Probabilidade e Estatística

Professores:

Otton Teixeira da Silveira Filho

Regina Célia Paula Leal Toledo

Fundação CECIERJ
Consórcio cederj

Aula 7

Professores:

Otton Teixeira da Silveira Filho
Regina Célia Paula Leal Toledo

Variáveis bidimensionais

Conteúdo:

7 Introdução
- 7.1 Função de probabilidade conjunta
- 7.2 Distribuição marginal de probabilidade
- 7.3 Probabilidade condicional de variáveis aleatórias discretas
- 7.4. Independência de variáveis aleatórias
- 7.5. Covariância
- 7.6 Correlação entre variáveis aleatórias discretas

7. Introdução

É comum termos interesse no comportamento conjunto de várias variáveis.

↓

Variáveis bidimensionais

7. Introdução

É comum termos interesse no comportamento conjunto de várias variáveis.

Variáveis bidimensionais

Exemplo:

Alunos de Rio Azul - alunos que fumam e horas de exercícios semanais
(Bibliografia [1], www.ime.usp.br/~noproest)

he / fuma	0	1	2	3	4	5	6	7	8	9	10	Total
sim	1	0	1	0	0	2	1	0	0	0	1	6
não	7	3	7	6	4	6	2	5	3	0	1	44
Total	8	3	8	6	4	8	3	5	3	0	2	50

Apresentamos a seguir uma tabela de dupla entrada que mostra a probabilidade conjunta das duas variáveis: *fuma* (sim ou não) e *h.e.* (horas semanais de exercício - de 0 a 10).

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

onde:

$$0,02 = \frac{1}{50}$$

$$0,14 = \frac{7}{50}$$

Apresentamos a seguir uma tabela de dupla entrada que mostra a probabilidade conjunta das duas variáveis: *fuma* (sim ou não) e *h.e.* (horas semanais de exercício - de 0 a 10).

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

→ Assim, a probabilidade do aluno fumar e fazer 5 horas semanais de exercício é *P(fuma=sim ; h.e.=5)=0,04* e,

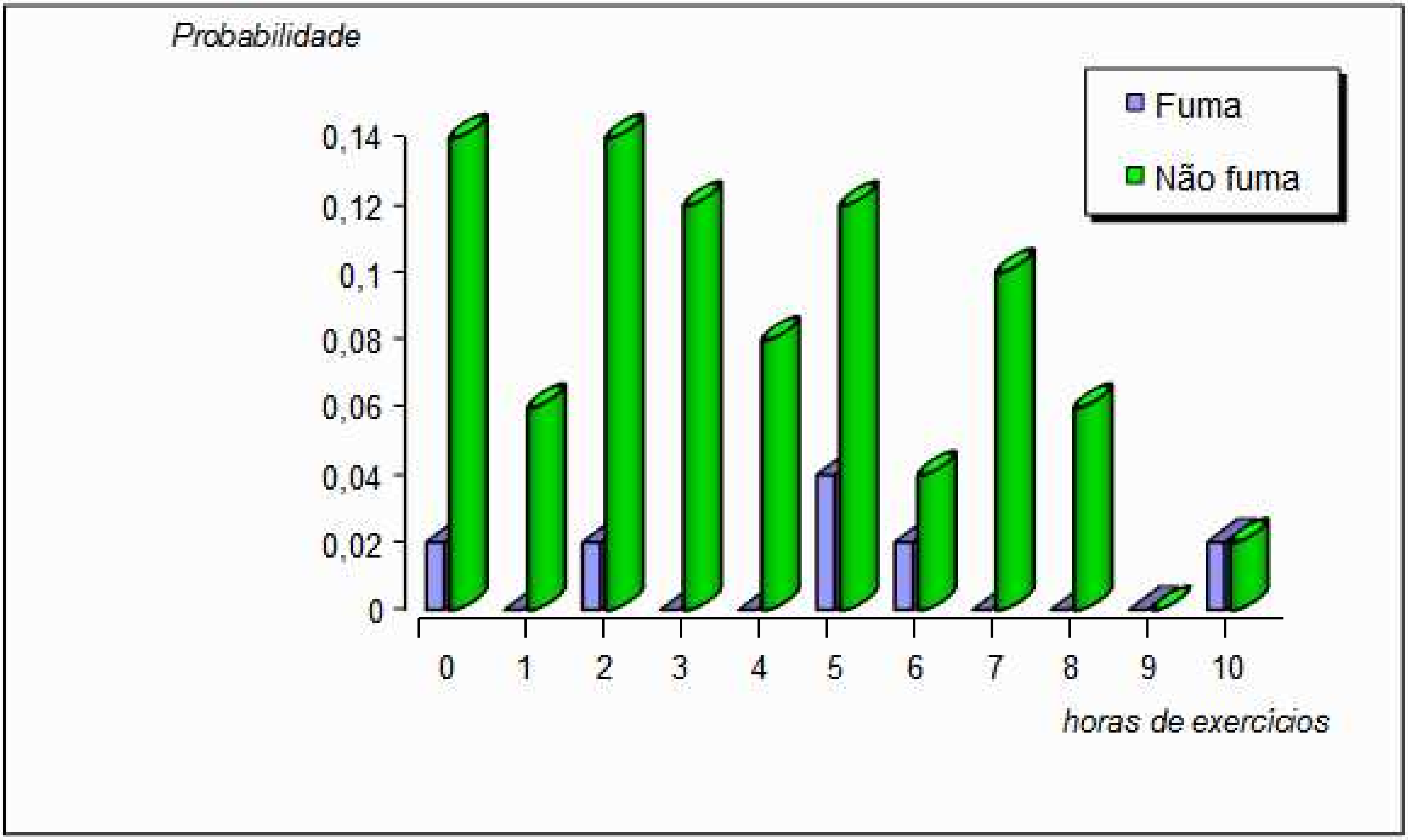
Apresentamos a seguir uma tabela de dupla entrada que mostra a probabilidade conjunta das duas variáveis: *fuma* (sim ou não) e *h.e.* (horas semanais de exercício - de 0 a 10).

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

→ Assim, a probabilidade do aluno fumar e fazer 5 horas semanais de exercício é *P(fuma=sim ; h.e.=5)=0,04* e,

→ a probabilidade do aluno que não fuma não fazer exercício é *P(fuma=não ; h.e.=0)=0,14*.

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1



7.1 Função de probabilidade conjunta

Seja X uma variável aleatória que assume os valores $x_1, x_2, \ldots, x_n$ e Y uma variável aleatória, do mesmo experimento, que assume os valores $y_1, y_2, \ldots, y_m$. A função de probabilidade conjunta associa a cada par (x_i, y_k), $i = 1, 2, \ldots, n$ e $k = 1, 2, \ldots, m$ a probabilidade $P(X=x_i, Y=y_k) = p(x_i, y_k)$.

Assim, temos que:

$$\boxed{\sum_{i=1}^{n}\sum_{k=1}^{m} P(X = x_i, Y = y_k) = 1}$$

$$\boxed{\sum_{i=1}^{n}\sum_{k=1}^{m} P(x_i, y_k) = 1}$$

Exemplo:

Alunos de Rio Azul idade e horas de exercícios semanais (Bibliografia [1], www.ime.usp.br/~noproest)

id \ he	0	1	2	3	4	5	6	7	8	9	10	Total
17	1	1	2	1	1	1	0	1	0	0	1	9
18	4	1	3	3	2	5	0	3	0	0	1	22
19	2	0	2	0	1	1	1	0	0	0	0	7
20	1	0	0	2	0	0	0	0	1	0	0	4
21	0	0	0	0	0	0	2	0	1	0	0	3
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	1	0	0	0	0	0	1	0	0	2
24	0	0	0	0	0	0	0	1	0	0	0	1
25	0	1	0	0	0	1	0	0	0	0	0	2
Total	8	3	8	6	4	8	3	5	3	0	2	50

