Aula de hoje

Aula sobre:

Modelos probabilísticos Discretos:

Distribuição Binomial

1º: uso do notebook Aula11_Exemplo

2º: uso do notebook Aula11_Atividade

3º: uso do notebook Aula11_Exercício

Insper

Ciência dos dados

Modelos Probabilísticos Discretos

Bernoulli e Binomial

Aula de hoje

Ao final desta aula, o aluno deve ser capaz de:

- Especificar as distribuições de probabilidades adequadas para variáveis aleatórias discretas considerando modelos probabilísticos discretos já bem definidos na literatura estatística.

Modelos probabilísticos

Modelagem probabilística de fenômenos

aleatórios que envolvem variáveis quantitativas

e que seguem padrões comuns.

Distribuição de Bernoulli

Contexto geral

Experimento que tem apenas dois resultados possíveis.

Por convenção denomina-se o resultado de interesse como ***sucesso*** e o outro como ***fracasso***.

Alguns exemplos

O que há em comum nos seguintes eventos:

- oferecer uma apólice de seguro em um telefonema;
- observar a ocorrência de sinistro em um carro segurado por uma companhia de seguros;
- perguntar a um telespectador se ele se lembra de um anúncio comercial exibido num determinado horário;
- verificar se um ativo se valorizou num determinado dia.

Como definir uma v.a. para modelar esses fenômenos?

Ensaio de Bernoulli

Experimento que tem apenas dois resultados possíveis. Por convenção denomina-se o resultado de interesse como ***sucesso*** e o outro como ***fracasso***.

Jacob Bernoulli - 1654-1705 - Suiça

http://www-groups.dcs.st-and.ac.uk/~history/Biographies/Bernoulli_Jacob.html

Ensaio de Bernoulli

Espaço amostral de um ensaio de Bernouli:

$$\Omega = \{Sucesso;\ Fracasso\}$$

Variável aleatória de Bernoulli:

$$X = \begin{cases} 0, \text{se fracassso} \\ 1, \text{se sucesso} \end{cases}$$

Distribuição de Bernoulli

Admitindo que p é a probabilidade de sucesso, temos:

$$P(X=1) = p \quad e \quad P(X=0) = 1 - p; \; ou$$

p é o parâmetro da distribuição que, uma vez conhecido, pode ser calcular qualquer probabilidade.

Distribuição de Bernoulli

Admitindo que p é a probabilidade de sucesso, temos:

$$P(X = 1) = p \quad e \quad P(X = 0) = 1 - p; \; ou$$

p é o parâmetro da distribuição que, uma vez conhecido, pode ser calcular qualquer probabilidade.

Notação: $X \sim Bern(p)$

Determine $E(X)$ e $Var(X)$.

$$E(X) = p \quad \text{e} \quad Var(X) = p(1 - p)$$

Distribuição de Bernoulli

Admitindo que p é a probabilidade de sucesso, temos:

$$P(X = 1) = p \quad e \quad P(X = 0) = 1 - p; \ ou$$

$$\boxed{P(X = x) = p^x (1 - p)^{1-x}; \ x = 0 \text{ ou } x = 1}$$

p é o parâmetro da distribuição que, uma vez conhecido, pode ser calcular qualquer probabilidade.

Notação: $X \sim Bern(p)$

Determine $E(X)$ e $Var(X)$.

$$E(X) = p \quad \text{e} \quad Var(X) = p(1 - p)$$

Distribuição Binomial

Contexto geral

Imagine, agora, que repetimos um ensaio de Bernoulli n vezes.

Suponha ainda que as repetições sejam independentes.

Uma amostra particular será constituída de uma sequência de sucessos e fracassos.

Intuito será contabilizar a quantidade de sucessos nessas n tentativas.

Alguns exemplos

Em muitas situações práticas, estamos interessados na probabilidade de **um evento ocorrer y vezes** em **n repetições do experimento**, por exemplo:

- a probabilidade de **vender 50 seguros** em **200 telefonemas**;
- a probabilidade de que **15 carros** de uma determinada marca sejam **roubados** na cidade, **dentre os 40 carros** desta marca segurados por uma companhia;
- a probabilidade de **100** em **300** telespectadores **entrevistados lembrarem quais produtos foram anunciados** em determinado programa;
- a probabilidade de **uma ação subir em 10** dos **21 dias avaliados**.

Experimento Binomial

- é uma sequência de ***n*** **repetições** (ou tentativas ou ensaios) idênticas;
- cada repetição tem apenas 2 resultados possíveis: um é denominado **sucesso** e o outro, **fracasso**;
- a **probabilidade de sucesso** para cada ensaio é denominada ***p*** e será constante em cada repetição. Então, a **probabilidade de fracasso** (*1-p*) também não varia de tentativa para tentativa;
- As tentativas são independentes.

Exemplo 1

Um resistor de 100K Ohms comprado na Santa Efigênia tem probabilidade de falha de 20%, segundo um fabricante bem ruim. Ainda, esses resistores podem falhar de forma independente uns dos outros.

