UNIVERSIDADE ESTADUAL DE MARINGÁ

BIOESTATÍSTICA

Teste de Hipótese

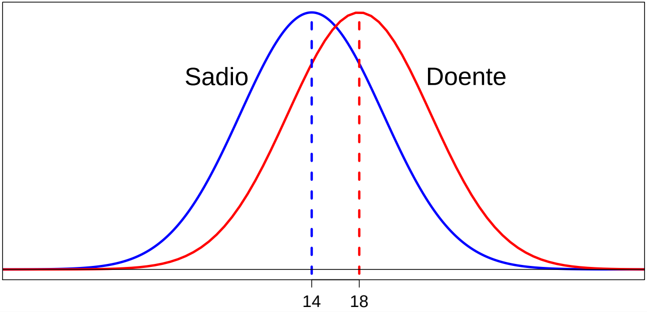
Isolde Previdelli
Omar Pereira

4 de maio de 2015

EXEMPLO

Para estudarmos sobre testes de hipóteses, seguiremos, em partes, o Capítulo 8 (Inferência Estatística - Testes de Hipóteses) do livro Noções de Probabilidade e Estatística.

- Suponha duas populações, uma composta de pessoas sadias e outra de doentes e que seguem distribuição Normal com médias 14 e 18 para uma determinada substância no sangue, ambas com desvio padrão igual a 6.



EXEMPLO

- Deseja-se saber se um certo tratamento é eficaz para combater a doença.
- Uma amostra aleatória com $n = 30$ foi selecionada entre os doentes e foram submetidos ao tratamento.
- Se a amostra fornecer valores próximos de 18, teremos evidências de que o tratamento não é eficaz.
- Se os valores fornecidos pela amostra forem próximos de 14, haverá evidências de que o tratamento produziu resultados satisfatórios.
- Estudaremos este problema por meio do *teste de hipóteses para média com variância conhecida*.

TESTES PARA A MÉDIA POPULACIONAL

- No exemplo, o interesse consiste em **testar** se a média populacional μ é **igual a 14** (população sadia) **contra** a alternativa de ser **igual a 18** (doentes).
- Para $n = 30$, a média amostral terá **distribuição** $N(\mu, 36/30)$.
- Para decidirmos sobre sobre o valor de μ, podemos determinar um **valor crítico**, x_c, tal que, se $\overline{X}$ for maior que x_c a amostra pertence à população com média $\mu = 18$, ou seja, o tratamento não é eficaz.
- Se a **média amostral for menor** que x_c, concluímos que a amostra pertence à população com média $\mu = 14$, e o **tratamento será considerado eficaz**.
- As **hipóteses** sobre a eficácia do tratamento são denotadas por H_0 e H_a, e denominadas **hipótese nula** e **hipótese alternativa**.

HIPÓTESES

Podemos escrever as hipóteses nula e alternativa como

H_0: O tratamento não é eficaz

H_a: O tratamento é eficaz

Podemos reescreve-lás como

H_0: $\mu = 18$

H_a: $\mu = 14$

Neste caso, as hipóteses não contêm **desigualdades** e são denominadas **hipóteses simples**. Entretanto, é comum o uso de **hipóteses compostas**, que ainda podem ser classificadas como **unilaterais** ou **bilaterais**, dependendo do interesse do estudo.

HIPÓTESES

No caso do tratamento ser eficaz, é razoável assumirmos que ele foi capaz de fazer com que os indivíduos amostrados mudassem para uma população cuja média é inferior a 18. Caso contrário, se o tratamento é ineficaz, μ não se alteraria. Neste caso, temos um teste de hipóteses unilateral.

H_0: $\mu = 18$

H_a: $\mu < 18$

Para verificarmos se o tratamento produz algum efeito, seja ele benéfico ($\mu < 18$) ou danoso ($\mu > 18$), devemos construir um teste de hipóteses bilateral

H_0: $\mu = 18$

H_a: $\mu \neq 18$

Por conveniência técnica, deixamos a igualdade na hipótese nula.

