CE-003: Estatística II - Turma: K/O, 1^a Prova (20/04/2016)

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

1. Um meteorologista acerta 80% dos dias em que chove e 90% dos dias em que faz bom tempo. Chove em 10% dos dias. Tendo havido previsão de chuva, qual a probabilidade de chover?

Solução:

$$
\begin{aligned}
C &: \text{chove em um dia}; P[C] = 0,10; P[\overline{C}] = 0.90 \\
P_C &: \text{previsão de chuva}; P[P_C|C] = 0,80 \longrightarrow P[\overline{P_C}|C] = 0,20 \\
\overline{P_C} &: \text{sem previsão de chuva}; P[\overline{P_C}|\overline{C}] = 0,90 \longrightarrow P[P_C|\overline{C}] = 0,10 \\
P[C|P_C] &= \frac{P[P_C \cap C]}{P[P_C]} = \frac{P[P_C \cap C]}{P[P_C \cap C] + P[P_C \cap \overline{C}]} = \frac{P[P_C|C] \cdot P[C]}{P[P_C|C] \cdot P[C] + P[P_C|\overline{C}] \cdot P[\overline{C}]} \\
&= \frac{(0,8)(0,10)}{(0,8)(0,10) + (0,10)(0,90)} = 0,471
\end{aligned}
$$

Solução alternativa, organizando dados na forma de uma tabela:

	Ocorrência		
Previsão	Chove (C)	Não chove ($\overline{C}$)	Probabilidade
com chuva (P_c)	$P[P_c \cap C] = 0,80 \cdot 0,10$	$P[P_c \cap \overline{C}] = 0,09$	0,17
sem chuva ($\overline{P_C}$)	$P[\overline{P_c} \cap C] = 0,02$	$P[\overline{P_c} \cap \overline{C}] = 0,90 \cdot 0,90$	0,83
Probabilidade	0,10	0,90	1

---

2. Um site de vendas pela internet registra 40% dos acessos do estado do PR, 50% de outros estados e 10% do exterior. 20% dos acessos do PR resultam em uma compra, enquanto que os percentuais para outros estados e exterior são de 10% e 30%, respectivamente.

   (a) Qual a probabilidade de um acesso resultar em compra?

   (b) Se foi feita uma compra, qual a probabilidade de ela ter sido do exterior?

Solução:
Eventos:

$$PR : \text{acesso do PR} \quad OE : \text{acesso de outros estados} \quad EX : \text{acesso do exterior}$$

$$C : \text{compra}$$

Probabilidades informadas:

$$P[PR] = 0,40 \quad P[OE] = 0,50 \quad P[EX] = 0,10$$

$$P[C|PR] = 0,20 \quad P[C|OE] = 0,10 \quad P[C|EX] = 0,30$$

(a) $P[C] = P[PR \cap C] + P[OE \cap C] + P[EX \cap C] = P[PR] \cdot P[C|PR] + P[OE] \cdot P[C|OE] + P[EX] \cdot P[C|EX] = (0,40)(0,20) + (0,50)(0,10) + (0,10)(0,30) = 0,16$

(b) $P[EX|C] = \frac{P[EX \cap C]}{P[C]} = \frac{P[EX] \cdot P[C|EX]}{P[C]} = \frac{(0,10)(0,30)}{(0,40)(0,20)+(0,50)(0,10)+(0,10)(0,30)} = 0,1875$

---

3. Um indivíduo vai participar de uma competição que consiste em responder questões que são lhe são apresentadas sequencialmente. Com o nível de conhecimento que possui, a chance de acertar uma questão escolhida ao acaso é de 75% . Neste contexto, para cada diferente situação apresentada a seguir, defina a variável aleatória, sua distribuição de probabilidades e calcule a probabilidade solicitada. Se preciso, faça suposições necessárias e adequadas em cada caso.

   (a) Se for responder até errar uma pergunta, qual a probabilidade de conseguir acertar quatro ou mais questões?

(b) Se for responder cinco perguntas, qual a probabilidade de acertar quatro ou mais?

(c) Se for responder até acertar a terceira pergunta, qual a probabilidade de errar apenas uma?

(d) Se o candidato selecionar aleatoriamente seis questões de um banco de 40 questões das quais o candidato sabe a resposta de 30 delas (75%), qual a probabilidade de acertar ao menos cinco delas.

