CE-003: Estatística II - Turma: K/O, Avaliações Periódica, 2o sem. 2017

GRR: ________________ **Nome:** ____ **Turma:** ________

Avaliação 01

1. Uma *playlist* vai ser montada com três músicas selecionadas a partir de uma lista de quatro músicas do(a) artista A, outra lista de três de B e outra de duas de C. A sequência de músicas na *playlist* é montada ao acaso porém não repete músicas e tem uma de cada artista sempre na sequência de A, B e C nesta ordem..
   (a) Explique se e por que a composição da *playlist* pode ser considerada um experimento aleatório.
   (b) Forneça o espaço amostral.
   (c) Caracterize o espaço amostral quanto a ser (i) finito ou infinito, (ii) enumerável ou não enumerável, (iii) equiprovável ou não equiprovável, justificando as respostas.
   (d) Considere o evento "a *playlist* inicia com a segunda música do(a) artista A". Qual o conjunto que define este evento e qual a sua probabilidade de ocorrência?
   (e) Considere o evento "a *playlist* não contém as primeiras músicas das listas de nenhum dos(as) artistas". Qual o conjunto que define este evento e qua a sua probabilidade de ocorrência?
   (f) Qual seria a probabilidade de ocorrência de ambos eventos definidos no itens eventos anteriores?
   (g) E qual seria a probabilidade de ocorrência de algum deles?
   (h) Quantas *playlists* seria possíveis se a ordem dos(as) artistas também fosse tomada ao acaso?
   (i) Quantas *playlists* seriam possíveis se fosse permitido o sorteio de mais de uma música do mesmo artista, porém ainda sem repetição de música?
   (j) Neste caso, qual seria a probabilidade da *playlist* não conter uma música do artista A?

   Solução:
   Notação:

$$\begin{aligned} A1, A2, A3 \text{ e } A4 &: \text{ músicas do(a) artista } A \\ B1, B2, \text{ e } B3 &: \text{ músicas do(a) artista } B \\ C1 \text{ e } C2 &: \text{ músicas do(a) artista } C \end{aligned}$$

   (a) Sim, pelo fato da ordem dos artistas e músicas de cada um ser escolhada ao acaso.
   (b) $\Omega_1 = \{(A1, B1, C1), (A1, B1, C2), (A1, B2, C1), (A1, B2, C2), (A1, B3, C1), (A1, B3, C2),$
   $(A2, B1, C1), (A2, B1, C2), (A2, B2, C1), (A2, B2, C2), (A2, B3, C1), (A2, B3, C2),$
   $(A3, B1, C1), (A3, B1, C2), (A3, B2, C1), (A3, B2, C2), (A3, B3, C1), (A3, B3, C2),$
   $(A4, B1, C1), (A4, B1, C2), (A4, B2, C1), (A4, B2, C2), (A4, B3, C1), (A4, B3, C2)\}$
   $n(\Omega_1) = 4 \cdot 3 \cdot 2 = 24$
   (c) Finito, enumerável e equiprovável. (Justificativa)
   (d) $E_1 = \{(A2, B1, C1), (A2, B1, C2), (A2, B2, C1), (A2, B2, C2), (A2, B3, C1), (A2, B3, C2)\}$
   $P[E_1] = n(E_1)/n(\Omega_1) = 6/24 = 1/4 = 0,25$
   (e) $E_2 = \{(A2, B2, C2), (A2, B3, C2), (A3, B2, C2), (A3, B3, C2), (A4, B2, C2), (A4, B3, C2)\}$
   $P[E_2] = n(E_2)/n(\Omega_1) = 6/24 = 1/4 = 0,25$
   (f) $P[E_1 \cap E_2] = 2/24 = 0.0833$
   (g) $P[E_1 \cup E_2] = P[E_1] + P[E_2] - P[E_1 \cap E_2] = 10/24 = 0.417$
   (h) $n(\Omega_2) = 24 * 6 = 144$
   (i) $n(\Omega_3) = 9 * 8 * 7 = 504$
   (j) $P[\overline{A}|\Omega_3] = (5 * 4)/504 = 0.0397$

---

1. Três algorítmos diferentes vão ser testados para a classificação do diagnóstico baseado em exames e imagens. Cada algorítmo pode acertar o diagnóstico da presença de certa doença e sabe-se que até o momento as taxas de acerto são de 85, 90 e 70%. Uma imagem/exames de um paciente com a doença é fornecida aos três algorítmos e avalia-se o acerto de cada um deles.

(a) Forneça o espaço amostral.
(b) Caracterize o espaço amostral quanto a ser (i) finito ou infinito, (ii) enumerável ou não enumerável, (iii) equiprovável ou não equiprovável, justificando as respostas.
(c) Defina dois eventos e forneça suas probabilidades.
(d) Qual a probabilidade de que A acerte o diagnóstico ou que apenas um algorítmo acerte?
(e) Qual a probabilidade de que A tenha acertado o diagnóstico sabendo que apenas um algorítmo acertou?
(f) Qual a probabilidade de que nem A nem B acertem o diagnóstico?
(g) Qual a probabilidade da doença ser detectada?
(h) Qual a probabilidade de B acertar o diagnóstico sabendo que a doença foi detectada?
(i) Defina uma variável aleatória sobre este espaço amostral e indique seus possíveis valores.
(j) Obtenha a distribuição de probabilidades da variável aleatória.

