Testes de Hipóteses

Rodrigo R. Pescim

Universidade Estadual de Londrina

09 de Maio de 2020

Teste de Hipóteses

- Em geral, intervalos de confiança é o processo de estimação mais informativo acerca de um parâmetro de interesse no estudo de certo fenômeno

- Entretanto, algumas vezes, existe um particular interesse em verificar certas afirmações ou conjecturas

- Cada uma dessas afirmações constitui uma hipótese que pode ser associada a uma regra de decisão (teste)

Exemplo 1

Um pesquisador deseja saber se o pH de um solo é ácido. Para isso, ele coletou uma amostra com cinco observações e obteve os seguintes valores de pH: $5,8$; $6,3$; $6,9$; $6,2$; $5,5$. Com base nessas informações, teste a hipótese de que o pH médio do solo é ácido com nível de significância de 5%. Qual é a conclusão ?

Conceitos Básicos

1. Hipótese Estatística:
   - É qualquer afirmação que se faça sobre um parâmetro populacional desconhecido
   - Observa-se que a partir de uma amostra, pode-se estabelecer uma regra de decisão na qual rejeita-se (ou não) a hipótese proposta
2. Tipos de Hipóteses:
   - Hipótese nula (H_0) $\rightarrow$ É a hipótese mais "importante"
   - Hipótese alternativa (H_1) $\rightarrow$ É qualquer outra hipótese diferente de H_0

- Em muitas aplicações da estatística, convenciona-se definir H_1 como a hipótese formulada pelo pesquisador, enquanto a H_0 é o seu complemento.

- A princípio, H_0 é considerada a verdadeira. Ao confrontarmos H_0 com os achados de uma amostra aleatória tomada de uma população de interesse, **verifica-se a sua plausibilidade em termos probabilísticos**, o que leva a rejeitar ou não H_0.

- Se não rejeitamos H_0, tomamo-la como verdadeira; caso contrário, tomamos H_1 como verdadeira.

- No entanto, **ao utilizar nesta tomada de decisão uma amostra** (uma parte da população) e não a população inteira, pode-se **cometer dois tipos de erros**.

3. Tipos de Erros:

    - Ao tomar uma decisão a favor ou contra uma hipótese, existem dois tipos de erros que podemos cometer
    - Podemos **rejeitar** a hipótese nula (H_0) quando de fato ela é **verdadeira** (erro tipo I)
    - Podemos **não-rejeitar** a hipótese nula (H_0) quando de fato ela é **falsa** (erro tipo II)

- Frequentemente, denota-se a probabilidade desses dois tipos de erros como α e β, respectivamente
- $\alpha = P(\text{Rejeitar } H_0 \mid H_0 \text{ é falsa}) \rightarrow$ Erro tipo I
- $\beta = P(\text{Não-Rejeitar } H_0 \mid H_0 \text{ é verdadeira}) \rightarrow$ Erro tipo II

Tabela 1: Tabela Aceitação/Rejeição da hipótese nula

	Não-Rejeitar H_0	Rejeitar H_0
H_0 é verdadeira	Decisão Correta	Erro tipo I (α)
H_0 é falsa	Erro tipo II (β)	Decisão Correta

- Portanto, o valor de α determina a chance de erro do teste de hipótese.
- Função poder $\rightarrow \pi = 1 - \beta$

Na maioria das situações práticas, não sabemos o verdadeiro valor do desvio-padrão populacional σ

Na maioria das situações práticas, não sabemos o verdadeiro valor do desvio-padrão populacional σ

Se não conhecemos σ, então **não podemos usar a distribuição normal** padrão para estimarmos a verdadeira média populacional μ

Na maioria das situações práticas, não sabemos o verdadeiro valor do desvio-padrão populacional σ

Se não conhecemos σ, então **não podemos usar a distribuição normal** padrão para estimarmos a verdadeira média populacional μ

Nesse caso, usaremos a **distribuição t de Student** que permite o uso da estimativa do **desvio-padrão amostral** *s* no lugar do valor desconhecido de σ.

