CE-003: Estatística II - Turma: K/O, Avaliações Semanais (2^o semestre/2012)

1.

(a) Uma urna contém doze bolas brancas e oito bolas vermelhas. Serão retiradas, sequencialmente, três bolas da urna. A cada bola anota-se a cor e, se a bola for vermelha ela é retornada à urna e se for branca ela é posta de lado.

i. Forneça o espaço amostral do experimento.
ii. Calcule probabilidade de cada elemento do espaço amostral.
iii. Qual a probabilidade de não se obter todas as bolas da mesma cor?
iv. Qual a probabilidade de se retirar ao menos duas bolas brancas?
v. Qual a probabilidade de retirar três vermelhas sabendo-se que ao menos uma das bolas é vermelha?
vi. Se a primeira bola for branca, qual a probabilidade de obter três bolas brancas?

Solução:

i. $\Omega = \{(B,B,B),(B,B,V),(B,V,B),(V,B,B),(B,V,V),(V,B,V),(V,V,B),(V,V,V)\}$

ii.

Evento	(B,B,B)	(B,B,V)	(B,V,B)	(B,B,V)	(B,V,V)	(V,B,V)	(V,V,B)	(V,V,V)
Probabilidade	$\frac{12}{20}\frac{11}{19}\frac{10}{18}$	$\frac{12}{20}\frac{11}{19}\frac{8}{18}$	$\frac{12}{20}\frac{8}{19}\frac{11}{19}$	$\frac{8}{20}\frac{12}{20}\frac{11}{19}$	$\frac{12}{20}\frac{8}{19}\frac{8}{19}$	$\frac{8}{20}\frac{12}{20}\frac{8}{19}$	$\frac{8}{20}\frac{8}{20}\frac{12}{20}$	$\frac{8}{20}\frac{8}{20}\frac{8}{20}$

iii. $P = 1 - P[(B,B,B)] - P[(V,V,V)] = 1 - \frac{12}{20}\frac{11}{19}\frac{10}{18} - \frac{8}{20}\frac{8}{20}\frac{8}{20} = 0.743$

iv. $P = P[(B,B,B)] + P[(B,B,V)] + P[(B,V,B)] + P[(V,B,B)] = \frac{12}{20}\frac{11}{19}\frac{10}{18} + \frac{12}{20}\frac{11}{19}\frac{8}{18} + \frac{12}{20}\frac{8}{19}\frac{11}{19} + \frac{8}{20}\frac{12}{20}\frac{11}{19} = 0.6326$

v. $P = \frac{P[(V,V,V)]}{1-P[(B,B,B)]} = \frac{\frac{8}{20}\frac{8}{20}\frac{8}{20}}{1-\frac{12}{20}\frac{11}{19}\frac{10}{18}} = 0.0793$

vi. $P = \frac{P[(B,B,B)]}{P[(B,B,B)]+P[(B,B,V)]+P[(B,V,B)]+P[(B,V,V)]} = \frac{\frac{12}{20}\frac{11}{19}\frac{10}{18}}{\frac{12}{20}\frac{11}{19}\frac{10}{18}+\frac{12}{20}\frac{11}{19}\frac{8}{18}+\frac{12}{20}\frac{8}{19}\frac{11}{19}+\frac{12}{20}\frac{8}{19}\frac{8}{19}} = 0.3216$

(b) Um candidato está fazendo uma prova de múltipla escolha com cinco alternativas das quais apenas uma é correta. A chance do candidato saber a solução de uma questão é de 40%. Quando ele sabe a solução ele sempre acerta a questão e quando não sabe ele escolhe uma das respostas ao acaso. Se o candidato acerta a questão, qual a probabilidade de ele saber resolver a questão?

