CE-003: Estatística II - Turma K/O
Avaliações Semanais (1^o semestre 2016)

Semana 2 (av-01)

1. (adaptado de Bussab & Morettin) Três jogadores, A, B e C disputam um torneio de tênis. Inicialmente A joga com B e o vencedor joga com C, e assim por diante. O torneio termina quando um jogador ganha duas vezes em seguida ou quando são disputadas, ao todo, quatro partidas.

   (a) A sequência de jogos que determina o resultado final do torneio pode ser considerada um experimento aleatório? Justifique.

   (b) Quais são os possíveis resultados?

   (c) O torneio é "justo" em relação às chances de vitória dos jogadores mediante a regra proposta?

   Considere agora que, sabendo-se dos restrospectos de resultados dos jogadores e suas classificações em *rankings* tem-se que A vence B com probabilidade de 0,72, vence C com probabilidade de 0,65 e B vence C com probabilidade de 0,38.

   (d) Qual a probabilidade de cada jogador ganhar o torneio?

Solução:

1o Jogo	Vencedor	2o Jogo	Vencedor	3o Jogo	Vencedor	4o Jogo	Vencedor	Campeão	Sequência*	Prob**
A x B	A	A x C	A	-	-	-	-	A	(AA)	1/4
	A	A x C	C	B x C	B	A x B	A	A	(ACBA)	1/16
	A	A x C	C	B x C	B	A x B	B	B	(ACBB)	1/16
	A	A x C	C	B x C	C	-	-	C	(ACC)	1/8
	B	B x C	B	-	-	-	-	B	(BB)	1/4
	B	B x C	C	A x C	A	A x B	A	A	(BCAA)	1/16
	B	B x C	C	A x C	C	A x B	B	B	(BCAB)	1/16
	B	B x C	C	A x C	C	-	-	C	(BCC)	1/8

*Sequência : sequência de vencedores dos jogos torneio

** Probabilidade supondo igualdade entre os competidores em cada jogo

(a) Sim. (justificativas serão analisadas)

(b)
$$\Omega = \{(AA), (ACBA), (ACBB), (ACC), (BB), (BCAA), (BCAB), (ACC)\}$$

(c) O jogo não é honesto pois sob a hipótese de igualdade de condições em cada jogo os jogadores possuem diferentes chances de vencer o torneio.

Vencedor	A	B	C
Sequências	{(AA), (ACBA), (BCAA)}	{(ACBB),(BB),(BCAB)}	{(BCC),(ACC)}
Probabilidade	(1/4)+(1/16)+(1/16)=3/8	(1/16)+(1/4)+(1/16)=3/8	(1/8)+(1/8) = 2/8

(d)

Campeão	Sequência	Probabilidade
A	(AA)	$0,72 \cdot 0,65 = 0.468$
A	(ACBA)	$0,72 \cdot 0,35 \cdot 0,38 \cdot 0,72 = 0.0689$
B	(ACBB)	$0,72 \cdot 0,35 \cdot 0,38 \cdot 0,28 = 0.0268$
B	(ACC)	$0,72 \cdot 0,35 \cdot 0,62 = 0.1562$
B	(BB)	$0,28 \cdot 0,38 = 0.1064$
A	(BCAA)	$0,28 \cdot 0,62 \cdot 0,65 \cdot 0,72 = 0.0812$
B	(BCAB)	$0,28 \cdot 0,62 \cdot 0,65 \cdot 0,28 = 0.0316$
C	(BCC)	$0,28 \cdot 0,62 \cdot 0,35 = 0.0608$

Vencedor	A	B	C
Sequências	{(AA), (ACBA), (BCAA)}	{(ACBB),(BB),(BCAB)}	{(BCC),(ACC)}
Probabilidade	0.6182	0.1648	0.217

1. Uma coleção de 100 programas de computador foi examinada para detectar erros de *"sintaxe"*, *"input/output"* e de *"outro tipo"* diferente dos anteriores. Desses 100 programas, 20 tinham erros de "sintaxe", 10 tinham erros de "input/output" e 5 tinham erros de "outro tipo", 6 tinham erros de "sintaxe" e de "input/output", 3 tinham erros de "sintaxe" e de "outro tipo", 3 tinham erros de "input/output" e de "outro tipo" e 2 tinham os três tipos de erros considerados. Um programa é seleccionado ao acaso desta coleção. Determine a probabilidade de que o programa seleccionado tenha:
   (a) Exclusivamente erros de "sintaxe".
   (b) Pelo menos um dos três tipos de erros

   Solução:
   Notação:

$$\begin{aligned} S &: \text{erro de sintaxe} \\ I &: \text{erro de input/output} \\ O &: \text{erro de outro tipo} \end{aligned}$$

   Dados:

$$\begin{aligned} P[S] = 0,20 \quad &; \quad P[I] = 0,10 \quad ; \quad P[O] = 0,05 \\ P[S \cap I] = 0,06 \quad &; \quad P[S \cap O] = 0,03 \quad ; \quad P[I \cap O] = 0,03 \\ P[S \cap I \cap O] = 0,02 & \end{aligned}$$

   (a) $P[S] - P[(S \cap I) \cup (S \cap O)] = P[S] - \{P[(S \cap I)] + (S \cap O)] - P[S \cap I \cap O]\} =$
   $= 0,20 - 0,06 - 0,03 + 0,02 = 0,13$
   (b) $P[S \cup I \cup O] = P[S] + P[I] + P[O] - P[S \cap I] - P[S \cap O] - P[I \cap O] + P[S \cap I \cap O] =$
   $= 0,20 + 0,10 + 0,05 - 0,06 - 0,03 - 0,03 + 0,02 = 0,25$

---

2. Suponha que 5% de uma população sofre de hipertensão e que, de entre estes, 75% ingerem bebidas alcoólicas. De entre os que não são hipertensos 50% ingerem bebidas alcoólicas. Suponha que um indivíduo é escolhido ao acaso da população.
   (a) Calcule a probabilidade de o indivíduo escolhido ingerir bebidas alcoólicas.
   (b) Sabendo que o indivíduo escolhido ingere bebidas alcoólicas, calcule a probabilidade de sofrer de hipertensão.

