AULA 11
Teste de Hipótese

Ernesto F. L. Amaral

16 de setembro de 2010
Metodologia de Pesquisa (DCP 854B)

Fonte:
Triola, Mario F. 2008. "Introdução à estatística". 10 ª ed. Rio de Janeiro: LTC. Capítulo 8 (pp.304-359).

ESQUEMA DA AULA

– Fundamentos do teste de hipótese.

– Teste de uma afirmativa sobre uma proporção.

– Teste de uma afirmativa sobre uma média: σ conhecido.

– Teste de uma afirmativa sobre uma média: σ desconhecido.

– Teste de uma afirmativa sobre um desvio padrão ou uma variância.

FUNDAMENTOS DO TESTE DE HIPÓTESE

HIPÓTESE

– **Inferência estatística** usa dados amostrais para duas atividades principais:

  – Estimar parâmetro populacional.

  – Testar hipótese ou afirmativa sobre parâmetro populacional.

– Em estatística, **hipótese** é uma afirmativa sobre uma propriedade da população.

– **Teste de hipótese** (teste de significância) é um procedimento padrão para se testar uma afirmativa sobre uma propriedade da população.

REGRA DO EVENTO RARO

– Métodos de teste de hipótese se baseiam na **regra do evento raro** em inferência estatística.

– Se, sob uma dada suposição, a probabilidade de um evento observado particular é excepcionalmente pequena, concluímos que a suposição provavelmente não é correta.

– Testamos uma afirmativa analisando dados amostrais na tentativa de distinguir entre resultados que podem facilmente ocorrer por acaso e resultados que são altamente improváveis de ocorrer por acaso.

FUNDAMENTOS DO TESTE DE HIPÓTESE

– É importante entender os **componentes individuais** de um teste de hipótese.

– **Conceitos básicos:** hipótese nula, hipótese alternativa, estatística de teste, região crítica, nível de significância, valor crítico, valor *P*, erro tipo I e erro tipo II.

– **Além do básico:** poder de um teste.

CONCEITOS BÁSICOS DE TESTES DE HIPÓTESES

– **Objetivos:**

– Dada uma afirmativa, identificar a hipótese nula e a hipótese alternativa e expressar ambas em forma simbólica.

– Dados uma afirmativa e dados amostrais, calcular o valor da estatística de teste.

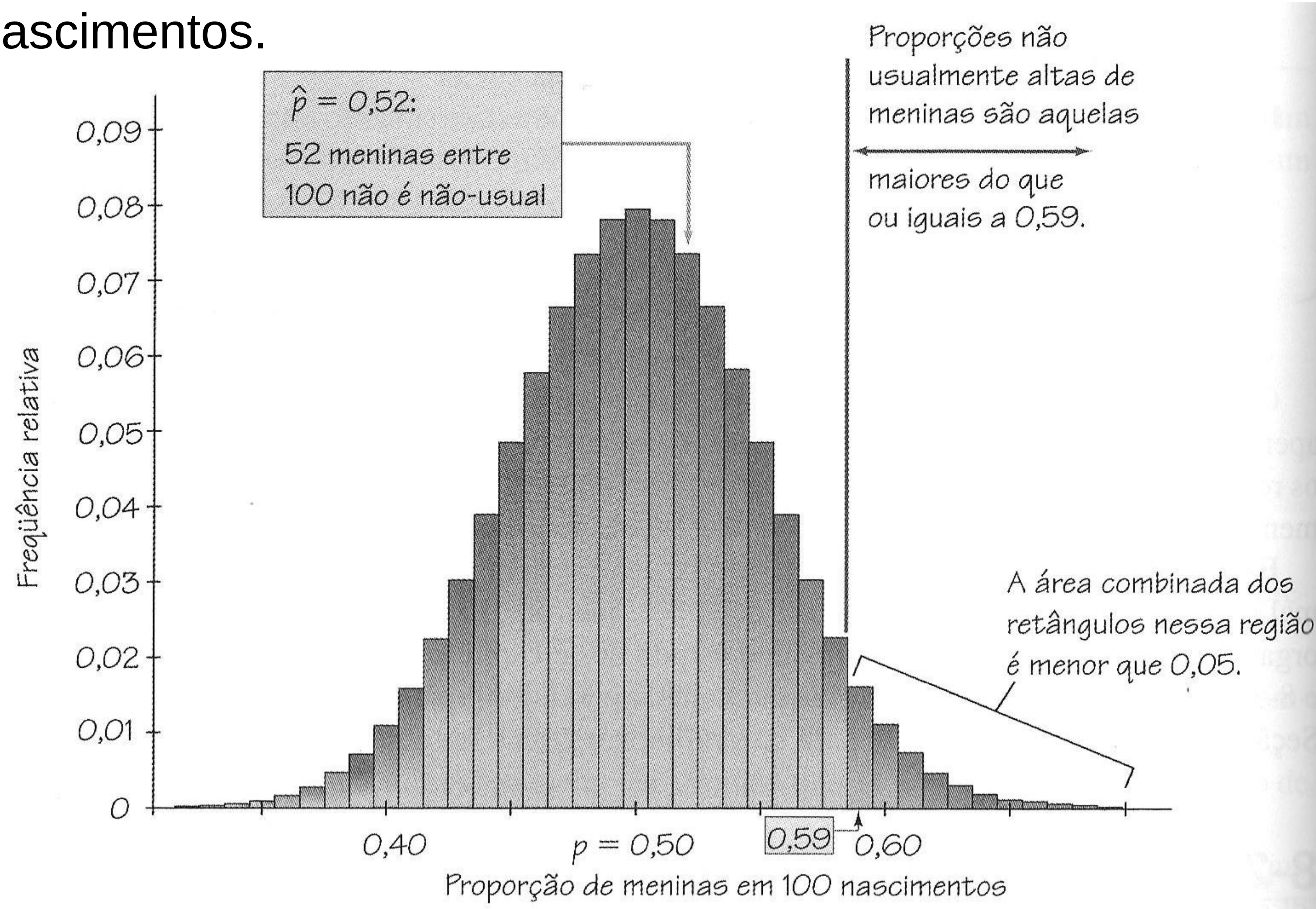
– Dado um nível de significância, identificar os valores críticos.

