Fundação Getúlio Vargas
Inferência Estatística
Wellington Silva

Abstract of statistical inference
Due: November 18, 2021

Resumo de Inferência Estatística

Sumário

Aula 1: O que é e para que serve Inferência Estatística?

Definição 1 (Modelo estatístico: informal) *Um **modelo estatístico** consiste na identificação de variáveis aleatórias de interesse (observáveis e potencialmente observáveis), na especificação de uma distribuição conjunta para as variáveis aleatórias observáveis e na identificação dos parâmetros (θ) desta distribuição conjunta. Às vezes é conveniente assumir que os parâmetros são variáveis aleatórias também, mas para isso é preciso especificar uma distribuição conjunta para θ.*

Definição 2 (Modelo estatístico: formal) *Seja $\mathcal{X}$ um espaço amostral qualquer, Θ um conjunto não-vazio arbitrário e $\mathcal{P}(\mathcal{X})$ o conjunto de todas as distribuições de probabilidade em $\mathcal{X}$. Um modelo estatístico **paramétrico** é uma função $P : \Theta \to \mathcal{P}(\mathcal{X})$ que associa a cada $\theta \in \Theta$ uma distribuição de probabilidade P_θ em $\mathcal{X}$.*

Definição 3 (Afirmação probabilística) *Dizemos que uma afirmação é **probabilística** quando ela utiliza conceitos da teoria de probabilidade para falar de um objeto.*

Definição 4 (Inferência Estatística) *Uma **inferência estatística** é uma **afirmação probabilística** sobre uma ou mais partes de um modelo estatístico.*

Definição 5 (Estatística) *Suponha que temos uma coleção de variáveis aleatórias $X_1, X_2, \ldots, X_n \subseteq \mathbf{R}^n$ e uma função $r : \mathbf{X} \to R^m$. Dizemos que a variável aleatória $T = r(X_1, X_2, \ldots, X_n)$ é uma **estatística**.*

Definição 6 (Permutabilidade) *Uma coleção finita de variáveis aleatórias $X_1, X_2, \ldots, X_n$ com densidade conjunta f é dita **permutável** se*

$$f(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) \tag{1}$$

para qualquer permutação $\pi = \{\pi(1), \pi(2), \ldots, \pi(n)\}$ dos seus elementos. Uma coleção finita é permutável se qualquer subconjunto finito é permutável.

Aula 2: Distribuição a priori e a posteriori

Definição 7 (Distribuição a priori) *Se tratamos o parâmetro θ como uma variável aleatória, então a **distribuição a priori** é a distribuição que damos a θ antes de observarmos as outras variáveis aleatórias de interesse. Vamos denotar a função de densidade/massa de probabilidade da priori por $\xi(\theta)$.*

Definição 8 (Distribuição a posteriori) *Considere o problema estatístico com parâmetros θ e variáveis aleatórias observáveis $X_1, X_2, \ldots, X_n$. A distribuição condicional de θ dados os valores observados das variáveis aleatórias, $\boldsymbol{x} := \{x_1, x_2, \ldots, x_n\}$ é a **distribuição a posteriori** de θ, denotamos por $\xi(\theta \mid \boldsymbol{x})$ a f.d.p./f.m.p. condicional a $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.*

Teorema 1 (Distribuição a posteriori: derivação) *Considere a amostra aleatória $X_1, X_2, \ldots, X_n$ de uma distribuição com f.d.p./f.m.p. $f(x \mid \theta)$. Se a distribuição a priori é $\xi(\theta)$, temos*

$$\xi(\theta \mid x) = \frac{\xi(\theta)\Pi_{i=1}^{n} f(x_i \mid \theta)}{g_n(x)}, \ \theta \in \Omega \tag{2}$$

Chamamos $g_n(x)$ de distribuição marginal de $X_1, X_2, \ldots, X_n$.

Definição 9 (Função de verossimilhança) *Quando encaramos a f.d.p./f.m.p. $f(x_1, x_2, \ldots, x_n \mid \theta)$ como uma função do parâmetro θ, chamamos esta função de **função de verossimilhança**, e podemos denotá-la como $L(\theta; x)$ ou, quando a notação não criar ambiguidade, simplesmente $L(\theta)$.*

Aula 3: Prioris conjugadas e função de perda

Definição 10 (Hiper-parâmetros) *Seja $\xi(\theta \mid \phi)$ a distribuição a priori para o parâmetro θ, indexada por $\phi \in \Phi$. Dizemos que ϕ é(são) o(s)* ***hiper-parâmetro(s)*** *da priori de θ.*

Definição 11 (Priori conjugada) *Suponha que $X_1, X_2, \dots$ sejam condicionalmente independentes dado θ, com f.d.p./f.m.p. $f(x \mid \theta)$. Defina*

$$\Psi = \left\{ f : \Omega \to (0, \infty), \int_\Omega f dx = 1 \right\} \tag{3}$$

onde Ω é o espaço de parâmetros. Dizemos que Ψ é uma ***família de distribuições conjugadas*** *para $f(x \mid \theta)$ se $\forall f \in \Psi$ e toda realização $\boldsymbol{x}$ de $X = X_1, X_2, \dots, X_n$*

$$\frac{f(\boldsymbol{x} \mid \theta) f(\theta)}{\int_\Omega f(\boldsymbol{x} \mid \theta) f(\theta) d\theta} \in \Psi \tag{4}$$

