# LISTA 0

1. a) Para $i = 1, \dots, n$, temos que

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i-\mu)^2\right]$$

Por ser uma amostra aleatória, as variáveis aleatórias são independentes. Logo, a densidade conjunta da amostra é dada por

$$f_X(x) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i-\mu)^2\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2-2x_i\mu+\mu^2)\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{\sum_{i=1}^{n}x_i^2}{2\sigma^2} + \frac{2\mu\sum_{i=1}^{n}x_i}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right]$$

$$= \frac{\exp[-n\mu^2/2\sigma^2]}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{\sum_{i=1}^{n}x_i^2}{2\sigma^2} + \frac{\mu\sum_{i=1}^{n}x_i}{\sigma^2}\right]$$

Se definirmos $a(\theta) = \dfrac{\exp[-n\mu^2/2\sigma^2]}{(2\pi\sigma^2)^{n/2}}$,

$b(x) = 1$, $c_1(\theta) = -\dfrac{1}{2\sigma^2}$, $c_2(\theta) = \dfrac{\mu}{\sigma^2}$,

$d_1(x) = \sum_{i=1}^{n} x_i^2$ e $d_2(x) = \sum_{i=1}^{n} x_i$, então

$$f_X(x) = a(\theta)\, b(x) \exp\left[\sum_{i=1}^{2} c_i(\theta)\, d_i(x)\right]$$

b) Vamos usar o Teorema da Fatoração para resolver esse problema

Teorema da Fatoração: Seja $f_X(x|\theta)$ a função densidade de probabilidade ou função de probabilidade conjunta de uma amostra $X$. Uma estatística $T(X)$ é suficiente para $\theta$ se, e somente se, existem funções $g(t|\theta)$ e $h(x)$ tais que, para qualquer ponto amostral $x$ e $\theta$ no espaço paramétrico, vale a igualdade

$$f_X(x|\theta) = g(T(x)|\theta)\, h(x)$$

Voltando ao resultado do item a), temos a forma

$$f_X(x|\theta) = \frac{\exp[-n\mu^2/2\sigma^2]}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i\right]$$

Se fizermos $c(\theta) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right)'$ e

$T(X) = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right)$, então

$$f_X(x|\theta) = g(T(x)|\theta)\, h(x),$$

onde $g(T(X)|\theta) = \frac{\exp[-n\mu^2/2\sigma^2]}{(2\pi\sigma^2)^{n/2}} \exp[T(x)c(\theta)]$

e $h(x) = 1$.

Logo, $T(X) = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right)$ é uma estatística

suficiente para $\theta = (\mu, \sigma^2)$

c) Seja $L(\theta|x) = \prod_{i=1}^{n} f_{X_i}(x_i)$

$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right]$ a função de

verossimilhança. O objetivo é maximizar $L(\theta|x)$
com relação a $\theta = (\mu, \sigma^2)$.

Isso é equivalente a encontrar os pontos $\theta = (\mu, \sigma^2)$
que maximizam $l(\theta|x) = \log(L(\theta|x))$.

$$l(\theta|x) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{n}{2}\log(\sigma^2)$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

- Encontrando o ponto em relação a $\mu$

Derivamos $l(\theta|x)$ em relação a $\mu$ e igualamos a 0 para encontrar o ponto $\mu$

$$\frac{\partial l(\theta|x)}{\partial \mu} = \frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\Rightarrow \sum_{i=1}^{n}x_i - n\mu = 0$$

$$\Rightarrow \mu = \frac{\sum_{i=2}^{n}x_i}{n} = \bar{x}$$

e o ponto é dado por $\hat{\mu} = \bar{x}$

- Encontrando o ponto em relação a $\sigma^2$

Derivamos $l(\theta|x)$ em relação a $\sigma^2$ e igualamos a 0 para encontrar o ponto $\sigma^2$

$$\frac{\partial l(\theta|x)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow n\sigma^2 = \sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

Substituindo o ponto que encontramos para $\mu$, temos que $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Portanto, os estimadores de máxima verossimilhança de $\theta = (\mu, \sigma^2)$ são $\hat{\theta} = (\bar{x}, \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2)$

d) Consideremos a estatística

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

onde $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$

Inicialmente, dividamos o numerador e denominador pelo desvio padrão $\sigma$ da população, e teremos

$$T = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{S/\sigma}$$

O numerador $Z = \sqrt{n}(\bar{X} - \mu)/\sigma$ tem distribuição $N(0,1)$. O quadrado do denominador pode ser escrito como

$$\frac{S^2}{\sigma^2} = \frac{n-1}{n-1} \cdot \frac{S^2}{\sigma^2} = \frac{Y}{n-1},$$

onde $Y = (n-1)S^2/\sigma^2$. Como os $X_i$ são normais, $Y$ tem distribuição $\chi^2(n-1)$. Logo, $T$ é o quociente entre uma v.a. $N(0,1)$ e a raiz quadrada de uma v.a. $\chi^2(n-1)$, dividada pelo número de graus de liberdade.

