CE-003: Estatística II - Turma AMB - Avaliações Semanais (1^o semestre 2013)

1. Considere que será feita uma pesquisa aplicando-se um questionário sobre o curso e aos alunos da Eng. Ambiental para avaliar opiniões e impressões dos alunos.

   (a) Liste possíveis questões deste questionário certificando-se que sejam incluídas ao menos duas de cada tipo de variáveis conforme discutido em aula (qualitativas nominal/ordinal e quantitativas discreta/contínua).

   (b) Imagine agora que o questionário foi aplicado e as respostas tabuladas para análises. Indique/esboce como seria analisada (separadamente) cada uma das variáveis do questionário.

   (c) Indique ao menos três questões de interesse envolvendo duas ou mais variáveis a serem investigadas no questionário e qual análise dos dados permitiria investigar estas questões.

---

2. Foram coletados dados[1] sobre indicadores sociais em 97 países. Os atributos[2] são: *Nat: taxa de natalidade (1.000 hab.)*, *Mort: taxa de mortalidade (1.000 hab.)*, *MI: mortalidade infantil (1.000 hab)*, *ExpM: expectativa de vida para homens*, *ExpF: expectativa de vida para mulheres*, *Renda: renda per capta anula* e *Região: região geográfica* sendo consideradas: "EUOr"(Europa Oriental),"SA"(América Latina e México),"PM"("Primeiro Mundo"),"OrMd"(Oriente Médio), "Asia"e "Africa". A renda *per capta* foi também dividida em classes: $[0, 500)$, $[500, 2.000)$, $[2.000, 10.000)$ e $[10.000, 35.000)$. Um cabeçalho do arquivo de dados e um resumo das variáveis são mostrados a seguir.

```
                   Nat Mort   MI ExpM ExpF Renda Regiao    GrupoRenda
Albania            24.7  5.7 30.8 69.6 75.5   600   EUOr   (500,2e+03]
Bulgaria           12.5 11.9 14.4 68.3 74.7  2250   EUOr (2e+03,1e+04]
Czechoslovakia     13.4 11.7 11.3 71.8 77.7  2980   EUOr (2e+03,1e+04]
Former_E._Germany  12.0 12.4  7.6 69.8 75.9    NA   EUOr          <NA>
Hungary            11.6 13.4 14.8 65.4 73.8  2780   EUOr (2e+03,1e+04]
Poland             14.3 10.2 16.0 67.2 75.7  1690   EUOr   (500,2e+03]
```

```
      Nat            Mort             MI              ExpM            ExpF
 Min.   : 9.7   Min.   : 2.2   Min.   :  4.5   Min.   :38.1   Min.   :41.2
 1st Qu.:14.5   1st Qu.: 7.8   1st Qu.: 13.1   1st Qu.:55.8   1st Qu.:57.5
 Median :29.0   Median : 9.5   Median : 43.0   Median :63.7   Median :67.8
 Mean   :29.2   Mean   :10.8   Mean   : 54.9   Mean   :61.5   Mean   :66.2
 3rd Qu.:42.2   3rd Qu.:12.5   3rd Qu.: 83.0   3rd Qu.:68.6   3rd Qu.:75.4
 Max.   :52.2   Max.   :25.0   Max.   :181.6   Max.   :75.9   Max.   :81.8
```

```
     Renda           Regiao              GrupoRenda
 Min.   :   80   EUOr  :11    (0,500]         :24
 1st Qu.:  475   SA    :12    (500,2e+03]     :24
 Median : 1690   PM    :19    (2e+03,1e+04]   :22
 Mean   : 5741   OrMd  :11    (1e+04,3.5e+04]:21
 3rd Qu.: 7325   Asia  :17    NA's           : 6
 Max.   :34064   Africa:27
 NA's   :6
```

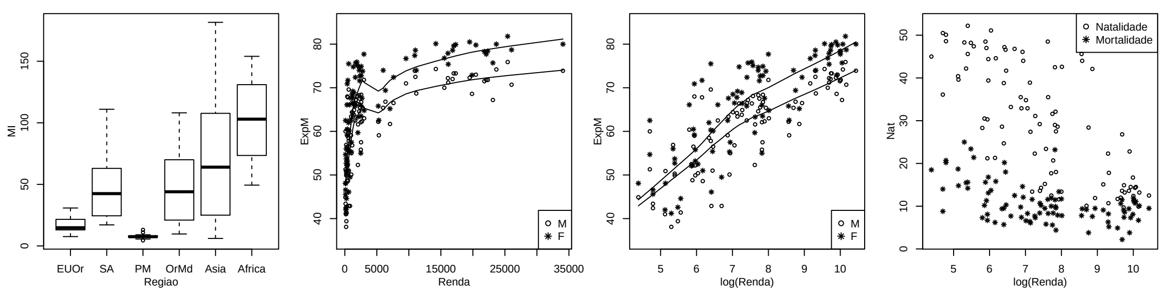
A seguir são mostrados alguns gráficos e resumos dos dados. Inicialmente são mostrados resumos das taxas de natalidade (NAT) para cada faixa de renda. A seguir uma tabela relaciona o grupo de renda com a região geográfica. Os gráficos ilustram relacionamentos entre algumas variáveis. As últimas matrizes são de correlação de Pearson e Spearman respectivamente.

   (a) Faça interpretações estatísticas, no contexto do problema, de cada um dos resultados mostrados.