Teorema 2 (Distribuição a posteriori da média de uma normal) *Suponha que $X_1, X_2, \dots, X_n$ formam uma amostra aleatória com distribuição normal e com média desconhecida θ e variância $\sigma^2 > 0$, conhecida e fixa. Suponha que $\theta \sim Normal(\mu_0, v_0^2)$ a priori. Então*

$$\xi(\theta \mid x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{(\theta - \mu_1)^2}{2v_1^2} \right), \tag{5}$$

onde

$$\mu_1 := \frac{\sigma^2 \mu_0 + n v_0^2 \overline{x}_n}{\sigma^2 + n v_0^2} \text{ e } v_1^2 := \frac{\sigma^2 v_0^2}{\sigma^2 + n v_0^2} \tag{6}$$

Definição 12 (Priori imprópria) *Seja $\xi : \Lambda \to (0, \infty), \Omega \subseteq \Lambda$, uma função tal que $\int_\Omega \xi(\theta) d\theta = \infty$. Se utilizamos ξ como uma p.d.f. [1] para θ, dizemos que ξ é uma* ***priori imprópria*** *para θ.*

Definição 13 (Estimador) *Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias com distribuição conjunta indexada por θ. Um* ***estimador*** *de θ é qualquer função real δ: $X_1, X_2, \dots, X_n \to \mathbb{R}^d, d \geq 1$.*

Definição 14 (Estimativa) *Dizemos que o valor de δ avaliado nas realizações de $X_1, X_2, \dots, X_n$, $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$, $\delta(\boldsymbol{x})\}$ é uma* ***estimativa*** *de θ.*

Definição 15 (Função de perda) *Uma função de perda é uma função real em duas variáveis*

$$L : \Omega \times \mathbb{R}^d \to \mathbb{R}, \tag{7}$$

em que dizemos que o estatístico perde $L(\theta, a)$ se o parâmetro vale θ e a estimativa dada vale a.

Aula 4: Estimadores de Bayes e EMV

Definição 16 (Estimador de Bayes) *Considere a perda esperada a posteriori:*

$$E_{\theta|x}[L(\theta, a)] = E[L(\theta, a) \mid x] = \int_\Omega L(\theta, a) \xi(\theta \mid x) d\theta \tag{8}$$

Dizemos que δ^* é um ***estimador de Bayes*** *se, para toda realização $X = x$,*

$$E[L(\theta, \delta^*(x)) \mid x] = \min_{a \in A} E[L(\theta, a) \mid x]. \tag{9}$$

Em outras palavras, um estimador de Bayes é uma função real dos dados que minimiza a perda esperada com respeito à posteriori dos parâmetros.

[1] p.d.f. - “probability density function” ou função de densidade de probabilidade

Teorema 3 (δ^* sob perda quadrática) *Seja θ um parâmetro tomando valores reais. Sob perda quadrática,*

$$\delta^*(x) = E[\theta \mid X = x] = \int_\Omega \theta \xi(\theta \mid x) d\theta \tag{10}$$

Teorema 4 (δ^* sob perda absoluta) *Suponha que a função de perda é dada por*

$$L(\theta, \delta^*) = |\theta - \delta^*|. \tag{11}$$

*Dizemos que a função de perda é **absoluta**. Seja θ um parâmetro tomando valores na reta. Sob perda absoluta, $\delta^*(x)$ é a **mediana** a posteriori, isto é,*

$$\int_\infty^{\delta^*(x)} \xi(\theta \mid x) d\theta = \frac{1}{2} \tag{12}$$

Definição 17 (Estimador consistente) *Seja $\delta_1, \delta_2, \ldots, \delta_n$ uma sequência de estimadores de θ. Se quando $n \to \infty$ a sequência convergente para θ, dizemos que esta é uma **sequência consistente de estimadores**.*

Definição 18 (Estimador de máxima verossimilhança) *Para cada possível vetor (de observações) x, seja $\delta(x) \in \Omega$ um valor de $\theta \in \Omega$ de modo que a função de verossimilhança, $L(\theta) \propto f(x \mid \theta)$ [2], atinge o máximo. Dizemos que $\hat{\theta} = \delta(\boldsymbol{X})$ é o **estimador de máximo verossimilhança** de θ (Fisher, 1922)[3]. Quando observamos $\boldsymbol{X} = x$, dizemos que $\delta(x)$ é uma estimativa de θ. Dito de outra forma:*

$$\max_{\theta \in \Omega} f(\boldsymbol{X} \mid \theta) = f(\boldsymbol{X} \mid \hat{\theta}). \tag{13}$$

Famílias Conjugadas

Se $X_1, \ldots, X_n$ são iid e seguem a distribuição da coluna "Dados" na tabela 1.
Notações: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i; \quad y = \sum_{i=1}^n x_i$

Dados	Priori	Posteriori
Bernoulli(θ)	Beta(α, β)	Beta($\alpha + y, \beta + n - y$)
Poisson(θ)	Gama(α, β)	Gama($\alpha + y, \beta + n$)
Normal(μ, σ^2)	Normal(μ_0, v_0^2)	Normal $\left(\frac{\sigma^2\mu_0 + nv_0^2\bar{x}_n}{\sigma^2 + nv_0^2}, \frac{\sigma^2 v_0^2}{\sigma^2 + nv_0^2}\right)$
Exp(θ)	Gama(α, β)	Gama($\alpha + n, \beta + y$)

Table 1: Famílias Conjugadas

Aula 5: EMV

Teorema 5 (Invariância do EMV) *Considere uma função $\phi : \Omega \to \mathbb{R}$. Se $\hat{\theta}$ é um EMV para θ, então $\phi(\hat{\theta})$ é um EMV para $\omega = \phi(\theta)$.*

Teorema 6 (Consistência do EMV) *Defina $l(\theta) := \log f_n(x \mid \theta)$ e assuma que $X_1, \ldots, X_n \sim f(\theta_0)$, isto é, que θ_0 é o valor verdadeiro do parâmetro. Denote $E_{\theta_0}[g] := \int_{\mathcal{X}} g(x, \theta_0) f(x \mid \theta_0)\, dx$. Suponha que*

- *$f(x_i \mid \theta)$ tem o mesmo suporte;*
- *θ_0 é o ponto inferior de Ω;*

[2] $\propto$ - é um operador matemático binário que indica que o valor esquerdo é proporcional ao valor direito.
[3] Ronald Aylmer Fisher (1890-1962), biólogo e estatístico inglês.

