# MAF 261 - Estatística Experimental

Prof. Fernando de Souza Bastos

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

03/05/2018

# Sumário

# Testes de hipóteses

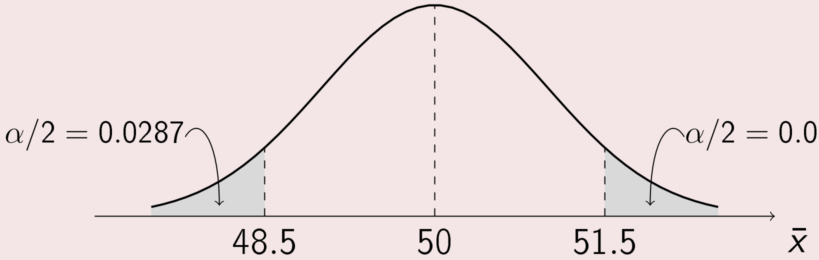


**Figura:** Região crítica para $H_0 : \mu = 50$ versus $H_1 : \mu \neq 50$ e $n = 10$

# Testes de hipóteses

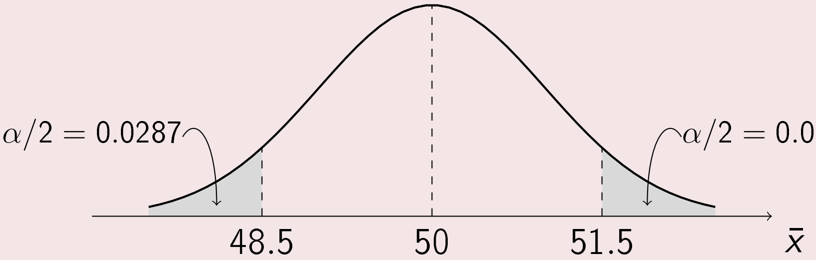


**Figura:** Região crítica para $H_0 : \mu = 50$ versus $H_1 : \mu \neq 50$ e $n = 10$

Podemos achar essa probabilidade como:

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

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

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

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:

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

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

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; 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. Especificamente, como o procedimento de teste funcionará ao rejeitar $H_0$, para um valor médio de $\mu = 52$ ou $\mu = 48$? Por causa da simetria, só é necessário avaliar um dos dois casos. Isto é, encontrar a probabilidade de aceitar a hipótese nula $H_0 : \mu = 50$ cm/s, quando a média verdadeira, por exemplo, for $\mu = 52$ cm/s.

A Figura 9-3 nos ajudará a calcular a probabilidade do erro tipo II, $\beta$. A distribuição normal no lado esquerdo da Figura 9-3 é a distribuição da estatística de teste $\bar{X}$, quando a hipótese nula $H_0 : \mu = 50$ for verdadeira (ou seja, isso é o que se entende pela expressão "sujeita a $H_0 : \mu = 50$"). A distribuição normal no lado direito é a distribuição de $\bar{X}$, quando a hipótese alternativa for verdadeira e o valor da média for 52 (ou "sujeita a $H_1 : \mu = 52$"). Agora, um erro tipo II será cometido, se a média amostral $\bar{x}$ cair entre 48,5 e 51,5 (os limites da região crítica), quando $\mu = 52$. Como visto na Figura 9-3, 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 Figura 9-3, encontramos que

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

Figura 9-3!!!

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 Figura 9-4, 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,

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

Figura!!!

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 $n = 16$, resulta a situação da Figura 9-5. A distribuição normal à esquerda é a distribuição de $\bar{X}$, quando a média $\mu = 50$, e a distribuição normal à direita é a distribuição de $\bar{X}$, quando $\mu = 52$. Conforme mostrado na Figura 9-5, a probabilidade do erro tipo II é

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

Quando $n = 16$, 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.
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. 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. 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. Valores menores de $\alpha$ refletiriam consequências mais severas, e valores maiores de $\alpha$ seriam consistentes com consequências menos severas. 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. No problema do propelente do foguete com $n = 10$, isso corresponderia aos valores críticos de 48,45 e 51,55.