Ainda neste contexto considere que o candidato responde, em média, 1,8 questões por minuto.

(e) Qual a probabilidade de conseguir responder ao menos três questões em três minutos?

(f) Qual a probabilidade de que o tempo para resposta de uma questão seja superior a 40 segundos?

Solução:

(a)

$$X : \text{Número de acertos até o primeiro erro}$$
$$X \sim \mathrm{G}(0,25)$$
$$P[X \geq 4] = 1 - P[X \leq 3] = 1 - \sum_{i=0}^{3}(1-0,25)^i(0,25) = 0,316$$

(b)

$$X : \text{Número de acertos em cinco perguntas}$$
$$X \sim \mathrm{B}(\mathrm{n}=5, \mathrm{p}=0,75)$$
$$P[X \geq 4] = P[X=4] + P[X=5] = \sum_{i=4}^{5}\binom{5}{i}0,75^i(1-0,75)^{5-i} = 0,633$$

(c)

$$X : \text{Número de erros até o terceiro acerto}$$
$$X \sim \mathrm{BN}(\mathrm{r}=3, \mathrm{p}=0,75)$$
$$P[X=1] = \binom{3+1-1}{3-1}0,75^3(1-0,75)^1 = 0,316$$

(d)

$$X : \text{Número de acertos nas seis questões selecionadas}$$
$$X \sim \mathrm{HG}(30,10,6)$$
$$P[X \geq 5] = P[X=5] + P[X=6] = \sum_{i=5}^{6}\frac{\binom{30}{i}\binom{10}{6-i}}{\binom{40}{6}} = 0,526$$

(e)

$$X : \text{Número de questões respondidas em 3 minutos}$$
$$X \sim \mathrm{P}(3 \cdot 1,8 = 5,4)$$
$$P[X \geq 3] = 1 - P[X \leq 2] = 1 - \sum_{i=0}^{2}\frac{\mathrm{e}^{-5,4}5,4^{\mathrm{i}}}{i!} = 0,905$$

(f)

$$X : \text{tempo (em min.) para responder uma questão}$$
$$X \sim \mathrm{Exp}(\lambda = 1,8)$$
$$P[X \geq 40/60] = \int_{40/60}^{\infty} 1,8\mathrm{e}^{-1,8\mathrm{x}}\mathrm{dx} = 0,301$$

---

4. Seja uma v.a. X com função de distribuição de densidade de probabilidade:

$$f(x) = \begin{cases} x/A, & \text{se } 0 \leq x < 3 \\ (6-x)/A, & \text{se } 3 \leq x \leq 6 \\ 0, & \text{caso contrário} \end{cases}$$

(a) encontre o valor de A

(b) encontre $P[X < 2]$ e $P[X < 1 \text{ ou } X > 4,5]$

(c) encontre a média e a mediana de X

Solução:

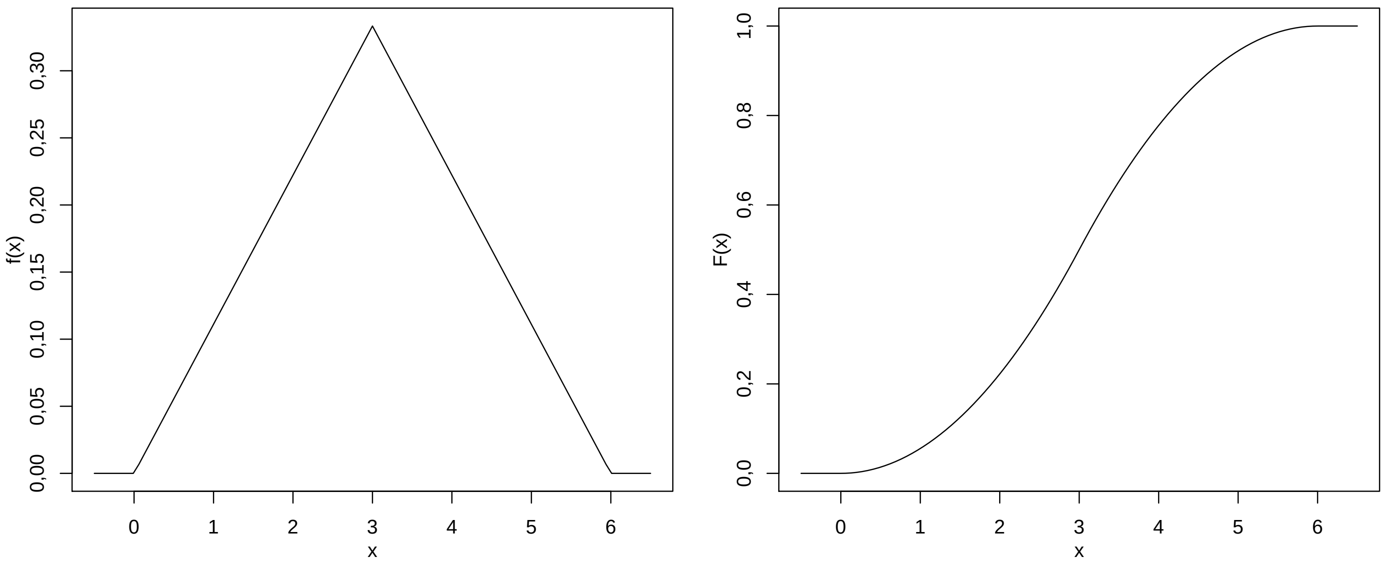

Figura 1: Funções de distribuição de probabilidades ($f(x)$) e acumulada ($F(x)$).

(a) $\int_0^6 f(x)\mathrm{d}x = 1 \longrightarrow A = 9$ (pode também ser encontrado pela área do triângulo)

(b) (considerar a solução geométrica equivalente por área de triângulos)

$$P[X < 2] = \int_0^2 f(x)\mathrm{d}x = 0,222$$

$$P[X < 1 \text{ ou } X > 4,5] = \int_0^1 f(x)\mathrm{d}x + \int_{4,5}^6 f(x)\mathrm{d}x = 0,181$$

(c) (notar que a distribuição é simétrica, a solução é portanto trivial)

$$E[X] = \int_0^6 x \cdot f(x)\mathrm{d}x = 3$$

$$\mathrm{md}[X] : \int_0^{\mathrm{md}} f(x)\mathrm{d}x = 0,5 \longrightarrow \mathrm{md}[X] = 3$$

Soluções computacionais (linguagem R):

```
> ## definindo f(x)
> ddist <- function(x){
+     y <- numeric(length(x))
+     y[x >= 0 & x < 3] <- x[x >= 0 & x < 3]/9
+     y[x >= 3 & x < 6] <- (6-x[x >= 3 & x < 6])/9
+     return(y)
+ }
> ## definindo F(x)
> pdist <- function(x){
+     y <- numeric(length(x))
+     ind <- x >= 0 & x < 3
+     y[ind] <- x[ind]^2/18
+     ind <- x >= 3 & x < 6
+     y[ind] <- 0.5 + (6*(x[ind]-3) - (x[ind]^2-3^2)/2)/9
+     y[x >= 6] <- 1
+     return(y)
+ }
> ## definindo F^{-1}(x) (forma numérica - geral)
> qdist <- function(prob){
+     uniroot(function(x) pdist(x) - prob, interval=c(0,6))$root
```

```
+ }
> ## definindo F^{-1}(x) (forma analítica - específica para este caso)
> qdist <- function(prob){
+     ifelse(prob <= 0.5, sqrt(18*prob), (12-sqrt(144-72*(prob+1)))/2)
+ }
> ## Soluções
> p1 <- pdist(2)
> p2 <- pdist(1) + (1-pdist(4.5))
> Exf <- function(x) x*ddist(x)
> Ex <- integrate(Exf, 0, 6)$value
> md <- qdist(0.5)
> ## Gráficos
> par(mfrow=c(1,2), mar=c(3.5,3.5,1,1), mgp=c(2,1,0))
> curve(ddist, from=-0.5, to=6.5, ylab="f(x)")
> curve(pdist, from=-0.5, to=6.5, ylab="F(x)")
```

---

5. Um grande número de alunos faz uma prova visando avaliar a sua formação e a partir da qual é atribuído a cada um deles um *escore de proficiência* (EP). Os resultados mostram que os EP's podem ser descritos por uma distribuição normal de média 300 e variância 400.

   (a) Qual a percentagem de alunos com escores acima de 320?

   (b) Qual a percentagem de pessoas com escores entre 250 e 350?

   (c) Qual a percentagem de pessoas com escores que não se afastem da média mais do que 50?

   (d) Qual valor deve ter 15% dos escores abaixo dele?