   (b) Comente ao menos mais duas (2) questões de interesse que poderiam ser investigadas e não foram abordadas nos resultados já mostrados. Indique como seriam utilizados os dados (tipo de análise) para abordar estas questões.

[1] http://www.amstat.org/publications/jse/datasets/poverty.dat.txt
[2] http://www.amstat.org/publications/jse/datasets/poverty.txt

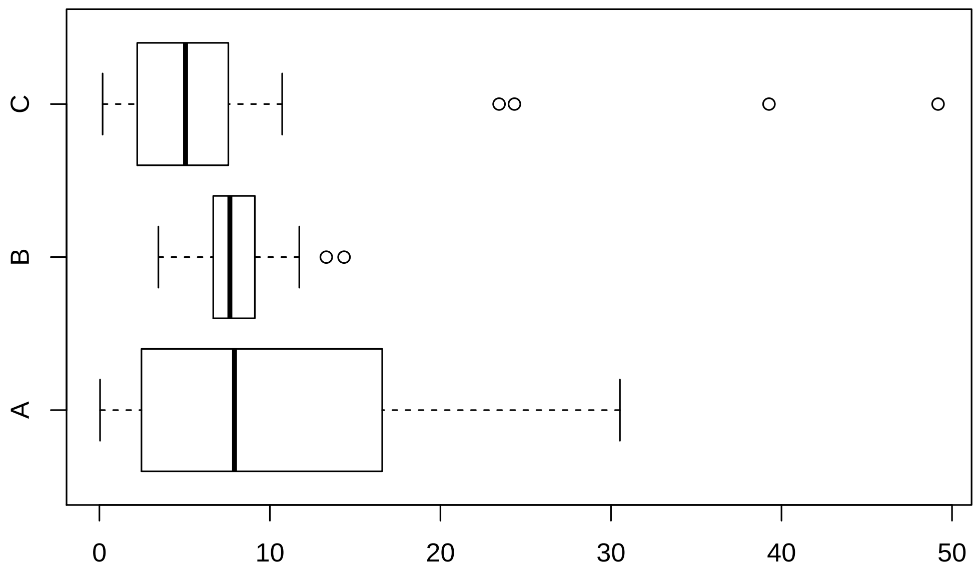
```
$`(0,500]`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   21.2    38.6    44.8    41.7    48.3    52.2

$`(500,2e+03]`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   13.4    24.4    32.9    31.8    39.6    47.2

$`(2e+03,1e+04]`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   10.1    15.8    28.5    27.7    40.4    48.5

$`(1e+04,3.5e+04]`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    9.7    12.0    13.6    14.7    14.9    26.8

                 Regiao
GrupoRenda        EUOr SA PM OrMd Asia Africa
  (0,500]            0  1  0    0    8     15
  (500,2e+03]        5  6  0    2    3      8
  (2e+03,1e+04]      4  5  3    5    1      4
  (1e+04,3.5e+04]    0  0 16    3    2      0

X-squared
    87.64
```




```
          Nat    Mort      MI    ExpM    ExpF   Renda
Nat    1.0000  0.4862  0.8584 -0.8665 -0.8944 -0.6291
Mort   0.4862  1.0000  0.6546 -0.7335 -0.6930 -0.3028
MI     0.8584  0.6546  1.0000 -0.9368 -0.9554 -0.6016
ExpM  -0.8665 -0.7335 -0.9368  1.0000  0.9826  0.6430
ExpF  -0.8944 -0.6930 -0.9554  0.9826  1.0000  0.6500
Renda -0.6291 -0.3028 -0.6016  0.6430  0.6500  1.0000

          Nat    Mort      MI    ExpM    ExpF   Renda
Nat    1.0000  0.4045  0.8861 -0.8823 -0.9018 -0.7342
Mort   0.4045  1.0000  0.4930 -0.5942 -0.5346 -0.4473
MI     0.8861  0.4930  1.0000 -0.9481 -0.9622 -0.8363
ExpM  -0.8823 -0.5942 -0.9481  1.0000  0.9784  0.8240
ExpF  -0.9018 -0.5346 -0.9622  0.9784  1.0000  0.8391
Renda -0.7342 -0.4473 -0.8363  0.8240  0.8391  1.0000
```

---

3. (a) Os tempos de atendimento e solução de problemas foram medidos em três *call-centers* distintos de uma mesma empresa e os dados foram representados no gráfico a seguir. Baseando-se no gráfico, avalie cada uma das afirmações a seguir, dizendo se está certa ou errada, justificando sua resposta e corrigindo as afirmações erradas.



( ) Os valores no local C possuem uma distribuição simétrica.
( ) Os dados discrepantes do local A afetam (aumentam) a mediana do local.
( ) Os locais B e C possuem médias e desvios padrão semelhantes.
( ) O local B possui o menor coeficiente de variação.
( ) As médias dos três locais devem ser semelhantes.

(b) Uma cidade recebeu críticas à sua excessiva descarga de esgoto não tratado em um rio. Um microbiologista tomou 45 amostras na água depois da passagem pela planta de tratamento de esgoto e mediu a quantidade de coliformes (bactéria) presente nas amostras.

