Ciência dos dados

Variáveis Aleatórias Discretas Bidimensionais

Distribuição de probabilidades conjuntas
Distribuição de probabilidades condicionais
Esperança e variância condicionais
Covariância
Correlação
Propriedades de Esperança e Variância

Magalhães e Lima (7ª. Edição): Capítulo 5

Distribuição conjunta

Normalmente estamos interessados na ocorrência de eventos envolvendo mais do que uma variável aleatória (v.a.).

Por exemplo, uma financeira gostaria de saber qual é a probabilidade de uma pessoa ser inadimplente, ganhar mais de 1000 reais, ter mais de 4 filhos, ser casada e possuir um nível de escolaridade médio.

Isso é um exemplo de probabilidade conjunta.

Exemplo 1

Pesquisa Nacional por Amostra de Domicílios (PNAD)
Obtém informações anuais sobre características demográficas e socioeconômicas da população, como sexo, idade, educação, trabalho e rendimento, e características dos domicílios, e, com periodicidade variável, informações sobre migração, fecundidade, nupcialidade, entre outras, tendo como unidade de coleta os domicílios. Temas específicos abrangendo aspectos demográficos, sociais e econômicos também são investigados. *Fonte: IBGE.*

A distribuição conjunta do número de carros (X) e do número de televisores (Y) dos moradores de um bairro de São Paulo é dada pela tabela abaixo.

X: Carros	Y: TV		
	0	1	2
0	0,10	0,21	0,11
1	0,05	0,20	0,33

Exemplo 1 (Continuação)

a. Qual a probabilidade de uma família ter televisores em casa? E se a família tem automóvel, quanto vale essa probabilidade? Esses resultados indicam algum tipo de associação entre essas duas variáveis (X e Y)? Justifique sua resposta.

b. Há dependência entre X e Y?

c. Qual o número esperado de televisores de uma família? E o desvio padrão?

d. Determine a correlação entre X e Y. Interprete o resultado.

e. Sabendo que uma família tem um automóvel, qual é o seu número esperado de televisores? E a variância?

Distribuição marginal

Caso Unidimensional:

Sejam *X* e *Y* duas variáveis aleatórias discretas.

Então, *X* tem uma distribuição marginal denotada por

$$\mathbf{P(X = \boldsymbol{x})}$$

E *Y* tem uma distribuição marginal denotada por

$$\mathbf{P(Y = \boldsymbol{y})}$$

Distribuição conjunta

Caso Bidimensional:

Sejam *X* e *Y* duas variáveis aleatórias discretas. Então, (*X*,*Y*) tem uma distribuição conjunta que será denotada

$$\mathbf{P}(\mathbf{X} = x; \mathbf{Y} = y)$$

Em alguns casos, é fácil obter a distribuição conjunta se tivermos a função de probabilidade de *X* e *Y*.

Distribuição condicional

Caso Condicional:

Sejam *X* e *Y* duas variáveis aleatórias discretas. Então, uma probabilidade condicional de *Y=y* dado *X=x* será denotada

$$\mathbf{P}(\mathbf{Y} = \boldsymbol{y} | \mathbf{X} = \boldsymbol{x})$$

Para obter, fazemos

$$\mathbf{P}(\mathbf{Y} = \boldsymbol{y} | \mathbf{X} = \boldsymbol{x}) = \frac{\mathbf{P}(\mathbf{X} = \boldsymbol{x}; \mathbf{Y} = \boldsymbol{y})}{\mathbf{P}(\mathbf{X} = \boldsymbol{x})}$$

Resumo - Definições

1. Distribuição conjunta de $\mathbf{X}$ e $\mathbf{Y}$:

   $\mathbf{P(X = \boldsymbol{x}; Y = \boldsymbol{y})}$, para todo $\boldsymbol{x}$ e $\boldsymbol{y}$.

2. Distribuições marginais de $\mathbf{X}$ e $\mathbf{Y}$:

   $\mathbf{P(X = \boldsymbol{x}) = \sum_{\boldsymbol{y}} P(X = \boldsymbol{x}; Y = \boldsymbol{y})}$, para todo $\boldsymbol{x}$.

   $\mathbf{P(Y = \boldsymbol{y}) = \sum_{\boldsymbol{x}} P(X = \boldsymbol{x}; Y = \boldsymbol{y})}$, para todo $\boldsymbol{y}$.

3. Distribuição condicional de $\mathbf{X}$ dado $\mathbf{Y} = \boldsymbol{y}$:

   $\mathbf{P(X = \boldsymbol{x} | Y = \boldsymbol{y}) = \frac{P(X=\boldsymbol{x}; Y=\boldsymbol{y})}{P(Y=\boldsymbol{y})}}$, para todo $\boldsymbol{x}$.

Variáveis Aleatórias Independentes

X e Y são <u>v.a.'s independentes</u> se, e somente se:

$$P(X = x|Y = y) = P(X = x)$$

Alternativamente,

$$P(Y = y|X = x) = P(Y = y)$$

Por consequência,

$$P(X = x; Y = y) = P(X = x)P(Y = y)$$

para todos os valores de x e y possíveis.

Valor Esperado (média ou esperança) de uma variável discreta

O valor esperado de uma **v.a. discreta X** é a soma dos produtos dos valores da variável pelas respectivas probabilidades da variável assumir tais valores, ou seja,

$$\mathrm{E(X)} = \mu_{\mathrm{X}} = \sum_{x} x\,\mathrm{P(X} = x)$$

Variância de uma variável discreta aleatória

A variância de uma **v.a. discreta X** é

$$\mathrm{Var(X)} = \sigma^2{}_{\mathrm{X}} = \sum_x (x - \mathrm{E(X)})^2\, \mathrm{P(X} = x)$$

ou ainda, $\mathrm{Var(X)} = \mathrm{E}[(\mathrm{X} - \mathrm{E(X)})^2] =$

$$= \mathrm{E(X^2)} - [\mathrm{E(X)}]^2$$

em que $\mathrm{E(X^2)} = \sum_x x^2\, \mathrm{P(X} = x)$

Distribuição Condicional de X dado Y=y P(X=x |Y=y)

