# MAF 261 - Estatística Experimental

Prof. Fernando de Souza Bastos

Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa
Campus UFV - Florestal

14/08/2018

# Sumário

# Testes de hipóteses

## Exemplo prático (Montgomery e Runger (2016)):

Suponha que um engenheiro esteja projetando um sistema de escape da tripulação de uma aeronave, que consiste em um assento de ejeção e um motor de foguete que energiza o assento. O motor de foguete contém um propelente. Para o assento de ejeção funcionar apropriadamente, o propelente deve ter uma taxa mínima de queima de 50 cm/s. Se a taxa de queima for muito baixa, o assento de ejeção poderá não funcionar apropriadamente, levando a uma ejeção não segura. Taxas maiores de queima podem implicar instabilidade no propelente ou um assento de ejeção muito potente, levando outra vez a insegurança da injeção. Dessa maneira, a questão prática de engenharia que tem de ser respondida é: a taxa média de queima do propelente é igual a 50 cm/s ou é igual a algum outro valor (maior ou menor)?

# Testes de hipóteses

## Hipótese Estatística:

Uma hipótese estatística é uma afirmação sobre os parâmetros de uma ou mais populações.

# Testes de hipóteses

Considere o sistema de escape da tripulação descrito no exemplo anterior. Suponha que estejamos interessados na taxa de queima do propelente sólido. Agora, a taxa de queima é uma variável aleatória que pode ser descrita por uma distribuição de probabilidades. Suponha que nosso interesse esteja focado na taxa média de queima (um parâmetro dessa distribuição). Especificamente, estamos interessados em decidir se a taxa média de queima é ou não 50 centímetros por segundo. Podemos expressar isso formalmente como:

$$\begin{aligned} H_0 &: \mu = 50cm/s \\ H_1 &: \mu \neq 50cm/s \end{aligned} \tag{1}$$

# Testes de hipóteses

A afirmação $H_0 : \mu = 50$ centímetros por segundo na Equação 1 é chamada de hipótese nula, e a afirmação $H_1 : \mu \neq 50$ centímetros por segundo é chamada de hipótese alternativa. Uma vez que a hipótese alternativa especifica valores de $\mu$ que poderiam ser maiores ou menores do que 50 centímetros por segundo, ela é chamada de hipótese alternativa bilateral.

# Testes de hipóteses

A afirmação $H_0 : \mu = 50$ centímetros por segundo na Equação 1 é chamada de hipótese nula, e a afirmação $H_1 : \mu \neq 50$ centímetros por segundo é chamada de hipótese alternativa. Uma vez que a hipótese alternativa especifica valores de $\mu$ que poderiam ser maiores ou menores do que 50 centímetros por segundo, ela é chamada de hipótese alternativa bilateral.

Em algumas situações, podemos desejar formular uma hipótese alternativa unilateral, como em:

$$\begin{cases} H_0 : \mu = 50cm/s \\ H_1 : \mu > 50cm/s \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \mu = 50cm/s \\ H_1 : \mu < 50cm/s \end{cases} \tag{2}$$

# Testes de hipóteses

Sempre estabeleceremos a hipótese nula como uma reivindicação de igualdade. Entretanto, quando a hipótese alternativa for estabelecida com o sinal $<$, a reivindicação implícita na hipótese nula será $\geq$ e quando a hipótese alternativa for estabelecida com o sinal $>$, a reivindicação implícita na hipótese nula será $\leq$ .

# Testes de hipóteses

É importante lembrar que hipóteses são sempre afirmações sobre a população ou distribuição sob estudo, não afirmações sobre a amostra. O valor do parâmetro especificado da população na hipótese nula (50 centímetros por segundo no exemplo anterior) é geralmente determinado em uma das três maneiras.

1. experiência passada, conhecimento do processo ou experimentos prévios. O objetivo nesse caso é determinar se o valor do parâmetro variou;

# Testes de hipóteses

É importante lembrar que hipóteses são sempre afirmações sobre a população ou distribuição sob estudo, não afirmações sobre a amostra. O valor do parâmetro especificado da população na hipótese nula (50 centímetros por segundo no exemplo anterior) é geralmente determinado em uma das três maneiras.