TIPOS DE ERROS

Os dois erros que podem ser cometidos ao se realizar um teste de hipóteses são:

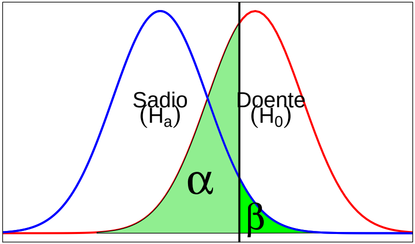
- Rejeitar a hipótese H_0, quando tal hipótese é verdadeira.
- Não rejeitar a hipótese H_0 quando ela deveria ser rejeitada.

Tabela de decisão

Realidade (desconhecida)	Decisão do teste	
	Não rejeita H_0	Rejeita H_0
H_0 verdadeira	Decisão correta ($prob = 1 - \alpha$)	Erro tipo I ($prob = \alpha$)
H_0 falsa	Erro tipo II ($prob = \beta$)	Decisão correta ($prob = 1 - \beta$)

- P(Erro Tipo I) $=$ P(rejeitar $H_0 | H_0$ é verdadeira) $= \alpha$
- P(Erro Tipo II) $=$ P(não rejeitar $H_0 | H_0$ é falsa) $= \beta$
- Poder do teste $=$ P(Rejeitar $H_0 | H_0$ é falsa) $= 1 - \beta$

REPRESENTAÇÃO GRÁFICA DE α e β



- $\alpha =$ P(concluir que o Trat **é eficaz** quando na verdade ele **NÃO É**)
- $\beta =$ P(concluir que o Trat **não é eficaz** quando na verdade ele **É**)

DETERMINAÇÃO DO VALOR CRÍTICO

Supondo α conhecido, vamos determinar o valor crítico x_c.

$$\begin{aligned}
\alpha &= P(\textit{erro tipo I}) \\
&= P(\textit{rejeitar } H_0 \mid H_0 \textit{ verdadeira}) \\
&= P(\overline{X} < x_c \mid \mu = 18) \\
&= P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{x_c - 18}{6/\sqrt{30}}\right) \\
&= P(Z < z_c)
\end{aligned}$$

com $Z \sim N(0,\ 1)$. Dado α obtemos z_c e calculamos x_c. Temos que

$$z_c = \frac{x_c - 18}{6/\sqrt{30}} \quad \Rightarrow \quad x_c = 18 + z_c \frac{6}{\sqrt{30}}$$

Por exemplo, para $\alpha = 0.05$ temos

$$0.05 = P(Z < z_c) \quad \Rightarrow \quad z_c = -1.64$$

Logo

$$x_c = 18 - 1.64 \frac{6}{\sqrt{30}} = 16.20$$

DETERMINAÇÃO DO VALOR CRÍTICO

- Uma vez escolhida a amostra, **se a estimativa** $\overline{x}_{obs}$ é tal que $\overline{x}_{obs} < 16.20$, **rejeitamos a hipótese nula** concluindo que o **tratamento é eficaz**.
- A região dada pelo conjunto dos números reais menores que 16.20 é denominada de **Região de Rejeição** ou **Região Crítica** (RC). Ou seja,

$$RC = \{x \in \mathbb{R} : x < 16.20\}$$

- Denominamos **Região de Aceitação** (RA) ao complementar de RC.
- Por exemplo, se a amostra obtida forneceu a estimativa $\overline{x}_{obs} = 16.04$ que pertence à RC, rejeitamos H_0 ao nível de significância $\alpha = 0.05$.

