Testes de Hipótese para uma única Amostra - parte I

Marcos Oliveira Prates

2012/02

Objetivos

Ao final deste capítulo você deve ser capaz de:

- Estruturar problemas de engenharia como testes de hipótese.
- Entender os conceitos de:
  - hipótese nula;
  - hipótese alternativa;
  - significância;
  - p-valor.

- Podemos estar interessados em aceitar ou rejeitar uma afirmação a cerca de um parâmetro.
- A afirmação é denominada uma **hipótese**.
- A tomada de decisão é denominado um **teste de hipótese**.
- Possui uma conexão com intervalo de confiança.

- Pode-se querer comparar a média da população com um valor especificado.
- Isso é chamado um **estudo comparativo**.
- Trataremos aqui apenas do caso com uma única população.
- Queremos testar afirmações sobre parâmetros dessa população.

Hipótese Estatística

Afirmação sobre o parâmetro de uma ou mais populações.

- Usamos distribuições para representar uma população.
- Uma hipótese é uma afirmação sobre essa distribuição de probabilidade.
- Geralmente envolve um ou mais parâmetros dessa distribuição.

Exemplo:

- Estamos analisando a taxa de queima de um propelente sólido.
- A taxa de queima é uma variável aleatória $\Rightarrow$ tem distribuição de probabilidade.
- Estamos interessados apenas na média dessa taxa.
- Queremos decidir se a taxa média é ou não 50 cm por segundo.
- Podemos escrever como

$$H_0 : \mu = 50 \text{ cm por segundo}$$

$$H_1 : \mu \neq 50 \text{ cm por segundo}$$

- A afirmação

$$H_0 : \mu = 50 \text{ cm por segundo}$$

  é chamada **hipótese nula.**
  (Sempre abrange a igualdade.)
- A afirmação

$$H_1 : \mu \neq 50 \text{ cm por segundo}$$

  é chamada **hipótese alternativa.**
  (Oposto da hipótese nula e representa a conclusão que será apoiada caso a H_0 seja rejeitada)
- Esse é um tipo de hipótese **bilateral.**

- Em alguns casos podemos formular hipóteses **unilaterais**:

$$H_0 : \mu = 50 \text{ cm por segundo}$$

$$H_1 : \mu < 50 \text{ cm por segundo}$$

ou

$$H_0 : \mu = 50 \text{ cm por segundo}$$

$$H_1 : \mu > 50 \text{ cm por segundo}$$

Observações:

- Hipóteses são afirmações sobre a população.
- Não são afirmações sobre a amostra.
- O valor do parâmetro especificado em H_0 (ex: 50) pode ser especificado de três formas:
  - de experimentos passados $\Rightarrow$ verificar se houve alteração;
  - a partir de uma teoria ou modelo $\Rightarrow$ verificar a teoria ou modelo;
  - considerações externar $\Rightarrow$ verificar se o produto obedece algumas especificações.

- Usamos a informação amostral para testar a hipótese.
- Se a informação da amostra é consistente com a hipótese $\Rightarrow$ não rejeitamos.
- Se a informação é inconsistente (pouco provável) $\Rightarrow$ rejeitamos.
- Só poderíamos saber se a hipótese é falsa ou não tivéssemos acesso a toda população.
- Temos apenas a probabilidade de tomar uma conclusão errada.

- Considere o exemplo da taxa de queima do propelente.
- Queremos testar

$$H_0 : \mu = 50 \text{ cm por segundo}$$

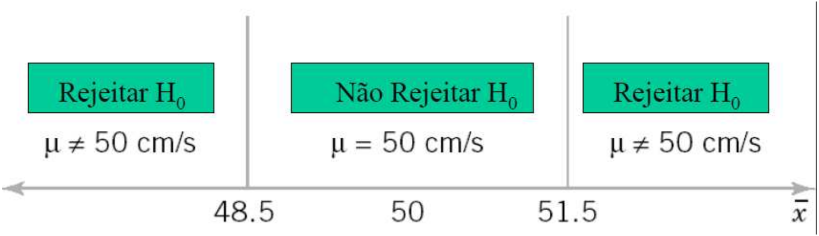
$$H_1 : \mu \neq 50 \text{ cm por segundo}$$