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 "falhar em 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 a potência 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 $n$.

Na construção de hipóteses, sempre vamos estabelecer a hipótese nula como uma igualdade, de modo que a probabilidade do erro tipo I, $\alpha$, pode ser controlada em um valor específico. A hipótese alternativa tanto pode ser unilateral como bilateral, dependendo da conclusão a ser retirada se $H_0$ é rejeitada. Se o objetivo é fazer uma alegação envolvendo afirmações, tais como **maior que**, **menor que**, **superior a**, **excede**, **no mínimo**, e assim por diante, uma alternativa unilateral é apropriada. Se nenhuma direção é implicada pela alegação, ou se a alegação "**não igual a**" for feita, uma alternativa bilateral deve ser usada.

Considere o problema da taxa de queima de um propelente. Suponha que, se a taxa de queima for menor do que 50 centímetros por segundo, desejamos mostrar esse fato com uma conclusão forte. As hipóteses deveriam ser estabelecidas como

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

Aqui, a região crítica está na extremidade inferior da distribuição de $X$. Visto que a rejeição de $H_0$ é sempre uma conclusão forte, essa afirmação das hipóteses produzirá o resultado desejado se $H_0$ for rejeitado. Note que, embora a hipótese nula seja estabelecida com um sinal de igual, deve-se incluir qualquer valor de $\mu$ não especificado pela hipótese alternativa. Desse modo, falhar em rejeitar $H_0$ não significa $\mu = 50$ centímetros por segundo exatamente, mas somente que não temos evidência forte em suportar $H_1$.

Em alguns problemas do mundo real, em que os procedimentos de testes unilaterais sejam indicados, é ocasionalmente difícil escolher uma formulação apropriada da hipótese alternativa. Por exemplo, suponha que um engarrafador de refrigerantes compre 10 garrafas de 10 reais de uma companhia de vidro. O engarrafador quer estar certo de que as garrafas satisfazem as especificações de pressão interna média ou resistência à explosão, que, para garrafas de 10 reais, a resistência mínima é 200 psi. O engarrafador decidiu formular o procedimento de decisão para um lote específico de garrafas como um problema de teste de hipóteses. Há duas formulações possíveis para esse problema:

$$\begin{cases} H_0 : \mu = 200psi \\ H_1 : \mu > 200psi \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \mu = 200psi \\ H_1 : \mu < 200psi \end{cases} \tag{1}$$

Considere a formulação com $H_1 : \mu > 200psi$. Se a hipótese nula for rejeitada, as garrafas serão julgadas satisfatórias; se $H_0$ não for rejeitada, a implicação é que as garrafas não obedecem às especificações e não devem ser usadas. Como rejeitar $H_0$ é uma conclusão forte, essa formulação força o fabricante de garrafas a "demonstrar" que a resistência média à explosão das garrafas excede a especificação. Agora considere a formulação $H_1 : \mu < 200psi$. Nessa situação, as garrafas serão julgadas satisfatórias, a menos que $H_0$ seja rejeitada. Ou seja, concluímos que as garrafas são satisfatórias, a menos que haja forte evidência do contrário.

Qual formulação é a correta? $H_1 : \mu > 200 psi$ ou $H_1 : \mu < 200 psi$? A resposta é “depende” do objetivo da análise.

Qual formulação é a correta? $H_1 : \mu > 200psi$ ou $H_1 : \mu < 200psi$? A resposta é "depende" do objetivo da análise.

Na formulação de hipóteses unilaterais, devemos lembrar que rejeitar $H_0$ é sempre uma conclusão forte. Consequentemente, devemos estabelecer uma afirmação acerca do que é importante para fazer uma conclusão forte na hipótese alternativa. Em problemas do mundo real, isso dependerá frequentemente de nosso ponto de vista e experiência com a situação.

Uma maneira de reportar os resultados de um teste de hipóteses é estabelecer que a hipótese nula foi ou não foi rejeitada com um valor especificado de $\alpha$, ou nível de significância. Isso é chamado de teste de **nível de significância fixo**.