Esperança condicional

Definição. **A esperança condicional de *X*, dado que *Y* = *y*, é definida por**

$$\mathrm{E}(\mathrm{X}|\mathrm{Y}=y)=\sum_{x} x\,\mathrm{P}(\mathrm{X}=x|\mathrm{Y}=y)$$

Distribuição Condicional de X dado Y=y P(X=x |Y=y)

Variância condicional

Definição. **Dadas as v.a. *X* e *Y*, a variância de *X*, condicionada a Y = y, é simplesmente a variância associada com a distribuição condicional de X dado que Y = y:**

$$\mathrm{Var}(X|Y=y) = \sum_{x} (x - E(X|Y=y))^2 \, P(X=x|Y=y)$$

ou ainda,

$$\mathrm{Var}(X|Y=y) = E(X^2|Y=y) - [E(X|Y=y)]^2$$

Características das Distribuições Conjuntas e Condicionais

Medidas de associação:
Covariância e Correlação

A distribuição conjunta de duas v.a. descreve completamente a relação entre elas, mas é interessante ter uma medida resumo de como, em média, as duas variáveis variam uma com a outra.

Covariância entre X e Y

Definimos a covariância entre X e Y por

$$\mathrm{Cov}(X, Y) = \sum_x \sum_y \big(x - E(X)\big)\big(y - E(Y)\big) P(X = x; Y = y)$$

Ou ainda,

$$\mathrm{Cov}(X, Y) = E[\{X - E(X)\}\{Y - E(Y)\}] =$$

$$= E(XY) - E(X)E(Y)$$

em que

$$E(XY) = \sum_x \sum_y xy\, P(X = x; Y = y)$$

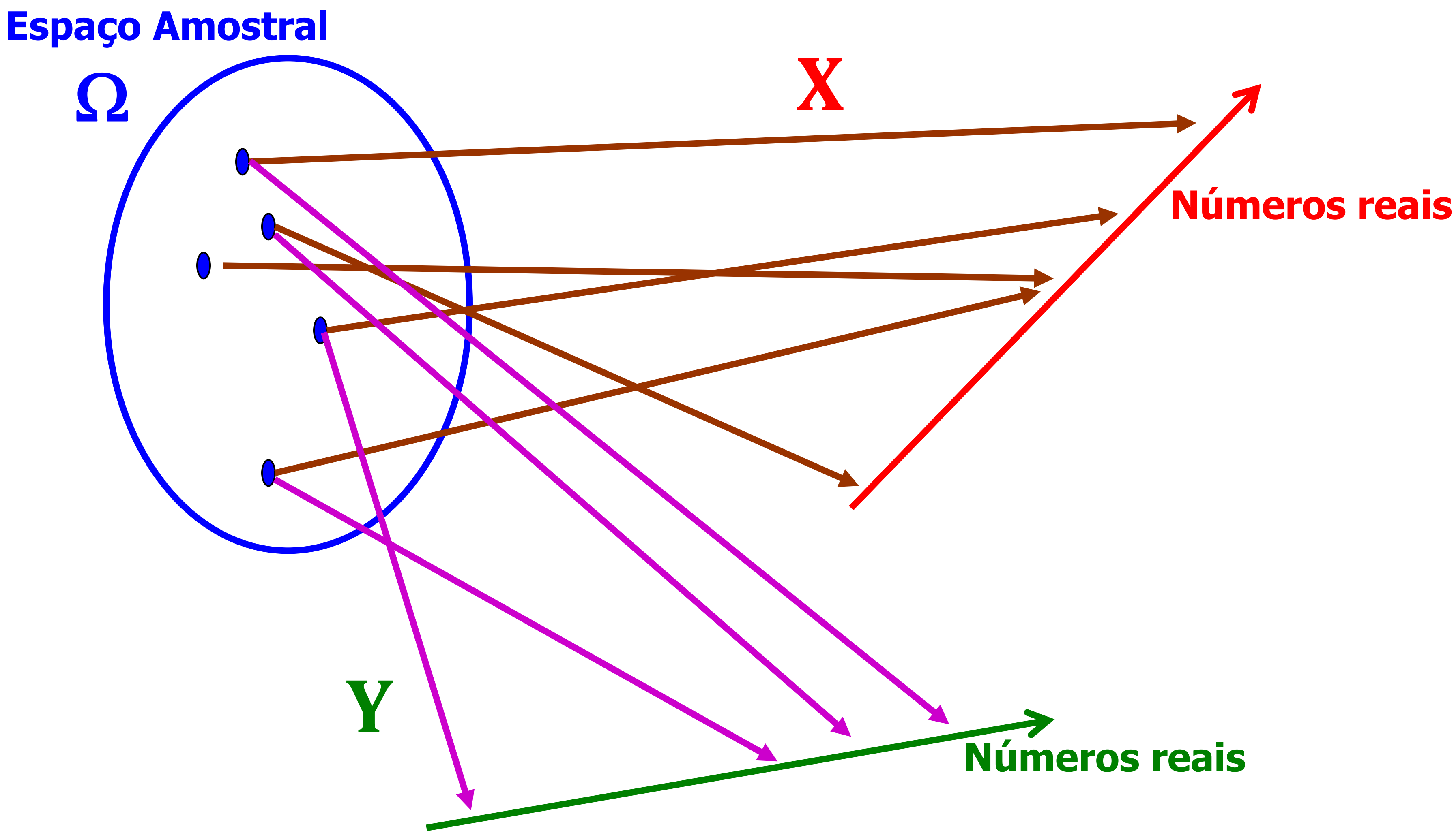
Correlação entre X e Y

A correlação linear entre X e Y é dada por

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

$$= \frac{\mathrm{Cov}(X,Y)}{\mathrm{DP}(X)\mathrm{DP}(Y)}$$

Variáveis Aleatórias Bidimensionais



Exemplo 2:

Há três investimentos disponíveis, com desempenhos independentes. A probabilidade de cada investimento ser bem sucedido é de 60%.

Quando um invest. é bem sucedido ganha-se 10% do valor investido, se não for, perde-se 10%. João aplicará $100 no invest. 1 e $100 no invest. 2. Maria aplicará $50 no primeiro e $50 no terceiro. Seja X o **ganho** obtido por João e Y por Maria. Responda:

a) Qual a probabilidade de João ter um ganho de $20? E seu ganho esperado?

b) Construa uma tabela com as probabilidades conjuntas que associam todos os valores de X e Y.

c) Qual a probabilidade de João ter ganhado zero, sabendo que Maria também teve um ganho zero?

