Parte 2: Probabilidade e Inferência

25. (Família de localização/posição) Seja $\mathcal{F}_\theta = \{f(\cdot;\theta), \theta \in \mathbb{R}\}$ uma família de densidades. O parâmetro θ é definido ser um *parâmetro de localização/posição* se e somente se a densidade $f(x;\theta)$ poder ser escrita da forma $f(x;\theta) = g(x-\theta)$, com g uma função conhecida, denominada de função *geradora*. Adicionalmente, diz-se que a densidade $f(\cdot;\theta)$ (ou a V.A. associada) é um membro da família de localização. Mostre que as seguintes V.A's pertencem à família de localização, identificando o parâmetro de localização e a respectiva função geradora:

i) $X \sim \mathcal{N}(\mu, \sigma^2)$, com σ conhecido.

ii) $X \sim \text{Cauchy}(\mu, \sigma)$, com σ conhecido, i.e.,

$$f_X(x) = \frac{\sigma}{\pi\left[\sigma^2 + (x-\mu)^2\right]}\mathbb{1}_{\mathbb{R}}(x).$$

iii) $X \sim t_k(\mu, \sigma)$, com k e σ conhecidos, i.e.,

$$f_X(x) = \frac{1}{\sqrt{k}B(1/2, k/2)\sigma}\left(1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{k}\right)^{-(k+1)/2}\mathbb{1}_{\mathbb{R}}(x).$$

iv) $X \sim \text{Laplace}(\mu, \sigma)$, com σ conhecido, i.e.,

$$f_X(x) = \frac{1}{2\sigma}e^{-\frac{|x-\mu|}{\sigma}}\mathbb{1}_{\mathbb{R}}(x).$$

26. (Família de escala) Seja $\mathcal{F}_\theta = \{f(\cdot;\theta), \theta \in \mathbb{R}\}$ uma família de densidades. O parâmetro θ é definido ser um *parâmetro de escala* se e somente se a densidade $f(x;\theta)$ poder ser escrita da forma $f(x;\theta) = \theta^{-1}g(x/\theta)$, com g uma função conhecida, denominada de função *geradora*. Adicionalmente, diz-se que a densidade $f(\cdot;\theta)$ (ou a V.A. associada) é um membro da família de escala. Mostre que as seguintes V.A's pertencem à família de escala, identificando o parâmetro de escala e a respectiva função geradora:

i) $X \sim \mathcal{U}(-\theta,\theta)$.

ii) $X \sim \exp(\lambda)$.

iii) $X \sim \mathcal{N}(\mu,\sigma^2)$, com μ conhecido.

iv) $X \sim t_k(\mu,\sigma)$, com k e μ conhecidos.

v) $X \sim \text{Laplace}(\mu,\sigma)$, com μ conhecido.

27.(Família de localização-escala) Seja $\mathcal{F}_\theta = \{f(\cdot;\theta_1,\theta_2), \theta_1 \in \mathbb{R},\ \theta_2 > 0\}$ uma família de densidades. Os parâmetros θ_1 e θ_2 são definidos, respectivamente, parâmetros de *localização* e de *escala* se e somente se a densidade $f(x;\theta_1,\theta_2)$ poder ser escrita da forma $f(x;\theta) = \theta_2^{-1}g((x-\theta_1)/\theta_2)$, com g uma função conhecida, denominada de função *geradora*. Adicionalmente, diz-se que a densidade $f(\cdot;\theta_1,\theta_2)$ (ou a V.A. associada) é um membro da família de localização-escala. Mostre que as seguintes V.A's pertencem à família de localização-escala, identificando os parâmetros de localização e de escala, e a respectiva função geradora:

i) $X \sim \mathcal{N}(\mu,\sigma^2)$.

ii) $X \sim t_k(\mu,\sigma)$, com k conhecido.

iii) $X \sim \text{Laplace}(\mu,\sigma)$.

iv) $X \sim \text{Logística}(\mu,\sigma)$, i.e.,

$$f_X(x) = \frac{e^{-\left(\frac{x-\mu}{\sigma}\right)}}{\sigma\left\{1+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right\}^2}\mathbb{1}_{\mathbb{R}}(x)$$

28.(Distribuições simétricas) Considere X uma variável aleatória com suporte em $\mathbb{R}$, com parâmetro de localização $\mu \in \mathbb{R}$ e de escala $\phi > 0$. Dizemos que X pertence a família de distribuições simétricas, se sua densidade é escrita da seguinte forma

$$f_X(x) = \frac{1}{\sqrt{\phi}} g\left\{\frac{(x-\mu)^2}{\phi}\right\} \mathbb{1}_{\mathbb{R}}(x),$$

para alguma função $g(\cdot)$ denominada função geradora de densidades, com $g(u) > 0$ se $u > 0$ e $\int_0^\infty u^{-1/2} g(u) d_u = 1$. Mostre que as seguintes V.A's pertencem à família de de distribuições simétricas, identificando o parâmetro de localização o de escala e a respectiva função geradora:

i) $X \sim \mathcal{N}(\mu, \sigma^2)$.

ii) $X \sim \text{Cauchy}(\mu, \sigma)$.

iii) $X \sim t_k(\mu, \sigma)$, com k conhecido.

iv) $X \sim \text{Logística}(\mu, \sigma)$.