Solução:
Evento $S:$ o candidato sabe a questão
Evento $\overline{S}:$ o candidato sabe a questão
Evento $A:$ o candidato acerta a questão
Evento $\overline{A}:$ o candidato acerta a questão Dados:

$$\begin{aligned}
P[S] &= 0,40 \quad ; P[\overline{S}] = 0,60 \\
P[A|S] &= 1,00 \quad ; P[\overline{A}|S] = 0,00 \\
P[S] &= 0,40 \quad ; P[\overline{S}] = 0,60 \\
P[A|\overline{S}] &= 0,20 \quad ; P[\overline{A}|\overline{S}] = 0,80 \\
P[S|A] &= ? \\
P[S|A] &= \frac{P[S \cap A]}{P[A]} = \frac{P[A|S] \cdot P[S]}{P[A|S] \cdot P[S] + P[A|\overline{S}] \cdot P[\overline{S}]} = \frac{1 \cdot 0,40}{(1 \cdot 0,40) + (0,20 \cdot 0,60)} = \frac{0,40}{0,52} = 0.769
\end{aligned}$$

---

2. Estamos interessados nos tempos de processamento para um certo procedimento de tratamento de imagens. O algoritmo de tratamento e classificação das imagens funciona em dois estágios. O primeiro estágio é realizado em 20 segundos e a experiência mostra que a classificação é encerrada nesse estágio para 25% das imagens. As demais são processadas em um segundo estágio e destas, o processamento de 80% delas é encerrado com mais 30 segundos e 60 segundos para as restantes.

Defina a variável aleatória (v.a.), forneça sua distribuição de probabilidades, a esperança e a variância da v.a. Informe ainda o tempo que espera-se gastar no processamento de 1500 imagens.

Solução:

Eventos:

A : Classifica no primeiro estágio

$\overline{A}$: Não classifica no primeiro estágio

B : Classifica em 30 seg no segundo estágio

$\overline{B}$: Classifica em 60 segundos no segundo estágio

$\Omega = \{(A), (\overline{A}, B), (\overline{A}, \overline{B})\}$

Dados:

$$P[A] = 0,25 \quad P[B|\overline{A}] = 0,80$$
$$P[\overline{A}] = 0,75 \quad P[\overline{B}|\overline{A}] = 0,20$$
$$P[B \cap \overline{A}] = P[B|\overline{A}]P[\overline{A}] = 0,80 \cdot 0,75 = 0,60$$
$$P[\overline{B} \cap \overline{A}] = P[\overline{B}|\overline{A}]P[\overline{A}] = 0,20 \cdot 0,75 = 0,15$$

X : tempo de processamento (s)

$x \in \{20, 50, 80\}$

Evento	(A)	$(\overline{A}, B)$	$(\overline{A}, \overline{B})$
x	20	50	80
P[X=x]	$0,25$	$0,60$	$0,15$

$$\mathrm{E}(X) = \sum x \cdot P[X = x] =$$
$$= 20(0,25) + 50(0,60) + 80(0,15) = 47s$$
$$\mathrm{Var}(X) = \sum (x - \mathrm{E}(X))^2 \cdot P[X = x] =$$
$$= (20 - \mathrm{E}(X))^2(0,25) + (50 - \mathrm{E}(X))^2(0,60) + (80 - \mathrm{E}(X))^2(0,15) = 351s^2$$
$$T = 1500 \cdot \mathrm{E}(X) = 70500s = 19.58hr(\text{tempo para processamento de 1500 imagens})$$

---

3. Seja uma variável aleatória X (v.a.) com função de densidade de probabilidade (f.d.p.) dada por:

$$f(x) = \begin{cases} |1 - x| & \text{se } 0 < x < 2 \\ 0 & \text{caso contrário.} \end{cases}$$

(a) Mostre que $f(x)$ é uma f.d.p..

(b) Obtenha o valor esperado de X.

(c) Obtenha a função de distribuição acumulada $F(x)$

(d) Obtenha $P[X > 1]$.

(e) Obtenha $P[X < 1,5|X > 1]$.

(f) Obtenha $P[X < 0.25 \text{ ou } X > 0,75]$.

Seja ainda Y uma outra v.a. definida a partir da variável X anterior tal que:

$$Y = \begin{cases} 200 & \text{se } x \leq 0,25 \\ 500 & \text{se } 0,25 < x \leq 0,75 \\ 1000 & \text{se } x > 0,75 \end{cases}$$

(g) Obtenha a função de probabilidade Y.

(h) Obtenha o valor esperado de Y.

(i) Obtenha a função de distribuição acumulada $F(y)$

(j) Obtenha $P[Y = 1000|Y \geq 500]$.