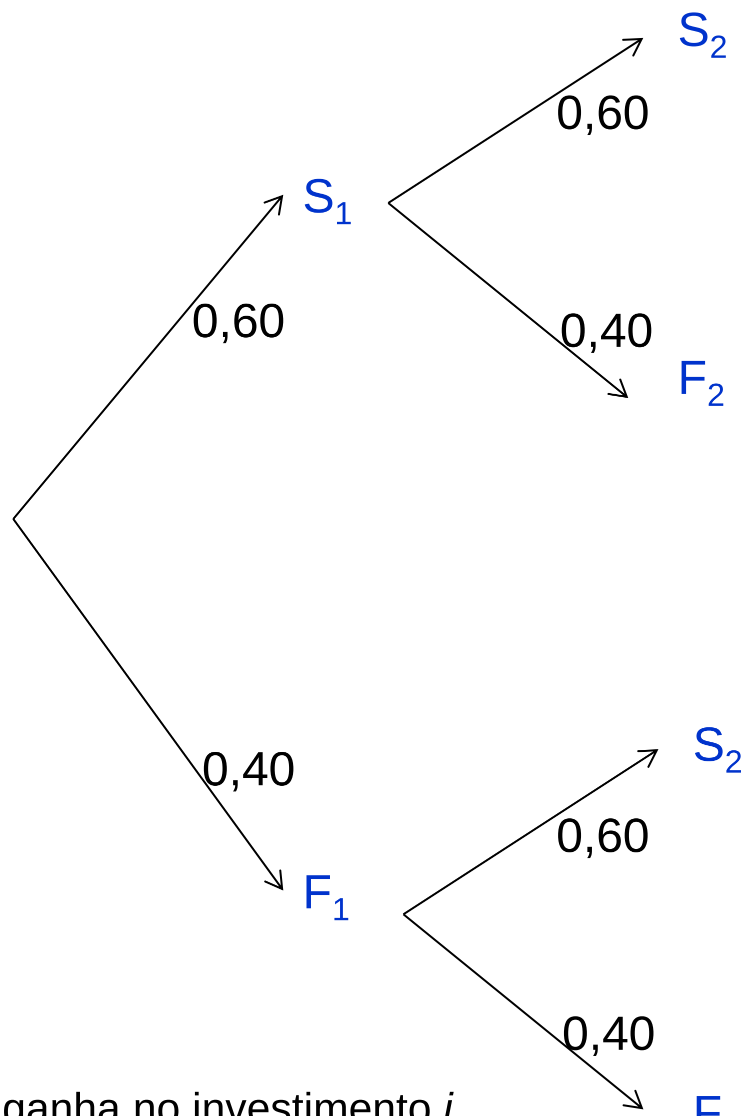
d) Qual o ganho esperado de João quando Maria tem zero de ganho?

e) X e Y são independentes? Calcule a correlação entre X e Y.

Espaço amostral

Item a)

Prob.



0,36

0,24

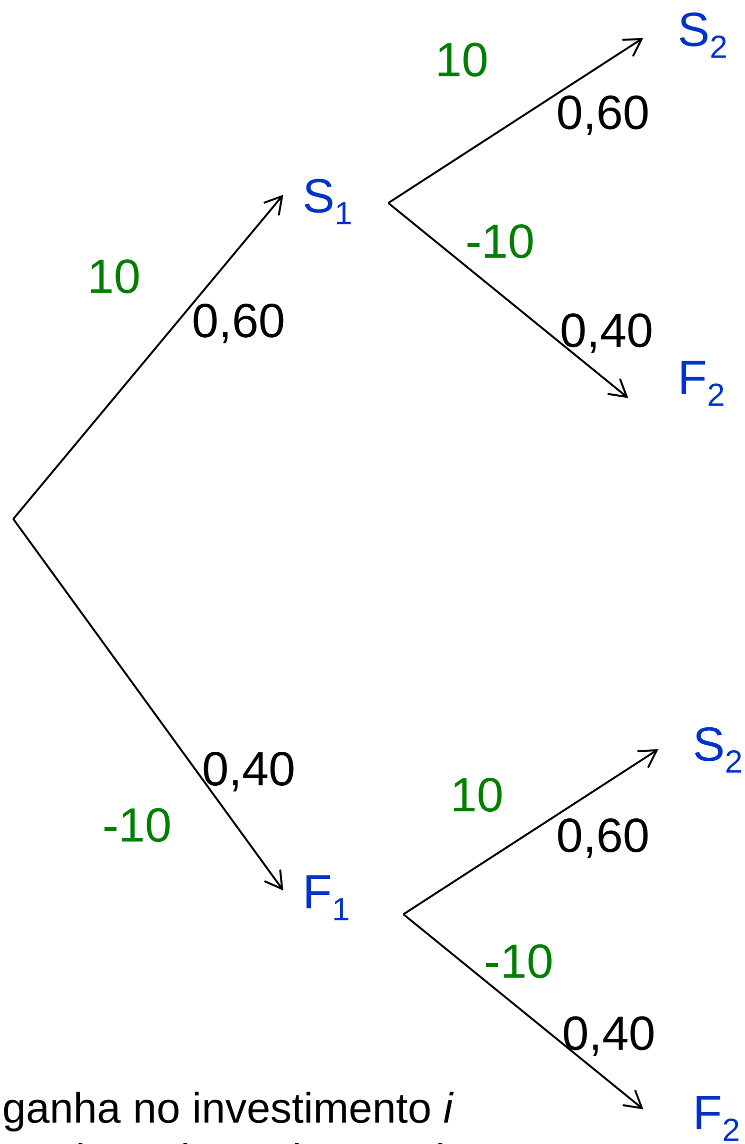
0,24

0,16

S_i: ganha no investimento *i*
F_i: perde no investimento *i*

Espaço amostral

Item a)



X	Prob.
20	0,36
0	0,24
0	0,24
- 20	0,16

S_i: ganha no investimento *i*

F_i: perde no investimento *i*

Item a)

