5. Inferência estatística

Problemas de inferência

Inferir significa fazer afirmações sobre algo **desconhecido**.

A inferência estatística tem como objetivo fazer afirmações sobre uma característica de uma **população** a partir do conhecimento de dados de uma parte desta população (isto é, uma **amostra** de **n** observações).

A população é representada por uma distribuição de probabilidade com **parâmetro**(s) cujo(s) valor(es) é (são) **desconhecido**(s).

Fazemos inferências sobre o(s) parâmetro(s).

Definições

Parâmetro: característica numérica (desconhecida) da distribuição dos elementos da população.

Espaço paramétrico: conjunto de todos os valores possíveis que o parâmetro pode assumir.

Estatística: qualquer função dos elementos de uma amostra aleatória, a qual não depende de parâmetros desconhecidos.

Estimador: Função da amostra, construída com a finalidade de representar, ou estimar um parâmetro de interesse na população.

Estimativa: Valor numérico que um estimador assume.

Erro padrão: é o desvio padrão de sua distribuição amostral ou uma estimativa desse desvio padrão.

Problemas de inferência

Se θ é um parâmetro da distribuição de uma v. a. X e $X_1, X_2, \ldots, X_n$ é uma **amostra aleatória** desta distribuição, encontramos três problemas típicos:

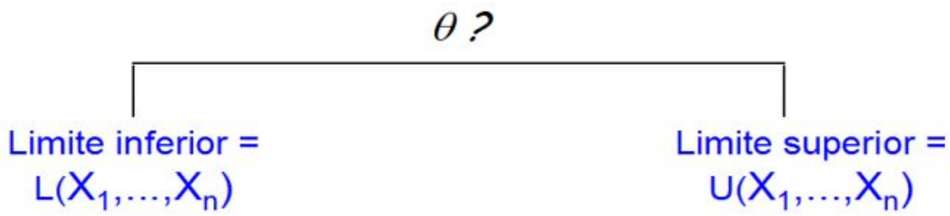
1. **Estimação pontual**
   Apresentar um valor para θ, que é uma função da amostra $X_1, X_2, \ldots, X_n$ ("cálculo" de θ), chamada de estimador de θ.

   Espera-se que o estimador tenha **boas propriedades**.

Problemas de inferência

2. **Estimação intervalar**

   Apresentar um **intervalo de possíveis valores** para θ, chamado de **intervalo de confiança**. Os limites do intervalo são funções da amostra $X_1,...,X_n$ (são aleatórios).



   A probabilidade de que o intervalo contenha θ deve ser **alta**.

   A **amplitude** do intervalo deve ser tão **pequena** quanto possível (intervalo mais **preciso**).

Problemas de inferência

3. **Teste de hipóteses**

   Uma **hipótese estatística (H)** é uma afirmação sobre o valor de θ. Pode ser **verdadeira** ou **falsa**.

   Por exemplo, se θ é a probabilidade de sucesso no modelo binomial, H: $\theta = ½$, H: $\theta \neq ½$ e H: $\theta > ¾$ são exemplos de hipóteses.

   Com base na amostra $X_1,\ldots,X_n$, formulamos uma **regra de decisão** que permita concluir pela **rejeição** ou **não rejeição** (aceitação) de H. A decisão pode ser **correta** ou **errada**.

PROPRIEDADES DOS ESTIMADORES

5.1

Estimação pontual

Como vimos, é importante que os estimadores possuam algumas características desejáveis. Consideremos uma amostra $X_1,X_2,...,X_n$ de uma população X. Seja θ o parâmetro de interesse da população que desejamos estimar, como por exemplo μ = E(X) ou σ^2 = Var(X).

1 - Estimador não viciado

Um estimador $\hat{\theta}$ = T(X_1,X_2,...,X_n) é dito **não viciado** (ou não viesado) para algum parâmetro populacional θ se

$$\mathbf{E(\hat{\theta}) = \theta,}$$

para todo θ.

