UNIVERSIDADE DE BRASÍLIA

DEPARTAMENTO DE ESTATÍSTICA

Disciplina: INFERÊNCIA ESTATÍSTICA
Curso: Graduação em Estatística
Código: EST0035
Semestre: 2025.1
Professor: Frederico Machado Almeida

LISTA DE EXERCÍCIOS #04

Observações:

- Questões para entregar: 2, 4, 7 (b), 8 e 11.
- Demais questões são apenas para estudar.
- Prazo de entrega: **06/06/2025**

Q1. Durante vários anos, uma determinada tarefa no processo de fabrico de um produto foi executada pelo Sr. Silva, que a levava a efeito num tempo médio de 35 minutos. O Sr. Silva abandonou a empresa, e foi substituído por um novo operário, o jovem Alberto que, apesar de não ter nenhuma experiência, frequentou um curso de formação profissional que o pode tornar mais eficiente. Admita-se que o tempo de execução da tarefa pelo novo operário segue distribuição aproximadamente normal, com desvio-padrão de 4 minutos.

(a) Se, nas últimas 25 observações, o Alberto demorou, em média, 34 minutos, como classificaria a performance do jovem operário?

Resposta: Queremos verificar se a duração do processo pelo Alberto (34 minutos) é estatisticamente significativa comparada com os 35 minutos do Sr. Silva. Desse modo, as hipóteses são:

$$H_0 : \mu = 35 \quad \text{(Alberto tem o mesmo desempenho)}$$
$$H_1 : \mu < 35 \quad \text{(Alberto é mais rápido)}$$

Segue do enunciado que o tempo de Alberto segue uma distribuição aproximadamente normal com desvio-padrão de 4 minutos. Dessa forma, como a variância populacional é conhecida, a estatística de teste para esse problema é dada por:

$$z_{obs} = \frac{\sqrt{n}\,(\bar{x}_{obs} - \mu)}{\sigma} \overset{sob\ H_0}{=} \frac{\sqrt{n}\,(\bar{x}_{obs} - \mu_0)}{\sigma} = \frac{\sqrt{25}\,(34 - 35)}{4} = -1,25.$$

Utilizando a tabela da distribuição normal padrão, o resultado é:

$$\mathbb{P}(Z < -1,25) \approx 0,1056$$

Como o p-valor (0,1056) é maior que o nível de significância ($\alpha = 0,05$), não se rejeita H_0. Ou seja, não há evidência estatística suficiente para afirmar que Alberto é mais rápido do que o Sr. Silva, ao nível de confiança de 95%.

(b) Ao decidir não rejeitar H_0, existe a consciência de se poder estar a cometer um erro. Qual a respetiva probabilidade, se for verdade que o Alberto demora só 34 minutos em média? E se, pelo contrário, for verdade que ele demora mais? Ou seja, que demora 37 minutos?

Resposta: O exercício pede para calcular a probabilidade de erro tipo II, ou seja, a probabilidade de não rejeitar H_0 quando este é falso.

Como é um teste unilateral à esquerda com $\alpha = 0,05$, tem-se valor crítico de $z_{obs} = -1,645$. Então rejeita-se H_0 em favor de H_1 se $Z \leq -Z_\alpha$, logo:

$$
\begin{aligned}
Z \leq -Z_\alpha &\Rightarrow \frac{\sqrt{25}(\bar{x}_n - 35)}{4} \leq -Z_\alpha \\
&\Rightarrow \bar{x}_n - 35 \leq -\frac{4Z\alpha}{\sqrt{25}} \\
&\Rightarrow \bar{x}_n \leq 35 - \frac{4 \times 1,645}{\sqrt{25}} \\
&\Rightarrow \bar{x}_n \leq 33,684
\end{aligned}
$$

Ou seja, H_0 será rejeitada se a média observada for inferior a 33,684

Caso 1: A média verdadeira é 34 minutos

Queremos:

$$\text{Erro tipo II} = \mathbb{P}\left(NRH_0|\mu = 34\right) = \mathbb{P}(\bar{x} \geq 33{,}684 \mid \mu = 34)$$

Calculando o z_{obs} :

$$z = \frac{\sqrt{25}(33,684 - 34)}{4} = -0,395$$

Portanto:

$$\mathbb{P}(Z \geq -0,395) = 1 - \mathbb{P}(Z < -0,395) \approx 1 - 0,346 = 0,654$$

Então, se a média real for 34 minutos, há aproximadamente 65,4% de probabilidade de não rejeitar H_0 (erro tipo II).

Caso 2: A média verdadeira é 37 minutos

Calculando o z_{obs} para esse caso:

$$z = \frac{\sqrt{25}(33,684 - 37)}{4} = -4,145$$

Portanto:

$$\mathbb{P}(Z \geq -4,145) = 1 - \mathbb{P}(Z < -4,145) \approx 1$$

Então, se a média real for 37 minutos, a probabilidade de não rejeitar H_0 (tomar a decisão correta) é praticamente 100%.

Q2. Assuma que a vida útil de uma marca de Pneus em milhas seja denotada por uma variável aleatória X, que é normalmente distribuída com média θ e desvio-padrão 4500. Experiências passadas indicam que $\theta = 30000$. A afirmação do fabricante é de que os pneus fabricados por meio de um novo processo tem uma vida útil maior que 30000. Supondo que uma amostra aleatória de tamanho n, digamos, $x_1, x_2, \cdots, x_n$ tenha sido observada. A H_0 será rejeitada à favor da H_1 se $\bar{x}_{obs} > c$. Determine n e c tal que, a função poder do teste seja, $\pi(30000) = 0,002$ e $\pi(35000) = 0,96$.

Resposta: Segue do enunciado que $X \sim \mathcal{N}(\theta, \sigma^2)$, com $\theta \in \mathbb{R}$ sendo desconhecido e $\sigma = 4500$ conhecido. As hipóteses de interesse são, $H_0 : \theta = 30000$ (ou $\theta \leq 30000$) contra $H_1 : \theta > 30000$. Portanto, para refutar ou não uma determinada hipótese, uma amostra de tamanho n foi selecionada da população em causa, e forneceu uma média de $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Desta forma, como a variância populacional é conhecida, a estatística de teste para esse problema é dada por:

$$z_{obs} = \frac{\sqrt{n}\left(\bar{x}_{obs} - \theta\right)}{\sigma} \overset{sob\ H_0}{=} \frac{\sqrt{n}\left(\bar{x}_{obs} - \theta_0\right)}{\sigma} = \frac{\sqrt{n}\left(\bar{x}_{obs} - 30000\right)}{4500}.$$