   Solução:
   Notação:

$$\begin{aligned} &H : \text{indivíduo é hipertenso} \quad \overline{H} : \text{indivíduo não é hipertenso} \\ &A : \text{indivíduo ingere bebida alcólica} \quad \overline{H} : \text{indivíduo não ingere bebida alcólica} \end{aligned}$$

   Dados:

$$P[H] = 0,05 \quad , \quad P[A|H] = 0,75 \quad , \quad P[A|\overline{H}] = 0,50$$

   Portanto

$$P[\overline{H}] = 1 - P[H] = 0,95$$

   (a) $P[A] = P[A \cap H] + P[A \cap \overline{H}] = P[A|H] \cdot P[H] + P[A|\overline{H}] \cdot P[\overline{H}] = 0,75 \cdot 0,05 + 0,50 * 0,95 = 0.5125$
   (b) $P[H|A] = \frac{P[H \cap A]}{P[A]} = \frac{P[A|H] \cdot [H]}{P[A]} = \frac{0,75 \cdot 0.05}{0.5125} = 0.0732$

   Observação: o problema poderia ser organizado e resolvido utilizando uma tabela 2×2:

	H	$\overline{H}$	
A	$P[A \cap H] = 0.0375$	$P[A \cap \overline{H}] = 0.475$	$P[A] = 0.5125$
$\overline{A}$	$P[\overline{A} \cap H] = 0.0125$	$P[\overline{A} \cap \overline{H}] = 0.475$	$P[\overline{A}] = 0.4875$
	$P[H] = 0,05$	$P[\overline{H}] = 0,95$	1

---

3. Registos efetuados levaram a concluir que os motoristas que circulam em determinada estrada podem ter transgressões classificadas em dois tipos ditas do tipo I ou do tipo II, não se notando nenhum caso em que o motorista cometa ambas as transgressões. Por cada 500 motoristas multados há 100 motoristas multados por transgressões do tipo I. Sabendo que 10% dos motoristas que cometem transgressões do tipo I são multados; que 1% cometem transgressões do tipo I e que 2% cometem transgressões do tipo II, calcule a probabilidade de que um motorista que circule nessa estrada e cometa uma transgressão do tipo II seja multado.

Solução:
Notação:

I : transgressão do tipo I ; II : transgressão do tipo II ; 0 : sem transgressão
M : motorista recebe multa ; $\overline{M}$: motorista não recebe multa

Dados:

$$
\begin{aligned}
P[I \cap II] &= 0 \\
P[I|M] &= 100/500 = 0,20 \\
P[M|I] &= 0,10 \\
P[I] &= 0,01 \text{ e } P[II] = 0,02
\end{aligned}
$$

Queremos calcular $P[M|II]$.
A partir dos dados temos que:

$$
\begin{aligned}
P[II|M] &= 1 - P[1|M] = 0,80 \\
P[M|I] \cdot P[I] &= 0,10 \cdot 0,01 = 0,001
\end{aligned}
$$

Podemos obter $P[M]$:

$$
\begin{aligned}
P[I|M] &= 0,20 \\
\frac{P[M \cap I]}{P[M]} &= 0,20 \\
\frac{0,001}{P[M]} &= 0,20 \\
P[M] &= 0,005
\end{aligned}
$$

Como $P[M] = P[M \cap I] + P[M \cap II]$ temos que

$$
\begin{aligned}
P[M \cap II] &= 0,004 \\
P[M|II] \cdot P[II] &= 0,004 \\
P[M|II] &= 0,004/0,02 = 0,20
\end{aligned}
$$

Alternativamente, o problema pode ser esquamatizado na tabela:

	0	I	II	
M	$P[0 \cap M] = 0$	$P[I \cap M] = 0,001$	$P[II \cap M] = 0,004$	$0,005$
$\overline{M}$	$P[0 \cap \overline{M}] = 0,97$	$P[I \cap \overline{M}] = 0,009$	$P[II \cap \overline{M}] = 0,016$	$0,995$
	$P[0] = 0,97$	$P[I] = 0,01$	$P[II] = 0,02$	1

Semana 4 (av-03)

Um estudante vai fazer um teste no qual cada questão tem cinco alternativas das quais apenas uma é correta. Nos dois contextos a seguir defina a variável aleatória, seus possíveis valores, sua distribuição de probabilidades e calcule a probabilidade pedida, supondo-se acerto ao acaso (ou seja, o estudante "chuta" todas as questões).

1. **Contexto 1:** O estudante vai fazer cinco questões e deseja-se a probabilidade de acertar três ou mais questões.
2. **Contexto 2:** As questões são apresentadas sequencialmente e o teste se encerra quando o estudante erra alguma questão. Qual a probabilidade do estudante acertar três ou mais questões.

Itens extras discutidos em sala:

3. **Contexto 3:** Considere o mesmo que no **Contexto 2:**, só que encerrando quando erra a *terceira* questão.
4. **Contexto 4:** Considere que há um banco de 30 questões das quais o estudante sabe 10. Selecionam-se cinco questões e deseja-se saber a probabilidade de mais que três acertos.

