Estatística - 2023.1

Todos os PDFs do semestre
juntados num PDFzão só

Eduardo Ochs - RCN/PURO/UFF
http://anggtwu.net/2023.1-ES.html

Estatística - 2023.1

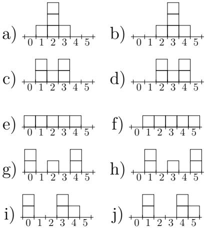
Primeira prova (P1)

Eduardo Ochs - RCN/PURO/UFF

http://anggtwu.net/2023.1-ES.html

Questão 1
(Total: 2.0 pts)

Calcule a média, o desvio médio e a variância de cada uma das distribuições abaixo.

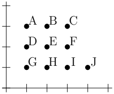


Não é preciso simplificar as frações. Arrume os seus resultados em três tabelas com essa cara aqui:

Média		Desvio médio		Variância	
a)	b)	a)	b)	a)	b)
c)	d)	c)	d)	c)	d)
e)	f)	e)	f)	e)	f)
g)	h)	g)	h)	g)	h)
i)	j)	i)	j)	i)	j)

Questão 2
(Total: 4.0 pts)

Na turma que eu usei num monte de exemplos as crianças se chamavam Ana, Bia, Carlos, Dani, Eduardo, Fábio, Geraldo, Heloá, Inês e Joana. O diagrama abaixo diz quantas balas "X" e quantas balas "Y" cada criança tem – por exemplo, a Joana tem quatro balas X e uma bala Y.



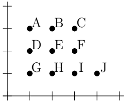
a) Transforme as informações do diagrama acima numa tabela. A sua tabela deve ter pelo menos estas colunas aqui: i, nome$_i$, X_i, Y_i, e $Z_i = (X_i - Y_i)$ – mas você pode acrescentar outras se você achar que isto vai te ajudar a responder os outros itens.

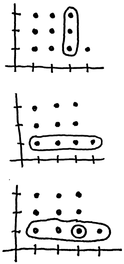
b) Calcule estas probabilidades:

$$\begin{array}{lll} P(X \leq 0), & P(Y \leq 0), & P(Z = 0), \\ P(X \leq 1), & P(Y \leq 1), & P(Z = 1), \\ P(X \leq 2), & P(Y \leq 2), & P(Z = 2), \\ P(X \leq 3), & P(Y \leq 3), & \\ P(X \leq 4), & P(Y \leq 4). & \end{array}$$

Questão 3
(Total: 4.0 pts)

Aqui nós vamos usar o mesmo diagrama da questão 2. Nesta questão você pode usar tanto os diagramas que só tem bolinhas da terceira coluna do anexo quanto os diagramas que têm bolinhas e blobs que eu desenhei à mão e pus na coluna da direita. O último desenho, com um blob pequenininho dentro de um blob maior, é um jeito de visualizar probabilidades condicionais.





a) Represente graficamente estes eventos:
$X \leq 2$, $X \geq 2$,
$Y \geq 2$, $Y \leq 2$,
$Z \geq 0$, $Z \geq 1$,
e diga a probabilidade de cada um deles.

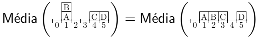
b) Para cada uma das probabilidades condicionais abaixo represente-a graficamente e diga quanto ela vale como um número:
$P(X \leq 2 | Z \geq 0)$,
$P(Z \geq 0 | X \geq 2)$.