- Coletamos uma amostra de tamanho $n = 10$.
- A taxa média observada $\bar{x}$ é uma estimativa para μ.
- Se $\bar{x} \approx 50 \Rightarrow$ evidência de que $\mu = 50$.
- Se $\bar{x}$ é muito diferente de 50 $\Rightarrow$ evidência de que $\mu \neq 50$.
- Precisamos definir para quais valores de $\bar{x}$ não é rezoável que $\mu = 50 \Rightarrow$ **região crítica.**

- Uma possibilidade:

$$\text{se } 48,5 \leq \bar{x} \leq 51,5 \Rightarrow \text{ não rejeitamos } H_0$$

$$\text{se } \bar{x} < 48,5 \text{ ou } \bar{x} > 51,5 \Rightarrow \text{ rejeitamos } H_0$$



- Veremos mais a frente como determinamos a região crítica.

- Podemos tirar conclusões erradas.
- Dizer que $\mu \neq 50$ quando $\mu = 50$:
  - mesmo com $\mu = 50$, por azar, econtrarmos uma amostra tal que

$$\bar{x} < 48,5 \text{ ou } \bar{x} > 51,5 \,.$$

    (Erro tipo I)
- Dizer que $\mu = 50$ quando $\mu \neq 50$:
  - mesmo com $\mu \neq 50$, por azar, econtrarmos uma amostra tal que

$$48,5 \leq \bar{x} \leq 51,5 \,.$$

    (Erro tipo II)

Erro tipo I

Rejeitar H_0 quando ela é verdadeira. Denotado por α.

Erro tipo II

Deixar de rejeitar H_0 quando ela é falsa. Denotado por β.

Decisão	H_0 é verdadeira	H_0 é falsa
Não rejeita H_0	nenhum erro	erro tipo II
Rejeita H_0	erro tipo I	nenhum erro

- O erro tipo I é também chamado **nível se significância** ou **tamanho do teste.**
- No exemplo da taxa suponha que $\sigma = 2,5$

$$\alpha = P(\bar{x} < 48,5 \text{ ou } \bar{x} > 51,5|\mu = 50)$$

$$= P(\bar{x} < 48,5|\mu = 50) + P(\bar{x} > 51,5|\mu = 50)$$

$$= P\left(\frac{\bar{x} - 50}{2,5/\sqrt{10}} < \frac{48,5 - 50}{2,5/\sqrt{10}}\right) + P\left(\frac{\bar{x} - 50}{2,5/\sqrt{10}} > \frac{51,5 - 50}{2,5/\sqrt{10}}\right)$$

$$= P(Z < -1,90) + P(Z > 1,90) = 0,0574$$

- 5,74% das amostras levariam à rejeição de H_0 mesmo com $\mu = 50$.

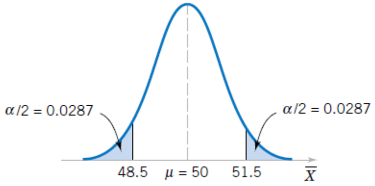


Figure 9-2 The critical region for H_0: $\mu = 50$ versus H_1: $\mu \neq 50$ and $n = 10$.

- Podemos dimunuir α aumentando a região de aceitação.
- Rejeitamos menos vezes.
- Tem menos probabilidade menor de errar ao rejeitar.
- Se rejeitamos H_0 quando

$$\bar{x} < 48 \text{ ou } \bar{x} > 52$$

temos que

$$\alpha = 0,0164 \, .$$

- Vamos analisar agora o erro do tipo II (β).
- Precisamos ter uma hipótese alternativa específica.
- Um valor particular de μ, exemplo:

$$\mu = 48 \quad \text{ou} \quad \mu = 52\,.$$

- Como o teste se funciona se queremos rejeitar H_0 para um valor de $\mu = 52$ ou $\mu = 48$?
- Pela simetria da normal os dois casos levam à mesma probabilidade de erro.

$$\beta = P\,(48,5 \leq \bar{x} \leq 51,5 | \mu = 52)$$

$$= P\left(\frac{48,5 - 52}{2,5/\sqrt{10}} \leq \frac{\bar{x} - 52}{2,5/\sqrt{10}} \leq \frac{51,5 - 52}{2,5/\sqrt{10}}\right)$$

$$P(-4,43 \leq Z \leq -0,63) = 0,263\,.$$

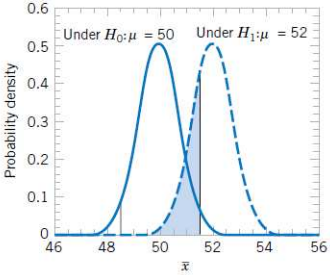


Figure 9-3 The probability of type II error when $\mu = 52$ and $n = 10$.