Solução:
Notação:

$$\begin{aligned}
A &: \text{ o algorítmo A acerta o diagnóstico } ; \overline{A} :: \text{ o algorítmo A erra o diagnóstico} \\
P[A] = 0,85 &\quad P[\overline{A}] = 0,15 \\
B &: \text{ o algorítmo B acerta o diagnóstico } ; \overline{B} :: \text{ o algorítmo B erra o diagnóstico} \\
P[B] = 0,90 &\quad P[\overline{B}] = 0,10 \\
C &: \text{ o algorítmo C acerta o diagnóstico } ; \overline{C} :: \text{ o algorítmo C erra o diagnóstico} \\
P[C] = 0,70 &\quad P[\overline{C}] = 0,30
\end{aligned}$$

(a) $\Omega = \{(A,B,C),(\overline{A},B,C),(A,\overline{B},C),(A,B,\overline{C}),(\overline{A},\overline{B},C),(\overline{A},B,\overline{C}),(A,\overline{B},\overline{C}),(\overline{A},\overline{B},\overline{C})\}$
(b) Finito, enumerável e não equiprovável. (Justificativa)
(c)

$$\begin{aligned}
E_1 &: \text{ A acerta o diagnóstico } = \{(A,B,C),(A,\overline{B},C),(A,B,\overline{C}),(A,\overline{B},\overline{C})\} \\
P[E_1] &= 0,85 \cdot 0,90 \cdot 0,70 + 0,85 \cdot 0,10 \cdot 0,70 + 0,85 \cdot 0,90 \cdot 0,30 + 0,85 \cdot 0,10 \cdot 0,30 = 0.85 \\
E_2 &: \text{ B erra o diagnóstico } = \{(A,\overline{B},C),(\overline{A},\overline{B},C),(A,\overline{B},\overline{C}),(\overline{A},\overline{B},\overline{C})\} \\
P[E_2] &= 0,85 \cdot 0,10 \cdot 0,70 + 0,15 \cdot 0,10 \cdot 0,70 + 0,85 \cdot 0,10 \cdot 0,30 + 0,15 \cdot 0,10 \cdot 0,30 = 0.304
\end{aligned}$$

(d)

$$\begin{aligned}
E_3 &: \text{ apenas um algorítmo acerta o diagnóstico } = \{(A,\overline{B},\overline{C}),(\overline{A},B,\overline{C}),(\overline{A},\overline{B},C)\} \\
P[E_3] &= 0,85 \cdot 0,10 \cdot 0,30 + 0,15 \cdot 0,90 \cdot 0,30 + 0,15 \cdot 0,10 \cdot 0,70 = 0.0765 \\
E_1 \cap E_3 &= \{(A,\overline{B},\overline{C})\} \ ; \quad P[E_1 \cap E_3] = 0,85 \cdot 0,10 \cdot 0,30 = 0.0255 \\
P[E_1 \cup E_3] &= P[E_1] + P[E_3] - P[E_1 \cap E_3] = 0.85 + 0.0765 - 0.0255 = 0.901
\end{aligned}$$

(e) $P[E_1|E_3] = \frac{P[E_1 \cap E_3]}{P[E_3]} = \frac{0.0255}{0.0765} = 0.333$
(f)

$$\begin{aligned}
E_4 &: \text{ nem A nem B acertam o diagnóstico } = \{(\overline{A},\overline{B},C),(\overline{A},\overline{B},\overline{C})\} \\
P[E_4] &= 0,15 \cdot 0,10 \cdot 0,70 + 0,15 \cdot 0,10 \cdot 0,30 = 0.015 \\
\text{note que: } & P[\overline{A}] \cdot P[\overline{B}] = 0,15 \cdot 0,10 = 0.015 (independentes)
\end{aligned}$$

(g)

$$\begin{aligned}
E_5 &: \text{ da doença ser detectada} \\
P[E_5] &= 1 - P[\overline{E}_5] = 1 - 0,15 \cdot 0,10 \cdot 0,30 = 0.996
\end{aligned}$$

(h)

$$\begin{aligned}
P[B] &= 1 - P[E_2] \\
B \cap E_5 &= \{(A,B,C),(\overline{A},B,C),(A,B,\overline{C}),(\overline{A},B,\overline{C})\} \\
P[B \cap E_5] &= 0,85 \cdot 0,90 \cdot 0.70 + 0,15 \cdot 0,90 \cdot 0.70 + 0,85 \cdot 0,90 \cdot 0.30 + 0,15 \cdot 0,90 \cdot 0.30 = 0.9 \\
P[B|E_5] &= \frac{P[B \cap E_5]}{P[E_5]} = \frac{0.9}{0.996} = 0.904
\end{aligned}$$

(i)

$$X: \text{ número de algorítmos que acertam o diagnóstico}$$
$$x \in \{0, 1, 2, 3\}$$

(j)

x	0	1	2	3
P[X=x]	0.0045	0.0765	0.383	0.535

Avaliação 03

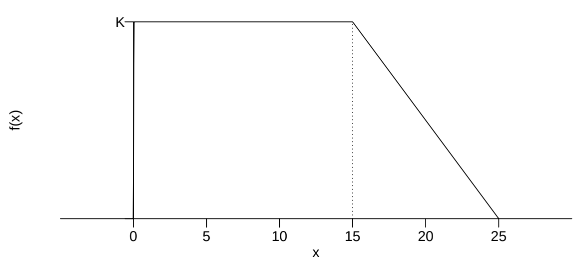
1. A localização de ocorrências em um trecho de 25 km de rodovia é considerada ser uma variável aleatória com função de densidade de probabilidade dada por:

$$f(x) = \begin{cases} K & \text{se } 0 < x \leq 15 \\ -\frac{K}{10}(x - 25) & \text{se } 15 < x \leq 25 \\ 0 & \text{se } x < 0 \text{ ou } x > 25 \end{cases}$$

   (a) Encontre o valor de K.
   (b) Qual a probabilidade de ocorrer uma ocorrência nos primeiros 10 km?
   (c) Se sabe-se que uma ocorrência ocorreu antes do km 20, qual a probabilidade de que tenha sido antes do km 10?
   (d) Qual a probabilidade de uma ocorrência ter ocorrido entre os km 12 e 22 ?
   (e) Se uma central de apoio for colocada no km 0, qual a distância que espera-se percorrer para atender 100 ocorrências?

