

Planejamento e Análise de Experimentos (EEE933)

Estudo de Caso 1

Pedro Vinícius, Samara Silva e Savio Vieira

10 de Agosto de 2020

Introdução

Uma versão atual de um software conhecido apresenta uma distribuição dos custos de execução com média populacional $\mu_c = 50$ e variância populacional $\sigma_c^2 = 100$. Posteriormente, uma nova versão do software é desenvolvida, no qual deseja-se investigar prováveis melhorias de desempenho. Com esse intuito, duas análises estatísticas são propostas: (i) uma sobre o custo médio e (ii) uma sobre a variância do custo.

Para ambos os casos, as hipóteses nulas foram definidas de maneira conservadora, partindo-se do pressuposto de que os parâmetros populacionais conhecidos foram mantidos na nova versão. A partir disso, diversas etapas foram conduzidas até a conclusão dos experimentos, entre elas a coleta de dados da distribuição dos custos do software novo, a análise exploratória dessa amostra, a inferência por meio dos testes estatísticos, a validação das premissas consideradas e as conclusões. As próximas seções contém o detalhamento técnico de cada uma dessas etapas.

Parte 1: Teste Sobre o Custo Médio

Planejamento dos Experimentos

No que se refere a primeira parte do estudo de caso, o teste terá que dispor de um nível de significância $\alpha = 0.01$, um tamanho de efeito de mínima relevância $\delta^* = 4$ e uma potência desejada $\pi = 1 - \beta = 0.8$. As hipóteses estatísticas foram definidas com o intuito de responder às questões propostas abaixo:

- Há alguma diferença entre o custo médio da versão nova do software e o custo médio da versão corrente?
- Caso haja, qual a melhor versão em termos de custo médio?

Em concordância com a proposta de comparação de custo médio entre as versões, as hipóteses de teste podem ser formuladas sobre o parâmetro média:

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

onde a hipótese nula implica na igualdade entre os custos médios das versões e a hipótese alternativa unilateral na superioridade da nova versão em média.

A fase subsequente desse experimento consiste em gerar uma amostra representativa do desempenho da nova versão do software. Para tal fim, é necessário especificar o tamanho dessa amostra, considerando as propriedades preestabelecidas do teste. A priori, o Poder do Teste é bastante conveniente, porém implica em um grande dilema. O cálculo do tamanho amostral requer uma estimativa da variância, que só é obtida através das observações contidas na amostra. As possibilidades mais práticas de se conduzir o experimento nesse caso são[1]:

1. Utilização de conhecimento do problema para se obter uma estimativa (inicial) da variância;
2. Condução do estudo com um tamanho amostral predefinido, como $N = 30$, o que poderia violar a potência desejada;

3. Realização de um estudo piloto para estimar a variância dos dados a partir do tamanho de efeito de mínima relevância δ^* .

Considerando as vantagens e desvantagens de cada uma, optou-se por utilizar a primeira abordagem. Por mais que essa estimativa seja sobre-estimada e os prováveis ganhos não sejam observados ao término do estudo, uma vez que se espera ganhos de variância da nova versão do software em relação à versão atual, pode-se considerar igualdade de variâncias como uma estimativa inicial, ou seja, $\sigma_n^2 \approx \sigma_c^2 = 100$. Entretanto, essa premissa será avaliada posteriormente na análise exploratória dos dados.

Diante da estimativa inicial da variância amostral, o Poder do Teste pode ser finalmente realizado. Esse teste é originalmente usado para mensurar o controle do teste de hipóteses sobre o erro do tipo II (β), isto é, $P(\text{rejeitar } H_0 | H_0 \text{ é falso})$. No entanto, tal teste também pode ser utilizado para estimar outros parâmetros amostrais, como tamanho de efeito δ^* , nível de significância α , tamanho da amostra N , potência π e desvio padrão amostral σ_n [2].

```
# Define the sample size to be used in this experiment
(params <- power.t.test(delta = delta_star,
  sd = sigma_n,
  sig.level = alpha,
  power = pi,
  type = "one.sample",
  alternative = "one.sided"))
```

```
##
##      One-sample t test power calculation
##
##              n = 65.45847
##              delta = 4
##              sd = 10
##      sig.level = 0.01
##              power = 0.8
##      alternative = one.sided
```

```
# Number of observations
n <- ceiling(params$n)
```

Assim, tem-se uma estimativa de $N = 66$ observações.