Um aluno de engenharia compra um pacote com 3 desses resistores.

Responda:

a) Encontre a distribuição de probabilidades da variável que conta número de resistores com falha nesse pacote contendo 3 resistores. Dica: Use árvore de probabilidades para essa construção.

b) Qual a probabilidade de exatamente dois falharem?

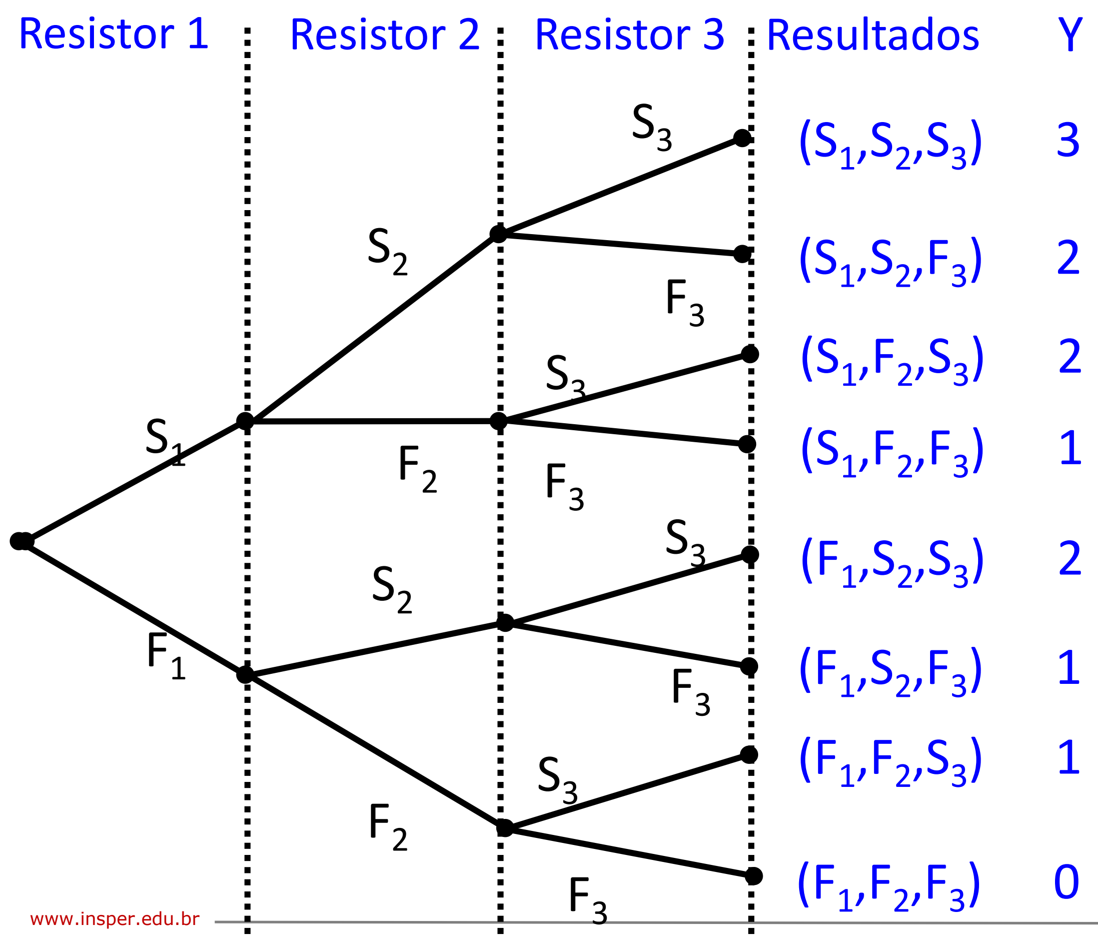
c) E se for um pacote com 100 resistores, qual a probabilidade de 20 falharem?

