# Estatística SEFAZ-CE e PCDF

Igor Gameleiro e Rafaela Fonte

Simulado 01 - Junho 2021

---

Um certo teste para COVID-19 é considerado correto 95% das vezes: se uma pessoa tem a doença, o resultado do teste é positivo com probabilidade de 0.95, e se a pessoa não tem a doença, o resultado do teste é negativo com probabilidade 0.95. Uma pessoa sorteada aleatoriamente em uma determinada população tem probabilidade de 0.001 de ter a doença.

1. A probabilidade de uma pessoa escolhida de forma aleatória teste positivo é maior que 5%.

2. Dado que uma pessoa realizou o teste e o resultado foi positivo, a probabilidade desta pessoa realmente ter a doença é menor que 2%.

---

Um dado honesto de 6 faces é lançado 5 vezes de forma independente. Seja $X$ o número de vezes que o lançamento resulta em 2 ou 3. Com base nessas informações, julgue os itens seguintes.

3. $X$ é uma variável aleatória com distribuição de Bernoulli.

4. O valor numérico para $P(X = 2.5)$ é maior que 20% .

5. O valor numérico para $P(X = 1)$ é menor que 1/3.

---

Está sendo feito um planejamento para um pesquisa que deverá estimar a proporção de eleitores que apoiam uma proposta de lei de controle de drogras. A estimativa deve estar dentro de um margem de erro de 2% com 95% de confiança. Como não há nenhum conhecimento prévio sobre o proporção de pessoas que podem apoiar a lei, deve-se utilizar a estimativa mais conservadora. Considere que a amostragem será realizada com reposição.

Sabendo que $P(Z < 1.96) = 0,975$, em que Z representa a distribuição normal padrão e que $E[.]$ é o operador linear do cálculo do valor esperado, julgue os itens que se seguem, em relação a essa situação hipotética.

6. A estimativa pontual mais conservativa para a proporção é aquela que acarreta na maior variância, a qual é obtida para $\tilde{p} = 0.5$.

7. O valor mínimo de pessoas que devem ser incluídas na amostra é 2401 para que os requisitos sejam atendidos.

8. Se $X = 2Z^2 + 3$, então $E[X^2] < (E[X])^2$.

---

| | | | | | |
|---|---|---|---|---|---|
| 1.5 | 2.3 | 3.2 | 3.5 | 4.1 | 4.9 |

Considerando que o conjunto de dados represente uma realização de uma amostra aleatória simples de tamanho $n = 6$ retirada de uma população $X$, cuja distribuição de probabilidade é Uniforme no intervalo $[0, \beta]$, em que $\beta$ é o parâmetro desconhecido. Julgue os itens que se seguem.

9. A estimativa de máxima verossimilhança para o parâmetro $\beta$ é 3.25 .

10. O valor máximo amostral é uma estatística suficiente para a estimação do parâmetro $\beta$.

**Gabarito**

1. C
2. C

Seja $A$ o evento que determinada pessoa tem a doença e $B$ o evento que o resultado do teste é positivo.

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c) = 0.001 \cdot 0.95 + 0.999 \cdot 0.05 = 0.0509 > 5\%$$

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.001 \cdot 0.95}{0.0509} < 2\%$$

3. E

X é uma variável aleatória formada por 5 ensaios de Bernoulli, logo tem distribuição Binomial.

4. E

Dado que X é o número de vezes (inteiro), só pode assumir valores de 0 a 5 e $P(X = 2.5) = 0$ .

5. C

Para cada dado lançado, temos um ensaio de Bernoulli com probabilidade $2/6 = 1/3$ de obter um 2 ou 3. Assim, a variável aleatória X possui distribuição Binomial com parâmetros $n = 5$ e $p = 1/3$, então

$$P(X = 1) = \binom{5}{1} \cdot (1/3)^1 \cdot (2/3)^4 = \frac{8}{243} < \frac{8}{240} = 1/3$$

6. C

7. C

A margem de erro, ou raio do intervalo de confiança, desejado é $ME = 0.02$. Para o nível de confiança de 95% usamos a informação fornecida $Z_{score} = 1.96$. Dado que não temos conhecimento prévio sobre a proporção de apoiadores, usamos a estimativa conservativa $\tilde{p} = 0.5$.

