Mestrado Profissional em Avaliação e
Monitoramento de Políticas Públicas

Métodos Quantitativos I

Aula 3: Testando hipóteses

Professores: Daniel Grimaldi e Arthur Bragança

3º Trimestre - 2025

Conceitos básicos

Inferência estatística

Teste de hipótese é o procedimento padrão para inferência estatística

- Na prática, ele é um procedimento que busca combinar a informação contida em um conjunto de dados observados com nosso conhecimento estatístico a respeito de distribuições téoricas (como as que vimos na aula anterior) para aprendermos sobre a realidade.
    - Existem duas abordagens clássicas para inferência estatística: clássica e bayesiana
    - Nossa aula se concentrará sobre a abordagem clássica - que se baseia na idéia de identificar alguns parâmetros populacionais verdadeiros.

Construindo a hipótese

- O primeiro passo é construir uma hipótese que seja **falseável sob uma perspectiva estatística**
- Na abordagem clássica, isso significa três coisas:
    i. Definir um parâmetro (ou conjunto de parâmetros) de interesse;
    ii. Definir uma hipótese a respeito do valor assumido pelo(s) parâmetro(s); e
    iii. Definir uma hipótese a respeito da distribuição probabilistica desse(s) parâmetro(s)
- O item ii. estabelece uma **hipótese nula (H_0)**, e o item iii. me permite testá-la.

Testando a hipótese

- Após a definição da hipótese nula, podemos (por conta de iii.) definir uma **região crítica** para o parâmetro de interesse
  - i.e: um intervalo de valores que, se observados empiricamente, tornam a nossa H_0 altamente improvável
- O teste, então, passa ser conseguir uma amostra de dados que nos permita estimar o nosso(s) parâmetro(s)
  - se o valor estimado estiver na região crítica, então devemos a **rejeitar nossa** H_O

Exemplo: Renda informal de beneficiários do PBV

- Quero testar se a renda média informal dos beneficiários do Programa Bolsa Verde (μ_y) é igual a um determinado valor (μ_0): $H_0 : \mu_y = \mu_0$
- Consigo, então uma amostra $(Y_1, Y_2, ..., Y_n)$ para a renda dos beneficiários do PBF
- Se H_0 estiver correta, pelo TLC, vale que em grandes amostras:

$$\sqrt{n}\,\frac{\bar{\mu}_y - \mu_0}{\hat{\sigma}_{\bar{Y}}} \sim N(0,1) \tag{1}$$

Região crítica

- A estatística apresentada em (1) é também conhecida como **estatística *t***. Ela é uma das mais utilizadas em econometria. Se $t \sim N(0,1)$, então:
  - $P(-1,96 \leq t \leq 1,96) = 95\%$
  - $P(-2,576 \leq t \leq 2,576) = 99\%$
- Ou seja, conforme o módulo da estatística t se afasta de 0, nossa H_0 se torna cada vez mais improvável.
  - A partir de um certo ponto, t se torna tão improvável que a única conclusão possível é que H_0 deve ser rejeitada
  - Chamamos de **região crítica** esse conjunto de valores que, quando observados, nos fazem rejeitar H_0

Erro tipo I

- Note que valor de t dentro da região crítica (qualquer que seja ela), não significa que H_0 é impossível, apenas que é improvável
- Portanto, sempre existe a possibilidade de rejeitarmos H_0 mesmo sendo ela verdadeira
- Chamamos essa possibilidade de Erro Tipo I, e podemos atribuir uma probabilidade a ele

$$P(\text{rejeitar } H_0 | H_0 \text{ é verdadeira}) = \alpha$$

Nível de significância

- Imagine que $|t| \geq 1,96$ caracteriza a região crítica para o nosso teste. Então vale que:

$$
\begin{aligned}
\alpha &= P(|t| \geq 1,96 \mid \mu_y = \mu) \\
&= 1 - P(|t| \leq 1,96 \mid \mu_y = \mu) \\
&= 5\%
\end{aligned}
\tag{2}
$$

- α é também chamado de **nível de significância** (ou tamanho) do teste.
  - **equivale ao grau de tolerância ao Erro do tipo I**