– Dado um valor da estatística de teste, identificar o valor *P*.

– Estabelecer a conclusão de um teste de hipótese em termos simples, não-técnicos.

EXEMPLO

– Distribuição amostral das proporções de meninas em 100 nascimentos.



COMPONENTES DE UM TESTE DE HIPÓTESE FORMAL

– **Hipótese nula (H_0)** é uma afirmativa de que o valor de um parâmetro populacional (proporção, média ou desvio padrão) é igual a algum valor especificado.

– Testamos a hipótese, supondo que ela seja verdadeira e chegamos à conclusão para rejeitar ou não rejeitar H_0.

– Por exemplo: H_0: p=0,5; ou H_0: μ=98,6; ou H_0: σ=15.

– **Hipótese alternativa (H_1 ou H_a ou H_A)** é a afirmativa de que o parâmetro tem um valor que difere da hipótese nula.

Proporções	**H_1: p>0,5**	**H_1: p<0,5**	**H_1: p≠0,5**
Médias	**H_1: μ>98,6**	**H_1: μ<98,6**	**H_1: μ≠98,6**
Desvios padrões	**H_1: σ>15**	**H_1: σ<15**	**H_1: σ≠15**

ALGUMAS OBSERVAÇÕES

– **Sobre o sinal de igualdade em H_0:**

– Alguns livros usam os símbolos ≤ ou ≥.

– Porém, Triola sugere fazer o teste de hipótese supondo que a proporção, média ou desvio padrão seja **igual** a algum valor especificado.

– **Sobre o estabelecimento de suas próprias hipóteses:**

– Se você usa um teste de hipótese para **apoiar sua afirmativa**, esta deve ser sua hipótese alternativa (hipótese de pesquisa).

– Deve ser escrita usando os símbolos **<** ou **>** ou **≠**.

– Não se deve usar teste de hipótese para apoiar afirmativa de que parâmetro seja igual a algum valor especificado.

IDENTIFICAÇÃO DE H_0 E H_1

– Identifique a afirmativa ou hipótese específica a ser testada e expresse-a em forma simbólica.

– Dê a forma simbólica que tem que ser verdadeira quando a afirmativa original é falsa.

– Das duas expressões simbólicas obtidas até agora:
  - – Faça da que não contém a igualdade, a hipótese alternativa H_1, de modo que H_1 use o símbolo $<$ ou $>$ ou $\neq$.
  - – Deixe que a hipótese nula H_0 seja a expressão simbólica que iguala o parâmetro ao valor fixo sendo considerado.

ESTATÍSTICA DE TESTE

– A **estatística de teste** é um valor usado para se tomar a decisão sobre a hipótese nula.

– Essa estatística é encontrada pela **conversão da estatística** amostral em um escore com a suposição de que a hipótese nula seja verdadeira.

– Estatística de teste para a **proporção**: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$

– Estatística de teste para a **média**:

$$z = \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \quad \text{ou} \quad t = \frac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$$

– Estatística de teste para o **desvio padrão**: $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$

REGIÃO CRÍTICA, NÍVEL DE SIGNIFICÂNCIA, VALOR CRÍTICO

– **Região crítica (região de rejeição)** é o conjunto de todos os valores da estatística de teste que nos fazem rejeitar a hipótese nula.

– **Nível de significância (*α*)** é a probabilidade da estatística de teste cair na região crítica quando a hipótese nula for verdadeira.

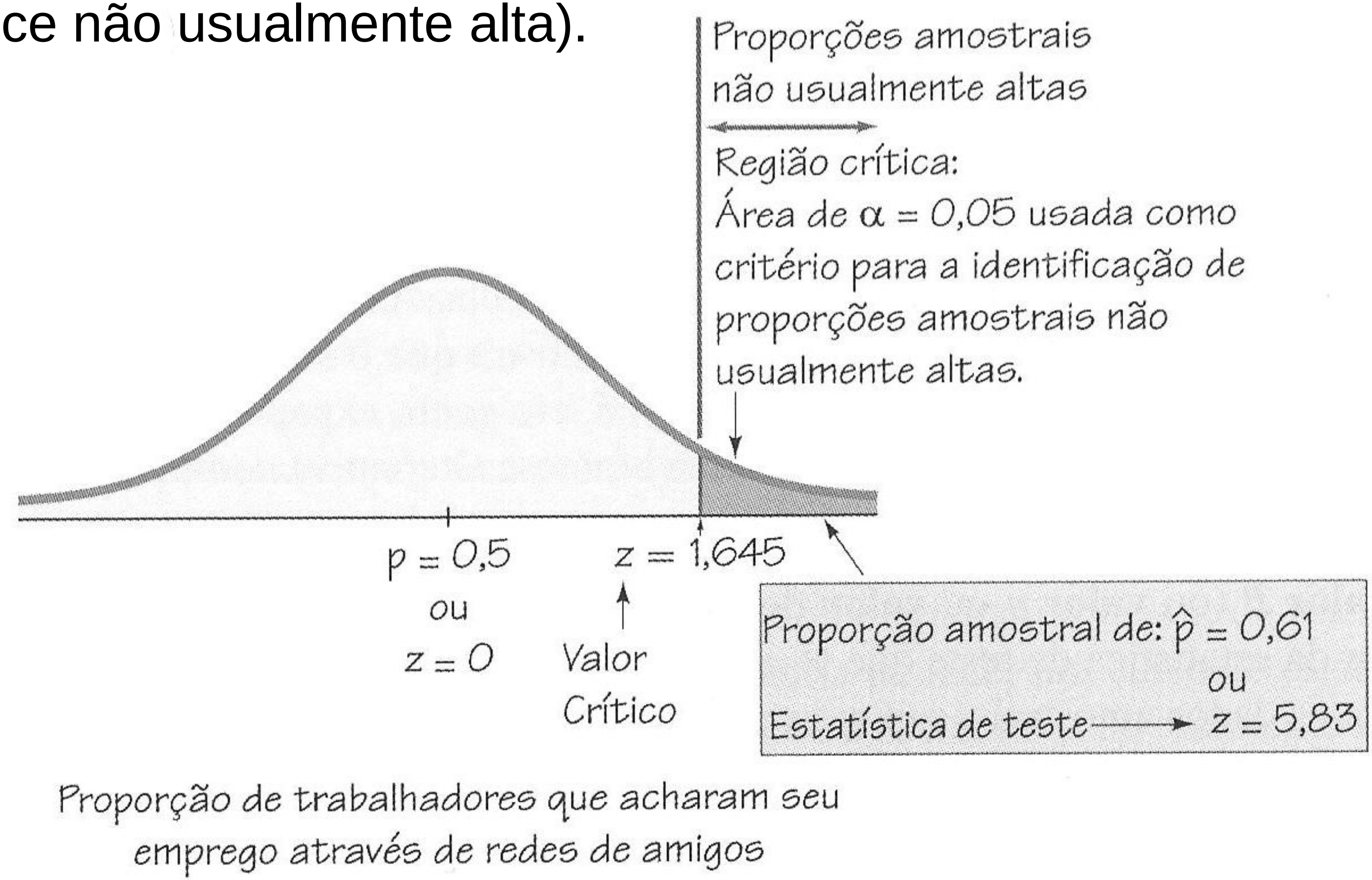
  – Se estatística de teste cair na região crítica, rejeitamos a hipótese nula, sendo α igual à probabilidade de cometer o erro de rejeitar a hipótese nula quando ela é verdadeira.