Número de Bactérias	Número de amostras
20-30	5
30-40	20
40-50	15
50-60	5

i. Obtenha a média
ii. Obtenha a mediana

Solução:

i. $\overline{x} = 39.44$
ii. $\text{md}(x) = 30 + \frac{10*(22{,}5-5)}{20} = 38.75$

(c) Em um levantamento geológico foram coletadas amostras de sedimentos de fundo de rios de uma bacia hidrográfica. Os teores obtidos de um certo elemento são mostrados a seguir.

`2.3 4.0 2.7 34.5 48.8 11.6 36.5 32.8 22.3 2.1 3.1 0.7 5.2`

`1.5 11.4 3.7 5.1 5.1 1.2 8.9 19.2 5.5 1.3 14.2 27.4`

i. obtenha o teor médio e o desvio padrão,
ii. obtenha os quantis e a amplitude,
iii. obtenha o coeficiente de variação,
iv. obtenha um histograma,
v. obtenha um box-plot,
vi. obtenha um diagrama de ramo-e-folhas,
vii. comente sobre o padrão da distribuição dos dados e se voce consideraria alguma outra forma de analisa-los.

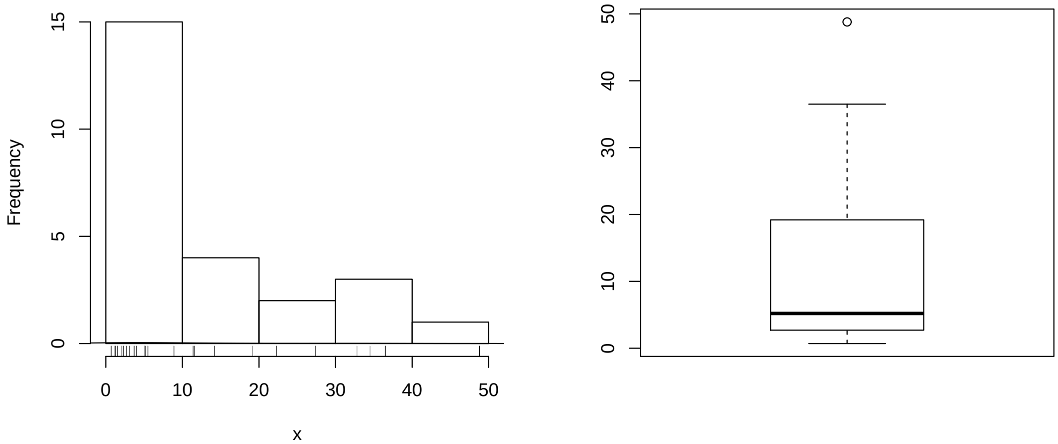


Figura 1: (d) histograma (esquerda) e (e) *box-plot* (direita) dos dados

Solução:

i. $\overline{x} = 12.44$ e $S_x = 13.6$

ii.
```
     Q1        md        Q3 Amplitude
    2.7       5.2      19.2      48.1
```

iii. $C.V. = 109\%$

iv.

v.

vi.
```
> stem(x)
The decimal point is 1 digit(s) to the right of the |

0 | 111222334455569
1 | 1249
2 | 27
3 | 357
4 | 9
```

vii. comentários

---

4. (a) Dois dados são lançados. Calcule a probabilidade de:
   i. sairem dois números iguais,
   ii. o produto dos números que sairam ser ímpar,
   iii. o produto dos números que sairam ser ímpar ou a soma ser maior ou igual a 10,
   iv. a soma dos valores ser maior ou igual a sete, sabendo-se que em um dos dados saiu três,
   v. a soma ser maior que sete sabendo que sairam dois números iguais.

Solução:

$$X_1 : \text{resultado do primeiro dado} \quad X_2 : \text{resultado do segundo dado}$$

$$\Omega = \{(1,1), (1,2), \dots (6,6)\} \quad n(\Omega) = 36$$

i. $P[X_1 = X_2] = \frac{n(X_1 = X_2)}{n(\Omega)} = \frac{6}{36} = 0.167$

$$\text{evento } A : X_1 = X_2$$

$$P[A] = \frac{n(X_1 = X_2)}{n(\Omega)} = \frac{6}{36} = \frac{1}{6} = 0.167$$

ii.

$$\text{evento } B : X_1 \cdot X_2 \text{ é impar}$$
$$P[B] = \frac{9}{36} = \frac{1}{4} = 0.25$$

iii.

$$\text{evento } C_1 : X_1 \cdot X_2 \text{ é impar}$$
$$\text{evento } C_2 : X_1 + X_2 \geq 10$$
$$P[C_1 \cup C_2] = P[C_1] + P[C_2] - P[C_1 \cap C_2] = \frac{9}{36} + \frac{6}{36} - \frac{1}{36} = \frac{14}{36} = 0.389$$

iv. a soma dos valores ser maior ou igual a sete, sabendo-se que em um dos dados saiu três,

$$\text{evento } D_1 : X_1 + X_2 \geq 7$$
$$\text{evento } D_2 : X_1 = 3 \cup X_2 = 3$$
$$P[D_1 \cup D_2] = P[D_1] + P[D_2] - P[D_1 \cap D_2] = \frac{21}{36} + \frac{11}{36} - \frac{6}{36} = \frac{26}{36} = 0.722$$

v. a soma ser maior que sete sabendo que sairam dois números iguais.

$$\text{evento } E_1 : X_1 + X_2 \geq 7$$
$$\text{evento } E_2 : X_1 = X_2$$
$$P[E_1|E_2] = \frac{P[E_1 \cap E_2]}{P[E_2]} = \frac{3/36}{6/36} = \frac{3}{6} = \frac{1}{2} = 0.5$$

(b) Em um teste múltipla escolha, marca-se uma alternativa em cada uma de quatro questões, cada uma com cinco alternativas da qual apenas uma é correta. Qual a probabilidade de um indivíduo acertar por mero acaso alguma questão?

