MAE 229 - Introdução à Probabilidade e Estatística II
Resolução Lista 2

Professor: Pedro Morettin

Exercício 1

Denotando por Y a variável aleatória que representa o comprimento dos cilindros de aço, temos que $Y \approx N(3,25; 0,0008)$. O comprimento dos dois cilindros justapostos é $Y_1 + Y_2$. Note que $\bar{Y} = \frac{Y_1+Y_2}{2} \sim N(3,25; 0.0004)$. Portanto, $P(Y_1 + Y_2 < 6,55) = P(\bar{Y} < 3,275) = P\left(Z < \frac{3,275-3,25}{\sqrt{0,0004}}\right) = P(Z < 1,25) = 89,44\%$.

Exercício 2

Denotando por X a variável aleatória que representa a vida de uma lâmpada em dias, temos que $X \sim N(50, 19^2)$. Considerando a colocação de uma lâmpada no dia 01 de janeiro, após 31 dias (dia 01 de fevereiro), teríamos $P(X < 31) = P\left(Z < \dfrac{31-50}{19}\right) = 15,86\%$. Logo, espera-se que $(15,86\%).5000 \approx 794$ lâmpadas devam ser substituídas neste dia.

Exercício 3

Pelo TLC, para n suficientemente grande, temos que $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. Logo, considerando X_i o peso do i-ésimo indivíduo em libras, temos que para $n = 50$, $\mu = 150$ e $\sigma = 25$, $P(X_1 + \cdots + X_{50} > 7800) = P(\bar{X} > 7800/50) = P\left(\frac{\bar{X}-150}{25/\sqrt{50}} > \frac{156-150}{25/\sqrt{50}}\right) = P(Z > 1,697) = 4,48\%$.
Para achar o aumento de capacidade necessário, temos que achar a capacidade C tal que $P(X_1 + \cdots + X_{50} > C) = 1\%$. Ou seja, $P\left(Z > \frac{(C/50)-150}{25/\sqrt{50}}\right) = 1\% \Rightarrow \frac{(C/50)-150}{25/\sqrt{50}} = 2,326 \Rightarrow C = 7911,244$. Portanto, o aumento de capacidade teria que ser de aproximadamente 111 libras.

Exercício 4

(a) Sim, pois $E(X) = 8 \cdot 0,25 + 10 \cdot 0,25 + 11 \cdot 0,5 = 10$, que é o verdadeiro valor para o comprimento.

(b) Sim, pois $E(Y) = 0,5 \cdot 4 + 0,5 \cdot 6 = 5$.

(c) Combinando todos os possíveis pares de largura e comprimento, e assumindo independência entre as variáveis aleatórias X e Y, chega-se à seguinte distribuição de probabilidade para a variável aleatória $A = XY$:

Distribuição de probabilidade para A

A	32	40	44	48	60	66
$P(A = a)$	0.125	0.125	0.25	0.125	0.125	0.25

Assim, como $E(A) = 50$, temos que $A = XY$ é um estimador não viesado para a verdadeira área. Com efeito, pela independência temos que $E(XY) = E(X) \cdot E(Y)$, e como ambos são estimadores não viesados individualmente, então o produto também é.

(d) $Var(A) = E(A^2) - E(A) = 2639 - 2500 = 139$.

Exercício 5

(a) Todos são não viesados, note que a soma dos pesos de $\hat{\mu_1}$ e $\hat{\mu_2}$ é igual a 1 em todos os casos, de modo que a esperança é sempre μ.

(b) O estimador mais eficiente é aquele que possui a menor variância. Considerando que $\hat{\sigma_1} = DesvPad(\hat{\mu_1}) = 5 \cdot DesvPad(\hat{\mu_2}) = 5 \cdot \hat{\sigma_2}$, temos:

$Var(\hat{W_1}) = (1/2)^2 \cdot \hat{\sigma_1}^2 + (1/2)^2 \cdot \hat{\sigma_2}^2 = 1/4 \cdot \hat{\sigma_1}^2 + 1/4 \cdot \hat{\sigma_2}^2 = (25/4 + 1/4) \cdot \hat{\sigma_2}^2$
$= 6,5 \cdot \hat{\sigma_2}^2$

$Var(\hat{W_2}) = (4/5)^2 \cdot \hat{\sigma_1}^2 + (1/5)^2 \cdot \hat{\sigma_2}^2 = 16/25 \cdot \hat{\sigma_1}^2 + 1/25 \cdot \hat{\sigma_2}^2 = (16 + 1/25) \cdot \hat{\sigma_2}^2$
$= 16,04 \cdot \hat{\sigma_2}^2$