– **Valor crítico** é qualquer valor que separa a região crítica (em que rejeitamos H_0) dos valores da estatística de teste que não levam à rejeição da hipótese nula.

  – Depende da hipótese nula, distribuição amostral e α.

REGIÃO CRÍTICA, VALOR CRÍTICO, ESTATÍSTICA DE TESTE

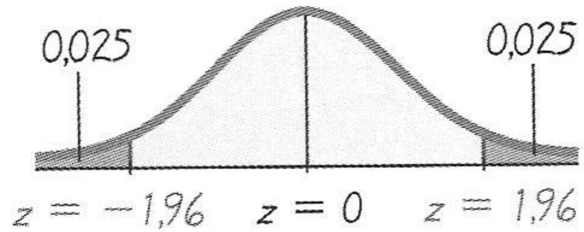
– Proporção amostral de 0,61 está na região de valores significativos, já que não têm chance de ocorrer por acaso (chance não usualmente alta).



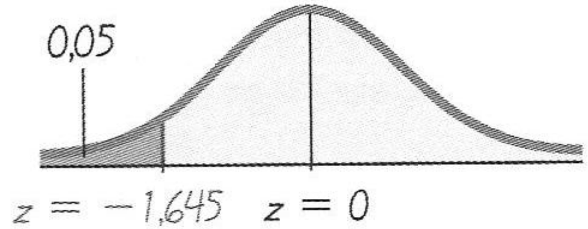
BILATERAL, UNILATERAL À ESQUERDA OU À DIREITA

– Caudas em uma distribuição são as regiões extremas limitadas pelos valores críticos e dependem de H_1.

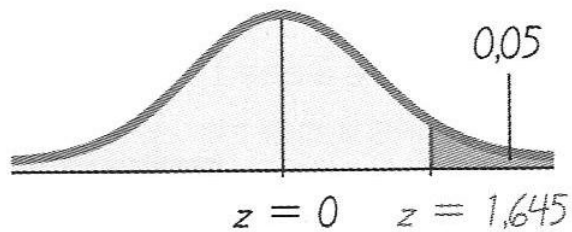
– **Teste bilateral:** região crítica está nas duas regiões extremas sob a curva.



– **Teste unilateral à esquerda:** região crítica está na região extrema esquerda sob a curva.



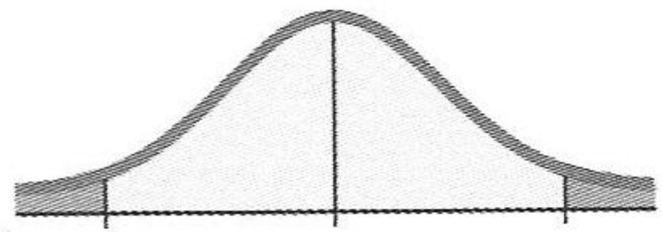
– **Teste unilateral à direita:** região crítica está na região extrema direita sob a curva.



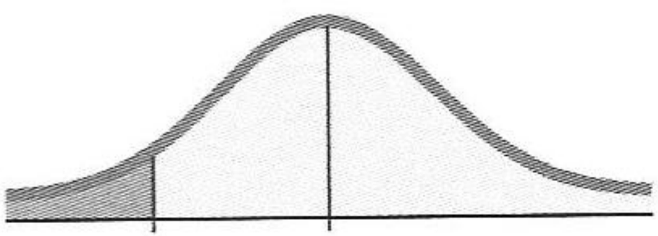
MAIS SOBRE TIPO DE TESTES

– Cauda será a região crítica com valores que entrarão em conflito significativo com hipótese nula.

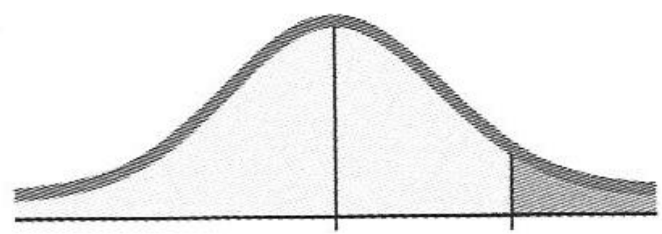
– O sinal de desigualdade em H_1 indica a direção da região crítica.