Solução:

A função $f(x)$ tem a forma conforme a seguinte figura:



(a) Para que $f(x)$ seja uma f.d.p. a área sob a figura deve ser 1, ou seja, $\int_0^{25} f(x)\mathrm{d}x = 1$.
Solução geométrica:

$$b_r \cdot h_r + \frac{b_t \cdot h_t}{2} = 1$$
$$15 \cdot K + \frac{10 \cdot K}{2} = 1$$
$$K = \frac{1}{20} = 0,05$$

Solução por integração:

$$\int_0^{25} f(x)\mathrm{d}x = 1$$
$$\int_0^{15} f(x)\mathrm{d}x + \int_{15}^{25} f(x)\mathrm{d}x = 1$$
$$K \cdot (15 - 0) - \frac{K}{10}\left(25 \cdot (25 - 15) - \frac{25^2 - 15^2}{2}\right) = 1$$
$$K = \frac{1}{20} = 0,05$$

(b) $P[X \leq 10]$
Solução geométrica:

$$P[X \leq 10] = b_r \cdot h_r = 10 \cdot 0,05 = 0.5 = 1/2$$

Solução por integração:

$$\int_0^{10} f(x)\mathrm{d}x = 0,05 \cdot (15-0) = 0.5 = 1/2$$

(c)

$$P[X \leq 10|X \leq 20] = \frac{P[(X \leq 10(\cap(X \leq 20])}{P[X \leq 20]} = \frac{P[X \leq 10]}{P[X \leq 20]}$$

é mais conveniente calcular: $P[X \leq 20] = 1 - P[X > 20]$

Solução geométrica:

$$P[X > 20] = \frac{b_t \cdot h_t}{2} = \frac{(25-20) \cdot f(20)}{2} = 0.9375 = 15/16$$

Solução por integração:

$$P[X > 20] = \int_{20}^{25} f(x)\mathrm{d}x = 0,005\left(25(25-20) - \frac{25^2 - 20^2}{2}\right) = 0.9375 = 15/16$$

Portanto,

$$P[X \leq 10|X \leq 20] = \frac{1}{0.9375} = 0.5333 = 8/15$$

(d) $P[12 \leq X \leq 22]$

Solução geométrica:

$$P[12 \leq X \leq 22] = b_r \cdot h_r + \frac{b_t \cdot h_t}{2} = (15-12) \cdot K + \frac{f(15) \cdot f(22)}{2} = 0.3775 = 6550/17351$$

Solução por integração:

$$\int_{12}^{22} f(x)\mathrm{d}x = \int_{12}^{15} f(x)\mathrm{d}x + \int_{15}^{22} f(x)\mathrm{d}x = 0,05(15-12) - 0,005\left(25 \cdot (22-15) - \frac{22^2 - 15^2}{2}\right) = 0.3775 = 6550/17351$$

(e) $100 \cdot E[X] = 100 \cdot \int_0^{25} x \cdot f(x)\mathrm{d}x = 1021$

Solução computacional:

```
> fx <- function(x){
+     y <- numeric(length(x))
+     y[x > 0 & x <= 15]  <- 0.05
+     y[x > 15 & x <= 25]  <- -0.005 * (x[x > 15 & x <= 25] - 25)
+     return(y)
+ }
> ## (a)
> (qa <- integrate(fx, 0, 25)$value)

[1] 1

> ## (b)
> (qb <- integrate(fx, 0, 10)$value)

[1] 0.5

> ## (c)
> (p20 <- integrate(fx, 0, 20)$value)

[1] 0.9375

> (qc <- integrate(fx, 0, 10)$value/integrate(fx, 0, 20)$value)

[1] 0.5333

> ## (d)
> (qd <- integrate(fx, 12, 22)$value)

[1] 0.3775

> ## (e)
>  Ex.f <- function(x){ x * fx(x)}
> (EX <- integrate(Ex.f, 0, 25)$value)

[1] 10.21
```

1. Seja uma função de densidade de probabilidades $f(x) = 0,2 - 0,02x \; I_{0,10}(x)$.
   (a) Esboce um gráfico da função.
   (b) Mostre que $f(x)$ é uma função de densidade de probabilidades (f.d.p.) válida.
   (c) Obtenha a expressão de $F(x)$ e seu gráfico.
   (d) Obtenha o valor médio $E[X]$ e a variância $\text{Var}[X]$
   Obtenha as probabilidades:
   (e) $P[X > 2]$,
   (f) $P[X < 7]$,
   (g) $P[X > 5]$,
   (h) $P[X > E[X]]$,
   (i) $P[X > 2|X < 7]$,
   (j) $P[3 < X < 8]$.
   (k) Obtenha os quantis 0,15, 0,25, 0,50, 0,60, 0,75 e 0,90,

Solução:

(a) Ver Figura (esquerda)

(b) Mostrar que:

$$(i) f(x) \geq 1 \forall x$$
$$(ii) \int_0^{10} f(x)\text{d}x = 1$$

(c) Ver Figura (direita)

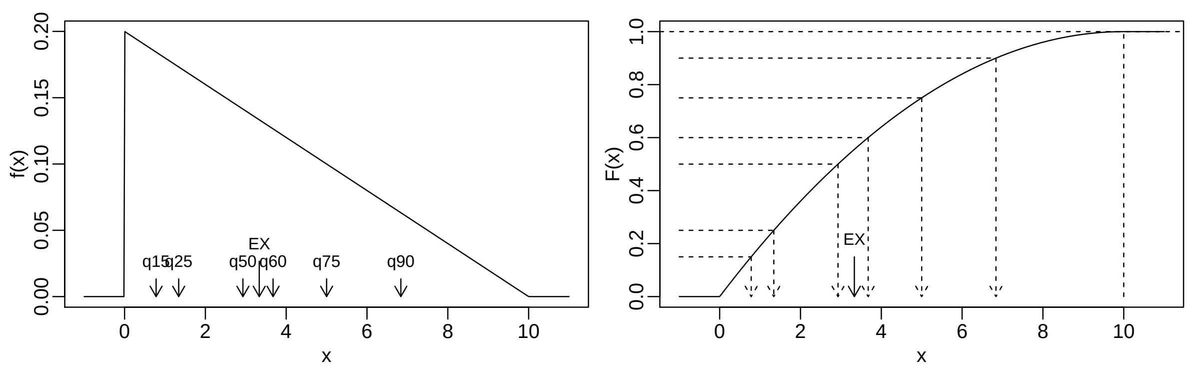


Figura 1: Função de densidade de probabilidade e acumulada com valores da esperança e quantis indicados.