Sugestão de leitura: Para maiores detalhes sobre esta classe de distribuições, bem como de aplicações em modelos de regressão veja: Cysneiros, F.J., Paula, G.A. e Galea, M. (2007). *Modelos simétricos aplicados.* ABE, 9ª Escola de Modelos de Regressão, Águas de São Pedro-SP, Brasil.

29. Considere $X_1, \ldots, X_n$ uma a.a. de uma densidade $f(\cdot)$. Mostre que se $f(\cdot)$ pertence a família de:

i) Localização, então $T = \bar{X} - \theta$ é uma quantidade pivotal.

ii) Escala, então $T = \bar{X}/\theta$ é uma quantidade pivotal.

iii) Localização (θ_1) e escala (θ_2), então $T = (\bar{X} - \theta_1)/\theta_2$ é uma quantidade pivotal.

30. Defina:

i) Estatística, estimador, estimativa, estimação e parâmetro.

ii) Faça uma comparação entre intervalo de confiança e teste de hipóteses, abordando os objetivos de cada técnica. Qual a grande vantagem de se fazer um teste de hipóteses com relação a construção de um IC?

31. Explique os erros tipo I e tipo II que podem ocorrer nas seguinte situações:

i) Um júri decide condenar ou não o réu.

ii) O juiz marcou penalidade máxima contra o time azul.

32. (**Gráfico do poder do teste**) Define-se como gráfico do poder do teste para um determinado teste de hipóteses, a curva que expressa o comportamento do poder $(1-\beta)$ em função das diversas hipóteses alternativas $\mathcal{H}_1$, fixando-se o nível de significância α. Desenhe o gráfico do poder do teste para $\mathcal{H}_0 : \mu = 10$ versus $\mathcal{H}_1 : \mu \neq 10$, em que $n = 9$, $\sigma^2 = 4$ e $\alpha = 0,05$. Para simplificar, assuma normalidade da população e admita: 7,0; 7,5; 8,0; 8,5; 9,0; 9,5; 10,00; 10,5; 11,0; 11,5; 12,0; 12,5 e 13,0 como possíveis valores de μ.

33. Refaça a questão anterior, considerando diferentes tamanhos de amostras, por exemplo, $n = 20$ e $n = 30$. Desenhe as três curvas em um mesmo gráfico e comente.

34. Considere que o interesse é testar a hipótese $\mathcal{H}_0 : \mu = \mu_0$, com σ conhecido e sob suposição de normalidade. Mostre que a probabilidade do erro tipo II para um teste de nível α, para os diferentes tipos de hipóteses alternativas, são dadas por:

Hipótese	**Probabilidade do erro tipo II**
$\mathcal{H}_1 : \mu = \mu' > \mu_0$	$\Phi\left(z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)$
$\mathcal{H}_1 : \mu = \mu' < \mu_0$	$1 - \Phi\left(-z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)$
$\mathcal{H}_1 : \mu = \mu' \neq \mu_0$	$\Phi\left(z_{\alpha/2} + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)$,

em que $\Phi(z) = \mathbb{P}(Z \leq z)$ e $z_\alpha = \Phi^{-1}(\alpha)$ representam, respectivamente, a função distribuição acumulada e o quantil de ordem $\alpha \in (0,1)$ da distribuição $\mathcal{N}(0,1)$. Adicionalmente, mostre que $\lim_{n\to\infty} \beta(\mu') = 0, \forall \mu' \neq \mu_0$. Você consegue explicar este resultado?

35. De forma análoga a questão anterior, encontre a probabilidade do erro tipo II, quando o interesse é testar $\mathcal{H}_0 : \mu = \mu_0$, com σ desconhecido e sob suposição de normalidade. Deixe a resposta em termos da função distribuição acumulada da distribuição t com k graus de liberdade: $\Phi_k(\cdot)$.

36.(Determinação do tamanho da amostra) Em situações práticas é desejável controlar a probabilidade de se cometer ambos os tipos de erros. Fixado α, temos que a probabilidade de se cometer o erro tipo II é uma função decrescente do tamanho da amostra (Questões # 31 e # 32). Baseado no resultado que você obteve na Questão # 33, mostre que o tamanho da amostra para o qual um teste de nível α possui probabilidade tipo II igual a $\beta(\mu')$ (perceba que usamos o valor μ' especificado) é dado por

Hipótese	**Tamanho da amostra**
$\mathcal{H}_1 : \mu = \mu' > \mu_0 (\mu' < \mu_0)$	$n = \left\lfloor \left(\frac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right)^2 \right\rfloor$,

em que $\lfloor x \rfloor$, representa o menor número inteiro maior ou igual a x.

