# MAE 229 - Introdução à Probabilidade e Estatística II
# Resolução Lista 1

Professor: Pedro Morettin

Exercício 1

(a) Fazer histograma usando os seguintes dados:

Distribuição de probabilidade da variável $X$:

| $X$ | 1 | 3 | 5 | 7 |
|---|---|---|---|---|
| P$(X = x)$ | 1/5 | 1/5 | 2/5 | 1/5 |

(b) $E(X) = (1/5) \cdot 1 + (1/5) \cdot 3 + (2/5) \cdot 5 + (1/5) \cdot 7 = 21/5 = 4,2$

$$Var(X) = E(X^2) - [E(X)]^2 \quad = (1/5) \cdot 1^2 + (1/5) \cdot 3^2 + (2/5) \cdot 5^2 + (1/5) \cdot 7^2 - (21/5)^2 = 4,16$$

Para calcular a $Md(X)$ usamos a seguinte definição: $P(X \leqslant Md(X)) \geqslant 1/2$ e $P(X \geqslant Md(X)) \geqslant 1/2$. Portanto, $Md(X) = 5$, pois $P(X \leqslant 5) = 4/5 > 1/2$ e $P(X \geqslant 5) = 3/5 > 1/2$.

(c) (A,A);(A,B);(A,C);(A,D);(A,E);
(B,A);(B,B);(B,C);(B,D);(B,E);
(C,A);(C,B);(C,C);(C,D);(C,E);
(D,A);(D,B);(D,C);(D,D);(D,E);
(E,A);(E,B);(E,C);(E,D);(E,E).

(d) Para cada par enunciado em (c) temos os seguintes tempos médios de serviço:
1;2;3;3;4;
2;3;4;4;5;
3;4;5;5;6;
3;4;5;5;6;
4;5;6;6;7.

Distribuição de probabilidade da variável $\bar{X}$

| $\bar{X}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $P(\bar{X} = \bar{x})$ | 1/25 | 2/25 | 1/5 | 6/25 | 6/25 | 4/25 | 1/25 |

Fazer histograma e comparar com letra (a).

(e) $E(\bar{X}) = (1/25)\cdot 1 + (2/25)\cdot 2 + (1/5)\cdot 3 + (6/25)\cdot 4 + (6/25)\cdot 5 + (4/25)\cdot 6 + (1/25)\cdot 7 = 4,2$

$$\begin{aligned} Var(\bar{X}) = E(X^2) - [E(X)]^2 &= (1/25)\cdot 1^2 + (2/25)\cdot 2^2 + (1/5)\cdot 3^2 \\ &+ (6/25)\cdot 4^2 + (6/25)\cdot 5^2 + (4/25)\cdot 6^2 + (1/25).7^2 - (4,2)^2 \\ &= 2,08 \end{aligned}$$

Da mesma forma que na letra (b), temos que $Med(\bar{X}) = 4$, pois $P(\bar{X} \leqslant 4) = 14/25 > 1/2$ e $P(\bar{X} \geqslant 4) = 19/25 > 1/2$.

Comparando estes resultados com a letra (b), temos que a esperanças são iguais, porém a mediana e variância mudam. Note que $Var(\bar{X}) = Var(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{Var(\sum_{i=1}^{n} X_i)}{n^2} = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \frac{Var(X_i)}{n}$. Como $n = 2$ neste caso, temos que $Var(\bar{X})$ é metade de $Var(X)$.

(f) No item (c) mostramos todos os pares de funcionários possíveis. Os pares de tempos de serviços correspondentes a estes funcionários são, respectivamente:

(1,1);(1,3);(1,5);(1,5);(1,7);
(3,1);(3,3);(3,5);(3,5);(3,7);
(5,1);(5,3);(5,5);(5,5);(5,7);
(5,1);(5,3);(5,5);(5,5);(5,7);
(7,1);(7,3);(7,5);(7,5);(7,7).

Sendo assim, as respectivas estatísticas $S^2$ são:

0;2;8;8;18;
2;0;2;2;8;
8;2;0;0;2;
8;2;0;0;2;
18;8;2;2;0.

Distribuição amostral da estatística $S^2$

| $S^2$ | 0 | 2 | 8 | 18 |
|---|---|---|---|---|
| $P(S^2 = s)$ | 7/25 | 2/5 | 6/25 | 2/25 |

Fazer histograma correspondente.