- *$I(\theta)$ é diferenciável;*

- *$\hat{\theta}_{EMV}$ é única solução de $I'(\theta) = 0$.*

Então

$$\hat{\theta}_{EMV} \to \theta$$

Aula 6: Método dos momentos e suficiência

Definição 19 (Método dos momentos) *Suponha que $X_1, \ldots, X_n$ formam uma sequânica aleatória com distribuição conjunta $f_n(X_1, \ldots, X_n \mid \theta), \theta \in \Omega \subseteq \mathbb{R}^k$ e que o k-ésimo momento existe. Defina $\mu_j(\theta) = E[X_1^j \mid \theta]$ e suponha que $\mu : \Omega \to \mathbb{R}^k$ é biunívoca, de modo que sua inversa é*

$$\theta = M(\mu_1(\theta), \ldots, \mu_k(\theta)).$$

*Dados os momentos amostrais $m_j := \frac{1}{n} \sum_{i=1}^{n} X_i^j, j = 1, \ldots, k$ o **estimador de momentos** (EMM) de θ é*

$$\hat{\theta}_{EMM} = M(m_1, \ldots, m_k).$$

Teorema 7 (Consistência do EMM) *Suponha que $X_1, \ldots, X_n$ formam uma amostra aleatória com distribuição conjunta $f_n(X_1, \ldots, X_n \mid \theta), \theta \in \Omega \subseteq \mathbb{R}^k$ e que o k-ésimo momento existe. Suponha que a inversa M existe e é continua. Então o EMM é **consistente** para θ.*

Definição 20 (Estatística suficiente) *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição indexada pelo parâmetro θ. Seja $T = r(X_1, \ldots, X_n)$ uma estatística. Dizemos que T é uma **estatística suficiente** para θ se e somente se*

$$f(X_1, \ldots, X_n \mid T = t, \theta) = f(X_1, \ldots, X_n \mid T = t, \theta'), \forall \theta, \theta' \in \Omega, \tag{14}$$

isto é, se a distribuição condicional da amostra dado o valor da estatística não depende de θ.

Definição 21 (Aleatorização auxiliar) *Suponha que T é suficiente para θ. O processo de simular $X_1', \ldots, X_n'$ dado que $T = r(X_1, \ldots, X_n)$ de modo que*

$$f(X_1, \ldots, X_n \mid \theta) = f(X_1', \ldots, X_n' \mid \theta), \forall \theta \in \Omega, \tag{15}$$

*é chamado de **aleatorização auxiliar** (em inglês, auxiliary randomisation).*

Teorema 8 (Teorema de fatorização) *Suponha que $X_1, \ldots, X_n$ perfazem uma amostra aleatória com f.d.p./f.m.p. $f(x \mid \theta), \theta \in \Omega$. Uma estatística $T = r(X_1, \ldots, X_n)$ é suficiente para θ se, e somente se, para todo $x \in \mathcal{X}$ e $\theta \in \Omega$ existem u e v não negativos tal que*

$$f_n(x \mid \theta) = u(x)v[r(x), \theta]. \tag{16}$$

Definição 22 (Suficiência conjunta) *Dizemos que um conjunto de estatísticas $T = \{T_1, \ldots, T_n\}$ é **suficiente** (conjuntamente) se que a distribuição condicional conjunta de $X_1, \ldots, X_n$ dado $T_1 = t_1, \ldots, T_n = t_n$ não dependentes de θ.*

Aula 7: Suficiência conjunta e mínima, teorema de Rao-Blackwell

Definição 23 (Estatísticas de ordem) *Seja $\boldsymbol{X} = X_1, \ldots, X_n$ uma amostra aleatória. Dizemos que $Y_1, \ldots, Y_n$ são **estatísticas de ordem** se Y_1 é o menor valor de $\boldsymbol{X}$, Y_2 é o segundo menor valor e assim sucessivamente.*

Teorema 9 (Estatísticas de ordem são suficientes conjuntas) *Seja $X_1, \ldots, X_n$ uma amostra aleatória com f.d.p./f.m.p. $f(x \mid \theta)$. As estatísticas de ordem $Y_1, \ldots, Y_n$ são suficientes conjuntas para θ.*

Definição 24 (Suficiência mínima) *Uma estatística $\boldsymbol{T}$ é dita **mínima suficiente** se $\boldsymbol{T}$ é suficiente e é função de qualquer outra estatística suficiente. Um vetor $\boldsymbol{T} = \{T_1, \ldots, T_n\}$ é dito **minimamente suficiente conjunto** se é função de qualquer outro valor de estatísticas suficientes conjuntas.*

Teorema 10 (EMV e Bayes são suficientes) *Se a função de verossimilhança admite fatorização pelo Teorema 8, os estimadores de Bayes e de máxima verossimilhança são estatísticas minimamente suficientes.*

Definição 25 (Notação conveniente) *É conveniente definir que para $g : \mathcal{X}^n \to \mathbb{R}$, escrevemos*

$$E_\theta[g] = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} g(\boldsymbol{x}) f_n(\boldsymbol{x} \mid \theta) dx_1 \cdots dx_n = \int_{\mathcal{X}} g(\boldsymbol{x}) f_n(\boldsymbol{x} \mid \theta) d\boldsymbol{x} \tag{17}$$