Exemplo 1

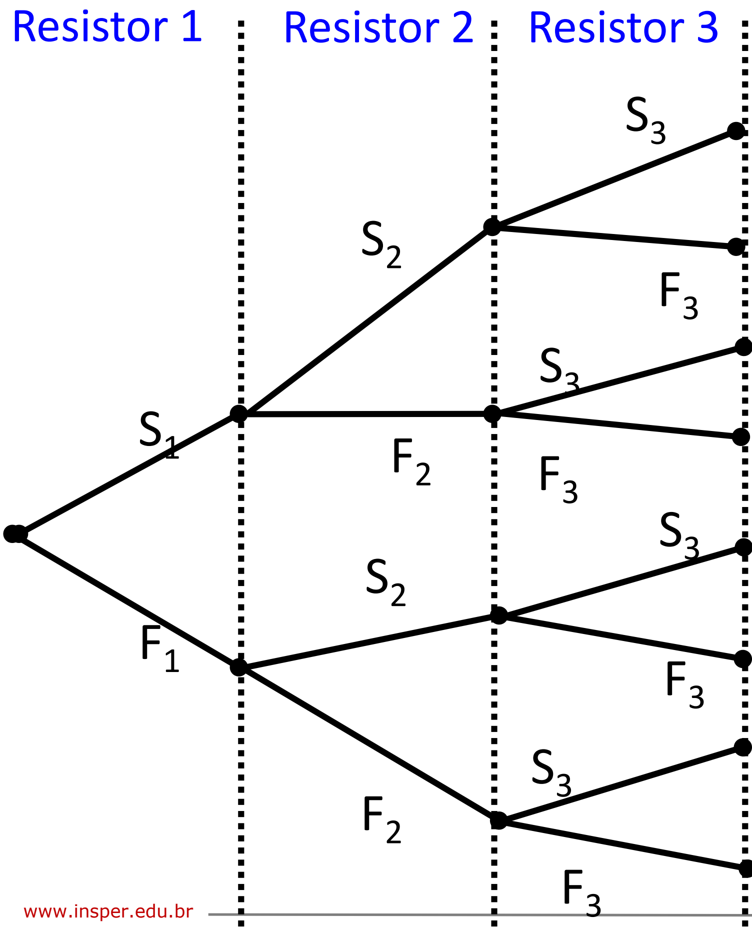
Tal experimento possui a distribuição binomial?

- Experimento consiste de 3 sorteios idênticos ➔ n = 3.
- Dois resultados possíveis em cada tentativa: falha (**sucesso**) ou não (**fracasso**).
- A probabilidade de falha é a mesma em cada tentativa ➔ p = 0,2
- As tentativas (resistores) são independentes.

Árvore de probabilidades (p=0,20)



Árvore de probabilidades (p=0,20)

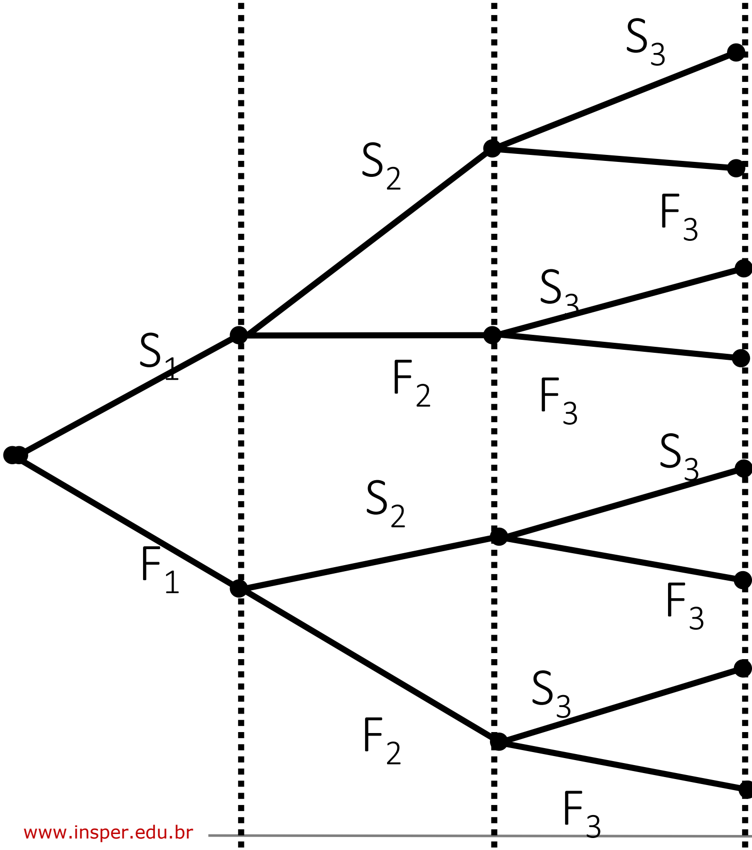


Resultados	Y
(S_1,S_2,S_3)	3
(S_1,S_2,F_3)	2
(S_1,F_2,S_3)	2
(S_1,F_2,F_3)	1
(F_1,S_2,S_3)	2
(F_1,S_2,F_3)	1
(F_1,F_2,S_3)	1
(F_1,F_2,F_3)	0

Probabilidades:

Árvore de probabilidades (p=0,20)

