Ciência dos Dados

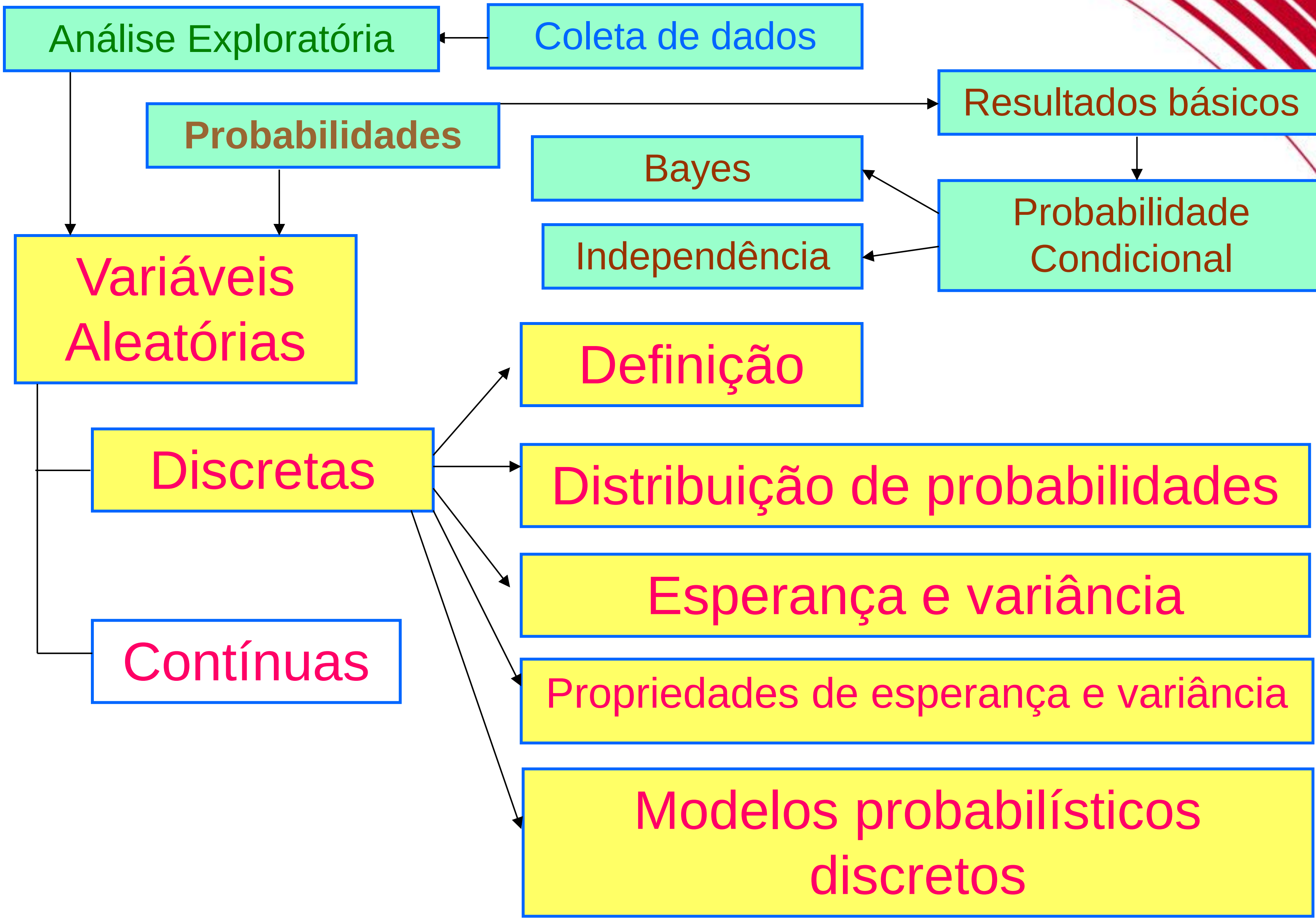
Aula 09

Variáveis aleatórias discretas

Distribuição de probabilidades

Esperança e variância

Magalhães e Lima, 7ª. Edição. Seção 3.1 e Definição 4.2 (pág. 110) e Definição 4.5 (pág. 121).