Solução:

1. **Contexto 1:**

$$\begin{aligned}
& X : \text{número de acertos em cinco questões} \\
& x \in \{0,1,2,3,4,5\} \\
& X \sim \mathrm{B}(n=4, p=1/5) \\
P[X=x] &= \binom{n}{x} p^x (1-p)^{n-x} \\
& p : \text{probabilidade de acertar cada questão (“sucesso”)} \\
P[X \geq 3] &= P[X=3] + P[X=4] + P[X=5] = \\
&= \binom{5}{3} 0,2^3 (1-0,2)^{5-3} + \binom{5}{4} 0,2^4 (1-0,2)^{5-4} + \binom{5}{5} 0,2^5 (1-0,2)^{5-5} = 0.0579
\end{aligned}$$

2. **Contexto 2:**

$$\begin{aligned}
& X : \text{número de acertos até o primeiro erro} \\
& x \in \{0,1,2,3,\ldots\} \\
& X \sim \mathrm{G}(n=4, p=4/5) \\
P[X=x] &= (1-p)^x p \\
& p : \text{probabilidade de errar a questão (“sucesso”)} \\
P[X \geq 3] &= 1 - (P[X=0] + P[X=1] + P[X=2]) = 1 - (0,2^0 0,8 + 0,2^1 0,8 + 0,2^2 0,8) = 0.008
\end{aligned}$$

3. **Contexto 3:**

$$\begin{aligned}
& X : \text{número de acertos até o terceiro erro} \\
& x \in \{0,1,2,3,\ldots\} \\
& X \sim \mathrm{BN}(k=3, p=4/5) \\
P[X=x] &= \binom{x+k-1}{x} (1-p)^x p \\
& p : \text{probabilidade de errar a questão (“sucesso”)} \\
P[X \geq 3] &= 1 - (P[X=0] + P[X=1] + P[X=2]) = \\
&= 1 - (\binom{2}{0} 0,2^0 0,8 + \binom{3}{1} 0,2^1 0,8 + \binom{4}{2} 0,2^2 0,8) = 0.0579
\end{aligned}$$

4. **Contexto 4:**

$$\begin{aligned}
& X : \text{número de acertos em cinco questões sorteadas} \\
& x \in \{0,1,2,3,4,5\} \\
& X \sim \mathrm{HG}(N=30, K=10, n=5) \\
P[X=x] &= \frac{\binom{k}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \\
P[X \geq 3] &= P[X=3] + P[X=4] + P[X=5] = \\
&= \frac{\binom{10}{3}\binom{20}{2}}{\binom{30}{5}} + \frac{\binom{10}{4}\binom{20}{1}}{\binom{30}{5}} + \frac{\binom{10}{5}\binom{20}{5}}{\binom{30}{0}} = 0.1912
\end{aligned}$$

Soluções computacionais (linguagem R):

```
> pa <- pbinom(2, size=5, prob=1/5, lower=FALSE)
> pb <- pgeom(2, prob=4/5, lower=FALSE)
> pc <- pnbinom(2, size=3, prob=4/5, lower=FALSE)
> pd <- phyper(2, m=10, n=20, k=5, lower=FALSE)
```

Semana 5 (av-04)

1. Considere que serão feitas inspeções em veículos em uma determinada área para identificar e orientar a correção de irregularidades. Supõe-se que os veículos inspecionados são escolhidos ao acaso. Considere os diferentes cenários descritos em cada um dos itens a seguir, identifique a variável aleatória em questão, seus possíveis valores, sua distribuição de probabilidades e responda à questão formulada.

(a) Em um lote de 50 veículos sabe-se que 8 deles possuem alguma irregularidade. Serão inspecionados sete veículos. Qual a probabilidade de encontrar mais de um com alguma irregularidade?

(b) Sabendo que 15% dos veículos na área apresentam irregularidade, serão inspecionados veículos até que seja encontrado o segundo com irregularidade. Qual a probabilidade de que sejam feitas no máximo cinco inspeções?

(c) A partir de experiências anteriores sabe-se que são encontrados, em média, 1,8 carros irregulares por hora de inspeção. Qual a probabilidade de que em uma hora não seja encontrado nenhum carro irregular? E qual a probabilidade de que sejam encontrados exatamente cinco irregulares em duas horas de inspeção? (supondo ainda que 15% dos carros da região são irregulares)

(d) A partir de um certo momento decide-se que a inspeção vai terminar quando for encontrado o próximo veículo irregular. Qual a probabilidade de que a partir deste momento sejam ainda inspecionados três ou mais carros?

(e) Um inspetor vai inspecionar sete carros. Qual a probabilidade de que encontre mais que um irregular? (supondo ainda que 15% dos carros da região são irregulares)

Solução:

(a) **Contexto 1:**

$$
\begin{aligned}
& X : \text{número veículos com irregularidade (dentre os sete selecionados)} \\
& x \in \{0,1,2,3,4,5,6,7\} \\
& X \sim \mathrm{HG}(N=200, K=30, n=7) \\
P[X=x] &= \frac{\binom{k}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \\
P[X>1] &= P[X=2]+P[X=3]+\ldots P[X=7] = 1-\{P[X=0]+P[X=1]\} \\
&= 1-\left\{\frac{\binom{8}{0}\binom{42}{7}}{\binom{50}{7}}+\frac{\binom{8}{1}\binom{42}{6}}{\binom{50}{7}}\right\} = 0.3098
\end{aligned}
$$

(b) **Contexto 2:**

$$
\begin{aligned}
& Y : \text{número de inspeções até o segundo irregular} \\
& y \in \{2,3,4,5,\ldots\} \\
& X : \text{número de regulares até o segundo irregular} \\
& x \in \{0,1,2,3,\ldots\} \\
& X \sim \mathrm{BN}(k=2, p=0,15) \\
P[X=x] &= \binom{x+k-1}{x} p^k (1-p)^x \\
& p : \text{probabilidade de encontrar um veículo irregular (“sucesso”)} \\
P[Y \leq 5] = P[X \leq 3] &= P[X=0]+P[X=1]+P[X=2]+P[X=3] = \\
&= \binom{1}{0}0,15^2 0,85^0 + \binom{2}{1}0,15^2 0,85^1 + \binom{3}{2}0,15^2 0,85^2 + \binom{4}{1}0,15^2 0,85^3 = 0.7765
\end{aligned}
$$