Intervalo de confiança

- O conjunto de valores para os quais "toleramos" H_0 é o complementar da região crítica
- Esse intervalo dentro do qual confiamos em H_0 é denominado intervalo de confiança.
  - Ele depende diretamente do nível de significância que estabelecermos para o teste
- No nosso exemplo, dado um nível de significância de 5%, o intervalo de confiança é dado por:

$$[\mu_0 - 1,96\ \hat{\sigma}_{\bar{y}}, \mu_0 + 1,96\ \hat{\sigma}_{\bar{y}}]$$

p-valor

- No nosso exemplo, suponha que tenhamos observado um valor t_i para a estatística t.
- Ao invés de definirmos diretamente um nível de significância, podemos calcular a **probabilidade de observarmos um valor igual ou mais extremo do t_i sendo H_0 verdadeira**. Chamamos essa probabilidade de **p-valor**
- Em um teste bicaudal, isso equivale a calcularmos $P(|t| \geq t_i)$

Erro Tipo II

- É também verdade que observar valores "toleráveis" para t, não quer dizer que H_0 seja verdadeira.
    - Outras hipóteses diferentes de H_0 também podem produzir valores de t dentro da **região de aceitação de** H_0
- Por isso, existe também a possibilidade de não rejeitarmos H_0, mesmo sendo H_0 falsa.
    - só pode ser calculado para uma hipótese alternativa definida
- Chamamos essa possibilidade de Erro Tipo II, cuja probibilidade é dada por:

$$P(\text{aceitar } H_0 | H_1 \text{ é verdadeira}) = \beta(H_1)$$

Poder do teste

- Chamamos de **poder do teste** a probabilidade de rejeitar H_0 quando H_0 é falsa. É equivalente a $1 - \beta(H_1)$
  - Teste não viesado: $1 - \beta \geq \alpha \; \forall \; \mu_0$;
  - Teste consistente: $\lim(1 - \beta)_{n \to \infty} = 1$
- Como regra geral, um teste vai ser não-viesado (consistente) se estiver baseado em um estimador não-viesado (consistente)

Estimador não-viesado

- $\hat{\theta}$ é não-viesado $\Leftrightarrow E(\hat{\theta}) = \theta$. Isso equivale dizer que o valor esperado do estimador equivale ao valor do parâmetro que se deseja estimar.
- Por exemplo, a média amostral $(\bar{Y})$ é um estimador não-viesado da média populacional (μ_y).

$$\begin{aligned} E(\bar{Y}) &= E(\frac{Y_1 + Y_2 + ... + Y_N}{N}) \\ &= \frac{E(Y_1 + Y_2 + ... + Y_N)}{N} \\ &= \frac{N\mu_y}{N} \end{aligned} \tag{3}$$

Estimador consistente

- $\hat{\theta}$ é consistente $\Leftrightarrow \hat{\theta} \xrightarrow{p} \theta$. Isso equivale dizer que a probabilidade do valor estimado diferir do parâmetro que se deseja estimar se aproxima de 0 conforme n aumenta.
- Por exemplo, a média amostral $(\bar{Y})$ é um estimador consistente da média populacional (μ_y).

$$\frac{1}{N}\sum_{i=1}^{N} y_i - \mu_y \xrightarrow{p} 0 \tag{4}$$

$$\frac{1}{N}\sum_{i=1}^{N} y_i \xrightarrow{p} \mu_y$$

Notas: [1] Falta de viés não implica em consistência. [2] Consistência não implica falta de viés.

Hands On

Serviço Militar Obrigatório

O exemplo de hoje é inspirado em *Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records". The American Economic Review 80 (3): 313–36.*

- Investiga o impacto do serviço militar sobre a renda futura dos indivíduos;
  - Havia debate a respeito de benefícios exagerados aos veteranos americanos...
  - ...mas comparações diretas sofriam com viés de endogeneidade

Contexto do alistamento militar no Brasil

- Há uma autoseleção no serviço militar obrigatório
  - Como existem menos vagas do que alistamentos, preferência é dada por quem deseja servir; e
  - Indivíduos com perspectiva de baixa remuneração na área civil, enxergam no alistamento uma grande oportunidade
- Por conta disso, uma comparação direta entre os que serviram no serviço militar obrigatório e os que não serviram, não permite inferência causal