Médias

Definição: $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

Digamos que temos duas variáveis, A e D – "antes" e "depois" – que dizem o número de paçocas de cada criança antes e depois do Carlos dar uma paçoca pra Beatriz. As distribuições de A e de D são diferentes, mas como o número total de paçocas não mudou as médias dessas duas distribuições são iguais: $\overline{A} = \overline{D}$. Por exemplo:

i	nome$_i$	A_i	D_i
1	Ana	1	1
2	Beatriz	1	2
3	Carlos	4	3
4	Dani	5	5



Desvio médio e variância

$\mathsf{dm}(X) = \frac{1}{N}\sum_{i=1}^{N} |X_i - \overline{X}|$
$\mathsf{var}(X) = \frac{1}{N}\sum_{i=1}^{N} (X_i - \overline{X})^2$

Lembre que $|42| = 42$, $|-42| = 42$, e que nas aulas a gente calculou o desvio médio e a variância usando "histogramas com numerozinhos", como esses aqui:

	2	1	0	1	2
0	1	2	3	4	5

	4	1	0	1	4
0	1	2	3	4	5

Probabilidade

Uma variável que só pode assumir os valores '**V**' (verdadeiro) ou '**F**' (falso) é uma variável *booleana*.
A operação $[\cdot]$ ("colchete") transforma booleanos nos valores 0 e 1. Por exemplo:

$$[2 < 3] = [\mathbf{V}] = 1,$$
$$[2 > 3] = [\mathbf{F}] = 0.$$

Se B é uma variável booleana e todas as linhas da nossa tabela são "equiprováveis" então a probabilidade de B, $P(B)$ é definida assim:

$$P(B) = \frac{1}{N}\sum_{i=1}^{N}[B_i].$$

Às vezes a gente intepreta expressões como '$A < 42$' como *variáveis com nomes longos* – e aí $(A < 42)_i = (A_i < 42)$. E às vezes a gente coloca definições na primeira linha da tabela. Por exemplo, em

i	A_i	$B_i = (A_i < 42)$	$C_i = [B_i]$
1	200	**F**	0
2	20	**V**	1
3	99	**F**	0

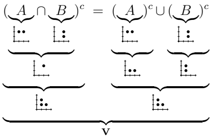
a segunda coluna diz que cada B_i vai ser definido como o resultado do $A_i < 42$ correspondente e lista os valores dos 'B_i's, e a terceira coluna faz a mesma coisa pros 'C_i's. Neste caso temos $\overline{C} = \overline{[B]} = \frac{1}{3}$ e:

$$P(B) = P(A < 42) = \frac{1}{3}.$$

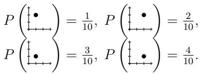
Conjuntos

Se $\Omega = \{1,2,3,4\}$, $A = \{1,2\}$ e $B = \{2,3\}$ então $A \cup B = \{1,2,3\}$, $A \cap B = \{2\}$, e $A^c = \{3,4\}$.

Se $\Omega =$ , $A =$  e $B =$ , então:



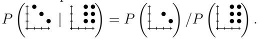
Às vezes a gente diz qual é a probabilidade de cada "evento". Nós usamos este exemplo aqui várias vezes:



Quando a gente não diz a probabilidade de cada evento fica implícito que eles são equiprováveis.

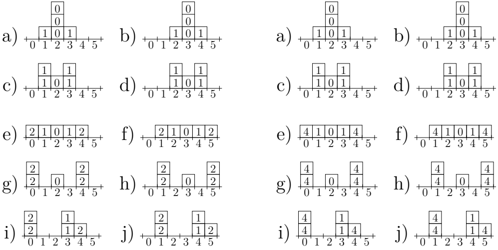
Probabilidade condicional

A definição é: $P(A|B) = P(A \cap B)/P(B)$.
Por exemplo:



Questão 1: gabarito

Dá pra passar de cada distribuição na coluna da esquerda pra distribuição abaixo dela transferindo uma paçoca de uma criança pra outra, e idem na coluna da direita... então todas as distribuições à esquerda têm a mesma média, que é 2, e todas as distribuições à esquerda também têm a mesma média, que é 3. E aí dá pra calcular os desvios médios e as variâncias fazendo histogramas com numerozinhos – veja os diagramas à direita.