1. experiência passada, conhecimento do processo ou experimentos prévios. O objetivo nesse caso é determinar se o valor do parâmetro variou;
2. alguma teoria ou modelo relativo ao processo sob estudo. Aqui, o objetivo do teste é verificar a teoria ou modelo;

# Testes de hipóteses

É importante lembrar que hipóteses são sempre afirmações sobre a população ou distribuição sob estudo, não afirmações sobre a amostra. O valor do parâmetro especificado da população na hipótese nula (50 centímetros por segundo no exemplo anterior) é geralmente determinado em uma das três maneiras.

1. experiência passada, conhecimento do processo ou experimentos prévios. O objetivo nesse caso é determinar se o valor do parâmetro variou;
2. alguma teoria ou modelo relativo ao processo sob estudo. Aqui, o objetivo do teste é verificar a teoria ou modelo;
3. considerações externas, tais como projeto ou especificações de engenharia, ou a partir de obrigações contratuais. Nessa situação, o objetivo usual é avaliar a correção das especificações.

# Testes de hipóteses

**Teste de hipóteses se apoiam no uso de informações de uma amostra aleatória proveniente da população de interesse**. É importante ressaltar que a verdade ou falsidade de uma hipótese particular pode nunca ser conhecida com certeza, a menos que possamos examinar a população inteira. Testar uma hipótese envolve:

- considerar uma amostra aleatória;

# Testes de hipóteses

**Teste de hipóteses se apoiam no uso de informações de uma amostra aleatória proveniente da população de interesse**. É importante ressaltar que a verdade ou falsidade de uma hipótese particular pode nunca ser conhecida com certeza, a menos que possamos examinar a população inteira. Testar uma hipótese envolve:

- considerar uma amostra aleatória;
- computar uma estatística de teste a partir de dados amostrais

# Testes de hipóteses

**Teste de hipóteses se apoiam no uso de informações de uma amostra aleatória proveniente da população de interesse**. É importante ressaltar que a verdade ou falsidade de uma hipótese particular pode nunca ser conhecida com certeza, a menos que possamos examinar a população inteira. Testar uma hipótese envolve:

- considerar uma amostra aleatória;
- computar uma estatística de teste a partir de dados amostrais
- e então usar a estatística de teste para tomar uma decisão a respeito da hipótese nula.

# Testes de Hipóteses Estatísticas

A hipótese nula corresponde à taxa média de queima ser igual a 50 centímetros por segundo e a alternativa corresponde a essa taxa não ser igual a 50 centímetros por segundo. Ou seja, desejamos testar

$$H_0 : \mu = 50cm/s \qquad \text{contra} \qquad H_1 : \mu \neq 50cm/s$$

Suponha que uma amostra de $n = 10$ espécimes seja testada e que a taxa média $\bar{x}$ seja observada. A média amostral é uma estimativa de $\mu$. Um valor de $\bar{x}$ que caia próximo a $\mu = 50$ cm/s é uma evidência de que $\mu$ é realmente 50 cm/s. Por outro lado, uma média amostral que seja consideravelmente diferente de 50 cm/s evidencia a validade da hipótese alternativa $H_1$. Assim, a média amostral é a estatística de teste nesse caso.

# Testes de hipóteses

A média amostral pode assumir muitos valores diferentes. Suponha que se $48,5 \leq \bar{x} \leq 51,5$, não rejeitaremos a hipótese nula $H_0 : \mu = 50$ e se $\bar{x} < 48,5$ ou $\bar{x} > 51,5$, rejeitaremos a hipótese nula em favor da hipótese alternativa $H_1 : \mu \neq 50$. Isso é ilustrado na Figura abaixo:

# Testes de hipóteses

A média amostral pode assumir muitos valores diferentes. Suponha que se $48,5 \leq \bar{x} \leq 51,5$, não rejeitaremos a hipótese nula $H_0 : \mu = 50$ e se $\bar{x} < 48,5$ ou $\bar{x} > 51,5$, rejeitaremos a hipótese nula em favor da hipótese alternativa $H_1 : \mu \neq 50$. Isso é ilustrado na Figura abaixo:

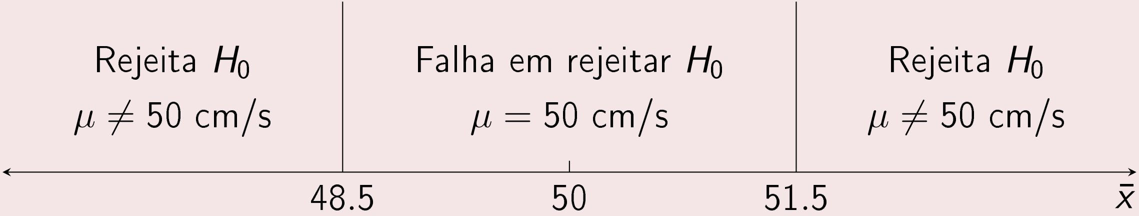



**Figura:** Critérios de decisão para testar $H_0 : \mu = 50$ cm/s versus $H_1 : \mu \neq 50$ cm/s.

# Testes de hipóteses

Os valores de $\bar{x}$ que forem menores do que $48,5$ e maiores do que $51,5$ constituem a **região crítica** para o teste, enquanto todos os valores que estejam no intervalo $48,5 \leq \bar{x} \leq 51,5$ formam uma região para a qual falharemos em rejeitar a hipótese nula. Por convenção, ela geralmente é chamada de **região de não rejeição**. Os limites entre as regiões críticas e a região de aceitação são chamados de valores críticos.

# Testes de hipóteses

Em nosso exemplo, os valores críticos são $48,5$ e $51,5$. É comum estabelecer conclusões relativas à hipótese nula $H_0$. Logo, rejeitaremos $H_0$ em favor de $H_1$, se a estatística de teste cair na região crítica e falhamos em rejeitar $H_0$ por sua vez se a estatística de teste cair na região de aceitação.

# Testes de hipóteses

Esse procedimento pode levar a duas conclusões erradas. Por exemplo, a taxa média verdadeira de queima do propelente poderia ser igual a 50 centímetros por segundo. Entretanto, para as amostras de propelente, selecionados aleatoriamente, que são testados, poderíamos observar um valor de estatística de teste $\bar{x}$ que caísse na região crítica. Rejeitaríamos então a hipótese nula $H_0$ em favor da alternativa $H_1$, quando, de fato, $H_0$ seria realmente verdadeira. Esse tipo de conclusão errada é chamado de **erro tipo I**.

# Testes de hipóteses

Esse procedimento pode levar a duas conclusões erradas. Por exemplo, a taxa média verdadeira de queima do propelente poderia ser igual a 50 centímetros por segundo. Entretanto, para as amostras de propelente, selecionados aleatoriamente, que são testados, poderíamos observar um valor de estatística de teste $\bar{x}$ que caísse na região crítica. Rejeitaríamos então a hipótese nula $H_0$ em favor da alternativa $H_1$, quando, de fato, $H_0$ seria realmente verdadeira. Esse tipo de conclusão errada é chamado de **erro tipo I**.

## Erro Tipo I

A rejeição da hipótese nula $H_0$ quando ela for verdadeira é definida como **erro tipo I**.

# Testes de hipóteses

Agora, suponha que a taxa média verdadeira de queima seja diferente de 50 centímetros por segundo, mesmo que a média amostral $\bar{x}$ caia na região de aceitação. Nesse caso, falharíamos em rejeitar $H_0$, quando ela fosse falsa. Esse tipo de conclusão errada é chamado de **erro tipo II**.

# Testes de hipóteses

Agora, suponha que a taxa média verdadeira de queima seja diferente de 50 centímetros por segundo, mesmo que a média amostral $\bar{x}$ caia na região de aceitação. Nesse caso, falharíamos em rejeitar $H_0$, quando ela fosse falsa. Esse tipo de conclusão errada é chamado de **erro tipo II**.

## Erro Tipo II

A falha em rejeitar a hipótese nula, quando ela é falsa, é definida como **erro tipo II**.

# Testes de hipóteses

Assim, testando qualquer hipótese estatística, quatro situações diferentes determinam se a decisão final está correta ou errada. Pelo fato de a nossa decisão estar baseada em variáveis aleatórias, probabilidades podem ser associadas aos erros tipo I e tipo II. A probabilidade de cometer o erro tipo I é denotada pela letra grega $\alpha$.

| **Decisão** | $H_0$ **é verdadeira** | $H_0$ **é falsa** |
|---|---|---|
| **Não rejeita** $H_0$ | Correta<br>**Probabilidade** $= (1-\alpha)$ | Erro Tipo II<br>**Probabilidade** $= \beta$ |
| **Rejeita** $H_0$ | Erro Tipo I<br>**Nível de significância** $\alpha$ | Correta<br>**Poder** $= (1-\beta)$ |

# Testes de hipóteses

A probabilidade do erro tipo I é chamada de **nível de significância**, ou **erro** $\alpha$, ou **tamanho do teste**. No exemplo da taxa de queima de propelente, um **erro tipo I** ocorrerá quando $\bar{x} > 51,5$ ou $\bar{x} < 48,5$, quando a taxa média verdadeira de queima do propelente for $\mu = 50$ cm/s.