Resistor 1 | Resistor 2 | Resistor 3

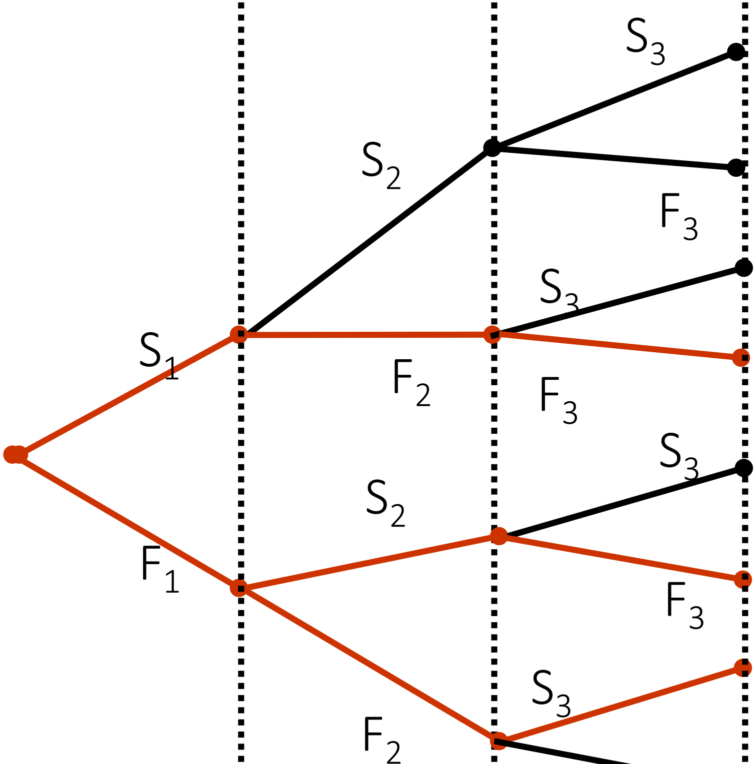


Resultados	Y	Probabilidades:
(S_1,S_2,S_3)	3	
(S_1,S_2,F_3)	2	
(S_1,F_2,S_3)	2	
(S_1,F_2,F_3)	1	
(F_1,S_2,S_3)	2	
(F_1,S_2,F_3)	1	
(F_1,F_2,S_3)	1	
(F_1,F_2,F_3)	0	

Árvore de probabilidades (p=0,20)

Quando Y=1

Resistor 1 | Resistor 2 | Resistor 3



Resultados	Y	Probabilidades:
(S_1,S_2,S_3)		
(S_1,S_2,F_3)		
(S_1,F_2,S_3)		
(S_1,F_2,F_3)	1	$p \times (1-p) \times (1-p)$
(F_1,S_2,S_3)		
(F_1,S_2,F_3)	1	$(1-p) \times p \times (1-p)$
(F_1,F_2,S_3)	1	$(1-p) \times (1-p) \times p$

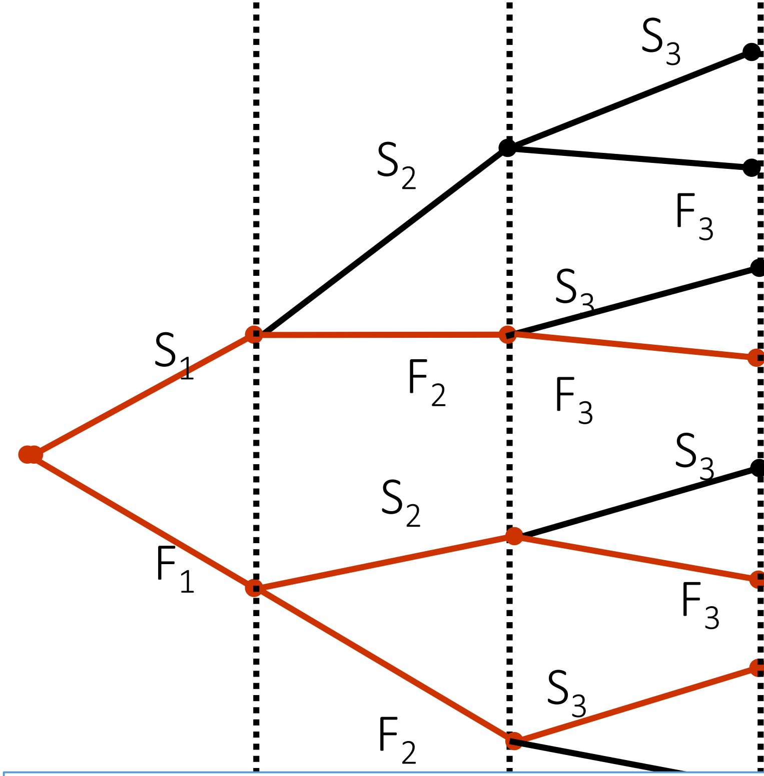
Independência

$$P(Y=1) = P(S_1F_2F_3) + P(F_1S_2F_3) + P(F_1F_2S_3) =$$

Árvore de probabilidades (p=0,20)

Quando Y=1

Resistor 1 | Resistor 2 | Resistor 3



Resultados	Y	Probabilidades:
(S_1,S_2,S_3)		
(S_1,S_2,F_3)		
(S_1,F_2,S_3)		
(S_1,F_2,F_3)	1	$p \times (1-p) \times (1-p)$
(F_1,S_2,S_3)		
(F_1,S_2,F_3)	1	$(1-p) \times p \times (1-p)$
(F_1,F_2,S_3)	1	$(1-p) \times (1-p) \times p$

