CE-003: Estatística II - Turma AMB/K/O
Avaliações Semanais (1^o semestre 2014)

Semana 3

1. Dois jogadores (A e B) vão jogar um jogo que consiste no lançamento de dois dados. Ambos começam com R$ 10,00. Se a soma dos dados for um número ímpar, A paga R$ 1,00 para B. Se a soma for par, B paga R$ 1,00 para A.
   (a) Quais os possíveis valores em dinheiro que os jogadores podem ter após 2 rodadas? A chance é a mesma para todos esses possíveis valores?
   (b) Quais os possíveis valores em dinheiro que os jogadores podem ter após 3 rodadas? A chance é a mesma para todos esses possíveis valores?
   (c) O jogo é honesto?

Solução:
No lançamento de dos dados exixtem 36 resultados possíveis $((1,1),(1,2),\ldots,(6,6))$ todos com a mesma probabilidade de ocorrência. As possíveis somas e respectivas probabilidades são:

S(soma)	2	3	4	5	6	7	8	9	10	11	12
$P[S=s]$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

O que resulta nas probabilidades de par e ímpar:

S(soma)	par	ímpar
$P[S=s]$	$\frac{18}{36}=\frac{1}{2}$	$\frac{18}{36}=\frac{1}{2}$

(a) Denotanto por A e B as vitórias de cada um destes jogadores os possíveis resultados são:

$$\Omega = (A,A),(A,B),(B,A),(B,B)$$

e como $P[A] = P[par] = P[\text{ímpar}] = P[B] = 1/2$, cada um dos possíveis resultados em Ω possui probabilidade 1/2. Denotando por X_A o saldo de A após duas rodadas (lembrando que $X_B = 20 - X_A$), temos que os possíveis valores e suas probabilidades após duas rodadas são:

x_A	8	10	12
$P[X_A = x_A]$	$\frac{1}{4}$	$\frac{2}{4}=\frac{1}{2}$	$\frac{1}{4}$

(b) De forma análoga à anterior, os possíveis resultados são:

$$\Omega = (A,A,A),(A,A,B),(A,B,A),(B,A,A),(A,B,B),(B,A,B),(B,B,A),(B,B,B)$$

e cada ponto do espeço amostral possui probabilidade 1/8. Os possíveis valores de X_A e suas probabilidades após duas rodadas são:

x_A	7	9	11	13
$P[X_A = x_A]$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

(c) Sim, a probabilidade de *par* é igual a de *mpar* o que resultou em probabilidades iguais para os ganhos possíveis dos jogadores.

As probabilidades poderiam ser ainda obtidas definindo-se:

$$
\begin{aligned}
& Y_B \text{número de vitórias de B} \\
& y_B \in \{0, 1, 2, 3\} \\
& Y_B \sim \text{Bin}(n, p = 1/2) \qquad \text{em que } n = 2 \text{ ou } 3 \text{ é o número de rodadas} \\
P[Y_B = y_B] = & \binom{n}{y_B} p^{y_B} (1-p)^{n-y_B}
\end{aligned}
$$

---

2. A cada rodada de um experimento sobre funções cognitivas, o animal recebe um estímulo e depois faz uma escolha entre duas opções, sendo que em uma deles ele recebe um "prêmio" e em outra não. O interesse está no resultados dos "acertos" do animal (ou seja se ele recebe ou não o "prêmio"). Considere três rodadas deste experimento.
   (a) Qual o espaço amostral do experimento?
   (b) Qual a probabilidade associada a cada ponto do espaço amostral?
   (c) Qual(ais) as suposições feitas para calcular estas probabilidades?
   (d) Qual a probabilidade do animal receber algum "prêmio" durante as três rodadas?
   (e) Defina algum evento simples e algum evento composto e forneça as probabilidades destes eventos.

Solução:
Notação:

$$A: \text{o animal recebe o prêmio} \quad P[A] = 0,50 \quad P[\overline{A}] = 0,50$$

(a) $\Omega = \{(A, A, A), (A, A, \overline{A}), (A, \overline{A}, A), (\overline{A}, A, A), (A, \overline{A}, \overline{A}), (\overline{A}, A, \overline{A}), (\overline{A}, \overline{A}, A), (\overline{A}, \overline{A}, \overline{A})\}$
(b) todos tem a mesma probabilidade, $P[(A, A, A)] = P[(A, A, \overline{A})] = \ldots = P[(\overline{A}, \overline{A}, \overline{A})] = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$
(c) Os resultados das três rodadas são independentes, a probabilidade de prêmio permanece constnate e igual a 1/2 em cada rodada.
(d) $P[\text{algum prêmio}] = 1 - P[\text{nenhum prêmio}] = 1 - P[(\overline{A}, \overline{A}, \overline{A})] = 1 - 1/8 = 7/8$
(e) Há várias possibilidades, o que se seguem são apenas possíveis opções:

