

# CE-003: Estatística II - Turma AMB - Avaliações Semanais (1º 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<sup>1</sup> sobre indicadores sociais em 97 países. Os atributos<sup>2</sup> 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 capita anual e *Regiao*: região geográfica sendo consideradas: "EUOr"(Europa Oriental),"SA"(América Latina e México),"PM"("Primeiro Mundo"),"OrMd"(Oriente Médio), "Asia"e "Africa". A renda per capita 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.

<sup>1</sup><http://www.amstat.org/publications/jse/datasets/poverty.dat.txt>

<sup>2</sup><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
```

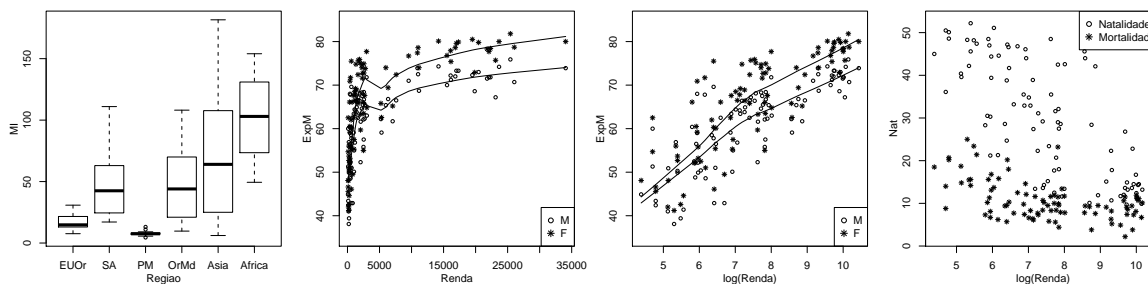
```
$` (500,2e+03]`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  13.4   24.4   32.9   31.8   39.6   47.2
```

```
$` (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
```

GrupoRenda	Regiao					
	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

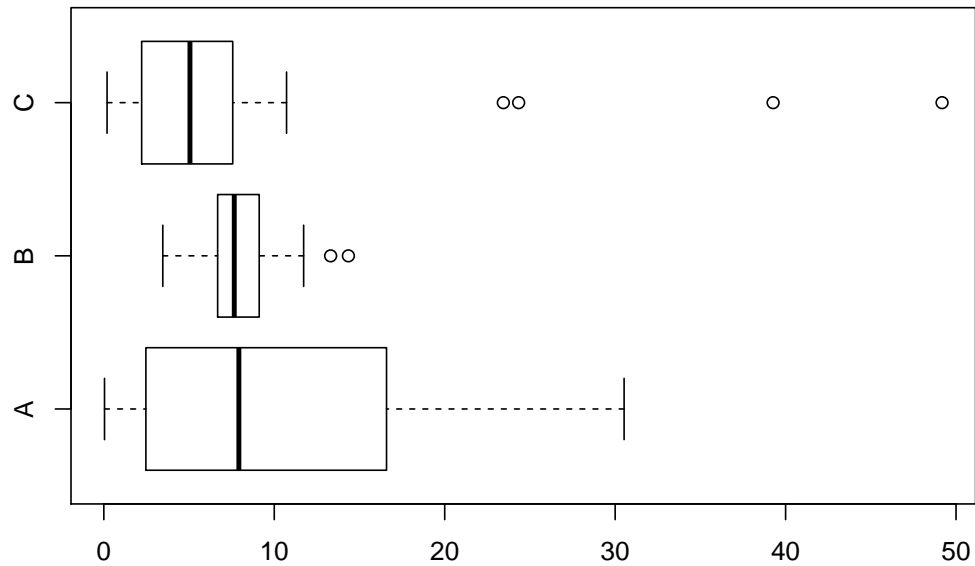
X-squared  
87.64



	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.  $\bar{x} = 39.44$
- ii.  $md(x) = 30 + \frac{10 \cdot (22,5 - 5)}{20} = 38.75$

(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.