Uma maneira de reportar os resultados de um teste de hipóteses é estabelecer que a hipótese nula foi ou não foi rejeitada com um valor especificado de $\alpha$, ou nível de significância. Isso é chamado de teste de **nível de significância fixo**.

A abordagem de nível de significância fixo para teste de hipóteses é muito interessante porque conduz diretamente aos conceitos de erro tipo II e potência, que são de valor considerável na determinação de tamanhos apropriados de amostras para usar em testes de hipóteses. Mas a abordagem de nível de significância fixo tem algumas desvantagens.

Por exemplo, no problema anterior, do propelente, podemos dizer que $H_0 : \mu = 50$ foi rejeitada com um nível de significância de $0,05$. Essa forma de conclusão é frequentemente inadequada, porque ela não dá ideia, a quem vai tomar a decisão, a respeito de se o valor calculado da estatística de teste estava apenas nas proximidades da região de rejeição ou se estava muito longe dessa região.

Na estatística clássica, o valor-p (também chamado de nível descritivo ou probabilidade de significância), é a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra, sob a hipótese nula. Por exemplo, em testes de hipótese, pode-se rejeitar a hipótese nula a 5% caso o valor-p seja menor que 5%. Assim, uma outra interpretação para o valor-p, é que este é menor nível de significância com que se rejeitaria a hipótese nula. Em termos gerais, um valor-p pequeno significa que a probabilidade de obter um valor da estatística de teste como o observado é muito improvável, levando assim à rejeição da hipótese nula. Assim, um valor P carrega muita informação sobre o peso da evidência contra $H_0$; logo, quem for tomar a decisão pode tirar uma conclusão com qualquer nível especificado de significância.

**O valor-P é o menor nível de significância que conduz à rejeição da hipótese nula $H_0$, com os dados fornecidos**. Em outras palavras, o valor-P é o **nível de significância observado**. Uma vez que o valor P seja conhecido, a pessoa que vai tomar a decisão pode determinar quão significativos são os dados, sem o analista de dados impor, formalmente, um nível pré-selecionado de significância.

Considere o teste bilateral de hipóteses para a taxa de queima

$$H_0 : \mu = 50cm/s$$
$$H_1 : \mu \neq 50cm/s$$

com $n = 16$ e $\sigma = 2,5$. Suponha que a média amostral observada seja $\bar{X} = 51,3$ centímetros por segundo. A Figura abaixo mostra uma região crítica para esse teste, com valores críticos em 51,3 e no valor simétrico 48,7. O valor P do teste é a probabilidade acima de 51,3 mais a probabilidade abaixo de 48,7. O valor P é fácil de calcular depois da estatística de teste ser observada.

$$
\begin{aligned}
Valor - p &= 1 - P(48.7 < \bar{X} < 51.3)\\
&= 1 - P\left(\frac{48.7 - 50}{2.5/\sqrt{16}} < Z < \frac{51.3 - 50}{2.5/\sqrt{16}}\right)\\
&= 1 - P(-2.08 < Z < 2.08)\\
&= 1 - 0.962 = 0.038
\end{aligned}
$$

O valor P nos diz que, se a hipótese nula $H_0 = 50$ for verdadeira, a probabilidade de se obter uma amostra aleatória, cuja média seja no mínimo tão longe de 50 quanto de 51,3 (ou de 48,7), será igual a 0,038. Por conseguinte, uma média amostral observada de 51,3 é um evento razoavelmente raro, se a hipótese nula $H_0$ for realmente verdadeira. Comparado com o nível de significância "padrão" de 0,05, nosso valor P observado é menor; desse modo, se estivéssemos usando um nível de significância fixo de 0,05, Operacionalmente, uma vez calculado o valor P, tipicamente o comparamos a um nível de significância predefinido usado para tomar decisão. Geralmente, esse nível de significância predefinido é 0,05. No entanto, na apresentação de resultados e conclusões, é prática padrão reportar o valor P observado, juntamente com a decisão que é feita em relação à hipótese nula.