Distribuição (marginal) de X

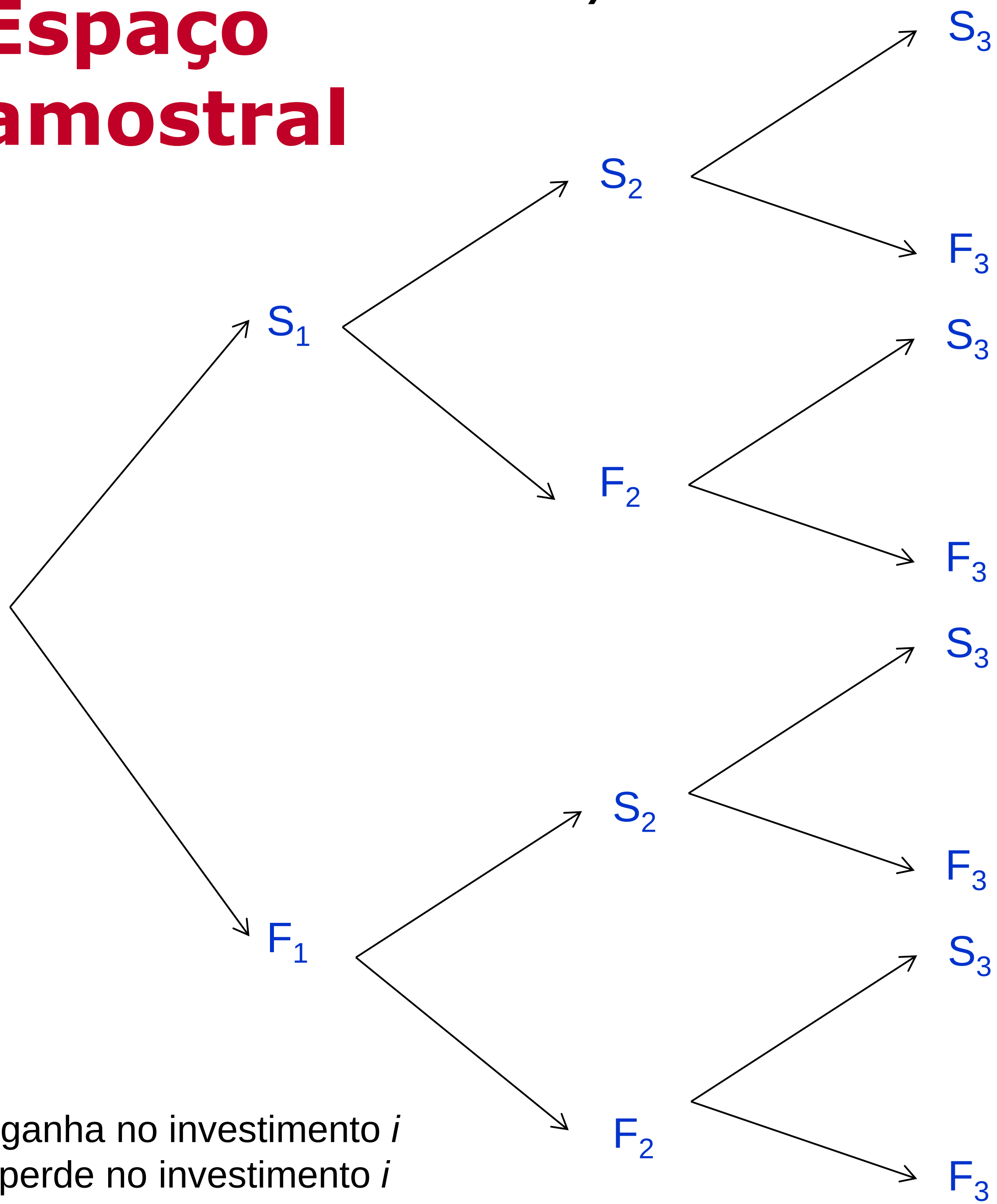
x	P(X=x)
-20	0,160
0	0,480
20	0,360

E(X) = ?

E(X) = 4,00

Espaço amostral

Item b)

