

Em relação aos objetivos, o teste da variância considerou: $\alpha = 0.05$ e $1 - \alpha = 0.95$

Os dois testes foram realizados com os mesmos dados coletados os dados de acordo com a descrição da próxima seção.

2.1.1 Descrição da coleta de dados

Para coletar os dados referente à nova versão do software, foi executada uma simulação no software R utilizando a biblioteca *ExpDE* ????. A coleta de dados foi declarada da seguinte forma:

```
# Set-up the data generating procedure
mre <- list(name = "recombination_bin", cr = 0.9)
mmu <- list(name = "mutation_rand", f = 2)
mpo <- 100
mse <- list(name = "selection_standard")
mst <- list(names = "stop_maxeval", maxevals = 10000)
mpr <- list(name = "sphere", xmin = -seq(1, 20), xmax = 20 + 5 * seq(5, 24))

set.seed(1234) # to generate always the same results

# define functions for data generation
get.single.sample <- function(mpo, mmu, mre, mse, mst, mpr){
  generator <- ExpDE(mpo, mmu, mre, mse, mst, mpr, showpars = list(show.its = "none"))
  return(generator$Fbest)
}

get.n.samples <- function(mpo, mmu, mre, mse, mst, mpr, N){
  my.sample <- numeric(N)
  for (i in seq(N)){
    my.sample[i] <- get.single.sample(mpo, mmu, mre, mse, mst, mpr)
  }
  return(my.sample)
}
```

As funções `get.single.sample` e `get.n.samples` foram criadas para facilitar o entendimento da função de geração de dados, sendo elas para coletar uma única amostra ou n amostras, respectivamente.

Com base nas amostras coletadas referente à segunda versão do software, foi realizada uma análise exploratória dos dados a fim de validar as premissas dos testes que foram realizados em seguida, conforme apresentado na seção seguinte.

3 Análise Exploratória dos Dados

Antes de proceder com as análises estatísticas e realizar as inferências sobre o problema, é importante realizar uma análise preliminar dos dados ????. destaca que a análise exploratória tem o papel de extrair informações dos dados antes de realizar as inferências estatísticas, a fim de obter os modelos plausíveis para cada estudo.

Para a realização das análises foi feita uma amostra de 30 elementos da população, pois essa quantidade já é suficiente boa para amostrar o comportamento dos dados ???.

```
N <- 30

my.sample <- numeric(N)
for (i in seq(N)){
  my.sample[i] <- ExpDE(mpo, mmu, mre, mse, mst, mpr,
    showpars = list(show.its = "none"))$Fbest
}
```

```
summary(my.sample)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   39.75  45.25   47.59   48.67  52.36   65.08
```