Análise Exploratória
Coleta de dados
Probabilidades
Resultados básicos
Bayes
Independência
Probabilidade Condicional
Variáveis Aleatórias
Definição
Discretas
Distribuição de probabilidades
Esperança e variância
Contínuas
Propriedades de esperança e variância
Modelos probabilísticos discretos

Objetivos de Aprendizagem

Os alunos devem ser capazes de:

- Descrever e aplicar as propriedades de distribuições probabilísticas em variáveis aleatórias discretas.

Variáveis aleatórias

Quando **tomamos decisões** em face da incerteza, **raramente elas se baseiam apenas** na probabilidade descritas em **eventos**.

Na maioria dos casos, devemos **também** saber algo sobre as consequências potenciais da **tomada de decisão descritas em variáveis quantitativas** (tempo de execução de uma tarefa, perdas, lucros, penalidades ou recompensas).

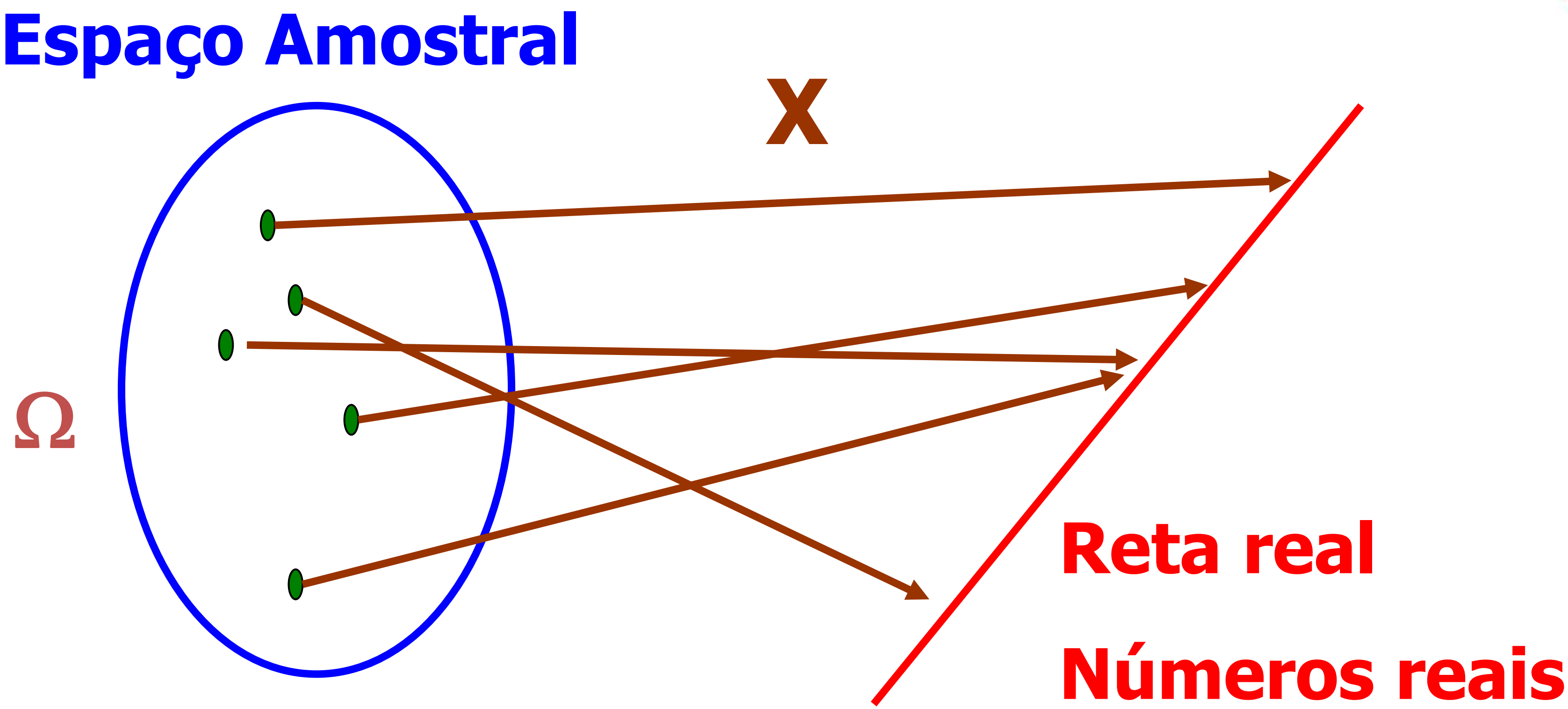
Variáveis aleatórias

Por exemplo, uma construtora precisa decidir se apresenta proposta para um projeto que lhe oferece a perspectiva de R$ 250.000,00 de lucro com probabilidade de 20% ou de um prejuízo de R$ 50.000,00 (em consequência de uma crise financeira no país) com 80% de probabilidade.