Se $Z$ é uma v.a. $N(0,1)$ e $Y$ uma v.a. $\chi^2(v)$, com $Z$ e $Y$ independentes, então

$$t = \frac{Z}{\sqrt{Y/v}} \sim t(v)$$

Usando esse resultado, temos que

$$T = \frac{(\bar{X}-\mu)}{S/\sqrt{n}} = \frac{Z}{\sqrt{Y/(n-1)}} \sim t(n-1)$$

Observe que $Z$ e $Y$ são independentes, pois $\bar{X}$ e $S^2$ são independentes.

Sob $H_0$, $\mu_0$ é tido como o verdadeiro valor de $\mu$, e a estatística ~~fica~~ é dada por

$$T = \frac{(\bar{X}-\mu_0)}{S/\sqrt{n}}$$, que sabemos agora ter uma distribuição $t$ de Student com $(n-1)$ graus de liberdade.

Sob $H_1$, considerando $\mu \neq \mu_0$, a estatística

$$T = \frac{(\bar{X}-\mu)}{S/\sqrt{n}}$$ também segue uma distribuição $t$ de Student com $(n-1)$ graus de liberdade.

Estamos, agora, em condições de testar as hipóteses

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

A estatística a ser usada, como visto anteriormente, é

$$T = \frac{\sqrt{n}\,(\bar{X} - \mu_0)}{S}$$

Fixado o nível de significância $\alpha$, podemos encontrar o valor $t_c$ tal que $P(|T| < t_c) = 1 - \alpha$

Colhida a amostra de $n$ indivíduos, calculamos os valores $\bar{x}_0$ e $s_0^2$ das estatísticas $\bar{X}$ e $S^2$, respectivamente, e depois o valor $t_0 = \sqrt{n}\,(\bar{x}_0 - \mu_0)/s_0$ de $T$. Se o valor dessa estatística for inferior a $-t_c$, ou superior a $t_c$, rejeita-se $H_0$. Caso contrário, aceita-se $H_0$.

Logo, a região crítica é dada por

$$RC = \{ t_0 \in \mathbb{R} \mid t_0 < -t_c \text{ ou } t_0 > t_c \}$$

e a região de aceitação é a região complementar à região crítica

$$RA = \{ t_0 \in \mathbb{R} \mid -t_c \leq t_0 \leq t_c \}$$

A função poder do teste é dada por

$$\pi(\mu) = 1 - P(\text{erro tipo II})$$
$$= 1 - P(\text{não rejeitar } H_0 \mid H_0 \text{ é falsa})$$
$$= 1 - P(|T| < t_c),$$

onde $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$, $\mu \neq \mu_0$ e $t_c$ é o valor definido anteriormente.

e) Um teste de razão de verossimilhança (TRV) é qualquer teste que tem a região de rejeição da forma $\{x : \lambda(x) \leq c\}$, onde $c$ é qualquer número satisfazendo $0 \leq c \leq 1$.

Aqui, $\lambda(x) = \Lambda = \dfrac{L(\hat{\theta}_0)}{L(\hat{\theta})}$

é a estatística do TRV.

- Calculando $L(\hat{\theta}_0)$

Sob $H_0$, $\theta_0 = (\mu_0, \sigma^2)$. Com isso, $L(\theta)$ é maximizada apenas em $\sigma^2$, pois $\mu_0$ é conhecido. Do item c), vimos que o valor $\sigma^2$ que maximiza $L(\theta)$ é

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n},$$

onde, nesse caso, $\mu$ é substituído por $\mu_0$.
Logo, $\hat{\theta}_0 = (\mu_0, \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2)$, e

$$L(\hat{\theta}_0) = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left[-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (x_i - \mu_0)^2\right]$$

$$= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left[-\frac{n}{2\sigma_0^2} \sigma_0^2\right]$$

$$= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left[-n/2\right]$$

• Calculando $L(\hat{\theta})$

Como não temos restrição em relação aos valores de $\mu$ e $\sigma^2$, do item c), sabemos que $L(\theta)$ é maximizada em

$$\hat{\mu} = \bar{x} \quad e \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Logo, $\hat{\theta} = \left( \bar{x}, \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)$, e

$$L(\hat{\theta}) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp\left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right]$$

$$= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp\left[ -\frac{n}{2\hat{\sigma}^2} \hat{\sigma}^2 \right]$$

$$= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp[-n/2]$$

Uma vez que ~~que~~ calculamos $L(\hat{\theta}_0)$ e $L(\hat{\theta})$, temos que

$$\Lambda = \frac{(2\pi\sigma_0^2)^{-n/2} \exp[-n/2]}{(2\pi\hat{\sigma})^{-n/2} \exp[-n/2]}$$

$$= (\hat{\sigma} / \hat{\sigma}_0)^{n/2}$$

$$= \left[ \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \mu_0)^2} \right]^{n/2}$$