Desejo de servir

- Soldo equivale a BRL 2.100 e serviço dura 1 ano, sem possibilidade de renovação.
- Salário civil (w_i^0) equivale a $\alpha_i + \epsilon_i^0$, sendo $\epsilon^0 \sim N(0, 800)$
- Existem 2 tipos de indivíduos: tipo A tem $\alpha_i = \alpha_a = 2.500$ e tipo B tem $\alpha_i = \alpha_b = 2.000$.
  - ambos indivíduos ocorrem com igual probabilidade
- Desejo de servir (D_i) depende da expectativa salarial ($E(w_{i0}^0)$), de tal forma que:

$$D_i = \begin{cases} 1 & \text{se } E(w_i^0) < 2.100 \\ 0 & \text{se } c.c. \end{cases}$$

Serviço militar

- Por restrições de vagas, apenas 50% dos que desejam servir conseguem ($P(S_i = 1|D_i = 1) = 0.5$)
  - escolha é aleatória
- δ_i mede o impacto do serviço militar sobre salário civil futuro (w_i^1), de tal forma que:

$$w_i^1 = \delta i \; \alpha_i + \epsilon_i^1 \tag{5}$$

$$\delta_i = \begin{cases} 1.1 & \text{se } S_i = 1 \\ 1 & \text{se } c.c. \end{cases}$$

PGD

```
pgd <- function(N){
  data <- data.frame("id"=paste0("cpf_", 1:N)) %>%
    mutate(alpha = sample(c(2500, 2000),
                          prob=c(0.5, 0.5),
                          replace=TRUE,
                          size=N),
           epsilon_0=rnorm(N, 0, 800),
           D = ifelse(alpha==2000, 1, 0),
           w_0 = alpha + epsilon_0,
           S = case_when(D==1 ~ rbern(N, 0.5),
                         TRUE ~ 0),
           delta = ifelse(S==1, 1.10, 1),
           epsilon_1=rnorm(N, 0, 800),
           w_1 = delta*alpha + epsilon_1
    )
}
```

O que sabemos?

- Dado o PGD, sabemos que:

$$E(w_i^1|S=0) = \frac{E(w_i^1|S=0 \cap \alpha_i = \alpha_a)P(\alpha_i = \alpha_a) + E(w_i^1|S=0 \cap \alpha_i = \alpha_b)P(\alpha_i = \alpha_b)}{P(S=0 \cap \alpha_i = \alpha_b \cup S=0 \cap \alpha_i = \alpha_a)} \tag{6}$$

$$E(w_i^1|S=0) = \frac{2500\ 0.5 + 2000 * 0.25}{0.75}$$

$$E(w_i^1|S=0) = 2333,33$$

$$E(w_i^1|S=1) = \frac{E(w_i^1|S=1 \cap \alpha_i = \alpha_a)P(\alpha_i = \alpha_a) + E(w_i^1|S=0 \cap \alpha_i = \alpha_b)P(\alpha_i = \alpha_b)}{P(S=0 \cap \alpha_i = \alpha_b \cup S=0 \cap \alpha_i = \alpha_a)} \tag{7}$$

$$E(w_i^1|S=1) = \frac{2200 * 0.25}{0.25}$$

$$E(w_i^1|S=1) = 2200$$

Teste de média simples (militares)

- Quero testar se salário civil dos que serviram é igual a BRL 2.500,00 ($H_0 : \mu_{w^1} = 2500$)
- Sob H_0, vale que:

$$\sqrt{n}\frac{\bar{\mu}_{w^1} - 2500}{\hat{\sigma}_{w^1}} \simeq N(0,1)$$

- Consideraremos **5% como nível de significância**

Gerando amostra

```r
set.seed(13)
data <- pgd(2000)
head(data)
```

```
##      id alpha   epsilon_0 D      w_0 S delta   epsilon_1       w_1
## 1 cpf_1  2500 -469.50076 0 2030.499 0   1.0  200.054589 2700.055
## 2 cpf_2  2000  892.29971 1 2892.300 1   1.1  -11.582969 2188.417
## 3 cpf_3  2000  999.73075 1 2999.731 1   1.1   -5.813859 2194.186
## 4 cpf_4  2000  -40.38681 1 1959.613 0   1.0 -221.801747 1778.198
## 5 cpf_5  2500  301.56497 0 2801.565 0   1.0 -112.126609 2387.873
## 6 cpf_6  2000 -972.90867 1 1027.091 0   1.0  548.351847 2548.352
```

Teste-t (comando em R)

```r
data_serviu <- filter(data, S==1)
t.test(data_serviu$w_1, mu=2500, alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  data_serviu$w_1
## t = -7.3809, df = 502, p-value = 6.571e-13
## alternative hypothesis: true mean is not equal to 2500
## 95 percent confidence interval:
##  2148.817 2296.474
## sample estimates:
## mean of x
##  2222.645
```