TESTE BILATERAL

A construção de testes de hipóteses bilaterais é feita de maneira similar ao caso unilateral, exceto que, agora, devemos considerar uma Região de Rejeição composta de duas partes disjuntas. Suponha que μ_0 seja uma constante conhecida, então as hipóteses são expressas como

H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$

A Região Crítica será dada por

$$RC = \{x \in \mathbb{R} : x < x_{c_1} \ ou \ x > x_{c_2}\}$$

e para um dado valor de α, determinamos x_{c_1} e x_{c_2} de modo que

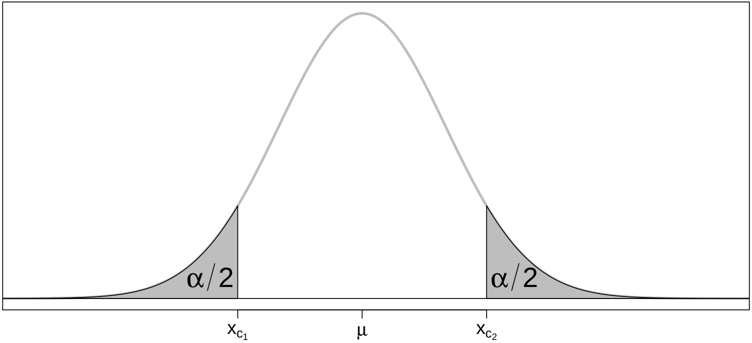
$$P(\overline{X} < x_{c_1} \ ou \ \overline{X} > x_{c_2}) = \alpha$$

TESTE BILATERAL

Dada a simetria da Normal, distribuímos a massa de α igualmente entre as duas partes da Região de Rejeição,

$$P(\overline{X} < x_{c_1}) = \frac{\alpha}{2} \quad e \quad P(\overline{X} > x_{c_2}) = \frac{\alpha}{2}$$

Região de Rejeição Bilateral



EXEMPLO

Agora faremos um teste de hipóteses bilateral e calcularemos a probabilidade do erro tipo II.

- Um experimento para determinar o tempo de reação de seres vivos a um estimulo elétrico sob o efeito de uma determinada substância foi realizado. Os valores obtidos foram: 9.1, 9.3, 7.2, 7.5, 13.3, 10.9, 7.2, 9.9, 8.0, 8.6 segundos. Admite-se que o tempo de reação, em geral, tem distribuição Normal com média 8 e desvio padrão igual a 2 segundos. O pesquisador desconfia que o tempo médio sofre alteração por influência da substância. Neste caso as hipóteses são:

H_0 : as cobaias apresentam tempo de reação padrão
H_1 : as cobaias têm o tempo de reação alterado

Estas hipóteses podem ser escritas como

$$H_0 : \mu = 8.0$$
$$H_1 : \mu \neq 8.0$$

EXEMPLO

Uma vez que o teste envolve a média populacional, consideramos a média amostral $\overline{X}$ para construir a estatística de teste e usamos que $\overline{X} \sim N(\mu,\ 4/10)$. A Região Crítica será da forma

$$RC = \{x \in \mathbb{R} : x < x_{c_1}\ \ ou\ x > x_{c_2}\}$$

Fixando $\alpha = 0.06$ temos:

$$\begin{aligned}
0.06 &= P(erro\ \ tipo\ \ I) \\
&= P(rejeitar\ \ H_0 \mid H_0\ \ verdadeira) \\
&= P(\overline{X}\ <\ x_c \mid \mu = 18) \\
&= P(\overline{X} \in RC \mid \mu = 8.0) \\
&= P(\overline{X} < x_{c_1}\ \ ou\ \overline{X} > x_{c_2} | \mu = 8.0) \\
&= P\left(\frac{\overline{X} - 8.0}{\sqrt{4/10}} < \frac{x_{c_1} - 8.0}{\sqrt{4/10}} \quad ou \quad \frac{\overline{X} - 8.0}{\sqrt{4/10}} > \frac{x_{c_2} - 8.0}{\sqrt{4/10}}\right) \\
&= P(Z < z_{c_1}\ \ ou\ \ Z > z_{c_2})
\end{aligned}$$

em que $z_{c_j} = (x_{c_j} - 8.0)/\sqrt{4/10}$, com $j = 1,\ 2$ e $Z \sim N(0,\ 1)$.