Assim, como estamos testando uma hipótese do tipo unilateral à direita, temos com regra de decisão: rejeitar a H_0 à favor da H_1 se $z_{obs} > z_\alpha \iff \frac{\sqrt{n}(\bar{x}_{obs}-30000)}{4500} > z_\alpha \iff \bar{x}_{obs} > 30000 + 4500\frac{z_\alpha}{\sqrt{n}}$. Ou seja, a região crítica do teste é dada por: $RC = \{\mathbf{x} : \bar{x}_{obs} > c\}$, com $c = 30000 + 4500\frac{z_\alpha}{\sqrt{n}}$. Observe que, por definição a função poder é dada pela probabilidade de tomar uma decisão correta, ou seja, não rejeitar H_1 quando ela for verdadeira. Desta forma, segue que,

$$\begin{aligned}
\pi(\theta^*) &= \mathbb{P}(NRH_1|\theta > 30000) = \mathbb{P}\left(\bar{X}_n > c|\theta > 30000\right) = \mathbb{P}\left(\bar{X}_n > c|\theta = \theta^*\right) \;\; \theta^* \in \Theta_1 \\
&= \mathbb{P}\left(\frac{\sqrt{n}\left(\bar{X}_n - \theta^*\right)}{\sigma} > \frac{\sqrt{n}\left(c - \theta^*\right)}{\sigma}\right) = \mathbb{P}\left(Z > \frac{\sqrt{n}\left(c - \theta^*\right)}{\sigma}\right) \\
&= 1 - \mathbb{P}\left(Z \leq \frac{\sqrt{n}\left(c - \theta^*\right)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}\left(c - \theta^*\right)}{\sigma}\right).
\end{aligned}$$

Portanto, avaliando nos pontos $\theta^* = \theta_0$ (com $\theta_0 = 30000$) e $\theta^* = 35000$ obtemos o seguinte resultado: (i) segue da definição dada na sala de aula que $\alpha(\theta_0) = \pi(\theta_0) = 1 - \mathbb{P}\left(Z \leq \frac{\sqrt{n}(c-\theta_0)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}(c-30000)}{4500}\right)$. Como $\pi(30000) = 0,002$ segue imediatamente que, $1 - \Phi\left(\frac{\sqrt{n}(c-30000)}{4500}\right) = 0,002 \iff \Phi\left(\frac{\sqrt{n}(c-30000)}{4500}\right) = 1 - 0,002 = 0,998 \iff \frac{\sqrt{n}(c-30000)}{4500} = \Phi^{-1}(0,998)$. Logo, $c = 30000 + 4500\Phi^{-1}(0,998)/\sqrt{n}$, com $\Phi^{-1}(0,998) = z_{(0,998)} = 2,878$.

(ii) De forma similar, para um valor de $\theta^* = 35000$ o valor da função poder é dado por: $\pi(\theta^*) = 1 - \Phi\left(\frac{\sqrt{n}(c-\theta^*)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}(c-35000)}{4500}\right) = 0,96 \iff \frac{\sqrt{n}(c-35000)}{4500} = \Phi^{-1}(0,04) = -1,751$. Logo, $c = 35000 - 1,751 \times 4500/\sqrt{n}$. Igualando as duas expressões de c obtidas nos itens (i) e (ii) segue que: $30000 + 4500 \times 2,878/\sqrt{n} = 35000 - 1,751 \times 4500/\sqrt{n}$. Resolvendo essa equação obtemos o seguinte resultado, $n = \left(\frac{20830,5}{5000}\right)^2 \approx 18$. Desta forma, pegando uma das equações do c obtemos o seguinte resultado, $c = 33052,58$. Consequentemente, a região crítica do teste de hipóteses pode ser reescrita da seguinte forma, $RC = \{\mathbf{x} : \bar{x}_{obs} > 33052,58\}$.

Q3. O Ministério da Saúde afirma que, com os meios agora postos à disposição dos Hospitais Civis, o número médio de dias de internamento é, no máximo, 15. Estas declarações foram postas em causa por alguns gestores hospitalares que decidiram proceder em conjunto à recolha de uma amostra de 225 doentes, onde se observou que o número médio de dias de internamento foi de 18. Com base nestes dados, e supondo que a variável em estudo segue uma distribuição normal com desvio-padrão de 15 dias:

(a) Terão os gestores hospitalares razão? Justifique convenientemente a sua resposta, utilizando o teste adequado, a 1% de significância. Na decisão que tomou, qual a probabilidade de estar a cometer um erro?

Resposta: Queremos testar se os gestores têm razão, ou seja, se a média é maior do que 15. Logo:

$$\begin{cases} H_0 : \mu = 15 \\ H_1 : \mu > 15 \end{cases}$$

Como o desvio-padrão populacional é conhecido e a variável segue distribuição normal. Então, como estamos perante uma hipótese unilateral à direita, RH_0 à favor de H_1 se $z > z_\alpha$:

$$\begin{aligned} z_{obs} = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} &> z_\alpha \\ \Rightarrow \frac{\bar{x}_n - 15}{5/\sqrt{225}} &> z_\alpha \\ \Rightarrow \bar{x}_n &> 15 + \frac{5z_\alpha}{\sqrt{225}} \end{aligned}$$

Temos que a região crítica será:

$$C_\alpha = \left\{(x_1, ..., x_n) : \bar{x}_n > 15 + \frac{5z_\alpha}{\sqrt{225}}\right\}$$

Com nível de significância de 1%, buscamos o valor crítico da normal padrão $z_{1-\alpha} = z_{0,99} \approx 2,33$. Logo:

$$\begin{aligned} C_\alpha &= \left\{(x_1, ..., x_n) : \bar{x}_n > 15 + \frac{5 \times 2,33}{\sqrt{225}}\right\} \\ &= \{(x_1, ..., x_n) : \bar{x}_n > 15,776\} \end{aligned}$$

Então, rejeitamos H_0 se $\bar{x}_n > 15,776$. Como temos que $\bar{x}_n = 18$, então rejeita-se H_0 e se dá razão aos gestores hospitalares.

Outra forma de tomar essa decisão é calculando o z associado aos dados, ou seja:

$$z_{obs} = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{18 - 15}{5/\sqrt{225}} = \frac{3}{5/15} = \frac{3}{1/3} = 9$$

Sabemos que o $z_{0,99} = 2,33$, então, como o valor observado $z = 9$ é muito maior que o valor crítico $2,33$, rejeitamos a hipótese nula. Portanto, há evidências estatísticas suficientes, ao nível de significância de 1%, para rejeitar a afirmação do Ministério. Logo, os gestores hospitalares têm razão em questionar que a média de dias de internamento é maior que 15.

Probabilidade de erro do tipo I: É a probabilidade de rejeitarmos H_0 sendo ela verdadeira, ou seja:

$$\mathbb{P}(\text{Erro Tipo I}) = \mathbb{P}(RH_0|H_0 \text{ Verdadeira}) = \alpha = 0,01 = 1\%$$

(*b*) Com que probabilidade e dada razão aos gestores hospitalares, se o verdadeiro número médio de dias de internamento for 17?