Sinal usado em $H_1: \neq$
Teste bilateral



Sinal usado em $H_1: <$
Teste unilateral à esquerda



Sinal usado em $H_1: >$
Teste unilateral à direita

VALOR *P*

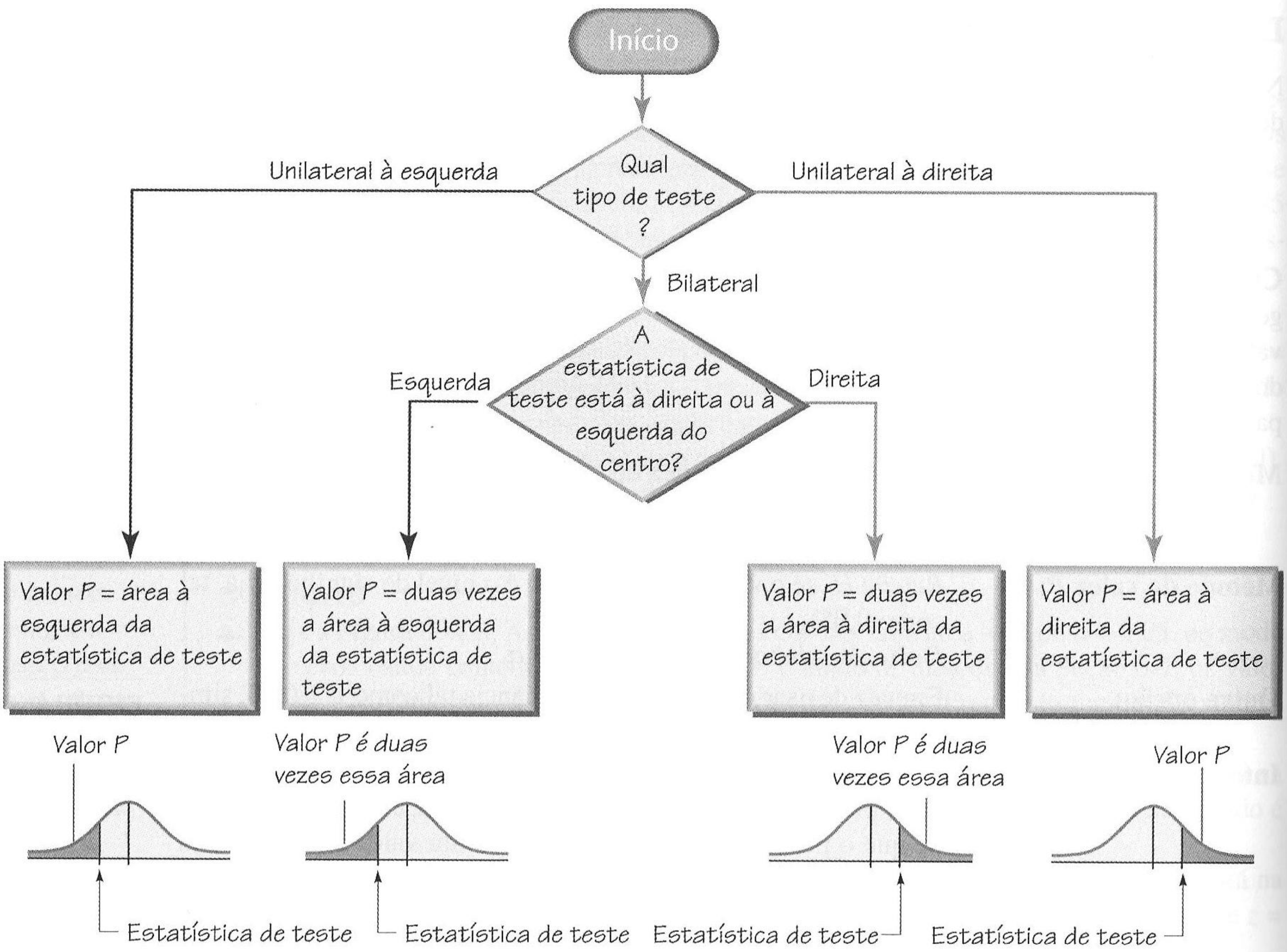
– **Valor *P* (ou valor *p* ou valor de probabilidade)** é a probabilidade de se obter um valor da estatística de teste que seja, no mínimo, tão extremo quanto aquele que representa os dados amostrais, supondo que a hipótese nula seja verdadeira.

– Hipótese nula é rejeitada se valor *P* for muito pequeno, por exemplo, igual ou menor a 0,05.

– Pequeno valor *P* indica que resultados amostrais têm pouca chance de ocorrer por acaso. Ou seja, dados apresentam tendência, rejeitando H_0.

– Podemos ainda pensar esse valor como sendo a probabilidade da hipótese nula não ser rejeitada.

PROCEDIMENTO PARA DETERMINAR VALORES *P*



DECISÕES E CONCLUSÕES

– Nosso procedimento padrão de teste de hipótese requer que testemos sempre a hipótese nula, de modo que nossa **conclusão inicial** será sempre uma das seguintes:

– Rejeitar a hipótese nula.

– Deixar de rejeitar a hipótese nula.

– A **decisão** de rejeitar ou não rejeitar H_0 é feita com:

– Método tradicional (clássico).

– Método do valor *P* (método mais usado atualmente).

– Intervalos de confiança.

CRITÉRIO DE DECISÃO

– **Método tradicional (clássico):**

  – Rejeite H_0: se estatística de teste ficar dentro da região crítica.

  – Deixe de rejeitar H_0: se estatística de teste não ficar dentro da região crítica.

– **Método do valor *P*:**

  – Rejeite H_0: se valor $P \leq \alpha$ (α é o nível de significância).

  – Deixe de rejeitar H_0: se o valor $P > \alpha$.

– **Outra opção:** em vez de usar valor para α, indique valor *P*.

– **Intervalos de confiança:** rejeite afirmativa de que parâmetro populacional tenha um valor que não esteja no IC.

REDAÇÃO DA CONCLUSÃO FINAL

– Devemos usar termos simples (não-técnicos) para escrever a conclusão final sobre o teste de hipótese.

– Se você deseja **apoiar** uma afirmativa, formule-a para ser a hipótese alternativa, de modo a **rejeitar** a hipótese nula.

– Alguns textos dizem "aceitar a hipótese nula" em vez de "deixar de rejeitar a hipótese nula":

  – Porém, devemos saber que não estamos provando H_0.

  – Termo "aceitar" é enganoso, pois implica que H_0 foi provada.

  – "Deixar de rejeitar" é mais apropriado, pois dizemos que evidência amostral não é forte o bastante para rejeitar H_0.

EVITE NEGATIVAS MÚLTIPLAS

– Em vez de dizer:

  – **Não** há evidência suficiente para garantir a **rejeição** da afirmativa de **nenhuma** diferença entre 0,5 e a proporção populacional.

