Testes de Hipóteses

Tópicos de Inferência Estatística

José Passos

ISEG-ULisboa

29 de Novembro de 2019

Tabela de conteúdos

- Vamos assumir que a população X tem distribuição na família $\mathcal{F} = \{f(.|\theta) : \theta \in \Theta\}$.
- Uma hipótese estatística é uma conjectura sobre aspectos desconhecidos da distribuição da população, em particular sobre o parâmetro que indexa $\mathcal{F}$
- Qualquer hipótese estatística induz uma partição em Θ do seguinte modo:
  - $\Theta = \Theta_0 \cup \Theta_1$
  - $\Theta_0 \cap \Theta_1 = \emptyset$

  onde Θ_0 representa o conjunto dos valores de θ que verifica a conjectura.
- Temos então as hipóteses:
  - $H_0 : \theta \in \Theta_0$
  - $H_1 : \theta \in \Theta_1$

- Um teste de hipóteses corresponde a uma estratégia para averiguar qual das duas hipóteses é melhor suportada pelos dados observados na amostra.
- Uma hipótese estatística diz-se simples quando especifica um único valor para o parâmetro desconhecido.
- Caso contrário, as hipóteses dizem-se compostas.

- Um teste de hipóteses é uma regra que permite especificar um subconjunto do espaço amostral, $W \subset \mathcal{X}$, tal que,
  - se $(x_1, x_2, \ldots, x_n) \in W$ rejeita-se H_0;
  - se $(x_1, x_2, \ldots, x_n) \notin W$ aceita-se H_0
- Ao conjunto W dá-se o nome de região crítica ou região de rejeição.
- Ao conjunto complementar $\overline{W}$ dá-se o nome de região de aceitação.
- Portanto, um teste estatístico induz uma partição do espaço amostra em dois subconjuntos, W e $\overline{W}$ que verificam,
  - $\mathcal{X} = W \cup \overline{W}$
  - $W \cap \overline{W} = \emptyset$

- Na generalidade dos casos existe uma estatística, T, e um conjunto W_T onde a região de rejeição é tal que,

$$(x_1, x_2, \ldots, x_n) \in W \Leftrightarrow T(x_1, x_2, \ldots, x_n) \in W_T$$

- Neste caso a decisão e rejeitar ou não a hipótese nula depende do valor da estatística e esta designa-se por estatística de teste. De modo análogo W_T designa-se por região de rejeição ou região crítica

- Em suma: ingredientes de um teste de hipóteses:
  - hipótese nula, H_0
  - hipótese alternativa, H_1
  - uma estatística de teste cujo valor observado determina a decisão
  - uma região crítica (valores da estatística que correspondem à rejeição de H_0)
- Quando se toma uma decisão podem-se cometer dois tipos de erros:
  - Erro de 1ª espécie (Tipo I): rejeitar H_0 sendo verdadeira
  - Erro de 2ª espécie (Tipo II): aceitar H_0 sendo falsa

Hipóteses Simples:

- Neste caso $\Theta = \{\theta_0, \theta_1\}$ e pretende-se testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$
- Dimensão do teste: é a probabilidade de cometer um erro tipo I e é dada por,

$$\begin{aligned} \alpha &= P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}) \\ &= P[(X_1, \ldots, X_n) \in W \mid \theta = \theta_0] \end{aligned}$$

- Potência do teste: é o complementar da probabilidade de um erro tipo II

$$\begin{aligned} \beta &= P(\text{rejeitar } H_0 \mid H_0 \text{ falsa}) \\ &= P[(X_1, \ldots, X_n) \in W \mid \theta = \theta_1] \end{aligned}$$

- Uma diminuição de α implica uma diminuição de β e portanto um aumento da probabilidade de erro tipo II $(1-\beta)$
- A estratégia de **Neyman-Pearson** consiste em fixar α e procurar o teste com menor probabilidade de erro tipo II (ou maior potência): teste mais potente de dimensão α
- **Definição (teste mais potente):** fixada a dimensão do teste, o teste mais potente é aquele cuja região crítica minimiza a probabilidade de erro de 2ª espécie, ou maximiza a potência. Diz-se que a região crítica associada é a mais potente.