- Se estivéssemos testando $\mu = 50$ e $\mu \neq 50$.
- O valor verdadeiro de $\mu = 52$.
- A probabilidade de erramos ao rejeitar H_0 é 0,2643.

- Se o valor verdadeiro fica mais próximo do valor de $H_0 \Rightarrow$ a probabilidade de erro aumenta.
- É mais difícil diferenciar médias muito próximas.
- Se o valor real fosse $\mu = 50,5$ temos

$$\beta = 0,8923\,.$$

- Probabilidade de erro bem maior.

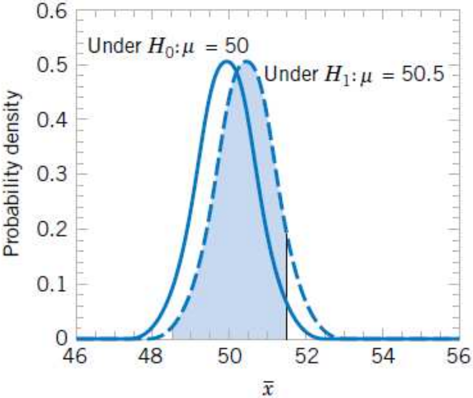


Figure 9-4 The probability of type II error when $\mu = 50.5$ and $n = 10$.

- O erro também depende do tamanho da amostra.
- Se aumentamos o tamanho da amostra $\Rightarrow$ mais informação.
- Teremos menor probabilidade de cometer erros.

- No exemplo anterior, se $n = 16$ e não 10.
- Temos que

$$\beta = P\,(48,5 \leq \bar{x} \leq 51,5|\mu = 52)$$

$$= P\left(\frac{48,5-52}{2,5/\sqrt{16}} \leq \frac{\bar{x}-52}{2,5/\sqrt{16}} \leq \frac{51,5-52}{2,5/\sqrt{16}}\right)$$

$$P(-5,60 \leq Z \leq -0,80) = 0,2119\,.$$

- O valor anterior era $\beta = 0,2643$.

Resumindo

- Selecionando valores críticos apropriados $\Rightarrow$ podemos diminuir a região crítica e probabilidade de erro tipo I.
- Os erros estão amarrados $\Rightarrow$ se um aumenta o outro diminui.
- Aumentando tamanho da amostra $\Rightarrow$ diminuímos as probabilidade de erro.
- Se H_0 é falsa e se o verdadeiro valor se aproxima do valor em $H_0 \Rightarrow \beta$ aumenta.
- Se H_0 é falsa e se o verdadeiro valor se afasta do valor em $H_0 \Rightarrow \beta$ diminui.

- O pesquisador controla o α quando seleciona os valores críticos.
- Rejeitar H_0 é uma conclusão forte.
- O β depende do verdadeiro valor do parâmetro e do tamanho da amostra.
- Aceitar H_0 é uma conclusão fraca.
- Preferimos dizer "deixar de rejeitar H_0" do que "aceitar H_0".
- Falhar em rejeitar H_0:
  - não encontramos evidências suficiente para rejeitar H_0;
  - não significa que haja alta probabilidade de que H_0 seja verdadeira.

Poder de um teste

Probabilidade de rejeitar H_0 quando H_0 é falsa.

- É a probabilidade de rejeitar **corretamente** uma hipótese nula.
- É calculada como

$$1 - \beta = 1 - P(\text{ deixar de rejeitar } H_0 | H_0 \text{ é falsa })$$

- Comparamos testes estatísticos comparando seu poder.

- Considere o exemplo da taxa de queima.
- Estávamos testando

$$H_0 : \mu = 50 \text{ cm por segundo}$$

$$H_1 : \mu \neq 50 \text{ cm por segundo}$$

- Suponha que o valor verdadeiro é $\mu = 52$.
- Se $n = 10$ vimos que

$$\beta = 0,2643.$$

- O poder é

$$1 - \beta = 1 - 0,2643 = 0,7357$$

quando $\mu = 52$.

- O poder é uma medida de sensibilidade do teste.
- Qual sua habilidade de detectar diferenças.
- Qual a sensibilidade do teste de detectar uma diferença entre $\mu = 50$ e $\mu = 52$.
- O poder é 0,7357.
- O teste detecta essa diferença em 73,57% das vezes.
- Se o poder é muito baixo, podemos:
    - aumentar o α;
    - aumentar o *n*.