EXEMPLO

Dai segue que $z_{c_1} = -1.88$ e $z_{c_2} = 1.88$. Logo

$$x_{c_1} = 8.0 - 1.88\sqrt{4/10} = 6.8$$

$$x_{c_2} = 8.0 + 1.88\sqrt{4/10} = 9.2$$

Então podemos expressar a Região Crítica para $\alpha = 0.06$ como

$$RC = \{x \in \mathbb{R} : x < 6.8 \ ou \ x > 9.2\}$$

Calculando a **média amostral** obtemos $\overline{x}_{obs} = 9.1$. Como este valor **não pertence** à RC, **não rejeitamos** a hipótese H_0 ao nível de significância de 6%. Concluímos que o **tempo de reação não fica alterado**.

ERRO TIPO II - β

Vamos calcular a probabilidade do erro tipo II (β). Note uqe para calcular α, μ está bem especificado, o que não é o caso para o erro tipo II. Como a hipótese alternativa é composta, existem diversos valores possíveis para μ. Dessa forma, β será função de qual valor de μ foi escolhido dentro da região definida pela hipótese H_a. Neste caso, a probabilidade do erro tipo II será denotada por $\beta(\mu)$. Por exemplo, para $\mu = 9.0$ teríamos

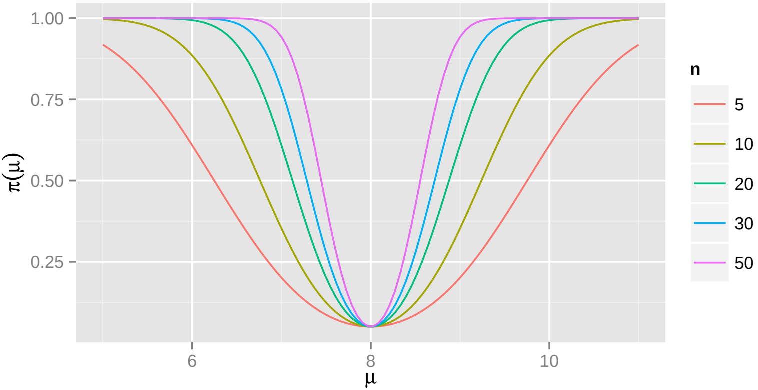
$$
\begin{aligned}
\beta(9.0) &= P(\textit{erro tipo II}) \\
&= P(\textit{Não rejeitar } H_0 \mid H_0 \textit{ falsa}) \\
&= P(\overline{X} \notin RC \mid \mu = 9.0) \\
&= P(6.8 \leq \overline{X} \leq 9.2 \mid \mu = 9.0) \\
&= P\left(\frac{6.8 - 9.0}{\sqrt{4/10}} \leq \frac{\overline{X} - 9.0}{\sqrt{4/10}} \leq \frac{9.2 - 9.0}{\sqrt{4/10}}\right) \\
&= P(-3.48 \leq Z \leq 0.32) \\
&= 0.4997 + 0.1255 \\
&= 0.6252
\end{aligned}
$$

Assim, com $\mu = 9.0$, e com probabilidade 0.6252 estaríamos concluindo, equivocadamente, que H_0 é verdadeira. Neste caso, o poder é baixo (0.3748)

PODER DO TESTE (1 - β)

Definimos a função poder do teste por $\pi(\mu) = P(\textit{rejeitar } H_0 \mid H_0 \textit{ falsa})$. Note que, se o valor de μ for aquele de H_0, $\pi(\mu)$ é igual ao nível de significância α. A função poder é apresentada abaixo para $n = 10$ e para outros valores de n.

Função Poder



UM POUCO MAIS SOBRE O PODER DO TESTE

Aqui trataremos um pouco sobre a função poder para o teste t.

Definimos poder de um teste estatístico como a probabilidade do teste rejeitar H_0 quando H_0 é realmente falsa, isto é, $1 - \beta$.

