Testes de hipóteses

Samuel Martins de Medeiros

Introdução

De maneira geral, existem duas grandes áreas na Estatística: a estimação de parâmetros, para mais informações sobre estimação, verifique nosso post sobre função de Verossimilhança, e o teste de hipóteses. Em particular, o teste de hipoteses consiste em avaliar uma afirmação a respeito de um parâmetro (média, variância, proporção, etc.) ou um conjunto de parâmetros. Tal afirmação recebe o nome de de Hipótese Nula (H_0), a afirmação alternativa recebe o nome de Hipótese Alternativa (H_1).

Para deixar essa ideia um pouco mais clara, suponha que queremos saber qual marca, A ou B, de um determinado tipo de blusa dura mais tempo. Conseguimos reescrever esse problema na forma de um sistema de hipóteses, a saber:

$$\begin{cases} H_0 : \text{Marca A dura mais que marca B} \\ H_1 : \text{Marca A dura menos que marca B} \end{cases}$$

Para realizar o teste, assumimos que é possível obter uma amostra aleatória $X_1, ..., X_n$ de uma distribuição $f(.;\theta)$.

Tambem é necessário definir a estatística de teste (T) e região de rejeição (R). Estatística de teste é um valor calculado a partir da amostra, seu valor define a regra de rejeição para uma hipótese, ele mostra o quanto seus dados observados correspondem à distribuição esperada sob a hipótese nula desse teste estatístico. Tomemos mais um exemplo: seja $X_1, ..., X_n$ uma amostra aleatória do produto de marca A, obtidos de uma distribuição $f(.;\theta)$, onde θ é o parâmetro tempo de duração médio. Estamos interessados em saber se o tempo de duração da marca A é igual ou maior que 5 anos, ou em outras palavras $H_0 : \theta \geq 5$. Um possível teste seria rejeitar H_0 se $\overline{x} < 5 - 10/\sqrt{n}$, onde $\overline{x}$ é a estatística de teste T e um estimador de θ. Formalmente, chamamos esse conjunto de possíveis resultados, para o qual se rejeita H_0, de região crítica. Assumiremos δ como representação do procedimento de testes de hipótese no dercorrer do post.

Um teste pode ser tanto aleatório quanto não aleatório. O exemplo anterior, por exemplo, é um ótimo exemplo de teste não aleatório. Já um teste aleatório poderia ser "jogue uma moeda para o alto, caso cara rejeite a hipótese nula".

Função Poder e Tipos de Erros Para cada teste aplicado sobre uma amostra obtida de uma distribuição $f(.;\theta)$ onde $\theta \in \Theta$ em que Θ representa o espaço paramétricos de possíveis valores para θ, teremos uma função poder associada.

- Função Poder: suponha um procedimento de teste δ. A função $\pi(\theta|\delta)$ é chamada função poder do teste δ. Se S_1 denota a região de rejeição de δ, então a função poder é determinada pela relação:

$$\pi(\theta|\delta) = \Pr(\boldsymbol{X} \in S_1|\theta)$$

Se δ é descrito em função da estatística de teste T e da região de rejeição R, então

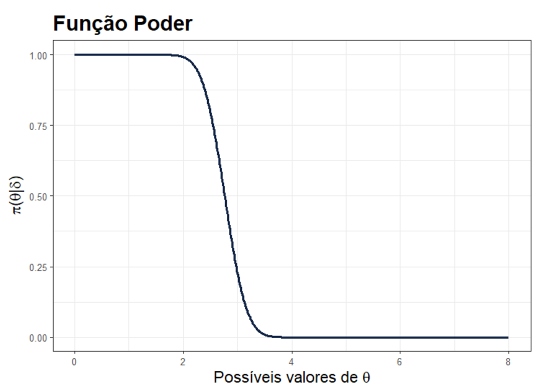
$$\pi(\theta|\delta) = \Pr(T \in R|\theta)$$

para todo $\theta \in \Theta$.