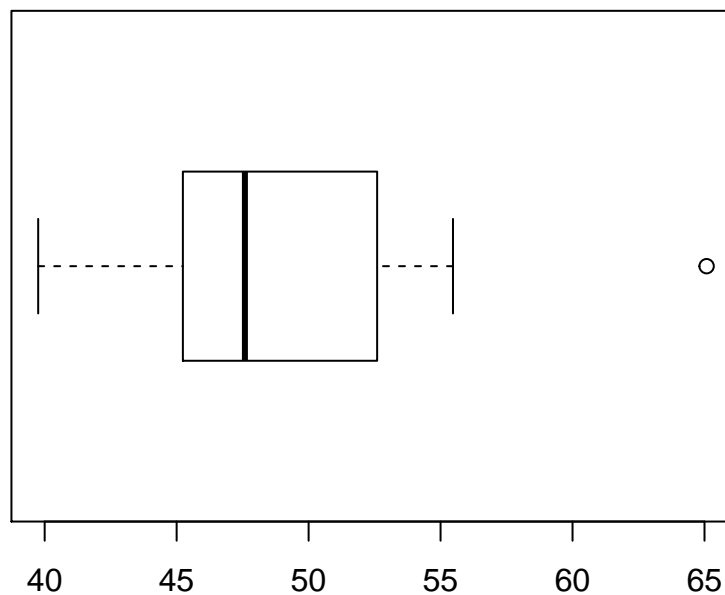
```
#var(my.sample)
```

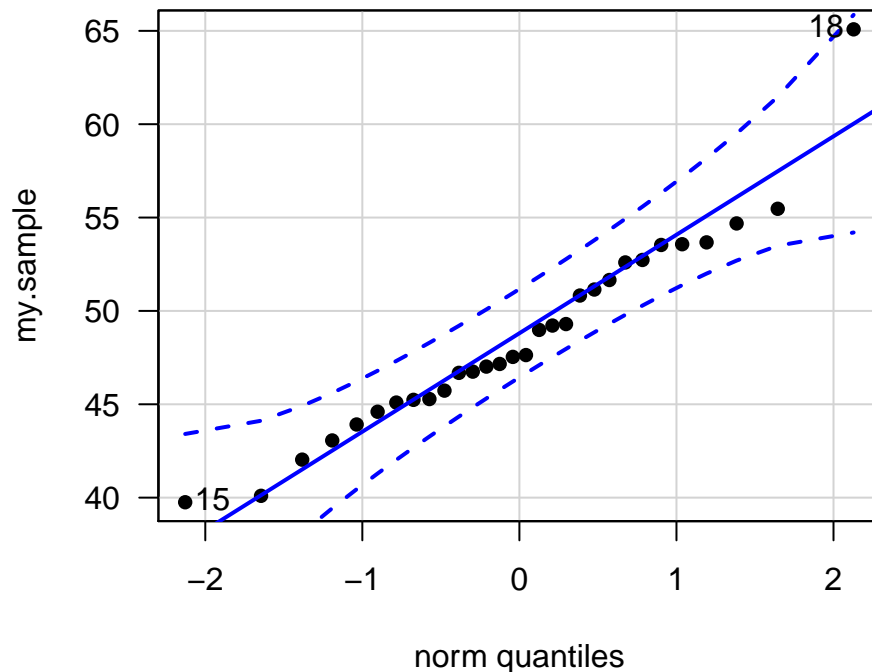
```
var(my.sample)
```

```
## [1] 27.72722
```

Pela análise dos dados verifica-se que a variância do novo software aparenta ser significamente menor que a versão anterior, enquanto em relação à média a diferença foi mais discreta. Para concluir com maior precisão sobre os dados, os testes estatísticos serão apresentados com detalhes nas seções seguintes.

No contexto da análise exploratória, é válido realizar o teste de outliers, através de um boxplot, e o teste da normalidade dos dados por meio do gráfico quantil x quantil.





```
## [1] 18 15
```

Pela análise do boxplot e do gráfico quantil X quantil, verifica-se que, embora o valor máximo observado (65.08) apresente-se como um outlier no boxplot, ele não interfere na normalidade dos dados. Sendo assim, é possível prosseguir com as análises estatísticas que serão apresentadas nas seções a seguir.

4 Resultados

4.1 Validação das premissas

4.2 Teste do custo médio

4.2.1 Teste da variância do custo

5 Análise Estatística

Nesta seção é realizado a análise estatística dos dados referentes ao experimento realizado no segundo software. Os testes realizados visam a testar as hipóteses nula e alternativa, descritas anteriormente, utilizando o teste t para o teste do custo médio e para o teste da variância.

5.1 Teste sobre a média do custo

Antes de realizar o teste sobre o custo médio realizou-se o cálculo do tamanho amostral baseado nos objetivos descritos na seção 2.1 e definidos da seguinte forma:

```
#Variaveis:
alpha<-0.01 #nivel de significancia
```

```

delta<-4      #tamanho minimo do efeito relevante
pi<-0.8       #potencia desejada
beta<-1-pi

#constantes atuais:
media_a<-50
variancia_a<-100

```

5.1.1 Cálculo do tamanho amostral

Para o cálculo do tamanho amostral foi utilizado a função `power.t.test` com os parâmetros de entrada sendo o tamanho mínimo de efeito δ^* desejado, a variância, o nível de significância α e a potência π , conforme mostrado abaixo.

```

#Calculo do tamanho amostral usando a função power.t.test:
sample_size <- power.t.test(delta = delta,
                             sd = sqrt(variancia_a),
                             sig.level = alpha,
                             power = pi,
                             alternative = "one.sided",
                             type = "one.sample")

#Tamanho amostral:
N <- ceiling(sample_size$n)

```

A função retornou o tamanho amostral de 66 unidades.

5.1.2 Teste de Hipoteses

Para testar a hipótese nula para o custo médio utilizou-se a função `t.test`, sobre os dados amostrados usando a função `get.n.samples`, descrita na seção 2.1.1, para N amostras. O teste está descrito a seguir.

```

#Gerando os dados:
my_samples<-get.n.samples(mpo,mmu,mre,mse,mst,mpr,N)

#teste para custo médio:
my_teste<-t.test(my_samples,
                  mu=media_a,                #hipotese nula
                  alternative = "less",      #hipotese alternativa
                  conf.level = 1-alpha)

```

O teste da média resultou em:

```

my_teste

##
##  One Sample t-test
##
## data:  my_samples
## t = -1.793, df = 65, p-value = 0.03882
## alternative hypothesis: true mean is less than 50
## 99 percent confidence interval:
##      -Inf 50.42078
## sample estimates:
## mean of x
##  48.7259

```

5.1.3 Cálculo do intervalo de confiança

Para o cálculo do intervalo de confiança utilizou-se a função `t.test` para o teste da hipótese nula, conforme mostrado a seguir.

```
intervalo<-t.test(my_samples,
                  mu = media_a,
                  conf.level=1-alpha)$conf.int
```

A função resultou no intervalo:

```
intervalo

## [1] 46.84021 50.61158
## attr(,"conf.level")
## [1] 0.99
```

5.1.4 Validação das premissas

5.2 Teste sobre a variância do custo

5.2.1 Teste de Hipóteses

5.2.2 Cálculo do intervalo de confiança

5.2.3 Validação das premissas

6 Discussão e Conclusões

7 Divisão das Atividades

Victor - Reporter Maressa - Coordenadora Gilmar - Verificador e Monitor

8 Referências