Definição 26 (Erro quadrático médio)

$$R(\theta, \delta) := E_\theta \left[\{\delta(\boldsymbol{X}) - \theta\}^2 \right]. \tag{18}$$

Definição 27 (Estimador condicionado)

$$\delta_0(\boldsymbol{T}) := E_\theta \left[\delta(\boldsymbol{X}) \mid \boldsymbol{T} \right]. \tag{19}$$

Teorema 11 (Teorema de Rao-Blackwell) *Seja $\delta(\boldsymbol{X})$ um estimador, $\boldsymbol{T}$ uma estatística suficiente para θ e seja $\delta_0(\boldsymbol{T})$ como na Definição 27. Então vale que*

$$R(\theta, \delta_0) \leq R(\theta, \delta)$$

Além disso, se $R(\theta, \delta) < \infty$ e $\delta(\boldsymbol{X})$ não é função de $\boldsymbol{T}$, vale a desigualdade estrita:

$$R(\theta, \delta_0) < R(\theta, \delta)$$

Aula 8: Admissibilidade e viés

Definição 28 (Admissibilidade) *Um estimador δ é dito **inadmissível** se existe outro estimador δ_0 tal que $R(\theta, \delta_0) \leq R(\theta, \delta), \forall \theta \in \Omega$ e existe $\theta' \in \Omega$ tal que $R(\theta', \delta_0) < R(\theta', \delta)$. Nesse caso, dizemos que δ_0 domina δ. O estimador δ_0 é **admissível** se (e somente se) não há nenhum estimador que o domine.*

Definição 29 (Estimador não-viesado) *Um estimador $\delta(\boldsymbol{X})$ de uma função $g(\theta)$ é dito **não-viesado** se $E_\theta[\delta(\boldsymbol{X})] = g(\theta)$, $\forall \theta \in \Omega$. Um estimador que não atende a essa condição é dito viesado. E o **víes** de δ é definido como $B_\delta(\theta) := E_\theta[\delta(\boldsymbol{X})] - g(\theta)$.*

Teorema 12 (Estimador não-viesado da variância) *Seja $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ uma amostra aleatória, com $E[X_1] = m$ e $Var(X_1) = v < \infty$. Então*

$$\delta_1(\boldsymbol{X}) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

é um estimador não-viesado de v.

Aula 9: Eficiência

Definição 30 (Informação de Fisher) *Seja X uma variável aleatória com f.d.p./f.m.p. $f(x \mid \theta)$, $\theta \in \Omega \subseteq \mathbb{R}$. Suponha que $f(x \mid \theta)$ é duas vezes diferenciável com respeito a θ. Defina $\lambda(x \mid \theta) = \log f(x \mid \theta)$ e*

$$\lambda'(x \mid \theta) = \frac{\partial \lambda(x \mid \theta)}{\partial \theta} \quad \text{e} \quad \lambda''(x \mid \theta) = \frac{\partial^2 \lambda(x \mid \theta)}{\partial \theta^2} \tag{20}$$

*Definimos a **informação de Fisher** como*

$$I(\theta) = E_\theta\left[\{\lambda'(x \mid \theta)\}^2\right] \overset{(1)}{=} -E_\theta\left[\lambda''(x \mid \theta)\right] = Var_\theta\left(\lambda'(x \mid \theta)\right). \tag{21}$$

Teorema 13 (Informação de Fisher em uma amostra aleatória) *Seja $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ uma amostra aleatória e seja $I_n = E_\theta[-\lambda_n''(\boldsymbol{X} \mid \theta)]$ a **informação de Fisher** da amostra. Então*

$$I_n(\theta) = nI(\theta)$$

Teorema 14 (Teorema de Cramér-Rao) *Seja $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ uma amostra aleatória, onde f.d.p./f.m.p. tem as mesmas premissas da Definição 30. Supondo que $T = r(\boldsymbol{X})$ é uma estatística com variância finita. Seja $m(\theta) = E_\theta(T)$ uma função diferenciável de θ. Então,*

$$Var_\theta(T) \geq \frac{[m'(\theta)]^2}{nI(\theta)}, \tag{22}$$

com igualdade apenas se existem u e v tal que

$$T = u(\theta)\lambda_n'(\boldsymbol{X} \mid \theta) + v(\theta).$$

Definição 31 (Estimador eficiente) *Um estimador $\delta(\boldsymbol{X})$ é dito **eficiente** de (sua esperança) $m(\theta)$ se*

$$Var_\theta(\delta) = \frac{[m'(\theta)]^2}{nI(\theta)}.$$

Aula 10: Distribuição de uma estatística amostral e qui-quadrado

Definição 32 (Distribuição qui-quadrado) *Dizemos que uma variável aleatória Y tem distribuição **qui-quadrado** com m graus de liberdade quando*

$$f_Y(y) = \frac{1}{2^{m/2}\Gamma(m/2)} y^{m/2-1} e^{-y/2}, y > 0 \tag{23}$$

Vemos que Y tem função geradora de momentos:

$$\psi(t) = \left(\frac{1}{1-2t}\right)^{m/2}, t < 1/2.$$

Teorema 15 (Soma de variáveis aleatórias qui-quadrado) *Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes com graus de liberdade m_i, então $W = \sum_{i=1}^n X_i$ tem distribuição qui-quadrado com graus de liberdade $m = \sum_{i=1}^n m_i$.*

Teorema 16 (Distribuição do quadrado de uma variável aleatória Normal padrão) *Se*

$$X \sim Normal(0,1), Y = X^2$$

então, tem distribuição qui-quadrado com $m = 1$.