Probabilidade (freqüência)

id \ he	0	1	2	3	4	5	6	7	8	9	10	Total
17	0,02	0,02	0,04	0,02	0,02	0,02	0	0,02	0	0	0,02	0,18
18	0,08	0,02	0,06	0,06	0,04	0,1	0	0,06	0	0	0,02	0,44
19	0,04	0	0,04	0	0,02	0,02	0,02	0	0	0	0	0,14
20	0,02	0	0	0,04	0	0	0	0	0,02	0	0	0,08
21	0	0	0	0	0	0	0,04	0	0,02	0	0	0,06
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0,02	0	0	0	0	0	0,02	0	0	0,04
24	0	0	0	0	0	0	0	0,02	0	0	0	0,02
25	0	0,02	0	0	0	0,02	0	0	0	0	0	0,04
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

7.2 Distribuição marginal de probabilidade

Seja X uma variável aleatória discreta que assume os valores x_1, x_2, ..., x_n. A distribuição marginal de probabilidade

– de $X=x_i$ é definida como:

$$P(X = x_i) = \sum_{k=1}^{m} P(X = x_i, Y = y_k), i = 1, 2, \ldots, n$$

7.2 Distribuição marginal de probabilidade

Seja X uma variável aleatória discreta que assume os valores x_1, x_2, ..., x_n. A distribuição marginal de probabilidade

– de $X=x_i$ é definida como:

$$P(X = x_i) = \sum_{k=1}^{m} P(X = x_i, Y = y_k), i = 1, 2, \ldots, n$$

– e de $Y=y_k$ é dada por:

$$P(Y = y_k) = \sum_{i=1}^{n} P(X = x_i, Y = y_k), k = 1, 2, \ldots, m$$

Assim, para a tabela de nosso exemplo, para as variáveis "fuma" e "h.e.", temos:

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

Assim, para a tabela de nosso exemplo, para as variáveis "fuma" e "h.e.", temos:

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

– para a variável fuma:

Fuma(X)	P(X)
sim(1)	0,12
não(2)	0,88
Total	1

Assim, para a tabela de nosso exemplo, para as variáveis "fuma" e "h.e.", temos:

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

– para a variável fuma:

Fuma(X)	P(X)
sim(1)	0,12
não(2)	0,88
Total	1

– para a variável horas semanais de exercícios (h.e.):

h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	Total
P(Y)	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

e, para a tabela "idade" x "horas de exercício", temos:

– para a variável idade:

idade(X)	P(X)
17	0,18
18	0,44
19	0,14
20	0,08
21	0,06
22	0
23	0,04
24	0,02
25	0,04
$\sum$	1

– para a variável horas semanais de exercícios (h.e.):

h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	Total
P(Y)	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

7.3 Probabilidade condicional de variáveis aleatórias discretas

Dados duas variáveis aleatórias discretas definidas no mesmo espaço amostral, a probabilidade condicional de *X*=*x* dado que *Y*=*y* ocorreu, é dado por:
Assim, temos que:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} \text{ com } P(Y = y) > 0$$

$$\text{caso } P(Y = y) = 0 \longrightarrow P(X = x|Y = y) = P(X = x)$$

7.4 Independência de variáveis aleatórias

Duas variáveis aleatórias discretas são independentes se a ocorrência de qualquer valor de uma delas não altera a probabilidade de ocorrência de valores da outra variável, para todos os valores *(x,y)* das variáveis *(X,Y)*, ou:

$$P(X = x, Y = y) = P(X = x) \times P(Y = y), \forall (x, y)$$

ou,

$$P(X, Y) = P(X) \times P(Y)$$

7.4 Independência de variáveis aleatórias

Duas variáveis aleatórias discretas são independentes se a ocorrência de qualquer valor de uma delas não altera a probabilidade de ocorrência de valores da outra variável, para todos os valores *(x,y)* das variáveis *(X,Y)*, ou:

$$P(X = x, Y = y) = P(X = x) \times P(Y = y), \forall (x, y)$$

ou,

$$P(X, Y) = P(X) \times P(Y)$$

Exemplos:

Verificar se a variável "fuma" é independente da variável horas de exercícios semanais.

he / fuma	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

$$\boxed{P(X=x, Y=y) = P(X=x) \times P(Y=y)}$$

he / fuma	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

fuma e não faz exercício

P(fuma=sim, h.e.=0) = 0,02

P(fuma=sim) x P(h.e.=0) = 0,12 x 0,16 = 0,0192

he / fuma	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

fuma e não faz exercício

P(fuma=sim, h.e.=0) = 0,02

P(fuma=sim) x P(h.e.=0) = 0,12 x 0,16 = 0,0192

não fuma e não faz exercício

P(fuma=não, h.e.=0) = 0,14

P(fuma=não) x P(h.e.=0) = 0,88 x 0,16 = 0,1408

fuma \ he	0	1	2	3	4	5	6	7	8	9	10	Total
sim	0,02	0	0,02	0	0	0,04	0,02	0	0	0	0,02	0,12
não	0,14	0,06	0,14	0,12	0,08	0,12	0,04	0,1	0,06	0	0,02	0,88
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

fuma e não
faz exercício

P(fuma=sim, h.e.=0) = 0,02

P(fuma=sim) x P(h.e.=0) = 0,12 x 0,16 = 0,0192

não fuma e não
faz exercício

P(fuma=não, h.e.=0) = 0,14

P(fuma=não) x P(h.e.=0) = 0,88 x 0,16 = 0,1408

fuma e faz 6h/sem
de exercício

P(fuma=sim, h.e.=6) = 0,02

P(fuma=sim) x P(h.e.=6) = 0,12 x 0,06 = 0,0072

Outro exemplo

id \ he	0	1	2	3	4	5	6	7	8	9	10	Total
17	0,02	0,02	0,04	0,02	0,02	0,02	0	0,02	0	0	0,02	0,18
18	0,08	0,02	0,06	0,06	0,04	0,1	0	0,06	0	0	0,02	0,44
19	0,04	0	0,04	0	0,02	0,02	0,02	0	0	0	0	0,14
20	0,02	0	0	0,04	0	0	0	0	0,02	0	0	0,08
21	0	0	0	0	0	0	0,04	0	0,02	0	0	0,06
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0,02	0	0	0	0	0	0,02	0	0	0,04
24	0	0	0	0	0	0	0	0,02	0	0	0	0,02
25	0	0,02	0	0	0	0,02	0	0	0	0	0	0,04
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

18 anos e 8 h/sem
de exercício

P(Id.=18, h.e.=9) = 0,0

P(Id.=18) x P(h.e.=9) = 0,44 x 0,0 = 0,0

Outro exemplo

id \ he	0	1	2	3	4	5	6	7	8	9	10	Total
17	0,02	0,02	0,04	0,02	0,02	0,02	0	0,02	0	0	0,02	0,18
18	0,08	0,02	0,06	0,06	0,04	0,1	0	0,06	0	0	0,02	0,44
19	0,04	0	0,04	0	0,02	0,02	0,02	0	0	0	0	0,14
20	0,02	0	0	0,04	0	0	0	0	0,02	0	0	0,08
21	0	0	0	0	0	0	0,04	0	0,02	0	0	0,06
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0,02	0	0	0	0	0	0,02	0	0	0,04
24	0	0	0	0	0	0	0	0,02	0	0	0	0,02
25	0	0,02	0	0	0	0,02	0	0	0	0	0	0,04
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

18 anos e 8 h/sem de exercício

P(Id.=18, h.e.=9) = 0,0

P(Id.=18) x P(h.e.=9) = 0,44 x 0,0 = 0,0

18 anos e 5 h/sem de exercício

P(Id.=18, h.e.=8) = 0,0

P(Id.=18) x P(h.e.=8) = 0,44 x 0,06 = 0,03

7.5 Covariância

Para entender: sendo (X,Y) uma variável bidimensional

– as esperanças das componentes X e Y fornecem uma medida de posição das distribuições de X e Y em seus respectivos eixos.

– as variâncias de X e Y dão uma medida da dispersão de X e Y em torno de suas médias $E(X)$ e $E(Y)$, respectivamente.

