Universidade de Brasília
Departamento de Estatística
Professor: Frederico Machado Almeida
Data: 12/06/2025
Aluno (a): ______________________

Disciplina: Inferência Estatística
Nível: Graduação em Estatística
Semestre: 2025.1, **Turma:** 01
Matrícula: ______________

Prova II

Esta prova contêm 2 páginas, 4 questões totalizando 10 pontos. A duração da prova será de 2 (duas) horas.

Quadro de pontuação (uso EXCLUSIVO do professor)

Questão	**1**	**2**	**3**	**4**	**Total**
Valor	3,0	1,0	2,0	4,0	10,0
Pontuação	3,0	1,0	2,0	4,0	10

Questão 1 (3,0 pontos). Assinale com V, as alternativas verdadeiras e F as falsas. Justifique as afirmações que julgar serem falsas.

(i) Assuma que cada um dos 45 pesquisadores de um grupo de investigação obteve separadamente e de forma independente um intervalo de confiança a 93% para a média μ da resposta (considerada normal) de um organismo sujeito à presença de determinada substância química. Assim, é correto afirmar que:

a) (F) Construindo um número grande de intervalos de confiança, espera-se que 93 intervalos contenham o verdadeiro valor do parâmetro μ.

Justificativa: Espera-se que 93% dos intervalos contenham o verdadeiro valor de μ.

b) (V) A amplitude de um intervalo de confiança aumenta quando se diminui o tamanho de amostra n, mantendo fixas as demais quantidades.

c) (F) O nível de confiança indica que a probabilidade de μ pertencer ao intervalo de confiança observado é de aproximadamente 0,93.

Justificativa: A probabilidade de $\mu \in IC_{0,93}$ é 0 ou 1.

d) (F) A amplitude de um intervalo de confiança aumenta quando se reduz a variância dos dados, mantendo constante as demais quantidades.

Justificativa: Segue da definição de um intervalo de confiança que $AIC = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Então, como as duas grandezas são diretamente proporcionais, então, reduzir a variância dos dados, σ^2, implica em reduzir a amplitude do IC.

e) (F) O tamanho de amostra aumenta em duas vezes, reduzindo a margem de erro pela metade, e mantendo constante as demais quantidades.

Justificativa: Observe que, $n = \left(\frac{z_{\alpha/2}\sigma}{\epsilon}\right)^2$. Então, se $\epsilon' = \epsilon/2$, segue que, $n' = \left(\frac{z_{\alpha/2}\sigma}{\epsilon'}\right)^2 = \left(\frac{z_{\alpha/2}\sigma}{\epsilon/2}\right)^2 = 4\left(\frac{z_{\alpha/2}\sigma}{\epsilon}\right)^2$. Ou seja, o tamanho de amostra aumenta quadruplica

f) (F) Aumentando o nível de significância, α a amplitude de um intervalo de confiança específico vai diminuir.

Justificativa: Observe que, aumentar o nível de significância, α, significa reduzir o nível de confiança, $1-\alpha$. Consequentemente, na redução do $z_{\alpha/2}$, o que implica a redução da amplitude do IC.

(ii) Em um teste estatístico de hipóteses, é correto afirmar que:

a) (F) Um teste de hipóteses mais poderoso pressupõem que, tanto a H_0 como a H_1 podem ser hipóteses simples ou compostas, a depender do problema.

Justificativa: Segue da definição dos testes mais poderosos (*vide nos slides de aula*).

b) (F) De acordo com a literatura, o erro tipo II é sempre o mais grave de ser cometido, por estar associado à função poder.

Justificativa: O erro tipo I é um mais grave (*vide nos slides de aula*).

c) (V) O erro tipo II é a probabilidade de não rejeitar uma H_0 falsa.

d) (V) Em um teste de hipóteses mais poderoso, a melhor região crítica será sempre aquela que apresentar maior poder, entre todas as regiões com tamanho α.

e) (F) O erro tipo I é a probabilidade de não rejeitar uma H_0 verdadeira.