Aula 11: Distribuição da média e variância amostrais

Teorema 17 (Independência da média e variância amostrais na Normal) *Seja $X_1, \dots, X_n$ uma amostra aleatória de uma distribuição Normal com parâmetros μ e σ^2, $\overline{X}_n$ e a variância amostral $\overline{S}_n^2$, são independentes. Ademais, $\overline{X}_n \sim Normal\left(\mu, \sigma^2\right)$ e $\overline{S}_n^2 \sim Gama\left(\frac{n-1}{2}, \frac{n}{2n^2}\right)$*

Aula 12: Distribuição t de Student e intervalos de confiança

Definição 33 (A distribuição t de Student) *Tome, $Y \sim \text{Qui} - \text{quadrado}(m)$ e $Z \sim \text{Normal}(0,1)$ e defina a variável aleatória*

$$X = \frac{Z}{\sqrt{\frac{Y}{m}}}.$$

*Dizemos que X tem distribuição **t de Student com m graus de liberdade**. E sabemos que*

$$f_X = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{m\pi}\Gamma(\frac{m}{2})}\left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad x \in (-\infty, +\infty).$$

Teorema 18 (Distribuição amostral do estimador não-viesado da variância) *Considere o estimador*

$$\hat{\sigma}' = \sqrt{\frac{\Delta^2}{n-1}},$$

onde $\Delta^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2$. Então, vale que

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\hat{\sigma}'} \sim \text{Student}(n-1)$$

Teorema 19 (Intervalo de confiança) *Seja $\boldsymbol{X} = \{X_1, \dots, X_n\}$ uma amostra aleatória, onde cada uma tem p.d.f. $f(x \mid \theta)$, e considere uma função real $g(\theta)$. Sejam $A(\boldsymbol{X})$ e $B(\boldsymbol{X})$ duas estatísticas de modo de valha*

$$P(A(\boldsymbol{X}) < g(\theta) < B(\boldsymbol{X})) \geq \gamma. \tag{24}$$

*Dizemos que $I(\boldsymbol{X}) = (A(\boldsymbol{X}, B(\boldsymbol{X}))$ é um **intervalo de confiança** de $100\gamma\%$ para $g(\theta)$. Se a desigualdade for uma igualdade para todo $\theta \in \Omega$, dizemos que o intervalo é **exato**.*

Aula 13: Intervalos de confiança e Quantidades Pivotais

Definição 34 (Intervalo de confiança unilateral) *Seja $\boldsymbol{X} = \{X_1. \dots, X_n\}$ uma amostra aleatória, onde cada uma tem p.d.f. $f(x \mid \theta)$, e considere uma função real $g(\theta)$. Seja $A(\boldsymbol{X})$ uma estatística que*

$$P(A(\boldsymbol{X}) < g(\theta)) \geq \gamma, \quad \forall \theta \in \Omega$$

*dizemos que o intervalo aleatório $(A(\boldsymbol{X}), \infty)$ é chamado de intervalo de confiança **unilateral** de $100\gamma\%$ para $g(\theta)$ (ou ainda, de intervalo de confiança **inferior** de $100\gamma\%$ para $g(\theta)$). O intervalo $(-\infty, B(\boldsymbol{X}))$, com*

$$P(g(\theta) < B(\boldsymbol{X})) \geq \gamma, \quad \forall \theta \in \Omega$$

*é definido de forma análoga, e é chamado de intervalo de confiança **superior** de $100\gamma\%$ para $g(\theta)$. Se a desigualdade é uma igualdade para todo $\theta \in \Omega$, os intervalos são chamados **exatos**.*

Definição 35 (Quantidade pivotal) *Seja* $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ *uma amostra aleatória com p.d.f.* $f(x \mid \theta)$. *Seja* $V(\boldsymbol{X}, \theta)$ *uma variável aleatória cuja distribuição é a mesma para todo* $\theta \in \Omega$. *Dizemos que* $V(\boldsymbol{X}, \theta)$ *é uma* ***quantidade pivotal****.*

Teorema 20 (Intervalo de confiança unilateral) *Seja* $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ *uma amostra aleatória com p.d.f.* $f(x \mid \theta)$. *Suponha que existe uma quantidade pivotal V, com c.d.f.* [4] *continua G. Assuma que existe* $r(v, \boldsymbol{x})$ *estritamente crescente em* v *para todo* $\boldsymbol{x}$. *Finalmente, tome* $0 < \gamma < 1$ *e* $\gamma_1 < \gamma_2$ *de modo que* $\gamma_2 - \gamma_1 = \gamma$. *Então as estatísticas*

$$A(\boldsymbol{X}) = r(G^{-1}(\gamma_1), \boldsymbol{X}),$$

$$B(\boldsymbol{X}) = r(G^{-1}(\gamma_2), \boldsymbol{X}),$$

são os limites de um intervalo de confiança de $100\gamma\%$ *para* $g(\theta)$.

Aula 14: Testes de hipótese I

Definição 36 (Hipótese nula e hipótese alternativa) *Considere o espaço de parâmetros* Ω *e defina* $\Omega_0, \Omega_1 \subset \Omega$ *de modo que* $\Omega_0 \cup \Omega_1 = \Omega$ *e* $\Omega_0 \cap \Omega_1 = \emptyset$. *Definimos*

$$H_0 := \theta \in \Omega_0,$$
$$H_1 := \theta \in \Omega_1,$$

E dizemos que H_0 *é a* ***hipótese nula*** *e* H_1 *é a* ***hipótese alternativa****. Se* $\theta \in \Omega_1$, *então dizemos que rejeitamos a hipótese nula. Por outro lado, se* $\theta \in \Omega_0$, *então dizemos que não rejeitamos ou falhamos em rejeitar* H_0.