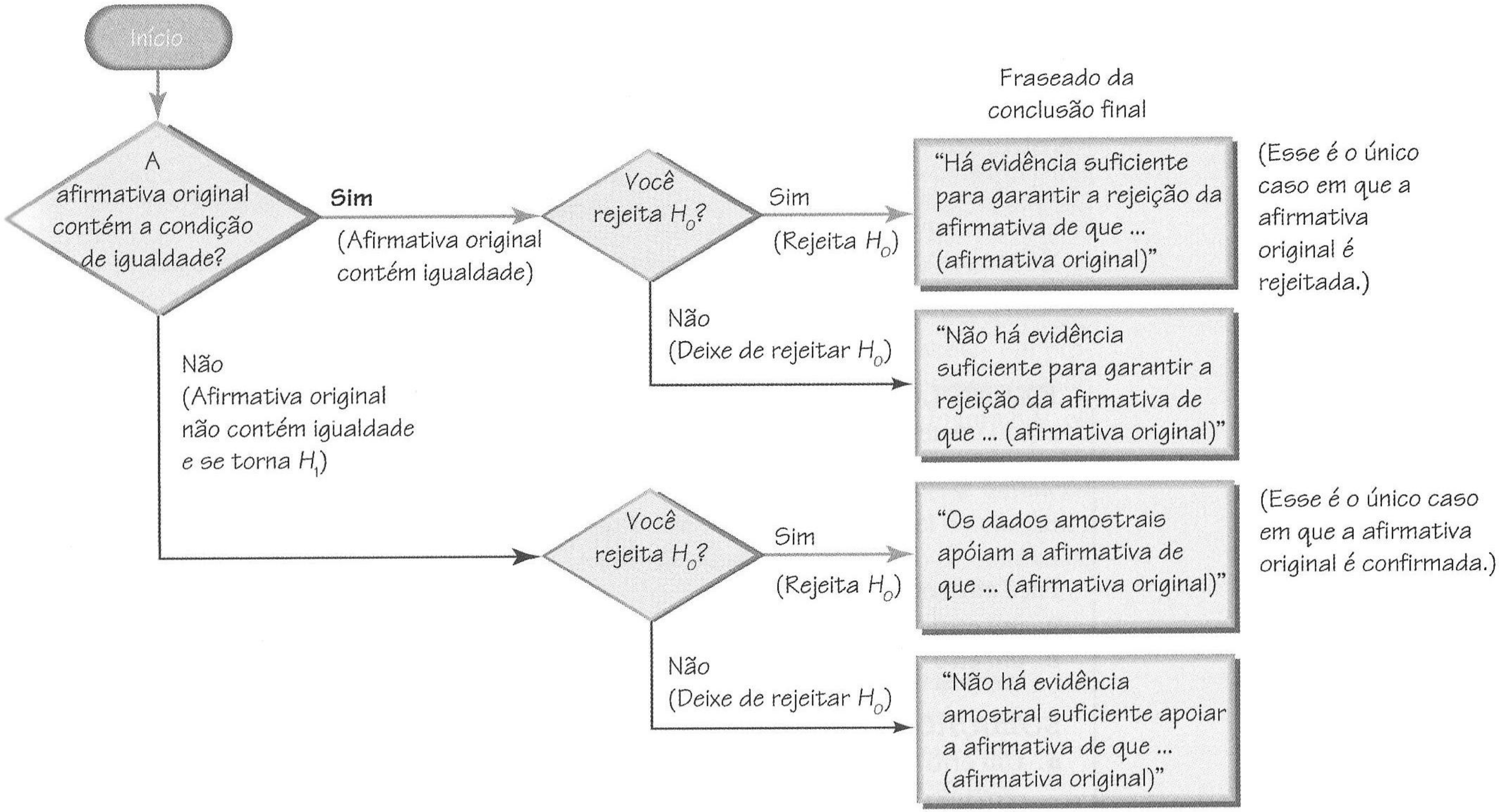
– Seria melhor usar:

  – Deixa-se de **rejeitar** a afirmativa de que a proporção populacional seja igual a 0,5.

ou

  – Até que se obtenha evidência mais forte, continuamos admitindo que a proporção populacional seja igual a 0,5.

PROCEDIMENTO PARA ESCREVER CONCLUSÃO FINAL



ERROS TIPO I E TIPO II

– Ao testar H_0, chegamos a uma conclusão de rejeitá-la ou de deixar de rejeitá-la.

– Tais conclusões pode estar corretas ou erradas.

		Estado verdadeiro da natureza	
		A hipótese nula é verdadeira	A hipótese nula é falsa
Decisão	Decidimos rejeitar a hipótese nula.	**Erro tipo I** (rejeitar uma hipótese nula verdadeira) α	Decisão Correta
	Deixamos de rejeitar a hipótese nula	Decisão Correta	**Erro tipo II** (deixar de rejeitar uma hipótese nula falsa) β

– α: probabilidade de erro tipo I (probabilidade de rejeitar hipótese nula quando ela é verdadeira).

– β: probabilidade de erro tipo II (probabilidade de deixar de rejeitar hipótese nula quando ela é falsa).

CONTROLE DOS ERROS TIPO I E TIPO II

– No procedimento para teste de hipóteses, selecionamos um nível de significância (*α*), que é a probabilidade de rejeitar a hipótese nula quando ela é verdadeira (erro tipo I).

– Porém, não selecionamos (*β*), que é a probabilidade de deixar de rejeitar H_0 quando ela é falsa (erro tipo II).

– Alfa (***α***), beta (***β***) e tamanho amostral (***n***) estão relacionados: se determinamos dois deles, o terceiro está determinado.

– Geralmente selecionamos primeiro *α* e *n*:

  – Para qualquer *α* fixo, aumento em *n* causará diminuição em *β*.

  – Para qualquer *n* fixo, diminuição em *α* causará aumento em *β* e vice-versa.

  – Para diminuir *α* e *β*, aumente *n*.

TESTE DE HIPÓTESE ABRANGENTE

– Foram descritos componentes individuais de um teste de hipótese.

– Podemos testar afirmativas sobre parâmetros populacionais com:

  – **Método do valor *P***.

  – **Método tradicional**.

  – **Método do intervalo de confiança**.

