ICMC – USP – SME122

Introdução à Inferência Estatística

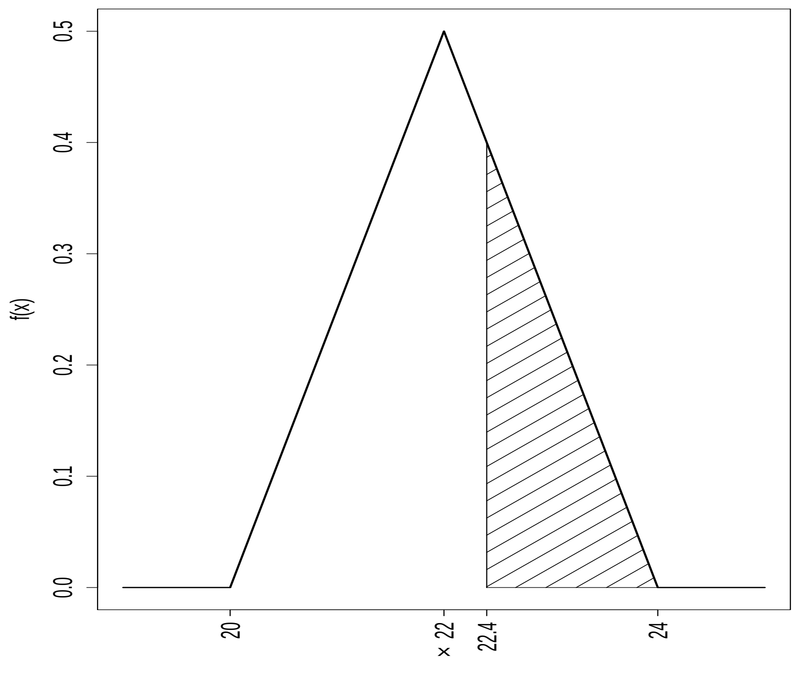
1ª PROVA – 2º/2007 – 18/10/2007

1. Um sistema computacional realiza determinada tarefa em um tempo (dado em segundos), que pode ser considerado uma variável aleaória (X) com função densidade de probabilidade

$$f(x) = \begin{cases} \frac{x}{4} - 5, & \text{se } 20 \leq x < 22, \\ 6 - \frac{x}{4}, & \text{se } 22 \leq x < 24, \\ 0, & \text{se } x \notin [20, 24]. \end{cases}$$

   (a) Em uma particular execução, qual a probabilidade de que o tempo ultrapasse 22,4 segundos?

   Solução. A probabilidade do evento $[X > 22,4]$, igual a $\int_{22,4}^{24} f(x)\,dx = \int_{22,4}^{24}(6 - \frac{x}{4})\,dx$, corresponde à área da região tracejada na figura abaixo e é igual a $P(X > 22,4) = 0,4\,(24 - 22,4)/2 = 0,32$.



   (b) Em 30 execuções, qual a probabilidade de o sistema demorar, em média, mais do que 22,4 segundos por execução?

   Solução. Apresentamos uma solução aproximada. De acordo com o teorema central do limite, a distribuição aproximada de $\overline{X} = \sum_{i=1}^{n} X_i/n$ é normal com média $\mu = E(X)$ e variância σ^2/n, sendo que $\sigma^2 = \text{var}(X)$ e $n = 30$. Calculamos

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_{20}^{22} x \left(\frac{x}{4} - 5\right) dx + \int_{22}^{24} x \left(6 - \frac{x}{4}\right) dx = 22,$$

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{20}^{22} x^2 \left(\frac{x}{4} - 5\right) dx + \int_{22}^{24} x^2 \left(6 - \frac{x}{4}\right) dx \\
&= \left(\frac{x^4}{16} - \frac{5x^3}{3}\right)\Bigg|_{20}^{22} + \left(2x^3 - \frac{x^4}{16}\right)\Bigg|_{22}^{24} = \frac{1454}{3}
\end{aligned}
$$

e $\sigma^2 = E(X^2) - \mu^2 = 2/3$, de modo que $\overline{X} \approx N(22, \frac{2}{90})$. Sendo assim, $Z = (\overline{X} - 22)/\sqrt{2/90} \approx N(0,1)$ e $P(\overline{X} > 22,4) \cong P(Z > (22,4 - 22)/\sqrt{2/90}) = P(Z > 2,68) = 0,0037$.

2. Uma amostra aleatória de 400 itens foi selecionada de um grande lote produzido durante certo período.

   (a) Foram detectados 20 itens que não satisfazem às especificações de qualidade. Apresente um intervalo de confiança de 95% para a proporção de itens que satisfazem às especificações.

   Solução. Como o lote é grande, podemos considerar que as $n = 400$ observações são independentes e têm a mesma distribuição de Bernoulli com parâmetro θ (igual à proporção de itens que satisfazem às especificações). De acordo com os dados coletados temos $\widehat{\theta} = (400 - 20)/400 = 0,95$. Com 95% de confiança temos $z_\gamma = 1,96$ e $E_{\max} = z_\gamma \sqrt{\widehat{\theta}(1 - \widehat{\theta})/n} = 0,0214$. Um intervalo de confiança aproximado para θ é dado por $\widehat{\theta} \mp E_{\max} = [0,929; 0,971]$.