(g) $E(S^2) = (7/25)\cdot 0 + (2/5)\cdot 2 + (6/25)\cdot 8 + (2/25)\cdot 18 = 4,16$

$$\begin{aligned} Var(S^2) &= E[(S^2)^2] - [E(S^2)]^2 = (7/25)\cdot 0^2 + (2/5)\cdot 2^2 + (6/25)\cdot 8^2 + (2/25)\cdot 18^2 - 4,16^2 \\ &= 25,5744 \end{aligned}$$

(h) Com base na letra (f), podemos calcular as respectivas frequencias dos pares ($X_1$,$X_2$):

Distribuição conjunta de $(X_1,X_2)$ e respectivas marginais

| $X_1 \backslash X_2$ | 1 | 3 | 5 | 7 | Total |
|---|---|---|---|---|---|
| 1 | 1/25 | 1/25 | 2/25 | 1/25 | 1/5 |
| 3 | 1/25 | 1/25 | 2/25 | 1/25 | 1/5 |
| 5 | 2/25 | 2/25 | 4/25 | 2/25 | 2/5 |
| 7 | 1/25 | 1/25 | 2/25 | 1/25 | 1/5 |
| Total | 1/5 | 1/5 | 2/5 | 1/5 | 1 |

(i) Sim, são independentes e identicamente distribuídas (note que os Totais na tabela acima equivalem às marginais). Para mostrar que são independentes basta provar que $P(X_1 = x_1) \cdot P(X_2 = x_2) = P[(X_1, X_2) = (x_1, x_2)]$ para todos os pares. Por exemplo, $P(X_1 = 1) \cdot P(X_2 = 1) = P[(X_1, X_2) = (1, 1)]$. De fato, $(1/5) \cdot (1/5) = 1/25$.

(j) Respondido no item anterior.

(k) Podem ser formadas $5^3 = 125$ triplas.

(l) O gráfico com os valores de $\bar{X}$ devem ficar com em torno de $E(X) = 4.2$. Além disso, o valor de $Var(\bar{X})$ deve ser igual a $Var(X)/3 = 1,386$.

(m) Da mesma forma que antes, teríamos que $E(S^2) = Var(X) = 4,16$, uma vez que $S^2$ é um estimador não viesado para a variância populacional. A variância de $S^2$, no entanto, será menor do que com $n = 2$.

(n) Também seriam independentes e identicamente distribuídas.

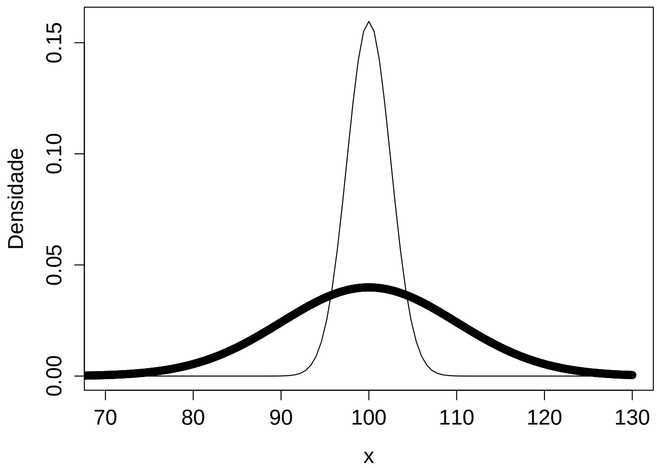
Exercício 2

(a) $P(90 \leqslant X \leqslant 110) = P(X \leqslant 110) - P(X \leqslant 90) = 0,6827$

(b) Como $X \sim N(100, 100)$, temos que $\bar{X} \sim N(100, 100/16)$, logo:

$$\begin{aligned} P(90 \leqslant \bar{X} \leqslant 110) &= P(\bar{X} \leqslant 110) - P(\bar{X} \leqslant 90) \\ &= P(Z \leqslant \frac{110 - 100}{2,5}) - P(Z \leqslant \frac{90 - 100}{2,5}) \\ &= P(Z \leqslant 4) - P(Z \leqslant -4) = 0,9999 \end{aligned}$$

(c) Temos que $E(X) = E(\bar{X})$, logo a distribuição de $\bar{X}$ tem a mesma média, porém é menos dispersa, uma vez que $Var(\bar{X}) = \frac{\sigma^2}{n}$.