$Var(\hat{W_3}) = (5/6)^2 \cdot \hat{\sigma_1}^2 + (1/6)^2 \cdot \hat{\sigma_2}^2 = (25/36) \cdot \hat{\sigma_1}^2 + (1/36) \cdot \hat{\sigma_2}^2 = (625/36 + 1/36) \cdot \hat{\sigma_2}^2$
$= 17,38 \cdot \hat{\sigma_2}^2$

$Var(\hat{W_4}) = \hat{\sigma_1}^2 = 25 \cdot \hat{\sigma_2}^2$

Assim, os estimadores do mais eficiente para o menos eficiente são, respectivamente: $\hat{W_1}$,$\hat{W_2}$,$\hat{W_3}$ e $\hat{W_4}$.

(c) Sejam w_1 o peso de $\hat{\mu_1}$ e w_2 o peso de $\hat{\mu_2}$. O estimador $\hat{W_0} = w_1 \cdot \hat{\mu_1} + w_2 \cdot \hat{\mu_2}$ possui menor variância quanto menor for w_1, uma vez que a variância de $\hat{\mu_1}$ é 25 vezes a variância de $\hat{\mu_2}$. Logo, a combinação mais eficiente possível ocorre quando $w_1 = 0$. Neste caso, $\hat{W_0} = \hat{\mu_2}$ e $Var(\hat{W_0}) = Var(\hat{\mu_2}) = \hat{\sigma_2}^2$.

Exercício 6

(a) O viés V de um estimador T é a diferença entre $E(T)$ e o verdadeiro valor do parâmetro a ser estimado. Suponha que valor verdadeiro para a área seja A. Portanto, no caso (1), temos:
$V_1 = E\{(\frac{X_1+X_2}{2})^2\} - A$.

Note que $E(\bar{X}^2) = Var(\bar{X}) + \{E(\bar{X})\}^2 = \frac{\sigma^2}{n} + \mu^2$, logo $V_1 = \frac{\sigma^2}{2} + \mu^2 - A$.
Já no caso (2), temos:
$V_2 = E(\frac{X_1^2+X_2^2}{2}) - A = \frac{E(X_1^2)}{2} + \frac{E(X_2^2)}{2} - A = \frac{Var(X_1)+\{E(X_1)\}^2}{2} + \frac{Var(X_2)+\{E(X_2)\}^2}{2} - A = 2 \cdot \frac{\sigma^2+\mu^2}{2} - A = \sigma^2 + \mu^2 - A$.
Sendo assim, a opção (1) tem menor viés.

(b) Da mesma forma o estimador (1) é $\bar{X}^2$ e o estimador (2) é $\frac{\sum_{i=1}^n X_i{}^2}{n}$.
Portanto, o viés de (1) é $V_1 = E(\bar{X}^2) - A = \frac{\sigma^2}{n} + \mu^2 - A$.
Já o viés de (2) é $V_2 = \frac{\sum_{i=1}^n E(X_i{}^2)}{n} - A = E(X^2) - A = \sigma^2 + \mu^2 - A$.
Assim, temos que a opção (1) tem menor viés.

(c) Como X_1 e X_2 são observações independentes, $V = E(X_1 \cdot X_2) - A = E(X_1) \cdot E(X_2) - A = \mu^2 - A < V_1 < V_2$.

Exercício 7

Denotando por X_i o consumo da i-ésima família (todas com mesmo padrão de renda, ou seja, identicamente distribuídas), temos para $n = 25$, $\bar{X} = 8,2$. Queremos construir um IC de 95 % para $E(X_i) = \mu$, considerando que $\sigma = 0,72$. Pelo TLC, temos para n suficientemente grande que $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, logo $P(-z < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z) = 95\% \Rightarrow z = 1,96$.

Assim, temos que $P\left(\bar{X} - \frac{1{,}96\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{1{,}96\sigma}{\sqrt{n}}\right) = 95\%$. Substituindo pelos valores mencionados para $\bar{X}$ e σ, chegamos ao intervalo [7,917 ; 8,482].

Exercício 8

Denotando por X_i o tempo de reação do i-ésimo motorista em segundos, temos $\bar{X} = 0,83$, com $n = 30$. Outro dado do problema é que o desvio padrão amostral $S = 0,2$. Usaremos a aproximação normal para calcular um IC de 95% para μ como no exercício anterior, utilizando como aproximação para $Var(\bar{X}) = \frac{S^2}{n}$. Veremos, mais adiante, que na realidade $\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$.