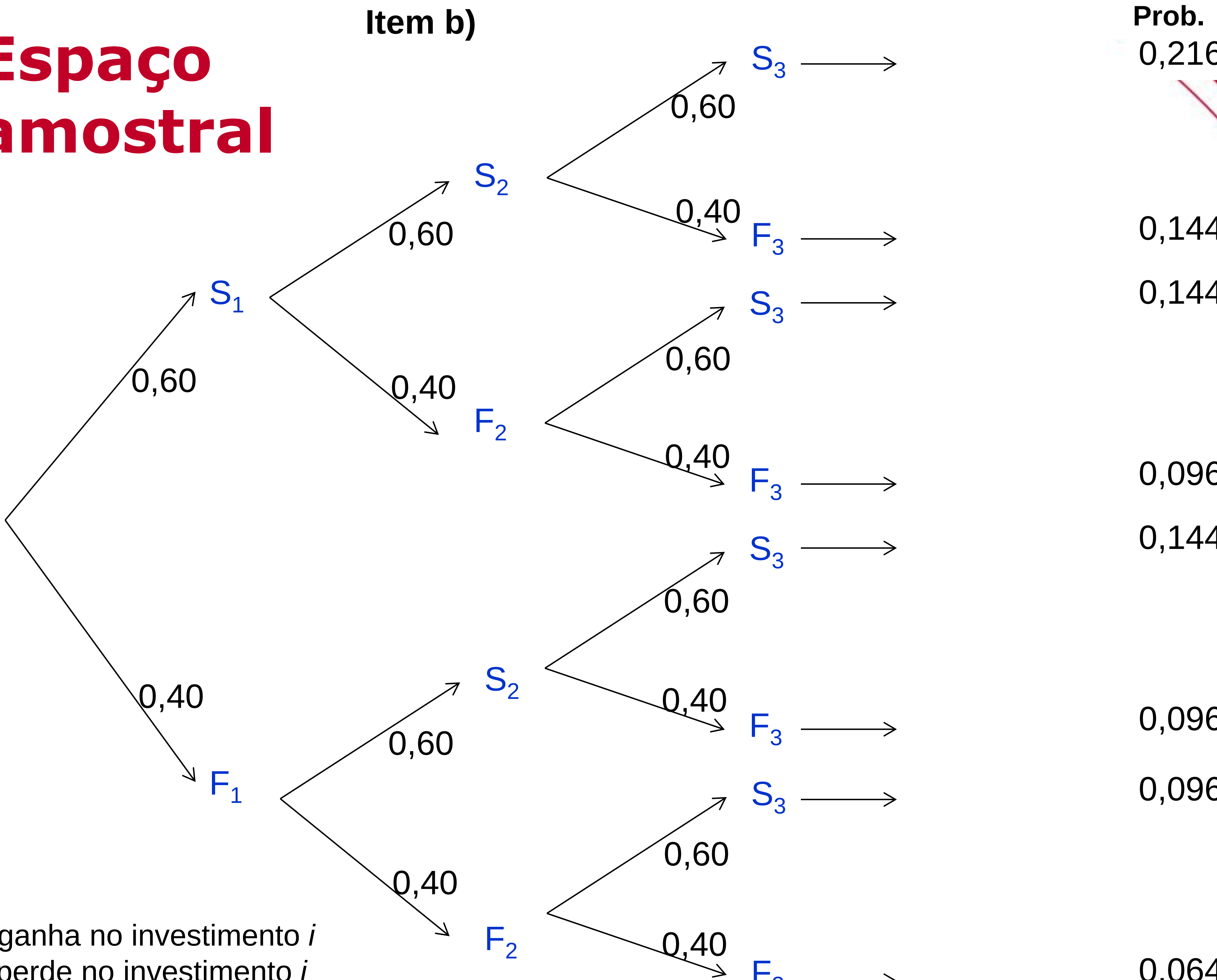


S_i: ganha no investimento *i*
F_i: perde no investimento *i*

Insper

Espaço amostral

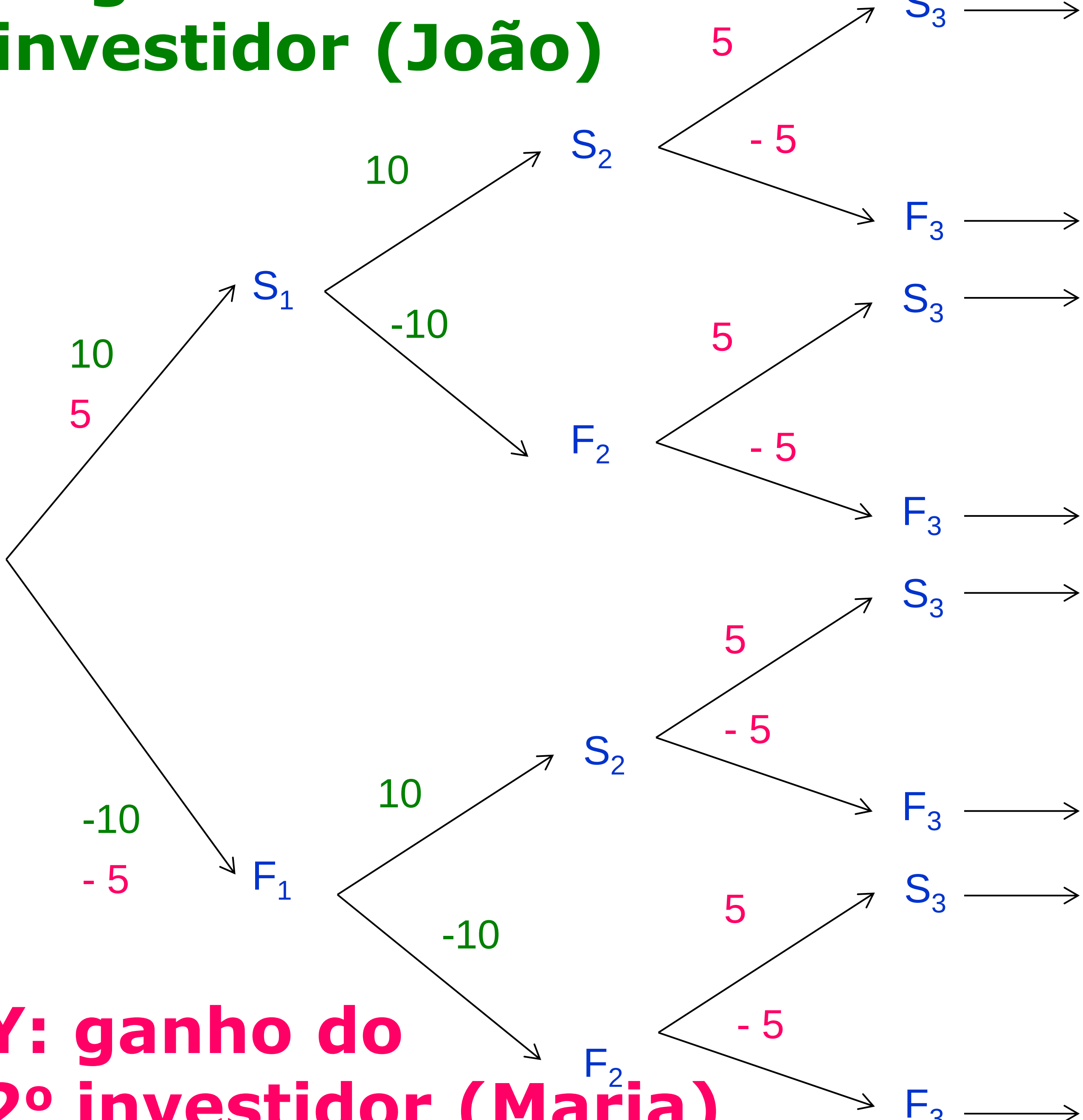
Item b)