Resposta: Neste item, queremos calcular o poder do teste, ou seja, a probabilidade de rejeitar H_0 se a média real fosse 17 dias.

$$\begin{aligned}
\pi_\tau(\mu^*) &= \mathbb{P}(NRH_1|H1 \text{ Verdadeira}) \\
&= 1 - \mathbb{P}(RH_1|H_1 \text{ Verdadeira}) = 1 - \beta(\mu) \\
&= 1 - \mathbb{P}\left(\frac{\bar{X}_n - \mu^*}{5/\sqrt{225}} \leq \frac{15,776 - \mu^*}{5/\sqrt{225}}\right) \quad \textbf{com } \mu^* \in \Theta_1 \\
&= 1 - \mathbb{P}\left(Z < \frac{15,776 - \mu^*}{5/\sqrt{225}}\right) \\
&= 1 - \Phi\left(\frac{15,776 - \mu^*}{5/\sqrt{225}}\right) \\
&= \Phi\left(\frac{\mu^* - 15,776}{5/\sqrt{225}}\right)
\end{aligned}$$

Portanto, avaliando no ponto $\mu^* = 17$, temos que:

$$\pi_\tau(\mu^*) = \mathbb{P}(\bar{X}_n > 15,776|\mu = 17) = \Phi\left(\frac{17 - 15,776}{5/\sqrt{225}}\right) = \Phi(3,672) = 0,9998$$

Portanto, se a verdadeira média de internamento for 17 dias, há aproximadamente 99,98% de probabilidade de rejeitarmos a hipótese do Ministério, ou seja, de dar razão aos gestores.

(*c*) Como variaria aquela probabilidade se a hipótese alternativa fosse superior ao valor especificado na alínea (*b*)? E se o tamanho da amostra aumentasse?

Resposta:

- Se $\mu > 17$: Quanto maior for o valor verdadeiro da média, maior será a distância entre μ e a hipótese nula ($\mu = 15$), o que implica em maiores valores de z. Isso torna o teste mais sensível, aumentando o poder do teste (isto é, a probabilidade de detectar corretamente que $\mu > 15$).
- Se o tamanho da amostra n aumentar: O erro padrão da média amostral, dado por $EP = \frac{\sigma}{\sqrt{n}}$, diminui. Isso faz com que a estatística z aumente (para a mesma diferença entre médias), o que também eleva o poder do teste.

Q4. Seja X uma variável aleatória com distribuição Poisson de média θ. Considere as seguintes hipóteses, $H_0 : \theta = 3/4$ contra $H_1 : \theta < 3/4$. Então, com base nas hipóteses, podemos afirmar que $\Theta = \{\theta : 0 < \theta \leq 3/4\}$. Seja $X_1, X_2, \cdots, X_{16}$ uma amostra aleatória de tamanho 16 da distribuição supracitada. Assuma que a H_0 será rejeitada se $T(\mathbf{X}) = \sum_{i=1}^{16} x_i \leq 2$. Se $\pi(\theta)$ denota a função poder, obtenha os poderes $\pi(1/8)$, $\pi(1/4)$, $\pi(1/2)$ e $\pi(3/4)$, e faça o respectivo gráfico. Qual é o tamanho desse teste?

Resposta: Segue do enunciado que, $X \sim Poisson(\theta)$, com $\theta > 0$. As hipóteses de interesse são, $H_0 : \theta = 3/4$ contra $H_1 : \theta < 3/4$, ou podemos apresentar as hipóteses da seguinte forma, $H_0 : \theta = \theta_0$ $(com\ \theta_0 = 3/4)$ e $H_1 : \theta = \theta_1$ $(com\ \theta_1 < \theta_0)$, onde H_0 é rejeitada à favor da H_1 se $\sum_{i=1}^{n} x_i \leq 2$. Isso significa que, $RC = \{\mathbf{x} : \sum_{i=1}^{n} x_i \leq 2\}$, ou seja, a função poder do teste será representado por:

$$
\begin{aligned}
\pi(\theta^*) &= \mathbb{P}\left(\sum_{i=1}^{n} X_i \leq 2|\theta < 3/4\right) = \mathbb{P}\left(\sum_{i=1}^{n} X_i \leq 2|\theta = \theta^*\right) = \mathbb{P}(T \leq 2|\theta = \theta^*) \\
&= \sum_{t=0}^{2} \mathbb{P}(T = t|\theta = \theta^*) = \sum_{t=0}^{2} e^{-n\theta^*} \frac{(n\theta^*)^t}{t!} = e^{-n\theta^*}\left(1 + n\theta^* + \frac{(n\theta^*)^2}{2}\right).
\end{aligned}
$$

Observe que, como $X_i \sim Poisson(\theta)$, segue imediatamente que $T = \sum_{i=1}^{n} X_i \sim Poisson(n\theta)$, com $n = 16$. Assim,

$$
\begin{aligned}
\text{Se } \theta^* = 1/8, \quad \pi(1/8) &= e^{-16/8}\left(1 + (16/8) + \frac{(16/8)^2}{2}\right) \approx 0,677 \\
\text{Se } \theta^* = 1/4, \quad \pi(1/4) &= e^{-16/4}\left(1 + (16/4) + \frac{(16/4)^2}{2}\right) \approx 0,238 \\
\text{Se } \theta^* = 1/2, \quad \pi(1/2) &= e^{-16/2}\left(1 + (16/2) + \frac{(16/2)^2}{2}\right) \approx 0,014 \\
\text{Se } \theta^* = 3/4, \quad \pi(3/4) &= e^{-16\times 3/4}\left(1 + (16\times 3/4) + \frac{(16\times 3/4)^2}{2}\right) \approx 0,0005.
\end{aligned}
$$

Observe que o tamanho do teste $\alpha = \alpha(\theta_0) \equiv \pi(\theta_0) = 0,0005$, pois, sob H_0 $\theta = \theta_0$ $(= 3/4)$. O gráfico da função poder $\pi(\theta^*)$ como função de θ^* é apresentado abaixo.

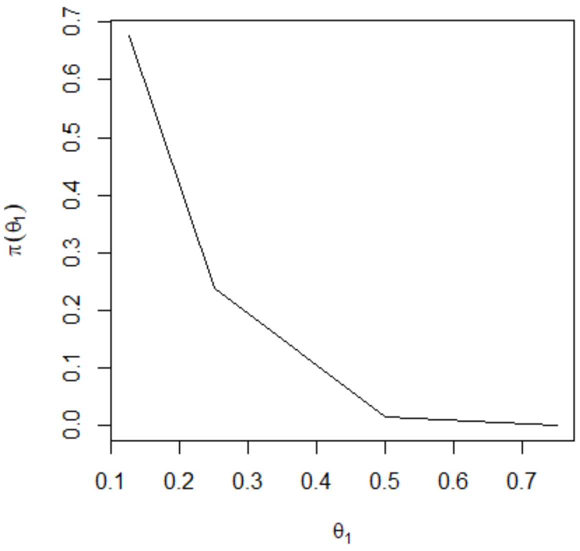

Figure 1: Curva da função poder para diferentes valores de θ^*.