Assim, o IC de 95% para μ é $\left[\bar{X} - \frac{1{,}96S}{\sqrt{n}}; \bar{X} + \frac{1{,}96S}{\sqrt{n}}\right]$. Substituindo pelos valores de $\bar{X}$, n e S, temos que o IC é [0,758 ; 0,901].

Exercício 9

Para essa amostra com $n = 40$ temos que $\bar{X} = 66$ e $\sigma = 11,83$. Assim, vamos usar a aproximação normal tal que $\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$ para n suficientemente grande. Precisamos encontrar z tal que $P\left(-z < \sqrt{n}\frac{\bar{X}-\mu}{\sigma} < z\right) = P(-z < Z < z) = 90\%$. Verificando a tabela da normal padrão, temos que $z = 1,645$. Assim, o intervalo de confiança para μ é $[\bar{X} - \frac{z\sigma}{\sqrt{n}}; \bar{X} + \frac{z\sigma}{\sqrt{n}}] = [62,923; 69,076]$.

Exercício 10

Seja X o tempo de execução do serviço em minutos da primeira fábrica e Y o tempo de execução do mesmo serviço em minutos da segunda fábrica. Seja $n = 120$ o tamanho de amostra da primeira fábrica, e m o tamanho de amostra da segunda fábrica. Assume-se que para n suficientemente grande $\bar{X} \sim N(22, 4/120)$ e $\bar{Y} \sim N(19, 10/120)$. Assim, assumindo independência entre X e Y, temos que $(\bar{X} - \bar{Y}) \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X{}^2}{n} + \frac{\sigma_Y{}^2}{m}\right) = N(3, 14/120)$. Portanto, um intervalo de confiança de 95% para a diferença $\mu_X - \mu_Y$ é $\left[(\bar{X} - \bar{Y}) - 1,96\sqrt{\frac{\sigma_X{}^2}{n} + \frac{\sigma_Y{}^2}{m}}; (\bar{X} - \bar{Y}) + 1,96\sqrt{\frac{\sigma_X{}^2}{n} + \frac{\sigma_Y{}^2}{m}}\right] = [2,331; 3,669]$.

Exercício 11

(a) Para encontrar o estimador de mínimos quadrados precisamos achar o ponto de mínimo de $f(\mu) = \sum_{i=1}^{n}(y_i - \mu)^2$. Derivando, obtemos $f'(\mu) = 2\sum_{i=1}^{n}(\mu - y_i) = 2n\mu - 2n\bar{y} = 2n(\mu - \bar{y})$.
Portanto, o ponto crítico é $\mu = \bar{y}$. Como $f''(\bar{y}) = 1 > 0$, $\bar{y}$ é ponto de mínimo, sendo assim o estimador de mínimos quadrados de μ.

(b) Com os valores do enunciado, temos que $\bar{y} = \frac{3+5+6+8+16}{5} = 7,6$.

Exercício 12

(a) O erro amostral do estimador $\hat{p} = X/n$ (sendo X o número de sucessos em n tentativas) é a diferença $e = \hat{p} - p$, em que p é o verdadeiro valor do parâmetro (no caso a proporção populacional dos eleitores favoráveis a um determinado candidato). Portanto, seja X o número de eleitores favoráveis em uma amostra de tamanho n. Sabemos que $X \sim Bin(n, p)$ e usando o TLC, temos que $Z = \frac{X-np}{\sqrt{np(1-p)}} \sim N(0, 1)$ ou ainda $\frac{(\hat{p}-p)\sqrt{n}}{\sqrt{p(1-p)}} \sim N(0, 1)$.
Portanto, considerando um erro absoluto de no máximo 0,01, temos $P(|e| \leqslant 0,01) = P(-0,01 \leqslant e \leqslant 0,01) = P\left(\frac{-0,01\sqrt{n}}{\sqrt{p(1-p)}} \leqslant \frac{e\sqrt{n}}{\sqrt{p(1-p)}} \leqslant \frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right)$.

Para o cálculo da $Var(\hat{p}) = \frac{p(1-p)}{n}$, podemos usar $p = 0,6$, conforme obtido na amostra piloto. Assim, calculamos n tal que $P\left(\frac{-0,01\sqrt{n}}{0,489} \leqslant Z \leqslant \frac{-0,01\sqrt{n}}{0,489}\right) = P(-0,0204\sqrt{n} \leqslant Z \leqslant 0,0204\sqrt{n}) = 80\%$. Logo, $n \approx 3942$.