$$
\begin{aligned}
& \text{E}_1 (\text{evento simples}) : \text{o animal recebe prêmio em todas as tentativas} \\
& \qquad P[\text{E}_1] = \frac{1}{8} \\
& \text{E}_2 (\text{evento simples}) : \text{o animal recebe prêmio em duas tentativas} \\
& \qquad P[\text{E}_2] = P[(A, A, \overline{A})] + P[(A, \overline{A}, A)] + P[(\overline{A}, A, A)] = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}
\end{aligned}
$$

Semana 4

1. Considere que um componente de um sistema de reconhecimento é composto de um código de três algarismos que podem ser fornecidos em qualquer ordem. O reconhecimento é positivo se ao menos dois algarismos corretos são fornecidos. Considere que são feitas tentativas aleatórias, sendo que cada tentativa consiste em fornecer três algarismos (de 0 a 9).
   (a) Qual a probabilidade do reconhecimento ser positivo em uma tentativa qualquer?
   (b) Qual a probabilidade de conseguir reconhecimento positivo em duas tentativas consecutivas?
   (c) Qual a probabilidade de conseguir reconhecimento positivo em exatamente duas entre 10 tentativas?
   (d) Se são feitas tentativas até conseguir o primeiro reconhecimento positivo, qual a probabilidade de que sejam necessárias mais do que quatro tentativas?
   (e) Voce consegue identificar distribuições de probabilidades discretas discutidas em aula os itens anteriores? Qual(ais)?

Solução:
O problema pode ser resolvido diretamente utilizando distribuições de probabilidades "conhecidas", mas vamos aqui detalhar a solução.
Denotando por C: *algarismo correto é fornecido* e $\bar{C}$: *algarismo incorreto é fornecido*, o espaço amostral do experimento definido por uma tentativa é:

$$\Omega = \{(\bar{C},\bar{C},\bar{C}),(\bar{C},\bar{C},C),(\bar{C},C,\bar{C}),(C,\bar{C},\bar{C}),(\bar{C},C,C),(C,\bar{C},C),(C,C,\bar{C}),(C,C,C)\}.$$

As probabilidades associadas a cada ponto podem ser obtidas considerando que $P[C] = 0,1$ e independência. Desta forma $P[(\bar{C},\bar{C},C)] = 0,9 \cdot 0,9 \cdot 0,1$ e de forma análoga para demais.

(a) E_1 : reconhecimento ser positivo em uma tentativa

$$pR = P[E_1] = P[(C,C,\bar{C})]+P[(C,\bar{C},C)]+P[(\bar{C},C,C)]+P[(C,C,C)] = 3\cdot(0,1)^2(0,9)+(0,1)^3 = 0.028$$

(b)

$$P[E_1 \cap E_1] \overset{ind}{=} 0.01^2 = 0.000784$$

(c) E_3: reconhecimento positivo em exatamente duas entre 10 tentativas

$$P[E_3] = \binom{10}{2}(pR)^2(1-pR)^{10-2} = 0.02811$$

(d) E_4: mais do que quatro tentativas até primeiro reconhecimento positivo.

$$P[E_3] = 1 - P[\bar{E}_3] = 1 - (pR + (1-pR)pR + (1-pR)^2pR + (1-pR)^3pR) = 0.89262$$

(e) Distribuição Binomial para (i) número de algarismos corretos (ii) número de tentativas positivas. Distribuição Geomérica para número de tentativas até o primeiro reconhecimento.

Utilizando distribuições de probabilidades "conhecidas" teríamos as soluções apresentadas a seguir. Definindo as variáveis aleatórias:

X_1 : número de algarísmos corretos em uma tentativa	$X_1 \sim \mathrm{B}(n=3, p=1/10)$
X_2 : número de tentativas corretas em 2 tentativas	$X_2 \sim \mathrm{B}(n=2, p=P[X_1 \geq 2])$
X_3 : número de tentativas corretas em 10 tentativas	$X_3 \sim \mathrm{B}(n=2, p=P[X_1 \geq 2])$
X_4 : número de não reconhecimentos até o primeiro positivo	$X_4 \sim \mathrm{G}(p=P[X_1 \geq 2)$

E os itens anteriores seriam respondidos como:

(a) $P[X_1 \geq 2] = P[X_1 = 2] + P[X_1 = 3] = 0.027 + 0.001 = 0.028$
(b) $P[X_2 = 2] = 0.000784$
(c) $P[X_3 = 2] = 0.02811$
(d) $P[X_4 \geq 4] = 1 - P[X \leq 3] = 1 - 0.10738 = 0.89262$