(c) **Contexto 3:**

$$
\begin{aligned}
& X_1 : \text{número de irregulares por hora} \\
& x_1 \in \{0,1,2,3,\ldots\} \\
& X_1 \sim \mathrm{P}(\lambda=1,8) \\
P[X=x] &= \frac{e^{-\lambda}\lambda^x}{x!} \\
P[X_1=0] &= \frac{e^{-1,8}1,8^0}{0!} = 0.1653 \\
& X_2 : \text{número de irregulares por período de duas (2) horas} \\
& x_2 \in \{0,1,2,3,\ldots\} \\
& X_2 \sim \mathrm{P}(\lambda = 2 \cdot 1,8 = 3,6) \\
P[X_2=5] &= \frac{e^{-3,6}3,6^5}{5!} = 0.1377
\end{aligned}
$$

(d) **Contexto 4:**

$$\begin{aligned}
&Y : \text{número de inspeções até o próximo irregular}\\
&y \in \{1, 2, 3, 4, \ldots\}\\
&X : \text{número de regulares até o próximo irregular}\\
&x \in \{0, 1, 2, 3, \ldots\}\\
&X \sim \mathrm{G}(p = 0,15)\\
P[X = x] &= (1-p)^x p\\
&p : \text{probabilidade de encontrar um veículo irregular (“sucesso”)}\\
P[Y \geq 3] = P[X \geq 2] &= 1 - \{P[X = 0] + P[X = 1]\} =\\
&= 1 - \left\{0,15 \cdot 0,85^0 + 0,15 \cdot 0,85^1\right\} = 0.7225
\end{aligned}$$

(e) **Contexto 5:**

$$\begin{aligned}
&X : \text{número de irregulares em sete (7) inspeções}\\
P[X = x] &= \binom{n}{x} p^x (1-p)^{n-x}\\
&p : \text{probabilidade de acertar cada questão (“sucesso”)}\\
P[X > 1] &= 1 - \{P[X = 0] + P[X = 1]\} = 1 - \left\{\binom{5}{0} 0,15^0 (1-0,15)^{7-0} + \binom{5}{1} 0,15^1 (1-0,15)^{7-1}\right\} = 0.2834
\end{aligned}$$

Soluções computacionais (linguagem R):

```
> pa <- phyper(1, m=8, n=42, k=7, lower=FALSE)
> pb <- pnbinom(4, size=2, prob=0.15, lower=FALSE)
> pc1 <- dpois(0, lambda=1.8)
> pc2 <- dpois(5, lambda=2*1.8)
> pd <- pgeom(1, prob=0.15, lower=FALSE)
> pe <- pbinom(1, size=7, prob=0.15, lower=FALSE)
```

Semana 6 (av-05)

1. Uma determinada indústria classifica ovos como: XL acima de 73 g, L 63 a 73 g, M 53 a 63 g, S abaixo de 53 g. Suponha que um produtor produza ovos cujos tamanhos (pesos) são descritos pela seguinte função de densidade de probabilidades:

$$f(x) = \begin{cases} \frac{(x-48)}{12k} & \text{se } 48 \leq x < 60 \\ -\frac{(x-78)}{18k} & \text{se } 60 \leq x \leq 78 \\ 0 & \text{se } x < 48 \text{ ou } x > 78 \end{cases}$$

(a) Qual o valor de k ?

(b) Qual a proporção de ovos que deve ser produzida em cada classificação?

(c) Se o produtor recebe R\$ 0,05 por ovo S, R\$ 0,10 por ovo M, R\$ 0,12 por ovo L e R\$ 0,18 por ovo XL, quanto deve receber em um lote de 10.000 ovos?

(d) Qual o tamanho (peso) mediano dos ovos?

(e) Forneça a expressão da distribuição acumulada $F(x)$.

(f) Qual o tamanho (peso) para o qual apenas 20% dos ovos estão acima dele?

Solução:

(a) $k = 15$

(b) Pode-se resolver de três formas diferentes: geometricamente (áreas dos polígonos indicados na figura), integrando-se $f(x)$ ou avaliando-se e fazendo as diferenças dos valores de $F(x)$ nos pontos que definem as classificações.

$$\begin{aligned}
&C : \text{valor por ovo}\\
&c \in \{0,05; 0,10; 0,12; 0,18\}
\end{aligned}$$

c_i	$0,05$	$0,10$	$0,12$	$0,18$
$P[C = c_i]$	$P[X < 53] =$ 0.0694	$P[53 \leq X < 63] =$ 0.5139	$P[63 \leq X \leq 73] =$ 0.3704	$P[X > 73] =$ 0.0463