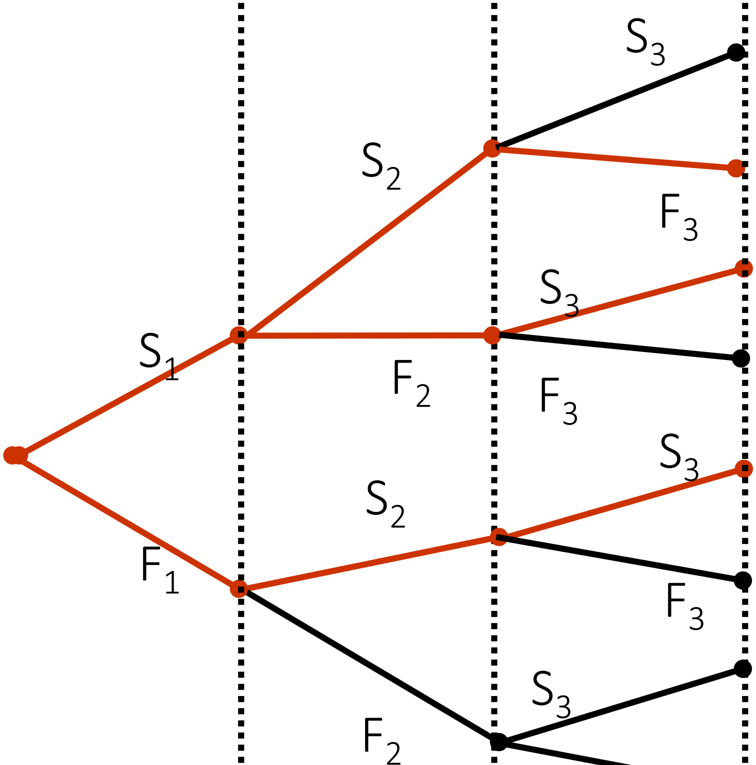
Independência

$$P(Y=1) = P(S_1F_2F_3) + P(F_1S_2F_3) + P(F_1F_2S_3) =$$
$$= 0{,}2\ 0{,}8\ 0{,}8 + 0{,}8\ 0{,}2\ 0{,}8 + 0{,}8\ 0{,}8\ 0{,}2 = 3\ 0{,}2\ 0{,}8^2 = 0{,}348$$

Árvore de probabilidades (p=0,20)

Quando Y=2

Resistor 1 | Resistor 2 | Resistor 3



Resultados	Y	Probabilidades:
(S_1,S_2,S_3)		
(S_1,S_2,F_3)	2	$p \times p \times (1-p)$
(S_1,F_2,S_3)	2	$p \times (1-p) \times p$
(S_1,F_2,F_3)		
(F_1,S_2,S_3)	2	$(1-p) \times p \times p$
(F_1,S_2,F_3)		
(F_1,F_2,S_3)		

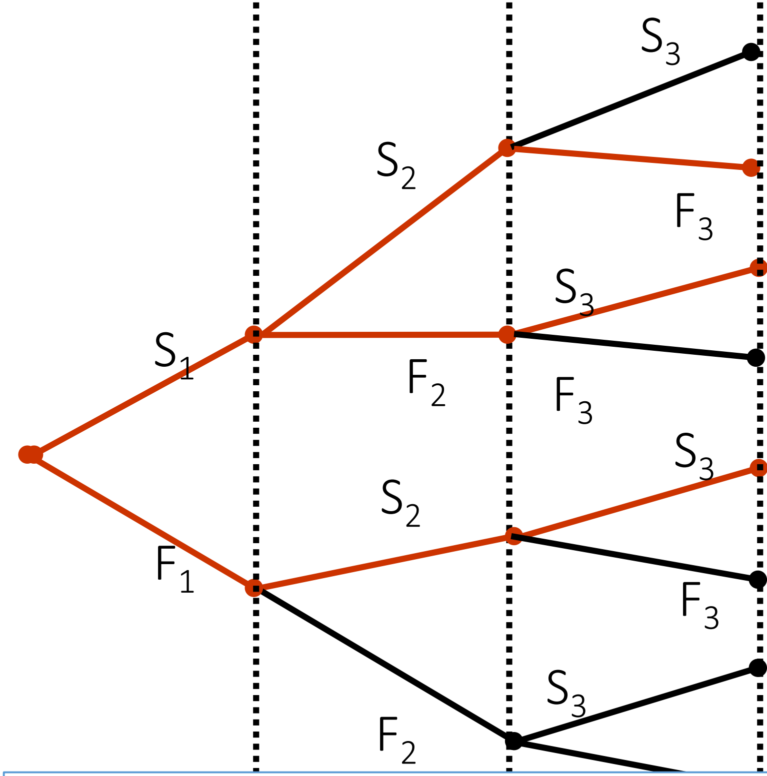
Independência

$$P(Y=2) = P(S_1S_2F_3) + P(S_1F_2S_3) + P(F_1S_2S_3) =$$

Árvore de probabilidades (p=0,20)