(d)

$$\text{E}[X] = \int_0^{10} x \cdot f(x)\text{d}x = \ldots = 3.33$$
$$\text{Var}[X] = \int_0^{10} (x - \text{E}[X])^2 \cdot f(x)\text{d}x$$
$$= \text{E}[X^2] - (\text{E}[X])^2 = \int_0^{10} x \cdot f(x)\text{d}x - (E[X])^2 = \ldots = 16.7 - (3.33)^2 = 5.56$$

(e) $P[X > 2] = 1 - F(2) = 1 - (0,2 \cdot 2 - 0,01 \cdot 2^2) = 0.64$

(f) $P[X < 7] = F(7) = 0,2 \cdot 7 - 0,01 \cdot 7^2 = 0.91$

(g) $P[X > 5] = 1 - F(5) = 1 - (0,2 \cdot 5 - 0,01 \cdot 5^2) = 0.25$,

(h) $P[X > E[X]] = 1 - F(10/3) = 1 - (0,2 \cdot 10/3 - 0,01 \cdot (10/3)^2) = 0.444$

(i) $P[X > 2|X < 7] = \frac{P[2<X<7]}{P[X<7]} = \frac{F(7)-F(2)}{F(7)} = 0.604$

(j) $P[3 < X < 8] = F(8) - F(3) = 0.45$

(k) Para qualquer quantil q_p, $F(q) = p$, ou seja, $q = F^{-1}(p)$, portanto para a $f(x)$ dada o quantil é a raiz da equação $0,2q_p - 0,01q_p^2 = p$ que estiver no intervalo $(0, 10)$ no qual a função está definida.

$$q_{0,15} = 0.78$$
$$q_{0,25} = 1.34$$
$$q_{0,50} = 2.93$$
$$q_{0,60} = 3.68$$
$$q_{0,75} = 5$$
$$q_{0,90} = 6.84$$

Soluções computacionais com o programa R:

```
> fx <- function(x){
+     y <- ifelse(x>0 & x<10, 0.2 - 0.02*x, 0)
+     return(y)
+ }
> Fx <- function(x){
+     y <- numeric(length(x))
+     I1 <- (x > 0 & x < 10)
+     y[I1]  <- 0.2*x[I1] - 0.01*(x[I1])^2
+     y[x>10]  <- 1
+     return(y)
+ }
> par(mfrow=c(1,2), mar=c(2.5,2.5,0.5, 0.5), mgp=c(1.5,0.5,0))
> xs <- seq(-1, 11, length=501)
> #
> plot(xs, fx(xs), type="l", xlab="x", ylab="f(x)")
> arrows(EX, fx(EX)/5, EX, 0, length=0.1)
> text(EX, fx(EX)/5, "EX" , pos=3,cex=0.8)
> x0 <- c(q15, q25, q50, q60, q75, q90)
> arrows(x0, fx(EX)/10, x0, 0, length=0.1)
> text(x0, fx(EX)/10, c("q15", "q25", "q50", "q60", "q75", "q90") , pos=3,cex=0.8)
> #
> plot(xs, Fx(xs), type="l", xlab="x", ylab="F(x)")
> segments(10,0,10,1, lty=2)
> abline(h=1, lty=2)
> y0 <- c(0.15, 0.25, 0.50, 0.60, 0.75, 0.90)
> segments(-1, y0, x0, Fx(x0), lty=2)
> arrows(x0, Fx(x0), x0, 0, length=0.1, lty=2)
> text(x0, 0, c("q15", "q25", "q50", "q60", "q75", "q90") , pos=1, cex=0.8)
> arrows(EX, 0.15, EX, 0, length=0.1)
> text(EX, 0.15, "EX" , pos=3, cex=0.8)
> Ex.f <- function(x){ x * fx(x)}
> (EX <- integrate(Ex.f, 0, 10)$value)

[1] 3.333

> Ex2.f <- function(x){ x^2 * fx(x)}
> (EX2 <- integrate(Ex2.f, 0, 10)$value)

[1] 16.67

> (VarX <- EX2 - (EX)^2)

[1] 5.556

> (sdX <- sqrt(VarX))

[1] 2.357

> 1 - Fx(2)

[1] 0.64

> Fx(7)

[1] 0.91

> 1 - Fx(5)

[1] 0.25
```

```
> 1 - Fx(EX)
[1] 0.4444
> (Fx(7)-Fx(2))/Fx(7)
[1] 0.6044
> Fx(8)-Fx(3)
[1] 0.45
> q15 <- Re(polyroot(c(-0.15, 0.2, -0.01))); (q15 <- q15[q15 > 0 & q15 < 10])
[1] 0.7805
> q25 <- Re(polyroot(c(-0.25, 0.2, -0.01))); (q25 <- q25[q25 > 0 & q25 < 10])
[1] 1.34
> q50 <- Re(polyroot(c(-0.50, 0.2, -0.01))); (q50 <- q50[q50 > 0 & q50 < 10])
[1] 2.929
> q60 <- Re(polyroot(c(-0.60, 0.2, -0.01))); (q60 <- q60[q60 > 0 & q60 < 10])
[1] 3.675
> q75 <- Re(polyroot(c(-0.75, 0.2, -0.01))); (q75 <- q75[q75 > 0 & q75 < 10])
[1] 5
> q90 <- Re(polyroot(c(-0.90, 0.2, -0.01))); (q90 <- q90[q90 > 0 & q90 < 10])
[1] 6.838
```

