

Análise estatística de várias populações: 2ª Aula Prática - Parte 2

Profa. Dra. Tatiane F. N. Melo

14/09/2024

Inferência estatística para duas populações:

Teste de hipótese para igualdade de duas proporções

Exemplo 4: O objetivo, neste exemplo, é saber se a proporção de vacinados contra COVID-19, com primeira dose, em Aparecida de Goiânia é igual à proporção de vacinados em Anápolis. Podemos realizar o teste z para igualdade de proporções. E, para isso, podemos implementar a função ou usar a função já implementada no R.

As hipóteses de interesse aqui, são:

$$H_0 : p_{AG} = p_A \text{ contra } H_1 : p_{AG} \neq p_A,$$

onde p_{AG} é a proporção de vacinados contra COVID-19, com primeira dose, em Aparecida de Goiânia e p_A é a proporção de vacinados, também com primeira dose, em Anápolis.

Abaixo, segue o código em R, com uma função baseada nas fórmulas que vimos na aula teórica.

```
# Dados de vacinação
vac_aparecida <- 494118
pop_aparecida <- 527796

vac_anapolis <- 327481
pop_anapolis <- 398869

# Função para o teste z de igualdade de proporções (aplicando as fórmulas)
teste_z_proporcoes <- function(x, nx, y, ny)
{
  # Calcular proporções
  prop.x <- x / nx
  prop.y <- y / ny

  # Calcular proporção combinada
  p.barra <- (x+y) / (nx+ny)

  # Calcular a estatística z
  z.obs <- (prop.x - prop.y) / sqrt(p.barra * (1 - p.barra) * (1/nx + 1/ny))
}
```

```

# Calcular o valor p (bicaudal)
valor.p <- 2 * (1 - pnorm(abs(z.obs)))

# Retornar os resultados
return(list(
  z = z.obs,
  p_value = valor.p,
  prop.1 = prop.x,
  prop.2 = prop.y
))
}

teste_z_proporcoes(vac_aparecida, pop_aparecida, vac_anapolis, pop_anapolis)

## $z
## [1] 173.1322
##
## $p_value
## [1] 0
##
## $prop.1
## [1] 0.9361913
##
## $prop.2
## [1] 0.8210239

```

Também podemos usar uma função que já está implementada no R. Porém, esta função para testar proporções usa o teste Qui-Quadrado que veremos adiante na aula.

```

# Teste bilateral
prop.test(x = c(vac_aparecida, vac_anapolis),
          n = c(pop_aparecida, pop_anapolis))

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(vac_aparecida, vac_anapolis) out of c(pop_aparecida, pop_anapolis)
## X-squared = 29974, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1138050 0.1165297
## sample estimates:
##   prop 1    prop 2
## 0.9361913 0.8210239

```

Em ambos os resultados, temos a indicação de rejeição da hipótese $H_0 : p_{AG} = p_A$. Logo, as proporções de vacinados nas duas cidades são diferentes. Então, realizamos o teste unilateral à direita - $H_0 : p_{AG} \leq p_A$ contra $H_0 : p_{AG} > p_A$. No R, temos:

```
# Teste unilateral à direita
prop.test(x = c(vac_aparecida, vac_anapolis),
          n = c(pop_aparecida, pop_anapolis), alternative = "greater")

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(vac_aparecida, vac_anapolis) out of c(pop_aparecida, pop_anapolis)
## X-squared = 29974, df = 1, p-value < 2.2e-16
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.1140236 1.0000000
## sample estimates:
##      prop 1      prop 2
## 0.9361913 0.8210239
```

Também rejeitamos a hipótese nula. Concluimos que, ao nível de 1%, a proporção de vacinados contra COVID-19, com primeira dose, em Aparecida de Goiânia é maior do que em Anápolis.

Análise de aderência e associação: Teste de Aderência

Exemplo 5: Considere um estudo cujo objetivo é determinar a aceitação de um novo analgésico pelos pacientes. Para isso, 100 médicos selecionaram uma amostra de pacientes para participar do estudo. Cada médico selecionou de 25 pacientes. Após experimentar o novo analgésico por um período de tempo especificado, cada paciente foi questionado se preferia o novo medicamento ou o analgésico utilizado regularmente no passado. Estamos interessados em determinar se esses dados são ou não compatíveis com a hipótese de que foram extraídos de uma população que segue uma distribuição binomial.