Definição 37 (Hipótese simples e hipótese composta) *Dizemos que uma hipótese* H_i, *é* ***simples****, se* $\Omega_i = \{\theta_i\}$, *isto é, se a partição correspondente é um único ponto. Uma hipótese é dita* ***composta*** *se não é simples.*

Definição 38 (Hipótese unilateral e hipótese bilateral) *Uma hipótese da forma* $H_0 : \theta \leq \theta_0$ *ou* $H_0 : \theta \geq \theta_0$ *é dita* ***unilateral*** *("one-sided"), enquanto hipóteses da forma* $H_0 : \theta \neq \theta_0$ *são ditas* ***bilaterais*** *("two-sided").*

Aula 15: Testes de hipótese II

Definição 39 (Região crítica) *O conjunto*

$$S_1 := \{\boldsymbol{x} : |\overline{X}_n - \mu_0| \geq c\}$$

é chamado de ***região crítica*** *do teste.*

Definição 40 (Região de rejeição) *Se* $R \subseteq \mathbb{R}$ *é tal que "rejeitamos* H_0 *se* $T \in R$*", então* R *é chamada uma* ***região de rejeição*** *para a estatística* T *e o teste associado.*

Definição 41 (Função poder) *Seja* δ *um procedimento de aceitação/rejeição como visto anteriormente. A* ***função poder*** *é definida como*

$$\pi(\theta \mid \delta) := P(\boldsymbol{X} \in S_1 \mid \theta) = P(T \in R \mid \theta), \ \theta \in \Omega \tag{25}$$

[4] c.d.f. - cumulative distribution function

Definição 42 (Tipos de erros) *Tipos de erros que podem ser cometidos*

Nome	*Erro cometido*
Erro tipo I	*Rejeitar H_0 quando ela é **verdadeira**.*
Erro tipo II	*Falhar em rejeitar H_0 quando ela é **falsa**.*

Definição 43 (Tamanho/nível de um teste) *Dizemos que um teste, δ, tem **tamanho** ou **nível de significância** $\alpha(\delta)$, com*

$$\alpha(\delta) := \sup_{\theta \in \Omega_0} \pi(\theta \mid \delta).$$

Aula 16: Testes de hipótese III

Definição 44 (O p-valor) *Para cada t, seja δ_t o teste que rejeita H_0 se $T \geq t$. Então, quando $T = t$, o **p-valor** vale*

$$p(t) := \sup_{\theta \in \Omega_0} \pi(\theta \mid \delta_t) = \sup_{\theta \in \Omega_0} P(T \geq t \mid \theta) \tag{26}$$

*ou seja, o **p-valor** é o tamanho do teste δ_t.*

Aula 17: Testes e conjuntos de confiança

Definição 45 (Intervalos de confiança e testes são equivalentes) *Suponha que dispomos de dados $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ com f.d.p. comum $f(x \mid \theta)$, e estamos interessados em testar as hipóteses:*

$$H_0 : g(\theta) = g_0,$$
$$H_1 : g(\theta) \neq g_0,$$

de modo que existe um teste δ_{g_0} com nível α_0 destas hipóteses. Para cada $\boldsymbol{X} = \boldsymbol{x}$, defina

$$w(\boldsymbol{x}) = \{g_0 : \delta_{g_0} \, não \, rejeita \, H_0 \, dado \, que \, \boldsymbol{X} = \boldsymbol{x}\}.$$

Fazendo o nível de confiança do intervalo $\gamma = 1 - \alpha_0$, temos

$$P(g(\theta_0) \in w(\boldsymbol{X}) \mid \theta = \theta_0) \geq \gamma, \ \forall \theta_0 \in \Omega.$$

Definição 46 (Conjunto de confiança) *Se um conjunto aleatório $w(\boldsymbol{X})$ satisfaz*

$$P(g(\theta_0) \in w(\boldsymbol{X}) \mid \theta = \theta_0) \geq \gamma,$$

*para todo $\theta_0 \in \Omega$, então chamamos $w(\boldsymbol{X})$ de um **conjunto de confiança** para $g(\theta)$.*

Teorema 21 (Testando hipóteses a partir de conjuntos de confiança) *Suponha que dispomos de dados $\boldsymbol{X} = \{X_1, \ldots X_n\}$ com f.d.p. comum $f(x \mid \theta)$ e que $w(\boldsymbol{X})$ é um conjunto de confiança para uma função de interesse $g(\theta)$. Então para todo valor g_0 assumido por $g(\theta)$ existe um teste δ_{g_0}, de nível α_0 que rejeita $H_0 : g(\theta) = g_0$ se e somente se $g(\theta_0) = g_0 \notin w(\boldsymbol{X})$.*

Teorema 22 (Teste de razão de verossimilhanças) *A estatística*

$$\wedge(\boldsymbol{x}) = \frac{\sup_{\theta \in \Omega_0} f_n(\boldsymbol{x}|\theta)}{\sup_{\theta \in \Omega} f_n(\boldsymbol{x}|\theta)}$$

*é chamada um **estatística de razão de verossimilhanças**. Um **teste de razão de verossimilhanças**, δ_k, é um teste que rejeita H_0 se $\wedge(\boldsymbol{x}) \leq k$ para uma constante k.*

Teorema 23 (Teorema de Wilks) *Suponha que temos um espaço de parâmetros com k coordenadas,* $\theta = (\theta_1, \ldots, \theta_n)$ *e desejamos testar a hipótese (simples) da forma*

$$H_0 : \theta_j = \theta_0^j,\ j = 1, \ldots, k,$$
$$H_1 : \theta_j \neq \theta_0^j,\ j = 1, \ldots, k.$$

