

Estatística - 2023.1

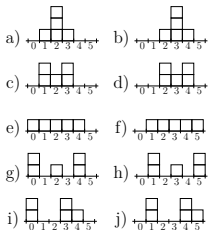
Primeira prova (P1)

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.

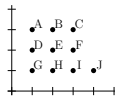


Não é preciso simplificar as frações. Arrume os seus resultados em três tabelas com essa cara aqui:

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.



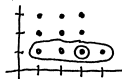
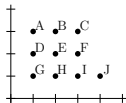
a) Transforme as informações do diagrama acima numa tabela. A sua tabela deve ter pelo menos estas colunas aqui: i , nome $_i$, X_i , Y_i , e $Z_i = (X_i - Y_i)$ – mas você pode acrescentar outras se você achar que isto vai te ajudar a responder os outros itens.

b) Calcule estas probabilidades:

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

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

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: $\bar{A} = \bar{D}$. 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

$$\text{Média} \left(\begin{array}{|c|c|c|c|c|} \hline \square & \square & \square & \square & \square \\ \hline 0 & 1 & 2 & 3 & 4 & 5 \\ \hline \end{array} \right) = \text{Média} \left(\begin{array}{|c|c|c|c|c|} \hline \square & \square & \square & \square & \square \\ \hline 0 & 1 & 2 & 3 & 4 & 5 \\ \hline \end{array} \right)$$

Desvio médio e variância

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

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

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:

$$\begin{array}{|c|c|c|c|c|c|} \hline 2 & 1 & 0 & 1 & 1 & 2 \\ \hline 0 & 1 & 2 & 3 & 4 & 5 \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|c|c|} \hline 4 & 1 & 0 & 1 & 1 & 4 \\ \hline 0 & 1 & 2 & 3 & 4 & 5 \\ \hline \end{array}$$

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 interpreta 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 $\bar{C} = [\bar{B}] = \frac{1}{3}$ 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 $\Omega = \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}$, $A = \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}$ e $B = \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}$, então:

$$\underbrace{\left(\underbrace{\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}}_A \cap \underbrace{\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}}_B \right)^c}_{\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}} = \left(\underbrace{\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}}_A \right)^c \cup \left(\underbrace{\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array}}_B \right)^c$$

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

$$P \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right) = \frac{1}{10}, \quad P \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right) = \frac{2}{10},$$

$$P \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right) = \frac{3}{10}, \quad P \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right) = \frac{4}{10}.$$

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

Probabilidade condicional

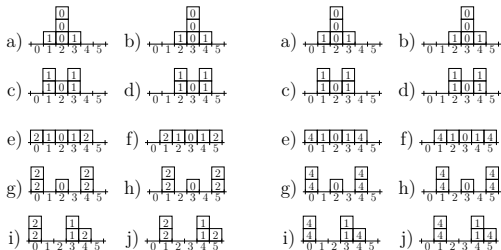
A definição é: $P(A|B) = P(A \cap B)/P(B)$.

Por exemplo:

$$P \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \mid \begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right) = P \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right) / P \left(\begin{array}{|c|} \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \bullet \\ \hline \end{array} \right).$$

Questão 1: gabarito

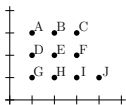
Dá pra passar de cada distribuição na coluna da esquerda pra distribuição abaixo dela transferindo uma paçoca de uma criança pra outra, e idem na coluna da direita... então todas as distribuições à esquerda têm a mesma média, que é 2, e todas as distribuições à esquerda também têm a mesma média, que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



Média		Desvio médio		Variância	
a) 2	b) 3	a) 2/5	b) 2/5	a) 2/5	b) 2/5
c) 2	d) 3	c) 4/5	d) 4/5	c) 4/5	d) 4/5
e) 2	f) 3	e) 6/5	f) 6/5	e) 10/5	f) 10/5
g) 2	h) 3	g) 8/5	h) 8/5	g) 16/5	h) 16/5
i) 2	j) 3	i) 8/5	j) 8/5	i) 14/5	j) 14/5

Questão 2: gabarito

Figura original:



O item (b) dá isto aqui:

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

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

Questão 3: gabarito

Item a:

$$X \leq 2 \Rightarrow \begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array} \quad P(X \leq 2) = 6/10 \quad X \geq 2 \Rightarrow \begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array} \quad P(X \geq 2) = 7/10$$

$$Y \geq 2 \Rightarrow \begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array} \quad P(Y \geq 2) = 6/10 \quad Y \leq 2 \Rightarrow \begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array} \quad P(Y \leq 2) = 7/10$$

$$Z \geq 0 \Rightarrow \begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array} \quad P(Z \geq 0) = 7/10 \quad Z \geq 1 \Rightarrow \begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array} \quad P(Z \geq 1) = 4/10$$

Item b:

$$P(\underbrace{X \leq 2}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}} | \underbrace{Z \geq 0}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}}) = P(\underbrace{(X \leq 2) \cap (Z \geq 0)}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}}) / P(\underbrace{Z \geq 0}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}}) = 3/7$$

$$P(\underbrace{Z \geq 0}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}} | \underbrace{X \geq 2}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}}) = P(\underbrace{(Z \geq 0) \cap (X \geq 2)}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}}) / P(\underbrace{X \geq 2}_{\begin{array}{|c|} \hline \cdot \\ \cdot \\ \cdot \\ \hline \end{array}}) = 6/7$$

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!!!* =)