Teste-t (manualmente)

```
# Estimador amostral do desvio-padrão
sd_serviu <- sd(data_serviu$w_1)
se_hat <- sd_serviu/sqrt(nrow(data_serviu))
se_hat
```

```
## [1] 37.57741
```

```
# Estimador amostral da média
mu_hat <- mean(data_serviu$w_1)
mu_hat
```

```
## [1] 2222.645
```

Teste-t (manualmente)

```r
# estatística t
t <- (mu_hat - 2500)/se_hat
t
```

```
## [1] -7.380891
```

```r
# IC (95%) para média amostral
ic <- c(mu_hat - 1.96*se_hat,
        mu_hat + 1.96*se_hat)
ic
```

```
## [1] 2148.994 2296.297
```

Erro tipo I: cálculo

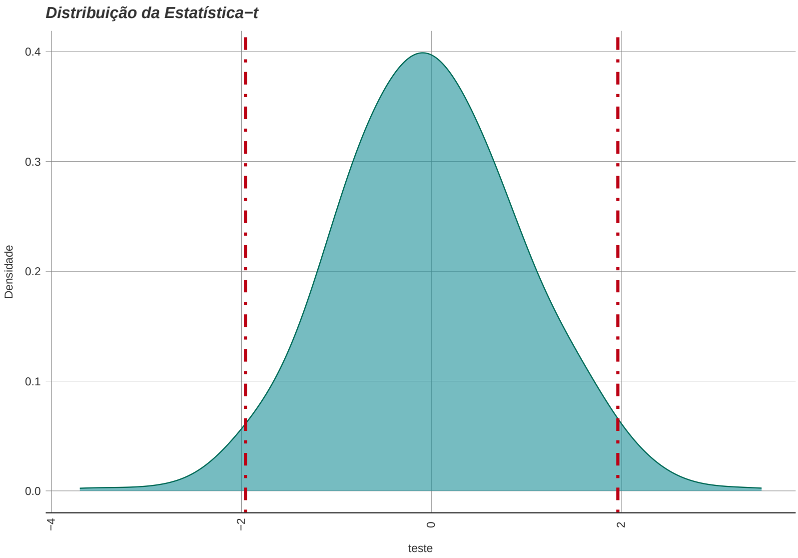
Se usamos 5% como nível de significância, significa que em amostragens repetidas $\approx$ 5% das vezes vamos rejeitar H_0, mesmo sendo ela verdadeira. **Também podemos usar nosso PGD para conferir isso.**

```
gen_t_stat <- function(seed.i, mu.i, Si){
  set.seed(seed.i)
  data_temp <- pgd(1000) %>% filter(S==Si)
  t.test(data_temp$w_1, mu=mu.i, alternative="two.sided")$statistic
}
teste <- sapply(1:1000, gen_t_stat, mu.i=2200, Si=1, simplify = TRUE)
prob_erro_I <- sum(abs(teste)>=1.96)/length(teste)
prob_erro_I
```

```
## [1] 0.044
```

Erro tipo I: visualização

```r
fig <- ggplot() +
  geom_density(aes(teste),
               color=cores$verde_escuro,
               fill=cores$verde_claro,
               alpha=0.6,
               adjust=1.5) +
  geom_vline(xintercept = 1.96,
             color=cores$vermelho_escuro, size=1.3, linetype=4) +
  geom_vline(xintercept = -1.96,
             color=cores$vermelho_escuro, size=1.3, linetype=4) +
  labs(title="Distribuição da Estatística-t",
       y="Densidade") +
  tema_base_fundobranco()
```


Distribuição da Estatística−t
0.4
0.3
0.2
0.1
0.0
Densidade
−4
−2
0
2
teste

Teste de média simples (civis)

```r
data_civil <- filter(data, S==0)
t.test(data_civil$w_1, mu=2500, alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  data_civil$w_1
## t = -7.3743, df = 1496, p-value = 2.725e-13
## alternative hypothesis: true mean is not equal to 2500
## 95 percent confidence interval:
##  2296.127 2381.799
## sample estimates:
## mean of x
##  2338.963
```

∗ Importante notar que os intervalos de confiança dos salários de civis e militares (após o serviço militar) se cruzam.