Claramente, o valor P fornece uma medida da credibilidade da hipótese nula. Especificamente, ele é o risco de você tomar uma decisão incorreta ao rejeitar a hipótese nula $H_0$. O valor P não é a probabilidade de a hipótese nula ser falsa, nem é a probabilidade $1-P$ de a hipótese nula ser verdadeira. A hipótese nula é verdadeira ou falsa (não há probabilidade associada a isso) e assim a interpretação apropriada do valor P é em termos do risco de rejeitar erroneamente a hipótese nula $H_0$. Não é sempre fácil calcular o valor exato de P para um teste estatístico. No entanto, a maioria dos softwares modernos reporta os resultados dos problemas de testes de hipóteses em termos de valores P. Usaremos extensivamente a abordagem do valor P.

## Simulação do valor - P!

O uso da seguinte sequência de etapas na metodologia de aplicação de testes de hipóteses é recomendado.

1. Parâmetro de interesse: A partir do contexto do problema, identifique o parâmetro de interesse.

O uso da seguinte sequência de etapas na metodologia de aplicação de testes de hipóteses é recomendado.

1. Parâmetro de interesse: A partir do contexto do problema, identifique o parâmetro de interesse.
2. Hipótese nula, $H_0$: Estabeleça a hipótese nula $H_0$.

O uso da seguinte sequência de etapas na metodologia de aplicação de testes de hipóteses é recomendado.

1. Parâmetro de interesse: A partir do contexto do problema, identifique o parâmetro de interesse.
2. Hipótese nula, $H_0$: Estabeleça a hipótese nula $H_0$.
3. Hipótese alternativa, $H_1$: Especifique uma hipótese alternativa apropriada, $H_1$.

O uso da seguinte sequência de etapas na metodologia de aplicação de testes de hipóteses é recomendado.

1. Parâmetro de interesse: A partir do contexto do problema, identifique o parâmetro de interesse.
2. Hipótese nula, $H_0$: Estabeleça a hipótese nula $H_0$.
3. Hipótese alternativa, $H_1$: Especifique uma hipótese alternativa apropriada, $H_1$.
4. Estatística de teste:Determine uma estatística apropriada de teste.

O uso da seguinte sequência de etapas na metodologia de aplicação de testes de hipóteses é recomendado.

1. Parâmetro de interesse: A partir do contexto do problema, identifique o parâmetro de interesse.
2. Hipótese nula, $H_0$: Estabeleça a hipótese nula $H_0$.
3. Hipótese alternativa, $H_1$: Especifique uma hipótese alternativa apropriada, $H_1$.
4. Estatística de teste:Determine uma estatística apropriada de teste.
5. Rejeita $H_0$ se: Estabeleça os critérios de rejeição para a hipótese nula.

O uso da seguinte sequência de etapas na metodologia de aplicação de testes de hipóteses é recomendado.

1. Parâmetro de interesse: A partir do contexto do problema, identifique o parâmetro de interesse.
2. Hipótese nula, $H_0$: Estabeleça a hipótese nula $H_0$.
3. Hipótese alternativa, $H_1$: Especifique uma hipótese alternativa apropriada, $H_1$.
4. Estatística de teste:Determine uma estatística apropriada de teste.
5. Rejeita $H_0$ se: Estabeleça os critérios de rejeição para a hipótese nula.
6. Cálculos:Calcule quaisquer grandezas amostrais necessárias, substitua-as na equação para a estatística de teste e calcule esse valor.

O uso da seguinte sequência de etapas na metodologia de aplicação de testes de hipóteses é recomendado.

1. Parâmetro de interesse: A partir do contexto do problema, identifique o parâmetro de interesse.
2. Hipótese nula, $H_0$: Estabeleça a hipótese nula $H_0$.
3. Hipótese alternativa, $H_1$: Especifique uma hipótese alternativa apropriada, $H_1$.
4. Estatística de teste:Determine uma estatística apropriada de teste.
5. Rejeita $H_0$ se: Estabeleça os critérios de rejeição para a hipótese nula.
6. Cálculos:Calcule quaisquer grandezas amostrais necessárias, substitua-as na equação para a estatística de teste e calcule esse valor.
7. Conclusões:Decida se $H_0$ deve ou não ser rejeitada e reporte isso no contexto do problema.