MÉTODO DO VALOR *P*

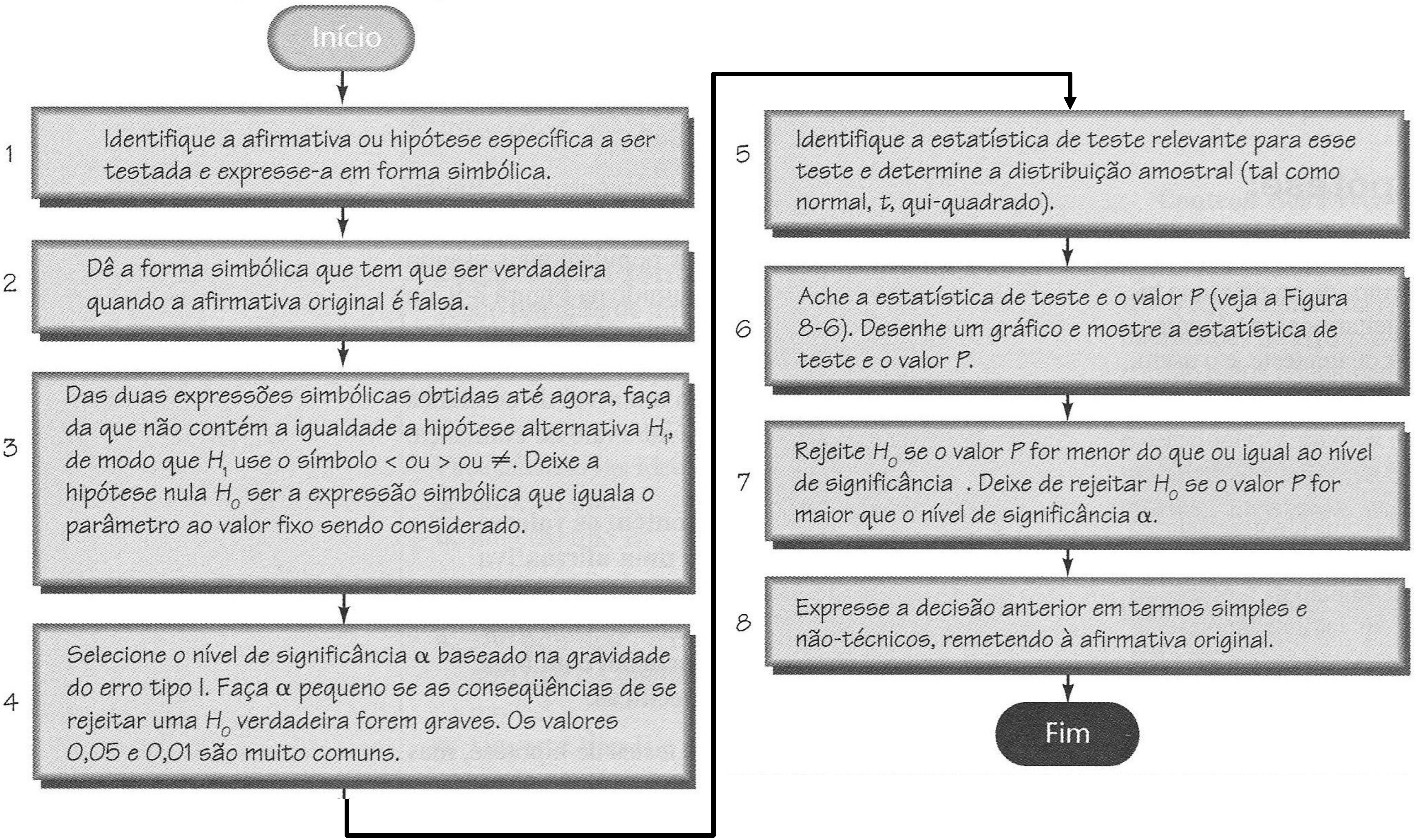
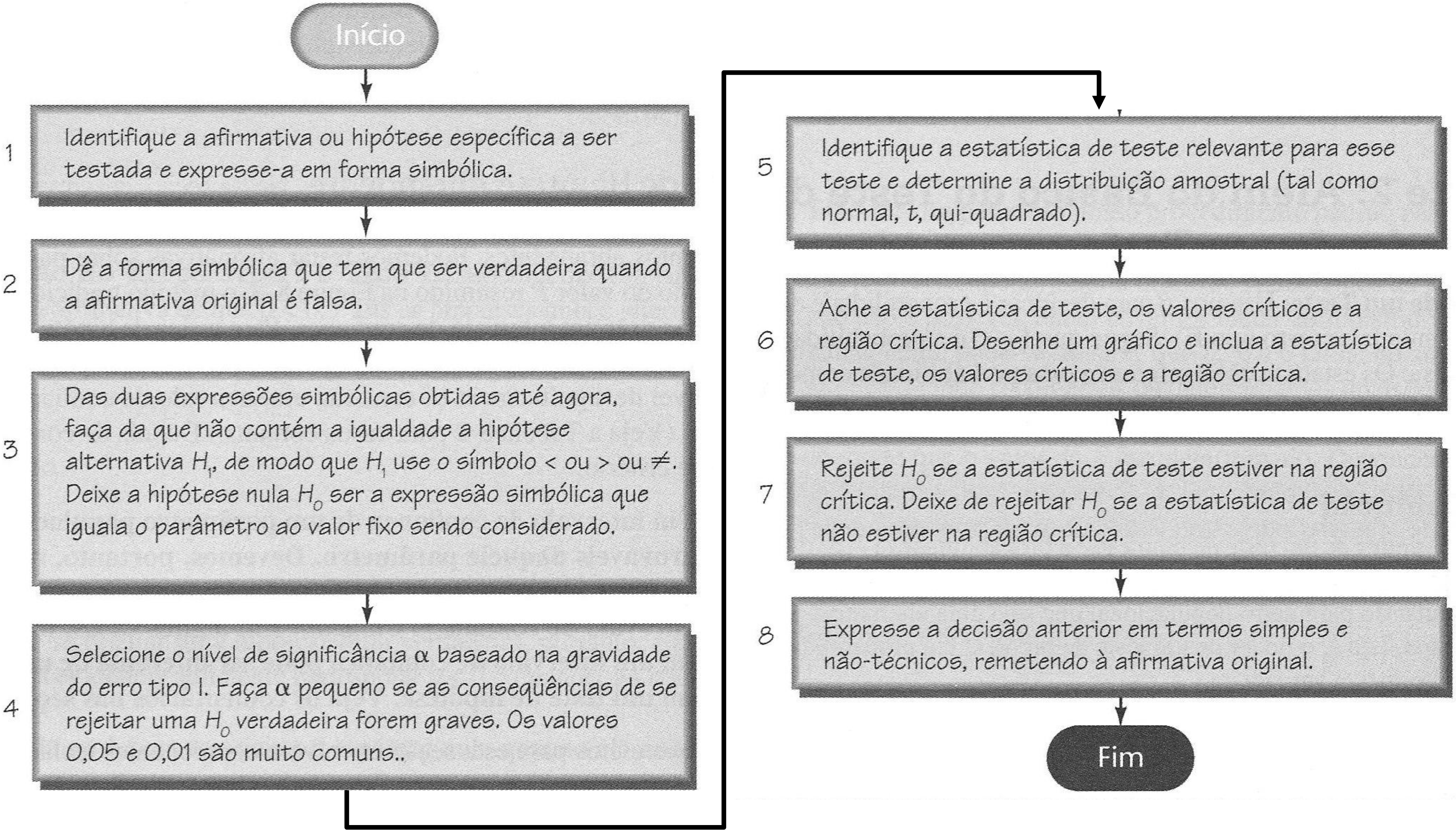