Q5. Seja $X_1, X_2, \cdots, X_n$ uma amostra aleatória de $X \sim \mathcal{N}(\theta, \sigma^2 = 100)$. Mostre, usando um teste *MP* que $RC = \{(x_1, x_2, \cdots, x_n) : c \leq \bar{x}\}$ é a melhor região crítica para testar as hipóteses $H_0 : \theta = 75$ contra $H_1 : \theta = 78$. Encontre n e c tal que, $\mathbb{P}_{H_0}(T(\mathbf{X} \in C_\alpha)) = \mathbb{P}_{H_0}\left(\bar{X}_n \geq c\right) = 0,05$ e $\mathbb{P}_{H_1}(T(\mathbf{X} \in C_\alpha)) = \mathbb{P}_{H_1}\left(\bar{X}_n \geq c\right) = 0,90$ aproximadamente.

Resposta: Como a amostra foi extraída de uma população normal, a fdp da variável aleatória X é dada por:

$$f(x|\theta) = (2\pi \times 100)^{-1/2} \exp\left\{-\frac{1}{2 \times 100}(x - \theta)^2\right\},$$

resultando na seguinte função de verossimilhança,

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta) = (2\pi \times 100)^{-n/2} \exp\left\{-\frac{1}{2 \times 100}\sum_{i=1}^{n}(x_i - \theta)^2\right\}.$$

O próximo passo seria obter as funções de verossimilhanças restritas às duas hipóteses em questão. Considerando $H_0 : \theta = \theta' = 75$ e $H_0 : \theta = \theta'' = 78$, temos que:

$$\text{Sob } H_0 : \quad L(\theta'|\mathbf{x}) = (2\pi \times 100)^{-n/2} \exp\left\{-\frac{1}{2 \times 100}\sum_{i=1}^{n}(x_i - 75)^2\right\},$$

$$\text{Sob } H_1 : \quad L(\theta''|\mathbf{x}) = (2\pi \times 100)^{-n/2} \exp\left\{-\frac{1}{2 \times 100}\sum_{i=1}^{n}(x_i - 78)^2\right\}.$$

Com base no resultado anterior, a estatística que define a razão das duas funções de verossimilhança é dada por:

$$R(\mathbf{x}) = \frac{L(\theta'|\mathbf{x})}{L(\theta''|\mathbf{x})} = \frac{(2\pi \times 100)^{-n/2} \exp\left\{-\frac{1}{2\times 100}\sum_{i=1}^{n}(x_i - 75)^2\right\}}{(2\pi \times 100)^{-n/2} \exp\left\{-\frac{1}{2\times 100}\sum_{i=1}^{n}(x_i - 78)^2\right\}} = \exp\left\{-\frac{6\sum_{i=1}^{n} x_i}{200} + \frac{459 \times n}{200}\right\}.$$

Para algum $k > 0$, rejeitamos H_0 a favor da H_1 se $R(\mathbf{x}) < k$. Resultando na seguinte região crítica:

$$C_\alpha = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i \geq k^*\right\},$$

sendo:

$$k = \exp\left\{-\frac{6\sum_{i=1}^{n} x_i}{200} + \frac{459 \times n}{200}\right\}$$

$$200\log(k) = -3\sum_{i=1}^{n} x_i + 459 \times n$$

$$200\log(k) - 459n = -3\sum_{i=1}^{n} x_i$$

$$\frac{-200\log(k) + 459n}{3} = \sum_{i=1}^{n} x_i = k^*$$

Sendo que a região crítica pode ser igualmente apresentada da seguinte forma:

$$C_\alpha = \{\mathbf{x} : \bar{x}_{obs} \geq k^{**}\},$$

com $k^{**} = \frac{-200\log(k)}{3n} + 153$.
Como os dados são provenientes de uma distribuição normal, segue das aulas anteriores que $\bar{X}_n \sim \mathcal{N}(\theta, 100/n)$. Ou seja, a quantidade k^{**} pode ser obtida facilmente usando os quantis da distribuição normal, usando o fato de que:

$$\alpha = \mathbb{P}_{H_0}\left(\bar{X}_n \geq k^{**}\right) = \mathbb{P}\left(\frac{\bar{X}_n - \theta_0}{\sqrt{100/n}} > \frac{k^{**} - 75}{\sqrt{100/n}} | \theta_0 = 75\right) = \mathbb{P}\left(Z > \frac{k^{**} - 75}{\sqrt{100/n}}\right)$$
$$= 1 - \mathbb{P}\left(Z \leq \frac{k^{**} - 75}{\sqrt{100/n}}\right) = 1 - \Phi\left[\frac{\sqrt{n}(k^{**} - 75)}{10}\right] = 0,05$$

Desse modo, segue que:

$$1 - \Phi\left[\frac{\sqrt{n}(k^{**} - 75)}{10}\right] = 0,05$$

$$\Rightarrow \Phi\left[\frac{\sqrt{n}(k^{**} - 75)}{10}\right] = 0,95$$

$$\Rightarrow \frac{\sqrt{n}(k^{**} - 75)}{10} = \Phi^{-1}(0,95) = z_{0,95} = 1,645$$

$$\Rightarrow k^{**} = \frac{16,45}{\sqrt{n}} + 75$$

Agora, também temos que $\mathbb{P}_{H_1}(\bar{X}_n \geq c) = 0,90$. Desse modo:

$$\mathbb{P}_{H_1}\left(\bar{X}_n \geq k^{**}\right) = \mathbb{P}\left(\frac{\bar{X}_n - \theta_1}{\sqrt{100/n}} > \frac{k^{**} - 78}{\sqrt{100/n}} | \theta_1 = 78\right) = \mathbb{P}\left(Z > \frac{k^{**} - 78}{\sqrt{100/n}}\right)$$
$$= 1 - \mathbb{P}\left(Z \leq \frac{k^{**} - 78}{\sqrt{100/n}}\right) = 1 - \Phi\left[\frac{\sqrt{n}(k^{**} - 78)}{10}\right] = 0,90$$

Desse modo, segue que:

$$1 - \Phi\left[\frac{\sqrt{n}(k^{**} - 78)}{10}\right] = 0,90$$

$$\Rightarrow \Phi\left[\frac{\sqrt{n}(k^{**} - 78)}{10}\right] = 0,10$$

$$\Rightarrow \frac{\sqrt{n}(k^{**} - 78)}{10} = \Phi^{-1}(0,10) = z_{0,10} = -1,28$$

$$\Rightarrow k^{**} = \frac{-12,8}{\sqrt{n}} + 78$$

Portanto, o que temos é:

- $k^{**} = \frac{16,45}{\sqrt{n}} + 75$
- $k^{**} = \frac{-12,8}{\sqrt{n}} + 78$

Igualando ambos os termos:

$$\frac{16,45}{\sqrt{n}} + 75 = \frac{-12,8}{\sqrt{n}} + 78$$

$$\frac{16,45}{\sqrt{n}} + \frac{-12,8}{\sqrt{n}} = 3$$

$$\left(\frac{16,45 + 12,8}{3}\right)^2 \approx 95,06 \Rightarrow 96$$

Desse modo, n deve ser, pelo menos, 96 para que os valores das probabilidades sejam atendidos.

