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?

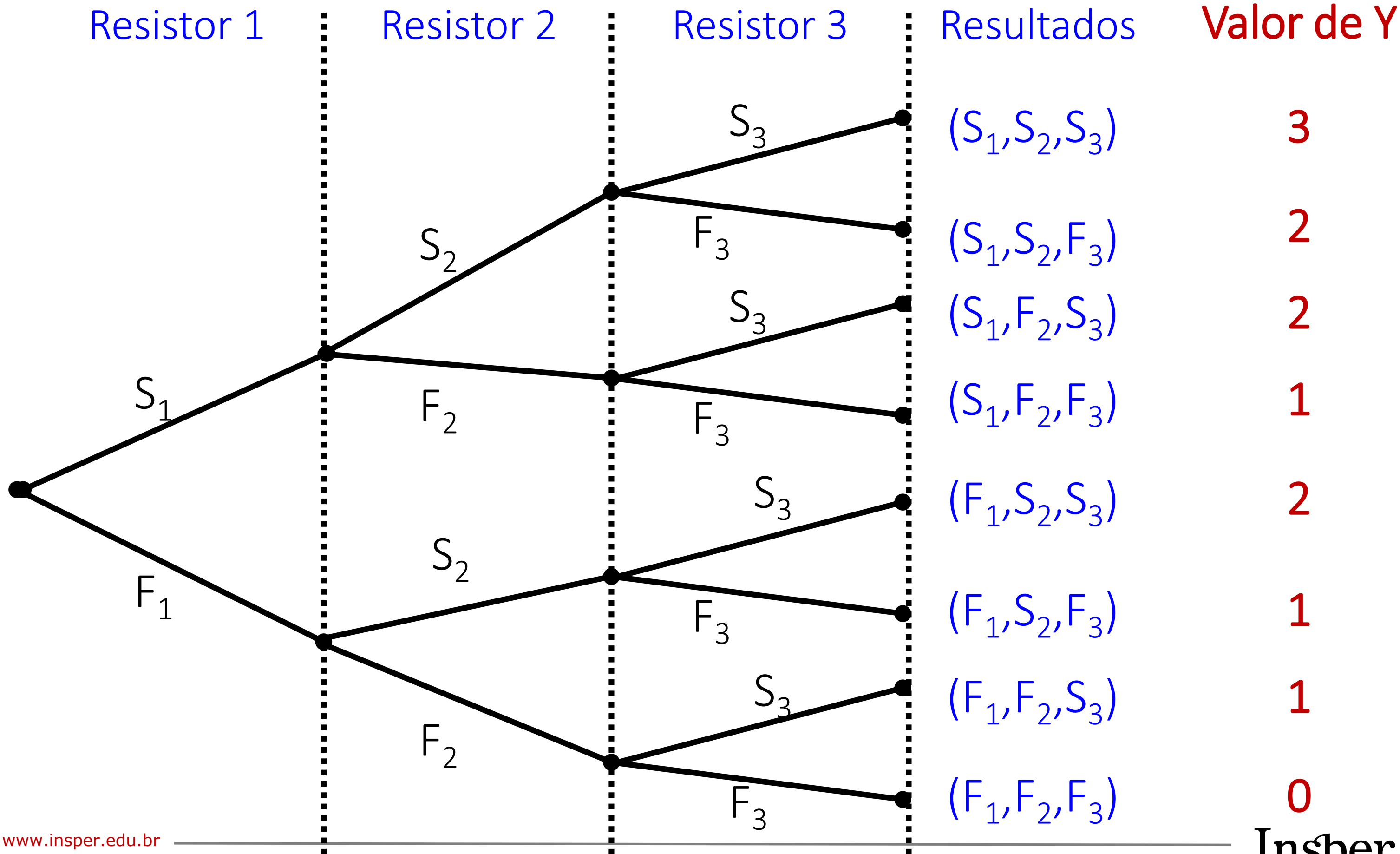
c) E se for um pacote com 100 resistores, qual a probabilidade de 20 falharem?