# Testes de hipóteses

Suponha que o desvio-padrão da taxa de queima seja $\sigma = 2,5$ centímetros por segundo e que a taxa de queima tenha uma distribuição para a qual as condições do **teorema central do limite** se aplicam; logo, a distribuição da média amostral é aproximadamente normal, com média $\mu = 50$ e desvio-padrão $\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{10}} = 0.79$.

A probabilidade de cometer o **erro tipo I** (ou o nível de significância de nosso teste) é igual à soma das áreas sombreadas nas extremidades da distribuição normal na Figura abaixo:

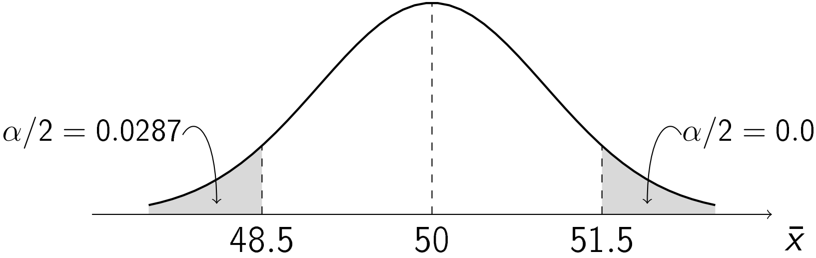



**Figura:** Região crítica para $H_0 : \mu = 50$ versus $H_1 : \mu \neq 50$ e $n = 10$

A probabilidade de cometer o **erro tipo I** (ou o nível de significância de nosso teste) é igual à soma das áreas sombreadas nas extremidades da distribuição normal na Figura abaixo:

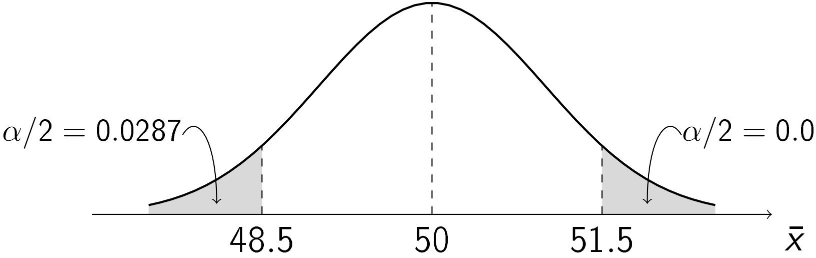



**Figura:** Região crítica para $H_0 : \mu = 50$ versus $H_1 : \mu \neq 50$ e $n = 10$

$$\alpha = P(\bar{X} < 48.5 \quad \text{quando} \quad \mu = 50) + P(\bar{X} > 51.5 \quad \text{quando} \quad \mu = 50).$$

Os valores de z que correspondem aos valores críticos $48,5$ e $51,5$ são

$$z_1 = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{48.5 - 50}{0.79} = -1.9 \quad \text{e} \quad z_2 = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{51.5 - 50}{0.79} = 1.9$$

Logo,

$$\alpha = P(z < -1.90) + P(z > 1.90) = 0.0287 + 0.0287 = 0.0574$$

Essa é a probabilidade do erro tipo I. Isso implica que $5,74\%$ de todas as amostras aleatórias conduziriam à rejeição da hipótese $H_0 : \mu = 50$ cm/s, quando a taxa média verdadeira de queima fosse realmente 50 centímetros por segundo.Da inspeção da Figura anterior, notamos que podemos reduzir $\alpha$ alargando a região de aceitação.

Por exemplo, se considerarmos os valores críticos 48 e 52, o valor de $\alpha$ será:

$$
\begin{aligned}
\alpha &= P\left(z < -\frac{48 - 50}{0.79}\right) + P\left(z > \frac{52 - 50}{0.79}\right) \\
&= P(z < -2.53) + P(z > 2.53) \\
&= 0.0057 + 0.0057 = 0.0114
\end{aligned}
$$

Por exemplo, se considerarmos os valores críticos 48 e 52, o valor de $\alpha$ será:

$$\alpha = P\left(z < -\frac{48-50}{0.79}\right) + P\left(z > \frac{52-50}{0.79}\right)$$
$$= P(z < -2.53) + P(z > 2.53)$$
$$= 0.0057 + 0.0057 = 0.0114$$

Poderíamos também reduzir $\alpha$, **aumentando o tamanho da amostra.** Se $n = 16$, então $\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} = 0.625$. Logo,

$$z_1 = \frac{48.5-50}{0.625} = -2.40 \quad \text{e} \quad z_2 = \frac{51.5-50}{0.625} = 2.40$$

e, $\alpha = P(z < -2.40) + P(z > 2.40) = 0.0082 + 0.0082 = 0.0164$.