Semana 5

1. Considere um estudo de padrões em texto no qual a ocorrência de algumas "expressões chave" é verificada. Responda cada um dos itens abaixo, identificando a variável aleatória envolvida e a respectiva distribuição de probabilidades.
   (a) Sabe-se que uma certa expressão E_1 ocorre em média 1,4 vezes por página. Qual a probabilidade de tal expressão
      i. não ser encontrada em uma determinada página?
      ii. ocorrer ao menos três vezes em duas páginas?
   (b) Uma outra expressão E_2 ocorre em 10% das páginas. Qual a probabilidade de:
      i. não ser encontrada em cinco páginas escolhidas ao acaso?
      ii. ser encontrada em ao menos uma de cinco páginas escolhidas ao acaso?

(c) Adota-se a estratégia de inspecionar páginas uma a uma até encontrar a primeira ocorrência da expressão E_2.

   i. Quantas páginas sem a ocorrência de E_2 espera-se encontrar?

   ii. Qual a probabilidade de serem encontradas mais que cinco páginas sem a ocorrência de E_2

(d) Em um texto completo de 80 páginas sabe-se que 30 apresentam algum erro no uso de expressões investigadas e as demais páginas não apresentam nenhum erro.

   i. Sendo escolhidas dez páginas do texto ao acaso, qual a probabilidade de que nenhuma delas contenha erro no uso das expressões?

   ii. Idem anterior para 15 páginas de texto.

(e) As páginas do texto serão inspecionadas até que se encontre a terceira página com algum erro.

   i. Qual a probabilidade que de sejam "varridas" pelo menos dez páginas?

   ii. Quantas páginas espera-se inspecionar antes de encontrar a terceira com erro?

   iii. Qual seria a probabilidade "de algo oposto", de ser necessário varrer mais de cinco páginas até que se encontre a terceira sem erro algum?

Solução:

(a)

$$X : \text{número de ocorrências de } E_1 \text{ em uma página}$$
$$X \sim \mathrm{P}(\lambda = 1,4)$$
$$P[X = x] = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}$$

   i. $P[X = 0] = \frac{\mathrm{e}^{-1,4}1,4^0}{0!} = 0.247$

   ii.

$$X : \text{número de ocorrências de } E_1 \text{ em duas página}$$
$$X \sim \mathrm{P}(\lambda = 2,8)$$
$$P[X \geq] = 1 - (P[X = 0] + P[X = 1] + P[X = 2]) = 0.167$$

(b)

$$X : \text{número de páginas em que } E_2 \text{ ocorre dentre cinco páginas}$$
$$X \sim \mathrm{B}(n = 5, p = 0,10)$$
$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}$$

   i. $P[X = 0] = \binom{5}{0}(0,1)^0(1-0,1)^{5-0} = 0.59$

   ii. $P[X \geq 1] = 1 - P[X = 0] = 0.41$

(c)

$$X : \text{número de páginas insecionadas sem encontrar a ocorrência de } E_2\text{, antes de encontrar a primeira}$$
$$X \sim \mathrm{G}(p = 0,10)$$
$$P[X = x] = (1-p)^x p$$

   i. $E[X] = \frac{1-p}{p} = \frac{1-0,1}{0,1} = 9$

   ii. $P[X > 5] = 1 - P[X \leq 5] = 0.531$

(d)

$$X : \text{número de página com erros em n páginas de texto verificadas}$$
$$X \sim \mathrm{HG}(N = 80, K = 30, n)$$
$$P[X = x] = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

i. $P[X=0|n=10] = \frac{\binom{30}{0}\binom{80-30}{10-0}}{\binom{80}{10}} = 0.00624$

ii. $P[X=0|n=15] = \frac{\binom{30}{0}\binom{80-30}{15-0}}{\binom{80}{15}} = 0.000339$

(e) As páginas do texto serão inspecionadas até que se encontre a terceira página com algum erro.

$$X : \text{número de páginas inspecionadas sem encontrar a ocorrência de } E_2\text{, antes de encontrar a terceira}$$
$$X \sim \text{BN}(r=3, p=0,10)$$

$$P[X=x] = \binom{x+r-1}{x} p^r (1-p)^x$$

i. $P[X+3 \geq 10] = P[X \geq 7] = 0.93$

ii. $E[X] = \frac{r(1-p)}{p} = \frac{3(1-0,1)}{0,1} = 27$

iii.