Notamos previamente que é muito útil reportar os resultados de um teste de hipóteses em termos do valor P, porque ele carrega mais informação que a simples afirmação "rejeita $H_0$" ou "falha em rejeitar $H_0$". Ou seja, a rejeição de $H_0$ com nível de significância igual a 0,05 é muito mais significativa se o valor da estatística de teste estiver bem na região crítica, excedendo em muito o valor crítico de 5%, do que se ele estiver excedendo pouco esse valor.

Mesmo um valor pequeno de P pode ser difícil de interpretar do ponto de vista prático, quando estamos tomando decisões, pois, enquanto um valor pequeno de P indica significância estatística no sentido de que $H_0$ deve ser rejeitada em favor de $H_1$, o desvio real de $H_0$ que foi detectado pode ter pouca (se alguma) significância prática (engenheiros gostam de dizer "significância de engenharia"). Isso é particularmente verdade quando o tamanho da amostra $n$ é grande.

Por exemplo, considere o problema da taxa de queima de propelente do Exemplo 9-1, em que testamos $H_0 : \mu = 50$ centímetros por segundo versus $H_1 : \mu \neq 50$ centímetros por segundo, com $\sigma = 2,5$. Se supusermos que a taxa média é realmente 50,5 centímetros por segundo, então esse não será um desvio sério de $H_0 : \mu = 50$ centímetros por segundo, no sentido de que se a média realmente for 50,5 centímetros por segundo, não haverá efeito prático observável no desempenho do sistema de escape da aeronave. Em outras palavras, concluir que $\mu = 50$ centímetros por segundo quando ela é realmente 50,5 centímetros por segundo é um erro que não é caro e não tem significância prática. Para um tamanho de amostra razoavelmente grande, um valor verdadeiro de $\mu = 50,5$ centímetros por segundo conduzirá a um $\bar{X}$ da amostra que está perto de 50,5 centímetros por segundo e não queremos que esse valor de $\bar{X}$ proveniente da amostra resulte na rejeição de $H_0$.

O quadro a seguir mostra o valor P para testar $H_0 : \mu = 50$, quando observamos $\bar{X} = 50,5$ centímetros por segundo e a potência do teste com $\alpha = 0,05$, quando a média verdadeira é 50,5 para vários tamanhos $n$ de amostra:

| $n$ | Valor-p para $\bar{x} = 51.5$ | poder ($\alpha = 0.05$) quando $\mu = 50.5$ for verdadeira |
|---|---|---|
| 10 | 0.527 | 0.097 |
| 25 | 0.317 | 0.170 |
| 50 | 0.157 | 0.293 |
| 100 | 0.046 | 0.516 |
| 400 | $6.3 \times 10^{-5}$ | 0.979 |
| 1000 | $2.5 \times 10^{-10}$ | 1.000 |

A coluna de valor P nesse quadro indica que, para tamanhos grandes de amostra, o valor amostral observado de $\bar{X} = 50,5$ fortemente sugere que $H_0 : \mu = 50$ deve ser rejeitada, embora os resultados observados da amostra impliquem que, de um ponto de vista prático, a média verdadeira não difere muito do valor usado na hipótese $H_0 : \mu = 50$. A coluna de potência indica que se testarmos uma hipótese com um nível de significância fixo, a, e mesmo se houver pouca diferença prática entre a média verdadeira e o valor usado na hipótese, uma amostra de tamanho grande conduzirá, quase sempre, à rejeição de $H_0$.

A moral dessa demonstração é clara:

A moral dessa demonstração é clara:

Seja cuidadoso quando interpretar os resultados do teste de hipóteses quando a amostra tiver tamanho grande, visto que qualquer pequeno desvio do valor usado na hipótese, $H_0$, será provavelmente detectado, mesmo quando a diferença for de pouca ou nenhuma significância prática.