Avaliação 05

1. Responda as questões a seguir declarando a variável aleatória e a sua distribuição.

   (a) Registros de um sistema mostram que 1 a cada 20 requisições de acesso de um determinado serviço não são completadas.

      i. Se forem feitas 15 requisições qual a probabilidade de que no máximo duas não sejam completadas.

      ii. Em um teste para avaliar o sistema requisições serão feitas sequencialmente até que a primeiro acesso não seja completado. Se este teste for feito diversas vezes, anotando-se o número de acessos a cada teste, qual deve ser o valor da média do número de acessos? Como voce calcularia a probabilidade de que o esse número de acessos não chegue a 5?

      iii. O teste anterior foi repetido porém até que o terceiro acesso não fosse completado. Em um particular ensaio foram feitas 10 análise desta forma. Qual a probabilidade desta ocorrência? Se este teste for repetido diversas vezes e o número de acessos anotado, qual deve ser a média do número de acessos?

   Solução:

$$p = 1/20$$

   i.

$$\begin{aligned}
&X : \text{número de requisições não atendidas em 15 solicitações} \\
&X \sim B(n = 15, p = 1/20) \\
&P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2] = 0.9638
\end{aligned}$$

   ii.

$$\begin{aligned}
&X : \text{número de requisições } \textbf{atendidas} \text{ até a primeira não completada} \\
&X \sim G(p = 1/20) \quad P[X = x] = p \cdot (1 - p)^x, x = 0, 1, \ldots \\
&E[X] + 1 = \frac{1 - p}{p} + 1 = \frac{1}{p} = 20 \\
&P[X < 4] = \sum_{i=0}^{3} P[X = i] = P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] = \\
&= \sum_{i=0}^{4} (1/20)(1 - 1/20)^i = 0.1855
\end{aligned}$$

iii.

X : número de requisições **atendidas** até a terceira não completada

$$X \sim BN(r=3, p=1/20) \quad ; \quad P[X=x] = \binom{x+3-1}{3}(1/20)^3 \cdot (1-1/20)^x, x=0,1,\ldots$$

10 requisições $\longrightarrow$ 7 atendidas

$$P[X=7] = 0.003143$$

$$E[X]+3 = r\frac{1-p}{p}+1 = \frac{r}{p} = 60$$

Soluções computacionais:

```
> (Pi <- pbinom(2, size=15, prob=1/20))
[1] 0.9638
> (Pii <- pgeom(3, prob=1/20))
[1] 0.1855
> (Piii <- dnbinom(7, size=3, prob=1/20))
[1] 0.003143
```

(b) Tem-se um conjunto de 40 sensores das quais 15 estão danificados. Uma transmissão é feita para 12 sensores foram selecionadas ao acaso. Qual a probabilidade da transmissão ter sido enviada para 4 ou mais sensores operantes?

Solução:

X : número de sensores operantes dentre os 12

$$X \sim HG(N=40, r=(40-15=25), n=12) \quad ; \quad P[X=x] = \frac{\binom{25}{x}\binom{15}{12-x}}{\binom{40}{12}}$$

$$P[X \geq 4] = 1 - P[X \leq 3] = 1 - \sum_{i=0}^{2} P[X=i] = 1 - P[X=0] - P[X=1] - P[X=2] - P[X=3] = 0.9978$$

Solução computacional:

```
> (pHG <- phyper(3, m=25, n=15, k=12, lower=F))
[1] 0.9978
```

(c) Considere agora uma transmissão de dados que tem uma taxa de falha de 5,2 falhas por hora. Qual a probabilidade de que em um intervalos de 15 minutos não haja nenhuma falhas de transmissão? E de que seja registradas mais do que 4 falhas?

Solução:

X : número de falhas em 15 minutos

$$X \sim P(\lambda = 5,2/4 = 1,3) \quad P[X=x] = \frac{e^{-1,3}1,3^x}{x!}, x=0,1,\ldots$$

$$P[X=0] = 0.2725$$

$$P[X>4] = 1 - P[X \leq 3] = 1 - (P[X=0] + P[X=1] + P[X=2] + P[X=3] + P[X=4]) =$$

$$= 1 - \sum_{i=0}^{4} \frac{e^{-1,3}1,3^i}{i!} = 0.01066$$

Solução computacional:

```
> (pPa <- dpois(0, lambda=1.3))
[1] 0.2725
> (pPb <- ppois(4, lambda=1.3, lower=F))
[1] 0.01066
```

---

2. A resistência de um papel é modelada por uma distribuição normal de média 35 libras por polegada quadrada (lb/in^2) e um desvio padrão de 2 lb/in^2.
   (a) qual a probabilidade de que a resistência de uma amostra seja menor que 40 lb/in^2?
   (b) qual a proporção de amostra que deve ter resistência entre 33 e 38 lb/in^2?
   (c) qual o valor de resistência para o qual se espera que 75% das amostras apresentem resistência inferior a ele?
   (d) se a especificação do material requer que a resistência seja superior a 33 lb/in^2, qual a proporção de amostras que espera-se descartar após inspeção?
   (e) qual deveria ser a resistência média para que esta proporção fosse inferior a 5%?