Então, sob condições de regularidade, temos que, à medida que $n \to \infty$,

$$-2 \log \wedge(\boldsymbol{x}) \xrightarrow{\mathrm{d}} X^2(k)$$

Aula 18: Teste t I

Definição 47 (Teste não viesado) *Suponha que desejamos testar a hipótese*

$$H_0 : \theta \in \Omega_0,$$
$$H_1 : \theta \in \Omega_1,$$

através do teste δ*. Dizemos que* δ *é* ***não-viesado*** *se (e somente se) para* $\theta \in \Omega_0$ *e* $\theta' \in \Omega_1$*, vale*

$$\pi(\theta \mid \delta) \leq \pi(\theta' \mid \delta),$$

ou seja, se a função poder é pelo menos tão grande no espaço onde H_0 *é falsa* (Ω_1) *quando no espaço em que* H_0 *é verdadeira* (Ω_0)*.*

Definição 48 (Teste t) *Um teste* δ_c *que rejeita* H_0 *se* $U \geq c$ *(equiv.* $U \leq c$*), com* $c = T^{-1}(1 - \alpha_0; n - 1)$ *é chamado de um* ***teste t*** *(unicaudal) de tamanho* α_0*.*

Teorema 24 (Propriedades do teste t) *Suponha que* δ_c *rejeita* H_0 *se* $U \geq c$*. Então*

- $\mu = \mu_0 \Longrightarrow \pi(\mu, \sigma^2 \mid \delta_c) = \alpha_0$
- $\mu < \mu_0 \Longrightarrow \pi(\mu, \sigma^2 \mid \delta_c) < \alpha_0$
- $\mu > \mu_0 \Longrightarrow \pi(\mu, \sigma^2 \mid \delta_c) > \alpha_0$
- $\lim_{\mu \to -\infty} \pi(\mu, \sigma^2 \mid \delta_c) = 0$
- $\lim_{\mu \to +\infty} \pi(\mu, \sigma^2 \mid \delta_c) = 1$
- δ_c *é não-viesado e tem tamanho* α_0*.*

Teorema 25 (P-valor para um teste t unicaudal) *Suponha que observarmos* $U = u$ *e seja* $T(\cdot . n - 1)$ *a f.d.a. de uma distribuição t de Student com n - 1 graus de liberdade. Para a hipótese*

$$H_0 : \mu \geq \mu_0,$$
$$H_1 : \mu < \mu_0,$$

o p-valor vale $T(u; n - 1)$*, enquanto para a hipótese*

$$H_0 : \mu \leq \mu_0,$$
$$H_1 : \mu > \mu_0,$$

o p-valor vale $1 - T(u; n - 1)$*.*

Aula 19: Teste t II

Teorema 26 (Teste pareado) *Sejam amostras* $\boldsymbol{X}$ *e* $\boldsymbol{Y}$ *(antes e depois), tais que* $X_i \sim \text{Normal}(\mu_1, \sigma^2)$ *e* $Y_i \sim \text{Normal}(\mu_2, \sigma^2)$*, a hipótese*

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Pode ser modelada com a variável $Z_i = X_i - Y_i$ *(*$Z_i \sim \text{Normal}(\mu_Z = \mu_1 - \mu_2, 2\sigma^2)$*), então podemos testar hipóteses sobre* μ_Z *a partir de* $\boldsymbol{Z}$

$$H_0 : \mu_Z \leq 0$$
$$H_1 : \mu_Z > 0$$

Teorema 27 (Teste t para duas amostras) *Considere* $\boldsymbol{X} = \{X_1, \ldots, X_m\}$ *e* $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$*, queremos estudar a diferença das médias. Modelando em distribuição normal* $X_i \sim \text{Normal}(\mu_1, \sigma_1^2),\ i = 1, \ldots, m$ *e* $Y_j \sim \text{Normal}(\mu_2, \sigma_2^2),\ j = 1, \ldots, n$*. Sob a premissa de homogeneidade* $\sigma_1^2 = \sigma_2^2 = \sigma^2$*, podemos testar a hipótese*

$$H_0 : \mu_1 \leq \mu 2$$
$$H_1 : \mu_1 > \mu_2$$

computando a estatística

$$U = \frac{\sqrt{m+n-2}(\overline{X}_m - \overline{Y}_n)}{\sqrt{(\frac{1}{m} + \frac{1}{n})(S_X^2 + S_Y^2)}}$$

onde $\overline{X}_m$ *e* $\overline{X}_m$ *são as médias e* S_X^2 *e* S_X^2 *são a soma das variâncias.*

Teorema 28 (Relaxando a premissa de homogeneidade) *Do teorema acima, podemos relaxar a premissa de igualdade das variâncias assumindo que* $\sigma_2^2 = k\sigma_1^2$*, então a estatística teste vale*

$$U = \frac{\sqrt{m+n-2}(\overline{X}_m - \overline{Y}_n)}{\sqrt{(\frac{1}{m} + \frac{k}{n})(S_X^2 + \frac{S_Y^2}{n})}}$$

Aula 20: Teste f

Definição 49 (A distribuição F) *Sejam* $Y \sim \text{Qui} - \text{quadrado}(m)$ *e* $W \sim \text{Qui} - \text{quadrado}(n)$*. Então*

$$X = \frac{Y/m}{W/n},$$

tem distribuição F com m e n graus de liberdade, com f.d.p.

$$f_X(x) = \frac{\Gamma(\frac{m+n}{2}) m^{m/2} n^{n/2}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \cdot \frac{x^{m/2-1}}{(mx+n)^{(m+n)/2}},\ x > 0,$$

Teorema 29 (Propriedades da distribuição F)

1. *Se* $X \sim F(m,n)$*, então* $\frac{1}{X} \sim F(m,n)$*;*
2. *Se* $Y \sim \text{Student}(n)$*, então* $Y^2 \sim F(1,n)$*.*