Justificativa: O erro tipo I é a probabilidade de rejeitar uma H_0 verdadeira (*vide nos slides de aula*).

f) (F) Em um teste da razão de verossimilhanças, a H_0 será rejeitada à favor da H_1 se $\Lambda \leq k$, com $k > 0$.

Justificativa: Em um razão de verossimilhanças a H_0 será rejeitada à favor da H_1 se $\Lambda \leq k$, com $0 \leq k \leq 1$.

Questão 2 (1,0 ponto). Seja $\bar{X}_n$ a média de uma amostra aleatória $X_1, X_2, \cdots, X_n$ proveniente da distribuição normal de média μ e variância $\sigma^2 = 9$. Encontre o tamanho mínimo de amostra n, tal que, a confiança aproximada de μ cair dentro do intervalo $\left(\bar{X}_n - 1/2 < \mu < \bar{X}_n + 1/2\right)$ seja de 0,90.

Resposta: Segue da definição de um intervalo de confiança que, $AIC = L_s - Li = \bar{X}_n + 1/2 - \bar{X}_n + 1/2 = 1$. Como consequência, a margem de erro é dada por: $\epsilon = AIC/2 = 1/2$. Logo, para $\sigma = 3$ e $\alpha = 0,1$ temos que, $z_{\alpha/2} = 1,645$. Assim,

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{\epsilon}\right)^2 = \left(\frac{1,645 \times 3}{0,5}\right)^2 = 97,417 \approx 98.$$

Ou seja, o tamanho mínimo de amostra para o qual a margem de erro não vai passar dos 0,5 unidades é de aproximadamente 98 observações.

Questão 3 (2,0 pontos). Seja $X_1, X_2, \cdots, X_n$ uma amostra aleatória de tamanho n proveniente da distribuição caracterizada por uma função densidade de probabilidade dada por:

$$f(x|\theta) = \frac{1}{2\theta} \exp\left(\frac{|x|}{\theta}\right), \quad \text{com } x \in \mathcal{R}, \ \ \theta > 0. \tag{1}$$

Encontre a melhor região crítica e de aceitação para um teste mais poderoso de tamanho α, associado as seguintes hipóteses, $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$.

Resposta: Como $X_1, X_2, \cdots, X_n$ são cópias *iid's* de uma v.a. $X \sim f(x|\theta)$, então, sua fdp conjunta, ou função de verossimilhança é dada por:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{2\theta} \exp\left(\frac{|x_i|}{\theta}\right) = (2\theta)^{-n} \exp\left(\frac{1}{\theta}\sum_{i=1}^{n}|x_i|\right).$$

Desta forma, segue imediatamente que,

$$\text{Sob } H_0: L(\theta_0|\mathbf{x}) = (2\theta_0)^{-n} \exp\left(\frac{1}{\theta_0}\sum_{i=1}^{n}|x_i|\right)$$

$$\text{Sob } H_1: L(\theta_1|\mathbf{x}) = (2\theta_1)^{-n} \exp\left(\frac{1}{\theta_1}\sum_{i=1}^{n}|x_i|\right).$$

Portanto, segue do lema de Neyman-Pearson que, a estatística de teste é dada por:

$$R(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\theta_1|\mathbf{x})} = \frac{(2\theta_0)^{-n} \exp\left(\frac{1}{\theta_0}\sum_{i=1}^{n}|x_i|\right)}{(2\theta_1)^{-n} \exp\left(\frac{1}{\theta_1}\sum_{i=1}^{n}|x_i|\right)} = \left(\frac{\theta_1}{\theta_0}\right)^n \exp\left[\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^{n}|x_i|\right].$$

Por fim, a H_0 será rejeitada à favor da H_1 se $R(\mathbf{x}) \leq k$, com $k > 0$. Ou seja, se $\left(\frac{\theta_1}{\theta_0}\right)^n \exp\left[\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^{n}|x_i|\right] \leq k \iff \exp\left[\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^{n}|x_i|\right] \leq k\left(\frac{\theta_0}{\theta_1}\right)^n \iff \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^{n}|x_i| \leq \log\left[k\left(\frac{\theta_0}{\theta_1}\right)^n\right]$. Desta forma, se $\theta_1 > \theta_0$ segue que, $RC = \{\mathbf{x} : \sum_{i=1}^{n}|x_i| \leq k^*\}$, e se $\theta_1 < \theta_0$ segue que, $RC = \{\mathbf{x} : \sum_{i=1}^{n}|x_i| > k^*\}$, com $k^* = \log\left[k\left(\frac{\theta_0}{\theta_1}\right)^n\right] / \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)$. Onde $RA = RC^c$.