S_i: ganha no investimento *i*

F_i: perde no investimento *i*

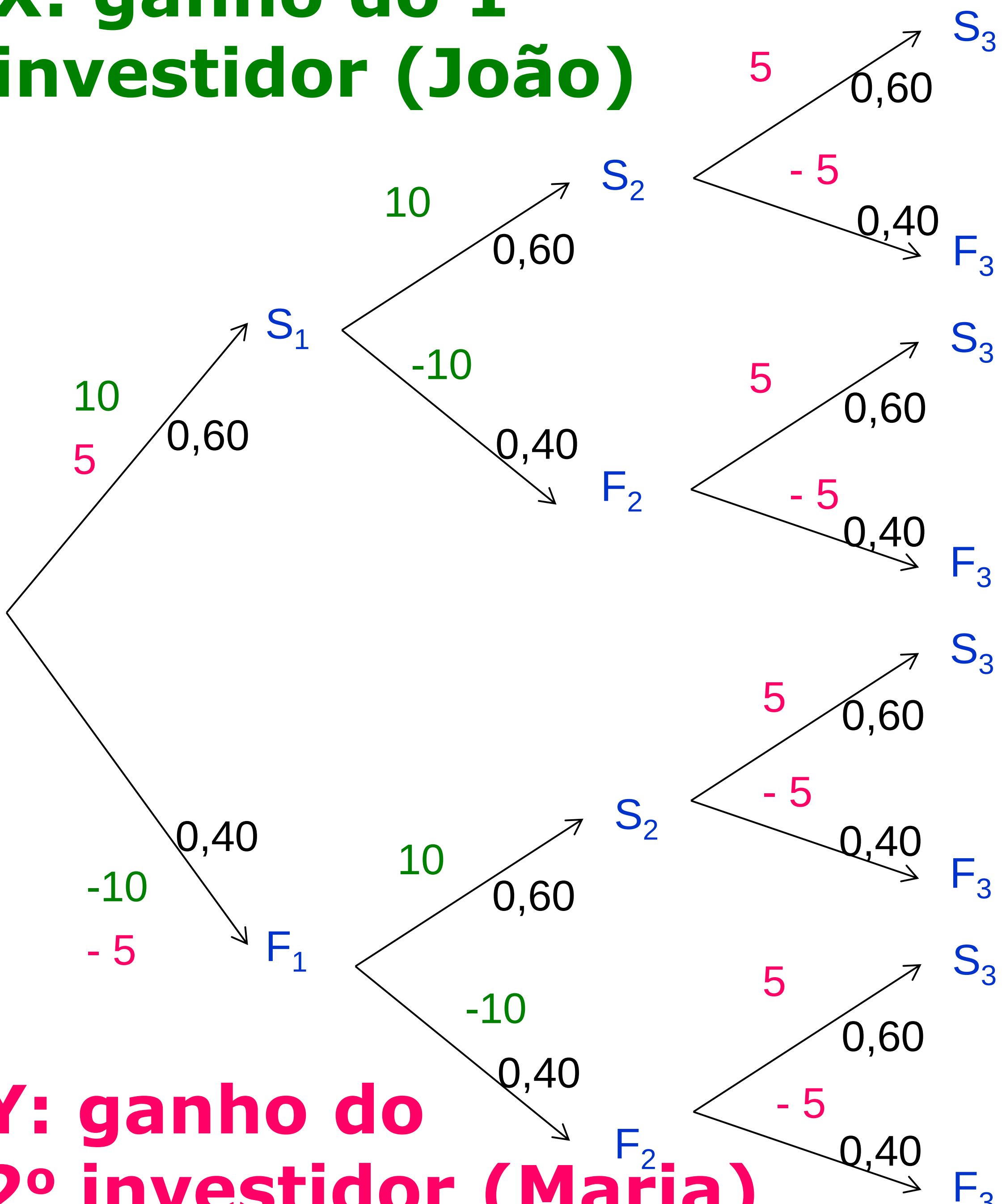
X: ganho do 1º investidor (João)



	X	Y	Prob.
S_3	20	10	0,216
F_3	20	0	0,144
S_3	0	10	0,144
F_3	0	0	0,096
S_3	0	0	0,144
F_3	0	-10	0,096
S_3	- 20	0	0,096
F_3	- 20	- 10	0,064

Y: ganho do 2º investidor (Maria)

X: ganho do 1º investidor (João)



X	Y	Prob.
20	10	0,216
20	0	0,144
0	10	0,144
0	0	0,096
0	0	0,144
0	-10	0,096
- 20	0	0,096
- 20	- 10	0,064

Y: ganho do 2º investidor (Maria)

Rentabilidade - Espaço amostral

Pontos de Ω	X	Y
$S_1S_2S_3$	20	10
$S_1S_2F_3$	20	0
$S_1F_2S_3$	0	10
$S_1F_2F_3$	0	0
$F_1S_2S_3$	0	0
$F_1S_2F_3$	0	-10
$F_1F_2S_3$	-20	0
$F_1F_2F_3$	-20	-10

Continuação

Probabilidade	Pontos de Ω	X	Y
0,216	$S_1S_2S_3$	20	10
0,144	$S_1S_2F_3$	20	0
0,144	$S_1F_2S_3$	0	10
0,096	$S_1F_2F_3$	0	0
0,144	$F_1S_2S_3$	0	0
0,096	$F_1S_2F_3$	0	-10
0,096	$F_1F_2S_3$	-20	0
0,064	$F_1F_2F_3$	-20	-10

Distribuição Conjunta de X e Y - P(X=x;Y=y)

	y			
x	-10	0	10	P(X=x)
-20	0,064	0,096	0,000	0,160
0	0,096	0,240	0,144	0,480
20	0,000	0,144	0,216	0,360
P(Y=y)	0,160	0,480	0,360	1,000

Exemplo 2 (cont.)

Distribuições (marginais) de X e Y

x	P(X=x)
-20	0,160
0	0,480
20	0,360

E(X) = ?

E(X) = 4,00

y	P(Y=y)
-10	0,160
0	0,480
10	0,360

E(Y) = ?

E(Y) = 2,00

Distribuição condicional de X dado Y=0

Sabendo que o segundo investidor (Maria) não ganhou e nem perdeu, qual seria o ganho esperado do primeiro investidor (João)?

x | Y = 0	-20	0	20	Total
P(X=x | Y = 0)	0,20	0,50	0,30	1

E(X | Y = 0) = ?

E(X | Y = 0) = -20 x 0,20 + 0 x 0,50 + 20 x 0,30 = 2,00

(ganho médio do primeiro investidor (João), sabendo que o segundo (Maria) não ganhou e nem perdeu)

Var(X | Y = 0) = ?