Quando Y=2

Resistor 1 Resistor 2 Resistor 3

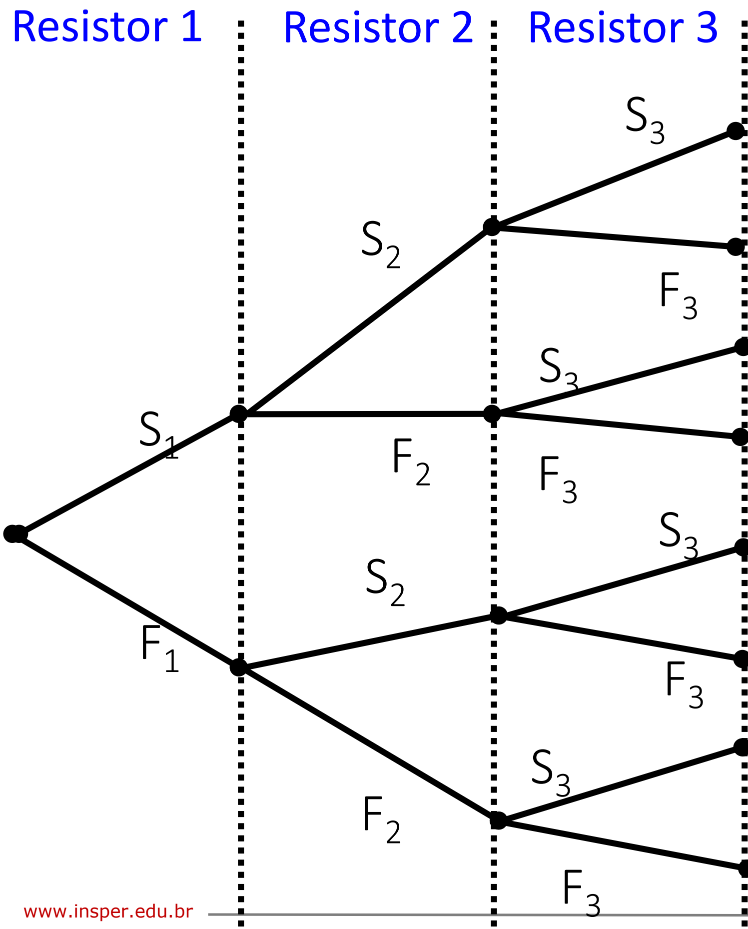


Resultados	Y	Probabilidades:
(S_1,S_2,S_3)		
(S_1,S_2,F_3)	2	$p \times p \times (1-p)$
(S_1,F_2,S_3)	2	$p \times (1-p) \times p$
(S_1,F_2,F_3)		
(F_1,S_2,S_3)	2	$(1-p) \times p \times p$
(F_1,S_2,F_3)		
(F_1,F_2,S_3)		

Independência

$$P(Y=2) = P(S_1S_2F_3) + P(S_1F_2S_3) + P(F_1S_2S_3) =$$
$$= 0{,}2\ 0{,}2\ 0{,}8 + 0{,}2\ 0{,}8\ 0{,}2 + 0{,}8\ 0{,}2\ 0{,}2 = 3\ 0{,}2^2 0{,}8$$

Árvore de probabilidades (p=0,20)



Resultados	Y	<u>Probabilidades:</u>
(S_1,S_2,S_3)	3	$p \times p \times p$
(S_1,S_2,F_3)	2	$p \times p \times (1-p)$
(S_1,F_2,S_3)	2	$p \times (1-p) \times p$
(S_1,F_2,F_3)	1	$p \times (1-p) \times (1-p)$
(F_1,S_2,S_3)	2	$(1-p) \times p \times p$
(F_1,S_2,F_3)	1	$(1-p) \times p \times (1-p)$
(F_1,F_2,S_3)	1	$(1-p) \times (1-p) \times p$
(F_1,F_2,F_3)	0	$(1-p) \times (1-p) \times (1-p)$

Probabilidade para sequências com 2 sucessos

$p = 0{,}20 \rightarrow$ probabilidade de um resistor falhar

Resultado	Probabilidade	
$S_1S_2F_3$	$pp(1-p)$	$p^2(1-p)$
$S_1F_2S_3$	$p(1-p)p$	$p^2(1-p)$
$F_1S_2S_3$	$(1-p)pp$	$p^2(1-p)$

$$P(Y = 2) = 3 \cdot p^2\,(1-p)$$

Probabilidade para sequências com 1 sucesso

$p = 0,20$ ➔ probabilidade de um resistor falhar

Resultado	Probabilidade	
$S_1F_2F_3$	$p(1-p)(1-p)$	$p(1-p)^2$
$F_1S_2S_3$	$(1-p)p(1-p)$	$p(1-p)^2$
$F_1F_2S_3$	$(1-p)\,(1-p)p$	$p(1-p)^2$