$$ME = Z_{score} \cdot \sqrt{\frac{\tilde{p}(1 - \tilde{p}}{n}}$$

$$n = \left(\frac{Z_{score}}{ME}\right)^2 \cdot \tilde{p}(1 - \tilde{p}) = \left(\frac{1.96}{0.02}\right)^2 \cdot 0.5 \cdot (1 - 0.5) = 2401$$

8.E

Perceba que X é uma variável aleatória e a variância de uma variável aleatória é sempre maior ou igual a zero.Temos que $\mathrm{var}(X) = E[X^2] - (E[X])^2 \geq 0$, logo $E[X^2] \geq (E[X])^2$.

9. E
10. C

Sabemos que o estimador de máxima verossimilhança da distribuição uniforme é o valor máximo amostral e que todo estimador de máxima verossimilhança é uma estimativa suficiente. Assim, a questão 10 está correta e a questão 9 está errada, pois o valor máximo amostral é 4.9.

Estimadores é um assunto muito importante dentro de estatística inferencial e estimador de máxima verossimilhança é um *hot topic* sem dúvida alguma. Segue uma tabela para você adicionar ao seu resumo.

Tabela 1: Estimadores de máxima verossimilhança

| **Distribuição de X** | **Estimador MV** |
|---|---|
| Bernoulli(p) | $\tilde{p} = \bar{X}$ |
| Poisson($\lambda$) | $\tilde{\lambda} = \bar{X}$ |
| $N(\mu, \sigma^2)$ | $\tilde{\mu} = \bar{X}$ e $\tilde{\sigma}^2 = \frac{(n-1) \cdot S^2}{n}$ |
| $Exp(\lambda)$ | $\tilde{\lambda} = \frac{1}{\bar{X}}$ |
| $U[0, \beta]$ | $\tilde{\beta} = max\{X_1, ..X_n\}$ |

Para quem não lembra como calcular o estimador de máxima verossimilhança da distribuição uniforme, vou deixar demonstrado a seguir. Na prova da Polícia Federal de 2021 caiu uma questão sobre estimador de máxima verossimilhança que você resolve de maneira muito parecida. A chave para resolver é lembrar que queremos maximizar a função de verossimilhança e quando temos amostras independentes basta maximizar a função densidade de probabilidade conjunta que é o produto das funções de distribuição de probabilidade.

Seja $X$ uma variável aleatória uniformemente distribuída em $[0, \beta]$ e $X_1, ...X_n$ uma amostra aleatória independente e igualmente distribuída de X. Esta distribuição tem função densidade de probabilidade dada por:

$$f(x|\beta) = \begin{cases} \frac{1}{\beta} & ,\text{se } 0 \leq x \leq \beta \\ 0 & \text{caso contrário} \end{cases}$$

A função de verossimilhança é dada por:

$$L(\beta, X_1, ...X_n) = \prod_{i=1}^{n} f(X_i|\beta) = \frac{1}{\beta^n} \mathbb{1}_{\{X_1,..X_n \in [0,\beta]\}} = \frac{1}{\beta^n} \mathbb{1}_{max\{X_1,..X_n\} \leq \beta}$$

Dado que $\frac{1}{\beta^n}$ é uma função decrescente de $\beta$, a função de verossimilhança será maximizada para o menor valor possível de $\beta$. Este valor é o $\beta = max\{X_1, ...X_n\}$, de onde obtemos o estimador de máxima verossimilhança de $\beta$ sendo $\tilde{\beta} = max\{X_1, ..X_n\}$. Se a notação $\mathbb{1}_{max\{X_1,..X_n\} \leq \beta}$ não for familiar, retorne ao seu material e procure por função indicadora (geralmente fica perto do conteúdo de distribuição de Bernoulli).

Agora que você viu a demonstração matemática, tente refletir sobre o significado da escolha deste estimador. Devemos escolher o valor de $\beta$ que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a "mais provável".

Abraços e bons estudos!