A probabilidade da construtora ter lucro não é muito grande, mas a quantia que ela pode ganhar é muito maior do que a que ela pode perder.

Este exemplo mostra a necessidade de um método que permita combinar probabilidades e consequências.

Variáveis aleatórias



Variável aleatória: função que associa um número real a cada ponto do espaço amostral (possível realização do experimento aleatório).

Tipos de Variáveis Aleatórias

- Variáveis Aleatórias Discretas
- Variáveis Aleatórias Contínuas

Variável aleatória discreta: conjunto de possibilidades é um conjunto finito ou enumerável

Exemplo: número de filhos, número de salas de aula, número de benefícios do Bolsa Família por família

Variável aleatória contínua: conjunto de possibilidades num intervalo ou conjuntos de intervalos contínuos

Exemplo: tempo de duração, temperatura

Variáveis Aleatórias Discretas

- ☆ Uma indústria de aviões recebe pedidos de um determinado tipo de jato comercial por ano (X: 0, 1, 2, 3, 4, ...)
- ☆ Um empresário com 20 escritórios comerciais quer saber o número de escritórios que estão alugados por mês (X: 0, 1, 2, 3, ..., 20)
- ☆ Número de vendas num dia de funcionamento de uma loja (X: 0, 1, 2, ...)
- ☆ Número de chamadas telefônicas recebidas numa central num dia (X: 0,1, 2, ...)
- ☆ Um vendedor de seguros aborda 5 clientes por dia, recebendo 50,00 de comissão a cada venda. A variável de interesse é o ganho diário do vendedor (X: 0,00; 50,00; 100,00; 150,00; 200,00; 250,00)
- ☆ Número de peças defeituosas num lote com 30 peças (X: 0, 1,..., 30)
- ☆ Número de papéis que fecharam em alta ao final de um pregão (X: 0, 1, ..., n)

Exemplo 1

Uma corretora de seguros paga uma comissão de R$50,00 a cada novo seguro que um corretor vende. A probabilidade de um cliente adquirir o seguro é de 0,20.

a) Descreva como pode se comportar a comissão se um corretor ao abordar 2 clientes de maneira independente um do outro.

b) Qual a probabilidade de um corretor ganhar apenas R$50,00?