A linha mais fina representa o gráfico de $\bar{X}$ e a linha mais grossa representa o gráfico de $X$.

(d) Vamos determinar $n$ tal que $P(90 \leqslant \bar{X} \leqslant 110) = 95\%$, i.e.,

$$P(\frac{90-100}{10/\sqrt{n}} \leqslant Z \leqslant \frac{110-100}{10/\sqrt{n}}) = 95\%$$
$$\therefore P(-\sqrt{n} \leqslant Z \leqslant \sqrt{n}) = 95\%$$
$$\therefore 2.P(0 \leqslant Z \leqslant \sqrt{n}) = 95\%$$
$$\therefore P(0 \leqslant Z \leqslant \sqrt{n}) = 47,5\%$$

Assim, temos que $\sqrt{n} = 1,96$, tal que $n = 1,96^2 = 3,84$. Como $n$ é inteiro, basta que $n = 4$ para que $\bar{X}$ esteja dentro do intervalo (90,110) com pelo menos 95% de confiança. Note que a probabilidade encontrada na letra (a) é bem superior a 95%, pois $n = 16 \gg 4$.

Exercício 3

(a) Seja $X_i$ a variável aleatória que toma o valor 1 se a peça é defeituosa, e o valor 0, caso contrário. Seja $Y = \sum_{i=1}^{n} X_i$ o número de peças defeituosas em uma amostra de tamanho $n$. Temos que $Y \sim Bin(n,p)$, em que $p = 20\%$ é a proporção real de peças defeituosas e $n = 8$ neste caso. Sabemos que $\hat{p} = Y/n$ é a proporção de peças defeituosas nesta amostra. Sendo assim, $P(Y = k) = P(Y/8 = k/8) = P(\hat{p} = k/8)$, o que nos permite usar a tabua da distribuição binomial para calcular a distribuição exata de $\hat{p}$.

Distribuição de $\hat{p}$

| $\hat{p}$ | 0 | 1/8 | 2/8 | 3/8 | 4/8 | 5/8 | 6/8 | 7/8 | 1 |
|---|---|---|---|---|---|---|---|---|---|
| $P(\hat{p} = k/8)$ | 0,168 | 0,336 | 0,294 | 0,147 | 0,046 | 0,009 | 0,001 | 0,000 | 0,000 |

(b) Note que $Y = n\bar{X} = \sum_{i=1}^{n} X_i$. Sabemos pelo TLC que $\bar{X} \approx N(p, \frac{p(1-p)}{n})$. Logo, como $Y/n = \hat{p} = \bar{X}$, temos que $\hat{p}$ também tem distribuição aproximadamente $N(0,2, 0,02)$, já substituindo os valores de $p = 0,2$ e $n = 8$.

(c) Para avaliar se é uma boa aproximação vamos calcular $P(\hat{p} \leqslant 1/8)$ usando a aproximação normal, i.e., $P(Z \leqslant \frac{1/8-0,2}{\sqrt{0,02}}) = P(Z \leqslant -0,53033) = 0,298$. Como podemos ver, usando a distribuição exata de $\hat{p}$ esta probabilidade é de 0,5, portanto, bem maior que o valor aproximado. Notamos, assim, que o valor de $n = 8$ é insatisfatório para atingir uma boa aproximação da distribuição da proporção usando a normal.

(d) A aproximação melhor, dado $n$ fixo ocorre quando $p = 1/2$, resultado que pode ser mostrado por simulações de diferentes valores de $p$.

Exercício 4

(a) Para calcular $E(X)$ e $Var(X)$ aproximadamente vamos usar como $X_i$ os pontos médios de cada intervalo. Se fizermos o histograma, poderemos ver que a distribuição é simétrica. Portanto, $E(X) = 6 \cdot 0,10 + 9 \cdot 0,2 + 12 \cdot 0,4 + 15 \cdot 0,2 + 18 \cdot 0,1 = 12$ e $Var(X) = E(X^2) - [E(X)]^2 = 6^2 \cdot 0,10 + 9^2 \cdot 0,2 + 12^2 \cdot 0,4 + 15^2 .0,2 + 18^2 .0,1 - 12^2 = 10,8$. Como a distribuição é simétrica, temos $E(X) = Med(X) = 12$.