Assim, podemos calcular a média ou valor esperado ou esperança da variável *Y* (horas semanais de exercícios) e para a variável *X* (idade):

h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	$\sum$
P(Y)	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

p_1 =0,16 (número de pessoas que não fazem exercícios)/50

$$E(Y) = \sum_{i=1}^{m} y_i p_i$$

Assim, podemos calcular a média ou valor esperado ou esperança da variável *Y* (horas semanais de exercícios) e para a variável *X* (idade):

h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	$\sum$
P(Y)	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

p_1 =0,16 (número de pessoas que não fazem exercícios)/50

p_2 =0,06 (número de pessoas que fazem 1h exercício)/50

$$E(Y) = \sum_{i=1}^{m} y_i p_i$$

Assim, podemos calcular a média ou valor esperado ou esperança da variável *Y* (horas semanais de exercícios) e para a variável *X* (idade):

h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	$\sum$
P(Y)	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1
$Y_i \cdot P_i(Y)$	0	0,06	0,32	0,36	0,32	0,8	0,36	0,7	0,48	0	0,4	3,8

$$\boxed{E(Y) = \sum_{i=1}^{m} y_i p_i}$$

Assim, podemos calcular a média ou valor esperado ou esperança da variável *Y* (horas semanais de exercícios) e para a variável *X* (idade):

h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	$\sum$
P(Y)	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1
$Y_i \cdot P_i\,(Y)$	0	0,06	0,32	0,36	0,32	0,8	0,36	0,7	0,48	0	0,4	3,8

idade(X)	P(X)	$X_i \cdot P_i\,(X)$
17	0,18	3,06
18	0,44	7,92
19	0,14	2,66
20	0,08	1,6
21	0,06	1,26
22	0	0
23	0,04	0,92
24	0,02	0,48
25	0,04	1
$\sum$	1	18,9

$$E(Y) = \sum_{i=1}^{m} y_i p_i$$

$$\boxed{E(X) = \sum_{i=1}^{n} x_i p_i}$$

Assim, podemos calcular a média ou valor esperado ou esperança da variável *Y* (horas semanais de exercícios) e para a variável *X* (idade):

h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	$\sum$
P(Y)	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1
$Y_i \cdot P_i(Y)$	0	0,06	0,32	0,36	0,32	0,8	0,36	0,7	0,48	0	0,4	3,8

idade(X)	P(X)	$X_i \cdot P_i(X)$
17	0,18	3,06
18	0,44	7,92
19	0,14	2,66
20	0,08	1,6
21	0,06	1,26
22	0	0
23	0,04	0,92
24	0,02	0,48
25	0,04	1
$\sum$	1	18,9

$$\boxed{E(Y) = \sum_{i=1}^{m} y_i p_i}$$

$$\boxed{E(X) = \sum_{i=1}^{n} x_i p_i}$$

momentos de primeira ordem

$$Var(X) = \sigma_X^2 = \sum_{i=1}^{k} (x_i - E(X))^2 p_i$$

idade(Y)	P(Y)	Y . P(Y)	$(Y-18,9)^2$. p(Y)
17	0,18	3,06	0,6498
18	0,44	7,92	0,3564
19	0,14	2,66	0,0014
20	0,08	1,6	0,0968
21	0,06	1,26	0,2646
22	0	0	0
23	0,04	0,92	0,6724
24	0,02	0,48	0,5202
25	0,04	1	1,4884
$\sum$	1	18,9	4,05

$$Var(X) = \sigma_X^2 = \sum_{i=1}^{k} (x_i - E(X))^2 p_i$$

idade(Y)	P(Y)	Y . P(Y)	$(Y-18,9)^2 . p(Y)$
17	0,18	3,06	0,6498
18	0,44	7,92	0,3564
19	0,14	2,66	0,0014
20	0,08	1,6	0,0968
21	0,06	1,26	0,2646
22	0	0	0
23	0,04	0,92	0,6724
24	0,02	0,48	0,5202
25	0,04	1	1,4884
$\sum$	1	18,9	4,05

$$\boxed{\sigma_X^2 = 4,05}$$

$$Var(Y) = \sigma_Y^2 = \sum_{i=1}^{k} (y_i - E(Y))^2 p_i$$

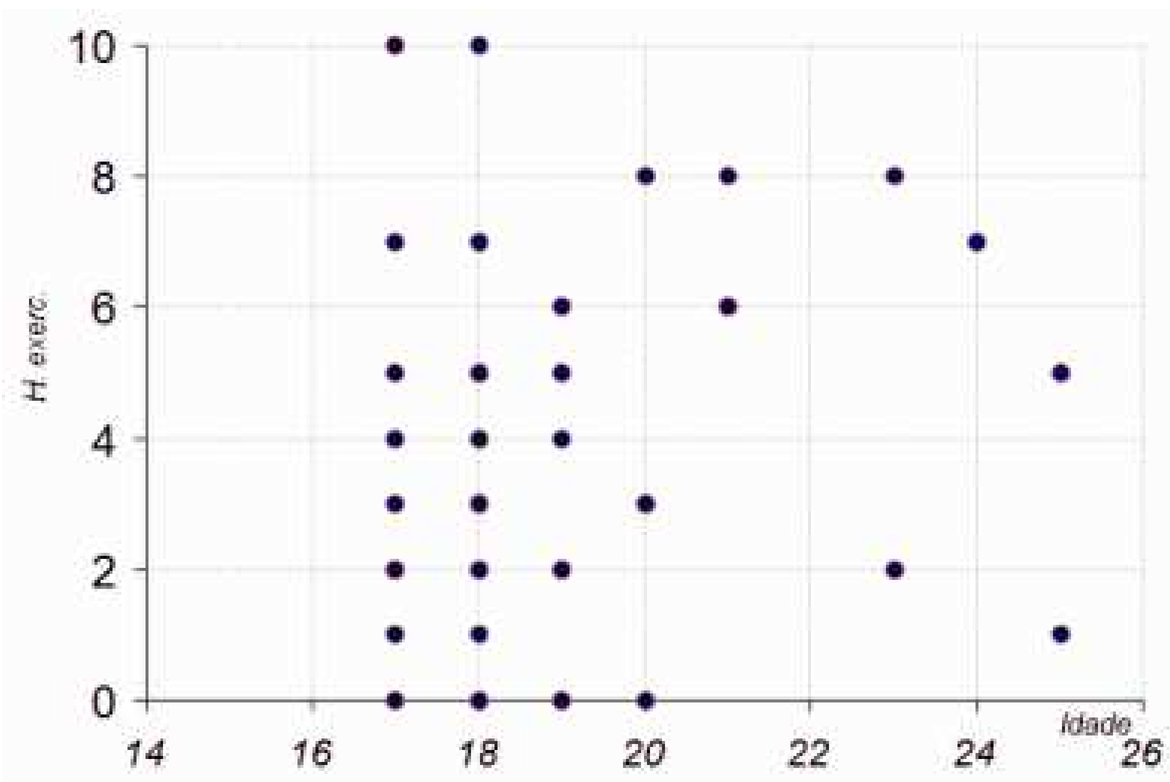
h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	$\sum$
P(Y)	0,160	0,060	0,160	0,120	0,080	0,160	0,060	0,100	0,060	0,000	0,040	1,000
Y. P(Y)	0,000	0,060	0,320	0,360	0,320	0,800	0,360	0,700	0,480	0,000	0,400	3,800
$(Y-3,8)^2 . p(Y)$	2,3104	0,470	0,518	0,077	0,003	0,230	0,290	1,024	1,058	0,000	1,538	7,520