Sendo a função poder, a probabilidade de rejeitar a hipótese nula dado os possíveis valores do parâmetro em estudo θ, buscamos a o teste δ que minimize $\pi()$ para os valores de θ pertencentes ao espaço paramétrico de H_0 e a maximize quando θ pertence ao espaço paramétrico de H_1, ou em outras palavras, $\pi(\theta \in \Theta_0|\delta) = 0$ e $\pi(\theta \in \Theta_1|\delta) = 1$, onde Θ_0 representa o espaço paramétrico sob a hipótese nula e Θ_1 o espaço paramétrico sob a hipótese alternativa. Retomando o exemplo inicial onde rejeitamos a hipótese nula para $\overline{x} < 5 - 10/\sqrt{n}$, suponha que uma amostra aleatória do produto A, $x_1, ..., x_{20}$ foi obtida de uma distribuição Normal(θ, σ^2), com σ^2 conhecido e igual a 2, onde $\overline{x}$ é definido como o estimador de máxima verossimilhaça para média amostral. Obtendo, assim, a seguinte função poder,

$$\pi(\theta|\delta) = P\left(\overline{x} < 5 - \frac{10}{\sqrt{20}} = c|\theta\right)$$
$$= P\left(\frac{\overline{x} - \theta}{\sqrt{2/20}} < \frac{c - \theta}{\sqrt{2/20}}|\theta\right) = P(Z < z_c|\theta)$$

onde Z segue uma distribuição Normal$(0, 1)$, tendo assim:



Como dito inicialmente, testamos se a hipótese nula é falsa e, portanto, se a hipótese alternativa é verdadeira, ou vice-versa. Nesse contexto, dois tipos de erros podem ser cometidos:

- Erro do Tipo I: rejeitar a hipótese nula quando a mesma é verdadeira.
- Erro do Tipo II: aceitar a hipótese nula quando a mesma é falsa.

Hipótese simples

As hipóteses de um teste podem ser da forma simples ou composta. Uma hipótese simples, é aquela onde o espaço de possíveis valores de θ é definido em apenas um ponto, onde a distribuição do parâmetro é completamente especificada ($H_0 : \theta = \theta_0, f(.;\theta_0)$ por exemplo). Por outro lado, uma hipótese composta é aquela cuja distribuição não é especificada completamente e θ pode assumir um conjunto de valores Θ ($H_0 : \theta \in \Theta,\ f(.;\Theta)$ por exemplo). A discussão acerca de hipótese simples versus hipótese simples não é muito vista na prática, porém serve como ótima introdução ao tema.

Testes de razão de verossimilhança simples Suponha que $X_1, ..., X_n$ uma amostra aleatória de uma distribuição $f_0(.;\theta_0)$ ou $f_1(.;\theta_1)$. Um teste de $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ é um teste da razão de verossimilhança se é definido como: onde k é uma constante não negativa e $L(x_1, \ldots, x_n)$ é a função de verossimilhança associada à função de densidade $f(\cdot)$. Rejeitamos a hipótese nula para um valor de $\lambda = L_0(\cdot)/L_1(\cdot)$ inferior a k, pois, seguindo a linha de raciocínio da razão das funções, $L_1(\cdot)$ é maior que $L_0(\cdot)$, nos dando mais indícios de que a amostra venha de uma população com distribuição $f_1(\cdot)$ em vez de uma $f_0(\cdot)$. Exemplo: seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição $N(\theta, 1)$ na qual queremos testar $H_0 : \theta = 0 \times H_1 : \theta = 1$. Tendo a função de verossimilhança como:

$$L(\theta; \underline{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n exp\left\{-\sum_{i=1}^{n} \frac{(x_i - \theta)^2}{2}\right\}$$

Obtendo o teste de razão de verossimilhança,

$$\lambda = \frac{L_0(X)}{L_1(X)} = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n exp\left\{-\sum_{i=1}^{n} \frac{x_i^2}{2}\right\}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n exp\left\{-\sum_{i=1}^{n} \frac{(x_i-1)^2}{2}\right\}}$$

$$= exp\left\{-\sum_{i=1}^{n} \frac{x_i^2}{2} + \sum_{i=1}^{n} \frac{(x_i - 1)^2}{2}\right\} < k$$

Que pode ser reescrito como

$$\sum_{i=1}^{n} x_i > \frac{n}{2} - logk = k*$$

Ou seja, rejeitamos H_0 para um somatório de x maior que alguma constante k*.