37. Para investigar a influência do tipo de ensino (Particular e Público, referente ao ultimo ano do ensino médio) sobre a média no curso de Introdução à Estatística de recém-ingressos na UFC, obteve-se a seguinte amostra:

Tabela 1: Notas de 20 alunos no curso de Introdução à Estatística.

Particular	Público
2,5	8,0
7,8	4,2
3,5	7,4
8,3	4,0
5,0	4,1
3,9	5,6
6,0	5,5
10,0	6,0
9,1	5,2
2,5	3,3

i) Denotando as notas dos estudantes oriundos de escolas particulares (públicas) por $x_1, \ldots, x_{10}$ $(y_1, \cdots, y_{10})$ obtenha $\bar{x}, \bar{y}, s_x^2$ e s_y^2.

ii) Calcule $\hat{\sigma}^2 = ((n_x - 1)s_x^2 + (n_y - 1)s_y^2)/(n_x + n_y - 2)$, em que n_x e n_y, representam, respectivamente, o número de alunos na amostra oriundos de escolas particulas e públicas.

iii) Recalcule $\bar{x}, \bar{y}, s_x^2, s_y^2$ e σ^2, usando ~~as~~ formas matriciais ~~apresentadas na questão # 19~~.

iv) Considerando válida as suposicções de Normalidade, independência e igualdade de variâncias, utilize o teste t para verificar se existe evidência a favor da hipótese de que os alunos oriundos de escolas particulares apresentam um melhor desempenho na disciplina.

38. (Uso de simulação para avaliar um teste de hipóteses) As seguintes hipóteses sobre a média são consideradas $\mathcal{H}_0 : \mu = 1$ vs $\mathcal{H}_1 : \mu \neq 1$. Fixado um nível de significância α, o teste irá a rejeitar a hipótese nula $\alpha\%$ das vezes, quando esta for verdadeira. Para estimar α empiricamente, utilize o seguinte procedimento de simulação:

i) Gerar 10.000 amostras de tamanho 20 da distribuição $\mathcal{N}(1, 2)$.

ii) Para cada amostra gerada determinar as estatísticas apropriadas para o teste e realizar o teste usando $\alpha = 0.05$. Determine se a hipótese nula é ou não rejeitada.

iii) Determinar a porcentagem de amostras em que a hipótese nula é rejeitada, digamos $\hat{\alpha}$) e comparamos com o verdadeiro valor α.

Você espera que $\hat{\alpha} = 0.05$? Justifique.

Adicionalmente, para avaliar a função poder do teste no ponto $\mu \neq 1$, você pode repetir o processo acima, trocando o item i) por

i') Gerar 10.000 amostras de tamanho 100 da distribuição $\mathcal{N}(\mu', 2)$.

Estime empiricamente o poder do teste para $\mu = 0.2, 0.4, 0.6, 0.8, 1.2, 1.4, 1.6, 1.8$. Comente os resultados.

39. Estime empiricamente o nível de significância do teste de hipóteses $\mathcal{H}_0 : \mu = 1$ vs $\mathcal{H}_1 : \mu \neq 1$, quando você gera valores de uma distribuição χ^2_1 (média 1 e variância 2) e use o teste obtido sob normalidade. Comente os resultados.

40. Considere $X_1, \ldots, X_n$ uma a.a. da distribuição Laplace$(\mu, 1)$. Obtenha o teste da razão de verossimilhanças generalizada para $\mathcal{H}_0 : \mu = 0$ vs $\mathcal{H}_1 : \mu \neq 0$ ao nível α. Estime empiricamente o tamanho e o poder do teste utilizando os valores $\mu = 0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8, \pm 1.2, \pm 1.4, \pm 1.6, \pm 1.8$. Plote a curva da função poder empírica. Considere $n = 5, 10, 15, 20, 30$ e 50 e discuta os resultados.

41. Considere $X_1, \ldots, X_n$ uma a.a. da distribuição $\mathcal{N}(\mu, 1)$, $\mu \geq 0$. Encontre o EMV de μ.

42. Considere X_1 uma a.a. da distribuição $\mathcal{N}(\mu, \sigma^2)$.

i) Mostre que os parâmetros são ortogonais. O que isso facilita no processo de estimação?

ii) Determine a distribuição assintótica conjunta de $(\hat{\mu}, \hat{\sigma}^2)^\top$, com $\hat{\theta}$ representando o EMV de θ.

43. Considere $Y_1, \ldots, Y_n$ independentes, tais que $Y_i \sim \mathcal{N}(\beta x_i, \sigma^2)$, x_i conhecido.

i) Determine os EMV de β e σ^2.

ii) Os parâmetros são ortogonais? Você esperaria esse resultado? Justifique.

ii) Determine a distribuição assintótica conjunta de $(\hat{\beta}, \hat{\sigma}^2)^\top$.

44. Apresente um resumo sobre:

i) A distribuição Normal multivariada e suas principais propriedades.

ii) Distribuições de formas lineares e quadráticas (sob suposição de normalidade).