O poder do teste depende de alguns fatores

- Em geral, num experimento, a probabilidade do Erro do Tipo I é controlada (α).
- A probabilidade do Erro do Tipo II (e consequentemente o poder do teste) não é, em geral, controlada.
- Da variabilidade da população estudada.
- Do tamanho da amostra retirada.

MOTIVAÇÃO - TAMANHO DA AMOSTRA

EXEMPLO 1: Influência das dietas A e B na glicemia

- PROBLEMA: Uma nutricionista quer comparar duas diferentes dietas, A e B, no aumento da glicemia de seus pacientes. Sua hipótese é que a dieta A seja melhor que a dieta B, ou seja, o grupo alimentado com a dieta A (G1) terá menor aumento na concentração de glicose no sangue quando comparado com o grupo da dieta B (G2). No fim do experimento (que durou 6 semanas), a glicemia será medida. Ela espera que a diferença entre as médias dos dois grupos seja ao menos 10mg/dl. Assume-se que os desvios padrão serão 15 e 17 para G1 e G2.
- QUESTÃO: Qual o número de pacientes necessários em cada grupo, assumindo que os dois grupos terão o mesmo tamanho?

MOTIVAÇÃO - PODER DO TESTE

EXEMPLO 2: Efeito do sexo no tempo de ação de um fármaco no organismo

- PROBLEMA: Um pesquisador quer saber o efeito do sexo no tempo de ação de um determinado fármaco no organismo. Sua hipótese é que nas mulheres esse tempo seja maior que nos homens. Ele escolhe aleatoriamente 20 homens e 20 mulheres para participar do estudo.
- QUESTÃO: Qual o poder do teste baseado nos 40 sujeitos para detectar diferença entre os sexos?

DOIS ASPECTOS DO PODER DO TESTE

- **PRIMEIRO**: calcular o **tamanho da amostra** necessário para um poder de teste específico (Exemplo 1).

- **SEGUNDO**: calcular o **poder do teste** quando o tamanho da amostra é dado (Exemplo 2).

EXEMPLO 1 - TAMANHO DA AMOSTRA

INFORMAÇÕES

- Espera-se diferença entre as médias, neste caso 10mg/dl.
- O desvios padrão dos grupos, $\sigma_{G1} = 15$ e $\sigma_{G2} = 17$.
- O nível $\alpha = 5\%$, que é a probabilidade do erro Tipo I (de rejeitarmos H_0 quando ela é verdadeira), será assumido.
- O poder pré estabelecido para o cálculo do tamanho da amostra será 0.8.

NOTA

- Neste exemplo, as médias não foram especificadas (apenas a diferença entre elas).

EXEMPLO 1 - TAMANHO DA AMOSTRA

ANÁLISE NO R project

- Usaremos o **pwr** package.

  library(pwr)

  pwr.t.test($d = 10/(15^2 + 17^2)/2$, $power = .8, sig.level = .05$, type="two.sample",alternative="two.sided")

Two-sample t test power calculation
n = 41.31968
d = 0.6238303
sig.level = 0.05
power = 0.8
alternative = two.sided
NOTE: n is number in *each* group

- O cálculo resulta em 42 pacientes para o G1 e 42 para G2.

EXEMPLO 1 - SUPONDO $\alpha > 0.05$

QUAL O CUSTO DE 84 PACIENTES?

- E se 84 paciente estiver além do orçamento para a pesquisa?
- Um caminho para reduzir o tamanho da amostra é aumentar o Erro Tipo I (α).

VAMOS SUPOR $\alpha = 7\%$

- pwr.t.test($d = 10/(15^2 + 17^2)/2$, $power = .8, sig.level = .07$, type="two.sample",alternative="two.sided")

Two-sample t test power calculation
n = 37.02896
d = 0.6238303
sig.level = 0.07
power = 0.8
alternative = two.sided
NOTE: n is number in *each* group

- Neste caso, reduzimos para 38 o número de pacientes em cada tratamento.