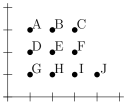
Média	
a) 2	b) 3
c) 2	d) 3
e) 2	f) 3
g) 2	h) 3
i) 2	j) 3

Desvio médio	
a) 2/5	b) 2/5
c) 4/5	d) 4/5
e) 6/5	f) 6/5
g) 8/5	h) 8/5
i) 8/5	j) 8/5

Variância	
a) 2/5	b) 2/5
c) 4/5	d) 4/5
e) 10/5	f) 10/5
g) 16/5	h) 16/5
i) 14/5	j) 14/5

Questão 2: gabarito

Figura original:



O item (a) dá esta tabela:

i	nome$_i$	X_i	Y_i	$Z_i = (X_i - Y_i)$
1	Ana	1	3	-2
2	Bia	2	3	-1
3	Carlos	3	3	0
4	Dani	1	2	-1
5	Eduardo	2	2	0
6	Fábio	3	2	1
7	Geraldo	1	1	0
8	Heloá	2	1	1
9	Inês	3	1	2
10	Joana	4	1	3

O item (b) dá isto aqui:

$$\begin{array}{lll}
P(X \le 0) = 0, & P(Y \le 0) = 0, & P(Z = 0) = 3/10, \\
P(X \le 1) = 3/10, & P(Y \le 1) = 4/10, & P(Z = 1) = 2/10, \\
P(X \le 2) = 6/10, & P(Y \le 2) = 7/10, & P(Z = 2) = 1/10, \\
P(X \le 3) = 9/10, & P(Y \le 3) = 10/10, & \\
P(X \le 4) = 10/10, & P(Y \le 4) = 10/10. &
\end{array}$$

Questão 3: gabarito

Item a:

$$\begin{array}{rcl@{\qquad}rcl}
X \leq 2 & \Rightarrow & & X \geq 2 & \Rightarrow & \\
P(X \leq 2) & = & 6/10 & P(X \geq 2) & = & 7/10 \\[1ex]
Y \geq 2 & \Rightarrow & & Y \leq 2 & \Rightarrow & \\
P(Y \geq 2) & = & 6/10 & P(Y \leq 2) & = & 7/10 \\[1ex]
Z \geq 0 & \Rightarrow & & Z \geq 1 & \Rightarrow & \\
P(Z \geq 0) & = & 7/10 & P(Z \geq 1) & = & 4/10
\end{array}$$

Item b:

$$P(\underbrace{X \leq 2} \,|\, \underbrace{Z \geq 0}) \;=\; \underbrace{\underbrace{P((\underbrace{X \leq 2}) \cap (\underbrace{Z \geq 0}))}}_{3/10} \,/\, \underbrace{P(\underbrace{Z \geq 0})}_{7/10} \;=\; 3/7$$

$$P(\underbrace{Z \geq 0} \,|\, \underbrace{X \geq 2}) \;=\; \underbrace{\underbrace{P((\underbrace{Z \geq 0}) \cap (\underbrace{X \geq 2}))}}_{6/10} \,/\, \underbrace{P(\underbrace{X \geq 2})}_{7/10} \;=\; 6/7$$

Avisos (do dia anterior à prova)

Eu ainda não fiz as questões da prova!

A próxima folha tem o anexo que eu prometi - uma folha com um monte de definições e exemplos que podem ajudar vocês a fazerem as questões da prova.

A parte em que as pessoas tiveram mais dificuldade foi a das "variáveis com nomes longos" da coluna do meio do anexo. A gente discutiu isso no dia 10/maio, nas páginas 16 e 17 do PDF com as fotos dos quadros. O link é este: 5gQ16. *Revisem isso!!!* =)

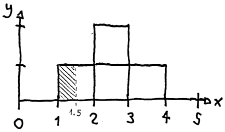
Estatística - 2023.1

Segunda prova (P2)

Eduardo Ochs - RCN/PURO/UFF
http://anggtwu.net/2023.1-ES.html

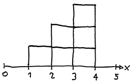
Questão 1

(Total: 5.0 pts)
Lembre que em distribuições contínuas as probabilidades são representadas por áreas. Por exemplo, se isto é a distribuição da variável A,



então $P(A \leq 1.5) = 1/8$.