Logo, rejeitamos $H_0$ se

$$\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \mu_0)} < c$$

~~Para o teste de hipóteses $H_0$~~

Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X$, com função de máxima verossimilhança $L(\theta)$. Seja $r_0$ o número de parâmetros livres sob a hipótese nula $H_0 : \theta \in \Theta_0$, e $r$ o número de parâmetros livres especificado por $\theta \in \Theta$, onde $\Theta$ é o espaço paramétrico de $\theta$. Então, a distribuição da estatística $-2 \log \Lambda$ converge para uma distribuição qui-quadrada conforme o tamanho amostral $n \to \infty$. O número de graus de liberdade da distribuição limite é $r - r_0$.

Nesse exercício, a hipótese nula especifica $\mu = \mu_0$, mas não especifica $\sigma^2$, então há um parâmetro livre, $r_0 = 1$. Para $\theta \in \Theta$, $\mu$ e $\sigma^2$ não são especificados, então há dois parâmetros livres, $r = 2$. Logo, sob $H_0$, conforme $n \to \infty$

$$-2 \log \Lambda \to \chi_1^2 \text{ em distribuição}$$

2\. a) Seja $Q = \sum_{i=1}^{n} (Y_i - \beta_1 X_{i1} - \cdots - \beta_K X_{iK})^2$

Para encontrar o estimador de mínimos quadrados de $\beta$, fazemos a derivada parcial de $Q$ em relação a cada elemento de $\beta$ e igualamos a zero. Em seguida, resolvemos a equação resultante para os valores de $\beta$, e os valores encontrados é o estimador de mínimos quadrados.

Dessa forma, dá muito trabalho resolver. Vamos deixar $Q$ na forma matricial para facilitar os cálculos.

$$Q = (Y - X\beta)'(Y - X\beta)$$

$$= (Y' - (X\beta)')(Y - X\beta)$$

$$= Y'Y - (X\beta)'Y - Y'X\beta + (X\beta)'(X\beta)$$

Como $(X\beta)'Y$ é um número, então
$(X\beta)'Y = ((X\beta)'Y)' = Y'X\beta$, logo

$$Q = Y'Y - 2Y'X\beta + \beta^t X^t X\beta$$

Se $x$ é um vetor e $A$ uma matriz de constantes, então

$$\frac{\partial x^t A x}{\partial x} = x'(A + A') \; ;$$

$$\frac{\partial Ax}{x} = A$$

Encontramos o estimador de mínimos quadrados fazendo $\frac{\partial Q}{\partial \beta} = 0$ e encontrando o ponto $\hat{\beta}$ que satisfaz essa igualdade.

$$\frac{\partial Q}{\partial \beta} = \frac{\partial}{\partial \beta}(Y'Y - 2Y'X\beta + \beta'X'X\beta)$$

$$= -2\frac{\partial Y'X\beta}{\partial \beta} + \frac{\partial \beta'X'X\beta}{\partial \beta}$$

$$= -2Y'X + \beta'(X'X + (X'X)')$$

$$= -2Y'X + \beta'(X'X + X'X)$$

$$= -2Y'X \quad 2\beta'X'X = 0$$

$$\Rightarrow \quad \beta'X'X = Y'X \Rightarrow (\beta'X'X)' = (Y'X)'$$

$$\Rightarrow \quad X'X\beta = X'Y \Rightarrow \beta = (X'X)^{-1}X'Y$$

Portanto, $\hat{\beta} = (X'X)^{-1}X'Y$

b) Seja $X_i$ a i-ésima linha da matriz $X$. Então, $Y_i = X_i\beta + \varepsilon_i$. Como $\varepsilon_i \sim N(0, \sigma^2)$, $Y_i \sim N(X_i\beta, \sigma^2)$.

Logo, $Y$ é um vetor de variáveis aleatórias normais

Fazendo $E(Y) = E(X\beta + \varepsilon) = X\beta$ e
$Var(Y) = Var(X\beta + \varepsilon) = Var(\varepsilon) = \sigma^2 I$,

Temos que $Y \sim N(X\beta, \sigma^2 I)$.

Como $\hat{\beta} = (X'X)^{-1}X'Y$, os componentes de $\hat{\beta}$ são ~~I~~ combinações ~~de~~ lineares dos elementos de $Y$. Como cada elemento de $Y$ segue uma distribuição, então os componentes de $\hat{\beta}$ também seguem uma distribuição normal.