Solução:

(a) Mostrar que $f(x) \geq 0 \; \forall x$ e que $\int_0^2 f(x)\mathrm{d}x = 1$.
Notar que:

$$f(x) = |1 - x| I_{(0,2)}(x) = \begin{cases} |1 - x| & \text{se } 0 < x < 2 \\ 0 & \text{caso contrário.} \end{cases} = \begin{cases} 1 - x & \text{se } 0 < x < 1 \\ x - 1 & \text{se } 1 < x < 2 \\ 0 & \text{caso contrário.} \end{cases}$$

(b) $E[X] = \int_0^2 f(x)\mathrm{d}x = \int_0^1 x(1-x)\mathrm{d}x + \int_1^2 x(x-1)\mathrm{d}x \ldots = 1$

(c)

$$F(x) = \begin{cases} \int_0^x 1 - x\mathrm{d}x = \frac{x(2-x)}{2} & \text{se } 0 < x \leq 1 \\ \int_0^1 1 - x\mathrm{d}x + \int_1^x x - 1\mathrm{d}x = \frac{x^2-2x+2}{2} & \text{se } 1 < x < 2 \\ 0 & \text{caso contrário.} \end{cases}$$

(d) $P[X > 1] = 1 - F(1) = 0,5$

(e) $P[X < 1,5|X > 1] = \frac{P[1<X<1,5]}{P[X>1]} = \frac{F(1,5)-F(1)}{1-F(1)}$

(f) $P[X < 0,25 \text{ ou } X > 0,75] = F(0,25) + (1 - F(0,75))$

(g)

Y	200	500	1000
$P[Y = y]$	$P[X \leq 0,25] =$ $F(0,25) =$ 0.21875	$P[0,25 < X \leq 0,75] =$ $F(0,75) - F(0,25) =$ 0.25	$P[X > 0,75] =$ $1 - F(0,75) =$ 0.53125

(h)

$$\mathrm{E}[Y] = \sum yP[Y = y] = 200 \cdot 0.21875 + 500 \cdot 0.25 + 1000 \cdot 0.53125 = 700$$

(i)

Y	$y < 200$	$200 \leq y < 500$	$500 \leq y1000$	$y \geq 1000$
$F(y) = P[Y \leq y]$	0	0.21875	0.46875	1

(j) $P[Y = 1000|Y \geq 500] = \frac{P[Y=1000]}{P[Y=500]+P[Y=1000]} = 17/25 = 0.68.$

---

4. 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)

$$X : \text{Número de acertos até o primeiro erro}$$
$$X \sim \mathrm{G}(0,25)$$
$$P[X \geq 4] = 1 - P[X \leq 3] = 1 - \sum_{i=0}^{1} (1 - 0,25)^i (0,25) = 0.316$$

(b)

$$X : \text{Número de acertos em cinco perguntas}$$
$$X \sim \mathrm{B(n = 5, p = 0,75)}$$
$$P[X \geq 4] = P[X = 4] + P[X = 5] = \sum_{i=4}^{5} \binom{5}{i} 0,75^i (1 - 0,75)^{5-i} = 0.633$$

(c)

$$
\begin{aligned}
&X : \text{Número de erros até o terceito acerto}\\
&X \sim \mathrm{BN(r=3, p=0,75)}\\
P[X=1] &= \binom{3+1-1}{3-1} 0,75^3 (1-0,75)^1 = 0.316
\end{aligned}
$$

(d)

$$
\begin{aligned}
&X : \text{Número de acertos nas seis questões selecionadas}\\
&X \sim \mathrm{HG}(30, 10, 6)\\
P[X \geq 5] = P[X=5] + P[X=6] &= \sum_{i=5}^{6} \frac{\binom{30}{i}\binom{10}{6-i}}{\binom{40}{6}} = 0.526
\end{aligned}
$$

(e)

$$
\begin{aligned}
&X : \text{Número de questões respondidas em 3 minutos}\\
&X \sim \mathrm{P}(3 \cdot 1,8 = 5,4)\\
P[X \geq 3] = 1 - P[X \leq 2] = 1 - &\sum_{i=0}^{2} \frac{\mathrm{e}^{-5,4} 5,4^{\mathrm{i}}}{i!} = 0.905
\end{aligned}
$$

(f)

$$
\begin{aligned}
&X : \text{tempo (em min.) para responder uma questão}\\
&X \sim \mathrm{Exp}(\lambda = 1,8)\\
P[X \geq 40/60] &= \int_{40/60}^{\infty} 1,8\mathrm{e}^{-1,8\mathrm{x}} \mathrm{dx} = 0.301
\end{aligned}
$$