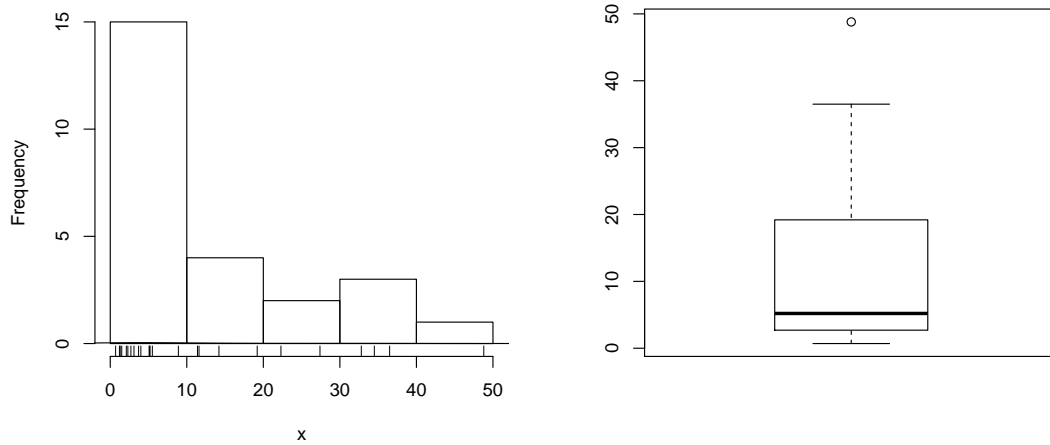


Figura 1: (d) histograma (esquerda) e (e) *box-plot* (direita) dos dados

**Solução:**

- i.  $\bar{x} = 12.44$  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. saírem dois números iguais,
  - ii. o produto dos números que saíram ser ímpar,
  - iii. o produto dos números que saíram 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 saíram dois números iguais.

**Solução:**

$X_1$  : resultado do primeiro dado     $X_2$  : resultado do segundo dado

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

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

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.

evento  $B : X_1 \cdot X_2$  é ímpar

$$P[B] = \frac{9}{36} = \frac{1}{4} = 0.25$$

iii.

evento  $C_1 : X_1 \cdot X_2$  é ímpar

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,

evento  $D_1 : X_1 + X_2 \geq 7$

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 saíram dois números iguais.

evento  $E_1 : X_1 + X_2 \geq 7$

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:**

$A_i$  : acerta a  $i$ -ésima questão  $i = 1, \dots, 4$

$$P(A_i) = 0,2 \quad P(\bar{A}_i) = 0,8 \quad \forall i$$

$$\begin{aligned} P(\text{acertar alguma}) &= 1 - P(\text{errar todas}) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3 \cap \bar{A}_4) \stackrel{\text{ind}}{=} 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) \cdot P(\bar{A}_3) \cdot P(\bar{A}_4) = \\ &= 1 - (0,8)^4 = 0.59 \end{aligned}$$

- 
- (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:

$$P[A] = 0,05 \quad ; \quad P[\bar{A}] = 0,95$$

$$P[B] = 0,065 \quad ; \quad P[\bar{B}] = 0,935$$

$$P[C] = 0,12 \quad ; \quad P[\bar{C}] = 0,88$$

$$P[\text{contaminação}] = P[A \cup B \cup C] = 1 - P[\bar{A} \cap \bar{B} \cap \bar{C}] \stackrel{\text{ind}}{=} 1 - P[\bar{A}] \cdot P[\bar{B}] \cdot P[\bar{C}] = 1 - 0.95 \cdot 0.935 \cdot 0.880.2183$$

- 
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.  $\Omega = (\overline{AAAA}), (\overline{AAAA}), (\overline{AAAA}), \dots, (AAAA), (AAAA)$   $n(\Omega) = 2^4 = 16$
- 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.  $X \sim B(n = 4, p = 0,2)$   $P[X = x] = \binom{4}{x}(0,2)^x(1 - 0,8)^{4-x}$
- 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 esperado de  $X$ .

**Solução:**

i.