$$X : \text{número de páginas inspecionadas com encontrar a ocorrência de } E_2\text{, antes de encontrar a terc}$$
$$X \sim \text{BN}(r=3, p=0,90)$$
$$P[X>5] = 1 - P[X \leq 4] = 0.000176$$

Semana 6

1. Considera-se que o número semanal de acidentes em uma certa rodovia segue uma distribuição de Poisson com média de 4,2.
   (a) Qual a probabilidade de que em uma determinada semana ocorram menos que três acidentes?
   (b) Indique como seria calculada a probabilidade de ocorrência de 10 ou mais acidentes.
   (c) Qual a probabilidade de que não ocorra acidentes em um único dia tomado ao acaso?
   (d) Qual seria a distribuição de probabilidades do número de acidentes por mês?
   (e) Qual a probabilidade de que, em quatro semanas, no máximo em uma delas ocorram três ou mais acidentes?

Solução:

$$X : \text{número semanal de acidentes}$$
$$X \sim \text{P}(\lambda = 4,2)$$
$$P[X=x] = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} = \frac{\mathrm{e}^{-4,2}4,2^x}{x!}$$

(a) $P[X \leq 3] = P[X=0] + P[X=1] + P[X=2] = 0.21$

(b) $P[X \geq 10] = 1 - P[X \leq 9] = 1 - (P[X=0] + P[X=1] + \ldots + P[X=9]) = 0.0111$

(c)

$$X_D : \text{número acidentes em 1 dia}$$
$$X_D \sim \text{P}(\lambda = 4,2/7)$$
$$P[X_D = 0] = \frac{\mathrm{e}^{-4,2/7}(4,2/7)^0}{0!} = \mathrm{e}^{-4,2/7} = 0.549$$

(d) Considerando o mês com 30 dias:

$$X_M : \text{número acidentes por mês}$$
$$X_M \sim \text{P}(\lambda = 30 \cdot 4,2/7 = 18)$$

(e)

$$Y: \text{número semanas, dentre 4 semanas, em que ocorrem 3 ou mais acidentes}$$
$$Y \sim \mathrm{B}(n=4, p=P[X \geq 3] = 0.79$$
$$P[Y=y] = \binom{n}{x} p^x (1-p)^{n-y} = \binom{4}{x} 0.79^x (1-0.79)^{4-y} P[Y \leq 1] = P[Y=0] + P[Y=1] = 0.0313$$

---

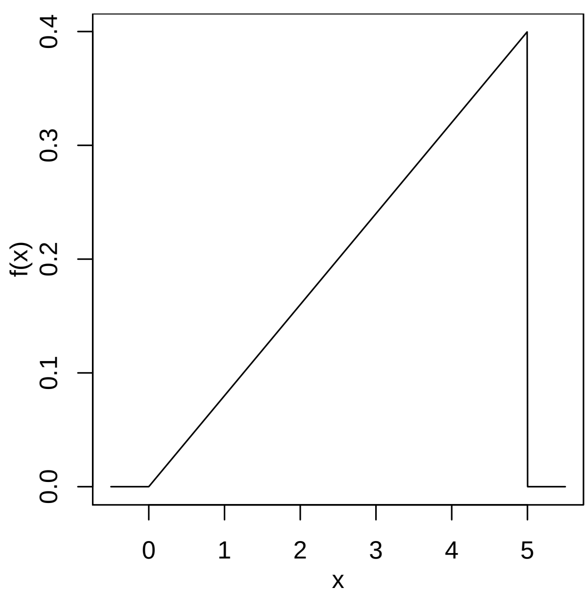
2. A função de densidade de probabilidades de uma variável aleatória X é dada por:

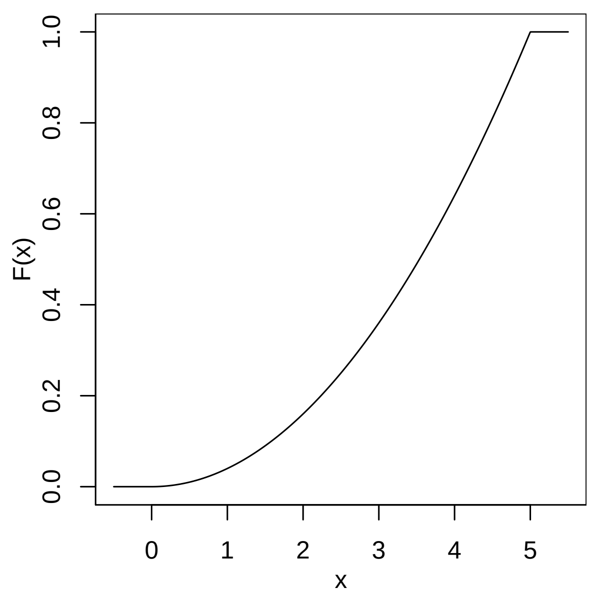
$$f(x) = \begin{cases} kx & \text{se } 0 \leq x \leq 5 \\ 0 & \text{para qualquer outro valor} \end{cases}$$

encontre:

(a) k,
(b) $P[1 \leq X \leq 3]$,
(c) $P[2 \leq X \leq 4]$,
(d) $P[X \geq 3]$,
(e) $P[X \leq 3 | X > 1]$,
(f) A expressão da função de densidade acumulada $F(x)$ e mostre como as probabilidades dos itens anteriores poderiam ser calculadas em função de $F(x)$,
(g) a média da variável X,
(h) os quartis (incluindo a mediana).

Solução:





(a)

$$Kx \geq 0, 0 \leq x \leq 5 \Rightarrow K > 0$$
$$\int_0^5 kx\mathrm{d}x = 1 \Rightarrow k\frac{5^2 - 0^2}{2} = 1 \Rightarrow K = 2/25 = 0,08$$

(b) $P[1 \leq X \leq 3] = \int_1^3 0,08x\mathrm{d}x = 0,08\frac{3^2-1^2}{2} = 0.32$
(c) $P[2 \leq X \leq 4] = \int_2^4 0,08x\mathrm{d}x = 0,08\frac{4^2-2^2}{2} = 0.48$
(d) $P[X \geq 3] = \int_3^5 0,08x\mathrm{d}x = 0,08\frac{5^2-3^2}{2} = 0.64$
(e) $P[X \leq 3 | X > 1] = \frac{P[1<X\leq 3]}{P[X>1]} = \frac{\int_1^3 0,08x\mathrm{d}x}{\int_1^5 0,08x\mathrm{d}x} = \frac{0,08\frac{3^2-1^2}{2}}{0,08\frac{5^2-1^2}{2}} = 0.333$

(f)

$$F(x) = \begin{cases} 0 & x < 0 \\ \int_0^x 0,08x\mathrm{d}x = 0,08\frac{x^2-0^2}{2} = 0,04x^2 & 0 \leq x \leq 5 \\ 1 & x > 5 \end{cases}$$

$$P[1 \leq X \leq 3] = F(3) - F(1) = 0,04(3^2 - 1^2) = 0.32$$
$$P[2 \leq X \leq 4] = F(4) - F(2) = 0,04(4^2 - 2^2) = 0.48$$
$$P[X \geq 3] = 1 - F(3) = 1 - 0,04(3^2) = 0.64$$
$$P[X \leq 3 | X > 1] = \frac{F(3) - F(1)}{1 - F(1)} = \frac{0,04(3^2 - 1^2)}{1 - 0,04(3^2)} = 0.333$$

(g) $E[X] = \int xf(x)\mathrm{d}x = \int_0^5 x0,08x\mathrm{d}x = 0,08\frac{5^3-0^3}{3} = 10/3 = 3,33$

(h)

$$M_d[X] : \int_0^{M_d} f(x)\mathrm{d} = 0,5 \Rightarrow M_d[X] = F^{-1}(0,5) = \sqrt{0,5/0,04} = 3.54$$
$$Q_1[X] : \int_0^{Q_1} f(x)\mathrm{d} = 0,25 \Rightarrow Q_1[x] = F^{-1}(0,25) = \sqrt{0,25/0,04} = 2.5$$
$$Q_3[X] : \int_0^{Q_3} f(x)\mathrm{d} = 0,75 \Rightarrow Q_3[x] = F^{-1}(0,75) = \sqrt{0,75/0,04} = 4.33$$