Questão 4 (4,0 pontos). Seja $X_1, X_2, \cdots, X_n$ cópias independentes e identicamente distribuídas de uma variável aleatória X com distribuição $\mathcal{N}(\mu, \sigma^2)$, onde $\sigma^2 = 100$.

(a) (**1,25 pontos**) Assuma que as hipóteses de interesse são dadas por, $H_0 : \mu = \mu_0$ contra $H_1 : \mu = \mu_1$. Pede-se para obter a melhor região crítica, e de aceitação para um teste mais poderoso associado de tamanho α.

Resposta: Segue do enunciado que, $X_1, X_2, \cdots, X_n$ são cópias $iid's$ de uma variável aleatória $X \sim \mathcal{N}(\mu, \sigma^2 = 100)$. Portando, para que possamos obter a melhor região crítica de tamanho α, vamos obter a função de verossimilhança (ou densidade conjunta do modelo normal), tal que, $f(x|\theta) = (2\pi 100)^{-1/2} \exp\left[-\frac{1}{2\times 100}(x-\mu)^2\right]$. Cosequentemente,

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} (200\pi)^{-1/2} \exp\left[-\frac{1}{200}(x_i-\mu)^2\right] = (200\pi)^{-n/2} \exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu)^2\right],$$

consequentemente, os valores da função de verossimilhança sob H_0 e H_1 são dados por:

$$\text{Sob } H_0 : L(\mu_0|\mathbf{x}) = (200\pi)^{-n/2} \exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_0)^2\right],$$

$$\text{Sob } H_1 : L(\mu_1|\mathbf{x}) = (200\pi)^{-n/2} \exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_1)^2\right].$$

Segue do lema de Neyma-Pearson que, a estatística de teste é dada por:

$$R(\mathbf{x}) = \frac{L(\mu_0|\mathbf{x})}{L(\mu_1|\mathbf{x})} = \frac{(200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_0)^2\right]}{(200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_1)^2\right]} = \exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_0)^2 + \frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_1)^2\right]$$

$$= \exp\left[\frac{1}{200}\sum_{i=1}^{n}\left\{x_i^2 - 2x_i\mu_1 + \mu_1^2 - x_i^2 + 2\mu_0 x_i - \mu_0^2\right\}\right] = \exp\left[\frac{n\bar{x}}{100}(\mu_0-\mu_1) + n\left(\mu_1^2-\mu_0^2\right)\right].$$

Assim, a H_0 será rejeitada à favor da H_1 se $R(\mathbf{x}) \leq k$, com $k > 0$. Ou seja se,

$$\exp\left[\frac{n\bar{x}}{100}(\mu_0-\mu_1) + n\left(\mu_1^2-\mu_0^2\right)\right] \leq k \Longleftrightarrow \exp\left[\frac{n\bar{x}}{100}(\mu_0-\mu_1)\right] \leq k\exp\left[-n\left(\mu_1^2-\mu_0^2\right)\right]$$

$$\leftrightarrow \frac{n\bar{x}}{100}(\mu_0-\mu_1) \leq \log(k) - n\left(\mu_1^2-\mu_0^2\right) \Longleftrightarrow \bar{x}(\mu_0-\mu_1) \leq \frac{100}{n}\left[\log(k) - n\left(\mu_1^2-\mu_0^2\right)\right].$$

Desta forma, se $\mu_1 > \mu_0$ segue que, $RC = \{\mathbf{x} : \bar{x}_{obs} > k^*\}$ e, se $\mu_1 < \mu_0$ temos que, $RC = \{\mathbf{x} : \bar{x}_{obs} \leq k^*\}$, com $k^* = \frac{100}{n(\mu_0-\mu_1)}\left[\log(k) - n\left(\mu_1^2-\mu_0^2\right)\right]$, com $RA = RC^c$.