Testes Mais Poderosos Antes de falar sobre os testes mais poderosos, uma definição deve ser esclarecida: o tamanho do teste. Vamos admitir um teste δ cuja hipótese nula seja $H_0 : \theta \in \Theta_0$ ($H_0 : \theta < \theta_0$, por exemplo), em que $\theta \subset \Theta_0$ (ou seja, Θ_0 é um subconjunto do espaço paramétrico Θ). Assim, o tamanho do teste é definido como . Esclarecida essa definição, daremos prosseguimento ao assunto. Assim como já comentado, queremos um teste δ em que $\pi(\theta_0|\delta) =$ P[Rejeitar H_0 | H_0 verdadeiro] seja a menor possível e que $\pi(\theta_1) =$ P[Rejeitar H_0 | H_0 falsa] seja a maior possível. Em um mundo perfeito, $\pi(\theta_1) = 1$ e $\pi(\theta_0) = 0$, isto é, quando os erros do tipo I e II são minimizados simultâneamente. Entretanto, na prática, uma das metodologias aplicadas de forma a definir o melhor teste possível é minimizar o erro do tipo II fixando o erro do tipo I.

Teste Mais Poderoso: Um teste $\delta*$ em que $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$ é definido como teste mais poderoso de tamanho α, com $0 < \alpha < 1$, se e somente se:

i. $\pi(\theta|\delta*) = \alpha$;

ii. $\pi(\theta_1|\delta*) > \pi(\theta_1|\delta)$, para qualquer outro teste δ onde $\pi(\theta_0|\delta) < \alpha$.

Ou seja, podemos considerar um teste $\delta*$ como sendo o teste mais poderoso se, para qualquer outro teste de tamanho α ou menor do que α, ele possuir o maior poder.

O lema (ou método) a seguir é muito útil para encontrar testes mais poderosos.

- **Lemma Neyman-Pearson:** seja $X_1, ..., X_n$ uma amostra aleatória de uma distribuição com densidade $f(x;\theta)$, onde θ pode assumir os valores θ_1 ou θ_0 e $0 < \alpha < 1$. Considere $k*$ uma constante positiva e C* um subconjunto do espaço de valores para X_i. Assim,

(i) $P_{\theta_0}[(X_1, \ldots, X_n) \in C^*] = \alpha.$

(ii) $\lambda = \dfrac{L(\theta_0; x_1, \ldots, x_n)}{L(\theta_1; x_1, \ldots, x_n)} = \dfrac{L_0}{L_1} \leq k^*$

e $\lambda >$ k* se $(x_1, ..., x_n) \in C*$.

Então, considerando um teste de hipóteses simples, temos que o teste para essa região crítica é o teste mais poderoso. Vamos mostrar um exemplo para melhor compreensão.

Exemplo: seja $X_1, ..., X_n$ uma amostra aleatória de uma distribuição Bernoulli(θ) e seja o teste $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, $\theta_1 > \theta_0$. O teste mais poderoso de tamanho α para testar H_0 contra H_1 é da forma

$$\varphi(\underset{\sim}{x}) = \begin{cases} 1, & caso\ \lambda(\underset{\sim}{x}) = \dfrac{\theta_1^{n\overline{x}}(1-\theta_1)^{n-n\overline{x}}}{\theta_0^{n\overline{x}}(1-\theta_0)^{n-n\overline{x}}} > k, \\ \gamma, & caso\ \lambda(\underset{\sim}{x}) = k, \\ 0, & caso\ \lambda(\underset{\sim}{x}) < k, \end{cases}$$