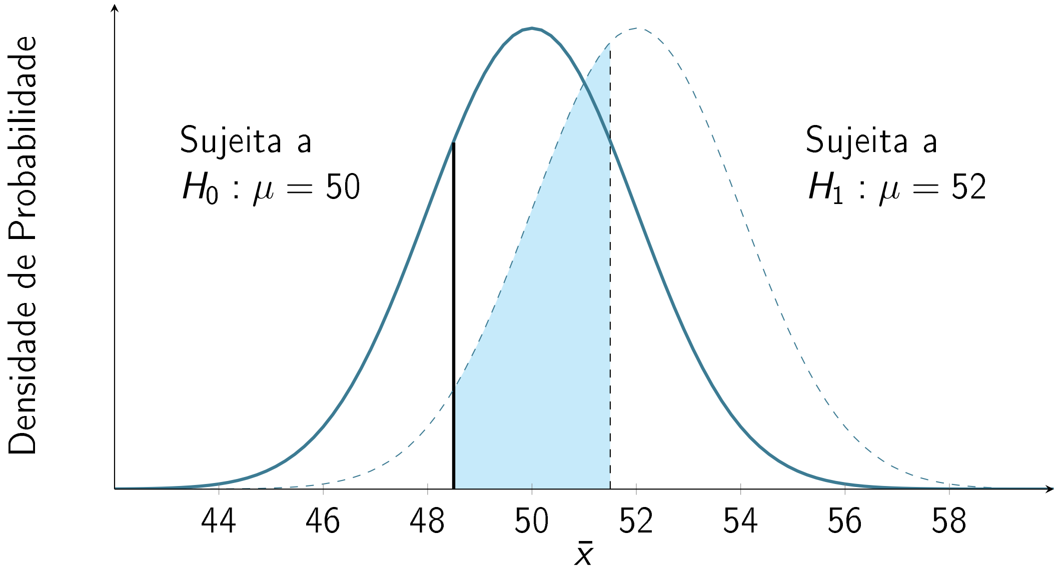
No entanto, na avaliação de um procedimento de teste de hipóteses, também é importante examinar a probabilidade do **erro tipo II**, que é denotado por $\beta$. Lembremos que,

$$\beta = P(\text{Erro tipo II}) = P(\text{Não rejeitar } H_0 \text{ dado que } H_0 \text{ é falsa})$$

Para calcular $\beta$ (algumas vezes chamado de erro $\beta$), temos de ter uma hipótese alternativa específica fixada; ou seja, temos de ter um valor particular de $\mu$. Por exemplo, suponha que seja importante rejeitar a hipótese nula $H_0 : \mu = 50$ toda vez que a taxa média de queima $\mu$ seja maior do que 52 cm/s ou menor do que 48 cm/s.

Poderíamos calcular a probabilidade de um erro tipo II, $\beta$, para os valores $\mu = 52$ e $\mu = 48$ e usar esse resultado para nos dizer alguma coisa acerca de como seria o desempenho do procedimento de teste. Por causa da simetria, só é necessário avaliar um dos dois casos. Isto é, encontrar a probabilidade de não rejeitar a hipótese nula $H_0 : \mu = 50$ cm/s, quando a média verdadeira, por exemplo, for $\mu = 52$ cm/s.

A próxima Figura nos ajudará a calcular a probabilidade do erro tipo II, $\beta$.