Distribuição t de student

Se uma população tem distribuição normal, então a distribuição da estatística

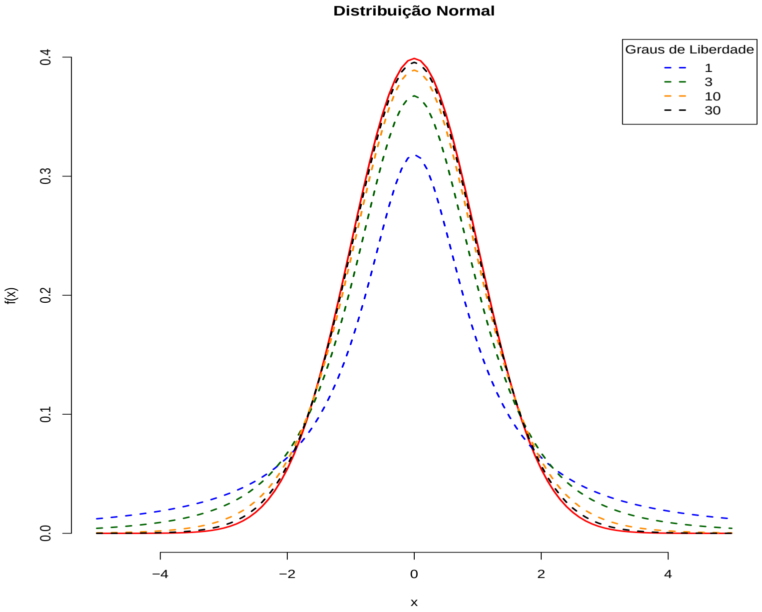
$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

é uma **distribuição t de Student** (ou simplesmente distribuição t), com $(n-1)$ **graus de liberdade**.

A distribuição t de Student

Características da distribuição t:

- É simétrica com média em $t = 0$ (assim como a distribuição normal $\mu = 0$)
- É diferente para tamanhos de amostra diferentes
- Possui **maior área nas caudas** e menor área no centro (quando comparada com a distribuição normal) $\rightarrow$ para incorporar a incerteza
- A medida que o **tamanho da amostra n aumenta**, a distribuição t se aproxima cada vez mais de uma distribuição normal padrão
- Por isso, para **amostras grandes** ($n > 30$) o resultado das duas é similar


Distribuição Normal
Graus de Liberdade
1
3
10
30
f(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
x

Teste de hipóteses para a média populacional

- **1^o passo**: Formular as hipóteses

$$H_0 : \mu = \mu_0 \quad \times H_1 : \begin{cases} \mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0 \end{cases}$$

- em que μ_0 é um valor numérico especificado

- **2^o passo**: Estatística do teste

- Variância (σ^2) desconhecida:

$$t_{calc} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}},$$

com $(n-1)$ graus de liberdade

- **3^o passo**: Região de Rejeição ou Crítica
  - É uma região ou conjunto de valores assumidos pela estatística de teste para os quais H_0 é rejeitada
  - Verificar a **probabilidade de significância (p-valor)**

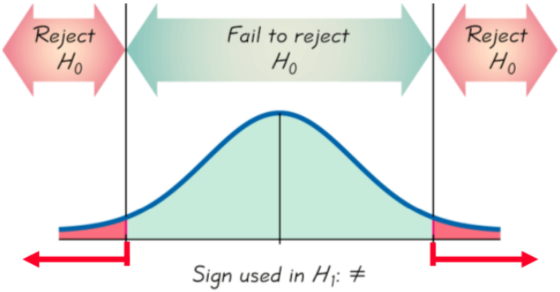
- **4^o passo**: Conclusão

Teste Bilateral

Uma **hipótese alternativa** dada por:

- $H_1 : \mu \neq \mu_0$

Possui uma **região de rejeição** dada por

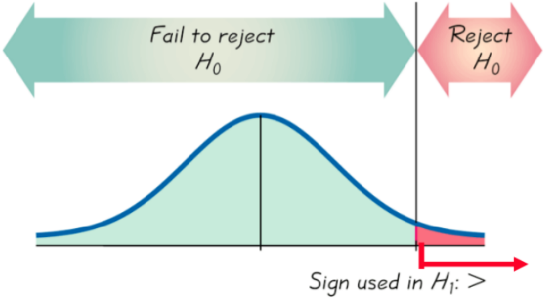


Teste Unilateral à direita

Uma **hipótese alternativa** dada por:

- $H_1 : \mu > \mu_0$

Possui uma **região de rejeição** dada por

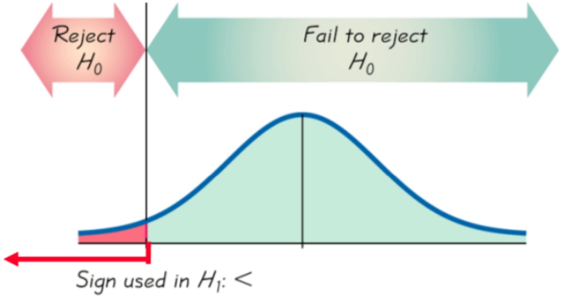


Teste Unilateral à esquerda

Uma **hipótese alternativa** dada por:

- $H_1 : \mu < \mu_0$

Possui uma **região de rejeição** dada por



Probabilidade de Significância (p-valor)

- O p-valor depende diretamente de uma dada amostra e fornece uma medida da "força" dos resultados de um teste, em contraste a uma simples rejeição ou não rejeição de H_0.