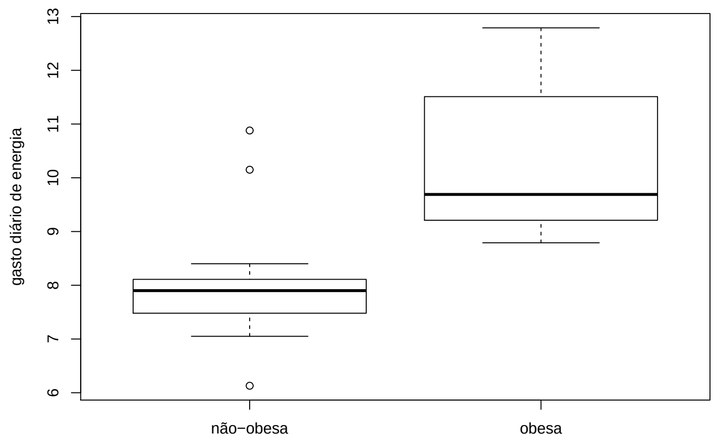
---

1. Considere a tabela de dados abaixo, que contém um extrato dos resultados da corrida de São Silvestre do ano de 2012[1]. As colunas dos dados correspondem a: 1 - classificação (geral) na prova, 2 - numeral do(a) atleta, 3 - nome do(a) atleta, 4 - idade, 5 - sexo e faixa etária para classificação por categoria de idade, 6 - equipe, 7 - tempo de prova (bruto), 8 - tempo de prova (corrigido). Considere que voce quer fazer um resumo dos resultados e também analisar algumas relações de possível interesse. Descreva ou esboce como seria o seu texto que resumiria os resultados, lembrando que o texto deverá fornecer: um perfil dos participantes e uma descrição das relações de possível interesse.

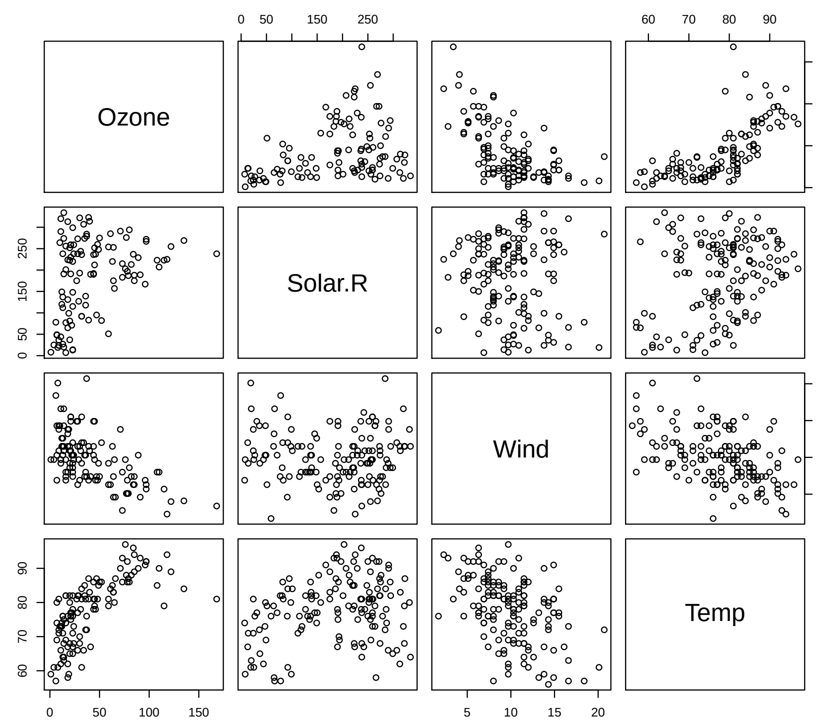
1º	223	EDWIN KIPSANG	24	M2024	COQUINHO FILA CAIXA	00:44:04	00:44:
2º	227	JOSEPH KACHAPIN APERUMOI	22	M2024	CRUZEIRO ESPORTE CLUBE	00:44:14	00:44:
3º	201	MARK KORIR	24	M2024		00:44:21	00:44:
4º	203	GIOVANI DOS SANTOS	31	M3034	PE DE VENTO CAIXA	00:44:50	00:44:
5º	231	HAFID CHANI	26	M2529	ATLAS MOUNTAIN	00:45:54	00:45:
6º	232	NAJIM EL QADY	32	M3034	ATLAS MOUNTAIN	00:46:03	00:46:
7º	224	ALPHONCE FELIX SIMBU	20	M2024	COQUINHO FILA CAIXA	00:46:05	00:46:
8º	204	UBIRATAN JOSE DOS SANTOS	31	M3034	USINA SAO JOSE	00:46:14	00:46:
9º	230	AHMED BADAY	38	M3539	ATLAS MOUNTAIN	00:46:18	00:46:
10º	234	PAULO ROBERTO DE ALMEIDA	33	M3034	CRUZEIRO CAIXA	00:46:26	00:46:
...	...	...	...	...	...	...	...
1º	20	MAURINE JELAGAT KIPCHUMBA	24	F2024	CRUZEIRO ESPORTE CLUBE	00:51:42	00:51:
2º	2	JACKLINE JUMA SAKILU	26	F2529	LUASA ESPORTE TANZANIA	00:52:11	00:52:
3º	1	RUMOKOL ELIZABEH CHEPKANAN	25	F2529	KENIA LUASA	00:52:50	00:52:
4º	19	FEKEDE ALMAZ NEGEDE	25	F2529	COQUINHO FILA CAIXA	00:53:36	00:53:
5º	18	ANASTAZIA MSANDAI MHOMI	20	F2024	COQUINHO FILA CAIXA	00:53:42	00:53:
6º	7	TATIELE ROBERTA CARVALHO	23	F2024		00:54:12	00:54:
7º	3	SUELI PEREIRA DA SILVA	35	F3539	EJA GRAN CURSO DF CAIXA	00:54:22	00:54:
8º	5	NACY JEPKOSGEI KIPRON	33	F3034	COQUINHO FILA CAIXA	00:54:43	00:54:
9º	15	ROSELAINE DE SOUSA SILVA	31	F3034	CRUZEIRO CAIXA	00:55:02	00:55:
10º	21	MARIZETE MOREIRA DO SANTOS	37	F3539	MARINHA DO BRASIL	00:55:25	00:55:
...	...	...	...	...	...	...	...