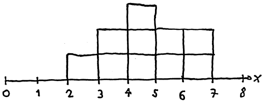
a) **(2.0 pts)** Seja B a distribuição contínua abaixo à esquerda. Complete a tabela abaixo à direita. Importante: nesta prova você não precisa simplificar ou calcular frações – se você chegar num resultado como $3 + \frac{2}{14}$ você pode responder $3 + \frac{2}{14}$ mesmo, não precisa transformá-lo em 3.142857.



x	$P(B \leq x)$
0.0	
0.5	
1.0	
1.5	
2.0	
2.5	
3.0	
3.5	
4.0	

No curso nós aprendemos o que são quantis usando principalmente exemplos, porque a definição de quantil que aparece no livro é meio complicada... por exemplo, para a variável A da figura à esquerda temos $P(A \leq 1.5) = 1/8 = 12.5\%$, e portanto $q(1/8) = q(12.5\%) = 1.5$. Note que os quantis dependem da distribuição – o $q(1/8)$ para a variável B pode ter um valor totalmente diferente do $q(1/8)$ para a variável A.

b) **(3.0 pts)** Seja C esta distribuição contínua:



Complete a tabela abaixo com os quantis para a variável C.

$q(10\%) =$
$q(20\%) =$
$q(30\%) =$
$q(40\%) =$
$q(50\%) =$
$q(60\%) =$
$q(70\%) =$
$q(80\%) =$
$q(90\%) =$

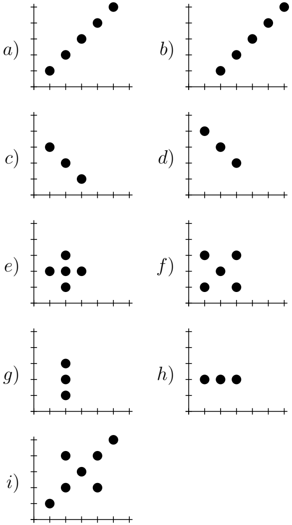
Questão 2

(Total: 5.0 pts)
No curso a gente viu vários jeitos de visualizar "váriaveis com nomes longos"; por exemplo, no último desenho da primeira coluna da folha de dicas da P1 – a página 5 do PDF da prova – eu desenhei o valor de cada $(X-\overline{X})^2$ dentro do quadradinho correspondente.

Nesta questão você vai ter que calcular $\mathsf{Cov}(X,Y)$ para várias distribuições em duas variáveis. A definição é:

$$\begin{array}{rcl} \mathsf{Cov}(X,Y) & = & \overline{(X-\overline{X})(Y-\overline{Y})} \\ & = & \frac{1}{N}\sum_{i=1}^{N}\big((X_i-\overline{X})(Y_i-\overline{Y})\big) \end{array}$$

Para cada uma das distribuições em duas variáveis à direita calcule $\mathsf{Cov}(X,Y)$ usando as dicas que eu vou pôr no quadro (e que depois viraram a página 4 do PDF da prova).



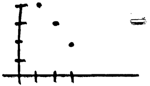
ESTATÍSTICA 5/JULHO/2023

• HOJE: P2!!!

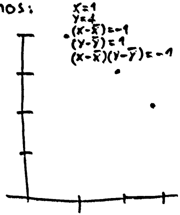
EU NÃO CONSEGUI DIGITAR TODAS AS FIGURAS DA PROVA A TEMPO... AÍ ACABEI FAZENDO UMAS MANUSCRITAS, E VOU TER QUE PÔR NO QUADRO ALGUMAS DAS DICAS QUE EU QUERIA TER POSTO NA PROVA!...

QUANDO A GENTE TEM UM HISTOGRAMA EM DUAS VARIÁVEIS A GENTE PODE VISUALIZAR O QUE CADA VARIÁVEL QUER DIZER - INCLUINDO AS VARIÁVEIS COM NOMES COMPRIDOS - ESCREVENDO O VALOR DELA DO LADO DE CADA PONTINHO/PESSOA.