$$f(x) \geq 0 \forall x \int_0^2 f(x)dx = 1$$

$$\frac{3}{8} \frac{2^3 - 0^3}{3} = 1$$

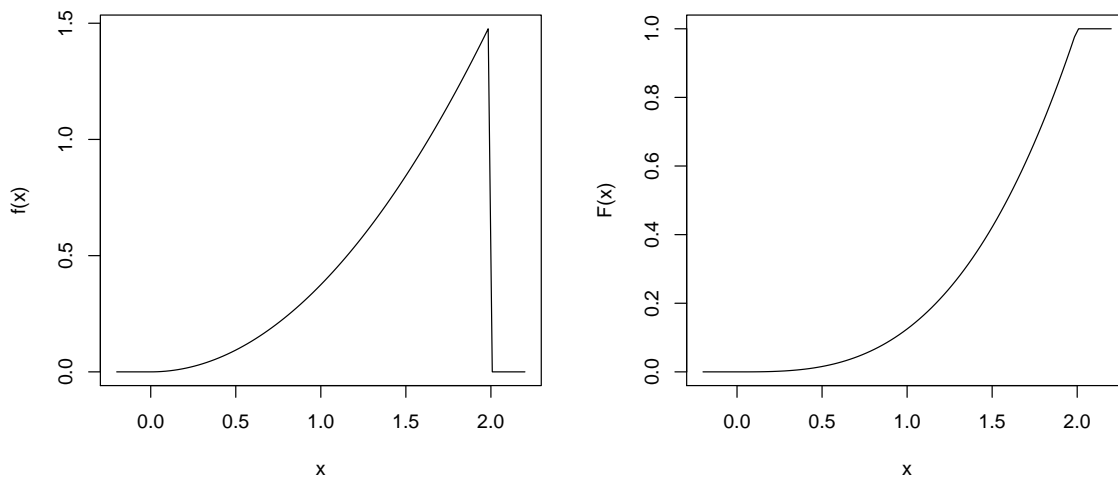
$$F(x) = \int_0^x f(x)dx = \frac{3}{8} \frac{1}{3} (x^3 - 0^3) = \frac{x^3}{8}$$

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

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

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 biólogo *A* possuía 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.

- Qual a probabilidade do biólogo *A* encontrar ao menos 2 animais sensíveis?
- Qual a probabilidade do biólogo *B* precisar testar no máximo 6 animais?
- Qual a probabilidade do biólogo *C* encontrar menos que dois sensíveis em um determinado dia?
- Qual a probabilidade do biólogo *D* encontrar mais que 3 animais sensíveis?
- Qual a probabilidade do biólogo *E* precisar testar mais que 3 animais?

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

**Solução:**

(a)

$X_a$  : número de sensíveis entre os 8 selecionados

$$x_a \in \{0, 1, 2, \dots, 8\}$$

$$X_a \sim HG(N = 30, K = 10, n = 8)$$

$$P[X_a = x] = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} = \frac{\binom{10}{x} \binom{20}{8-x}}{\binom{30}{8}}$$

$$P[X_a \geq 2] = 1 - P[X_a \leq 1] = 1 - (P[X_a = 0] + P[X_a = 1]) = 0.846$$

(b)

$X_b$  : número de não sensíveis examinados até encontrar o terceiro sensível

$$x_b \in \{0, 1, \dots\}$$

$$X_b \sim BN(k = 3, p = 0, 1)$$

$$P[X_b = x] = \binom{x+k-1}{x} p^k (1-p)^x = \binom{x+2}{x} 0,1^3 0,9^x$$

$$P[X_b \leq 3] = P[X_b = 0] + P[X_b = 1] + P[X_b = 2] + P[X_b = 3] = 0.016$$

(c)

$X_c$  : número de sensíveis encontrados em um dia

$$x_c \in \{0, 1, \dots\}$$

$$X_c \sim P(\lambda = 2, 8)$$

$$P[X_c = x] = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-2,8} 2,8^x}{x!}$$

$$P[X_c < 2] = P[X_c = 0] + P[X_c = 1] = 0.231$$

(d)

$X_d$  : número de sensíveis entre os 10 examinados

$$x_d \in \{0, 1, \dots, 10\}$$

$$X_d \sim B(n = 10, p = 0, 1)$$

$$P[X_d = x] = \binom{n}{x} p^x (1-p)^{n-x} = \binom{10}{x} 0,1^x 0,9^{10-x}$$

$$P[X_d > 3] = 1 - P[X_d \leq 3] = 1 - (P[X_d = 0] + P[X_d = 1] + P[X_d = 2] + P[X_d = 3]) = 0.013$$

(e)

$X_e$  : número de não sensíveis examinados até encontrar o primeiro sensível

$$x_e \in \{0, 1, \dots\}$$

$$X_e \sim G(p = 0, 1)$$

$$P[X_e = x] = p(1-p)^x = 0,1 \cdot 0,9^x$$

$$P[X_e > 3] = 1 - P[X_e \leq 3] = 1 - (P[X_e = 0] + P[X_e = 1] + P[X_e = 2] + P[X_e = 3]) = 0.656$$