

Oficina de Estatística Tutorial 5

Virgilio Mendes

19/07/2019

Tutorial 5

Neste tutorial 5, continuaremos a usar os dados simulados no Tutorial 4. Contudo, vamos reduzir os tamanhos das nossas amostras. Agora, cada amostra da pesquisa de aprovação para prefeito tem somente 250 observações, e a nossa amostra do *pool* de candidatos vai conter somente 45 deles.

```
set.seed(1234)

# Simulacao pesquisa aprovacao do prefeito

amostras_pesquisas <- list()

n_amostra <- 400

for (i in 1:100) {
  banco_pesquisa <- rbinom(n_amostra, 1, prob = 0.7)
  amostras_pesquisas[[i]] <- banco_pesquisa
}

amostra_sorteada <- sample(amostras_pesquisas, 1)[[1]]

# Simulacao votos para vereador

set.seed(1234)

banco_candidatos <- rnorm(1300, mean = 600, sd = 40)

amostra_vereadores <- sample(banco_candidatos, 45, replace = F)
```

Testes de hipóteses

Antes de começarmos a testar algumas hipóteses sobre esses dados, lembrem-se dos 5 passos para realizarmos testes de hipótese/significância:

- **Suposição:** qual é a minha variável de interesse (quantitativa ou categórica)? Minha amostra é aleatória/probabilística? Qual é o tamanho da minha amostra?
- **Hipóteses:** definir H_0 e H_a .
- **Estatística-teste:** Calcule a estatística teste (seja ela Z ou t) para a sua estimativa
- **Valor p:** qual é o peso da evidência contra a sua H_0 ?

- **Conclusão:** relate a sua estimativa, estatística-teste, valor p e tome uma decisão formal sobre a rejeição ou não de H_0 .

Vamos começar testando a hipótese de que a maioria do número de cidadãos de Belo Horizonte aprova o trabalho do prefeito Alexandre Kalil (em nossos dados simulados).

Sendo assim, nossa variável de interesse é a aprovação/desaprovação do trabalho do prefeito, uma variável categórica. Nossa amostra é aleatória simples (devido às características da simulação deste dado), de $n = 400$.

A hipótese que desejamos testar é se pelo menos metade dos cidadãos reprova o trabalho do prefeito. Ou seja, H_0 : Proporção_Aprovação ≤ 0.5 e H_a : Proporção_Aprovação > 0.5 .

Ao testarmos essa hipótese no R, não precisamos calcular nossa estatística-teste ou o valor p. Temos uma função que faz isso por nós, a `prop.test`.

Ela necessita algumas especificações nos seus argumentos:

- `x`: o número de “sucessos” na definição da nossa proporção (no caso, indivíduos que aprovam).
- `n`: o tamanho da nossa amostra;
- `p`: o vetor de probabilidade de sucesso, ou seja, qual a H_0 que eu desejo testar;
- `alternative`: se o nosso teste será TBL (“two.sided”, que é o padrão da função), TULCI (“less”) ou TULCS (“greater”);
- `conf.level`: especifica algum nível de significância.

O próprio output da função nos retorna todas as informações que precisamos para concluir sobre o nosso teste de hipótese

```
table(amostra_sorteada)
```

```
amostra_sorteada 0 1 123 277
```

```
prop.test(277, n = 400, p = 0.5, alternative = "greater", conf.level = 0.90, correct = F)
```

1-sample proportions test without continuity correction

data: 277 out of 400, null probability 0.5 X-squared = 59.29, df = 1, p-value = 6.803e-15 alternative hypothesis: true p is greater than 0.5 90 percent confidence interval: 0.6621937 1.0000000 sample estimates: p 0.6925

```
# 1-sample proportions test with continuity correction
```

```
#data: 277 out of 400, null probability 0.5
#X-squared = 58.523, df = 1, p-value = 0.00000000000001005
#alternative hypothesis: true p is greater than 0.5
#95 percent confidence interval:
## 0.6520769 1.0000000
#sample estimates:
#      p
#0.6925
```

Além disso, também podemos salvar esse resultado do teste num objeto e acessar elementos específicos deste resultado, como o p-valor, a estimativa da proporção, o intervalo de confiança (e o nível de significância escolhido).

```
teste_proporcao <- prop.test(277, n = 400, p = 0.5, alternative = "greater", conf.level = 0.90)
teste_proporcao$p.value
```

```
[1] 1.004897e-14
```

```
teste_proporcao$estimate
```

```
p
```

```
0.6925
```

```
teste_proporcao$conf.int
```

```
[1] 0.6609158 1.0000000 attr(,"conf.level") [1] 0.9
```

Nosso teste de hipótese revelou que o valor p desse nosso teste de hipóteses é bastante baixo, o que significa que temos uma baixíssima chance de errar, caso rejeitemos a hipótese nula.

Vamos testar agora a hipótese de que a média de votos dos candidatos a vereador em Belo Horizonte foi menor ou igual a 400 votos.

Iniciando nosso teste de hipótese, verificamos que a nossa variável de interesse é quantitativa (número de votos), que nossa amostra é aleatória (pelas características da simulação) e de tamanho $n = 45$.

Nossa H_0 portanto é $H_0: \text{Média_votos} \leq 400$ e $H_a: \text{Média_votos} > 400$.

Para realizar esse teste de hipótese no R, podemos usar a função `t.test` (lembre-se que a distribuição t se aproxima da normal com $n \geq 30$). Assim como na função `prop.test`, o output da função nos dá as informações que precisamos saber.

```
t.test(amostra_vereadores, alternative = "greater", mu = 400)
```

One Sample t-test

data: amostra_vereadores t = 36.544, df = 44, p-value < 2.2e-16 alternative hypothesis: true mean is greater than 400 95 percent confidence interval: 594.5333 Inf sample estimates: mean of x 603.9087

```
#One Sample t-test

#data: amostra_vereadores
#t = 36.544, df = 44, p-value < 0.00000000000000022
#alternative hypothesis: true mean is greater than 400
#95 percent confidence interval:
# 594.5333      Inf
#sample estimates:
#mean of x
# 603.9087
```

E também conseguimos salvar os resultados do teste num objeto para acessar alguns elementos dele.

```
teste_media <- t.test(amostra_vereadores, alternative = "greater", mu = 400)
```

```
teste_media$p.value
```

```
[1] 7.344218e-35
```

```
teste_media$estimate
```

```
mean of x 603.9087
```

```
teste_media$conf.int
```

```
[1] 594.5333 Inf attr(,"conf.level") [1] 0.95
```