Q6. Assuma que o peso de cereal em uma caixa de 10kg segue uma distribuição normal de média μ, e variância σ^2. Para testar as hipóteses $H_0 : \mu = 10,1$ contra $H_1 : \mu > 10,1$ uma amostra de tamanho $n = 16$ foi extraída da distribuição de X, tendo se observado $\bar{x} = 10,4$ e $s = 0,40$.

(a) O que se pode dizer quanto a rejeição ou não da H_0 ao nível de significância de 5%?

Resposta: Como nessa questão σ^2 é desconhecido, então usaremos o estimador pontal s^2. Dessa forma, H_0 será rejeitada à favor de H_1 se $t_{obs} > t_{(n-1,\alpha)}$, em que, sob H_0:

$$t_{obs} = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{S} \sim t_{n-1}$$

Ou seja, substituindo os valores da questão:

$$\begin{aligned} t_{obs} &= \frac{\sqrt{16}(\bar{X}_n - 10,1)}{0,4} > t_{(15;0,95)} \\ &\Rightarrow \bar{X}_n > 10,1 + \frac{1,753 \times 0,4}{4} \\ &\Rightarrow \bar{X}_n > 10,275 \end{aligned}$$

Logo, temos a região crítica $RC = \{\mathbf{x} > \bar{x}_{obs} > 10,275\}$. Como tem-se que $\bar{x}_n = 10,4$, então rejeita-se H_0, ao nível de significância de 95%. Outra forma de tomar essa decisão é calculando o z associado aos dados, ou seja:

$$z_{obs} = \frac{\bar{x}_n - \mu_0}{s/\sqrt{n}} = \frac{10,4 - 10,1}{0,4/\sqrt{16}} = \frac{0,3}{0,4/4} = \frac{0,3}{0,1} = 3$$

Sabemos que o $t_{(15;0,95)} = 1,753$, então, como o valor observado $t_{obs} = 3$ é maior que o valor crítico $1,753$, novamente rejeitamos a hipótese nula.

(b) Qual é o valor-p aproximado do teste?

Resposta: O p-valor da questão vai ser dado por:

$$\text{valor-p} = \mathbb{P}(T_{n-1} > t_{obs}) = P(T_{15} > 3) = 1 - P(T_{15} \leq 3) = 1 - 0,9955 = 0,0045$$

Como p-valor$= 0,0045 < 0,05 = \alpha$, então, de fato rejeita-se H_0, ao nível de significância de 95%.

Q7. Considerando amostras aleatórias das distribuições a seguir, obtenha o teste da razão de verossimilhanças para testar $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$. Se possível, obtenha a distribuição exata da estatística do teste, caso contrário, obtenha a distribuição aproximada. Com a distribuição exata ou aproximada, calcule a função poder exata (ou aproximada) dos testes que você obteve.

(a) $f(x|\theta) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}} \mathbf{1}_{(-\infty,\infty)}(x)$, com $\theta > 0$.

Resposta: Observe que, para essa questão estamos interessados em computar uma estatística da razão de verossimilhanças para um teste de hipótese bilateral, ou seja, $H_:\theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$. Observe que, conforme apresentado em sala de aulas, o primeiro passo para computar esse tipo de teste, é a obtenção da função de verossimilhança, ou seja,

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

Observe que, conforme apresentado nas aulas anteriores, o estimador de máxima verossimilhança para θ é obtido derivando o logaritmo da expressão anterior como função de θ. Ou seja, $\ell(\theta) = -n\log(2\theta) - \frac{1}{\theta}\sum_{i=1}^{n}|x_i| \iff \ell'(\theta) = \frac{-n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}|x_i| = 0 \iff \hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}|x_i|$ denota o estimador de máxima verossimilhança para θ. Portanto, os valores da função de verossimilhança no modelo restrito (sob H_0) e irrestrito, são dados por:

$$\begin{aligned}
\text{Sob } H_0: &\quad L(\theta_0|\mathbf{x}) = (2\theta_0)^{-n}\exp\left(-\frac{1}{\theta_0}\sum_{i=1}^{n}|x_i|\right)\\
\text{Sob } H_1: &\quad L\left(\hat{\theta}_n|\mathbf{x}\right) = (2\hat{\theta}_n)^{-n}\exp\left(-\frac{1}{\hat{\theta}_n}\sum_{i=1}^{n}|x_i|\right).
\end{aligned}$$

Observe que, sob H_0 é suficiente assumir que a $L(\theta|\mathbf{x})$ é maximizada no ponto $\theta = \theta_0$, porque essa é uma hipótese simples (isto é, o θ assume um único valor). Portanto, a estatística do teste da razão de verossimilhanças é dada por:

$$\begin{aligned}
\Lambda &= \frac{L(\theta_0|\mathbf{x})}{L\left(\hat{\theta}_n|\mathbf{x}\right)} = \frac{(2\theta_0)^{-n}\exp\left(-\frac{1}{\theta_0}\sum_{i=1}^{n}|x_i|\right)}{(2\hat{\theta}_n)^{-n}\exp\left(-\frac{1}{\hat{\theta}_n}\sum_{i=1}^{n}|x_i|\right)}\\
&= \left(\frac{\hat{\theta}_n}{\theta_0}\right)^n \exp\left(\frac{\sum_{i=1}^{n}|x_i|}{\hat{\theta}_n} - \frac{\sum_{i=1}^{n}|x_i|}{\theta_0}\right) = \left(\sum_{i=1}^{n}|x_i|\right)^n \exp\left(-\frac{\sum_{i=1}^{n}|x_i|}{\theta_0}\right)\left[\frac{e}{n\theta_0}\right]^n.
\end{aligned}$$