Observação: Caso a igualdade acima não ocorra, dizemos que o estimador $\hat{\theta}$ é um estimador viciado (viesado) e a diferença V($\hat{\theta}$,θ) =E($\hat{\theta}$) - θ é chamada de vício (viés) de $\hat{\theta}$

2 - Estimador consistente

Um estimador $\hat{\theta}$ é chamado **consistente** se a probabilidade dele diferir do verdadeiro valor θ em menos do que c, onde c é um número arbitrário positivo e pequeno, tende a 1, quando o tamanho da amostra (n) aumenta; ou seja, se

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta|) = 1.$$

- **Proposição:**
  As condições suficientes para um estimador ser consistente são:
  $\lim_{n \to \infty} E(\hat{\theta}) = \theta$ e $\lim_{n \to \infty} Var(\hat{\theta}) = 0.$

  - **Observação:**
    . Observe que, se $\hat{\theta}$ for um estimador não viciado de θ, então a primeira condição estará claramente satisfeita.

    . Esta é uma propriedade assintótica de um estimador. Ela é aplicada a amostras "suficientemente grandes".

3 - Estimador eficiente

Um estimador é **eficiente** se **for não viesado** e entre os estimadores não viesados, apresentar a **menor variância**.

Ou seja, suponha que T e T' sejam dois estimadores não viciados de um mesmo parâmetro θ. Se

$$Var(T) < Var(T')$$

então dizemos que T é um estimador **mais eficiente** do que T'.

Observação: Quanto menor for a diferença entre o estimador T e o parâmetro θ, menor será o erro cometido ao estimar o parâmetro θ pelo estimador T. Esta diferença $e = T - \theta$ é chamada de **erro amostral**.

Estimação pontual – método de substituição

Estimação pontual – método de substituição

1. Distribuição binomial. X ~ B(n, p). Vimos que **E(X) = np**.
   Um estimador para $p: \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i =$ proporção amostral de sucessos.

2. Distribuição de Poisson. X ~ Po(μ). Vimos que **E(X) = $\boldsymbol{\mu}$**.
   Um estimador para $\mu: \overline{X}$

3. Distribuição exponencial. X ~ Ex(λ). Vimos que **E(X) = 1 / $\boldsymbol{\lambda}$**.
   Um estimador para $\lambda: = \frac{1}{\overline{X}}$.

4. Distribuição normal. X ~ N(μ,σ^2). Vimos que **E(X) = $\boldsymbol{\mu}$** e **Var(X) = $\boldsymbol{\sigma^2}$**.
   Um estimador para $\mu: \overline{X}$.   Um estimador para $\sigma^2 : s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

   Observação: Existem outros métodos de estimação.

Resultado

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória independente e igualmente distribuída de uma população com média μ e variância $\sigma^2 < \infty$. Então

a) $E(\overline{X}) = \mu$,

b) $E(s^2) = \sigma^2$.

Demonstração:
Para a média amostral temos que

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}n\mathbb{E}(X_1) = \mu.$$

De maneira similar, para a variância amostral, temos que

$$\mathbb{E}(s^2) = \mathbb{E}\left(\frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]\right) = \frac{1}{n-1}(n\mathbb{E}(X_1^2) - n\mathbb{E}(\overline{X}^2)$$

ou seja,

$$\mathbb{E}(s^2) = \frac{1}{n-1}\left(n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \sigma^2.$$

Desta forma, podemos concluir que $\overline{X}$ e s^2 são estimadores não viciados da média populacional μ e da variância populacional σ^2.

Resultado

A partir do resultado anterior, é evidente que a média amostral $\overline{X}$ é estimador consistente da média populacional μ.

Resultado

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população com distribuição normal N(μ, σ^2).

O estimador $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ é viesado, pois E(s) $\approx \sigma\left(1 - \frac{1}{4n}\right)$. Logo, o vício é aproximandamente $\frac{\sigma}{4n}$.

Portanto, temos que o estimador s é viesado embora s^2 não seja, pois como sabemos $E(s^2) = \sigma^2$.

Exemplo

A quantidade de tempo, em minutos, que um passageiro gasta esperando na fila de check-in de um aeroporto é uma variável aleatória com média e variância desconhecidos e distribuição normal.
Suponha que uma amostra aleatória de 10 passageiros foi observada, em que os tempos foram: 10; 9; 11; 8,5; 7,1; 9; 9,5; 8; 10; 7,8.

a) Encontre a estimativa da média e variância.
b) Essas estimativas são não viciadas?

a) $\bar{X} = 8{,}99$

$S^2 = 1{,}394$

estimador para μ: $\hat{\mu} = \bar{X} = 8{,}99$

estimador para σ^2: $\hat{\sigma}^2 = S^2 = 1{,}394$

b) São não viciados (utilizando do resultado)