Um erro tipo II será cometido, se a média amostral $\bar{x}$ cair entre 48,5 e 51,5, quando $\mu = 52$. Como visto na Figura anterior, essa é apenas a probabilidade de $48,5 \leq \bar{X} \leq 51,5$, quando a média verdadeira for $\mu = 52$, ou a área sombreada sob a distribuição normal centralizada em $\mu = 52$. Consequentemente, referindo-se à anterior, encontramos que

$$\beta = P(48.5 \leq \bar{X} \leq 51.5, \text{ quando } \mu = 52)$$

Os valores $z$, correspondentes a $48,5$ e $51,5$, quando $\mu = 52$, são

$$z_1 = \frac{48.5 - 52}{0.79} = -4.43 \quad \text{e} \quad z_2 = \frac{51.5 - 52}{0.79} = -0.63$$

logo,

$$\beta = P(-4.43 \leq z \leq -0.63) = 0.2643.$$

Assim, se estivermos testando $H_0 : \mu = 50$ contra $H_1 : \mu \neq 50$, com $n = 10$ e o valor verdadeiro da média for $\mu = 52$, a probabilidade de falharmos em rejeitar a falsa hipótese nula é $0,2643$. Por simetria, se o valor verdadeiro da média for $\mu = 48$, o valor de $\beta$ será também $0,2643$. A probabilidade de cometer o erro tipo II, $\beta$, aumenta rapidamente à medida que o valor verdadeiro de $\mu$ se aproxima do valor da hipótese feita. Por exemplo, veja a próxima Figura, em que o valor verdadeiro da média é $\mu = 50,5$ e o valor da hipótese é $H_0 : \mu = 50$. O valor verdadeiro de $\mu$ está muito perto de 50 e o valor para $\beta$ é

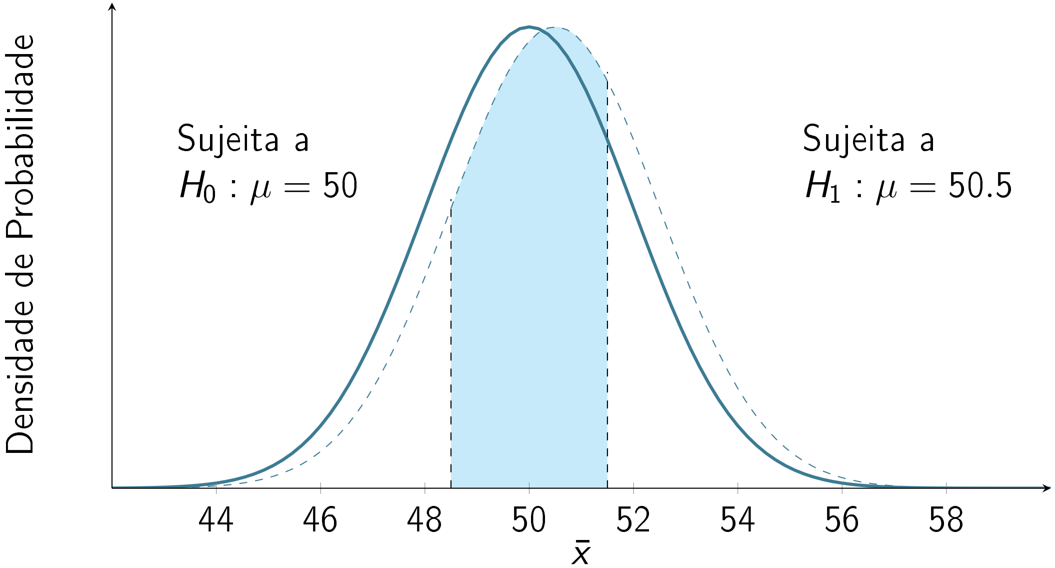
$$\beta = P(48.5 \leq \bar{X} \leq 51.5, \text{ quando } \mu = 50.5)$$

Os valores $z$, correspondentes a $48,5$ e $51,5$, quando $\mu = 50.5$, são

$$z_1 = \frac{48.5 - 50.5}{0.79} = -2.53 \quad \text{e} \quad z_2 = \frac{51.5 - 50.5}{0.79} = 1.27$$

logo,

$$\beta = P(-2.53 \leq z \leq 1.27) = 0.8923.$$



Densidade de Probabilidade
Sujeita a
$H_0 : \mu = 50$
Sujeita a
$H_1 : \mu = 50.5$
44
46
48
50
52
54
56
58
$\bar{x}$

Assim, a probabilidade do erro tipo II é muito maior para o caso em que a média verdadeira é 50,5 centímetros por segundo do que para o caso em que a média é 52 cm/s. Naturalmente, em muitas situações práticas, não estaríamos preocupados em cometer o erro tipo II se a média fosse "próxima" do valor utilizado na hipótese. Estaríamos muito mais interessados em detectar grandes diferenças entre a média verdadeira e o valor especificado na hipótese nula.

