Poder do teste e tamanho do efeito

Felipe Barletta

30/03/2016

Contents

Tipos de erros

Situação	Conclusão do teste	
Real	Rejeitar H_0	Não rejeitar H_0
H_0 Verdadeira	erro tipo I	decisão correta
H_0 Falsa	decisão correta	erro tipo II

- Nível de significância, α: probabilidade de cometer o erro do tipo I
- β : probabilidade de cometer o erro do tipo II

Poder do teste

É a capacidade de um teste identificar diferenças que realmente existem, ou seja, de rejeitar Ho quando é realmente falsa, definida como 1-β

$$\mathbb{P}(\text{Erro do tipo II}) = \mathbb{P}(\text{não rejeitar } H_0 | H_0 \text{é falsa}) = \beta$$

O poder de um teste de hipóteses é afetado por três fatores:

- Tamanho da amostra: Quanto maior o tamanho da amostra, maior o poder do teste.
- Nível de Significância: Quanto maior o nível de significância, maior o poder do teste. Se você aumenta o nível de significância, você reduz a região de aceitação. Como resultado, você tem maior chance de rejeitar a hipótese nula. Isto significa que você tem menos chance de aceitar a hipótese nula quando ela é falsa, isto é, menor chance de cometer um erro do tipo II. Então, o poder do teste aumenta.
- O verdadeiro valor do parâmetro a ser testado: Quanto maior a diferença entre o "verdadeiro" valor do parâmetro e o valor especificado pela hipótese nula, maior o poder do teste.

Poder do teste para média

Suponha que a hipótese nula é falsa e que o verdadeiro valor da média é $\mu = \mu_0 + \delta$ $Z_0 = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X}-(\mu_0+\delta)}{\sigma/\sqrt{n}} + \frac{\delta}{\sigma/\sqrt{n}}$

$$Z_0 \sim N\left(\frac{\delta}{\sigma/\sqrt{n}}, 1\right)$$

- Teste bilateral:
$$\beta = \Phi\left(LS - \frac{\mu_0+\delta}{\sigma/\sqrt{n}}\right) - \Phi\left(LI - \frac{\mu_0+\delta}{\sigma/\sqrt{n}}\right)$$
- Testes unilaterais:

$\Phi\left(LS - \frac{\mu_0+\delta}{\sigma/\sqrt{n}}\right) \quad 1 - \Phi\left(LI - \frac{\mu_0+\delta}{\sigma/\sqrt{n}}\right)$

Poder do teste - σ^2 conhecida

Considere $\delta = 1$, $n = 30$, $\alpha = 0,05$, $\mu = 8$ e $\sigma = 2,1$

Ou seja vamos testar:

$$H_0 : \mu = 8$$
$$H_1 : \mu \neq 8$$

$$\beta = \Phi\left(LS - \frac{\mu_0+\delta}{\sigma/\sqrt{n}}\right) - \Phi\left(LI - \frac{\mu_0+\delta}{\sigma/\sqrt{n}}\right) = \Phi(-0,9633) - \Phi(-4,2531) = 0,1677$$

Poder $= 1 - \beta = 1 - 0,1677 = 0,8323$

> $\Phi \sim$ *Normal padrão acumulada.*

Poder do teste - σ^2 conhecida

Primeiro calculamos o intervalo de confiança para a média:

```
###### Calculando Intervalo de confiança(95\%)
media <- 8
sd <- 2.1
n <- 30
zep <- qnorm(0.95)*sd/sqrt(n)
li_z <- media-zep
ls_z <- media+zep
cbind(limite_inferior=round(li_z,2),limite_superior=round(ls_z,2))
```

```
##      limite_inferior limite_superior
## [1,]            7.37            8.63
```

Poder do teste - σ^2 conhecida

Agora substituímos os valores na definição:

```
delta <- 1
zli <- (li_z-(media+delta))/(sd/sqrt(n))
zls <- (ls_z-(media+delta))/(sd/sqrt(n))
betaz <- pnorm(zls)-pnorm(zli)
betaz
```

```
## [1] 0.1676757
```

```
poderz <- 1-betaz
poderz
```

```
## [1] 0.8323243
```

A probabilidade de rejeitar H_0 é de aproximadamente 83%

Poder do teste - σ^2 desconhecida

Supondo que não conhecemos o valor da variância populacional, utilizamos a distribuição *t-student*

Repetimos os passos anteriores.

```
###### Calculando Intervalo de confiança(95\%)
media <- 8
sd <- 2.1
n <- 30
ept <- qt(0.95,df=n-1)*sd/sqrt(n)
li_t <- media-ept
ls_t <- media+ept
cbind(limite_inferior=round(li_t,2),limite_superior=round(ls_t,2))
```

```
##      limite_inferior limite_superior
## [1,]            7.35            8.65
```

Poder do teste - σ^2 desconhecida

Agora substituímos os valores na definição:

```
delta <- 1
tli <- (li_t-(media+delta))/(sd/sqrt(n))
tls <- (ls_t-(media+delta))/(sd/sqrt(n))
betat <- pt(tls,df=n-1)-pt(tli,df=n-1)
betat
```

```
## [1] 0.1853156
```

```
podert <- 1-betat
podert
```

```
## [1] 0.8146844
```

A probabilidade de rejeitar H_0 é de aproximadamente 81%

Tamanho do efeito

- **O que é o tamanho do efeito?**

É a magnitude da diferença entre os grupos

- **Por que calcular o tamanho do efeito?**

P-valor não revela o tamanho do efeito - relata apenas, se o efeito existe

- **Por que o *P-valor* não é suficiente?**

P-valor significativo nos diz que a diferença observada entre dois grupos não é devido ao acaso;

Quando a amostra é muito grande, o teste quase sempre demonstra diferença significativa;

Diferenças pequenas, mesmo que significativa, geralmente não fazem sentido;

Métodos para calcular o tamanho do efeito

- Cohen's d

Primeiro calculamos o desvio padrão agrupado

$d = (\bar{x_1} - \bar{x_2})/s_{pool}$

Depois o índice de Cohen's d

$s_{poll} = (s_1 + s_2)/2$

No R

```
treatment <- rnorm(100,mean=10)
control <- rnorm(100,mean=12)
#install.packages('effsize',repos='http://cran-r.c3sl.ufpr.br')
require(effsize)
```

```
## Loading required package: effsize
```

```
## calculando Cohen's d
cohen.d(treatment,control,pooled=TRUE)
```

```
##
## Cohen's d
##
## d estimate: -1.962212 (large)
## 95 percent confidence interval:
##       inf       sup
## -2.303907 -1.620516
```

Métodos para calcular o tamanho do efeito

- Cliff's Delta

```
require(effsize)
d <- (c(treatment,control))
f <- rep(c("Treatment","Control"),each=100)
## Calculando Cliff's Delta
cliff.delta(d,f,pooled=TRUE)
```

```
##
## Cliff's Delta
##
## delta estimate: 0.8274 (large)
## 95 percent confidence interval:
##       inf       sup
## 0.7368814 0.8887724
```

Métodos para calcular o tamanho do efeito

- Vargha and Delaney

Exemplo no R - Hedges' g

```
## compute Hedges' g
cohen.d(d,f,hedges.correction=TRUE, pooled=TRUE)
```

```
##
## Hedges's g
```

```
##
## g estimate: 1.95477 (large)
## 95 percent confidence interval:
##      inf      sup
## 1.613498 2.296041
```