$$Var(Y) = \sigma_Y^2 = \sum_{i=1}^{k} (y_i - E(Y))^2 p_i$$

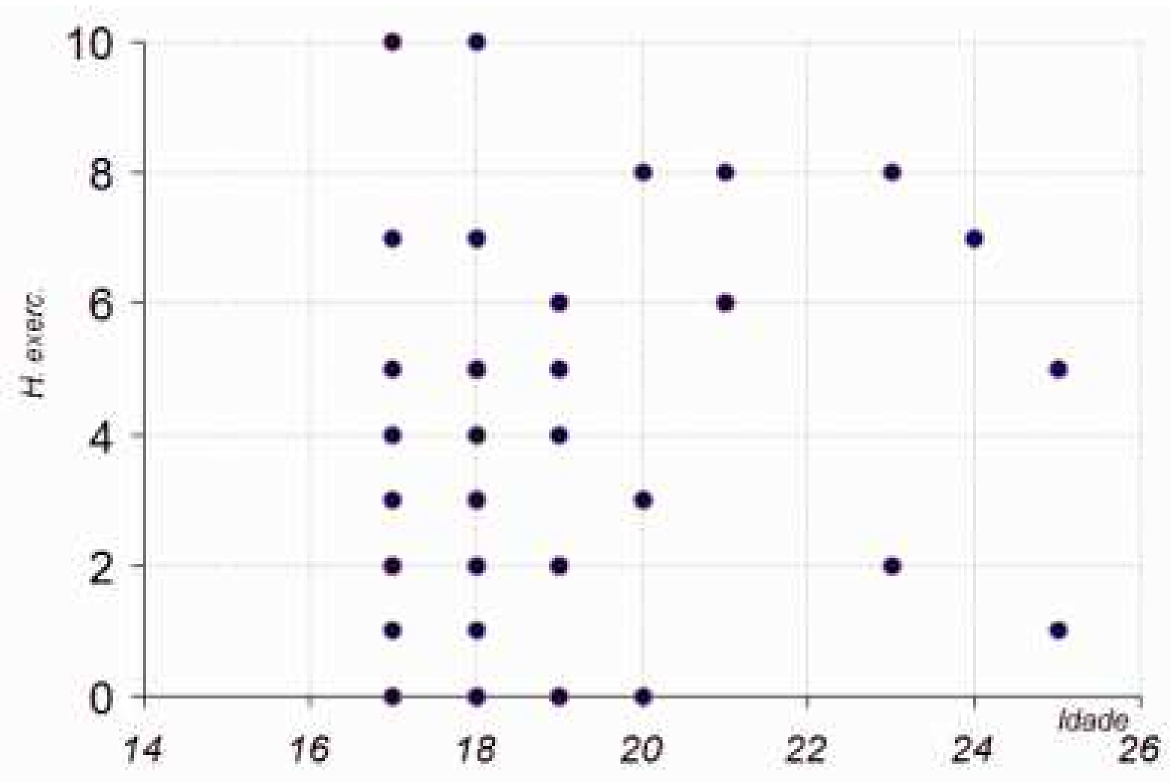
h.e.(Y)	0	1	2	3	4	5	6	7	8	9	10	$\sum$
P(Y)	0,160	0,060	0,160	0,120	0,080	0,160	0,060	0,100	0,060	0,000	0,040	1,000
Y. P(Y)	0,000	0,060	0,320	0,360	0,320	0,800	0,360	0,700	0,480	0,000	0,400	3,800
$(Y-3,8)^2 \cdot p(Y)$	2,3104	0,470	0,518	0,077	0,003	0,230	0,290	1,024	1,058	0,000	1,538	7,520

$$\boxed{\sigma_Y^2 = 7,52}$$

id \ he	0	1	2	3	4	5	6	7	8	9	10	Total
17	1	1	2	1	1	1	0	1	0	0	1	9
18	4	1	3	3	2	5	0	3	0	0	1	22
19	2	0	2	0	1	1	1	0	0	0	0	7
20	1	0	0	2	0	0	0	0	1	0	0	4
21	0	0	0	0	0	0	2	0	1	0	0	3
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	1	0	0	0	0	0	1	0	0	2
24	0	0	0	0	0	0	0	1	0	0	0	1
25	0	1	0	0	0	1	0	0	0	0	0	2
Total	8	3	8	6	4	8	3	5	3	0	2	50