[1]Fonte: http://www.saosilvestre.com.br.

1. Um estudo[2] coletou dados de gasto diário de energia de dois grupos de mulheres classificadas como *obesas* (9 casos) e *não-obesas* (13 casos). O gráfico a seguir mostra um resumo dos resultados. Identifique as variáveis em estudo, o tipo de cada uma e discuta os resultados mostrados no gráfico. Que tipo de medidas seriam utilizadas para verificar se há relação entre as variáveis?



2. A figura a seguir mostra relações de medidas diários de qualidade do ar em Nova York coletadas entre Maio e Setembro de 1973. Foram medidos: nivel de **Ozônio** (*Ozone*), a **radiação solar** (*Solar.R*), a velocidade do **vento** (*Wind*) e a **temperatura** (*Temp*). Discuta a relação das variáveis duas a duas, indicando como qual(ais) medida(s) pode(m) ser calculada(s) para refletir a associação.



Semana 9

1. Foi feito um estudo na área de uma baía para verificar a salinidade (superficial) da água no local. No estudo foram tomadas amostras da água em 47 pontos da baía. Cada amostra foi analisada e o resultado de interesse aqui era o nível de salinidade. O valor médio das 47 amostras foi de 28,35 $^{o}/^{oo}$. Neste contexto, identifique e descreva os seguintes elementos segundo os conceitos de estatística.

[2] D.G. Altman (1991), Practical Statistics for Medical Research, Table 9.4, Chapman & Hall. Dados obtidos no pacote ISwR do R.

(a) a população,
(b) a variável aleatória de interesse,
(c) o parâmetro de interesse,
(d) a amostra,
(e) o estimador,
(f) a estimativa,
(g) a distribuição amostral.

Discuta ainda se, ou sob quais condições, a amostra pode ser considerada uma *amostra aleatória simples.*

Semana 10

1. Uma locadora de veículos que possui uma grande frota decide fazer um estudo sobre vários aspectos relacionados ao desempenho. Para isto vai tomar uma amostra aleatória de 25 de seus veículos para inspeções detalhadas. Várias características serão medidas, mas vamos aqui nos ater apenas ao consumo de combustível, supondo que a variância do consumo de toda a frota é de $6,25 km/l$.
   (a) Qual a probabilidade do consumo médio aferido nos 25 veículos, diferir do consumo médio de toda a frota em mais que $0,5 km/l$? E em mais que $1 km/l$?
   (b) Qual a margem de erro na estimação do consumo médio da frota para uma confiança de 95%?
   (c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse a metade da calculada no item anterior?
   (d) Se uma amostra (com $n = 25$) fornecer uma estimativa intervalar de $(11,1\,;\,12,3) km/l$, qual a confiança desta estimativa?
   (e) identifique no problema: a população variável aleatória de interesse, o parâmetro de interesse, o estimador, a distribuição amostral, a estimativa pontual e a estimativa intervalar.

Solução:

$$\begin{aligned}
& X : \text{consumo de veículo da frota} \\
\text{Distribuição da variável aleatória (população):} & \\
& X \sim \text{Dist.}(\mu_X = \text{E}[X], \sigma_X^2 = \text{Var}[X]) \\
\text{distribuição amostral:} & \\
& \overline{X} \approx \text{N}(\mu_{\overline{X}} = \mu_X, \sigma_{\overline{X}} = \sigma_X^2/n) \\
\text{ou, equivalentemente,} & \\
& \overline{X} - \mu \approx \text{N}(0, \sigma^2/n) \\
& \sigma_X^2 = 6,25\ ;\ \sigma_{\overline{X}}^2 = 0,25\ ;\ \sigma_{\overline{X}} = 0,5
\end{aligned}$$

(a)

$$P[|\overline{X} - \mu| > 0,5] = P[\frac{|\overline{X} - \mu|}{\sigma} > 0,5/\sigma] = P[|Z| > 1] = 0.317$$

$$P[|\overline{X} - \mu| > 1] = P[\frac{|\overline{X} - \mu|}{\sigma} > 1/\sigma] = P[|Z| > 2] = 0.0455$$