(b) Sabemos que $P(-1,96 \leqslant Z \leqslant 1,96) = 95\%$, logo usando a aproximação utilizada em (a), temos que $P\left(-1,96 \leqslant \frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}} \leqslant 1,96\right) = 95\%$.

Assim, usando novamente que $p = 0,6$ e dado que $\hat{p} = 0,55$ e $n = 3942$, o IC de 95% para a proporção populacional p é $\left[\hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}}; \hat{p} + 1,96\sqrt{\frac{p(1-p)}{n}}\right] = [0,534; 0,565]$.

Exercício 13

(a) Como visto, o IC de 95% para a proporção populacional p é $\left[\hat{p} - 1,96\sqrt{\frac{p(1-p)}{n}}; \hat{p} + 1,96\sqrt{\frac{p(1-p)}{n}}\right]$.
Conforme enunciado, $\hat{p} = 1/3$ e $n = 300$, porém não temos amostra piloto. Neste caso, podemos usar o fato que $p(1-p) \leqslant 1/4$. Assim, usamos que o IC é $\left[\hat{p} - \frac{1,96}{4n}; \hat{p} + 1,96\frac{1,96}{4n}\right] =$

$[0,276;0,389]$. Podemos interpretar esse resultado da seguinte maneira: com 95% de confiança o intervalo encontrado contém o verdadeiro valor do parâmetro.

(b) Pode-se demonstrar que $n = \frac{z_\gamma^2 p(1-p)}{e^2} \approx \frac{z_\gamma^2}{4e^2}$, em que z_γ é tal que $P(-z_\gamma \leqslant Z \leqslant z_\gamma) = \gamma$ e $e = (\hat{p} - p)$ é o erro de estimação (note que utilizou-se novamente que $p(1-p) \leqslant 1/4$ para aproximar). Usando esta fórmula, com $z_\gamma = 1,96$, temos que $n \approx 2401$. Isso quer dizer que se o tamanho da amostra for maior ou igual a 2401 o erro de estimação $\hat{p} - p$ será no máximo 0,02 em valor absoluto com 95% de confiança.

Exercício 14

Para encontrar os estimadores de mínimos quadrados de α e β, precisamos minimizar a soma $\sum_{i=1}^{n}(y_i - \alpha - \beta {x_i}^2)^2 = f(\alpha, \beta)$. Derivando em relação à α e igualando a zero, temos:

$$\frac{\partial}{\partial \alpha} f(\alpha, \beta) = 2(-1)\sum_{i=1}^{n}(y_i - \alpha - \beta {x_i}^2) = 0 \tag{1}$$

$$\therefore \hat{\alpha} = \frac{n\bar{y} - \beta \sum_{i=1}^{n} {x_i}^2}{n}. \tag{2}$$

Derivando em relação à β e igualando a zero, temos:

$$\frac{\partial}{\partial \beta} f(\alpha, \beta) = 2(-1)\sum_{i=1}^{n}(y_i - \alpha - \beta {x_i}^2)({x_i}^2) = 0 \tag{3}$$

$$\therefore \hat{\beta} = \frac{\sum_{i=1}^{n} y_i {x_i}^2 - \alpha \sum_{i=1}^{n} {x_i}^2}{\sum_{i=1}^{n} {x_i}^4}. \tag{4}$$

Substituindo o valor de $\alpha = \hat{\alpha}$ encontrado em (2) na equação (4), chegamos a

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} y_i {x_i}^2 - n\bar{y}\sum_{i=1}^{n} {x_i}^2}{n\sum_{i=1}^{n} {x_i}^4 - (\sum_{i=1}^{n} {x_i}^2)^2}.$$

O valor de $\hat{\alpha}$ pode ser encontrado substituindo na equação (2), o valor de $\hat{\beta}$.

Exercício 15

(a) Neste caso, $\hat{p} = 0,6$, e usando a fórmulas do Exercício 13, temos que o IC de 95% é [0,543;0,656].