   (b) Verifique se essa amostra é suficiente para a obtenção de um intervalo de 99% de confiança para a proporção de itens que satisfazem às especificações, com erro amostral máximo de 1,0%. Caso contrário, determine o tamanho amostral.

   Solução. Segundo o enunciado, $\gamma = 99\%$ (logo, $z_\gamma = 2,58$) e $E_{\max} = 0,01$. O tamanho amostral requerido pode ser calculado pela expressão $n = z_\gamma^2 \widehat{\theta}(1 - \widehat{\theta})/E_{\max}^2$. Utilizando a estimativa de θ obtida no item (2a) resulta em $n = 3162$, mostrando que uma amostra com $n = 400$ itens não atende às exigências do item (2b). Se optarmos pela solução conservadora (que neste problema não é adequada) resulta em $n = (\frac{z_\gamma}{2E_{\max}})^2 = 16641$.

3. Realizou-se uma pesquisa com um certo número de usuários de computadores, perguntando a eles quantas verificações de vírus fizeram nos últimos 15 dias, obtendo-se média e mediana iguais a 0,5 e 0, respectivamente. As afirmações abaixo são verdadeiras ou falsas?

   (a) Na amostra coletada a moda é igual a 0.

   (b) Mais da metade dos usuários pesquisados não fez verificação de vírus nos últimos 15 dias.

   (c) Nos últimos 30 dias os usuários efetuaram em média uma verificação.

   (d) Talvez pelo menos um usuário tenha feito mais de uma verificação nos últimos 15 dias.

SOLUÇÃO. Se n denota o número de usuários entrevistados, o número total de verificações é dado por $n \times$média $= n/2$, que é um número inteiro; portanto, n é par. Além disso, devemos ter $n \geq 4$, pois do contrário a mediana não seria igual a 0. Como a mediana é igual a 0, na amostra do número de verificações ordenada podemos garantir que pelo menos os $n/2+1$ primeiros valores são iguais a 0, significando que as afirmações dos itens (3a) e (3b) são verdadeiras. Supondo que o comportamento dos usuários nos 15 dias anteriores à pesquisa seja o mesmo do período mais recente, o número médio de verificações em 30 dias é igual a $2 \times 0,5 = 1$; logo, o item (3c) está correto. Conclui-se das respostas aos itens (3a) e (3b) que no máximo $n/2-1$ $(= n - n/2 - 1)$ usuários fizeram verificação, de forma que *certamente* pelo menos um usuário fez mais de uma verificação nos últimos 15 dias, pois do contrário a média não seria igual a $0,5$ (seria menor); a afirmação no item (3d) é falsa.

4. Os dados da tabela abaixo se referem a uma amostra aleatória da espessura (em mm) de um item produzido.

418	421	421	422	425	427	431
434	437	439	446	447	448	453
454	463	465				

(a) Apresente um intervalo de confiança de 95% para a espessura média. Que suposição fundamental foi utilizada na obtenção deste intervalo?

SOLUÇÃO. Denotamos por X a espessura. Inicialmente, com base na amostra de $n = 17$ observações, estimamos a média e o desvio padrão de X: $\overline{X} = \sum_{i=1}^{n} X_i/n = 438,29$ mm e $s = \{\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)\}^{1/2} = 15,14$ mm. Com 16 graus de liberdade e 95% de confiança, $t_\gamma = 2,120$. Conseqüentemente, $E_{\max} = t_\gamma \, s/\sqrt{n} = 7,78$ mm. Um intervalo de confiança aproximado para a espessura média é dado por $\overline{X} \mp E_{\max} = [430,51; 446,07]$, em mm. Nesta solução supomos que a distribuição da média amostral $(\overline{X})$ é aproximadamente normal.

(b) O intervalo do item anterior sugere que 440 mm seja um valor plausível para a espessura média?

SOLUÇÃO. Notando que o intervalo de 95% de confiança apresentado no item (4a) contém 440, a resposta é sim.

5. Uma amostra aleatória $X_1, \ldots, X_n$ é obtida de uma população tal que $X \sim N(\mu, \sigma^2)$. Pode ser provado que a distribuição aproximada da mediana amostral é normal com média μ e variância $\frac{\pi\sigma^2}{2n}$. Pretendemos estimar o parâmetro μ. Você utilizaria a média amostral ou a mediana amostral?

SOLUÇÃO. Como $X \sim N(\mu, \sigma^2)$, segue que $\overline{X} = \sum_{i=1}^{n} X_i/n \sim N(\mu, \frac{\sigma^2}{n})$. Assim, os dois estimadores têm a mesma média μ (são não viesados) e como $\lim_{n\to\infty} \frac{\sigma^2}{n} = \lim_{n\to\infty} \frac{\pi\sigma^2}{2n} = 0$, os dois estimadores são consistentes. Serão comparados pelas suas variâncias (estas coincidem com os erros quadráticos médios, pois os estimadores são não viesados). Notando que $\frac{\pi\sigma^2}{2n} / \frac{\sigma^2}{n} = \pi/2 > 1$, optamos pela média amostral por ser um estimador mais preciso (menor variância).