- Neste sentido, **o p-valor é uma medida do quanto de evidência tem-se contra a hipótese nula** (H_0).

- Quanto menor o p-valor, mais evidência tem-se contra H_0, isto é, maior é a chance de rejeitarmos H_0.

- Deve-se combinar **o p-valor com o nível de significância (α) para tomar decisão sobre um teste de hipótese**. Em tal caso, se o p-valor for menor que um corte (usualmente 0,05 ou 5%) então rejeita-se a hipótese nula.

Probabilidade de Significância (p-valor)

- O p-valor depende diretamente de uma dada amostra e fornece uma medida da "força" dos resultados de um teste, em contraste a uma simples rejeição ou não rejeição de H_0.

- Neste sentido, **o p-valor é uma medida do quanto de evidência tem-se contra a hipótese nula** (H_0).

- Quanto menor o p-valor, mais evidência tem-se contra H_0, isto é, maior é a chance de rejeitarmos H_0.

- Deve-se combinar **o p-valor com o nível de significância (α) para tomar decisão sobre um teste de hipótese**. Em tal caso, se o p-valor for menor que um corte (usualmente 0,05 ou 5%) então rejeita-se a hipótese nula.

- P-valor $\leq \alpha \rightarrow$ Rejeita-se H_0
- P-valor $> \alpha \rightarrow$ Não Rejeita H_0

Testes de normalidade

- Existem muitos testes para verificar a normalidade dos dados, como por exemplo os testes de Shapiro-Wilk, Lilliefors, Kolmogorov-Smirnov, Anderson-Darling, etc.

- Para os testes de normalidade, independentemente do teste, as hipóteses nula e alternativa são

- H_0 : os dados seguem distribuição normal
- H_1 : os dados não seguem distribuição normal

- O teste é realizado ao nível de significância α

Teste para Uma Média Populacional - Exemplo 1

- A função t.test() realiza o teste t-student

```
#####################################
## Teste t para media populacional###
#####################################

rm(list=ls())

### Exemplo 1

# entrada dos dados

ph <- c(5.8, 6.3, 6.9, 6.2, 5.5)

## verificando a normalidade dos dados

shapiro.test(ph)

## Aplicando o teste t para H0: mu = 7 x H1: mu < 7

t.test(ph, alternative = c("less"), mu=7, conf.level = 0.95)
```

Saída do R

- Teste t para uma amostra

```
One Sample t-test

data:  ph
t = -3.6148, df = 4, p-value = 0.01123
alternative hypothesis: true mean is less than 7
95 percent confidence interval:
-Inf 6.647182
sample estimates:
mean of x
6.14
```

Exemplo 2

Um artigo no periódico *Materials Engineering* (1989, Vol. II, No. 4, pp. 275-281) descreve os resultados de teste de tensão quanto à adesão em 22 corpos de prova de liga U-700. A carga no ponto de falha do corpo de prova é dada a seguir (em MPa):

19.8	18.5	17.6	16.7	15.8	15.4	14.1	
13.6	11.9	11.4	11.4	8.8	7.5	15.4	
15.4	19.5	14.9	12.7	11.9	11.4	10.1	7.9

O objetivo é verificar se os dados sugerem que a carga média na falha excede 10 MPa. Considerar que a carga na falha tem uma distribuição normal e utilizar um nível de significância de 5%

No R

```
## Exemplo 2

carga <- c(19.8, 18.5, 17.6, 16.7, 15.8, 15.4, 14.1, 13.6,
+ 11.9, 11.4, 11.4, 8.8, 7.5, 15.4, 15.4, 19.5,
+ 14.9, 12.7, 11.9, 11.4, 10.1, 7.9)

# verificar pressuposicao de normalidade
shapiro.test(carga)

t.test(carga, alternative = c("greater"),
       mu=10, conf.level = 0.95)
```