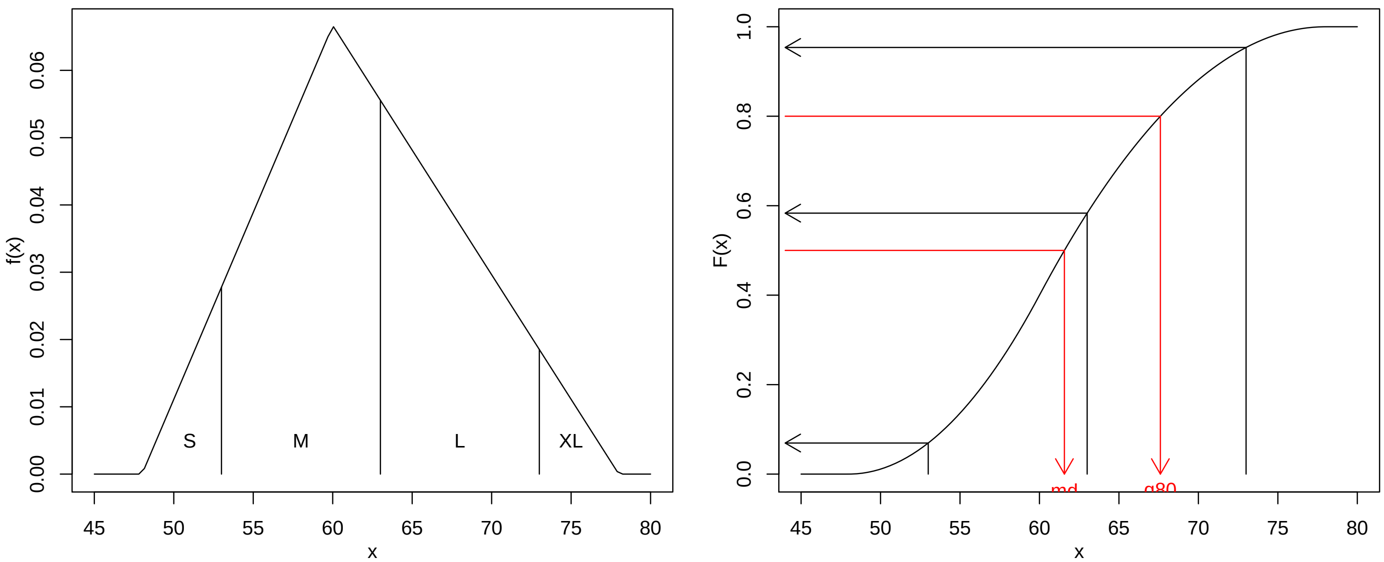


Figura 1: Funções de distribuição de probabilidades ($f(x)$) e acumulada ($F(x)$) do problema. Segmentos e setas são as soluções de alguns dos items do problema.

(c) $10,000 \cdot E[C] = 10.000 \cdot (0,05 \cdot 0.06944 + 0,10 \cdot 0.5139 + 0,12 \cdot 0.3704 + 0,18 \cdot 0.0463) = 10.000 \cdot 0.10764 = 1076.4$

(d) $md : \int_{md}^{78} f(x)dx = 0,5 \longrightarrow md = 61.6$

(e)

$$F(x) = \int_{48}^{x} f(x)dx = \begin{cases} 0 & \text{se } x < 48 \\ \frac{1}{180}\left[\frac{(x^2-48^2)}{2} - 48(x-48)\right] & \text{se } 48 \leq x < 60 \\ 0,4 - \frac{1}{270}\left[\frac{(x^2-60^2)}{2} - 78(x-60)\right] & \text{se } 60 \leq x < 78 \\ 1 & \text{se } x > 78 \end{cases}$$

(f) $q_{0,80} : \int_{q_{0,80}}^{78} f(x)dx = 0,20 \longrightarrow q_{0,80} = 67.6$

Soluções computacionais (linguagem R):

```
> ## definindo f(x)
> ddist <- function(x){
+     y <- numeric(length(x))
+     y[x >= 48 & x < 60] <- (x[x >= 48 & x < 60]-48)/180
+     y[x >= 60 & x < 78] <- -(x[x >= 60 & x < 78]-78)/270
+     return(y)
+ }
> ## definindo F(x)
> pdist <- function(x){
+     y <- numeric(length(x))
+     ind <- x >= 48 & x < 60
+     y[ind] <- ((x[ind]^2-48^2)/2 - 48*(x[ind]-48))/180
+     ind <- x >= 60 & x < 78
+     y[ind] <- 0.4 - ((x[ind]^2-60^2)/2 - 78*(x[ind]-60))/270
+     y[x >= 78] <- 1
+     return(y)
+ }
> ## definindo F^{-1}(x)
> qdist <- function(q){
+     uniroot(function(x) pdist(x) - q, interval=c(48,78))$root
+ }
> ## b) Proporções em cada classe
> ## integrando f(x)
> (PrS <- integrate(ddist, 48, 53)$value)
[1] 0.06944
> (PrM <- integrate(ddist, 53, 63)$value)
[1] 0.5139
> (PrL <- integrate(ddist, 63, 73)$value)
```

```
[1] 0.3704

> (PrXL <- integrate(ddist, 73, 78)$value)

[1] 0.0463

> ## utilizando F(x)
> (PC <- diff(pdist(c(48,53,63,73,78))))

[1] 0.06944 0.51389 0.37037 0.04630

> ## c) Valor médio por ovo
> (EC <- drop(crossprod(c(0.05, 0.10, 0.12, 0.18), PC)))

[1] 0.1076

> ## d) mediana
> (md <- qdist(0.5))

[1] 61.57

> ## f) quantil 0,80
> (q80 <- qdist(0.8))

[1] 67.61

> ## Gráficos
> par(mfrow=c(1,2), mar=c(3.5,3.5,1,1), mgp=c(2,1,0))
> curve(ddist, from=45, to=80, ylab="f(x)")
> segments(c(53,63,73), 0, c(53,63,73), ddist(c(53,63,73)))
> text(c(51, 58, 68, 75), 0.005, c("S","M","L","XL"))
> curve(pdist, from=45, to=80, ylab="F(x)")
> segments(c(53,63,73), 0, c(53,63,73), pdist(c(53,63,73)))
> arrows(c(53,63,73), pdist(c(53,63,73)), 44, pdist(c(53,63,73)), length=0.15)
> segments(44, 0.5,md, 0.5, col=2)
> arrows(md, 0.5, md,0, length=0.15, col=2)
> text(md, 0, "md", pos=1, col=2)
> segments(44, 0.8, q80, 0.8, col=2)
> arrows(q80, 0.8, q80,0, length=0.15, col=2)
> text(q80, 0, expression(q80), pos=1, col=2)
```

Semana 8 (av-06)

1. O conjunto de dados *studentdata* do pacote **LearnBayes** do programa **R** contém os registros de 657 questionários aplicados à estudantes. A tabela a seguir mostra os 10 primeiros registros dos questionários.