(b) Considerando um erro absoluto de no máximo 0,001, temos $P(|e| \leqslant 0,001) = P(-0,001 \leqslant e \leqslant 0,001) = P\left(\frac{-0{,}001\sqrt{n}}{\sqrt{p(1-p)}} \leqslant \frac{e\sqrt{n}}{\sqrt{p(1-p)}} \leqslant \frac{0{,}001\sqrt{n}}{\sqrt{p(1-p)}}\right)$. Como não conhecemos p, podemos usar o fato que $p(1-p) \leqslant 1/4$.
Assim, para $n = 300$ temos $P\left(\frac{-0{,}001\sqrt{300}}{\sqrt{1/4}} \leqslant Z \leqslant \frac{-0{,}001\sqrt{300}}{\sqrt{1/4}}\right) = P(-0,0346 \leqslant Z \leqslant 0,0346) = 2,76\%$.

(c) Podemos achar um tamanho de amostra $n \approx \frac{z_\gamma^2}{4e^2}$, em que z_γ é tal que $P(-z_\gamma \leqslant Z \leqslant z_\gamma) = \gamma$ e $e = (\hat{p} - p) = 0,0005$ é o erro de estimação. Para $\gamma = 95\%$, temos $z_\gamma = 1,96$. Assim, $n \approx 3.841.459$. Como este tamanho de amostra é muito alto o custo tende a ser inviável.

Exercício 16

Suponha que tenhamos uma amostra $x = x_1, \ldots, x_n$ de $X \sim$Poi(λ), então a log-verossimilhança é dada por

$$\begin{aligned} l(\lambda \mid X) &= \log\left(\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right) \\ &= \log\left(\frac{e^{-n\lambda}\lambda^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}\right) \\ &= -n\lambda + n\bar{x}\log(\lambda) + \sum_{i=1}^{n}\log(x_i!) \end{aligned}$$

Derivando obtemos:

$$\frac{\partial}{\partial\lambda} l(\lambda \mid x) = -n + \frac{n\bar{x}}{\lambda},$$

Igualando a zero, chegamos ao estimador de máxima verossimilhança:

$$\hat{\lambda}_{MV} = \bar{X}.$$

Note que $\frac{\partial^2}{\partial\lambda^2} l(\lambda \mid x) = -\frac{1}{\lambda^2} n\bar{x} < 0$. Logo, $\bar{x}$ é ponto de máximo.

Exercício 17

Precisamos achar o estimador de mínimos quadrados para o modelo $y_i = \alpha + \beta x_i$ com base numa amostra de n pares com $(x_1, y_1), \ldots, (x_n, y_n)$. Então, precisamos encontrar o ponto de mínimo de $f(\alpha, \beta) = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$. Derivando em relação a α, obtemos

$$\begin{aligned} \frac{\partial}{\partial\alpha} f(\alpha, \beta) &= -2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) \\ &= 2n\alpha + 2n\beta\bar{x} - 2n\bar{y} \\ &= 2n(\alpha + \beta\bar{x} - \bar{y}) \end{aligned}$$

Derivando em relação a β, temos que

$$\frac{\partial}{\partial\beta} f(\alpha, \beta) = 2n\alpha\bar{x} + 2\beta\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i y_i$$

Igualando à zero as duas equações anteriores, chegamos, portanto, ao seguinte sistema linear

$$\begin{cases} \alpha + \beta\bar{x} & = \bar{y} \\ n\bar{x}\alpha + \beta\sum_{i=1}^{n} x_i^2 & = \sum_{i=1}^{n} x_i y_i. \end{cases}$$

A solução do sistema linear, é dada por

$$\begin{aligned}\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}\\ \hat{\beta} &= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}.\end{aligned}$$

Usando os dados do enunciando e realizando os cálculos chegamos a $\hat{\alpha} = 4$ e $\hat{\beta} = 3$.

Exercício 18

Como já visto, o intervalo de confiança de $\gamma\%$ para p é $[\hat{p} - \frac{z_\gamma}{\sqrt{4n}}; \hat{p} + \frac{z_\gamma}{\sqrt{4n}}]$. Substituindo na fórmula os valores de $\hat{p} = 0,6$, $n = 100$ e dado que a amplitude A do intervalo é 0,09, temos que $A = \frac{2z_\gamma}{20} = 0,09$. Segue que $z_\gamma = 0,9$ é tal que $P(-0,9 \leqslant Z \leqslant 0,9) = 63,19\% = \gamma$.

Exercício 19

O intervalo de confiança de 95% para μ é $[\bar{X} - \frac{1{,}96\sigma}{\sqrt{n}}; \bar{X} + \frac{1{,}96\sigma}{\sqrt{n}}]$. Substituindo pelos valores de $\bar{X} = 510,6$, $n = 100$ e $\sigma = 4$, temos que IC=[509,816 ; 511,384].