Lema de Neyman-Pearson: seja $X_1, \ldots, X_n$ uma amostra casual de uma população com densidade $f(. \mid \theta)$, $\theta \in (\theta_0, \theta_1)$. Seja $c > 0$, $\alpha \in (0, 1)$ e considere-se $W \subset \mathcal{X}$ definido pelas condições,

$$\frac{\prod_{i=1}^{n} f(x_i \mid \theta_1)}{\prod_{i=1}^{n} f(x_i \mid \theta_0)} = \frac{L(\theta_1 \mid x_1, \ldots, x_n)}{L(\theta_0 \mid x_1, \ldots, x_n)} > c \Leftrightarrow (x_1, \ldots, x_n) \in W$$

$$P[(X_1, \ldots, X_n) \in W \mid \theta = \theta_0] = \alpha$$

Então o teste associado à região W é o mais potente de dimensão α para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

Nota: se $T = T(X_1, \ldots, X_n)$ é estatística suficiente para θ, tem-se pelo teorema da factorização,

$$L(\theta|x_1, \ldots, x_n) = G(t; \theta)h(x_1, \ldots, x_n)$$

e portanto, a razão de verosimilhanças simplifica-se para,

$$\frac{L(\theta_1 \mid x_1, \ldots, x_n)}{L(\theta_0 \mid x_1, \ldots, x_n)} = \frac{G(t; \theta_1)}{G(t; \theta_0)}$$

Notas:

- A aplicação do lema faz-se em duas fases:
  - determinar a estatística de teste e a forma da região crítica;
  - encontrar a constante "c"utilizando a condição de o teste ser de dimensão α.
- A estatística de teste é suficiente;
- O valor de θ_1 não é relevante para definir W. Apenas é necessário saber se $\theta_0 > \theta_1$ ou se $\theta_0 < \theta_1$;
- A potência de teste MP depende do valor de θ_1
- Como regra prática, se θ é uma média, variância ou proporção, a estatística de teste é um estimador de θ e a região de rejeição está do lado da hipótese alternativa.

Hipótese simples contra hipótese composta:

- Neste caso H_0 é simples e H_1 é composta mas unilateral,

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta > \theta_0 \text{ (ou versus } H_1 : \theta < \theta_0)$$

- A estratégia também é de fixar o erro tipo I
- Neste caso o erro tipo II é uma função de θ, para $\theta \in \Theta_1$
- Neste caso fala-se em função potência e em teste uniformemente mais potente
- **Definição (Função Potência):** a função potência do teste $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ (ou versus $H_1 : \theta < \theta_0$), com região crítica W é dada por,

$$\begin{aligned}\beta(\theta) &= P(\text{rejeitar } H_0 \mid H_0 \text{ falsa})\\ &= P[(X_1, \ldots, X_n) \in W \mid \theta], \ \text{com } \theta \in \Theta_1\end{aligned}$$

onde $\Theta_1 =]\theta_0, +\infty[$ (ou $\Theta_1 =]-\infty, \theta_0[$).

- **Definição (Teste uniformemente mais potente - UMP):** ao testar a hipótese nula $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ (ou versus $H_1 : \theta < \theta_0$), considerem-se dois testes de dimensão α : T_1 com função potência $\beta(\theta)$ e T_2 com função potência $\beta^*(\theta)$. Dizemos que T_1 é uniformemente mais potente do que T_2 se $\beta(\theta) \geq \beta^*(\theta)$, $\forall \theta \in \Theta_1$.
  Um teste T de dimensão α diz-se uniformemente mais potente para as hipóteses se for uniformemente mais potente do que qualquer outro teste de dimensão α.