(b) (**0,75 ponto**) Supondo agora que as hipóteses de interesse são, $H_0 : \mu = 75$ contra $H_1 : \mu = 78$, mostre que $RC = \{(x_1, x_2, \cdots, x_n) : c \leq \bar{x} = \sum_{i=1}^{n} x_i/n\}$ é a melhor região crítica para o teste mais poderoso de tamanho α.

Resposta: Com base no resultado do item anterior, do enunciado e das hipóteses de interesse, temos que, $f(x|\mu) = (2\pi 100)^{-1/2}\exp\left[-\frac{1}{2\times 100}(x-\mu)^2\right]$. Cosequentemente, os valores da função de verossimilhança sob H_0 e H_1 são dados por:

$$\text{Sob } H_0 : L(\mu_0 = 75|\mathbf{x}) = (200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_0)^2\right] = (200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-75)^2\right],$$

$$\text{Sob } H_1 : L(\mu_1 = 78|\mathbf{x}) = (200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu_1)^2\right] = (200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-78)^2\right].$$

Entretanto, segue do teorema de Neyman-Pearson que, a estatística $R(\mathbf{x})$ é dada pelo seguinte rácio:

$$R(\mathbf{x}) = \frac{L(\mu_0 = 75|\mathbf{x})}{L(\mu_1 = 78|\mathbf{x})} = \frac{(200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-75)^2\right]}{(200\pi)^{-n/2}\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-78)^2\right]} = \frac{\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-75)^2\right]}{\exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-78)^2\right]}$$

$$= \exp\left[-\frac{1}{200}\sum_{i=1}^{n}(x_i-75)^2\right]\exp\left[\frac{1}{200}\sum_{i=1}^{n}(x_i-78)^2\right]$$

$$= \exp\left[-\frac{1}{200}\sum_{i=1}^{n}\left(x_i^2 - 2\times 75x_i + 75^2\right) + \frac{1}{200}\sum_{i=1}^{n}\left(x_i^2 - 2\times 78x_i + 78^2\right)\right]$$

$$= \exp\left[-\frac{1}{200}\sum_{i=1}^{n}x_i^2 + \frac{75}{100}\sum_{i=1}^{n}x_i - \frac{75^2 n}{200} + \frac{1}{200}\sum_{i=1}^{n}x_i^2 - \frac{78}{100}\sum_{i=1}^{n}x_i + \frac{78^2 n}{200}\right]$$

$$= \exp\left[-0,03\sum_{i=1}^{n}x_i + \left(\frac{78^2 n}{200} - \frac{75^2 n}{200}\right)\right].$$

Assim, a H_0 será rejeitada à favor da H_1, se para algum $k > 0$, tivermos que, $R(\mathbf{x}) \leq k \iff \exp\left[-0,03\sum_{i=1}^{n} x_i + \left(\frac{78^2 n}{200} - \frac{75^2 n}{200}\right)\right] \leq k \iff \exp\left(-0,03\sum_{i=1}^{n} x_i\right) \leq k\exp\left[-\left(\frac{78^2 n}{200} - \frac{75^2 n}{200}\right)\right]$. Aplicando logaritmo natural nos dois membros obtemos: $-0,03\sum_{i=1}^{n} x_i \leq \log(k) - \left(\frac{78^2 n}{200} - \frac{75^2 n}{200}\right) \iff 0,03\sum_{i=1}^{n} x_i \geq \log(k) - \left(\frac{78^2 n}{200} - \frac{75^2 n}{200}\right) \iff \sum_{i=1}^{n} x_i > \frac{\log(k)-\left(\frac{78^2 n}{200}-\frac{75^2 n}{200}\right)}{0,03} = k^*$. Portanto, podemos escrever a região crítica do teste mais poderoso como $RC = \{\mathbf{x} : \bar{x} > c = k^*/n\}$. Ou seja, a H_0 será rejeitada à favor da H_1 sempre que $\bar{x}_{obs}$ for maior que $c = k^*/n$.