Portanto, o teste rejeita a H_0 se para algum $0 \leq k \leq 1$, a estatística de teste Λ for tal que, $\Lambda \leq k \iff \left(\sum_{i=1}^{n}|x_i|\right)^n \exp\left(-\frac{\sum_{i=1}^{n}|x_i|}{\theta_0}\right)\left[\frac{ne}{\theta_0}\right]^n \leq k \iff \left(\sum_{i=1}^{n}|x_i|\right)^n \exp\left(-\frac{\sum_{i=1}^{n}|x_i|}{\theta_0}\right) \leq k\left[\frac{\theta_0}{ne}\right]^n$. Aplicando logaritmo nos dois membros obtemos:

$$n\log\left(\sum_{i=1}^{n}|x_i|\right) - \frac{\sum_{i=1}^{n}|x_i|}{\theta_0} \leq \log\left(\frac{\theta_0^n}{n^n e^n}\right).$$

Assim, a região crítica para esse teste é dada por: $RC = \{\mathbf{x} : t(\mathbf{x}) = n\log\left(\sum_{i=1}^{n}|x_i|\right) - \sum_{i=1}^{n}|x_i|/\theta_0 \leq k^*\}$, com $k^* = n\log\left(\frac{\theta_0}{ne}\right)$. Portanto, segue do Teorema 1 (SLIDE 53) que, $-2\log\Lambda \sim \chi^2_{(1)}$, sendo essa uma distribuição aproximada. Ou seja, ela vale para n suficientemente grande.

(*b*) $f(x|\theta) = \frac{x}{\theta^2}e^{-\frac{x}{\theta}}\mathbf{1}_{(0,\infty)}(x)$, com $\theta > 0$.

Resposta: Observe que para responder esse item, podemos usar a mesma abordagem apresentada no item (*a*) Ou seja, computando primeiro a função de verossimilhança, tal que,

$$L\left(\theta|\mathbf{x}\right)=\prod_{i=1}^{n}f\left(x_i|\theta\right)=\prod_{i=1}^{n}\left[\frac{x_i}{\theta^2}e^{-\frac{x_i}{\theta}}\mathbf{1}_{(0,\infty)}(x_i)\right]=\theta^{-2n}\left(\prod_{i=1}^{n}x_i\right)\exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}x_i\right).$$

Note agora que o estimador de máxima verossimilhança para θ (verifique ou veja o Gabrito da Prova I), é dado por $\hat{\theta}_n=\bar{X}_n/2$. Portanto, tal como no item anterior, os valores da função de verossimilhança no modelo restrito (sob H_0) e irrestrito, são dados por:

$$\text{Sob } H_0: \quad L\left(\theta_0|\mathbf{x}\right)=\theta_0^{-2n}\left(\prod_{i=1}^{n}x_i\right)\exp\left(-\frac{1}{\theta_0}\sum_{i=1}^{n}x_i\right)$$
$$\text{Sob } H_1: \quad L\left(\hat{\theta}_n|\mathbf{x}\right)=\hat{\theta}_n^{-2n}\left(\prod_{i=1}^{n}x_i\right)\exp\left(-\frac{1}{\hat{\theta}_n}\sum_{i=1}^{n}x_i\right).$$

Observe que, sob H_0 é suficiente assumir que a $L\left(\theta|\mathbf{x}\right)$ é maximizada no ponto $\theta=\theta_0$, porque essa é uma hipótese simples (isto é, o θ assume um único valor). Portanto, a estatística do teste da razão de verossimilhanças é dada por:

$$\begin{aligned}\Lambda &= \frac{L\left(\theta_0|\mathbf{x}\right)}{L\left(\hat{\theta}_n|\mathbf{x}\right)}=\frac{\theta_0^{-2n}\left(\prod_{i=1}^{n}x_i\right)\exp\left(-\frac{1}{\theta_0}\sum_{i=1}^{n}x_i\right)}{\hat{\theta}_n^{-2n}\left(\prod_{i=1}^{n}x_i\right)\exp\left(-\frac{1}{\hat{\theta}_n}\sum_{i=1}^{n}x_i\right)}\\ &= \left(\frac{\hat{\theta}_n}{\theta_0}\right)^{2n}\exp\left(\frac{\sum_{i=1}^{n}x_i}{\hat{\theta}_n}-\frac{\sum_{i=1}^{n}x_i}{\theta_0}\right)=\bar{x}^{2n}\exp\left(-\frac{n\bar{x}}{\theta_0}\right)\left[\frac{e}{2\theta_0}\right]^{2n}.\end{aligned}$$

Portanto, o teste rejeita a H_0 se para algum $0\leq k\leq 1$, a estatística de teste Λ for tal que, $\Lambda\leq k \iff \bar{x}^{2n}\exp\left(-\frac{n\bar{x}}{\theta_0}\right)\left[\frac{e}{2\theta_0}\right]^{2n}\leq k \iff \bar{x}^{2n}\exp\left(-\frac{n\bar{x}}{\theta_0}\right)\leq k\left[\frac{2\theta_0}{e}\right]^{2n}$. Aplicando logaritmo nos dois membros obtemos:

$$t(\mathbf{x})=2n\log\left(\bar{x}\right)-\frac{n\bar{x}}{\theta_0}\leq\log\left(k\frac{e}{2\theta_0}\right)^{2n}.$$

Assim, a região crítica para esse teste é dada por: $RC=\{\mathbf{x}:t(\mathbf{x})\leq k^*\}$, com $k^*=\log\left(\frac{ke}{2\theta_0}\right)^{2n}$. Portanto, segue igualmente do Teorema 1 (SLIDE 53) que, $-2\log\Lambda\sim\chi^2_{(1)}$, sendo essa uma distribuição aproximada para n suficientemente grande.

Q8. Seja $X_1,X_2,...,X_n$ cópias *iid's* de uma variável aleatória X com distribuição exponencial, tal que, $f\left(x|\theta\right)=(1/\theta)\,e^{-x/\theta}$, $0<x<\infty$, e zero caso contrário.

(a) Supondo $n=2$, mostre que o teste com melhor região crítica de $H_0:\theta=2$ contra $H_1:\theta=4$ é dado pela seguinte estatística $R(X_1,X_2)=X_1+X_2$.