Logo, $\hat{\beta} \sim N(E(\hat{\beta}), Var(\hat{\beta}))$

Se $A$ é uma matriz de constantes e $W$ um vetor de variáveis aleatórias, então

$$E(AW) = A\,E(W)$$
$$Var(AW) = A\,Var(W)\,A'$$

Com isso,

$$\begin{aligned} E(\hat{\beta}) &= E((X'X)^{-1}X'Y) = (X'X)^{-1}X'E(Y) \\ &= (X'X)^{-1}X'X\beta = \beta \end{aligned}$$

$$\begin{aligned} Var(\hat{\beta}) &= Var((X'X)^{-1}X'Y) \\ &= (X'X)^{-1}X'\,Var(Y)((X'X)^{-1}X')' \\ &= (X'X)^{-1}X'\sigma^2 I\, X((X'X)^{-1})' \\ &= \sigma^2 (X'X)^{-1}X'X\,((X'X)')^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

Portanto,

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

c) Uma vez que $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$, sabemos que para cada componente de $k$ de $\hat{\beta}$,

$$\hat{\beta}_k \sim N(\beta_k, \sigma^2 S_{kk}),$$

onde $S_{kk}$ é o $k$-ésimo elemento da diagonal de $(X'X)^{-1}$. Logo,

$$Z_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 S_{kk}}} \sim N(0,1)$$

Vamos usar o seguinte Teorema:

Se $x \sim N(0, I)$ e $A$ é uma matriz simétrica $(A' = A)$ e idempotente $(AA = A)$, então $x'Ax$ segue uma distribuição $\chi^2_v$, onde $v$ é o posto de $A$.

Seja $\hat{\varepsilon} = Y - X\hat{\beta}$ o vetor de resíduos, logo

$$\begin{aligned}\hat{\varepsilon} &= Y - X(X'X)^{-1}X'Y = (I - X(X'X)^{-1}X')Y \\ &= MY,\end{aligned}$$

onde $M = I - X(X'X)^{-1}X'$.

Temos que

$$M' = (I - X(X'X)^{-1}X')' = I' - (X(X'X)^{-1}X')'$$

$$= I - (X')'(X(X'X)^{-1})' = I - X((X'X)^{-1})'X'$$

$$= I - X(((X'X)')^{-1}X' = I - X(X'X)^{-1}X' = M$$

e

$$MM = (I - X(X'X)^{-1}X')(I - X(X'X)^{-1}X')$$

$$= II - X(X'X)^{-1}X'I - IX(X'X)^{-1}X'$$

$$+ (X(X'X)^{-1}X')(X(X'X)^{-1}X')$$

$$= I - 2X(X'X)^{-1}X' + X\underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1}X'$$

$$= I - 2X(X'X)^{-1}X' + X(X'X)^{-1}X'$$

$$= I - X(X'X)^{-1}X = M$$

logo, $M$ é simétrica e idempotente

Seja $S^2 = \dfrac{\hat{\varepsilon}\hat{\varepsilon}}{n-p}$ um estimador para $\sigma^2$

$\boxed{\text{aqui é } K \text{ ao invés de } p}$

Precisamos, agora, de alguns resultados de matrizes:

- O ~~ram~~ posto de uma matriz idempotente é o seu traço

- $Tr(A+B) = Tr(A) + Tr(B)$

- $Tr(AB) = Tr(BA)$, onde $A$ é $m \times n$ e $B$ $n \times m$

então,

$$\text{posto}(M) = Tr(M)$$

$$= Tr(I - X(X'X)^{-1}X')$$

$$= Tr(I) - Tr(X(X'X)^{-1}X')$$

$$= Tr(I) - Tr(X'X(X'X)^{-1})$$

$$= \underbrace{Tr(I)}_{n \times n} - \underbrace{Tr(I)}_{K \times K} = n - K$$

Com isso,

$$V = \frac{(n-K)S^2}{\sigma^2} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2}$$

Temos que

$$MX = (I - X(X^TX)^{-1}X^T)X$$

$$= X - X\underbrace{(X'X)^{-1}X'X}_{I} = X - X = 0$$

a partir disso,

$$\hat{\varepsilon} = MY = M(XB + \varepsilon) = MX\beta + M\varepsilon = M\varepsilon$$

e $V = \frac{\hat{\varepsilon}'\varepsilon}{\sigma^2} = \frac{(M\varepsilon)'M\varepsilon}{\sigma^2} = \frac{\varepsilon' M' M\varepsilon}{\sigma^2}$

$$= \frac{\varepsilon' M M \varepsilon}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)' M \left(\frac{\varepsilon}{\sigma}\right)$$

Como $\varepsilon/\sigma \sim N(0, I)$, aplicando o Teorema mencionado, sabemos que $V \sim \chi^2_{n-K}$

Se $Z_K$ e $V$ são duas variáveis aleatórias independentes, pelo resultado descrito em 1. d),

$$T = \frac{Z_K}{\sqrt{V/(n-K)}} \sim t_{n-K}$$