$$P(Y = 1) = 3 \cdot p\,(1-p)^2$$

Probabilidades para sequências com 0 e 3 sucessos

$p = 0{,}20$ ➔ probabilidade de um resistor falhar

Resultado	Probabilidade	
$F_1F_2F_3$	$(1-p)(1-p)(1-p)$	$p^0(1-p)^3$

$$P(Y = 0) = 1 \cdot (1-p)^3$$

Resultado	Probabilidade	
$S_1S_2S_3$	ppp	$p^3(1-p)^0$

$$P(Y = 3) = 1 \cdot p^3$$

Distribuição de Probabilidades

Y: *número de sucessos, com* $p = 0{,}20$

Espaço amostral	y	$P(Y = y)$
$S_1S_2S_3$	3	0,008
$S_1S_2F_3$	2	0,032
$S_1F_2S_3$	2	0,032
$F_1S_2S_3$	2	0,032
$S_1F_2F_3$	1	0,128
$F_1S_2F_3$	1	0,128
$F_1F_2S_3$	1	0,128
$F_1F_2F_3$	0	0,512

Distribuição de Probabilidades de Y

y	$P(Y = y)$	$P(Y = y)$
0	$1\ 0{,}2^0\ 0{,}8^3$	$= 0{,}512$
1	$3\ 0{,}2^1\ 0{,}8^2$	$= 0{,}384$
2	$3\ 0{,}2^2\ 0{,}8^1$	$= 0{,}096$
3	$1\ 0{,}2^3\ 0{,}8^0$	$= 0{,}008$
soma		**1,000**

$$P(Y = y) = C_{n,y}\, p^y (1 - p)^{n-y}$$

Distribuição Binomial

Modela experimentos binomiais.

Y: número de sucessos em um experimento binomial com n tentativas

Y ~ Bin(n;p)

Para y = 0,1,..., n, temos:

$$P(Y = y) = \binom{n}{y} p^{y} (1 - p)^{n-y}$$

Em n repetições, em quantas combinações aparecem y sucessos?

Probabilidade de sucesso ocorrendo y vezes

Probabilidade de fracasso ocorrendo (n-y) vezes

Forma geral

$$P(Y = y) = \binom{n}{y} p^{y} (1 - p)^{n-y}$$

- P(Y=y): probabilidade de y sucessos em n tentativas
- n: número de tentativas
- p: probabilidade de sucesso em cada tentativa
- $\binom{n}{y} = \frac{n!}{y!(n-y)!}$

Esperança e Variância da Binomial

Quando **Y ~ Bin(n;p)**, então

$$\mu = E(Y) = np$$

$$\sigma^2 = Var(Y) = np(1-p)$$

Esperança da Binomial

$$\boldsymbol{Y \sim Bin(n,p)}$$

Esperança (veja definição de esperança na Aula 06):

$$\mathrm{E}(Y) = \sum_{y=0}^{n} y \cdot P(Y = y) =$$

$$= \sum_{y=0}^{n} y \cdot \binom{n}{y} p^{y} (1-p)^{n-y} = \cdots = np$$

$$\boldsymbol{\mathrm{E}(Y) = np}$$

Variância da Binomial

$$\boldsymbol{Y \sim Bin(n, p)}$$

Variância (veja definição de variância na Aula 08):

$$\mathrm{Var}(Y) = \sum_{y=0}^{n} (y - \mathrm{E}(Y))^2 \cdot P(Y = y) =$$

$$= \sum_{y=0}^{n} (y - \mathrm{E}(Y))^2 \cdot \binom{n}{y} p^y (1-p)^{n-y} = \cdots$$

$$= np(1-p)$$

$$\mathbf{Var}(\boldsymbol{Y}) = \boldsymbol{np}(\mathbf{1} - \boldsymbol{p})$$

Exemplo 1 – item (c)

Um resistor de 100K Ohms comprado na Santa Efigênia tem probabilidade de falha de 20%, segundo um fabricante bem ruim.

Um aluno de engenharia compra um pacote com 3 desses resistores.

Responda:

c) E se for um pacote com 100 resistores, qual a probabilidade de 20 falharem?

Vamos calcular no Jupyter

A probabilidade de um determinado resistor falhar é sempre de 0,20.

Suponha que os resistores se comportem de maneira independente.

a) Se um pacote tem 100 resistores, qual a probabilidade de 20 falharem?

b) Se um pacote tem 100 resistores, qual a probabilidade de no máximo 20 falharem?

c) Se um pacote tem 100 resistores, qual a probabilidade de pelo menos 20 falharem?