Resposta: Observe que, para esse problema em particular temos duas variáveis, digamos X_1 e X_2 que são cópias $iid's$ de $X\sim f\left(x|\theta\right)=(1/\theta)\,e^{-x/\theta}$, com $x>0$ e $\theta>0$. As hipóteses de interesse são, $H_0:\theta=2$ e $H_1:\theta=4$. Desta forma, a função de verossimilhança é dada por:

$$L\left(\theta|x_1,x_2\right)=f\left(x_1|\theta\right)f\left(x_2|\theta\right)=\left(1/\theta\right)^2\exp\left[-\left(x_1+x_2\right)/\theta\right].$$

Os valores que a função de verossimilhança assume sob H_0 e H_1 são dados por:

$$\begin{aligned}\text{Sob } H_0: \quad L(\theta = 2|x_1, x_2) &= (1/2)^2 \exp\left[-(x_1 + x_2)/2\right]; \ \ \theta' = 2.\\ \text{Sob } H_1: \quad L(\theta = 4|x_1, x_2) &= (1/4)^2 \exp\left[-(x_1 + x_2)/4\right]; \ \ \theta'' = 4.\end{aligned}$$

Entretanto, a estatística do teste $R(x_1, x_2)$ é dada pela seguinte expressão:

$$\begin{aligned}R(x_1, x_2) &= \frac{L(\theta = 2|x_1, x_2)}{L(\theta = 4|x_1, x_2)} = \frac{(1/2)^2 \exp\left[-(x_1 + x_2)/2\right]}{(1/4)^2 \exp\left[-(x_1 + x_2)/4\right]}\\ &= 4\exp\left[\frac{-2(x_1 + x_2)}{4} + \frac{(x_1 + x_2)}{4}\right] = 4\exp\left[-(x_1 + x_2)/4\right].\end{aligned}$$

Assim, para algum $k > 0$, a H_0 será rejeitada à favor da H_1 se $R(x_1, x_2) \leq k \iff 4\exp\left[-(x_1 + x_2)/4\right] \leq k \iff -(x_1 + x_2)/4 \leq \log(k/4) \iff t(x_1, x_2) = x_1 + x_2 > -4\log(k/4)$. Ou seja, a região crítica (ou de rejeição) desse teste é dada por, $RC = \{\mathbf{x} : t(x_1, x_2) > k^* = -4\log(k/4)\}$, onde $T(X_1, X_2) = X_1 + X_2$.

(b) Para a amostra $X_1, X_2, ..., X_n$ obtenha o teste da razão de verossimilhança para testar as hipóteses $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.

Resposta: Temos que $X \sim f(x|\theta) = (1/\theta)e^{-x/\theta}$, com $x > 0$ e $\theta > 0$. As hipóteses de interesse são, $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$. Desta forma, a função de verossimilhança é dada por:

$$L(\theta|\mathbf{X}) = \prod_{i=1}^{n} f(X_i|\theta) = (1/\theta)^n \exp\left[-\left(\sum_{i=1}^{n} X_i\right)/\theta\right].$$

Precisamos encontrar o estimador de máxima verossimilhança (EMV) de θ, seguindo o passo a passo visto em aulas anteriores. Tomamos o logaritmo da função de verossimilhança:

$$\ell(\theta) = \log L(\theta) = n\log\left(\frac{1}{\theta}\right) - \frac{1}{\theta}\sum x_i = -n\log\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

Derivamos a log-verossimilhança em relação a θ:

$$\frac{d\ell}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum x_i$$

Igualamos a derivada a zero:

$$-\frac{n}{\theta} + \frac{1}{\theta^2}\sum x_i = 0$$

Multiplicando ambos os lados por θ^2:

$$-n\theta + \sum x_i = 0 \Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Logo, no caso da maximização irrestrita, o EMV em Θ é $\hat{\theta}_n = \bar{X}_n$, e, portanto, obtemos:

$$L\left(\hat{\theta}_n|\mathbf{X}\right) = L\left(\bar{X}_n|\mathbf{X}\right) = \left(1/\bar{X}_n\right)^n \exp\left[-\left(\sum_{i=1}^{n} X_i\right)/\bar{X}_n\right].$$

Os valores que a função de verossimilhança assume sob H_0 e H_1 são dados por:

$$\text{Sob } H_1 : L\left(\hat{\theta}_n|\mathbf{X}\right) = \left(1/\bar{X}_n\right)^n \exp\left[-\left(\sum_{i=1}^{n} X_i\right)/\bar{X}_n\right] = \left(1/\bar{X}_n\right)^n e^{-n}.$$

Sob H_0: O EMV para θ é $\hat{\theta}_0 = \bar{X}_n$, se $\bar{X}_n < \theta_0$, e $\hat{\theta}_0 = \theta_0$ se $\bar{X}_n \geq \theta_0$. Desta forma,

$$L(\hat{\theta}_0|\mathbf{X}) = \begin{cases} \left(1/\bar{X}_n\right)^n \exp\left[-\left(\sum_{i=1}^n X_i\right)/\bar{X}_n\right] = \left(1/\bar{X}_n\right)^n e^{-n}, & \text{se } \bar{X}_n < \theta_0 \\ (1/\theta_0)^n \exp\left[-\left(\sum_{i=1}^n X_i\right)/\theta_0\right] = (1/\theta_0)^n \exp\left[-n\bar{X}_n/\theta_0,\right] & \text{se } \bar{X}_n \geq \theta_0. \end{cases}$$

Logo, a estatística do TRV é dada pela seguinte expressão:

$$\Lambda = \begin{cases} \frac{L(\bar{X}_n|\mathbf{X})}{L(\bar{X}_n|\mathbf{X})} = \frac{\left(1/\bar{X}_n\right)^n e^{-n}}{\left(1/\bar{X}_n\right)^n e^{-n}} = 1, & \text{se } \bar{X}_n < \theta_0 \\ \frac{L(\theta_0|\mathbf{X})}{L(\bar{X}_n|\mathbf{X})} = \frac{(1/\theta_0)^n \exp\left[-\sum_{i=1}^n X_i/\theta_0\right]}{\left(1/\bar{X}_n\right)^n \exp\left[-\sum_{i=1}^n X_i/\bar{X}_n\right]} = \left(\frac{\bar{X}_n}{\theta_0}\right)^n \exp\left[-\frac{n\bar{X}_n}{\theta_0} + \frac{n\bar{X}_n}{\bar{X}_n}\right] & \text{se } \bar{X}_n \geq \theta_0. \end{cases}$$

Observe que, se $\bar{X}_n < \theta_0$, a estatística de teste $\Lambda = 1$, ou seja, com probabilidade 1 a H_0 não será rejeitada. Ademais, para valores de $\bar{X}_n \geq \theta_0$ segue que,

$$\Lambda = \left(\frac{\bar{X}_n}{\theta_0}\right)^n \exp\left[-\frac{n\bar{X}_n}{\theta_0} + \frac{n\bar{X}_n}{\bar{X}_n}\right] = (e/\theta_0)^n \left(\bar{X}_n\right)^n \exp\left[-\frac{n\bar{X}_n}{\theta_0}\right].$$

Portanto, para algum $0 < k < 1$, rejeita-se H_0 à favor de H_1 se:

$$\Lambda \leq k \iff (e/\theta_0)^n \left(\bar{X}_n\right)^n \exp\left[-\frac{n\bar{X}_n}{\theta_0}\right] \leq k,$$

ou seja, se $\left[\bar{X}_n e^{-\frac{\bar{X}_n}{\theta_0}}\right]^n \leq k^* = k\,(e/\theta_0)^{-n}$. Assim, a região crítica será obtida isolando-se o $\bar{X}_n$ na expressão acima (técnicas avançadas devem ser consideradas para tal propósito).