Exemplo 2 (cont)

x	y			P(X=x)
	-10	0	10	
-20	0,064	0,096	0,000	**0,160**
0	0,096	0,240	0,144	**0,480**
20	0,000	0,144	0,216	**0,360**
P(Y=y)	**0,160**	**0,480**	**0,360**	**1,000**

→ Qual a distribuição do ganho do primeiro investidor (João), sabendo que o segundo investidor (Maria) perdeu \$10? Obtenha o ganho médio do primeiro investidor (João) nesta situação.

→ Qual a distribuição do ganho do primeiro investidor (João), sabendo que o segundo investidor (Maria) ganhou \$10? Obtenha o ganho médio do primeiro investidor (João) nesta situação.

Exemplo 2 (cont)

Distribuição Condicional de X dado Y= -10

X | Y= -10	*-20*	*0*	*20*	*Total*
P(X=x | Y=y)	0,40	0,60	0,00	1,00

E(X | Y = -10) = -8,00

Distribuição Condicional de X dado Y= 10

X | Y= 10	*-20*	*0*	*20*	*Total*
P(X=x | Y=y)	0,00	0,40	0,60	1,00

E(X | Y = 10) = 12,00

Esperança Condicional

y	$E(X \mid Y=y)$
-10	-8
0	2
10	12

Exemplo 3

Num determinado momento em um certo país, a taxa de juros (X) pode variar 0,25 pontos percentuais (pp), para cima ou para baixo, ou manter-se constante.

Já a taxa de câmbio (Y) pode variar para mais ou para menos em 1 pp, ou manter-se constante.

A tabela seguinte reflete as distribuições marginais e conjunta dessas duas taxas representadas, aqui, por X e Y.

Exemplo 3

X (juros)	Y (câmbio)			P(X=x)
	-1	0	1	
-0,25	0,050	0,070	0,260	0,380
0	0,075	0,210	0,120	0,405
0,25	0,125	0,070	0,020	0,215
P(Y=y)	0,250	0,350	0,400	1

$\mathbf{E(X)} = -0{,}25 * 0{,}380 + 0 * 0{,}405 + 0{,}25 * 0{,}215 = \mathbf{-0{,}041}$

$\mathbf{Var(X)} = (-0{,}25 + 0{,}041)^2 * 0{,}380 + \ldots = \mathbf{0{,}035}$

$\mathbf{E(Y)} = -1 * 0{,}250 + 0 * 0{,}350 + 1 * 0{,}400 = \mathbf{0{,}150}$

$\mathbf{Var(Y)} = (-1 - 0{,}150)^2 * 0{,}250 + \ldots = \mathbf{0{,}628}$

Exemplo 3

X (juros)	Y (câmbio)			P(X=x)
	-1	0	1	
-0,25	0,050	0,070	0,260	0,380
0	0,075	0,210	0,120	0,405
0,25	0,125	0,070	0,020	0,215
P(Y=y)	0,250	0,350	0,400	1

a. **Qual é a probabilidade do câmbio variar positivamente?** $P(Y>0)=P(Y=1)=0{,}400$

b. **Sabendo que a taxa de juros caiu, qual é a probabilidade do câmbio variar positivamente?**

$P(Y>0|X<0)=P(Y=1|X=-0{,}25)=0{,}260/0{,}380=0{,}684$

c. **Qual a taxa média do câmbio quando a taxa de juros cai?** $E(Y|X<0)=-1*0{,}050/0{,}380+0*0{,}070/0{,}380+1*0{,}260/0{,}380=0{,}553$

Insper

Exemplo 3

X (juros)	Y (câmbio) -1	0	1	P(X=x)
-0,25	**0,050**	**0,070**	**0,260**	**0,380**
0	**0,075**	**0,210**	**0,120**	**0,405**
0,25	**0,125**	**0,070**	**0,020**	**0,215**
P(Y=y)	**0,250**	**0,350**	**0,400**	**1**

É possível, a partir do modelo acima, afirmar que a variação da taxa de câmbio (Y) <u>independe</u> da variação da taxa de juros (X)?

Quantifique a dependência entre essas variáveis.

Exemplo 3

X (juros)	Y (câmbio) -1	0	1	P(X=x)
-0,25	0,050	0,070	0,260	0,380
0	0,075	0,210	0,120	0,405
0,25	0,125	0,070	0,020	0,215
P(Y=y)	0,250	0,350	0,400	1

$E(X) = -0{,}041$

$Var(X) = 0{,}035$

$E(Y) = 0{,}150$

$Var(Y) = 0{,}628$

Com os resultados abaixo, o que podemos dizer sobre a relação dessas duas taxas? Explique.

$P(Y>0)=P(Y=1)=40\%$

$P(Y>0|X<0)=P(Y=1|X=-0{,}25)=68{,}4\%$

$Cov(X,Y)=-0{,}07256$ $Corr(X,Y)=-0{,}486$

Exemplo 3 (Continuação)

X (juros)	Y (câmbio)			P(X=x)
	-1	0	1	
-0,25	0,050	0,070	0,260	0,380
0	0,075	0,210	0,120	0,405
0,25	0,125	0,070	0,020	0,215
P(Y=y)	0,250	0,350	0,400	1

$E(X) = -0{,}041$

$Var(X) = 0{,}035$

$E(Y) = 0{,}150$

$Var(Y) = 0{,}628$

$Cov(X,Y) = -0{,}07256$

Um investidor aplica a mesma quantia num fundo que acompanha a variação da taxa de juros e num fundo que acompanha a variação cambial. Ao final do dia ele resgatará seu investimento.

Construa a distribuição de probabilidades do ganho (em variação %) desse investidor.

Calcule esperança e variância.

Insper

Propriedades da Esperança

Sejam a e b constantes e X e Y variáveis aleatórias:

a) $E(X+Y) = E(X) + E(Y)$

b) $E(aX+bY) = a\,E(X) + b\,E(Y)$

Se X e Y são **independentes** ou **dependentes**, os resultados acima são sempre válidos!!