onde k e γ é determinada de maneira que . Agora, se

$$\lambda(\underset{\sim}{x}) = \left(\frac{\theta_1}{\theta_0}\right)^{n\overline{x}} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-n\overline{x}}$$

dado que $\theta_1 > \theta_0$ e $\lambda(x)$ é uma função crescente de $\overline{x}_n$,segue que $\lambda(x) > k$ se e somente se $\overline{x}_n > k_1$, sendo k_1 uma constante. Então, o teste mais poderoso de tamanho α é da forma

$$\varphi(\underset{\sim}{x}) = \begin{cases} 1, & caso\ \overline{x} > k_1 \\ \gamma, & caso\ \overline{x} = k_1, \\ 0, & caso\ contrário \end{cases} .$$

Ainda, k_1 e γ são determinados da forma

$$\begin{aligned} \alpha &= \mathrm{E}_{\theta_0}[\varphi(X)] = P_{\theta_0}(\overline{x} > k_1) + \gamma P_{\theta_0}(\overline{x} = k_1) \\ &= \sum_{r=k_1+1}^{n} \binom{n}{r} \theta_0^{nr}(1-\theta_0)^{n-nr} + \gamma \binom{n}{k_1} \theta_0^{nk_1}(1-\theta_0)^{n-nk_1}. \end{aligned}$$

Observe que o teste mais poderoso de tamanho α é independente de θ_1 quando $\theta_1 > \theta_0$, e é, portanto, o teste mais poderoso para verificar se $\theta = \theta_0$ contra $\theta > \theta_0$.

Testes para hipóteses compostas

Generalizaremo-los para os teste de hipóteses compostos. A princípio, começaremos com o método mais geral para testar hipóteses, que, geralmente, não é o que fornece resultados mais precisos, mas é aplicável em todo tipo de situação. Considere $X_1, ..., X_n$ uma amostra aleatória obtida de uma função de densidade $f(x;\theta)$, $\theta \in \Theta$, e um teste do tipo $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$.

- **Teste de Razão de Verossimilhança Generalizada**: suponha $L(\theta; X_1, ..., X_n)$ a função de verossimilhança para a amostra $X_1, ..., X_n$. O teste de razão de verossimilhança generalizada, denotado por λ, é definido como:

$$\lambda = \lambda_n = \lambda(x_1, \ldots, x_n) = \frac{\sup_{\theta \in \underline{\Theta}_0} L(\theta; x_1, \ldots, x_n)}{\sup_{\theta \in \underline{\Theta}} L(\theta; x_1, \ldots, x_n)}.$$

onde λ se torna uma função da amostra definida no intervalo [0,1]. Assim como no Teste de Razão de Verossimilhança para hipóteses simples, rejeitamos a H_0 para algum $\lambda_0 > \lambda$, em que λ_0 é uma constante definida no intervalo [0,1].

- **Testes Uniformemente Mais Poderosos (TUMP)**: um teste $\delta*$ do tipo $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1 = \theta - \Theta_0$ é definido como TUMP de tamanho α se e somente se

$$(i) \quad \sup_{\theta \in \Theta^0} [\pi_{\delta*}(\theta)] = \alpha$$

$$(ii) \quad \pi_{\delta*}(\theta) > \pi_{\delta}(\theta)$$

para todo $\theta \in \Theta - \Theta_0$ e para qualquer teste δ de tamanho menor ou igual a α.

Conclusão

Na literatura, podemos encontrar formas diferentes de testar hipóteses das vistas neste tutorial, mas elas fogem do escopo deste post e por isso não foram abordadas. Ainda assim, fomos capazes de aprender alguns dos métodos para testar hipóteses estatísticas mais utilizados, além de métodos para achar o melhor tipo de teste. Espero que o texto tenha sido esclarecedor e de ajuda ao leitor. Para mais informações ou dúvidas, escreva-nos em : comunicacao@observatorioobstetricobr.org