S_i: o cliente *i* compra o seguro

N_i: o cliente *i* não compra o seguro

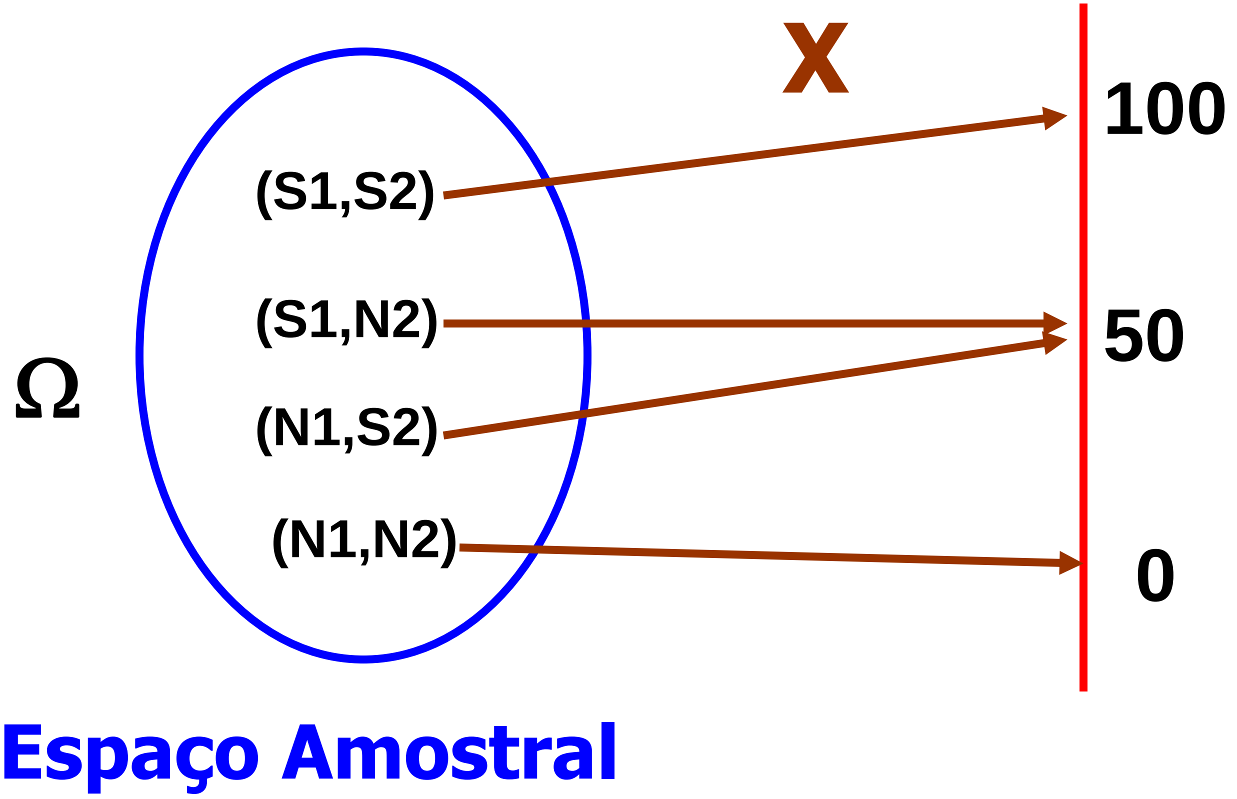
Exemplo 1 (item a)

Espaço Amostral		Comissão
1o cliente	2o cliente	
S_1	S_2	100
S_1	N_2	50
N_1	S_2	50
N_1	N_2	0

S_i: o cliente *i* compra o seguro
N_i: o cliente *i* não compra o seguro

Exemplo 1 (item a)

X: comissão recebida por um corretor com a tentativa de venda de duas apólices de seguro



Distribuição de probabilidades de uma variável aleatória

Exemplo 1 (item b)

Uma corretora de seguros paga uma comissão de R$50,00 a cada novo seguro que um corretor vende.

A probabilidade de um cliente adquirir o seguro é de 0,20.

a) Descreva como pode se comportar a comissão se um corretor abordar 2 clientes de maneira independente um do outro.

b) Qual a probabilidade de um corretor ganhar apenas R$50,00?

S_i: o cliente *i* compra o seguro;

N_i: o cliente *i* não compra o seguro.

$$P(S_i) = 0{,}20 \qquad\qquad P(N_i) = 0{,}80$$

Exemplo 1 (item b)

Abordando dois clientes de forma aleatória, os eventos possíveis são:

Espaço amostral			
1º Cliente	2º Cliente	Comissão	P(X=x)
S_1	S_2	100	0,04
S_1	N_2	50	0,16
N_1	S_2	50	0,16
N_1	N_2	0	0,64

$$P(X = 50) =$$
$$0{,}16 + 0{,}16 = 0{,}32$$

A comissão que um corretor ganha é uma variável aleatória discreta.

Distribuição de probabilidades de uma v.a. discreta

É uma função que associa uma probabilidade a cada valor de uma variável aleatória.

Exemplo: **X: Comissão de um corretor**

Distribuição de probabilidades de X

x	P(X=x)
0	0,64
50	0,32
100	0,04
Soma	**1,00**

Definição:

(Função de Probabilidade – f.p.) :

Seja uma variável aleatória (v.a.) discreta X, que assume valores *x*´s.

A função que associa a probabilidade de ocorrência em cada valor x, isto é,

$$f(x) = P(X = x) ,$$

chama-se **função de probabilidade**.

Definição: (Distribuição de uma v.a.) –

Seja uma **variável aleatória** (v.a.) **discreta** X, que assume valores *x*´s.