A probabilidade do erro tipo II depende também do tamanho da amostra, $n$. Suponha que a hipótese nula seja $H_0 : \mu = 50$ centímetros por segundo e que o valor verdadeiro da média seja $\mu = 52$. Se o tamanho da amostra for aumentado de $n = 10$ para $n = 16$, temos que o desvio-padrão de $\bar{X}$ é $\dfrac{\sigma}{\sqrt{n}} = \dfrac{2.5}{\sqrt{16}} = 0.625$. Logo,

$$z_1 = \frac{48.5 - 52}{0.625} = -5.60 \quad \text{e} \quad z_2 = \frac{51.5 - 52}{0.625} = -0.80$$

e, $\beta = P(-5.60 \leq z \leq -0.80) = 0.2119$.

Lembre-se de que quando $n = 10$ e $\mu = 52$, encontramos que $\beta = 0,2643$; consequentemente, o aumento do tamanho da amostra resulta em uma diminuição na probabilidade de erro tipo II. Os resultados vistos até agora e outros poucos cálculos similares estão sumarizados na tabela abaixo. Os valores críticos são ajustados para manter $\alpha's$ iguais para n = 10 e n = 16. Esse tipo de cálculo é discutido mais adiante nas aulas.

| RNR$H_0$ | n | $\alpha$ | $\beta$ em $\mu = 52$ | $\beta$ em $\mu = 50.5$ |
|---|---|---|---|---|
| $48.5 \leq \bar{X} \leq 51.5$ | 10 | 0.0576 | 0.2643 | 0.8923 |
| $48 \leq \bar{X} \leq 52$ | 10 | 0.0114 | 0.5000 | 0.9705 |
| $48.81 \leq \bar{X} \leq 51.19$ | 16 | 0.0576 | 0.0966 | 0.8606 |
| $48.42 \leq \bar{X} \leq 51.58$ | 16 | 0.0114 | 0.2515 | 0.9578 |

A tabela anterior e a discussão anterior revelam quatro pontos importantes:

1. O tamanho da região crítica, e consequentemente a probabilidade do erro tipo I, $\alpha$, pode sempre ser reduzido por meio da seleção apropriada dos valores críticos.

A tabela anterior e a discussão anterior revelam quatro pontos importantes:

1. O tamanho da região crítica, e consequentemente a probabilidade do erro tipo I, $\alpha$, pode sempre ser reduzido por meio da seleção apropriada dos valores críticos.
2. Os erros tipo I e tipo II estão relacionados. Uma diminuição na probabilidade de um tipo de erro sempre resulta em um aumento da probabilidade do outro, desde que o tamanho da amostra, n, não varie.

A tabela anterior e a discussão anterior revelam quatro pontos importantes:

1. O tamanho da região crítica, e consequentemente a probabilidade do erro tipo I, $\alpha$, pode sempre ser reduzido por meio da seleção apropriada dos valores críticos.
2. Os erros tipo I e tipo II estão relacionados. Uma diminuição na probabilidade de um tipo de erro sempre resulta em um aumento da probabilidade do outro, desde que o tamanho da amostra, n, não varie.
3. Um aumento no tamanho da amostra reduzirá $\beta$, desde que $\alpha$ seja mantido constante.

A tabela anterior e a discussão anterior revelam quatro pontos importantes:

1. O tamanho da região crítica, e consequentemente a probabilidade do erro tipo I, $\alpha$, pode sempre ser reduzido por meio da seleção apropriada dos valores críticos.
2. Os erros tipo I e tipo II estão relacionados. Uma diminuição na probabilidade de um tipo de erro sempre resulta em um aumento da probabilidade do outro, desde que o tamanho da amostra, n, não varie.
3. Um aumento no tamanho da amostra reduzirá $\beta$, desde que $\alpha$ seja mantido constante.
4. Quando a hipótese nula é falsa, $\beta$ aumenta à medida que o valor verdadeiro do parâmetro se aproxima do valor usado na hipótese nula. O valor de $\beta$ diminui à medida que aumenta a diferença entre a média verdadeira e o valor utilizado na hipótese.

Geralmente, o(a) analista controla a probabilidade $\alpha$ do erro tipo I quando ele ou ela seleciona os valores críticos. Assim, geralmente é fácil para o analista estabelecer a probabilidade de erro tipo I em (ou perto de) qualquer valor desejado.

Geralmente, o(a) analista controla a probabilidade $\alpha$ do erro tipo I quando ele ou ela seleciona os valores críticos. Assim, geralmente é fácil para o analista estabelecer a probabilidade de erro tipo I em (ou perto de) qualquer valor desejado.