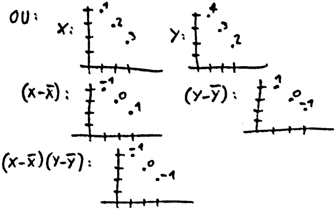
POR EXEMPLO, NESTE HISTOGRAMA



TEMOS:



OU:

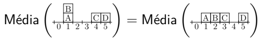


Médias

Definição: $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

Digamos que temos duas variáveis, A e D – “antes” e “depois” – que dizem o número de paçocas de cada criança antes e depois do Carlos dar uma paçoca pra Beatriz. As distribuições de A e de D são diferentes, mas como o número total de paçocas não mudou as médias dessas duas distribuições são iguais: $\overline{A} = \overline{D}$. Por exemplo:

i	nome_i	A_i	D_i
1	Ana	1	1
2	Beatriz	1	2
3	Carlos	4	3
4	Dani	5	5



Desvio médio e variância

$\mathsf{dm}(X) = \frac{1}{N}\sum_{i=1}^{N} |X_i - \overline{X}|$
$\mathsf{var}(X) = \frac{1}{N}\sum_{i=1}^{N} (X_i - \overline{X})^2$

Lembre que $|42| = 42$, $|-42| = 42$, e que nas aulas a gente calculou o desvio médio e a variância usando “histogramas com numerozinhos”, como esses aqui:

Probabilidade

Uma variável que só pode assumir os valores ‘**V**’ (verdadeiro) ou ‘**F**’ (falso) é uma variável *booleana*.
A operação $[\cdot]$ (“colchete”) transforma booleanos nos valores 0 e 1. Por exemplo:

$$[2<3] = [\mathbf{V}] = 1,$$
$$[2>3] = [\mathbf{F}] = 0.$$

Se B é uma variável booleana e todas as linhas da nossa tabela são “equiprováveis” então a probabilidade de B, $P(B)$ é definida assim:

$$P(B) = \frac{1}{N}\sum_{i=1}^{N} [B_i].$$

Às vezes a gente intepreta expressões como ‘$A < 42$’ como *variáveis com nomes longos* – e aí $(A < 42)_i = (A_i < 42)$. E às vezes a gente coloca definições na primeira linha da tabela. Por exemplo, em

i	A_i	$B_i = (A_i < 42)$	$C_i = [B_i]$
1	200	**F**	0
2	20	**V**	1
3	99	**F**	0

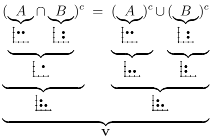
a segunda coluna diz que cada B_i vai ser definido como o resultado do $A_i < 42$ correspondente e lista os valores dos ‘B_i’s, e a terceira coluna faz a mesma coisa pros ‘C_i’s. Neste caso temos $\overline{C} = \overline{[B]} = \frac{1}{3}$ e:

$$P(B) = P(A < 42) = \frac{1}{3}.$$

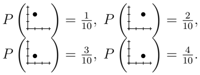
Conjuntos

Se $\Omega = \{1,2,3,4\}$, $A = \{1,2\}$ e $B = \{2,3\}$ então $A \cup B = \{1,2,3\}$, $A \cap B = \{2\}$, e $A^c = \{3,4\}$.

Se $\Omega =$ , $A =$  e $B =$ , então:



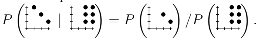
Às vezes a gente diz qual é a probabilidade de cada “evento”. Nós usamos este exemplo aqui várias vezes:



Quando a gente não diz a probabilidade de cada evento fica implícito que eles são equiprováveis.