A **distribuição da v.a. X** (ou distribuição de probabilidades da v.a. X) é o conjunto de todos os pares formados por

$$\{x, P(X = x)\},$$

isto é, pelos valores de X e as respectivas probabilidades da variável assumir tais valores.

Propriedades da Função de Probabilidades de uma Variável Aleatória Discreta

Seja X uma v.a. discreta assumindo valores x´s:

$$0 \leq P(X = x) \leq 1$$

$$\sum_{x} P(X = x) = 1$$

Média e Variância de uma variável aleatória

Voltando ao Exemplo 1

Distribuição de probabilidades de X

(X: Comissão)

x	P(X=x)
0	0,64
50	0,32
100	0,04
Soma	**1,00**

Qual a comissão média recebida por um corretor ao abordar 2 clientes?

E o desvio padrão dessa comissão?

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,

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

Voltando ao Exemplo 1

Distribuição de probabilidades de X

x	P(X=x)
0	0,64
50	0,32
100	0,04
Total	1,00

E(X) = 0x0,64 + 50x0,32 + 100x0,04 = 20 reais

(valor médio da comissão do corretor)

Voltando ao Exemplo 1

Distribuição de probabilidades de X

(X: Comissão)

x	P(X=x)
0	0,64
50	0,32
100	0,04
Soma	**1,00**

Qual a comissão média recebida por um corretor ao abordar 2 clientes?

E o desvio padrão dessa comissão?

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 $\mathrm{E(X^2)} = \sum_{x} x^2\, \mathrm{P(X} = x)$

Voltando ao Exemplo 1

Distribuição de probabilidades de X

x	P(X=x)
0	0,64
50	0,32
100	0,04
Total	1,00

E(X) = 20 reais

Var(X) = $(0-20)^2 \times 0{,}64 + (50-20)^2 \times 0{,}32 + (100-20)^2 \times 0{,}04$
= 800 reais^2

σ_X = 28,28 reais (desvio padrão da variável aleatória X)

3º. Hands On para próxima aula

Distribuição de probabilidades de X: Comissão atual de um corretor

x	P(X=x)
0	0,64
50	0,32
100	0,04
Total	1,00

Imagine que a corretora de seguros irá fornecer um aumento na comissão dos corretores. Entretanto, cada corretor poderá escolher uma das seguintes opções:

1. Nova comissão será a comissão atual mais um fixo de R$ 20,00.
2. Nova comissão será o dobro da atual comissão

Calcule o valor esperado e a variância de cada opção.

Escolha qual delas é melhor para aumentar o ganho de um corretor. Justifique sua resposta.

3º. Hands On para próxima aula

Distribuição de probabilidades de X: Comissão atual de um corretor

x	P(X=x)
0	0,64
50	0,32
100	0,04
Total	1,00

Distribuição de probabilidades Y: Comissão atual mais um fixo de R$ 20,00.

y	P(Y=y)
20	0,64
70	0,32
120	0,04
Total	1,00

Distribuição de probabilidades W: Dobro da atual comissão

w	P(W=w)
0	0,64
100	0,32
200	0,04
Total	1,00

Propriedades da Esperança

Seja **X uma variável aleatória qualquer**, então

(i) $\mathbf{E}(\mathbf{X} + \boldsymbol{d}) = \mathbf{E}(\mathbf{X}) + \boldsymbol{d}$**, onde** $\boldsymbol{d}$ **é uma constante.**

(ii) $\mathbf{E}(\boldsymbol{c}\,\mathbf{X}) = \boldsymbol{c}\,\mathbf{E}(\mathbf{X})$**, onde** $\boldsymbol{c}$ **é uma constante.**

(iii) Combinando (i) e (ii):

$$\mathbf{E}(\boldsymbol{c}\,\mathbf{X} + \boldsymbol{d}) = \boldsymbol{c}\,\mathbf{E}(\mathbf{X}) + \boldsymbol{d},$$

onde $\boldsymbol{c}$ **e** $\boldsymbol{d}$ **são constantes.**

Propriedades da Variância

Seja **<u>X uma variável aleatória qualquer</u>**, então

(i) $\mathbf{Var}(\mathbf{X} + \boldsymbol{d}) = \mathbf{Var}(\mathbf{X})$**, onde** $\boldsymbol{d}$ **é uma constante.**