---

5. Seja uma v.a. X com distribuição normal de média $\mu = 250$ e variância $\sigma^2 = 225$. Obtenha:

(a) $P[X > 270]$.
(b) $P[X < 220]$.
(c) $P[|X - \mu| > 25]$.
(d) $P[|X - \mu| < 30]$.
(e) $P[X < 270|X > 250]$.
(f) o valor x_1 tal que $P[X > x_1] = 0,80$.
(g) o valor x_2 tal que $P[X < x_2] = 0,95$.
(h) qual deveria ser um novo valor da média μ para que $P[X < 240] \leq 0,10$?
(i) com $\mu = 250$ qual deveria ser um novo valor da variância σ^2 para que $P[X < 240] \leq 0,10$?
(j) qual deveria ser um novo valor da variância σ^2 para que $P[|X - \mu| > 15] \leq 0,10$?

$$X \sim \mathrm{N}(250, 15^2)$$

(a) $P[X > 270] = P[Z > \frac{270-250}{15}] = P[Z > 1.3333] = 0.0912$
(b) $P[X < 220] = P[Z < \frac{220-250}{15}] = P[Z < -2] = 0.0228$
(c) $P[|X - \mu| > 25] = P[X < 225 \cup X > 275] = P[X < -1.667] + P[X > 1.667] = 0.0956$
(d) $P[|X - \mu| < 30] = P[220 < X < 280] = P[-2 < X < 2] = 0.9545$
(e) $P[X < 270|X > 250] = \frac{P[250<X<270]}{P[X>250]} = \frac{0.4088}{0.5} = 0.8176$
(f) $z = \frac{x_1-250}{15} = -0.842 \longrightarrow x_1 = 237.4$
(g) $z = \frac{x_2-250}{15} = 1.645 \longrightarrow x_2 = 274.7$
(h) $z = \frac{240-\mu}{15} = -1.282 \longrightarrow \mu = 259.2$
(i) $z = \frac{240-250}{\sigma} = -1.282 \longrightarrow \sigma = 7.8 \longrightarrow \sigma^2 = 60.8$
(j) $P[|X < \mu| > 15] = P[X < \mu - 15 \cup X > \mu + 15] \leq 0,10 \longrightarrow z = \frac{15}{\sigma} = 1.645 \longrightarrow \sigma = 9.1 \longrightarrow \sigma^2 = 83.1$

Comandos em R para soluções:

```
> (qa <- pnorm(270, mean=250, sd=15, lower=FALSE))

[1] 0.09121

> (qb <- pnorm(220, mean=250, sd=15))

[1] 0.02275

> (qc <- 2*pnorm(250-25, mean=250, sd=15))

[1] 0.09558

> (qd <- diff(pnorm(c(250-30,250+30), mean=250, sd=15)))

[1] 0.9545

> (qe <- diff(pnorm(c(250,270), mean=250, sd=15))/pnorm(250, mean=250, sd=15, lower=FALSE))

[1] 0.8176

> (qf <- qnorm(0.80, mean=250, sd=15, lower=FALSE))

[1] 237.4

> (qg <- qnorm(0.95, mean=250, sd=15))

[1] 274.7

> (qh <- 240 - 15 * round(qnorm(0.10), dig=3))

[1] 259.2

> (qi <- (240 - 250)/round(qnorm(0.10), dig=3))

[1] 7.8

> (qj <- 15/round(qnorm(0.95), dig=3))

[1] 9.119
```