Probabilidade condicional

A definição é: $P(A|B) = P(A \cap B)/P(B)$.
Por exemplo:



Estatística - 2023.1

Prova de reposição (VR)

Eduardo Ochs - RCN/PURO/UFF
http://anggtwu.net/2023.1-ES.html

Mais dicas

Todas as questões desta prova vão testar se você sabe visualizar e representar graficamente o que certos conceitos e certas expressões matemáticas querem dizer. Dá pra fazer as questões sem saber como visualizar quase nada, mas aí você vai ter que fazer um monte de contas e tabelas e vai levar horas; se você souber como visualizar e desenhar tudo você vai conseguir fazer cada item da prova em poucos segundos.

Todos os truques de visualização estão explicados – com exemplos! – na “folha com muitas dicas” que é exatamente igual à que eu pus na P1 e na P2 e no resto desta folha daqui.

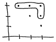
Se você ainda não souber algum truque de visualização tente entender os exemplos!!!

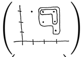
Sobre frações

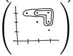
Nesta prova você não precisa simplificar ou calcular frações – se você chegar num resultado como $3+\frac{2}{14}$ você pode responder $3+\frac{2}{14}$ mesmo, não precisa transformá-lo em $3+\frac{1}{7}$, em $\frac{22}{7}$ ou em 3.142857.

Probabilidade condicional

Esse aqui é o “método dos blobs” pra visualizar probabilidades condicionais:

Se $A =$  e $B =$ 

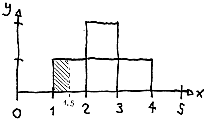
então $P(A|B) = 3/5$ 

e $P(B|A) = 3/4$  .

Ele é equivalente ao método do final da folha com muitas dicas mas ele é bem mais prático.

Probabilidades acumuladas e quantis

Lembre que em distribuições contínuas as probabilidades são representadas por áreas. Por exemplo, se isto é a distribuição da variável A,



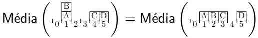
então:
$P(A \leq 1.5) = 1/8 = 12.5\%$ e
$q(1/8) = q(12.5\%) = 1.5$.

Médias

Definição: $\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$.

Digamos que temos duas variáveis, A e D – "antes" e "depois" – que dizem o número de paçocas de cada criança antes e depois do Carlos dar uma paçoca pra Beatriz. As distribuições de A e de D são diferentes, mas como o número total de paçocas não mudou as médias dessas duas distribuições são iguais: $\overline{A} = \overline{D}$. Por exemplo:

i	nome$_i$	A_i	D_i
1	Ana	1	1
2	Beatriz	1	2
3	Carlos	4	3
4	Dani	5	5



Desvio médio e variância

$\mathsf{dm}(X) = \frac{1}{N} \sum_{i=1}^{N} |X_i - \overline{X}|$

$\mathsf{var}(X) = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})^2$

Lembre que $|42| = 42$, $|-42| = 42$, e que nas aulas a gente calculou o desvio médio e a variância usando "histogramas com numerozinhos", como esses aqui:

Probabilidade

Uma variável que só pode assumir os valores '**V**' (verdadeiro) ou '**F**' (falso) é uma variável *booleana*.

A operação $[\cdot]$ ("colchete") transforma booleanos nos valores 0 e 1. Por exemplo:

$$[2 < 3] = [\mathbf{V}] = 1,$$
$$[2 > 3] = [\mathbf{F}] = 0.$$

Se B é uma variável booleana e todas as linhas da nossa tabela são "equiprováveis" então a probabilidade de B, $P(B)$ é definida assim:

$$P(B) = \frac{1}{N} \sum_{i=1}^{N} [B_i].$$

Às vezes a gente intepreta expressões como '$A < 42$' como *variáveis com nomes longos* – e aí $(A < 42)_i = (A_i < 42)$. E às vezes a gente coloca definições na primeira linha da tabela. Por exemplo, em