(b) Para isso, temos que listar todos os pares possíveis usando os pontos médios de cada intervalo e calcular $\bar{X} = \frac{\sum_{i=1}^{2} X_i}{n}$ Assim, todos os pares possíveis são descritos na tabela abaixo.

Todos os possíveis pares

| | | | | |
|---|---|---|---|---|
| (6,6) | (6,9) | (6,12) | (6,15) | (6,18) |
| (9,6) | (9,9) | (9,12) | (9,15) | (9,18) |
| (12,6) | (12,9) | (12,12) | (12,15) | (12,18) |
| (15,6) | (15,9) | (15,12) | (15,15) | (15,18) |
| (18,6) | (18,9) | (18,12) | (18,15) | (18,18) |

Tirando a média de cada um dos pares acima, obtemos a distribuição abaixo.

Distribuição de probabilidade de $\bar{X}$

| $\bar{X}$ | 6 | 7,5 | 9 | 10,5 | 12 | 13,5 | 15 | 16,5 | 18 |
|---|---|---|---|---|---|---|---|---|---|
| $P(\bar{X} = \bar{x})$ | 0,01 | 0,04 | 0,12 | 0,2 | 0,26 | 0,2 | 0,12 | 0,04 | 0,01 |

Note que como a distribuição é simétrica, a distribuição amostral de $md(X)$ será igual à de $\bar{X}$.

(c) Nota-se que $E(\bar{X}) = E(md(X)) = Med(X) = 12$, pois:

$$\begin{aligned} E(\bar{X}) &= 6 \cdot 0,01 + 7,5 \cdot 0,04 + 9 \cdot 0,12 + 10,5 \cdot 0,2 + 12 \cdot 0,26 + \\ &+ 13,5 \cdot 0,2 + 15 \cdot 0,12 + 16,5 \cdot 0,04 + 18 \cdot 0,01 \\ &= 12 = E(X) \end{aligned}$$

e as distribuições de probabilidade são iguais.

(d) Como ambos são estimadores não viesados de $Med(X)$, além de serem iguais, poderia ser usado qualquer um dos dois para este fim.

(e) Usando a distribuição anterior como base, temos:

Distribuição de Z

| $z$ | -2,582 | -1,936 | -1,291 | -0,645 | 0 | 0,645 | 1,291 | 1,936 | 2,582 |
|---|---|---|---|---|---|---|---|---|---|
| $P(Z = z)$ | 0,01 | 0,04 | 0,12 | 0,2 | 0,26 | 0,2 | 0,12 | 0,04 | 0,01 |

(f) $E(Z) = 0$; $Var(Z) \approx 1$

(g) Usando os pares da letra (b), com $n = 2$, temos:

Distribuição de $S^2$

| $s^2$ | 0 | 4,5 | 18 | 40,5 | 72 |
|---|---|---|---|---|---|
| $P(S^2 = s^2)$ | 0,26 | 0,4 | 0,24 | 0,08 | 0,02 |

(h) $E(S^2) = 10,8$; $Var(S^2) = 294,12$. Note que $E(S^2) = Var(X)$.

(i) Segue distribuição de $t$, com base nas distribuições de $\bar{X}$ e $S^2$ (e usando o fato de serem independentes):

Distribuição de $t$

| $t$ | -3 | -1 | -0,33 | 0 | 0,33 | 1 | 3 |
|---|---|---|---|---|---|---|---|
| $P(T = t)$ | 0,04 | 0,24 | 0,04 | 0,10 | 0,04 | 0,24 | 0,04 |

(j) $E(t) = 0$; $Var(t) \approx 1,21$.

(k) $P(|t| < 2) = P(-2 < t < 2) \approx 0,66$; $P(|t| < 4.3) \approx 0,74$.

Exercício 5

(a) Como $X \sim N(10, 16)$, então $\bar{X} \sim N(10, 16/16) = N(10, 1)$ e $Z = \frac{\bar{X}-10}{1} \sim N(0, 1)$, considerando $n = 16$. Portanto, a probabilidade do participante ganhar o jogo é $P(\bar{X} > 12) = P(Z > 2) = 1 - P(Z \leqslant 2) = 2,275\%$.

(b) Seja $n = 9$ o novo tamanho de amostra. Então, a probabilidade de ganhar o jogo é $P(\bar{X} > 12) = P(Z > \frac{12-10}{4/\sqrt{9}}) = P(Z > 3/2) = 6,68\%$.