Solução:

(a) $P[X < 40] = P[Z < \frac{40-35}{2}] = P[Z < 2.5] = 0.9938$
(b) $P[33 < X < 38] = P[\frac{33-35}{2} < Z < \frac{38-35}{2}] = P[-1 < -Z < 1.5] = 0.7745$
(c)

$$P[X < x] = P[Z < \frac{x-35}{2}] = 0,75$$
$$z = \frac{x-35}{2} = 0.6745$$
$$x = 36.3$$

(d) $P[X < 33] = P[Z < \frac{33-35}{2}] = P[Z < -1] = 0.1587$
(e) Devemos descobrir um valor para média μ que satisfaça:

$$P[X < 33] = 0.05$$
$$P[Z < \frac{33-\mu}{2}] = P[Z < z_{0.05}] = 0.05$$
$$\mu = 33 - 2 \cdot z_{0.05} = 33 - 2 \cdot (-1.64) = 36.3$$

Solução computacional:

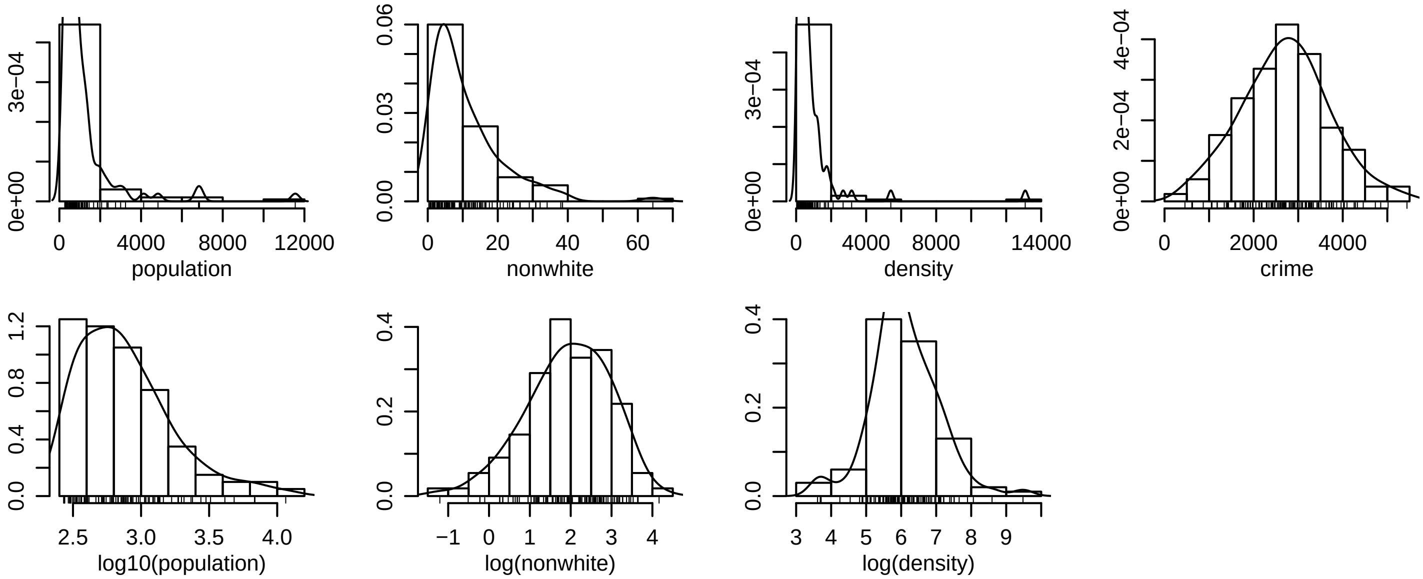
```
> (pa <- pnorm(40, m=35, sd=2))
[1] 0.9938
> (pb <- diff(pnorm(c(33,38), m=35, sd=2)))
[1] 0.7745
> (pc <- qnorm(0.75, m=35, sd=2))
[1] 36.35
> (pd <- pnorm(33, m=35, sd=2))
[1] 0.1587
> (pe <- 33 - 2 * qnorm(0.05))
[1] 36.29
```

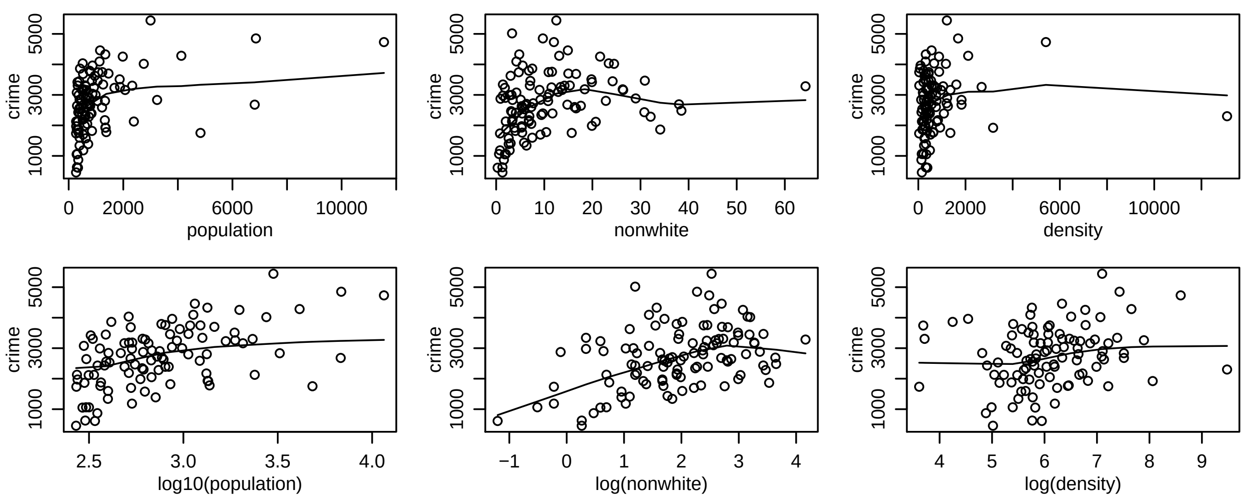
Avaliação 06

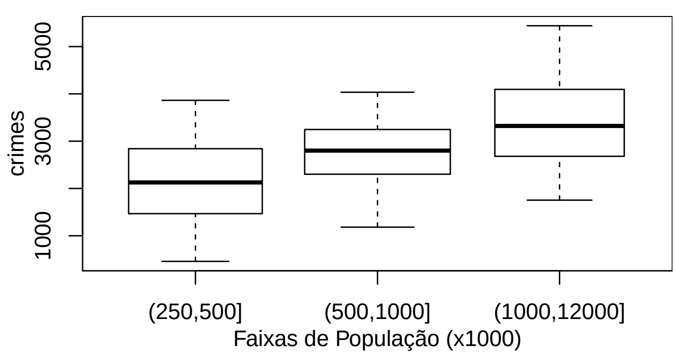
1. O conjunto de dados `car::Freedman` do programa **R** possui registros da população (`population`) em milhares de habitantes, porcentagem de não brancos (`nonwhite`), densidade populacional (`density`) e número de crimes (`crimes`) em 110 áreas metropolitanas com população acima de 250 mil habitantes dos Estados Unidos no ano de 1968. A tabela de medidas estatísticas e gráficos abaixo apresentam resumos dos dados a serem interpretados. Comece esboçando como seria o formato da tabela dos dados. Identifique os tipos de variáveis e discuta todos os resultados. Inclua ainda nos comentários o que voce espera dos valores de correlação entre número de crimes e demais variáveis.[1]