Q9. Seja $X_1, X_2, \cdots, X_n$ uma amostra aleatória de tamanho n da distribuição $\mathbb{N}(\mu, 1)$. Considere as hipóteses $H_0 : \mu = 0$ contra $H_1 : \mu = 2$. Mostre que a região crítica para um teste *MP* para testar H_0 contra H_1 é dada por: $RC = \{\mathbf{x} : \sum_{i=1}^{n} x_i \geq c\}$.

Resposta: Como a amostra foi extraída de uma população normal, a fdp da variável aleatória X é dada por:

$$f(x|\mu) = (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\},$$

resultando na seguinte função de verossimilhança,

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

O próximo passo seria obter as funções de verossimilhanças restritas às duas hipóteses em questão. Considerando $H_0 : \mu = \mu' = 0$ e $H_0 : \mu = \mu'' = 2$, temos que:

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

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

Com base no resultado anterior, a estatística que define a razão das duas funções de verossimilhança é dada por:

$$R(\mathbf{x}) = \frac{L(\mu'|\mathbf{x})}{L(\mu''|\mathbf{x})} = \frac{(2\pi)^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\}}{(2\pi)^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (x_i-1)^2\right\}} = \exp\left\{-\frac{\sum_{i=1}^{n} x_i}{2} + \frac{n}{2}\right\}.$$

Para algum $k > 0$, rejeitamos H_0 a favor da H_1 se $R(\mathbf{x}) < k$. Resultando na seguinte região crítica:

$$RC = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i \geq c\right\},$$

em que $c = n - 2\log(k)$.

Q10. Seja $X_1, X_2, \cdots, X_n$ uma amostra aleatória de tamanho n proveniente da distribuição Poisson(θ). Encontre um teste *MP* para testar as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$, com $\theta_1 > \theta_0$.

Resposta: Como a amostra foi extraída de uma população de distribuição Poisson(θ), então remos que:

$$P(X = x \mid \theta) = \frac{e^{-\theta}\theta^x}{x!}, \quad x = 0, 1, 2, \ldots$$

tendo sua função de verossimilhança como sendo:

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{1}{x_i!}.$$

Deve-se obter as funções de verossimilhança restritas às duas hipóteses em questão, tendo:

$$R(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\theta_1|\mathbf{x})} = \frac{e^{-n\theta_0}\theta_0^{\sum_{i=1}^{n} x_i}\prod_{i=1}^{n}\frac{1}{x_i!}}{e^{-n\theta_1}\theta_1^{\sum_{i=1}^{n} x_i}\prod_{i=1}^{n}\frac{1}{x_i!}} = e^{-n(\theta_0-\theta_1)}\theta_0^{\sum_{i=1}^{n} x_i}\theta_1^{-\sum_{i=1}^{n} x_i}$$

Para algum $k > 0$, rejeitamos H_0 a favor de H_1 se $R(\mathbf{x}) < k$. Ou seja, se

$$e^{-n(\theta_0-\theta_1)}\theta_0^{\sum\limits_{i=1}^{n} x_i}\theta_1^{-\sum\limits_{i=1}^{n} x_i} < k \iff \left(\frac{\theta_0}{\theta_1}\right)^{\sum\limits_{i=1}^{n} x_i} < ke^{-n(\theta_1-\theta_0)}$$

$$\log\left(\frac{\theta_0}{\theta_1}\right)\sum_{i=1}^{n} x_i < ke^{-n(\theta_1-\theta_0)}. \text{ Como } \theta_1 > \theta_0 \implies \log\left(\frac{\theta_0}{\theta_1}\right) < 0, \text{ portanto,}$$

$$RC = \left\{ \mathbf{x} : \sum_{i=1}^{n} x_i > k^* \right\},$$

com $k^* = ke^{-n(\theta_1 - \theta_0)} \log \left(\frac{\theta_1}{\theta_0} \right)$.

Q11. Seja X uma única observação da função de densidade $f(x|\theta) = (2\theta x + 1 - \theta)$, para $0 < x < 1$ e $\theta > 0$. Queremos testar as seguintes hipóteses $H_0 : \theta = 1$ contra $H_1 : \theta = 2$.

(a) Obtenha o teste MP com nível de significância α.

Resposta: Para esse caso temos uma única variável X com função densidade de probabilidade dada por, $f(x|\theta) = (2\theta x + 1 - \theta)$. As hipóteses de interesse são dadas por, $H_0 : \theta = 1$ contra $H_1 : \theta = 2$. Nisso, a fdp avaliada nos pontos $\theta = 1$ (sob H_0) e $\theta = 2$ (sob H_1) são dadas por:

$$\begin{aligned} \text{Sob } H_0 : &\quad f_0(x|\theta = 1) = 2x\mathbf{1}_{(0,1)}(x) \\ \text{Sob } H_1 : &\quad f_1(x|\theta = 2) = (4x - 1)\,\mathbf{1}_{(0,1)}(x). \end{aligned}$$

Assim, a estatística de teste é dada por,

$$R(x) = \frac{f_0(x|\theta = 1)}{f_1(x|\theta = 2)} = \frac{2x}{4x - 1}.$$

Logo, para algum $k > 0$, a H_0 será rejeitada à favor da H_1 se $R(x) \leq k \iff \frac{2x}{4x-1} \leq k \iff x \geq \frac{k}{4k-2}$. Consequentemente, a região crítica será dada por: $RC = \{x : x \geq k^* = \frac{k}{4k-2}\}$.

(b) Se $\alpha = 0,05$ e $x = 0,8$ qual a sua conclusão?

Resposta: Do item (a) vimos que $RC = \{x : x > k^* = \frac{k}{4k-2}\}$, então, para valores de $x = 0,80$ e $\alpha = 0,05$ obtemos: $\alpha = \mathbb{P}(X > k^*|\theta = 1) = \int_{k^*}^{1} f_0(x|\theta = 1)\,dx = \int_{k^*}^{1} 2x dx = x^2\Big|_{k^*}^{1} = 1 - (k^*)^2$. Como $\alpha = 0,05$ segue que, $(k^*)^2 = 1 - 0,05 = 0,95$, tendo $k^* = \sqrt{0,95} = 0,975$. Logo, $k = \frac{k^*}{4k^*-2} = 0,513$, desta forma, a região crítica será dada por: $RC = \{x : x \geq 0,975\}$ e região de aceitação dada por $RA = \{x : x < 0,975\}$. Portanto, como o ponto $x = 0,80$ cai na região de aceitação, concluímos que não há evidências suficientes para rejeitar a H_0.