Figura 8-6 = Slide 17

MÉTODO TRADICIONAL



MÉTODO DO INTERVALO DE CONFIANÇA

– Construa um intervalo de confiança (IC) com o nível de confiança (α) selecionado.

Tabela 8-2 Nível de Confiança para o Intervalo de Confiança

		Teste Bilateral	Teste Unilateral
Nível de Significância para o Teste de Hipótese	0,01	99%	98%
	0,05	95%	90%
	0,10	90%	80%

– A estimativa de intervalo de confiança de um parâmetro populacional contém os valores prováveis do parâmetro.

– Rejeite uma afirmativa de que o parâmetro populacional tem um valor que não está incluído no intervalo de confiança.

O PODER DE UM TESTE

– Usamos *β* para designar a probabilidade de deixar de rejeitar uma hipótese nula falsa (**erro tipo II**).

– **Poder de um teste** de hipótese é a probabilidade (1–*β*) de se rejeitar uma hipótese nula falsa.

  – Essa probabilidade é calculada usando um **nível de significância** específico (*α*) e um valor particular do parâmetro populacional que seja uma alternativa (H_1) ao valor assumido na hipótese nula (H_0).

– O **poder de um teste** de hipótese é a probabilidade de se apoiar uma hipótese alternativa (H_1) verdadeira.

– **Dependendo dos valores** particulares escolhidos como alternativos à hipótese nula, poder do teste será diferente.

– Geralmente é exigido poder de teste entre 0,8 e 0,9.

TAMANHO DA AMOSTRA E PODER DE TESTE NO STATA

– Utilize o comando:

sampsi #1 #2, sd(#) alpha(#) power(#) n(#) onesamp

– *#1*: média na população (hipótese nula).

– *#2*: média alternativa (hipótese alternativa).

– *sd*: desvio padrão da população.

– *alpha*: nível de significância adotado.

– *power*: poder de teste.

– *n*: tamanho da amostra.

– *onesamp*: teste de uma amostra.

DEFININDO TAMANHO DA AMOSTRA (*n*)

- Uma pesquisa verificou 40% de intenção de voto no candidato A, com desvio padrão de 10%. Hipótese alternativa é que a pesquisa subestimou intenção de voto em 5%. Qual o tamanho da amostra a ser coletada para que H_1 seja provada com margem confiável?
- H_0: V_A=40%
- H_1: V_A=45%
- Desvio padrão=10%
- Utilizamos: α=0,05 (prob. rejeitar H_0 quando é verdadeira)
- Utilizamos: (1–β)=0,90 (prob. rejeitar uma H_0 falsa).

sampsi #1 #2, sd(#) alpha(#) power(#) n(#) onesamp

sampsi 40 45, sd(10) alpha(.05) power(.9) onesamp

RESULTADO DO TAMANHO DA AMOSTRA (*n*)

```
. sampsi 40 45, sd(10) alpha(0.05) power(0.9) onesamp

Estimated sample size for one-sample comparison of mean
  to hypothesized value

Test Ho: m =     40, where m is the mean in the population

Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9000
 alternative m =       45
            sd =       10

Estimated required sample size:

             n =       43
```

– Quanto maior desvio padrão, maior *n*.

– Quanto maior α, menor *n*.

– Quanto maior poder de teste ($1–\beta$), maior *n*.

– Quanto maior diferença entre H_0 e H_1, menor *n*.

INTERPRETAÇÃO DO TAMANHO DA AMOSTRA (*n*)

– O resultado indica que, com nível de significância de 0,05 e poder de teste de 90%, seriam necessárias 43 entrevistas selecionadas aleatoriamente para detectar um aumento da intenção de voto no candidato A de 40% para 45%.

DEFININDO PODER DE TESTE (1–*β*)

- Uma pesquisa verificou 40% de intenção de voto no candidato A, com desvio padrão de 10%. Hipótese alternativa é que a pesquisa subestimou intenção de voto em 5%. Se testamos essa pesquisa com uma amostra de tamanho 20, qual o poder de teste neste caso?
- H_0: V_A=40%
- H_1: V_A=45%
- Desvio padrão=10%
- Tamanho da amostra (n)=20
- Utilizamos: α=0,05 (prob. rejeitar H_0 quando é verdadeira)

sampsi #1 #2, sd(#) alpha(#) power(#) n(#) onesamp

sampsi 40 45, sd(10) alpha(.05) n(20) onesamp

RESULTADO DO PODER DE TESTE (1–β)

```
. sampsi 40 45, sd(10) alpha(.05) n(20) onesamp

Estimated power for one-sample comparison of mean
  to hypothesized value

Test Ho: m =     40, where m is the mean in the population

Assumptions:

         alpha =   0.0500  (two-sided)
 alternative m =       45
            sd =       10
 sample size n =       20

Estimated power:

         power =   0.6088
```

– Quanto maior desvio padrão, menor poder de teste.