	n	media	desvioP	min	max	amplitude	Q0.25	Q0.5	Q0.75	CV
population	100	1136.0	1560.14	270.0	11551.0	11281	398.8	664.0	1167.75	137.34
nonwhite	110	10.8	10.26	0.3	64.3	64	3.4	7.2	14.88	94.97
density	100	765.7	1441.95	37.0	13087.0	13050	266.5	412.0	773.25	188.33
crime	110	2714.1	991.40	458.0	5441.0	4983	2066.8	2698.0	3305.00	36.53

[1] `log10()` é o logarítmo na base 10 enquanto que `log()` é o logarítmo neperiano







Avaliação 07

1. Os dados a seguir se referem ao diâmetro e altura de 31 cerejeiras.

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]
Diametro	8.3	8.6	8.8	10.5	10.7	10.8	11	11	11.1	11.2	11.3	11.4	11.4	11.7	12	12.9
Altura	70.0	65.0	63.0	72.0	81.0	83.0	66	75	80.0	75.0	79.0	76.0	76.0	69.0	75	74.0

	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]	[,23]	[,24]	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]
Diametro	12.9	13.3	13.7	13.8	14	14.2	14.5	16	16.3	17.3	17.5	17.9	18	18	20.6
Altura	85.0	86.0	71.0	64.0	78	80.0	74.0	72	77.0	81.0	82.0	80.0	80	80	87.0

(a) Obtenha um diagrama ramo-e-folhas dos diâmetros.

(b) Faça um diagrama *box-plot* da ambas variáveis/atributos.

(c) Descreva o comportamento de cada um dos atributos.

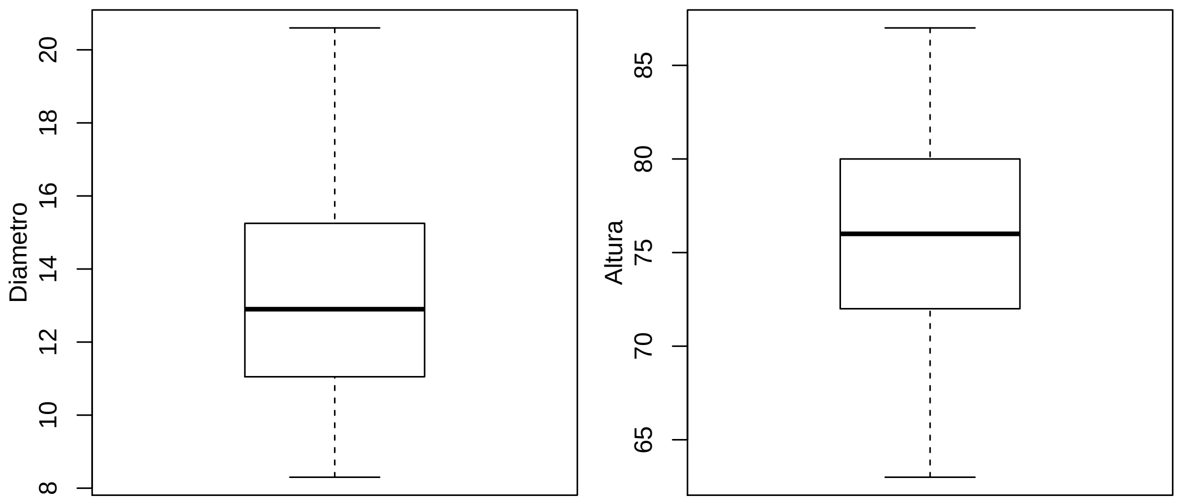
(d) Voce espera (a princípio) que os atributos estejam correlacionados? Justifique. Faça alguma análise (gráfico, tabela ou medida) que permita avaliar sua conjectura inicial e tire suas conclusões.

Solução:

(a)

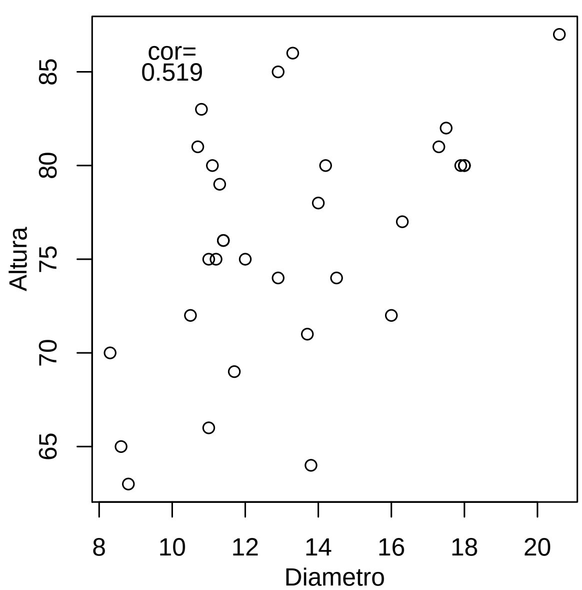
```
The decimal point is at the |

 8 | 368
10 | 57800123447
12 | 099378
14 | 025
16 | 03359
18 | 00
20 | 6
```

(b)