(c) (**2 pontos**) Com base no resultado do item (b), encontre as constantes n e c tal que, $\alpha(\mu_0) = \mathbb{P}\left(\bar{X}_n \geq c|\mu_0 = 75\right) = 0,05$ e $\pi(\mu_1) = \mathbb{P}\left(\bar{X}_n \geq c|\mu_1 = 78\right) = 0,94$. Com $\alpha(\mu_0)$ e $\pi(\mu_1)$ denotando o tamanho e o poder do teste, respectivamente.

Resposta: Com base no enunciado, queremos encontrar n e c tal que,
$\mathbb{P}(\text{Erro tipo I}) = \mathbb{P}_{H_0}\left(\bar{X}_n \geq c|\mu = 75\right) = 0,05$ e $\pi(\mu_1) = 1-\mathbb{P}(\text{Erro tipo II}) = \mathbb{P}_{H_1}\left(\bar{X}_n \geq c|\mu = 78\right) = 0,90$. Desta forma, segue imediatamente que:

$$\begin{aligned}
(i)\ \mathbb{P}_{H_0}\left(\bar{X}_n \geq c|\mu = 75\right) &= \mathbb{P}_{H_0}\left(\frac{\bar{X}_n - \mu}{10/\sqrt{n}} \geq \frac{c-\mu}{10/\sqrt{n}}\middle|\mu = 75\right) = \mathbb{P}_{H_0}\left(Z \geq \frac{c-75}{10/\sqrt{n}}\right) \\
&= 1 - \mathbb{P}_{H_0}\left(Z \leq \frac{c-75}{10/\sqrt{n}}\right) = 1 - \Phi\left(\frac{c-75}{10/\sqrt{n}}\right) = 0,05 \\
&= \Phi\left(\frac{c-75}{10/\sqrt{n}}\right) = 0,95 \iff \frac{c-75}{10/\sqrt{n}} = \Phi^{-1}(0,95) = 1,645.
\end{aligned}$$

$$\begin{aligned}
(ii)\ \pi(\mu) = \mathbb{P}_{H_1}\left(\bar{X}_n \geq c|\mu = 78\right) &= \mathbb{P}_{H_0}\left(\frac{\bar{X}_n - \mu}{10/\sqrt{n}} \geq \frac{c-\mu}{10/\sqrt{n}}\middle|\mu = 78\right) = \mathbb{P}_{H_0}\left(Z \geq \frac{c-78}{10/\sqrt{n}}\right) \\
&= 1 - \mathbb{P}_{H_0}\left(Z \leq \frac{c-78}{10/\sqrt{n}}\right) = 1 - \Phi\left(\frac{c-78}{10/\sqrt{n}}\right) = 0,94 \\
&= \Phi\left(\frac{c-78}{10/\sqrt{n}}\right) = 0,06 \iff \frac{c-78}{10/\sqrt{n}} = \Phi^{-1}(0,06) = -1.555,
\end{aligned}$$

sendo $\Phi(\cdot)$ a fda da normal padrão. Assim, isolando c nas expressões (i) e (ii) obtemos o seguinte resultado, $c = 75+1,645\times\frac{10}{\sqrt{n}}$ e $c = 78-1,555\times\frac{10}{\sqrt{n}}$, respectivamente. Portanto, igualando as duas equações anteriores e resolvendo como função de n obtemos: $75+1,645\times\frac{10}{\sqrt{n}} = 78-1,555\times\frac{10}{\sqrt{n}} \iff \sqrt{n} = 9,757 \iff \left(\sqrt{n}\right)^2 = 10,667^2 \iff n = 113,778 \approx 114$. Logo, o valor da constante c será dado por, $c = 75 + 1,645 \times \frac{10}{\sqrt{114}} \approx 76,54$. Observe que, se considerássemos a outra equação do c, o resultado seria o mesmo!!!

FIM!