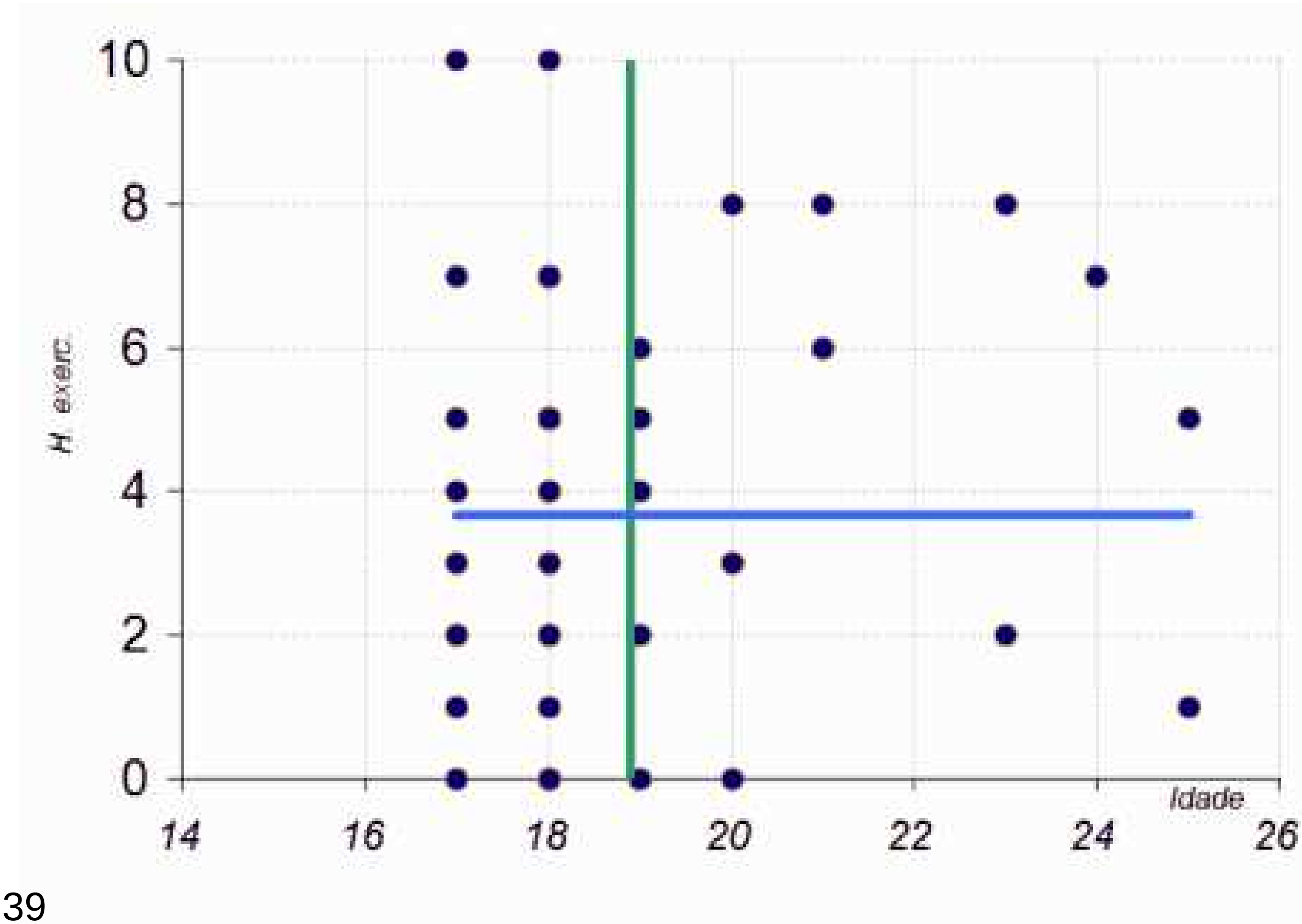
$$E(X) = \sum_{i=1}^{n} x_i p_i$$



$$E(X) = 18,9$$

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

$$E(Y) = \sum_{i=1}^{m} y_i p_i$$



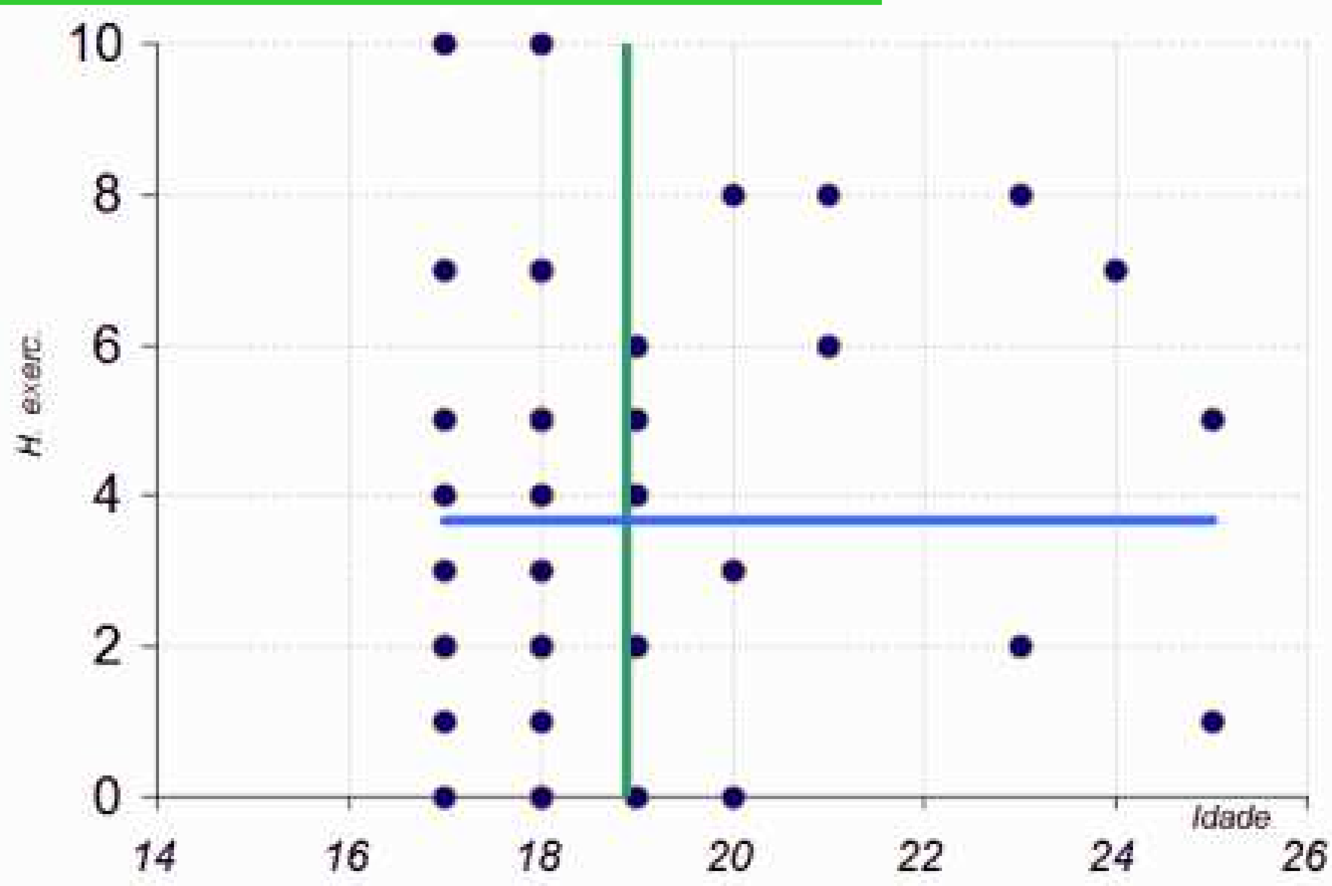
$$E(X) = 18,9$$

$$E(Y) = 3,8$$

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

$$E(Y) = \sum_{i=1}^{m} y_i p_i$$

$$\boxed{Var(X) = \sigma_X^2 = \sum_{i=1}^{k} (x_i - E(X))^2 p_i}$$



$$E(X) = 18,9$$

$$E(Y) = 3,8$$

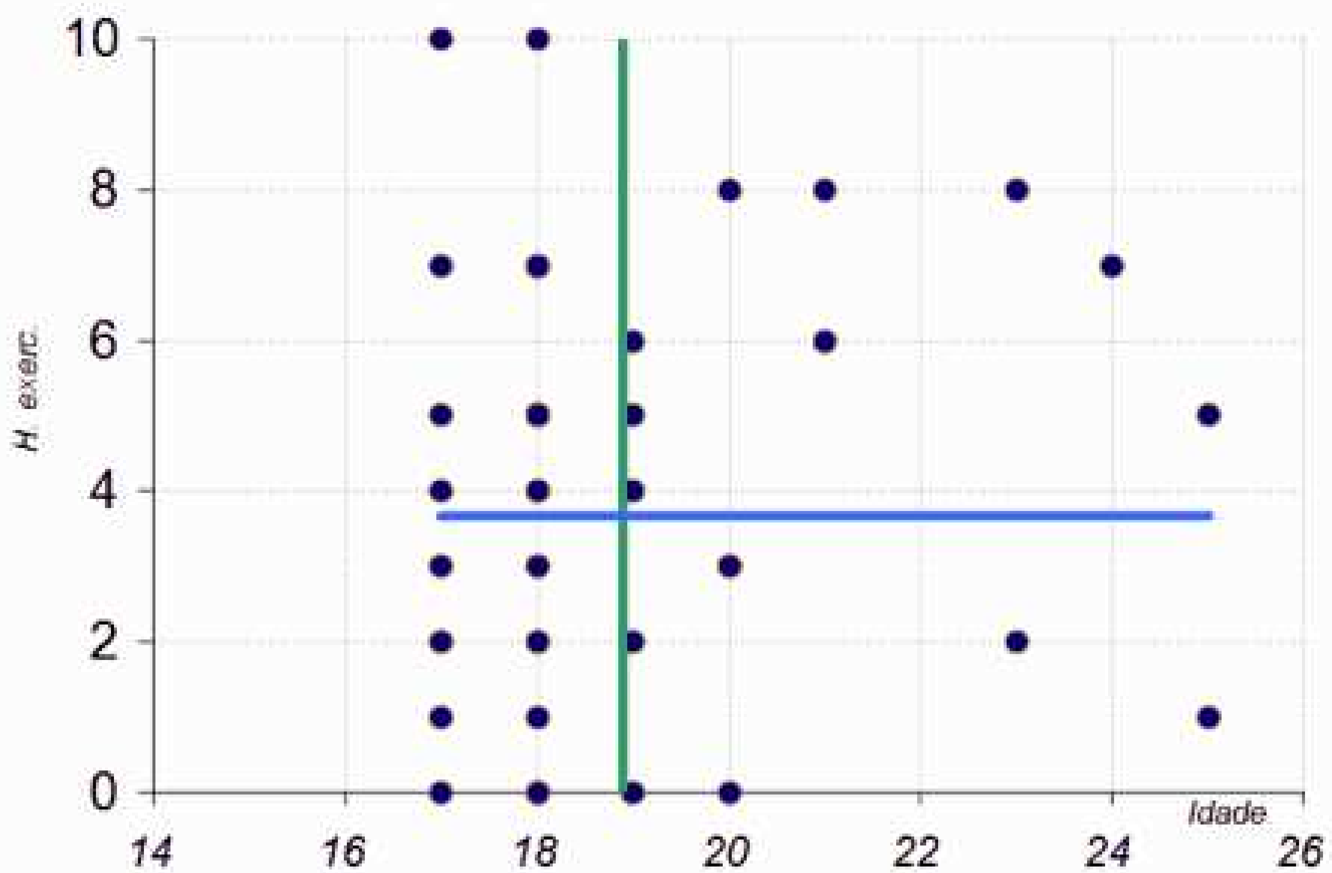
$$\boxed{\sigma_X^2 = 4,05}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