(ii) $\mathbf{Var}(\boldsymbol{c}\,\mathbf{X}) = \boldsymbol{c}^{2}\,\mathbf{Var}(\mathbf{X})$**, onde** $\boldsymbol{c}$ **é uma constante.**

(iii) Combinando (i) e (ii):

$$\mathbf{Var}(\boldsymbol{c}\,\mathbf{X} + \boldsymbol{d}) = \boldsymbol{c}^{2}\,\mathbf{Var}(\mathbf{X}),$$

onde $\boldsymbol{c}$ **e** $\boldsymbol{d}$ **são constantes.**

Exemplo 2

Uma empresa de segurança visita, em um dia de trabalho, dois potenciais clientes para oferecer seus serviços.

A probabilidade de fechar contrato com o primeiro cliente visitado no dia é da ordem de 10%. Quando a primeira visita resulta em contrato, a probabilidade de se fechar contrato na segunda visita quadruplica, caso contrário, ela se mantém em 10%.

Admitindo que o **custo do dia de trabalho** seja da ordem de **R$30,00** e que a receita obtida com **cada contrato fechado** seja da ordem de **R$500,00**, pergunta-se:

a) Encontre a distribuição de probabilidades da variável Lucro.
b) Qual a probabilidade de se ter prejuízo num dia?
c) Vale a pena manter essas visitas?
d) Qual a variância do lucro?
e) Qual valor esperado e variância do Lucro, se esse cair em 10%?

Exemplo 2 (item a)

Espaço Amostral		Probabilidade
1ª visita	2ª visita	
F_1	F_2	0,04
F_1	N_2	0,06
N_1	F_2	0,09
N_1	N_2	0,81

F_i: fechar contrato na visita *i*
N_i ($=F_i^C$): não fechar contrato na visita i

Espaço Amostral		Receita	Custo Fixo	Lucro
1a visita	2a visita			
F_1	F_2	1000	30	970
F_1	N_2	500	30	470
N_1	F_2	500	30	470
N_1	N_2	0	30	-30

Exemplo 2 (item a)

Defina a v.a. *X* como sendo o lucro obtido durante o período de interesse. Assim,

Espaço Amostral		x	P(X=x)
1a visita	2a visita		
F_1	F_2	970	0,04
F_1	N_2	470	0,06
N_1	F_2	470	0,09
N_1	N_2	-30	0,81

Exemplo 2 (item a)

Finalmente, a distribuição de probabilidades da v.a. *X*, definida anteriormente, fica dada por

x	P(X=x)
970	0,04
470	0,06+0,09=0,15
-30	0,81
Total	1,00

Exemplo 2

b) Qual a probabilidade de se ter prejuízo num dia?

$P(X < 0) = P(X = -30) = 0{,}81$

c) Vale a pena manter essas visitas?

$E(X) = 970 \times 0{,}04 + 470 \times 0{,}15 + (-30) \times 0{,}81 = 85$

d) Qual a variância do lucro?

$Var(X) = (970-85)^2 x 0{,}04 + (470-85)^2 x 0{,}15 + (-30-85)^2 x 0{,}81$

$= 64.275$ reais^2

$DP(X) = 253{,}53$ reais

e) Esperança e variância se lucro cair em 10%?

$Y=0{,}9X \rightarrow E(Y) = 0{,}9 \times 85 = 76{,}50$ reais

$Var(Y) = 0{,}9^2 \times 64.275 = 52.062{,}75$ reais^2

$DP(Y) = 228{,}17$ reais

Exemplo 3

Historicamente, as vendas mensais de uma loja têm média de $25.000 e desvio-padrão $4.000.

Os lucros correspondem a 30% das vendas menos um custo fixo de $6.000.

Encontre o valor esperado e o desvio-padrão do lucro.

Exercícios

Exercício 1

Um rapaz está pensando em convidar sua namorada para sair. O problema é que as despesas correm por sua conta. Eles podem ir ao cinema ou ao teatro. 70% das vezes ela prefere ir ao cinema, nesse caso, ele gasta $70,00 com os ingressos. Quando eles vão ao teatro, o gasto fica em $190,00. Se eles forem ao cinema, ele sabe que em 80% das vezes ela pede para ir jantar, a despesa adicional do jantar fica em $130,00; 20% das vezes, eles vão direto para casa. Levando a namorada ao teatro, em 40% das vezes ela pede para ir jantar e 60% das vezes eles vão direto para casa.