Solução:

$$A_i : \text{acerta a } i\text{-ésima questão} \;\; i = 1, \ldots, 4$$
$$P(A_i) = 0,2 \;\; P(\overline{A}_i) = 0,8 \;\; \forall i$$
$$P(\text{acertar alguma}) = 1 - P(\text{errar todas}) = 1 - P(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap \overline{A_4}) \stackrel{ind}{=} 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(\overline{A_3}) \cdot P(\overline{A_4})) =$$
$$= 1 - (0,8)^4 = 0.59$$

(c) Um reservatório recebe água de três fontes diferentes. A primeira tem 5% de chance de apresentar alguma contaminação, a segunda tem 6,5% e a terceira tem 12%. Qual a probabilidade do reservatório ser contaminado?

Solução:
Evento A : a água da primeira fonte é contaminada
Evento B : a água da segunda fonte é contaminada
Evento C : a água da terceira fonte é contaminada
Dados:

$$P[A] = 0,05 \quad ; P[\overline{A}] = 0,95$$
$$P[B] = 0,065 \quad ; P[\overline{B}] = 0,935$$
$$P[C] = 0,12 \quad ; P[\overline{C}] = 0,88$$
$$P[\text{contaminação}] = P[A \cup B \cup C] = 1 - P[\overline{A} \cap \overline{B} \cap \overline{C}] \stackrel{ind}{=} P1 - [\overline{A}] \cdot P[\overline{B}] \cdot P[\overline{C}] = 1 - 0.95 \cdot 0.935 \cdot 0.880.2183$$

---

5. (a) Considere o problema a seguir de uma avaliação semanal anterior.
Em um teste múltipla escolha, marca-se uma alternativa em cada uma de quatro questões, cada uma com cinco alternativas da qual apenas uma é correta. Qual a probabilidade de um indivíduo acertar por mero acaso alguma questão?

i. Indique como fica o espaço amostral do experimento (sem necessariamente listar todos os elementos).
ii. Defina a variável aleatória (v.a) adequada ao interesse do problema.
iii. Monte uma tabela com a distribuição de probabilidades desta variável
iv. Caso possível identifique a distribuição de probabilidades desta variável e fornecendo a equação da distribuição.
v. Mostre como obter a probabilidade solicitada a partir do resultado de alguns dos itens anteriores.
vi. Qual o valor esperado va v.a ? Como este valor deve ser interpretado?

Solução:

i. $\Omega = (\overline{AAAA}), (\overline{AAA}A), (\overline{AA}A\overline{A}), \ldots (AAA\overline{A}), (AAAA) \quad n(\Omega) = 2^4 = 16$
ii. X : número de acertos
iii.

x	0	1	2	3	4
$P[X = x]$	$(0,8)^4$	$\binom{4}{1}(0,2)^1(0,8)^3$	$\binom{4}{2}(0,2)^2(0,8)^2$	$\binom{4}{3}(0,2)^3(0,8)^2$	$(0,2)^4$

iv. $X \sim B(n = 4, p = 0,2) \quad P[X = x] = \binom{4}{x}(0,2)^x(1 - 0,8)^{4-x}$
v. $P[X \geq 1] = 1 - P[X = 0] = 1 - \binom{4}{0}(0,2)^0(1 - 0,2)^{4-0} = 0.59$
vi. interpretação de $E[X]$

(b) Seja a função:

$$f(x) = \begin{cases} 3x^2/8 & 0 < x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

i. Mostre que $f(x)$ é uma função de densidade de probabilidade válida.
ii. Obtenha $P[0,5 < X < 1,5]$.
iii. Obtenha $P[X > 1,2]$.
iv. Obtenha $P[X > 1,2 | X > 0,5]$.
v. Obtenha o valor experado de X.

Solução:

i.

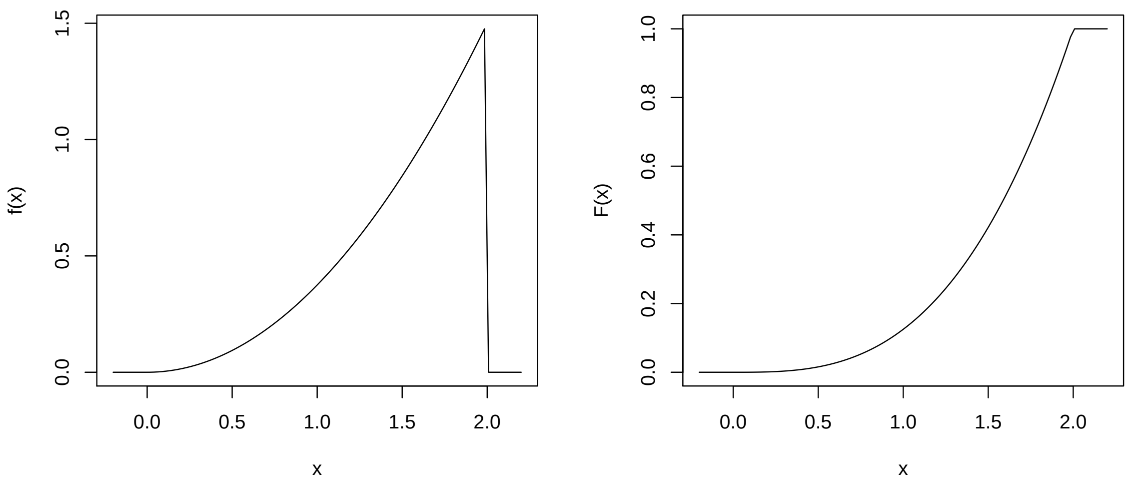
$$\text{Mostrar que: } f(x) \geq 0 \forall x \quad \text{e} \quad \int_0^2 f(x)\mathrm{d}x = 1$$
$$\frac{3}{8}\frac{2^3 - 0^3}{3} = 1$$
$$F(x) = \int_0^x f(x)\mathrm{d}x = \frac{3}{8}\frac{x^3 - 0^3}{3} = \frac{x^3}{8}$$