Uma vez que o analista pode controlar diretamente a probabilidade de rejeitar erroneamente $H_0$, sempre pensamos na rejeição da hipótese nula $H_0$ como uma **conclusão forte.**

Uma vez que podemos controlar a probabilidade de cometer um erro tipo I (ou nível de significância), uma questão lógica é que valor deve ser usado?

Uma vez que podemos controlar a probabilidade de cometer um erro tipo I (ou nível de significância), uma questão lógica é que valor deve ser usado?

A probabilidade do **erro tipo I** é uma medida de risco, especificamente o risco de concluir que a hipótese nula é falsa quando ela realmente não é. Assim, o valor de $\alpha$ deve ser escolhido para refletir as consequências (econômicas, sociais etc.) de rejeitar incorretamente a hipótese nula. Frequentemente, isso é difícil de fazer, e o que tem evoluído muito na prática científica e de engenharia é usar o valor $\alpha = 0,05$ na maioria das situações, a menos que haja alguma informação disponível que indique que essa é uma escolha não apropriada.

Por outro lado, a probabilidade $\beta$ do erro tipo II não é constante, mas depende do valor verdadeiro do parâmetro. Ela depende também do tamanho da amostra que tenhamos selecionado. Pelo fato de a probabilidade $\beta$ do erro tipo II ser uma função do tamanho da amostra e da extensão com que a hipótese nula $H_0$ seja falsa, costuma-se pensar na aceitação de $H_0$ como uma conclusão fraca, a menos que saibamos que $\beta$ seja aceitavelmente pequena.

Consequentemente, em vez de dizer "aceitar $H_0$", preferimos a terminologia "**Não rejeitar** $H_0$". Falhar em rejeitar $H_0$ implica que não encontramos evidência suficiente para rejeitar $H_0$, ou seja, para fazer uma afirmação forte. Falhar em rejeitar $H_0$ não significa necessariamente que haja uma alta probabilidade de que $H_0$ seja verdadeira. Isso pode significar simplesmente que mais dados são requeridos para atingir uma conclusão forte, o que pode ter implicações importantes para a formulação das hipóteses.

Existe uma analogia útil entre teste de hipóteses e um julgamento por jurados. Em um julgamento, o réu é considerado inocente (isso é como considerar a hipótese nula verdadeira). Se forte evidência for encontrada do contrário, o réu é declarado culpado (rejeitamos a hipótese nula). Se não houver **suficiente** evidência, o réu é declarado não culpado. Isso não é o mesmo de provar a inocência do réu; assim, tal qual falhar em rejeitar a hipótese nula, essa é uma conclusão fraca.

Um importante conceito de que faremos uso é a **potência** ou **poder** de um teste estatístico:

A potência ou poder de um teste estatístico é a probabilidade de rejeitar a hipótese nula H0, quando a hipótese alternativa for verdadeira.

A potência é calculada como $1 - \beta$, e pode ser interpretada como a probabilidade de rejeitar corretamente uma hipótese nula falsa. Frequentemente, comparamos testes estatísticos por meio da comparação de suas propriedades de potência. Por exemplo, considere o problema da taxa de queima de propelente, quando estamos testando $H_0 : \mu = 50$ cm/s contra $H_1 : \mu \neq 50$ cm/s. Suponha que o valor verdadeiro da média seja $\mu = 52$. Quando $n = 10$, encontramos que $\beta = 0.2643$; logo, a potência desse teste é $1 - \beta = 1 - 0.2643 = 0.7357$, quando $\mu = 52$.

A potência é uma medida muito descritiva e concisa da sensibilidade de um teste estatístico, em que por sensibilidade entendemos a habilidade do teste de detectar diferenças. Nesse caso, a sensibilidade do teste para detectar a diferença entre a taxa média de queima de 50 centímetros por segundo e 52 centímetros por segundo é 0.7357. Isto é, se a média verdadeira for realmente 52 centímetros por segundo, esse teste rejeitará corretamente $H_0 : \mu = 50$ e "detectará" essa diferença em $73,57\%$ das vezes. Se esse valor de potência for julgado como muito baixo, o analista poderá aumentar tanto $\alpha$ como o tamanho da amostra $n$.

# Referências Bibliográficas

D. C. Montgomery e G. C. Runger. *Estatística Aplicada E Probabilidade Para Engenheiros.* Grupo Gen-LTC, São Paulo, 6 edition, 2016.