Hipótese nula unilateral contra alternativa unilateral:

- Neste caso H_0 é composta unilateral esquerda (direita) e H_1 é composta unilateral direita (esquerda),

$$H_0 : \theta \leq \theta_0 \;\text{ versus }\; H_1 : \theta > \theta_0$$

ou,

$$H_0 : \theta \geq \theta_0 \;\text{ versus }\; H_1 : \theta < \theta_0$$

- A probabilidade de erro de 1ª espécie é também função de $\theta \in \Theta_0$
- A dimensão do teste define-se como o erro máximo de 1ª espécie,

$$\alpha = \sup_{\theta \in \Theta_0} P(\text{rejeitar } H_0 \mid \theta)$$

- Generalização do conceito de função potência:

$$\beta(\theta) = P[(X_1, \ldots, X_n) \in W \mid \theta], \text{ com } \theta \in \Theta$$

- Para $\theta \in \Theta_0$, $\beta(\theta)$ é a probabilidade de erro de 1ª espécie
- Para $\theta \in \Theta_1$, $\beta(\theta)$ é a probabilidade de não cometer um erro de 2ª espécie
- **Nota:** quando as hipóteses são compostas há situações em que a aplicação do Lema de Neyman-Pearson conduz a regiões UMP. Os resultados seguintes dizem-nos em que famílias existem testes UMP.

Definição (Razão de verosimilhança monótona - RVM): a família $\mathcal{F} = \{f(.|\theta) : \theta \in \Theta\}$ onde Θ é um intervalo real, tem RVM na estatística T quando para quaisquer $\theta' > \theta$, com $\theta', \theta \in \Theta$, a razão de verosimilhanças,

$$\frac{L(\theta' \mid x_1, \ldots, x_n)}{L(\theta \mid x_1, \ldots, x_n)}$$

é função monótona não decrescente de T.

- Nota: o importante é que a razão de verosimilhança seja uma função monótona de uma estatística; se for não crescente em T será não decrescente em $S = -T$.

Teorema (Karlin-Rubin): se $(X_1, \ldots, X_n)$ é amostra casual de uma população com RVM na estatística T, então para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$, os testes com região de rejeição da forma $W = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) > k\}$ são UMP da sua dimensão.

- Nota: o teorema cobre também os casos em que a razão de verosimilhanças é monótona não crescente, ou em que as hipóteses são do tipo $H_0 : \theta \geq \theta_0$ contra $H_1 : \theta < \theta_0$. Basta considerar a parametrização $\phi = -\theta$ ou $\phi = 1/\theta$ e a estatística $W = -T$ ou $W = 1/T$

Hipótese simples contra composta bilateral:

- Neste caso H_0 é simples e H_1 é composta bilateral,

$$H_0 : \theta = \theta_0 \ \text{ versus } \ H_1 : \theta \neq \theta_0$$

- Regra geral, não existem testes UMP
- Uma solução consiste em limitar a classe de testes onde se procura o óptimo, ou seja, procurar o teste UMP entre a classe de testes de tamanho α centrados. Esse teste, quando existe, denomina-se de teste UMPU (uniformly most powerful unbiased).
- Na prática consideram-se duas regiões de distribuição nas abas da distribuição da estatística de teste, distribuindo metade do erro de 1ª espécie ($\alpha/2$) por cada uma delas.

Ensaios de significância:

- Os ensaios de significância distinguem-se dos testes de hipóteses pela ausência da hipótese alternativa
- Baseados numa estatística de teste T adequada à hipótese nula,
  - a distribuição por amostragem de T é conhecida sob H_0
  - determinados valores de T lançam dúvidas sobre a validade de H_0
  - dado um valor observado da estatística, t_{obs}, calcular a probabilidade de observar, sob H_0, valores de T tão ou mais incompatíveis com H_0 do que t_{obs}, $p = P(T < t_{obs})$ (valor-p)
  - se p pequeno, ou H_0 é verdadeira e obteve-se por acaso um acontecimento pouco provável sob H_0, ou os dados observados não foram efectivamente gerados por H_0
  - em geral opta-se pela segunda explicação e rejeita-se H_0 se p pequeno

- Na prática procede-se do seguinte modo:
  - fixar α (em geral 1%, 5% ou 10%)
  - calcular o valor observado da estatística de teste, t_{obs}
  - calcular o valor-p
  - se $p \leq \alpha$, rejeitar H_0 ao nível α
  - se $p > \alpha$ não rejeitar H_0 ao nível α