

# MAF 261 - Estatística Experimental

Prof. Fernando de Souza Bastos

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

16/08/2018

# Sumário

- 1 Hipóteses Unilaterais e Bilaterais
- 2 Valor-p ou p-Valor
- 3 Procedimento geral para Testes de Hipóteses
- 4 Significância Estatística versus Significância Prática

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

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

$$H_1 : \mu < 50cm/s$$

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 uma companhia de vidro. O engarrafador quer estar certo de que as garrafas satisfazem as especificações de pressão interna média, que, para as tais garrafas, a resistência mínima é 200g/l. 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:

$$\begin{cases} H_0 : \mu = 200g/l \\ H_1 : \mu > 200g/l \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \mu = 200g/l \\ H_1 : \mu < 200g/l \end{cases} \quad (1)$$

Considere a formulação com  $H_1 : \mu > 200g/l$ . 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  $H_1 : \mu < 200g/l$ . 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?  $H_1 : \mu > 200g/l$  ou  $H_1 : \mu < 200g/l$ ?

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



Qual formulação é a correta?  $H_1 : \mu > 200\text{g/l}$  ou  $H_1 : \mu < 200\text{g/l}$ ? 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.

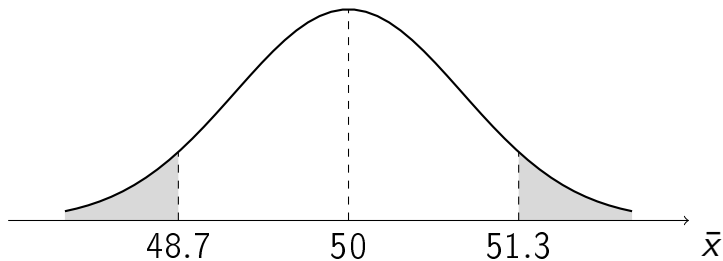
Considere o teste bilateral de hipóteses para a taxa de queima

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

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

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.





**Figura:** O valor-p é a área da região sombreada, quando  $\bar{x} = 51.3$

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

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, a hipótese nula seria rejeitada. De fato, a hipótese nula  $H_0 : \mu = 50$  seria rejeitada em qualquer nível de significância maior ou igual a 0,038.

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  $H_0$ .

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, 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, 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.

Essa aula foi retirada, com pequenas modificações, do livro:

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