(c) Dados os resultados acima, concluímos que quanto menor a amostra, maior é a probabilidade de ganhar o jogo. Assim, entre as duas opções acima, escolheria a amostra de tamanho $n = 9$.

Exercício 6

(a) $E(e) = E(\bar{X} - \mu) = E(\bar{X}) - \mu = 0$

$Var(e) = Var(\bar{X}) + Var(\mu) - 2Cov(\bar{X}, \mu) = \frac{\sigma^2}{n} + 0 - 0 = \frac{400}{n}$

(b) Para $n = 25$, temos $P(|\bar{X} - \mu| > 2) = 1 - P(|\bar{X} - \mu| \leqslant 2) = 1 - P(-2 \leqslant \bar{X} - \mu \leqslant 2) = 1 - P(\frac{-2.\sqrt{n}}{\sigma} \leqslant Z \leqslant \frac{2.\sqrt{n}}{\sigma}) = 1 - P(\frac{-10}{20} \leqslant Z \leqslant \frac{10}{20}) = 1 - P(-0,5 \leqslant Z \leqslant 0,5) = 1 - 38,29\% = 61,7\%$.

(c) Para $n = 100$, temos $P(|\bar{X} - \mu| > 2) = 1 - P(|\bar{X} - \mu| \leqslant 2) = 1 - P(-2 \leqslant \bar{X} - \mu \leqslant 2) = 1 - P(\frac{-2.\sqrt{n}}{\sigma} \leqslant Z \leqslant \frac{2.\sqrt{n}}{\sigma}) = 1 - P(\frac{-20}{20} \leqslant Z \leqslant \frac{20}{20}) = 1 - P(-1 \leqslant Z \leqslant 1) = 31,73\%$.

(d) Considerando novamente $n = 100$, temos que $P(|\bar{X} - \mu| > d) = 1 - P(|\bar{X} - \mu| \leqslant d) = 1 - P(-d \leqslant \bar{X} - \mu \leqslant d) = 1 - P(\frac{-d.\sqrt{n}}{\sigma} \leqslant Z \leqslant \frac{d.\sqrt{n}}{\sigma}) = 1 - P(\frac{-10d}{20} \leqslant Z \leqslant \frac{10d}{20}) = 1 - P(-0,5d \leqslant Z \leqslant 0,5d) = 1\%$. O valor $d$ é tal que $P(Z \leqslant 0,5d) = 99,5\%$, ou seja, $d = 5,15$.

(e) $P(|\bar{X} - \mu| < 1) = P(-1 < \bar{X} - \mu < 1) = P(-1.\frac{\sqrt{n}}{20} < Z < 1.\frac{\sqrt{n}}{20}) = 95\% \Rightarrow 2.P(Z < \frac{\sqrt{n}}{20}) - 1 = 95\% \Rightarrow P(Z < \frac{\sqrt{n}}{20}) = 97,5\% \Rightarrow \frac{\sqrt{n}}{20} = 1,9599 \Rightarrow n \approx 1537$. Ou seja, para que isto aconteça a amostra tem que ser grande.

Exercício 7

(a) O processo para se ocorre o evento A: $\bar{X} \geqslant 53,7$ ou ocorre o evento B: $\bar{X} \leqslant 46,3$. Logo, a probabilidade de o processo parar é $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Se $\mu = 50$, $n = 4$ e $\sigma = 2,5$, $P(A \cup B) = P(\bar{X} \geqslant 53,7) + P(\bar{X} \leqslant 46,3) - 0 = P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geqslant \frac{53,7-50}{2,5/\sqrt{4}}) + P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leqslant \frac{46,3-50}{2,5/\sqrt{4}}) = P(Z \geqslant 2,96) + P(Z \leqslant -2,96) = 2.P(Z \leqslant -2,96) = 0,308\%$.

(b) Se $\mu = 53,7$, $n = 4$ e $\sigma = 2,5$, $P(46,3 \leqslant \bar{X} \leqslant 53,7) = P(\frac{46,3-53,7}{2,5/\sqrt{4}} \leqslant Z \leqslant \frac{53,7-53,7}{2,5/\sqrt{4}}) = P(-5,92 \leqslant Z \leqslant 0) = P(Z \leqslant 0) - P(Z \leqslant -5,92) = 0,5 - 0 = 50\%$.