Saída do R no Exemplo 2

```
One Sample t-test

data:  carga
t = 4.9017, df = 21, p-value = 3.781e-05
alternative hypothesis: true mean is greater than 10
95 percent confidence interval:
12.40996      Inf
sample estimates:
mean of x
13.71364
```

1. Teste t não-pareado (Amostras independentes)

- **1º passo**: Formular as hipóteses

$$H_0 : \mu_1 = \mu_2 \quad \times H_1 : \begin{cases} \mu_1 \neq \mu_2 \\ \\ \mu_1 > \mu_2 \\ \\ \mu_1 < \mu_2 \end{cases}$$

- **2^o passo**: Estatística do teste
- Variância amostral homogênea :

$$t_{calc} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left[\frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1+n_2-2}\right]}},$$

com $(n_1 + n_2 - 2)$ graus de liberdade

- **3^o passo**: Região de Rejeição ou Crítica
- **4^o passo**: Conclusão

Exemplo 3 : Suponha, em um experimento industrial, que um engenheiro está interessado em saber como a absorção média de uma mistura em concreto varia entre dois agregados de concreto. As amostras foram expostas à mistura por 48 horas. Decidiu-se que 6 amostras seriam testadas para cada agregado. Os dados estão registrados na tabela abaixo:

Agregado 1	551	457	450	731	499	632
Agregado 2	639	615	511	573	648	601

Realize um teste de hipóteses ao nível de significância de 1% para verificar se existe diferença entre os agregados 1 e 2 com relação a absorção média de uma mistura em concreto.

- Teste t Não - Pareado

```
##############################################
## Teste t para duas medias populacionais ###
##############################################

## Exemplo 3

A1 <- c(551,457,450,731,499,632)

A2 <- c(639,615,511,573,648,601)

shapiro.test(A1) # normalidade para os dados da amostra 1
shapiro.test(A2) # normalidade para os dados da amostra 2

var.test(A1,A2) # teste de homogeneidade de duas variâncias

t.test(A1, A2, alternative = c("two.sided"), paired=F,
      var.equal=T, conf.level = 0.99)
```

2. Teste t pareado (Amostras dependentes)

- **1^o passo**: Formular as hipóteses

$$H_0 : \mu_1 - \mu_2 = 0 \quad \times H_1 : \begin{cases} \mu_1 \neq \mu_2 \\ \mu_1 > \mu_2 \\ \mu_1 < \mu_2 \end{cases}$$

- **2^o passo**: Estatística do teste

$$t_{calc} = \frac{\bar{d}}{s_d/\sqrt{n}},$$

com $(n-1)$ graus de liberdade, em que $n =$ número de pares

- $\bar{d}$ é a média das diferenças, s_d é o desvio-padrão das diferenças e n é o números de pares

- **3^o passo**: Região de Rejeição ou Crítica

- **4^o passo**: Conclusão

- **Exemplo 4** : Um produto foi desenvolvido com o objetivo de reduzir a temperatura média do funcionamento de motores. Para testar o efeito do produto, foram selecionados ao acaso 8 motores e após 10 minutos de funcionamento, em cada condição, foram obtidos os dados (em ^oC) do quadro abaixo.

Motor	Sem Produto	Com Produto
1	80.5	75.8
2	99.6	98.8
3	99.4	77.6
4	100.2	99.9
5	81.5	74.2
6	84.6	80.5
7	85.0	83.6
8	105.8	105.8

Podemos afirmar que a temperatura média dos motores diminuiu, com a aplicação do produto, ao nível de 5% de significância?

Exemplo 4 - No R

- Teste t Pareado (Amostras Dependentes)
- Porém, pode-se contruir uma função para o cálculo da média por meio do comando function()

```
##############################################
## Teste t para duas medias populacionais ###
##############################################

## Exemplo 5:

Semprod <- c(80.5,99.6,99.4,100.2,81.5,84.6,85.0,105.8)
Comprod <- c(75.8,98.8,77.6,99.9,74.2,80.5,83.6,105.8)

shapiro.test(Semprod)
shapiro.test(Comprod)

t.test(Semprod, Comprod, alternative = c("greater"), paired=T,
        conf.level = 0.95)
```

Exemplo 5

A Tabela abaixo indica duas amostras de rendimentos (%) de uma reação química em função do catalisador utilizado.

Catalisador A	43	45	51	50	62	42	53	50	48	55
Catalisador B	45	35	43	59	48	45	41	43	49	39

Realize um teste de hipóteses ao nível de significância de 1% supondo que o rendimento médio do catalisador A é superior ao rendimento médio do catalisador B. O que você conclui ?