Hipótese unilaterais e bilaterais

- A hipótese alternativa pode ser unilateral ou bilateral.
- Depende da conclusão a ser retirada se rejeitamos H_0.
- Se queremos fazer afirmações como:
  - maior que, inferior a, superior a
  - o teste é **unilateral**.
- Se não temos nenhuma direção (apenas diferença):
  - o teste é **bilateral**.

- Como podemos descrever um teste?
  - Dizendo se rejeitamos ou não H_0 e qual o nível de significância α.
- Dizemos que $H_0 : \mu = 50$ foi rejeitada com 0,05 de significância.
- Não dá informação sobre o valor calculado da estatística de teste.
- Estabelecemos o nível de significância antes de observarmos a amostra.
- Para evitar essas dificuldades podemos usar o **valor P** ou nível descritivo do teste.
- É a probabilidade de encontramos um valor tão ou mais extremo que a estatística de teste.

Valor P

Menor nível de significância que conduz à rejeição de H_0, com os dados fornecidos.

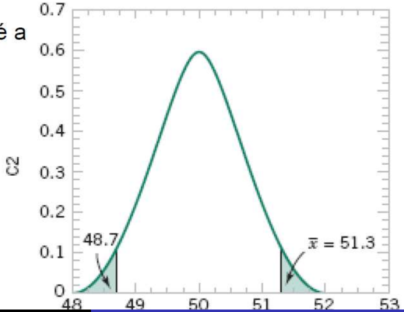
- Chamamos a estatística de teste de significante quando rejeitamos H_0.
- O valor P é o menor nível α em que a estatística é significante.
- Pode determinar quão significante os dados são.
- Valor P **pequeno** $\Rightarrow$ muito provável que H_0 é **falsa**.
- Valor P **grande** $\Rightarrow$ muito provável que H_0 é **verdadeira**.

- Considere o exemplo da taxa de queima.
- Suponha que $n = 16$ e $\sigma = 2,5$.
- Queremos testar

$$\mu = 50 \text{ vs } \mu \neq 50$$

- Observamos $\bar{x} = 51,3$.
- A figura mostra uma região crítica com valores simétricos 51,3 e 48,7.

Figura 9-6 O valor P é a área da região sombreada quando $\bar{x} = 51.3$.



- Qualquer valor menor para α diminui a região crítica.
- E rejeitamos H_0 quando $\bar{x} = 51,3$;
- O valor P é calculado como

$$ValorP = 1 - P(48,7 \leq \bar{X} \leq 51,3)$$

$$= 1 - P\left(\frac{48,7-50}{2,5/\sqrt{16}} \leq \frac{\bar{x}-50}{2,5/\sqrt{16}} \leq \frac{51,3-50}{2,5/\sqrt{16}}\right)$$

$$= 1 - P(-2,08 \leq Z \leq 2.08) = 0,038.$$

- H_0 é rejeitada para qualquer $\alpha > 0,038$.
- Não seria rejeitada para $\alpha = 0,01$.

Testes de Hipótese vs Intervalo de Confiança

- Estamos testando

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0 \, .$$

- Seja $[l, u]$ um intervalo com confiança $100(1-\alpha)\%$ para θ.
- Se $\theta_0 \in [l, u]$ não rejeitamos H_0.
- Se $\theta_0 \notin [l, u]$ rejeitamos H_0.

Exemplo:

- Considere o exemplo do propelente.
- Queremos testar

$$H_0 : \mu = 50 \quad vs \quad H_1 : \mu \neq 50\,.$$

- Vimos que com

$$\bar{x} = 51,3 \quad \sigma = 2,5 \quad n = 16$$

  rejeitamos a hipótese nula com $\alpha = 0,05$.
- O intervalo de 95% de confiança pra μ fica

$$50,075 \leq \mu \leq 52,525\,.$$

- Como $50 \notin [50,075; 52,525]$ a hipótese nula é rejeitada.

Procedimento geral para teste de hipótese

- Identifique o parâmetro de interesse a partir do problema.
- Especifique a hipótese nula H_0.
- Escolha um nível de significância α.
- Determine uma estatística de teste.
- Calcule o valor P e veja se é menor que α.
- **Ou** determine a região crítica e veja se a estatística cai nessa região.
- Decida se deve rejeitar ou não H_0 e conclua.