$$E(Y) = \sum_{i=1}^{m} y_i p_i$$

$$Var(X) = \sigma_X^2 = \sum_{i=1}^{k} (x_i - E(X))^2 p_i$$

$$\boxed{Var(Y) = \sigma_Y^2 = \sum_{i=1}^{k} (y_i - E(Y))^2 p_i}$$



$$E(X) = 18,9$$

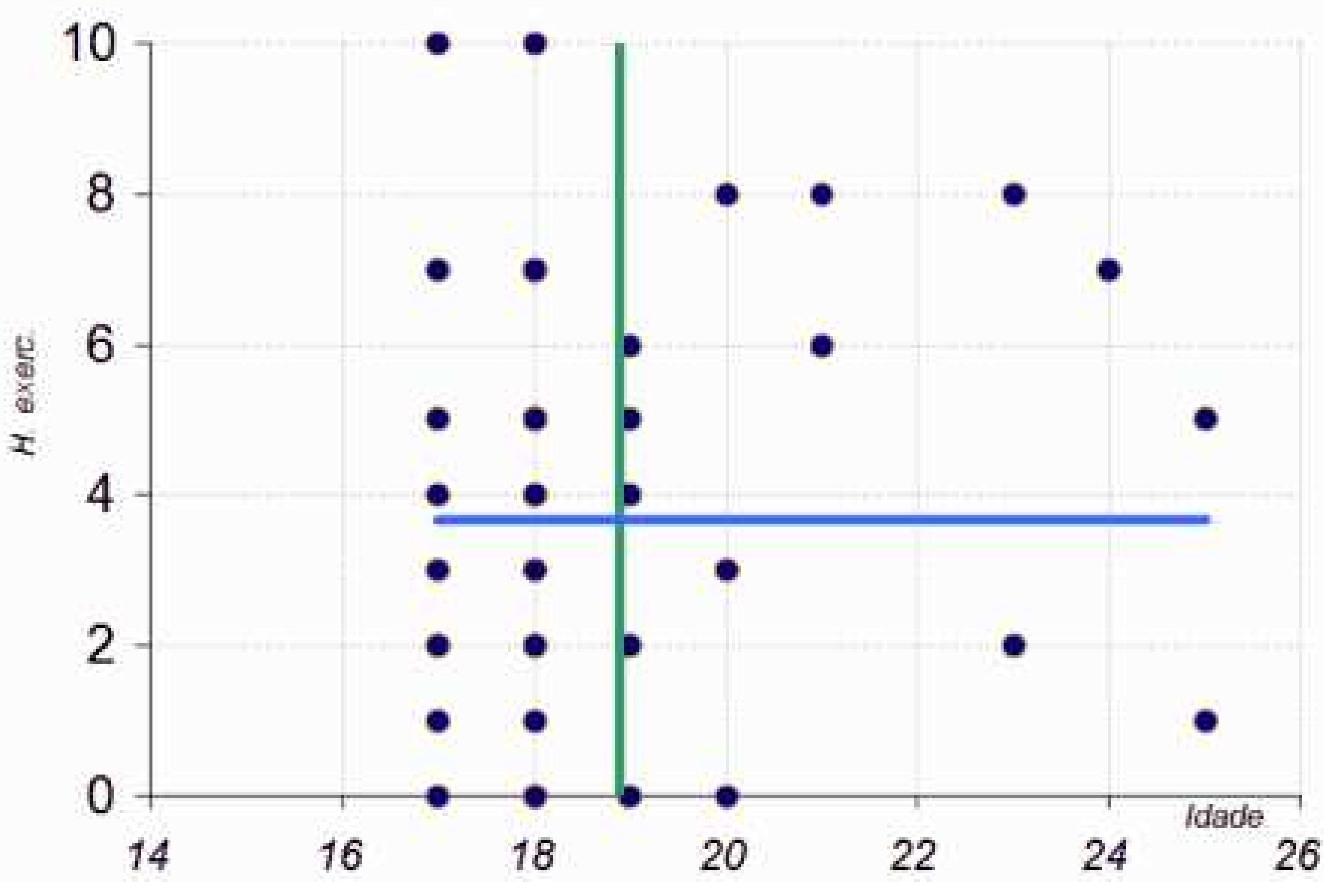
$$E(Y) = 3,8$$

$$\sigma_X^2 = 4,05$$

$$\boxed{\sigma_Y^2 = 7,52}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \qquad E(Y) = \sum_{i=1}^{m} y_i p_i$$

$$Var(X) = \sigma_X^2 = \sum_{i=1}^{k} (x_i - E(X))^2 p_i \qquad Var(Y) = \sigma_Y^2 = \sum_{i=1}^{k} (y_i - E(Y))^2 p_i$$



$$E(X) = 18,9$$

$$E(Y) = 3,8$$

$$\sigma_X^2 = 4,05$$

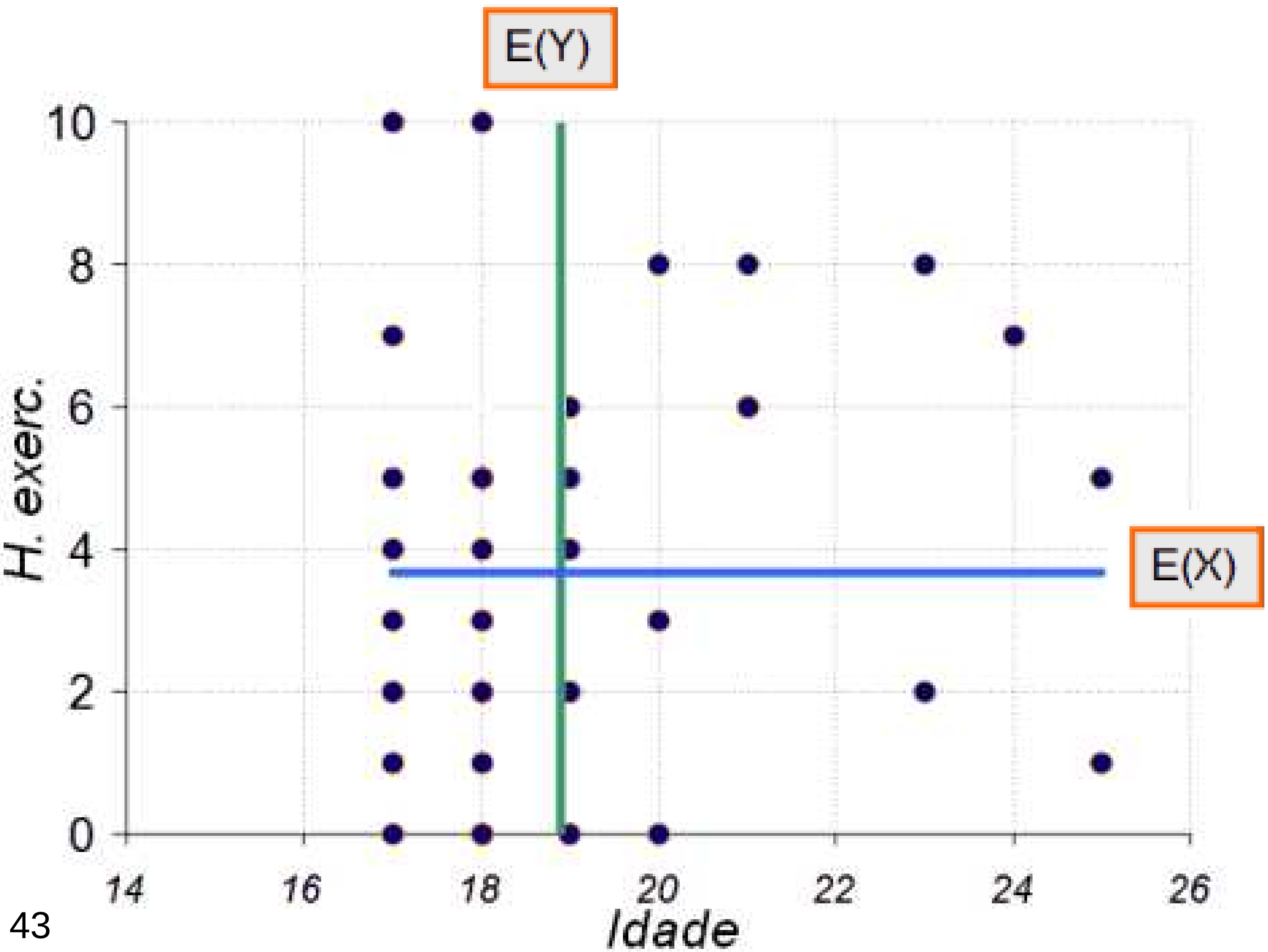
$$\boxed{\sigma_X = 2,01}$$

$$\sigma_Y^2 = 7,52$$

$$\boxed{\sigma_Y = 2,74}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \qquad E(Y) = \sum_{i=1}^{m} y_i p_i$$

$$Var(X) = \sigma_X^2 = \sum_{i=1}^{k} (x_i - E(X))^2 p_i \qquad Var(Y) = \sigma_Y^2 = \sum_{i=1}^{k} (y_i - E(Y))^2 p_i$$



$$\boxed{E(X) = 18,9}$$

$$\boxed{E(Y) = 3,8}$$

$$\sigma_X^2 = 4,05$$

$$\sigma_X = 2,01$$

$$\sigma_Y^2 = 7,52$$

$$\sigma_Y = 2,74$$

Fundação CECIERJ Consórcio cederj

Continuando ... Covariância

A covariância dá uma idéia da dispersão dos valores da variável bidimensional *(X,Y)* em relação ao ponto *(E(X),E(Y))*.