EXEMPLO 1 - PODER DO TESTE

FIXANDO O TAMANHO DA AMOSTRA

- Suponha que a nutricionista tenha apenas 60 pacientes para estudar, ou seja, 30 em cada grupo.

SUPONDO $n = 30$ e $\alpha = 5\%$

- pwr.t.test($d = 10/(15^2 + 17^2)/2,\ n = 30, sig.level = .05$, type="two.sample",alternative="two.sided")

Two-sample t test power calculation
n = 30
d = 0.6238303
sig.level = 0.05
power = 0.6612888
alternative = two.sided
NOTE: n is number in *each* group

- Obtivemos *power* $= 0.6613$

EXEMPLO 1 - EFFECT SIZE

EFFECT SIZE

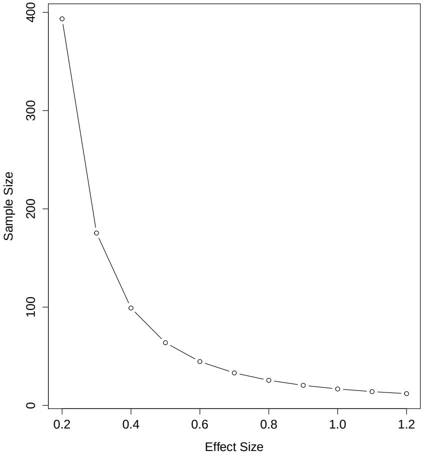
- O que realmente importa no cálculo do poder ou do tamanho da amostra é a diferença entre as médias e os desvios padrão.

VARIANDO A MÉDIA COM $SD = 1$

- ```
  ptab = cbind(NULL, NULL)
  for(i in seq(.2,1.2,.1)) {
  pwrt = pwr.t.test(d=i, power=.8, sig.level=.05,
  type ="two.sample", alternative = "two.sided")
  ptab = rbind(ptab, cbind(pwrt$d, pwrt$n)) }
  rownames(ptab) = c(1:11)
  colnames(ptab) = c("Effect Size", "Sample Size")
  ptab
  ```
  ```

EXEMPLO 1 - TAMANHO DA AMOSTRA

	Effect Size	Sample Size
1	0.2	393.40570
2	0.3	175.38467
3	0.4	99.08032
4	0.5	63.76561
5	0.6	44.58579
6	0.7	33.02458
7	0.8	25.52457
8	0.9	20.38633
9	1.0	16.71473
10	1.1	14.00190
11	1.2	11.94226

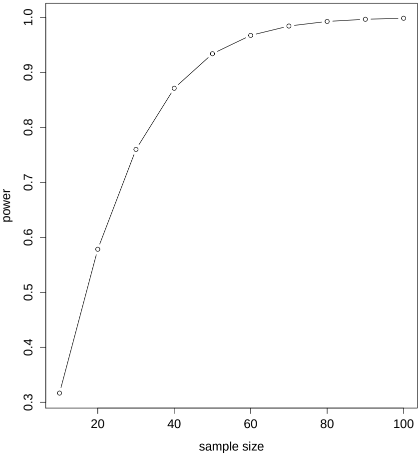


- Quanto menor o effect size, maior deve ser o tamanho da amostra.

EXEMPLO 1 - TAMANHO DA AMOSTRA

Também podemos gerar o gráfico do poder versus tamanho da amostra para um dado effect size, por exemplo, $d = 0.7$

- pwrt = pwr.t.test(d = .7, n = seq(10,100,10), sig.level = .05, type = "two.sample", alternative = "two.sided")
- plot(pwrt$n, pwrt$power, type = "b", xlab = "sample size", ylab = "power")



REFERÊNCIAS

- http://www.ats.ucla.edu/stat/r/dae/t_test_power2.htm
- MAGALHÃES, Marcos Nascimento; DE LIMA, Antonio Carlos Pedroso. Noções de probabilidade e estatística. IME-USP, 2000.
- PAGANO, Marcello; GAUVREAU, Kimberlee. Princípios de bioestatística. Thomson Learning, 2004.