Propriedades da Variância

Sejam X e Y variáveis aleatórias:

a) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

b) $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

sendo $Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$

Se X e Y são **independentes**, então

$$Var(X+Y) = Var(X) + Var(Y)$$

$$Var(X-Y) = Var(X) + Var(Y)$$

Propriedades da Variância

Sejam a e b constantes e X e Y variáveis aleatórias:

a) $Var(aX+bY) = a^2Var(X) + b^2Var(Y) + 2ab\, Cov(X,Y)$

b) $Var(aX-bY) = a^2Var(X) + b^2Var(Y) - 2abCov(X,Y)$

sendo $Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$

Se X e Y são **independentes**, então

$$Var(aX+bY) = a^2Var(X) + b^2Var(Y)$$

$$Var(aX-bY) = a^2Var(X) + b^2Var(Y)$$

Propriedades da Covariância

a. $Cov(X,Y) = E(XY) - E(X)E(Y)$

b. $Cov(aX,bY) = ab\, Cov(X,Y)$

c. Se X e Y forem **independentes**, então

$$Cov(X,Y)=0 \Rightarrow E(XY) = E(X)E(Y)$$

ATENÇÃO:

Há casos particulares em que X e Y são dependentes, mas Cov(X,Y)=0. Ver Exercício 4 desta aula.

Exemplo 4

Em uma determinada loja de roupas, o preço médio da calça jeans é de R$ 118,00, com um desvio-padrão associado a essa variável de R$ 22,00.

a) Defina a variável aleatória X em termos do problema.

b) Uma mãe de trigêmeas entra nessa loja e deseja comprar uma calça jeans para cada filha, mas para evitar brigas, comprará todas iguais. Qual o gasto total esperado dessa mãe e respectivo desvio padrão?

c) Uma outra mãe de três filhas entra nessa loja e deseja comprar uma calça jeans para cada filha. Como suas filhas não têm gostos iguais, fará escolhas independentes e não necessariamente iguais. Qual o gasto total esperado dessa outra mãe e respectivo desvio padrão?

DEIXE TODOS OS RESULTADOS TEÓRICOS EXPLICITAMENTE DEMONSTRADOS NA SUA RESOLUÇÃO.

Propriedades da Esperança e Variância

Sejam X, Y e Z variáveis aleatórias quaisquer, então

$$E(X+Y+Z) = E(X) + E(Y) + E(Z)$$

$$Var(X+Y+Z) = Var(X) + Var(Y) + Var(Z) + \\ + 2Cov(X,Y) + 2Cov(X,Z) + 2Cov(Y,Z)$$

Propriedades da Esperança e Variância

Sejam X_1, X_2, ..., X_k variáveis aleatórias **independentes**, então

$$E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k)$$

$$Var(X_1 + \cdots + X_k) = Var(X_1) + \cdots + Var(X_k)$$

Exercícios

Exercício 1

Dentre os clientes de uma seguradora, 25% são mulheres. Do histórico da empresa, sabe-se que 20% das mulheres acabam acionando o seguro para alguma indenização, enquanto que apenas 10% dos homens acionam o seguro para isso.

Para definir o preço do seguro a seguradora usa a informação de que a indenização de um acidente provocado por mulheres é da ordem de R$800,00, enquanto que a indenização de acidentes causados por homens é da ordem de R$2000,00. Obviamente se não ocorrer acidentes, não há indenização.

a) Qual é a distribuição de probabilidades do gasto com indenizações de uma apólice qualquer? Se há 10.000 pessoas na carteira, quanto se espera gastar com indenização considerando independência entre as pessoas? Com que desvio-padrão?

R: E(X) = 1,9 milhões e DP(X) = 54,4 mil

b) Admitindo que, em cada apólice, incida um custo fixo de R$50, qual deve ser o preço médio de uma apólice para que a seguradora garanta um lucro médio de R$100 por apólice?

R: 340 reais

Exercício 2

Dados:

- **X_1 e X_2: retornos de dois ativos**
- **$0 \leq a \leq 1$**
- **$R = aX_1 + (1-a)X_2$: retorno de uma carteira**
- **$Var(X_1) = Var(X_2) = \sigma^2 > 0$**
- **$Cov(X_1, X_2) = \sigma_{12} < 0$**

Qual o valor de a que minimiza o risco da carteira? R: 1/2

Exercício 3

Uma companhia de seguro tem informações sobre deduções em seguros de casa e de carro. A tabela a seguir apresenta as informações, incompletas, desta seguradora em que

X representa as deduções com seguros de carro e

Y as deduções com seguro de casa

Exercício 3 (cont.)

X	Y			P(X=x)
	R$0,00	**R$100,00**	**R$200,00**	
R$100,00	0,20		0,20	
R$250,00		0,15		**0,50**
P(Y=y)	**0,25**			

(i) Complete a tabela

(ii) As deduções são independentes? Porque? R: Não

(iii) Obtenha a correlação entre as deduções e comente. R: 0,3015

(iv) Dado que as deduções do seguro dos carros é de R$250,00, qual a distribuição das deduções de seguro de casa. Neste caso, qual é a dedução média do seguro de casa? R: R$150

Exercício 4

A variável aleatória (v.a.) X assume valores x=0,1 e 2.
A v.a. Y assume os valores y=1, 2 e 3.
A distribuição conjunta dessas duas v.a.'s se dá conforme a tabela abaixo.

Y	X			P(Y=y)
	0	1	2	
1	3/20	3/20	2/20	8/20
2	1/20	1/20	2/20	4/20
3	4/20	1/20	3/20	8/20
P(X=x)	8/20	5/20	7/20	1

Calcule a covariância entre X e Y, ou seja, Cov(X,Y). R: 0
As v.a.'s X e Y são independentes? R: Não
O que você pode concluir com isso?

Exercício 5

Um processo industrial pode ser executado em duas etapas independentes. O tempo gasto em cada etapa segue uma distribuição com média de 5 horas e desvio-padrão de 10 horas.

Um engenheiro resolveu modificar o processo de produção fazendo com que seja executado numa única fase, cujo tempo de execução é o dobro da primeira etapa do processo atual.

Vale a pena adotar o processo proposto pelo engenheiro?

R: Atual: E(T) = 10 e DP(T) = 14,14
Novo: E(T) = 10 e DP(T) = 20