	Estudante	Altura	Sexo	Sapatos	Numero	DVDs	Dormiu	Acordou	Cabelo	Trabalho	Bebida
1	1	67	female	10	5	10	-2.5	5.5	60	30.0	water
2	2	64	female	20	7	5	1.5	8.0	0	20.0	pop
3	3	61	female	12	2	6	-1.5	7.5	48	0.0	milk
4	4	61	female	3	6	40	2.0	8.5	10	0.0	water
5	5	70	male	4	5	6	0.0	9.0	15	17.5	pop
6	6	63	female	NA	3	5	1.0	8.5	25	0.0	water
7	7	61	female	12	3	53	1.5	7.5	35	20.0	water
8	8	64	female	25	4	20	0.5	7.5	25	0.0	pop
9	9	66	female	30	3	40	-0.5	7.0	30	25.0	water
10	10	65	male	10	7	22	2.5	8.5	12	0.0	milk

As colunas se referem às seguintes questões:

- Estudante: número do estudante
- Altura: altura em polegadas
- Sexo: sexo (Masculino/Feminino)
- Sapatos: número de pares de sapato que possui
- Numero: um número escolhido entre 0 e 10
- DVDs: número de DVD's de filmes que possui
- Dormiu: hora que foi dormir na noite anterior (em relação à meia noite)
- Acordou: hora que acordou na manhã seguinte
- Cabelo: custo do último corte de cabelo
- Trabalho: número de horas (semanais) de trabalho