(b)

$$\begin{aligned}
ME &= z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}} \\
ME &= 1.96\frac{2,5}{\sqrt{25}} \\
ME &= 0.98
\end{aligned}$$

(c)

$$\frac{ME}{2} = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n^*}}$$
$$n^* = \lceil(\frac{2}{ME})^2 z_{0,95}^2\sigma_X^2$$
$$n^* = \lceil(\frac{2}{0.98})^2 1.96^2 6,25$$
$$n^* = 100$$

(d)

$$(11,1\,;\,12,3) \equiv 11,7 \pm 0,6$$
$$ME = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}}$$
$$0,6 = z_{1-\alpha}0,5$$
$$z_{1-\alpha} = \frac{0,6}{0,5}$$
$$1-\alpha(\text{confiança}) = 0.77(77\%)$$

(e)

Semana 11

1. Considere um estudo na qual deseja-se estimar a proporção de solicitações atendidas e resolvidas de uma central do usuário através de uma amostra aleatória simples.
   (a) Qual a população, variável aleatória e parâmetro de interesse?
   (b) Mostre como obter o estimador de máxima verossimilhança deste parâmetro.
   (c) Se a amostra for de 4000 solicitações, qual será a margem de erro (com confiança de 95%) para a estimação da proporção de resolvidas?
   (d) Para este mesmo tamanho de amostra, qual seria a confiança associada a uma margem de erro de $\pm 0,01$ (1 %)?
   (e) Qual deveria ser o tamanho da amostra para se obter a estimativa com uma margem de erro de 2,5% com 95% de confiança?
   (f) E para uma margem de erro de 3% com 99% de confiança?

Solução:

(a) População: solicitações atendidas, representada pela v.a.:

$$X : \text{resolução de solicitações atendidas}$$

que é binária e pode ser codificada assumindo valores 0, para não atendidas e, 1, para atendidas. O parâmetro de interesse é a proporção de solicitações atendidas, que representa a probabilidade p de uma solicitação ser atendida. Assume-se então que:

$$X \sim B(p)$$

e os resultados das análises se baseiam no Teorema:

$$\text{Teo 2:} \quad \hat{p} = \overline{X} \approx N(p, \frac{p(1-p)}{n})$$

do qual se extraem os resultados:

$$\hat{p} \pm z_{(1-\alpha)}\sqrt{\frac{p(1-p)}{n}}$$

$$\hat{p} \pm \text{ME}$$

$$\text{ME} = z_{(1-\alpha)}\sqrt{\frac{p(1-p)}{n}}$$

$$n = \frac{(z_{(1-\alpha)})^2}{\text{ME}^2}p(1-p)$$

$$p(1-p) \text{ é limitado superiormente para } p = 0,5$$

$$n = \frac{(z_{(1-\alpha)})^2}{\text{ME}^2}0,25$$

(b) Seja a v.a. Y : número de solicitações resolvidas dentre n atendidas.

$$Y \sim \text{Bin}(n, p))$$

A função de verossimilhança é dada pela expressão de distribuição de Y, portanto:

$$L(p) = \binom{n}{y} p^y (1-p)^(n-y)$$

e a log-verossimilhança:

$$l(p) = \log\binom{n}{y} + y\log(p) + (n-y)\log(1-p).$$

O estimador de máxima verossimilhança de p é o ponto de máximo função (de p) acima e portanto encontrado derivando-se a expressão acima,

$$\frac{\mathrm{d}l(p)}{\mathrm{d}p} = \frac{y}{p} - \frac{n-y}{1-p}$$

que igualada a zero nos leva a:

$$\hat{p} = \frac{y}{n}.$$

Note-se que a derivada segunda fornece um valor negativo e valor $\hat{p}$ é então o estimador de máxima verossimilhança do parâmetro p.

(c) $M.E. = z_{(1-\alpha)}\sqrt{\frac{p(1-p)}{n}} = 1.96\sqrt{\frac{0,5(1-0,5)}{4000}} = 0.0155$

(d)

$$M.E. = z_{(1-\alpha)}\sqrt{\frac{p(1-p)}{n}}$$

$$z_{(1-\alpha)} = 0,01/\sqrt{\frac{0,5(1-0,5)}{4000}} = 1.26$$

$$1-\alpha = 79.4\%$$

(e) $n = \frac{(1.96)^2}{(0,025)^2}0,25 = 1537$

(f) $n = \frac{(2.576)^2}{(0,03)^2}0,25 = 1844$