Distribuição de Binomial

Como calcular no Python: $\mathbf{Y \sim Binomial(}\boldsymbol{n, p}\mathbf{)}$

```
In [ ]: from scipy import stats
```

$\mathrm{P}(\mathrm{Y}=y) \Rightarrow \mathrm{stats.binom.pmf}(y, \mathrm{n}, \mathrm{p})$

$\mathrm{P}(\mathrm{Y}\leq y) \Rightarrow \mathrm{stats.binom.cdf}(y, \mathrm{n}, \mathrm{p})$

$\mathrm{E}(\mathrm{Y}) \Rightarrow \mathrm{stats.binom.mean}(\mathrm{n}, \mathrm{p})$

$\mathrm{Var}(\mathrm{Y}) \Rightarrow \mathrm{stats.binom.var}(\mathrm{n}, \mathrm{p})$

$\mathrm{DP}(\mathrm{Y}) \Rightarrow \mathrm{stats.binom.std}(\mathrm{n}, \mathrm{p})$

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.binom.html

Exemplo

- Download do notebook
- Fazer individual e discutir em grupo
- Usar arquivo:

`Aula11_Exemplo_....ipynb`

Atividade

- Download do notebook
- Fazer individual e discutir em grupo
- Usar arquivo:

`Aula11_Atividade_....ipynb`

Exercício

- Download do notebook
- Fazer individual e discutir em grupo
- Usar arquivo para APS6:

`Aula11_Exercicio_....ipynb`

Exercícios

Exercício 1 – APS6

A mortalidade de microempresas no 1º ano de funcionamento é da ordem de 55%. Deseja-se avaliar as causas da mortalidade. Numa amostra de 10 empresas,

a. qual á a probabilidade de exatamente 5 terem falido ao final do 1º ano?

b. qual é a probabilidade de pelo menos 3 virem a falir no 1º ano?

c. sabe-se que pelo menos 3 faliram no 1º ano, qual é a probabilidade de no máximo 5 terem falido?

d. qual o número esperado de empresas que irão à falência no 1º ano na amostra? E o desvio padrão?

Exercício 2

O número médio de clientes satisfeitos com o atendimento de uma loja, em amostras de 40 clientes escolhidos ao acaso, é de 5,5.

Qual é a probabilidade de, numa amostra de 40 clientes escolhidos ao acaso, encontrarmos pelo menos 2 satisfeitos?

Resp.: p=0,1375 ➔ P(X>=2)=98,02%

Exercício 3 – APS6

Um supermercado que tem um cartão de fidelidade classifica os clientes em regular e diamante. Atualmente 70 % dos clientes são regulares. A direção do supermercado deseja premiar os clientes que, num determinado mês, comprem mais de 1 mil reais. Sabemos que se o cliente é diamante a probabilidade dele fazer uma compra acima de 1 mil reais é de 75%; já para o cliente regular, essa probabilidade é de 5%

a) Sorteada uma amostra de 20 clientes diamante, qual é a probabilidade de exatamente 12 fazerem uma compra acima de 1 mil reais?

b) Sorteados 20 clientes com a mesma classificação ao acaso, qual a probabilidade de exatamente 12 fazerem uma compra acima de 1 mil reais?

c) Sorteados 20 clientes ao acaso, qual a probabilidade de exatamente 12 fazerem uma compra acima de 1 mil reais?

d) Um gerente está analisando 200 compras efetuadas em um caixa, dessas 120 foram superiores a 1 mil reais. Dessas compras sob análise, deseja-se sortear uma amostra de 20 compras, sendo que a mesma compra não pode ser sorteada mais de uma vez, qual é a probabilidade de 15 terem sido superiores a 1 mil reais?

Exercício 4

Um agricultor cultiva laranjas e também produz mudas para vender. Após alguns meses a muda pode ser atacada por fungos com probabilidade 0,05 e, nesse caso, ela é escolhida para ser recuperada com probabilidade 0,5. Admita que o processo de recuperação é infalível. O custo de cada muda produzida é R$ 1,00; acrescido de R$1,50 se for para o processo de recuperação. Cada muda é vendida a R$ 4,00 e são descartadas as mudas que não foram para o processo de recuperação de ataque de fungos.

Responda:

a) Qual é a distribuição de probabilidades do lucro (ou seja, pode haver ganho ou prejuízo) por muda produzida pelo agricultor? Justifique claramente todas as contas feitas para obtenção do resultado.

x	-1,00	1,50	3,00
P(X=x)	0,025	0,025	0,95

sendo X a variável aleatória lucro

Exercício 4 - continuação

b) O agricultor seleciona aleatoriamente 3 mudas da sua produção de mudas de laranjas para verificar se houve ataque por fungos após alguns meses. Nesse caso:

b.1) Qual é a probabilidade dele encontrar pelo menos 2 mudas que foram atacadas por fungos? Justifique claramente todas as contas feitas para obtenção do resultado. R: 0,725%

b.2) Sabendo que o agricultor já encontrou pelo menos 2 mudas que foram atacadas por fungos, qual é a probabilidade de encontrar menos de 3 atacadas por fungos? Justifique claramente todas as contas feitas para obtenção do resultado. R: 98,3%