Erro tipo II: cálculo

- Vamos testar se salário civil dos que não serviram é igual a BRL 2200 ($H_0 : \mu_{w^1_{S=0}} = 2200$)
  - Dado o PGD, sabemos que H_0 é falsa

```
data_civil <- filter(data, S==0)
t.test(data_civil$w_1, mu=2200, alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  data_civil$w_1
## t = 6.3634, df = 1496, p-value = 2.617e-10
## alternative hypothesis: true mean is not equal to 2200
## 95 percent confidence interval:
##  2296.127 2381.799
## sample estimates:
## mean of x
##  2338.963
```

Erro tipo II: cálculo

Mantido 5% como nível de significância, nosso PGD indica que há uma probabilidade de $\approx$ 0,7% de aceitarmos H_0, mesmo sendo ela falsa.

```
teste <- sapply(1:1000, gen_t_stat, mu.i=2200, Si=0, simplify = TRUE)
prob_erro_II <- sum(abs(teste)<1.96)/length(teste)
prob_erro_II
```

```
## [1] 0.007
```

Erro tipo II: cálculo

Importante notar que o Erro Tipo II aumenta, conforme H_0 se aproxima da hipótese verdadeira.

```
teste <- sapply(1:1000, gen_t_stat, mu.i=2300, Si=0, simplify = TRUE)
prob_erro_II <- sum(abs(teste)<1.96)/length(teste)
prob_erro_II
```

```
## [1] 0.814
```

Teste de diferença de médias

- Vamos testar diretamente se salários dos civis é igual ao salário de militares após o serviço ($H_0 : \mu_{S=1} = \mu_{S=0} \equiv \mu_{S=1} - \mu_{S=0} = 0$)
- Sob H_0, temos que:

$$\sqrt{n_{S=0} + n_{S=1}} \frac{\bar{\mu}_{w^1_{S=0}} - \bar{\mu}_{w^1_{S=1}} - 0}{\hat{\sigma}_{w^1_{S=0,S=1}}} \simeq N(0,1)$$

Teste-t (comando em R)

```r
t.test(data_serviu$w_1, data_civil$w_1,
       alternative="two.sided", var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  data_serviu$w_1 and data_civil$w_1
## t = -2.6729, df = 1998, p-value = 0.007581
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -201.66150  -30.97386
## sample estimates:
## mean of x mean of y
##  2222.645  2338.963
```

Teste-t (manualmente)

```
# Estimador amostral do desvio-padrão
sd_civil <- sd(data_civil$w_1)
n_civil <- nrow(data_civil)
sd_serviu <- sd(data_serviu$w_1)
n_serviu <- nrow(data_serviu)
sd_pool <- sqrt(((n_civil - 1) * sd_civil^2 + (n_serviu - 1) * sd_serviu^2) / (
se_hat <- sd_pool * sqrt(1/n_civil + 1/n_serviu)
se_hat
```

```
## [1] 43.51719
```

```
# Estimador amostral da média
mu_civil <- mean(data_civil$w_1)
mu_serviu <-  mean(data_serviu$w_1)
mu_hat <- mu_serviu - mu_civil
mu_hat
```

```
## [1] -116.3177
```

Teste-t (manualmente)

```r
# estatística t
t <- (mu_hat - 0)/se_hat
t
```

```
## [1] -2.672913
```

```r
# IC (95%) para média amostral
ic <- c(mu_hat - 1.96*se_hat,
        mu_hat + 1.96*se_hat)
ic
```

```
## [1] -201.61137  -31.02399
```

Interpretação dos resultados

- Pelo conjunto de eviências, fica claro que a expectativa salarial daqueles que serviram é menor do que daqueles que não serviram
  - e sabemos que isso é verdade pelo PGD
- Não podemos, a partir desse fato, concluir que serviço militar afeta negativamente a renda futura dos indivíduos
  - e sabemos, pelo PGD, que essa afirmação seria falsa
- Isso é um exemplo do que chamamos em econometria de **víes de seleção**
  - Indivíduos se autoselecionam para políticas públicas e uma mera comparação de beneficiários e não-beneficiários não permite conclusões causais.
  - Vamos ver mais sobre isso nas próximas aulas.