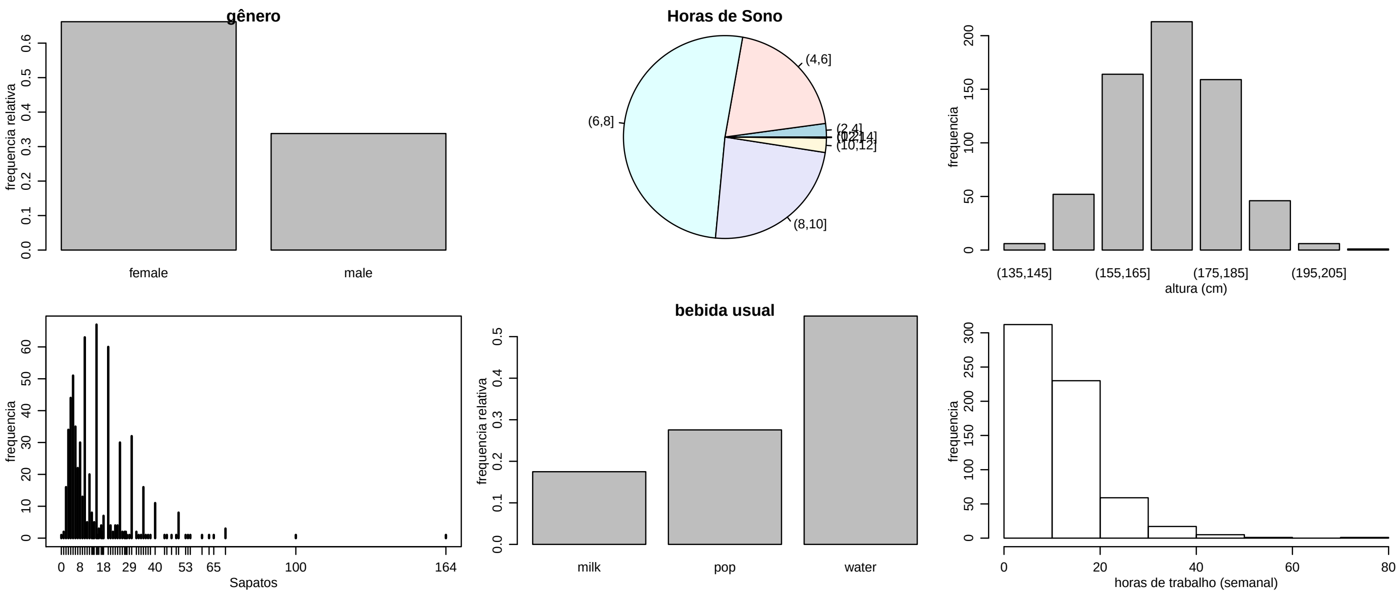


Figura 2: Gráficos do questionário aplicado aos estudantes

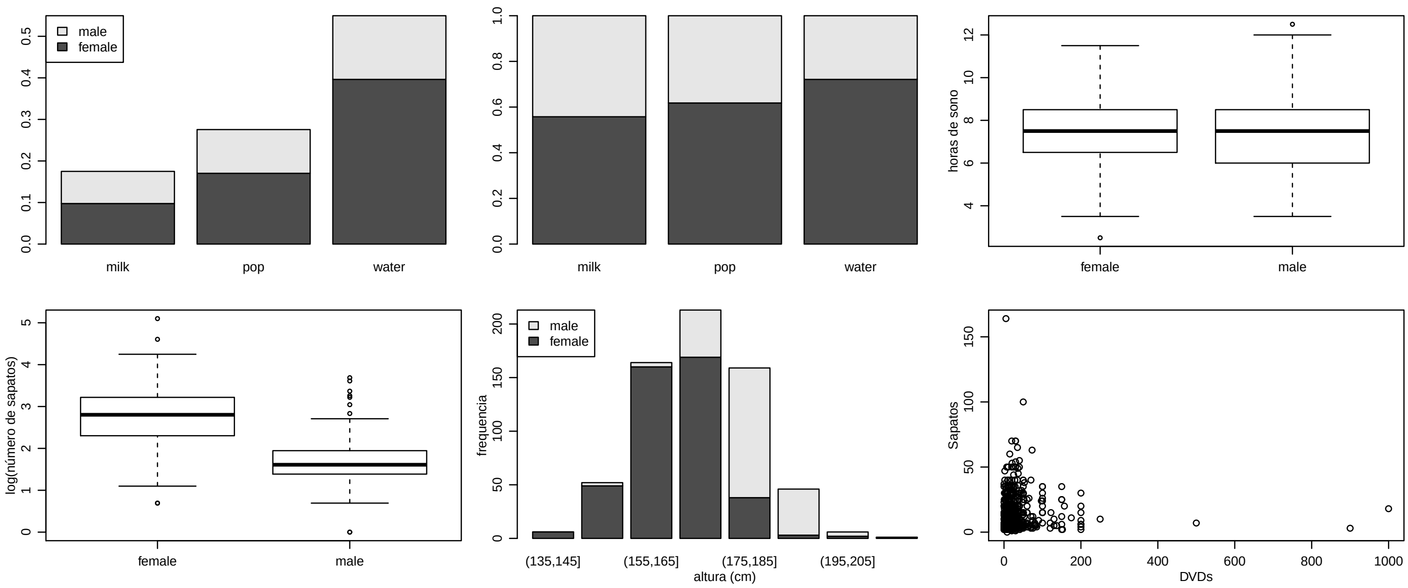


Figura 3: Gráficos do questionário aplicado aos estudantes

- Bebida: bebida usual na janta (água, leite, suco/refrigerante)

(a) Considere os gráficos mostrados a seguir. Para cada um deles comente sua interpretação, se o gráfico é ou não o mais adequado e, caso não seja, esboce o gráfico que seria mais adequado.

(b) Interprete os gráficos e resultados neles mostrados.

Semana 9 (av-07)

1. O conjunto de dados *chickwts* disponível no programa estatístico **R** apresenta o peso de frangos submetidos a diferentes dietas. Durante as análises foi construído o gráfico da figura 4. Discuta os resultados e possíveis recomendações práticas.

2. 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.

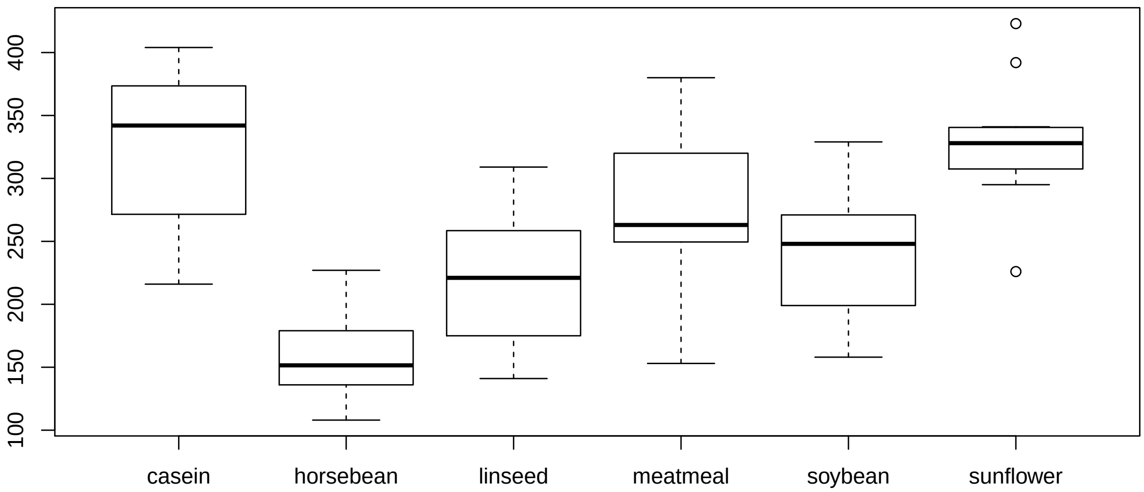


Figura 4: Peso final de frangos submetidos à diferentes dietas

(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.

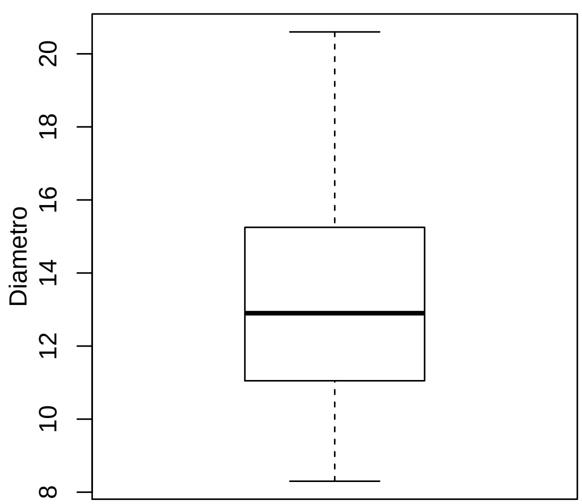
Solução:

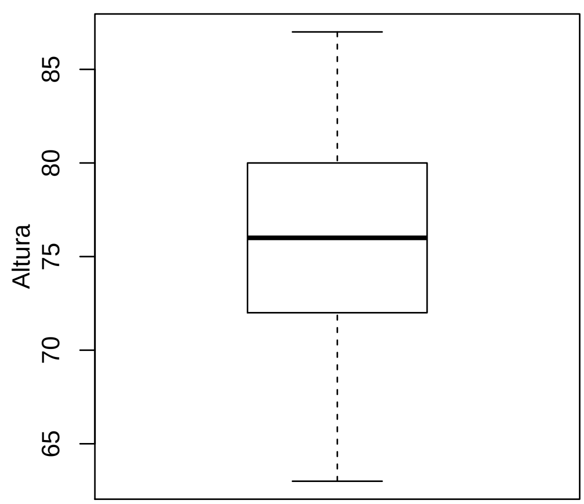
(a)

```
The decimal point is at the |

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

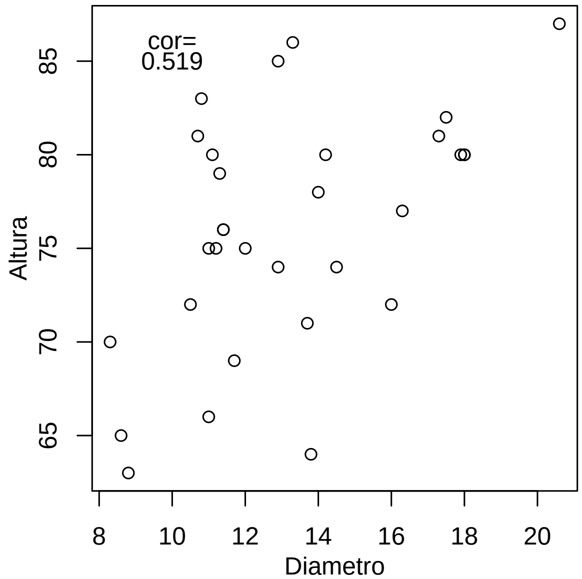
(b)





(c)

(d)



Semana 10 (av-08)

1. Em uma avaliação de um novo algoritmo de classificação foi analisada uma amostra de 1200 cenários dentre os quais 780 foram classificados corretamente.
   (a) Identifique no contexto do problema: a população ("informal" e estatística), a amostra, o parâmetro, o estimador, a distribuição amostral e a estimativa (pontual).
   (b) Obtenha a estimativa pontual e intervalar (95% de confiança) para a proporção de classificações corretas.
   (c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse de 1,5% (com 95% de confiança)?
   (d) Um algoritmo atualmente utilizado possui um percentual de acerto de 62%. Há evidências baseadas no estudo de que o novo algoritmo tem desempenho diferento do utilizado atualmente? Justifique sua resposta.
   (e) Quais as suposições relavantes para os cálculos feitos nos item anteriores?

Solução:

$$X : \text{resultado da classificação (correto/incorreto)}$$
$$X \sim \mathrm{B}(p) \quad \mathrm{E}[X] = p \quad \mathrm{Var}[X] = p(1-p)$$
$$\hat{p} \sim \mathrm{N}(p, p(1-p)/n)$$

(a)

(b)

$$\hat{p} = \frac{780}{1200} = 0.65$$

I.C. assintótico:

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow (0.623 \;\; ; \;\; 0.677)$$

I.C.conservador :

$$IC_{95\%} : \hat{p} \pm z\sqrt{\frac{1}{4n}} \longrightarrow (0.622 \;\; ; \;\; 0.678)$$

(c) i. Utilizando $p = \hat{p}$

$$\mathrm{ME} = z_{95\%}\sqrt{\frac{p(1-p)}{n}}$$
$$0.015 = 1.96\sqrt{\frac{0.65(1-0.65)}{n}}$$
$$n = \lceil \frac{1.96^2}{0,015^2} 0.65(1-0.65)$$
$$n = 3885$$

ii. Utilizando $p = 0,5$

$$
\begin{aligned}
\text{ME} &= z_{95\%}\sqrt{\frac{0,5(1-0,5)}{n}} = z_{95\%}\sqrt{\frac{1}{4n}} \\
0.015 &= 1.96\sqrt{\frac{0.5(1-0.5)}{n}} \\
n &= \lceil \frac{1.96^2}{0,015^2}\frac{1}{4} \\
n &= 4269
\end{aligned}
$$

(d) Resposta e justificativa baseada no valor estar ou não contido no I.C..

(e)

Semana 2 (av-09)

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:

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

(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) $P[\overline{X}_{16} < 95 | \overline{X}_{16} > 105] = P[Z < \frac{110-100}{12/\sqrt{16}} | Z > \frac{110-100}{12/\sqrt{16}}] = P[Z < -1.667 | Z > 1.667] = 0.904$

(e)

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

Soluções computacionais (linguagem R):

```
> 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 <- diff(pnorm(c(95,105), m=100, sd=12/sqrt(16)))
> n <- ceiling((qnorm(0.99)*12/(105-100))^2)
```

---

2. Explique os conceitos de ”estimador não viesado”e ”estimador mais eficiente”.

3. Uma amostra aleatória de uma população com distribuição de Poisson forneceu os seguintes valores: 1, 2, 0, 2, 2, 2, 5, 2, 3, 1, 2, 8. Obtenha o estimador de máxima verossimilhança para o parâmetro desta distribuição e o valor da estimativa para a amostra em questão.

Solução:

$$
\begin{aligned}
X &\sim \mathrm{P}(\lambda) \\
P[X = x_i] &= \frac{\mathrm{e}^{-\lambda}\lambda^{x_i}}{x_i!} \\
l(\lambda) &= \log\{L(\lambda)\} = \log\{\prod_{i=1}^{n} P[X = x_i]\} \\
&= \sum_{i=1}^{n} \log\{P[X = x_i]\} = \sum_{i=1}^{n} \log\{\frac{\mathrm{e}^{-\lambda}\lambda^{x_i}}{x_i!}\} \\
&= \sum_{i=1}^{n} [-\lambda + x_i \log(\lambda) - \log(x_i!)] = -n\lambda + \log(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log(x_i!)] \frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} \qquad = -n + \frac{\sum_{i=1}^{n} x_i}{\hat{\lambda}} = 0 \\
\hat{\lambda} &= \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}
\end{aligned}
$$

Para a amostra dada a estimativa é:

$$\hat{\lambda} = \frac{1 + 2 + \cdot + 2 + 8}{12} = 2,5$$