ii. $P[0,5 < X < 1,5] = \int_{0,5}^{1,5} f(x)\mathrm{d}x = Fx(1,5) - Fx(0,5) = 0.406$
iii. $P[X > 1,2] = \int_{1,2}^{2} f(x)\mathrm{d}x = 1 - Fx(1,2) = 0.784$
iv. $P[X > 1,2 | X > 0,5] = \frac{\int_{1,2}^{2} f(x)\mathrm{d}x}{\int_{0,5}^{2} f(x)\mathrm{d}x} = \frac{1 - Fx(1,2)}{1 - Fx(0,5)} = 0.796$
v.

$$E[X] = \int_0^2 x \cdot f(x)\mathrm{d}x = \frac{3}{8}[\frac{2^4 - 0^4}{4}] = \frac{3}{2} = 1,5$$

Resoluções computacionais:

```
> require(MASS)
> ## a)
> fx <- function(x) ifelse(x > 0 & x <= 2, (3*x^2)/8, 0)
> integrate(fx, 0, 2)$value

[1] 1

> Fx <- function(x) ifelse(x>0, ifelse(x<=2, (x^3)/8,1), 0)
> Fx(2)

[1] 1

> ## b)
> integrate(fx, 0.5, 1.5)$value
```



H!

Figura 2: Função de densidade de probabilidade (esquerda) e função de distribuição.

```
[1] 0.4062
> Fx(1.5)-Fx(0.5)
[1] 0.4062
> ##c)
> integrate(fx, 1.2, 2)$value
[1] 0.784
> 1-Fx(1.2)
[1] 0.784
> ## d)
> integrate(fx, 1.2, 2)$value/integrate(fx, 0.5, 2)$value
[1] 0.7964
> (1-Fx(1.2))/(1-Fx(0.5))
[1] 0.7964
> ## e)
> efx <- function(x) ifelse(x > 0 & x <= 2, x*(3*x^2)/8, 0)
> integrate(efx, 0, 2)$value
[1] 1.5
```

---

6. Alguns biólogos fizeram estudos de laboratório sobre o comportamento de animais quando submetidos a um estímulo, o quais poderiam apresentar ou não resposta positiva. Em particular estavam interessados nas respostas positivas os estímulo. Considera-se que na população destes animais, 10% sejam sensíveis ao estímulo.

   O biologo A possuia um grupo em que 10 animais eram sensíveis e 20 eram insensíveis. Ele selecionou ao acaso 8 animais para teste.

   O biólogo B dispunha de um grande número de animais e foi testando um a um até encontrar o terceiro sensível ao estímulo.

   O biólogo C tomou fazia testes diários e encontrava uma média de 2,8 animais sensíveis a cada dia.

   O biólogo D submeteu 10 animais ao estímulo.

   O biólogo E dispunha de um grande número de animais e foi testando um a um até encontrar um sensível ao estímulo.

   (a) Qual a probabilidade do biólogo A encontrar ao menos 2 animais sensíveis?
   (b) Qual a probabilidade do biólogo B precisar testar no máximo 6 animais?
   (c) Qual a probabilidade do biólogo C encontrar menos que dois sensíveis em um determinado dia?
   (d) Qual a probabilidade do biólogo D encontrar mais que 3 animais sensíveis?
   (e) Qual a probabilidade do biólogo E precisar testar mais que 3 animais?

   Sugestão: especifique a(s) variável(eis) aleatória, sua(s) distribuição(ções) e pressuposições feitas.

Solução:

(a)

$$
\begin{aligned}
X_a &: \text{número de sensíveis entre os 8 selecionados} \\
x_a &\in \{0, 1, 2, \dots 8\} \\
X_a &\sim HG(N = 30, K = 10, n = 8) \\
P[X_a = x] &= \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = \frac{\binom{10}{x}\binom{20}{8-x}}{\binom{30}{8}} \\
P[X_a \geq 2] &= 1 - P[X_a \leq 1] = 1 - (P[X_a = 0] + P[X_a = 1]) = 0.846
\end{aligned}
$$

(b)

$$
\begin{aligned}
X_b &: \text{número de não sensíveis examinados até encontrar o terceiro sensível} \\
x_b &\in \{0, 1, \dots\} \\
X_b &\sim BN(k = 3, p = 0,1) \\
P[X_b = x] &= \binom{x+k-1}{x} p^k (1-p)^x = \binom{x+2}{x} 0,1^3 0,9^x \\
P[X_b \leq 3] &= P[X_b = 0] + P[X_b = 1] + P[X_b = 2] + P[X_b = 3] = 0.016
\end{aligned}
$$