   (e) Deseja-se dividir os escores em três grupos (alto, médio e baixo) com a mesma proporção de alunos. Quais os valores de corte que dividem as categorias/grupos?

   (f) Mantendo-se a mesma média, quanto deveria ser o desvio padrão para que se tenha não mais que 10% dos escores abaixo de 270?

   (g) Mantida a variância de 400 quanto deveria ser o escore médio para que se tenha não mais que 10% dos escores abaixo de 270?

$$X : \text{escores no exame e proficiência}$$
$$X \sim N(400, 45^2)$$

(a) $P[X > 320] = P[Z > \frac{320-300}{20}] = P[Z > 1] = 0,1587 \longrightarrow 15,87\%$

(b) $P[250 < X < 350] = P[\frac{250-300}{20} < Z < \frac{350-300}{20}] = P[-2,5 < Z < 2,5] = 0,9876 \longrightarrow 98,76\%$

(c) $P[|X-300| < 50] = P[250 < X < 350] = P[\frac{250-300}{20} < Z < \frac{250-300}{20}] = P[2,5 < Z < 2,5] = 0,9876 \longrightarrow 98,76\%$

(d) $P[X < a] = 0,15 \longrightarrow z = -1,04 = \frac{a-300}{20} \longrightarrow a = 279,3$

(e)

$$P[X < x_1] = 1/3 \longrightarrow z = -0,431 = \frac{x_1 - 300}{20} \longrightarrow x_1 = 291,4$$
$$P[X < x_2] = 2/3 \longrightarrow z = 0,431 = \frac{x_2 - 300}{20} \longrightarrow x_2 = 308,6$$

(f) $P[X < 270|\mu = 300, \sigma] = 0,10 \longrightarrow z = -1,28 = \frac{270-300}{\sigma} \longrightarrow \sigma = 23,4$

(g) $P[X < 270|\mu, \sigma = 20] = 0,10 \longrightarrow z = -1,28 = \frac{270-\mu}{20} \longrightarrow \mu = 295,6$

Solução computacional com o programa *R*:

```
> (ita <- round(pnorm(320, 300, 20, low=F),dig=4))

[1] 0,1587

> (itb <- round(diff(pnorm(c(250, 350), 300, 20)),dig=4))

[1] 0,9876
```

```
> (itc <- round(diff(pnorm(c(250, 350), 300, 20)),dig=4))
[1] 0,9876
> (itd <- round(qnorm(0.15, 300, 20), dig=1))
[1] 279,3
> (ite <- round(qnorm(c(1/3,2/3), 300, 20), dig=1))
[1] 291,4 308,6
> (itf <- round((270-300)/qnorm(0.10), dig=1))
[1] 23,4
> (itg <- round(270-qnorm(0.10)*20, dig=1))
[1] 295,6
```

---

6. 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.

   (a) Qual a probabilidade de haver duas conexões com intervalo inferior a 1,5 segundos?
   (b) Qual a probabilidade de se passarem 5 segundos sem conexão alguma?
   (c) Tendo havido uma conexão, qual a probabilidade da próxima conexão ocorrer entre 0,5 e 2,5 segundos?
   (d) Se já se passou 1 segundo sem conexão, qual a probabilidade de se passar mais 0,5 segundos adicionais sem conexão?
   (e) 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.

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

(a) $P[X < 1,5] = \int_0^{1,5} f(x)\mathrm{d}x = F(1,5) = 0,63$
(b) $P[X > 5] = \int_5^{\infty} f(x)\mathrm{d}x = 1 - F(5) = 0,036$
(c) $P[X < 0,5] = \int_{0,5}^{2,5} f(x)\mathrm{d}x = F(2,5) - F(0,5) = 0,53$
(d) $P[X > 1,5|X > 1] = \frac{\int_{1,5}^{\infty} f(x)\mathrm{d}x}{\int_1^{\infty} f(x)\mathrm{d}x} = {}^1P[X > 0,5] = 1 - F(0,5) = 0,72$
(e) $P[X < 3,5|X > 2] = \frac{\int_2^{3,5} f(x)\mathrm{d}x}{\int_2^{3,5} f(x)\mathrm{d}x} = \frac{F(3,5)-F(2)}{1-F(2)} = {}^1P[X < 1,5] = F(1,5) = 0,63$

---

[1]propriedade de falta de memória da exponencial