Implementando a função no R:

```
# Pacote necessario para leitura dos dados
library(readxl)

## Warning: package 'readxl' was built under R version 4.3.3

dados.Ex5 <- read_excel("Exemplo5Aula2Parte2.xls")

print(dados.Ex5)

## # A tibble: 11 x 2
##   'Nr Pacientes' 'Nr Medicos'
##   <chr>          <dbl>
## 1 0              5
## 2 1              6
## 3 2              8
```

```
## 4 3 10
## 5 4 10
## 6 5 15
## 7 6 17
## 8 7 10
## 9 8 10
## 10 9 9
## 11 10 ou mais 0
```

```
# Frequências observados
freq.obs <- dados.Ex5$`Nr Medicos`

# Probabilidade: Um total de 500 pacientes dos
# 2.500 participantes do estudo disseram que,
# preferiam o novo analgésico, de modo que nossa
# estimativa pontual é
p <- 500/2500;

print(p)
```

```
## [1] 0.2
```

```
# 25 pacientes
n <- 25

# Total de médicos
total <- sum(freq.obs); total
```

```
## [1] 100
```

```
# Probabilidades esperadas
vetor = c()
vetor[1] = dbinom(0, n, p)
for(i in 1:9){vetor[i+1]=dbinom(i, n, p)}
probabilidades.esperadas <- c(vetor,1-pbinom(9, n, p))

# Frequências esperados
freq.esp <- total * probabilidades.esperadas

# Unindo as duas primeiras classes
freq.obs[1:2] <- sum(freq.obs[1:2])
freq.obs <- freq.obs[-1]; freq.obs
```

```
## [1] 11 8 10 10 15 17 10 10 9 0
```

```
freq.esp[1:2] <- sum(freq.esp[1:2])
freq.esp <- freq.esp[-1]; freq.esp
```

```
## [1] 2.738973 7.083550 13.576804 18.668105 19.601510 16.334592 11.084187
## [8] 6.234855 2.944237 1.733187
```

```
probabilidades.esperadas[1:2] <- sum(probabilidades.esperadas[1:2])
probabilidades.esperadas <- probabilidades.esperadas[-1]; probabilidades.esperadas
```

```
## [1] 0.02738973 0.07083550 0.13576804 0.18668105 0.19601510 0.16334592
## [7] 0.11084187 0.06234855 0.02944237 0.01733187
```

```
# Valor da estatística de teste
q.obs = sum(((freq.obs - freq.esp)^2)/(freq.esp));
print(q.obs)
```

```
## [1] 47.67772
```

```
nc = length(freq.obs); nc
```

```
## [1] 10
```

```
# Graus de liberdade
gl = nc-2

# Valor-p
pchisq(q.obs, df = gl, lower.tail = FALSE)
```

```
## [1] 1.13847e-07
```

Podemos utilizar diretamente a função existente no R:

```
# Teste Chi-Quadrado
teste = chisq.test(freq.obs, p = probabilidades.esperadas)
```

```
## Warning in chisq.test(freq.obs, p = probabilidades.esperadas): Aproximação do
## qui-quadrado pode estar incorreta
```

```
print(teste)
```

```
##
## Chi-squared test for given probabilities
##
## data: freq.obs
## X-squared = 47.678, df = 9, p-value = 2.934e-07
```

```
# O valor_p não é o mesmo, pois no R não considera que estamos
# estimando um parâmetro, então os gl = nc-1-1 = nc-2
# e no R é considerado gl = nc - 1 - 0 = nc - 1
valor_p_ajustado <- pchisq(teste$statistic, df = gl, lower.tail = FALSE)
print(valor_p_ajustado)
```

```
## X-squared
## 1.13847e-07
```

Portanto, rejeitamos a hipótese nula de que os dados vieram de uma distribuição binomial.

Exemplo 6: Agora usaremos os dados referentes à distribuição das doses das vacinas contra COVID-19, em Goiânia, aplicadas em diferentes faixas etárias. O objetivo é verificar se essa distribuição segue uma distribuição esperada baseada nas proporções da população. As hipóteses de interesse são:

H_0 : as vacinas são distribuídas proporcionalmente à população em cada faixa etária.

H_1 : as vacinas não são distribuídas proporcionalmente à população em cada faixa etária.

No R, temos:

```
# Dados referentes ao numero de vacinas contra COVID19
# aplicadas Goiânia, por faixa etária
dados.Ex6 <- read_excel("Exemplo6Aula2Parte2.xls")

# Dados observados
vac_aplicadas <- dados.Ex6$`Total de Doses Aplicadas`
pop_faixa_etaria <- dados.Ex6$`População de Goiânia`

# Calcular a população total de Goiânia
pop_total <- sum(pop_faixa_etaria)

# Calcular a proporção da população em cada faixa etária
prop_populacao <- pop_faixa_etaria / pop_total; prop_populacao

## [1] 0.02795803 0.02524008 0.08654310 0.07413534 0.02795454 0.07928589
## [7] 0.08421822 0.08318307 0.08544781 0.08458122 0.07072915 0.06342647
## [13] 0.05586291 0.04810981 0.03756811 0.02766707 0.01775205 0.02033712

# Realizar o teste qui-quadrado
chisq.test(vac_aplicadas, p = prop_populacao)

##
## Chi-squared test for given probabilities
##
## data: vac_aplicadas
## X-squared = 43552, df = 17, p-value < 2.2e-16
```

Rejeitamos a hipótese nula, ao nível de 1%, indicando assim, que a distribuição das vacinas aplicadas, por faixa etária, em Goiânia não corresponde à distribuição da população por faixa etária.