Coleta dos Dados

```
data_generation <- function(n){

  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))

  sample <- c()
  # Generate n observations
  for (i in 1:n){
    observation <- ExpDE(mpo, mmu, mre, mse, mst, mpr,
      showpars = list(show.iters = "none"))$Fbest
```

```

    sample <- c(sample, observation)
  }

  return(sample)
}

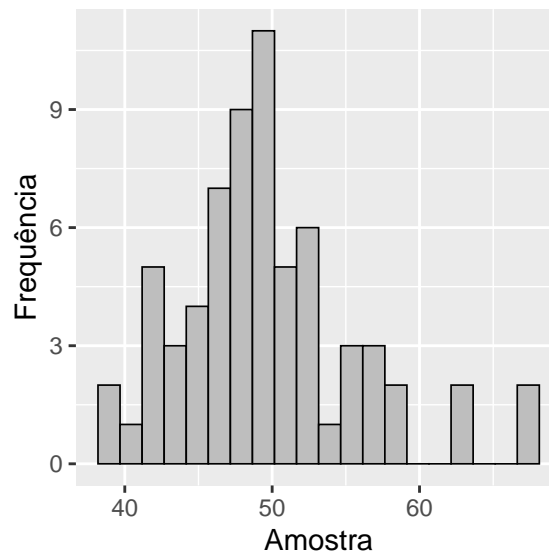
```

Análise Exploratória de Dados

```

# Histogram
histogram <- ggplot(data = as.data.frame(sample), mapping = aes(x = sample))
histogram + geom_histogram(lwd = 0.3, bins = 20, color = 'black', fill = 'gray') +
  scale_x_continuous(name = 'Amostra') +
  scale_y_continuous(name = 'Frequência')

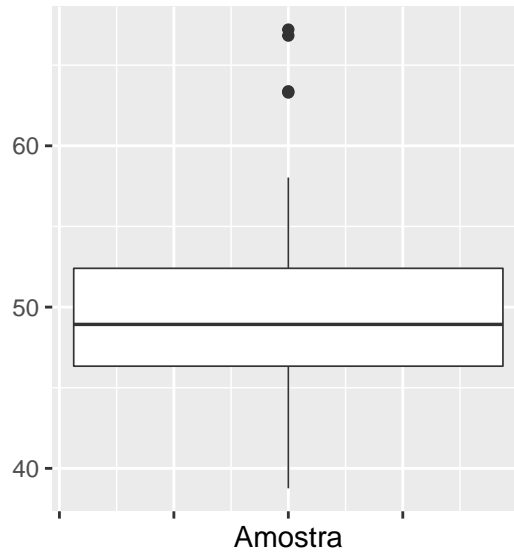
```



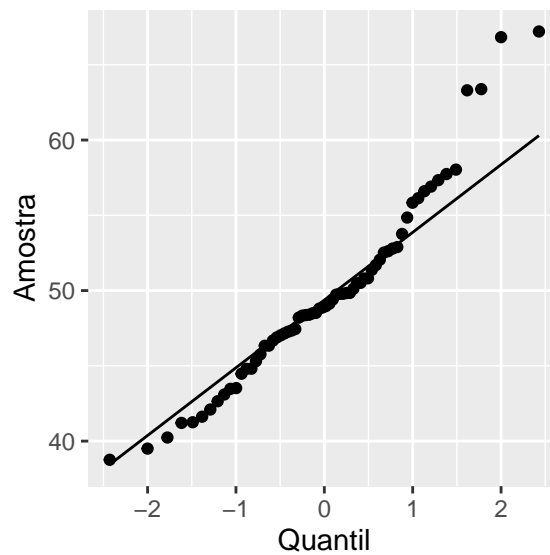
```

# Boxplot
boxplot <- ggplot(data = as.data.frame(sample), mapping = aes(y = sample))
boxplot + geom_boxplot(lwd = 0.3) +
  scale_x_continuous(name = 'Amostra') +
  scale_y_continuous(name = '') +
  theme(axis.text.x = element_blank())

```



```
# QQ-Plot
qqplot <- ggplot(data = as.data.frame(sample), mapping = aes(sample = sample))
qqplot + geom_qq_line() +
  geom_qq() +
  scale_y_continuous(name = 'Amostra') +
  scale_x_continuous(name = 'Quantil')
```



Análise Estatística

```
# ----- Hypothesis Testing ----- #
(t_test <- t.test(x = sample,
  mu = mu_c,
  alternative = "less",
  conf.level = conf_level))
```

```
##
## One Sample t-test
##
## data: sample
## t = -0.4853, df = 65, p-value = 0.3145
## alternative hypothesis: true mean is less than 50
## 99 percent confidence interval:
##      -Inf 51.4154
## sample estimates:
## mean of x
## 49.63844
```

```
# Confidence Interval
CI <- t_test$conf.int[1:2]
```

Validação de Premissas

```
# Wilcoxon Signed-Ranks Test, because it does not assume normality
(wilcoxon_test <- wilcox.test(x = sample,
                             alternative = "less",
                             mu = mu_c,
                             paired = FALSE,
                             exact = NULL,
                             correct = TRUE,
                             conf.int = FALSE,
                             conf.level = conf_level))
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: sample
## V = 903, p-value = 0.09846
## alternative hypothesis: true location is less than 50
```

Parte 2: Teste Sobre a Variância do Custo

Planejamento dos Experimentos

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

Conclusões

Referências

- [1] Felipe Campelo. Lecture Notes on Design and Analysis of Experiments. <http://git.io/v3Kh8>, 2018. Version 2.12; Creative Commons BY-NC-SA 4.0.
- [2] R Development Core Team. Power Calculations For One And Two Sample T Tests. <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/power.t.test>, 2020. Documentation reproduced from package stats, version 3.6.2, License: Part of R 3.6.2.