Qual a distribuição de probabilidades do gasto que o rapaz tem com a namorada? Qual o gasto médio? E o seu desvio-padrão?

R: E(G) = 194,40 reais e DP(G) = 63,88 reais

Com a inflação deste ano, o gasto aumentou até agora $9, mas com a crise geral, o casal resolveu reduzir esse novo gasto total em 15%. Calcule o novo gasto médio e respectivo desvio padrão. R: E(Y) = 172,89 reais e DP(Y) = 54,30 reais

Exercício 2

A AirBrazil tem uma política de, rotineiramente, superlotar os vôos, porque, por experiências passadas, alguns passageiros não comparecem para o embarque. A variável aleatória X representa o número de passageiros que não podem ser embarcados por haver mais passageiros do que assentos.

x	0	1	2	3	4
P(X=x)	0,807	0,115	0,057	0,019	0,002

Considerando a distribuição de probabilidades da v.a. X, calcule:

a) Calcule a E(X) e a Var(X). Resp: E(X) = 0,294 e Var(X) = 0,4596

b) A cada passageiro não embarcado, a companhia aérea paga uma multa de R$500,00. Qual a despesa média com multas da AirBrazil? E a variância?

Resp: Y=500X; E(Y)=147; Var(Y)=114900

c) A Infraero pretende alterar a taxação da multa, dobrando o valor pago por passageiro não embarcado. Neste caso, qual seria o valor médio gasto com multas? E a variância? Resp: Z=2Y; E(Z)=294; Var(Z)=459600

d) A Infraero pretende alterar a taxação da multa, dobrando o valor pago por passageiro não embarcado no item b) e cobrando um valor fixo de R$2000, sempre que houver overbooking. Neste caso, qual seria o valor médio gasto com multas? E a variância?

Resp: E(W)=680; Var(W)=2.031.600

Exercício 3

A) O tempo *T*, em minutos, necessário para um operário processar certa peça, é uma v.a. com a seguinte distribuição de probabilidades:

t	2	3	4	5	6	Total
P (T=t)	0,1	0,1	0,4	0,2	0,2	1,0

O gerente de produção necessita estudar o tempo de produção das peças para decidir se contrata mais funcionários ou se pode dar um bônus em dinheiro para que os funcionários sejam mais rápidos. Qual o tempo médio de processamento e o desvio padrão do tempo gasto para processar uma peça?

R: E(T) = 4,3 minutos e DP(T) = 1,19 minutos

B) Para cada peça processada, o operário ganha um fixo de 2,00 u.m. (unidade monetária), mas se ele processa a peça em menos de 5 minutos, ganha 0,50 u.m. por cada minuto poupado. Por exemplo, se ele processa a peça em 4 minutos, recebe a quantia adicional de 0,50 u.m.

Qual a média e o desvio padrão da quantia (em u.m.) ganha por peça?

R: E(G) = 2,45 u.m.
DP(G) = 0,47 u.m.

t	2	3	4	5	6	Total
P (T=t)	0,1	0,1	0,4	0,2	0,2	1,0
Ganho (u.m.)	3,50	3,00	2,50	2,00	2,00	

Exercício 4

Um funcionário de uma corretora ganha um bônus de 30 u.m. a cada investimento bem sucedido e é penalizado em 15 u.m. a cada investimento mal sucedido.

Admita que ele tenha que propor 2 investimentos por dia. Sabe-se que se ele tem sucesso no 1o. investimento, a probabilidade de sucesso no 2o. é 0,80; caso ele fracasse no 1o., a probabilidade de sucesso no 2o. é 0,50.

Sabendo que a probabilidade de ser bem sucedido no 1o. é 0,60, determine a distribuição de probabilidades do seu ganho diário.

A. Qual é o ganho médio diário? E o seu desvio-padrão?

B. O funcionário comunica ao seu chefe que se demitirá caso seu ganho médio diário seja inferior a R$32,00. Qual deveria ser a penalização máxima para que o funcionário não se demita?

A: E(X) = 27,6 u.m. e DP(X) = 34,90 u.m. [Var(X) = 1218,24 u.m.^{2}]
B: 8,88