(c)

$$
\begin{aligned}
X_c &: \text{número de sensíveis encontrados em um dia} \\
x_c &\in \{0, 1, \dots\} \\
X_c &\sim P(\lambda = 2,8) \\
P[X_c = x] &= \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} = \frac{\mathrm{e}^{-2,8} 2,8^x}{x!} \\
P[X_c < 2] &= P[X_c = 0] + P[X_c = 1] = 0.231
\end{aligned}
$$

(d)

$$
\begin{aligned}
X_d &: \text{número de sensíveis entre os 10 examinados} \\
x_d &\in \{0, 1, \dots 10\} \\
X_d &\sim B(n = 10, p = 0,1) \\
P[X_d = x] &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{10}{x} 0,1^x 0,9^{10-x} \\
P[X_d > 3] &= 1 - P[X_d \leq 3] = 1 - (P[X_d = 0] + P[X_d = 1] + P[X_d = 2] + P[X_d = 3]) = 0.013
\end{aligned}
$$

(e)

$$
\begin{aligned}
X_e &: \text{número de não sensíveis examinados até encontrar o primeiro sensível} \\
x_e &\in \{0, 1, \dots\} \\
X_e &\sim G(p = 0,1) \\
P[X_e = x] &= p(1-p)^x = 0,1 \cdot 0,9^x \\
P[X_e > 3] &= 1 - P[X_e \leq 3] = 1 - (P[X_e = 0] + P[X_e = 1] + P[X_e = 2] + P[X_e = 3]) = 0.656
\end{aligned}
$$

---

7. (a) Um biólogo percorre uma trilha de 5 km procurando avistar um exemplar de uma determinada ave. A chance de avistar a ave durante uma passada no percurso é de 25% e constante em todo o percurso.
   - i. Qual a probabilidade avistar a ave e que seja nos primeiros 2 km do percurso?
   - ii. Se ele avista a ave, qual a probabilidade de que seja nos últimos 500 metros do percurso?
   - iii. Se ele avista a ave, qual a probabilidade de que seja no primeiro ou último quilometro do percurso?
   - iv. Se ele avistou a ave e sabe-se que não foi nos primeiros 2 km qual a probabilidade de que tenha sido nos últimos 1.500 metros do percurso?
   - v. É adotada a seguinte classificação para uma campanha: A: ave avistada nos primeiros 1.500 metros; B: ave avistada entre 1.500 e 4.000 metros; C: ave avistada nos últimos 1.000 metros; X: ave não avistada. Monte a distribuição de probabilidades da classificação da campanha.

Solução:

$$\begin{aligned} A &: \text{a ave é vista} \quad P[A] = 0,25 \\ X &: \text{posição onde a ave é avistada} \\ X|A &\sim U_c[0,5] \\ f(x|A) &= \frac{1}{5-0} = \frac{1}{5}\mathrm{I}_{[0,5]}(x) \qquad F(x) = \frac{x-0}{5-0} = \frac{x}{5} \end{aligned}$$

i. $P[A \cap X < 2] = P[A] \cdot P[X < 2|A] = 0,25 \cdot \frac{2}{5} = \frac{1}{10} = 0.1$

ii. $P[X > 4,5|A] = \frac{0,5}{5} = \frac{1}{10} = 0.1$

iii. $P[X < 1 \cup X > 4|A] = \frac{1}{5} + \frac{1}{5} = \frac{2}{5} = 0.4$

iv. $P[X > 3,5|A \cap X > 2] = \frac{\frac{5-3,5}{5}}{\frac{5-2}{5}} = \frac{1}{2} = 0.5$

v.

$$Y \sim \text{classificação da campanha} \quad y \in \{A, B, C, X\}$$

y	A	B	C	X
$P[Y = y]$	$P[A \cap X < 1,5]$	$P[A \cap 1,5 < X < 4]$	$P[A \cap X > 4]$	$P[\overline{A}]$
$P[Y = y]$	$P[A] \cdot P[X < 1,5\|A]$	$P[A] \cdot P[1,5 < X < 4\|A]$	$P[A] \cdot P[X > 4\|A]$	$P[\overline{A}]$
$P[Y = y]$	$0,25 \cdot 1,5/5 = 0.075$	$0,25 \cdot 2,5/5 = 0.125$	$0,25 \cdot 1/5 = 0.05$	0,75

(b) Assume-se que o tempo entre acessos a um blog tem uma distribuição com média de 1,5 segundos. Assumindo alguma distribuição responda os itens a seguir.

i. Qual a probabilidade de haver duas conexões com intervalo inferior a 1,5 segundos?

ii. Qual a probabilidade de se passarem 5 segundos sem conexão alguma?

iii. Tendo havido uma conexão, qual a probabilidade da próxima conexão ocorrer entre 0,5 e 2,5 segundos?

iv. Se já se passou 1 segundo sem conexão, qual a probabilidade de se passar mais 0,5 segundos adicionais sem conexão?

v. Qual a probabilidade do intervalo entre conexões não superar 3,5 segundos se já se passaram 2 segundos sem conexão?