---

6. Os dados abaixo são provenientes de uma base de dados referentes a especificações técnicas de diversos modelos de automóveis[1]. Os dados mostrados são um extrato de 6 de um total de 93 modelos de veículos disponíveis na tabela de dados e alguns dos atributos foram omitidos.

```
  Manufacturer   Model    Type Price MPG.city MPG.highway            AirBags DriveTrain Cylinders EngineSize Horsepower
1        Acura Integra   Small  15.9       25          31               None      Front         4        1.8        140
2        Acura  Legend Midsize  33.9       18          25 Driver & Passenger      Front         6        3.2        200
3         Audi      90 Compact  29.1       20          26        Driver only      Front         6        2.8        172
4         Audi     100 Midsize  37.7       19          26 Driver & Passenger      Front         6        2.8        172
5          BMW    535i Midsize  30.0       22          30        Driver only       Rear         4        3.5        208
6        Buick Century Midsize  15.7       22          31        Driver only      Front         4        2.2        110
  Man.trans.avail Fuel.tank.capacity Passengers Length Width Rear.seat.room Luggage.room Weight  Origin
1             Yes               13.2          5    177    68           26.5           11   2705 non-USA
2             Yes               18.0          5    195    71           30.0           15   3560 non-USA
3             Yes               16.9          5    180    67           28.0           14   3375 non-USA
4             Yes               21.1          6    193    70           31.0           17   3405 non-USA
5             Yes               21.1          4    186    69           27.0           13   3640 non-USA
6              No               16.4          6    189    69           28.0           16   2880     USA
```

(a) Caracterize cada um dos atributos (variáveis) quanto ao seu tipo

(b) Esboce como seria um gráfico adequado para representar cada variável

(c) Escolha quatro relações de possível interesse entre duas variáveis e indique que tipo de análise seria feita para investigar cada uma das relações.

(d) Mostre como poderia ser feito um único gráfico que contivesse informações entre *Type*, *Weight* e *MPG.city*.

7.

[1] http://www.amstat.org/publications/jse/v1n1/datasets.lock.html

(a) 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
iii. Obtenha os percentis 10 e 90.

(b) A concentração de bactérias foi medida em um conjunto de amostras e os resultados foram resumidos na tabela a seguir.

Concentração	Número de amostras
$[0, 200)$	50
$[200, 400)$	65
$[400, 800)$	70
$[800, 1200)$	10
$[1200, 2000]$	5
Total	200

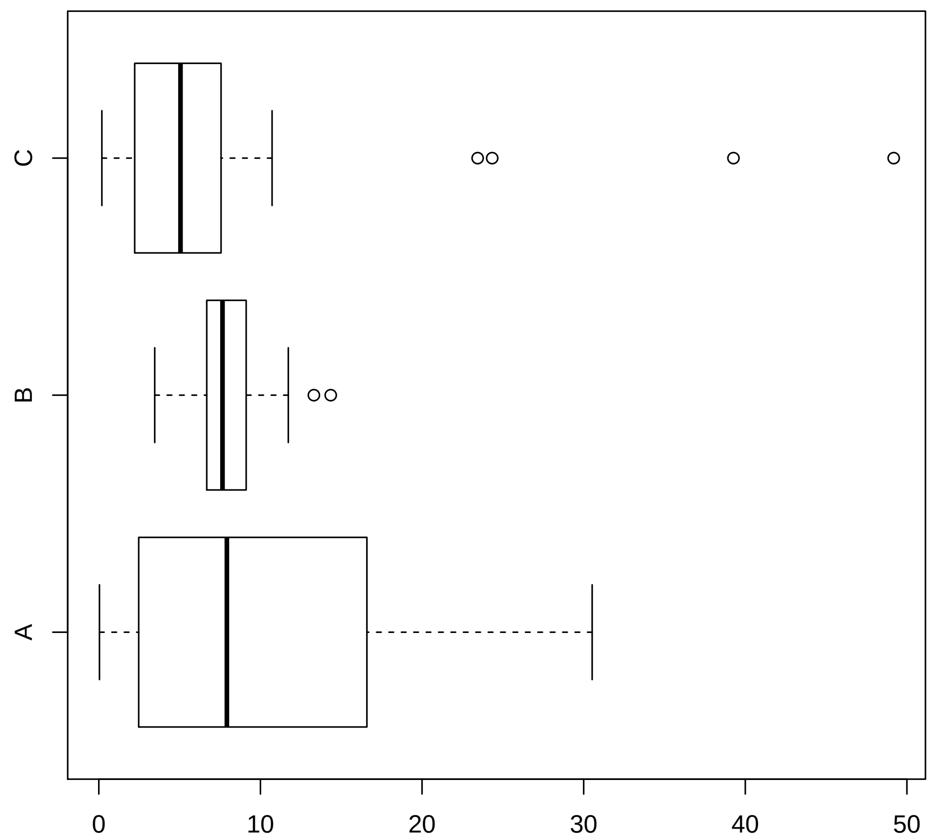
Assinale a alternativa verdadeira

a) a concentração média é de aproximadamente 600 unidades
b) a concentração média é de aproximadamente 354 unidades
c) a moda da concentração é de aproximadamente 600 unidades
d) a concentração mediana é de aproximadamente 354 unidades
e) a concentração mediana é de 600 unidades