i	A_i	$B_i = (A_i < 42)$	$C_i = [B_i]$
1	200	**F**	0
2	20	**V**	1
3	99	**F**	0

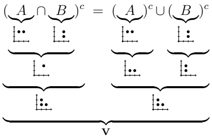
a segunda coluna diz que cada B_i vai ser definido como o resultado do $A_i < 42$ correspondente e lista os valores dos 'B_i's, e a terceira coluna faz a mesma coisa pros 'C_i's. Neste caso temos $\overline{C} = \overline{[B]} = \frac{1}{3}$ e:

$$P(B) = P(A < 42) = \frac{1}{3}.$$

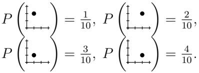
Conjuntos

Se $\Omega = \{1, 2, 3, 4\}$, $A = \{1, 2\}$ e $B = \{2, 3\}$ então $A \cup B = \{1, 2, 3\}$, $A \cap B = \{2\}$, e $A^c = \{3, 4\}$.

Se $\Omega =$ , $A =$  e $B =$ , então:



Às vezes a gente diz qual é a probabilidade de cada "evento". Nós usamos este exemplo aqui várias vezes:

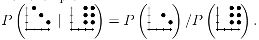


Quando a gente não diz a probabilidade de cada evento fica implícito que eles são equiprováveis.

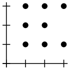
Probabilidade condicional

A definição é: $P(A|B) = P(A \cap B)/P(B)$.

Por exemplo:



Questão 1.
(Total: 3.5 pts)
Seja A esta distribuição em duas variáveis,



e sejam $Z = X + Y$ e $W = X - Y$.
Para cada um dos itens abaixo mostre como representá-lo graficamente e diga o seu resultado como um número. Os itens mais fáceis valem bem poucos pontos mas eles vão ajudar você a fazer os itens mais difíceis, que valem muito.

a) $P(X = 1)$
b) $P(X = 2)$
c) $P(X = 3)$
d) $P(X \leq 1)$
e) $P(X \leq 2)$
f) $P(X \leq 3)$

g) $P(Z = 0)$
h) $P(Z = 1)$
i) $P(Z = 2)$
j) $P(Z \leq 0)$
k) $P(Z \leq 1)$
l) $P(Z \leq 2)$

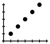
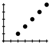
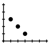
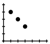
k) $P(W = 0)$
n) $P(W = 1)$
o) $P(W = 2)$

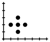
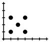
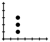
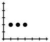
p) $P(X \leq 2 | Y \leq 2)$
q) $P(Y \leq 2 | X \leq 2)$
r) $P(Z \geq 1 | Y \geq 1)$

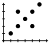
Questão 2.
(Total: 3.0 pts)
A covariância de duas variáveis, X e Y, é definida desta forma, como a média de uma "variável com nome longo":

$$\begin{aligned}\mathsf{Cov}(X,Y) &= \overline{(X-\overline{X})(Y-\overline{Y})} \\ &= \frac{1}{N}\sum_{i=1}^{N}\big((X_i-\overline{X})(Y_i-\overline{Y})\big)\end{aligned}$$

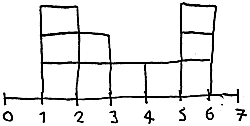
Calcule $\mathsf{Cov}(X,Y)$ para cada uma das distribuições em duas variáveis abaixo. Dica: você pode fazer isso em duas etapas – primeiro você representa graficamente $(X-\overline{X})(Y-\overline{Y})$ como um número do lado de cada pontinho e depois você calcula a média disso.

a)  *b)*  *c)*  *d)* 

e)  *f)*  *g)*  *h)* 

i) 

Questão 3.
(Total: 3.5 pts)
Seja A esta distribuição contínua:



Complete as duas tabelas abaixo:

x	$P(A \leq x)$
0.0	
0.5	
1.0	
1.5	
2.0	
2.5	
3.0	
3.5	
4.0	
4.5	
5.0	

$q(10\%) =$
$q(20\%) =$
$q(30\%) =$
$q(40\%) =$
$q(50\%) =$
$q(60\%) =$
$q(70\%) =$
$q(80\%) =$
$q(90\%) =$