– Quanto maior *α*, maior poder de teste.

– Quanto maior *n*, maior poder de teste.

– Quanto maior diferença entre H_0 e H_1, maior poder de teste.

INTERPRETAÇÃO DO PODER DE TESTE (1–β)

– O resultado indica que, com nível de significância de 0,05 e 20 entrevistas selecionadas aleatoriamente, a pesquisa teria um poder de teste de 61% para detectar um aumento da intenção de voto no candidato A de 40% para 45%.

TESTE DE UMA AFIRMATIVA SOBRE UMA PROPORÇÃO

PROPORÇÃO POPULACIONAL (*prtest*)

– **Requisitos** para testar afirmativas sobre uma proporção populacional *p*:

  – Amostra aleatória simples.

  – Distribuição binomial satisfeita (número fixo de tentativas independentes tendo probabilidades constantes; duas categorias de resultados).

  – Distribuição binomial das proporções amostrais pode ser aproximada por uma distribuição normal ($np \geq 5$ e $nq \geq 5$).

– **Utilize**:

  – Valores *P*: distribuição normal padrão.

  – Valores críticos: distribuição normal padrão.

  – Estatística de teste: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$

TESTE DE UMA AFIRMATIVA SOBRE UMA MÉDIA: σ CONHECIDO

MÉDIA POPULACIONAL COM σ CONHECIDO

– **Requisitos** para testar afirmativas sobre uma média populacional com σ conhecido:

  – Amostra aleatória simples.

  – Valor do desvio padrão populacional σ é conhecido.

  – População é normalmente distribuída e/ou n>30.

– **Utilize**:

  – Valores P: distribuição normal padrão.

  – Valores críticos: distribuição normal padrão.

  – Estatística de teste:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

TESTE DE UMA AFIRMATIVA SOBRE UMA MÉDIA: σ DESCONHECIDO

MÉDIA POPULACIONAL COM *σ* DESCONHECIDO (*ttest*)

– **Requisitos** para testar afirmativas sobre uma média populacional com *σ* desconhecido:

– Amostra aleatória simples.

– Valor do desvio padrão populacional *σ* não é conhecido.

– População é normalmente distribuída e/ou $n>30$.

– **Utilize**:

– Valores *P*: distribuição *t* de Student, com (n–1) graus de liberdade (gl).

– Valores críticos: distribuição *t* de Student, com (n–1) graus de liberdade (gl).

– Estatística de teste:

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

TESTE DE UMA AFIRMATIVA SOBRE UM DESVIO PADRÃO OU UMA VARIÂNCIA

DESVIO PADRÃO OU VARIÂNCIA (*sdtest*)

– **Requisitos** para testar afirmativas sobre um desvio padrão ou uma variância:

- – Amostra aleatória simples.
- – População é normalmente distribuída (exigência mais estrita do que a exigência de normalidade para médias).

– **Utilize**:

- – Valores *P*: distribuição qui-quadrado, com (n–1) graus de liberdade (gl).
- – Valores críticos: distribuição qui-quadrado, com (n–1) graus de liberdade (gl).
- – Estatística de teste:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

APLICAÇÃO NO STATA

PROPORÇÕES

– Com base na amostra do *World Values Survey*, a proporção de homens na população é igual a 50% (hipótese nula)?

gen homem=x001

replace homem=0 if x001==2

prtest homem=.5, level(95)

```
. prtest homem=.5, level(95)

One-sample test of proportion                  homem: Number of obs =    79946
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       homem |   .4969604   .0017683                      .4934946    .5004263
------------------------------------------------------------------------------
    p = proportion(homem)                                         z =  -1.7188
Ho: p = 0.5

     Ha: p < 0.5                 Ha: p != 0.5                  Ha: p > 0.5
 Pr(Z < z) = 0.0428         Pr(|Z| > |z|) = 0.0856          Pr(Z > z) = 0.9572
```

– Probabilidade de não rejeitar H_0 é pequena [Pr(Z<z)=0,04]. Isso indica que proporção de homens é menor que 0,5.

MÉDIAS COM σ DESCONHECIDO

– Com base na amostra do *World Values Survey*, a média do índice de valores racionais (tradicional/secular) é igual a 0,2250021 (hipótese nula)?

sum tradrat5

ttest tradrat5==.2250021, level(95)

```
. sum tradrat5

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    tradrat5 |      81589    .2250021    .8854222  -.9999222   3.844721

. ttest tradrat5==.2250021, level(95)

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
tradrat5 |   81589    .2250021    .0030998    .8854222    .2189265    .2310777
------------------------------------------------------------------------------
    mean = mean(tradrat5)                                         t =  -0.0000
Ho: mean = .2250021                              degrees of freedom =    81588

 Ha: mean < .2250021          Ha: mean != .2250021           Ha: mean > .2250021
 Pr(T < t) = 0.5000         Pr(|T| > |t|) = 1.0000          Pr(T > t) = 0.5000
```

– Probabilidade de não rejeitar H_0 é grande em todas situações, por isso não rejeitamos a hipótese nula.

DESVIOS PADRÕES

– Com base na amostra do *World Values Survey*, o desvio padrão do índice de valores racionais (tradicional/secular) é igual a 1 (hipótese nula)?

sdtest tradrat5==1, level(95)

```
. sdtest tradrat5==1, level(95)

One-sample test of variance
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
tradrat5 |   81589    .2250021    .0030998    .8854222    .2189265    .2310777
------------------------------------------------------------------------------
    sd = sd(tradrat5)                                        c = chi2 =  6.4e+04
Ho: sd = 1                                        degrees of freedom =    81588

     Ha: sd < 1                 Ha: sd != 1                  Ha: sd > 1
  Pr(C < c) = 0.0000       2*Pr(C < c) = 0.0000          Pr(C > c) = 1.0000
```

– Probabilidade de não rejeitar H_0 é pequena [Pr(C<c)=0,00]. Então, rejeitamos H_0. Ou seja, isso é evidência de que desvio padrão do índice é menor do que 1.