Efeito causal

- Pelo PGD, sabemos que o serviço militar aumenta a renda futura em BRL 200

$$
\begin{aligned}
E(w_i^1|S=0) - E(w_i^1|S=1) &= E(\delta_i\ \alpha_i + \epsilon_i^1) - E(\alpha_i + \epsilon_i^1) \\
&= E(\delta_i\ \alpha_i) + E(\epsilon_i^1) - E(\alpha_i) - E(\epsilon_i^1) \\
&= \delta_i\ E(\alpha_i) - E(\alpha_i) \\
&= (\delta_i - 1)\ E(\alpha_i) \\
&= 0.1 * 2000
\end{aligned} \tag{8}
$$

- Há algum teste-t que me permita estimar esse efeito causal?

Efeito causal

```r
data_rct_treated <- filter(data, D==1, S==1)
data_rct_control <- filter(data, D==1, S==0)
rct <- t.test(data_rct_treated$w_1, data_rct_control$w_1,
                alternative="two.sided", var.equal=TRUE)
rct
```

```
##
##  Two Sample t-test
##
## data:  data_rct_treated$w_1 and data_rct_control$w_1
## t = 4.1892, df = 1002, p-value = 3.046e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  115.5801 319.2797
## sample estimates:
## mean of x mean of y
##  2222.645  2005.215
```

Efeito causal

- O efeito pontual estimado é de 217.43, com intervalo de confiança de 95% indo de 115.58 a 319.28.
  - Inclui o valor correto, portanto.
- Dentre os que desejam se alistar, o acesso ao serviço militar (tratamento) foi definido de maneira aleatória.
  - Com isso garantimos que o tratamento não depende de outros fatores que também explicam o salário;
  - Controla-se, assim, o viés de seleção
- Falaremos mais sobre a diferença entre efeito causal e correlação nas próximas aulas...

Lista II

Orientações gerais

- Objetivo: replicar um teste de diferença de médias e uma análise de erros tipo I e II, tal qual fizemos nessa aula.
  - mas usaremos um cenário alternativo
- Novamente, vocês vão partir de um arquivo *modelo*, que é um arquivo tipo Rmd.
  - Lembrem de renomear o arquivo para um nome com padrão “Lista_1_nome” antes de submeter;
  - submetam o relatório gerado (html) e o código (Rmd) antes da próxima aula, enviando-os para o email daniel.sgrimaldi@outlook.com.br;
  - o assunto do email deve seguiro padrão “Lista_1_nome”

Cenário alternativo

- Dessa vez, imaginemos que o impacto do serviço militar acaba prejudicando a trajetória futura dos indivíduos, de tal forma que:

$$\delta_i = \begin{cases} 1.1 & \text{se } S_i = 1 \\ 1.3 & \text{se } c.c. \end{cases}$$

- Isso pode ocorrer, por exemplo, porque os que não entram para o serviço militar seguem estudando e os estudos têm impacto maior sobre a produtividade futura.

Lista II: especificações

- Vocês vão escrever a função exec_lista_2, que terá **obrigatoriamente 3 *inputs***
  - nome (nome do aluno);
  - sample_size=300 (tamanho da amostra a ser gerada para teste-t)
  - n_samples=100 (qtd. de amostras a serem geradas para cálculo de erros tipo I e II)
- A função exec_lista_2 terá **obrigatoriamente 2 *outputs***
  - nome: um objeto tipo *character* com o nome do aluno;
  - resultados: um objeto tipo vetor com parâmetros de interesse

Lista II: especificações

- O teste-t
  - você deve testar se os civis e militares têm o mesmo salário (W_1)
- Erro Tipo I
  - calcule o Erro Tipo I em um teste-t a respeito do salário dos militares $(\mu_{W^1_{S=1}})$
  - lembre-se que para esse teste, o valor de H_0 deve ser equivalente ao valor real de $\mu_{W^1_{S=1}}$, dado pelo PGD
- Erro Tipo II
  - calcule o Erro Tipo II em um teste-t a respeito do salário dos civis $(\mu_{W^1_{S=0}})$
  - para esse cálculo considere como H_0 o valor de BRL 3.000,00

Implementa função

```
lista_2 <- exec_lista2("Daniel Grimaldi", sample_size = 100, n_samples=100)
```

Aluno

```
lista_2$nome
```

```
## [1] "Daniel Grimaldi"
```

Parâmetros estimados

```
lista_2$resultados
```

```
##   Erro_tipo_I  Erro_tipo_II    estimativa       p-valor  estatistic-t
##  2.000000e-02  9.400000e-01 -7.511055e+02  5.259254e-04 -3.585897e+00
```