Teorema 30 (Igualdade de duas variâncias) *Suponha* $X_i \sim \text{Normal}(\mu_1, \sigma_1^2),\ i = 1, \ldots, m$ *e* $Y_j \sim \text{Normal}(\mu_2, \sigma_2^2),\ j = 1, \ldots, n$. *Queremos testar*

$$H_0 : \sigma_1^2 \leq \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$

Para isso, vamos computar a estatística de teste

$$V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$$

onde $S_X^2 = \sum_{i=1}^m (X_i - \overline{X}_m)^2$ *e* $S_Y^2 = \sum_{j=1}^m (Y_j - \overline{Y}_m)^2$

Definição 50 (O teste F) *O teste F de homogeneidade (igualdade de variâncias) é o teste* δ_c *que rejeita* H_0 *de* $V \geq c$, *para uma constante positiva c.*

Teorema 31 (A distribuição de V) *Seja* $V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)}$, *então:*

$$\frac{\sigma_2^2}{\sigma_1^2} V \sim F(m-1, n-1).$$

Além disso, se $\sigma_1^2 = \sigma_2^2$, $V \sim F(m-1, n-1)$.

Aula 21: Regressão Linear I

Teorema 32 (A linha de mínimos quadrados) *Sejam* $(x_1, y_1), \ldots, (x_n, y_n)$ *uma coleção de n pontos. Suponha que estamos interessados na reta*

$$y_i = \beta_0 + \beta_1 x_i. \tag{27}$$

Os valores dos coeficientes que minimizam a soma de quadrados são

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x},$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}.$$

Definição 51 (Modelo linear) *Podemos construir um modelo estatístico explícito para a relação entre as variáveis X e Y:*

$$E[Y \mid X = x_1, \ldots, x_P] = \beta_0 + \beta_1 x_1 + \ldots + \beta_P x_P \tag{28}$$

Podemos então idealizar o seguinte modelo

$$Y_i = \beta_0 \sum_{j=1}^P \beta_j x_{ij} + \varepsilon_i,\ \varepsilon \sim \text{Normal}(0, \sigma^2).$$

Aula 22: Regressão Linear II

Teorema 33 (EMV para os coeficientes de uma regressão linear (simples)) *Sob as premissas já listadas, os estimadores de máxima verossimilhança para $\theta = (\beta_0, \beta_1, \sigma^2)$ são*

$$\hat{\beta}_{0EMV} = \overline{y} - \hat{\beta}_{1EMV}\overline{x},$$

$$\hat{\beta}_{1EMV} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

$$\hat{\sigma^2}_{EMV} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (\hat{\beta}_{0EMV} + \hat{\beta}_{1EMV}x_i)\right)^2,$$

ou seja, os estimadores de máxima verossimilhança dos coeficientes minimizam a soma de quadrados da reta estimada.

Teorema 34 (Distribuição amostral dos estimadores dos coeficientes)

$$\hat{\beta}_{0EMV} \sim \text{Normal}\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{s_X^2}\right)\right),$$

$$\hat{\beta}_{1EMV} \sim \text{Normal}\left(\beta_0, \frac{\sigma^2}{s_X^2}\right),$$

$$\text{Cov}(\hat{\beta}_{0EMV}, \hat{\beta}_{1EMV}) = -\frac{\overline{x}\sigma^2}{s_X^2},$$

onde $s_x = \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}$.

Teorema 35 (Intervalos de confiança para os coeficientes de uma regressão linear) *Podemos computar intervalos de confiança para os coeficientes da regressão linear de maneira muito similar ao que já vimos para o caso da média da Normal*

$$\hat{\beta}_0 \pm \hat{\sigma}' c\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}} \quad \text{e} \quad \hat{\beta}_1 \pm c\frac{\hat{\sigma}'}{s_x},$$

$$\hat{\beta}_0 + \hat{\beta}_1 x_{pred} \pm c\hat{\sigma}'\sqrt{\frac{1}{n} + \frac{(x_{pred} - \bar{x})^2}{s_x^2}}$$

onde $c = T^{-1}(1 - \frac{\alpha_0}{2}; n-2)$ *e*

$$\hat{\sigma}' := \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{n-2}}.$$

Definição 52 (Testes de hipóteses para o coeficiente angular) *Em geral, estamos interessados em testar a hipótese*

$$H_0 : \beta_1 = \beta^\star,$$
$$H_1 : \beta_1 \neq \beta^\star.$$

Para tanto, podemos computar a estatística

$$U_1 = s_x\frac{\hat{\beta}_1 - \beta^\star}{\hat{\sigma}'}, \tag{29}$$

e computar o p-valor como

$$P(U_1 \geq |u_1|) + P(U_1 \leq -|u_1|). \tag{30}$$

Notando que U_1 tem distribuição t de Student com $n-2$ graus de liberdade sob H_0, podemos computar o p-valor exatamente.
Resultados bem similares valem para testar hipóteses sobre β_0 ou $\hat{Y}$.

Teorema 36 (Predição pontual) *Suponha que queremos prever o valor de Y para um certo x_{pred} que não foi observado no experimento. Podemos compor nossa predição (pontual) como*

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_{pred}. \tag{31}$$

Onde a predição tem erro quadrático médio (EQM) igual a

$$E\left[(\hat{Y} - Y)^2\right] = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{pred} - \overline{x})^2}{s_X^2}\right).$$

Teorema 37 (Intervalos de predição para $\hat{Y}$) *A probabilidade de $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_{pred}$ estar no intervalo*

$$\hat{Y} \pm T^{-1}(1 - \frac{\alpha_0}{2}; n-2)\hat{\sigma}' \sqrt{\left[1 + \frac{1}{n} + \frac{(x_{pred} - \overline{x})^2}{s_X^2}\right]},$$

é $1 - \alpha_0$.