Solução:
Não se especificou a distribuição e vamos assumir a distribuição exponencial considerando: (i) que devem ser valores positivos, (ii) pela possibilidade de cálculos com as informações fornecidas.

$$\begin{aligned} X &: \text{intervalo de tempo entre conexões (segundos)} \\ X &\sim \mathrm{Exp}(\lambda = 1/1,5 = 2/3) \\ f(x) &= \frac{2}{3}\mathrm{e}^{-2x/3}\ \ \mathrm{I}_{(0,\infty)}(x) \qquad F(x) = 1 - \mathrm{e}^{-2x/3} \end{aligned}$$

i. $P[X < 1,5] = \int_0^{1,5} f(x)\mathrm{d}x = F(1,5) = 0.63$

ii. $P[X > 5] = \int_5^{\infty} f(x)\mathrm{d}x = 1 - F(5) = 0.036$

iii. $P[X < 0,5] = \int_{0,5}^{2,5} f(x)\mathrm{d}x = F(2,5) - F(0,5) = 0.53$

iv. $P[X > 1,5|X > 1] = \frac{\int_{1,5}^{\infty} f(x)\mathrm{d}x}{\int_1^{\infty} f(x)\mathrm{d}x} =^3 P[X > 0,5] = 1 - F(0,5) = 0.72$

v. $P[X < 3,5|X > 2] = \frac{\int_2^{3,5} f(x)\mathrm{d}x}{\int_2^{3,5} f(x)\mathrm{d}x} = \frac{F(3,5)-F(2)}{1-F(2)} =^3 P[X < 1,5] = F(1,5) = 0.63$

---

8. Um determinado elemento é medido regularmente em amostras de água e os valores possuem distribuição normal de média 500 e desvio padrão de 20 unidades.

(a) Se um grande número de amostras é processada, qual a proporção esperada com teores acima da 525 unidades?

(b) Se 6.000 amostras são processadas, quantas devem apresentar teores entre 470 e 510 unidades?

(c) Se uma amostra é considerada suspeita e enviada para reanálise se apresenta teor acima de 535 unidades, quantas dentre as 6.000 serão enviadas para reanálise?

(d) Considera-se como teores *usuais* 80% dos valores ao redor da média. Quais os valores que determinam a faixa de teores *usuais*.

(e) Qual deveria ser o teor médio para que 90% das amostras estivessem abaixo de 500 unidades?

[3] propriedade de falta de memória da exponencial

Solução:

$$X : \text{teores das amostras (unidades)}$$
$$X \sim N(\mu = 500, \sigma^2 = 20^2)$$

(a) $P[X > 525] = P[Z > \frac{525-500}{20}] = P[Z > 1.25] = 0,5 - P[0 < Z < 1.25] = 0,5 - 0.3944 = 0.1056$

(b)

$$P[470 < X < 510] = P[\frac{470-500}{20} < Z < \frac{510-500}{20}] = P[-1.5 < Z < 0.5] = P[0 < Z < 1.5] + P[0 < Z < 0.5] =$$
$0.4332 + 0.1915 = 0.6247$

$$\text{Número esperado} : n = 6.000 \cdot 0.6247 = 3748$$

(c)

$$P[X > 535] = P[Z > \frac{535-500}{20}] = P[Z > 1.75] = 0,5 - P[0 < Z < 1.75] = 0,5 - 0.4599 = 0.04006$$
$$\text{Número esperado} : n = 6.000 \cdot 0.04006 = 240$$

(d)

$$P[x_1 < X < x_2] = P[z_1 < Z < z_2] = 0,80$$
$$P[-1.282 < Z < 1.282] = 0,80$$
$$z = \frac{x-\mu}{s}$$
$$-1.282 = \frac{x_1 - 500}{20}$$
$$x_1 = 474.4$$
$$1.282 = \frac{x_2 - 500}{20}$$
$$x_2 = 525.6$$

(e)

$$P[X < 500] = 0,90$$
$$P[Z < 1.282] = 0,90$$
$$z = \frac{x-\mu}{s}$$
$$1.282 = \frac{500-\mu}{20}$$
$$\mu = 474.4$$

Solução Computacional:

```
> p.a <- pnorm(525, m=500, sd=20, low=F)
> p.b <- diff(pnorm(c(470,510), m=500, sd=20)); na <- 6000*p.b
> p.c <- pnorm(535, m=500, sd=20, low=F) ; n.c <- 6000*p.c
> q.d <- qnorm(c(0.10, 0.90), m=500, sd=20)
> mu.e  <- 500 - qnorm(0.90) * 20
```

---

9. Um determinado elemento é medido regularmente em amostras água coletadas um reservatório. Assume-se que os valores individuais das coletas possuem distribuição normal de média 500 e desvio padrão de 20 unidades. Em um programa de monitoramento são feitas periodicamente como rotina a amostragem com análise de cinco coletas. Se algum anomalia é detectada é feita uma inspeção detalhada. Adota-se considerar uma *anomalia* se o teor médio de uma amostra ultrapassa valor crítico de 512 unidades.

   (a) Qual a proporção de coletas que devem apresentar teores individuais entre 490 e 515 unidades?

(b) Qual a proporção de médias amostrais (das cinco coletas) que devem apresentar teores individuais entre 490 e 515 unidades?

(c) Entre quais valores de teores das coletas individuais ao redor da média de 500 unidades espera-se obter 90% das medidas?

(d) Entre quais valores ao redor da média de 500 unidades de teores médios das amostras de cinco coletas espera-se obter 90% das médias amostrais?