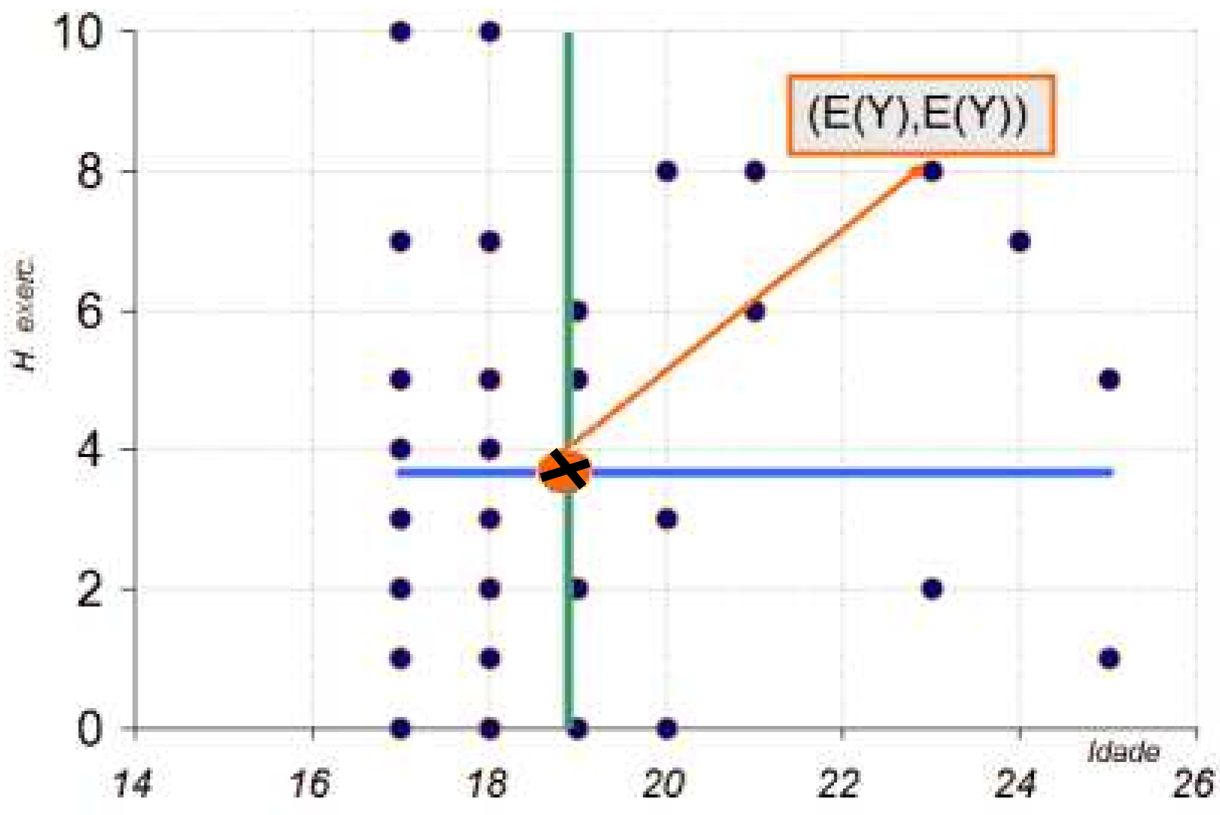
$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)]}$$

Fundação CECIERJ Consórcio cederj

Continuando ... Covariância

A covariância dá uma idéia da dispersão dos valores da variável bidimensional *(X,Y)* em relação ao ponto *(E(X),E(Y))*.

$$Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]$$



Continuando ... Covariância

A covariância dá uma idéia da dispersão dos valores da variável bidimensional *(X,Y)* em relação ao ponto *(E(X),E(Y))*.

$$Cov(X,Y) = \sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)]$$

Ou seja, a covariância é o valor esperado do produto dos desvios de cada variável em relação à sua média. Pode também ser escrita como:

$$Cov(X,Y) = \sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)] = E(X,Y) - E(X)E(Y)$$

(Idade – 18,9)

(horas de exerc. – 3,8)

	-3,8	-2,8	-1,8	-0,8	0,2	1,2	2,2	3,2	4,2	5,2	6,2	Total
-1,9	0,02	0,02	0,04	0,02	0,02	0,02	0	0,02	0	0	0,02	0,18
-0,9	0,08	0,02	0,06	0,06	0,04	0,1	0	0,06	0	0	0,02	0,44
0,1	0,04	0	0,04	0	0,02	0,02	0,02	0	0	0	0	0,14
1,1	0,02	0	0	0,04	0	0	0	0	0,02	0	0	0,08
2,1	0	0	0	0	0	0	0,04	0	0,02	0	0	0,06
3,1	0	0	0	0	0	0	0	0	0	0	0	0
4,1	0	0	0,02	0	0	0	0	0	0,02	0	0	0,04
5,1	0	0	0	0	0	0	0	0,02	0	0	0	0,02
6,1	0	0,02	0	0	0	0,02	0	0	0	0	0	0,04
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

$$Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]$$

$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)]}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \quad \boxed{X = X.y_k}$$

$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \xrightarrow{X = X.y_k} E(X \cdot y_k) = \sum_{i=1}^{n} x_i \; y_k \; p_{ik} \;\rightarrow\; k = 1, 2, \cdots, m$$

$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \quad \Rightarrow \quad E(X \cdot y_k) = \sum_{i=1}^{n} x_i \; y_k \; p_{ik} \; \rightarrow \; k = 1, 2, \cdots, m$$

$$\boxed{E(XY) = \sum_{i=1}^{n} \sum_{k=1}^{m} x_i \; y_k \; p_{ik}}$$

$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \Rightarrow E(X \cdot y_k) = \sum_{i=1}^{n} x_i\ y_k\ p_{ik} \rightarrow k = 1, 2, \cdots, m$$

$$\boxed{E(XY) = \sum_{i=1}^{n} \sum_{k=1}^{m} x_i\ y_k\ p_{ik}}$$

Se fizermos $X = X - \mu_X$ e $Y = Y - \mu_Y$

$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \Rightarrow E(X \cdot y_k) = \sum_{i=1}^{n} x_i \; y_k \; p_{ik} \rightarrow k = 1, 2, \cdots, m$$

$$E(XY) = \sum_{i=1}^{n} \sum_{k=1}^{m} x_i \; y_k \; p_{ik}$$

Se fizermos $X = X - \mu_X \text{ e } Y = Y - \mu_Y$

$$\boxed{E((X-\mu_X)(Y-\mu_Y)) = \sum_{i=1}^{n} \sum_{k=1}^{m} (x_i - \mu_X)(y_k - \mu_Y) p_{ik}}$$

Fundação CECIERJ Consórcio cederj

Exemplo

Covariância - idade e horas de exercícios semanais

Tabela de probabilidade conjunta

Id. \ h.e.	0	1	2	3	4	5	6	7	8	9	10	Total
17	0,02	0,02	0,04	0,02	0,02	0,02	0	0,02	0	0	0,02	0,18
18	0,08	0,02	0,06	0,06	0,04	0,1	0	0,06	0	0	0,02	0,44
19	0,04	0	0,04	0	0,02	0,02	0,02	0	0	0	0	0,14
20	0,02	0	0	0,04	0	0	0	0	0,02	0	0	0,08
21	0	0	0	0	0	0	0,04	0	0,02	0	0	0,06
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0,02	0	0	0	0	0	0,02	0	0	0,04
24	0	0	0	0	0	0	0	0,02	0	0	0	0,02
25	0	0,02	0	0	0	0,02	0	0	0	0	0	0,04
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