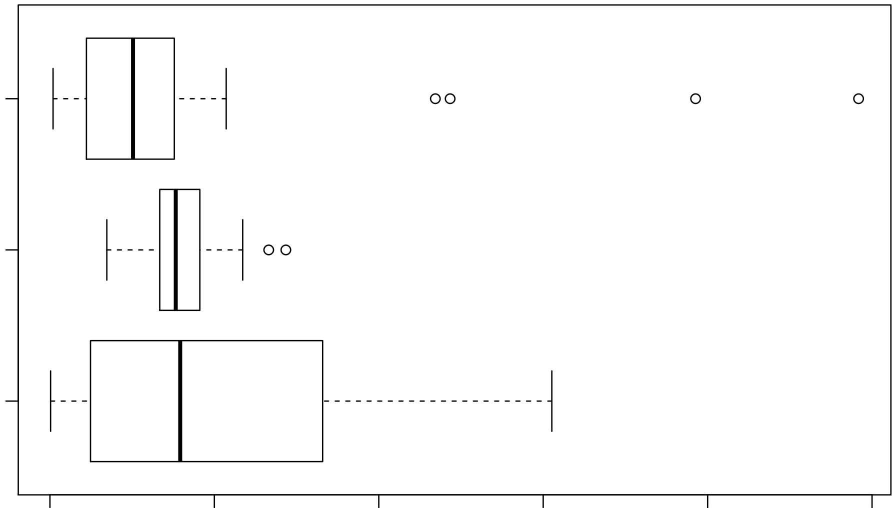
(c)

(d)



---

2. 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.

Avaliação 08

1. Considere um exame no qual as notas possuem uma média de 100 pontos e um desvido padrão de 12 pontos. Obtenha a probabilidade de que
   (a) a nota de um estudante selecionado ao acaso seja superior a 110,
   (b) a nota média de um grupo de 25 estudantes tomados ao acaso seja superior a 105,
   (c) a nota média de um grupo de 64 estudantes tomados ao acaso seja superior a 105,
   (d) a nota média de um grupo de 16 estudantes tomados ao acaso esteja abaixo de 95 ou acima de 105,
   (e) o número de estudantes que deveriam ser selecionados ao acaso para que a nota média estivesse entre 95 e 105 pontos com probabilidade de 0,98.

Solução:

$$\begin{aligned} X &: \text{nota de um estudante} \\ X &\sim N(\mu = 100, \sigma^2 = 12^2) \\ \overline{X}_n &: \text{nota média de um grupo de } n \text{ estudantes} \\ \overline{X}_n &\sim N(\mu_{\overline{X}} = 100, \sigma^2_{\overline{X}} = 12^2/n) \end{aligned}$$

Soluções computacionais (linguagem R):

(a) $P[X > 110] = P[Z > \frac{110-100}{12}] = P[Z > 0.8333] = 0.202$
(b) $P[\overline{X}_{25} > 105] = P[Z > \frac{110-100}{12/\sqrt{25}}] = P[Z > 2.083] = 0.0186$
(c) $P[\overline{X}_{64} > 105] = P[Z > \frac{110-100}{12/\sqrt{64}}] = P[Z > 3.333] = 0.000429$
(d) $P[\overline{X}_{16} < 95 \cup \overline{X}_{16} > 105] = P[Z < \frac{110-100}{12/\sqrt{16}} \cup Z > \frac{110-100}{12/\sqrt{16}}] = P[Z < -1.667 \cup Z > 1.667] = 0.0956$
(e)

$$\begin{aligned} P[95 < \overline{X}_n < 105] &= 0,98 \\ P[|\overline{X}_n - 100| < 5] &= 0,98 \\ z &= \frac{5}{12/\sqrt{n}} \\ n &= \lceil \left(\frac{z \cdot 12}{5}\right)^2 = 32 \end{aligned}$$

```
> pa <- pnorm(110, m=100, sd=12, lower=FALSE)
> pb <- pnorm(105, m=100, sd=12/sqrt(25), lower=FALSE)
> pc <- pnorm(105, m=100, sd=12/sqrt(64), lower=FALSE)
> pd <- 1-diff(pnorm(c(95,105), m=100, sd=12/sqrt(16)))
> n <- ceiling((qnorm(0.99)*12/(105-100))^2)
```

---

2. 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, parâmetro de interesse e o estimador?
   (b) Se a amostra for de 4000 solicitações, qual será a margem de erro (com confiança de 95%) para a estimação da proporção de resolvidas?
   (c) Para este mesmo tamanho de amostra, qual seria a confiança associada a uma margem de erro de $\pm 0,01$ (1 %)?
   (d) Qual deveria ser o tamanho da amostra para se obter a estimativa com uma margem de erro de 2,5% com 95% de confiança?
   (e) 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 na população, que representa a probabilidade p de uma solicitação ser atendida. O estimador é a proporção de atendidas na amostra. 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) $M.E. = z_{(1-\alpha)}\sqrt{\frac{p(1-p)}{n}} = 1.96\sqrt{\frac{0,5(1-0,5)}{4000}} = 0.0155$

(c)

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

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

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

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

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

Avaliação 09

1. Seja um conjunto de observações independentes (dados) $y_1, y_2, ..., y_n$. Mostre que a média amostral $\overline{y} = \sum_{i=1}^{n} y_i/n$ é o estimador de mínimos quadrados para μ sob o modelo $y_i = \mu + \epsilon_i$.

2. Seja um conjunto de observações independentes $y_1, y_2, ..., y_n$ de uma distribuição com densidade $f(y) = \lambda \mathrm{e}^{-\lambda y}$. Lembrando que a função de log-verossimilhança é dada por:

$$l(\lambda) = \sum_{i=1}^{n} \log[f(y_i)] \ ,$$

(a) Obtenha a expressão do estimador de máxima verossimilhança de λ

(b) Obtenha a estimativa para a amostra: $(12, 15, 8, 10, 32, 7, 10, 13)$