(c) A média de uma distribuição de uma variável aleatória é 50, a mediana é 60 e a moda é 65. É mais provável que a distribuição seja:

a) assimétrica à esquerda
b) assimétrica à direita
c) bimodal
d) simétrica
e) assintótica

(d) Teores de um certo poluente foram medidos em três localidades distintas e os dados foram representados no gráfico a seguir. Baseando-se no gráfico, qual das afirmações claramente **não** pode ser feita?



i. Os valores no local C possuem uma distribuição assimétrica.
ii. Os dados discrepantes do local A afetam (aumentam) a média do local.
iii. Os locais B e C possuem médias semelhantes e desvios padrão distintos
iv. O local B possui o maior coeficiente de variação.
v. As médias dos três locais devem ser semelhantes.

8.

(a) Considere uma pesquisa eleitoral na qual deseja-se estimar a intenção de voto de um candidato através de uma amostra aleatória simples.

i. Qual deveria ser o tamanho da amostra para se obter uma margem de erro de 1,5% com 95% de confiança?
ii. E para uma margem de erro de 3% com 90% de confiança?

$$
\begin{aligned}
X &\sim B(p) \\
\text{Teo2}: \ \hat{p} = \overline{X} &\approx N(p, \frac{p(1-p)}{n}) \\
\hat{p} &\pm z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} \\
\hat{p} &\pm \text{ME} \\
\text{ME} &= z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}} \\
n &= \frac{(z_{1-\alpha})^2}{\text{ME}^2}p(1-p) \\
p(1-p) &\text{ é limitado superiormente para } p = 0,5 \\
n &= \frac{(z_{1-\alpha})^2}{\text{ME}^2}0,25
\end{aligned}
$$

i. $n = \frac{(1{,}96)^2}{(0{,}015)^2} 0,25 = 4269$

ii. $n = \frac{(1{,}645)^2}{(0{,}03)^2} 0,25 = 752$

(b) Um veículo transporta até 12 passageiros e uma carga máxima de 1200 kg, incluindo pesos dos passageiros e bagagens. O peso dos passageiros possui distribuição normal de média 76 kg e desvio padrão de 15 kg. O peso das bagagens dos passageiros possui distribuição normal de média 22 kg e desvio padrão de 5 kg.

i. Qual a probabilidade de um passageiro (com sua bagagem) ultrapassar 100 kg?

ii. Qual a probabilidade do peso total de 12 passageiros (com sua bagagem) ultrapassar a carga máxima?

iii. Qual a probabilidade do peso total de 5 passageiros (com sua bagagem) ultrapassar 600 kg?

$$\begin{aligned}
X_1 &: \text{ peso do passageiro; } X_1 \sim N(\mu_1 = 76, \sigma_1^2 = 15^2) \\
X_2 &: \text{ peso da bagagem; } X_2 \sim N(\mu_2 = 22, \sigma_2^2 = 5^2) \\
&\quad \text{assumindo independência,} \\
X = X_1 + X_2 &: \text{ peso do passageiro com sua bagagem; } X \sim N(\mu = 98, \sigma^2 = 15^2 + 5^2 = 250) \\
\text{Teo1} &: \hat{\mu} = \overline{X}_n \sim N(\mu = 98, \sigma^2/n = 250/n)
\end{aligned}$$

i. $P[X > 100] = P[Z > \frac{100-98}{\sqrt{250}}] = 0.4497$

ii. $P[\sum_{i=1}^{12} X_i > 1200] = P[\overline{X}_{12} > 100] = P[Z > \frac{100-98}{\sqrt{250/12}}] = 0.3306$

iii. $P[\sum_{i=1}^{5} X_i > 600] = P[\overline{X}_5 > 120] = P[Z > \frac{120-98}{\sqrt{250/5}}] = 0.00093$

iv. $P[\sum_{i=1}^{6} X_i > 600] = P[\overline{X}_6 > 100] = P[Z > \frac{100-98}{\sqrt{250/6}}] = 0.3783$