Exemplo

Covariância - idade e horas de exercícios semanais

Tabela de probabilidade conjunta

Id. \ h.e.	0	1	2	3	4	5	6	7	8	9	10	Total
17	0,02	0,02	0,04	0,02	0,02	0,02	0	0,02	0	0	0,02	0,18
18	0,08	0,02	0,06	0,06	0,04	0,1	0	0,06	0	0	0,02	0,44
19	0,04	0	0,04	0	0,02	0,02	0,02	0	0	0	0	0,14
20	0,02	0	0	0,04	0	0	0	0	0,02	0	0	0,08
21	0	0	0	0	0	0	0,04	0	0,02	0	0	0,06
22	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0,02	0	0	0	0	0	0,02	0	0	0,04
24	0	0	0	0	0	0	0	0,02	0	0	0	0,02
25	0	0,02	0	0	0	0,02	0	0	0	0	0	0,04
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

$$\boxed{Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$\boxed{E[(X-\mu_X)(Y-\mu_Y)] = \sum_{i=1}^{n}\sum_{k=1}^{m}(x_i-\mu_X)(y_k-\mu_Y)p_{ik}}$$

(Idade – 18,9)

(horas de exerc. – 3,8)

	-3,8	-2,8	-1,8	-0,8	0,2	1,2	2,2	3,2	4,2	5,2	6,2	Total
-1,9	0,02	0,02	0,04	0,02	0,02	0,02	0	0,02	0	0	0,02	0,18
-0,9	0,08	0,02	0,06	0,06	0,04	0,1	0	0,06	0	0	0,02	0,44
0,1	0,04	0	0,04	0	0,02	0,02	0,02	0	0	0	0	0,14
1,1	0,02	0	0	0,04	0	0	0	0	0,02	0	0	0,08
2,1	0	0	0	0	0	0	0,04	0	0,02	0	0	0,06
3,1	0	0	0	0	0	0	0	0	0	0	0	0
4,1	0	0	0,02	0	0	0	0	0	0,02	0	0	0,04
5,1	0	0	0	0	0	0	0	0,02	0	0	0	0,02
6,1	0	0,02	0	0	0	0,02	0	0	0	0	0	0,04
Total	0,16	0,06	0,16	0,12	0,08	0,16	0,06	0,1	0,06	0	0,04	1

$$Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]$$

$$E[(X-\mu_X)(Y-\mu_Y)] = \sum_{i=1}^{n}\sum_{k=1}^{m}(x_i-\mu_X)(y_k-\mu_Y)p_{ik}$$

$(\text{Idade}_{(i)} - 18{,}9)$

$(\text{h.e.}_{(k)} - 3{,}8)$

$(\text{Idade}(i) - 18{,}9) * (\text{h.e.}(k) - 3{,}8) * P(\text{id}(i), \text{he}(k)) =$

	-3,8	-2,8	-1,8	-0,8	0,2	1,2	2,2	3,2	4,2	5,2	6,2	Total
-1,9	0,1444	0,1064	0,1368	0,0304	-0,0076	-0,0456	0	-0,1216	0	0	-0,2356	0,0076
-0,9	0,2736	0,0504	0,0972	0,0432	-0,0072	-0,108	0	-0,1728	0	0	-0,1116	0,0648
0,1	-0,0152	0	-0,0072	0	0,0004	0,0024	0,0044	0	0	0	0	-0,0152
1,1	-0,0836	0	0	-0,0352	0	0	0	0	0,0924	0	0	-0,0264
2,1	0	0	0	0	0	0	0,1848	0	0,1764	0	0	0,3612
3,1	0	0	0	0	0	0	0	0	0	0	0	0
4,1	0	0	-0,1476	0	0	0	0	0	0,3444	0	0	0,1968
5,1	0	0	0	0	0	0	0	0,3264	0	0	0	0,3264
6,1	0	-0,3416	0	0	0	0,1464	0	0	0	0	0	-0,1952
Total	0,3192	-0,1848	0,0792	0,0384	-0,0144	-0,0048	0,1892	0,032	0,6132	0	-0,3472	0,72

(Idade(i) – 18,9)* (h.e.(k) – 3,8)*P(id(i),he(k))=

= 0,1 * 1,2 * 0,02 = 0,0024

	-3,8	-2,8	-1,8	-0,8	0,2	1,2	2,2	3,2	4,2	5,2	6,2	Total
-1,9	0,1444	0,1064	0,1368	0,0304	-0,0076	-0,0456	0	-0,1216	0	0	-0,2356	0,0076
-0,9	0,2736	0,0504	0,0972	0,0432	-0,0072	-0,108	0	-0,1728	0	0	-0,1116	0,0648
0,1	-0,0152	0	-0,0072	0	0,0004	0,0024	0,0044	0	0	0	0	-0,0152
1,1	-0,0836	0	0	-0,0352	0	0	0	0	0,0924	0	0	-0,0264
2,1	0	0	0	0	0	0	0,1848	0	0,1764	0	0	0,3612
3,1	0	0	0	0	0	0	0	0	0	0	0	0
4,1	0	0	-0,1476	0	0	0	0	0	0,3444	0	0	0,1968
5,1	0	0	0	0	0	0	0	0,3264	0	0	0	0,3264
6,1	0	-0,3416	0	0	0	0,1464	0	0	0	0	0	-0,1952
Total	0,3192	-0,1848	0,0792	0,0384	-0,0144	-0,0048	0,1892	0,032	0,6132	0	-0,3472	0,72

	-3,8	-2,8	-1,8	-0,8	0,2	1,2	2,2	3,2	4,2	5,2	6,2	Total
-1,9	0,1444	0,1064	0,1368	0,0304	-0,0076	-0,0456	0	-0,1216	0	0	-0,2356	0,0076
-0,9	0,2736	0,0504	0,0972	0,0432	-0,0072	-0,108	0	-0,1728	0	0	-0,1116	0,0648
0,1	-0,0152	0	-0,0072	0	0,0004	0,0024	0,0044	0	0	0	0	-0,0152
1,1	-0,0836	0	0	-0,0352	0	0	0	0	0,0924	0	0	-0,0264
2,1	0	0	0	0	0	0	0,1848	0	0,1764	0	0	0,3612
3,1	0	0	0	0	0	0	0	0	0	0	0	0
4,1	0	0	-0,1476	0	0	0	0	0	0,3444	0	0	0,1968
5,1	0	0	0	0	0	0	0	0,3264	0	0	0	0,3264
6,1	0	-0,3416	0	0	0	0,1464	0	0	0	0	0	-0,1952
Total	0,3192	-0,1848	0,0792	0,0384	-0,0144	-0,0048	0,1892	0,032	0,6132	0	-0,3472	0,72

$$\boxed{\sigma_{X,Y} = 0,72}$$

7.6 Correlação entre variáveis aleatórias

O coeficiente de correlação entre variáveis aleatórias discretas X e Y é dado por:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Coeficiente de correlação fica entre +1 e -1 e valores próximos de 1 indicam correlação forte! Servem para comparação.

7.6 Correlação entre variáveis aleatórias

O coeficiente de correlação entre variáveis aleatórias discretas X e Y é dado por:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Coeficiente de correlação fica entre +1 e -1 e valores próximos de 1 indicam correlação forte! Servem para comparação.

Para nosso exemplo ...

$$\boxed{\rho_{X,Y} = \frac{0,72}{2,01 \times 2,74} = 0,13}$$

Fundação CECIERJ
Consórcio cederj

Aula 7

Professores:

Otton Teixeira da Silveira Filho
Regina Célia Paula Leal Toledo

Variáveis bidimensionais

Conteúdo:

7 Introdução
- 7.1 Função de probabilidade conjunta
- 7.2 Distribuição marginal de probabilidade
- 7.3 Probabilidade condicional de variáveis aleatórias discretas
- 7.4. Independência de variáveis aleatórias
- 7.5. Covariância
- 7.6 Correlação entre variáveis aleatórias discretas