(e) Verificou-se em uma amostra o valor médio de 510 unidades. Voce considera que há uma anomalia e recomendaria uma inspeção detalhada? Justifique.

(f) Qual a probabilidade de encontrar um teor médio na amostra acima do valor crítico de 512 unidades, quando os valores do lago estão nos níveis usuais?

(g) Qual deveria ser o *valor crítico* para que a probabilidade de encontrar teores médios acima dele não ultrapasse 0,02?

(h) Se o *valor crítico* é mantido em 512 unidades qual deveria ser o número de coletas por amostra para que a probabilidade de encontrar teores médios acima de 512 unidades não ultrapasse 0,02?

(i) Suponha que tenha havido uma alteração no reservatório elevando a concentração média de 500 para 515 unidades. Qual seria a probabilidade de obter uma média amostras abaixo de 500 unidades?

(j) Ainda no contexto do item anterior, qual a probabilidade de não se recomendar uma inspeção detalhada?

Solução:

$$X : \text{tempo de processamentos (segundos)} \quad X \sim N(\mu_X = 500, \sigma^2_X = 20^2)$$
$$\overline{X} : \text{tempo de procesamentos (segundos)} \quad \overline{X} \sim N(\mu_{\overline{X}} = \mu_X, \sigma^2_{\overline{X}} = \sigma^2_X) \sim N(\mu_{\overline{X}} = 500, \sigma^2_{\overline{X}} = 20^2/5)$$

(a) $P[490 < X < 515] = P[\frac{490-500}{20} < Z < \frac{515-500}{20}] = P[-0.5 < Z < 0.75] = P[0 < Z < 0.5] + P[0 < Z < 0.75] = 0.4648$

(b) $P[490 < \overline{X} < 515] = P[\frac{490-500}{20/\sqrt{5}} < Z < \frac{515-500}{20/\sqrt{5}}] = P[-1.118 < Z < 1.677] = P[0 < Z < 1.118] + P[0 < Z < 1.677] = 0.8215$

(c)

$$\begin{aligned} P[500 - \Delta_X < X < 500 + \Delta_X] &= 0,90 \\ z = \frac{(500 + \Delta_X) - 500}{20} &= 1.64 \\ \Delta_X &= 32.9 \\ P[467.1 < X < 532.9] &= 0,90 \end{aligned}$$

(d)

$$\begin{aligned} P[500 - \Delta_{\overline{X}} < \overline{X} < 500 + \Delta_{\overline{X}}] &= 0,90 \\ z = \frac{(500 + \Delta_{\overline{X}}) - 500}{20/\sqrt{5}} &= 1.64 \\ \Delta_{\overline{X}} &= 14.71 \\ P[485.3 < \overline{X} < 514.7] &= 0,90 \end{aligned}$$

(e) $P[\overline{X} > 510] = P[Z > \frac{510-500}{20/\sqrt{5}}] = P[Z > 1.118] = 0,5 - P[0 < Z < 1.118] = 0,5 - 0.368 = 0.132$

(f) $P[\overline{X} > 512] = P[Z > \frac{512-500}{20}] = P[Z > 1.342] = 0,5 - P[0 < Z < 1.342] = 0,5 - 0.41 = 0.0899$

(g)

$$\begin{aligned} P[\overline{X} < \overline{x}_c] &= 0,02 \\ z = \frac{\overline{x}_c - 500}{20/\sqrt{5}} &= 2.05 \\ \overline{x}_c &= 518.4 \end{aligned}$$

(h)

$$\begin{aligned} P[\overline{X}_n < 512] &= 0,02 \\ z = \frac{512 - 500}{20/\sqrt{n}} &= 2.05 \\ n = \left\lceil \frac{2.05^2 20^2}{(512 - 500)^2} \right. &= 12 \end{aligned}$$

(i) $P[\overline{X} < 500|\mu = 515] = P[Z < \frac{500-515}{20/\sqrt{5}}] = P[Z < -1.677] = 0,5 - P[0 < Z < 1.677] = 0,5 - 0.4532 = 0.04677$

(j) $P[\overline{X} < 512|\mu = 515] = P[Z < \frac{512-515}{20/\sqrt{5}}] = P[Z < -0.3354] = 0,5 - P[0 < Z < 0.3354] = 0,5 - 0.1313 = 0.3687$

Solução Computacional:

```
> (it.a <- diff(pnorm(c(490,515), m=500, sd=20)))

[1] 0.4648

> (it.b <- diff(pnorm(c(490,515), m=500, sd=20/sqrt(5))))

[1] 0.8215

> (it.c <- qnorm(c(0.05,0.95), m=500, sd=20))

[1] 467.1 532.9

> (it.d <- qnorm(c(0.05,0.95), m=500, sd=20/sqrt(5)))

[1] 485.3 514.7

> (it.e <- pnorm(510, m=500, sd=20/sqrt(5), low=F))

[1] 0.1318

> (it.f <- pnorm(512, m=500, sd=20/sqrt(5), low=F))

[1] 0.08986

> (it.g <- qnorm(0.98, m=500, sd=20/sqrt(5)))

[1] 518.4

> (it.h <- ceiling((qnorm(0.98)^2)*20^2/(512-500)^2))

[1] 12

> (it.i <- pnorm(500, m=515, sd=20/sqrt(5)))

[1] 0.04677

> (it.j <- pnorm(512, m=515, sd=20/sqrt(5)))

[1] 0.3687
```