Exemplo 1

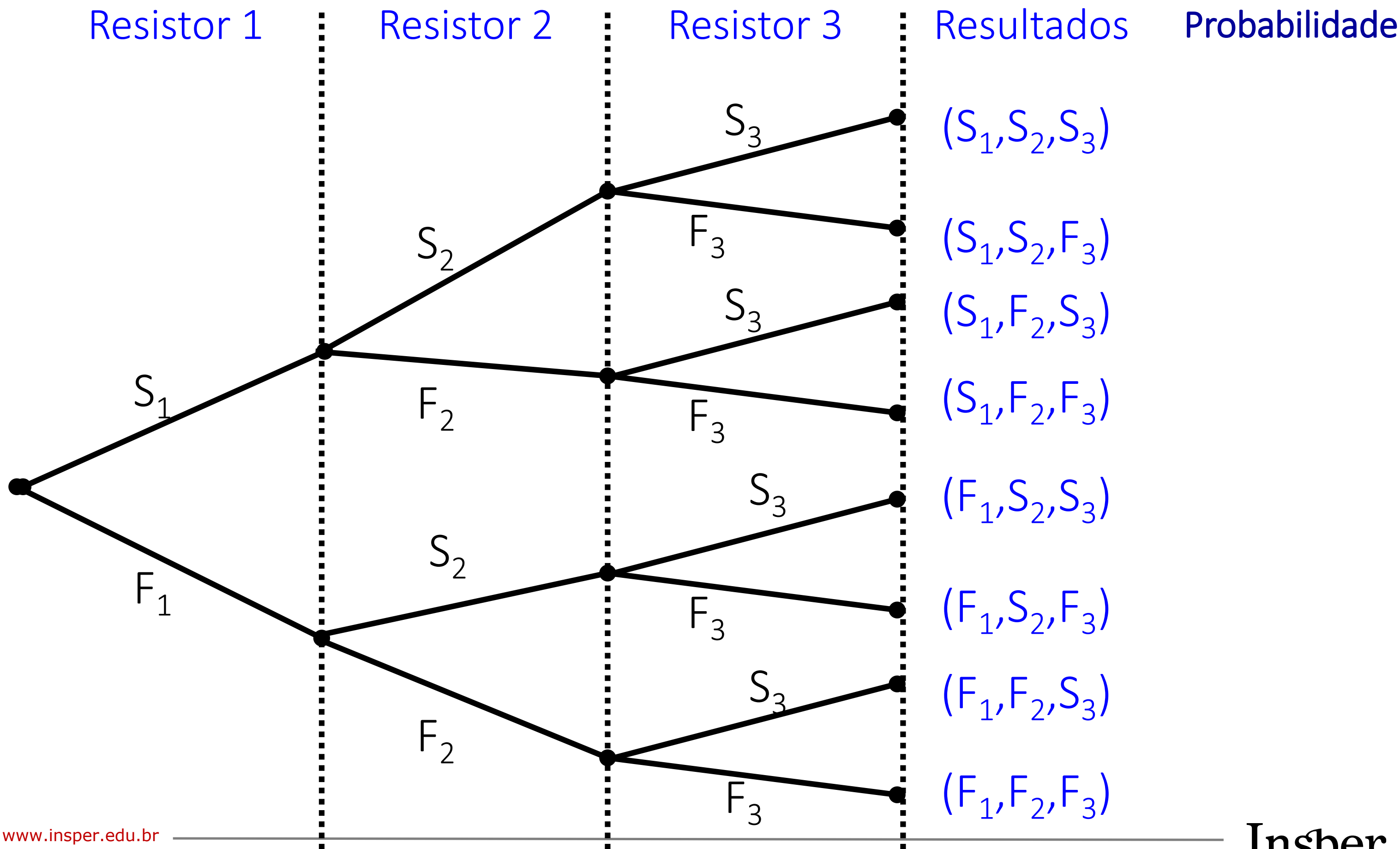
Tal experimento possui a distribuição binomial?

- 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.

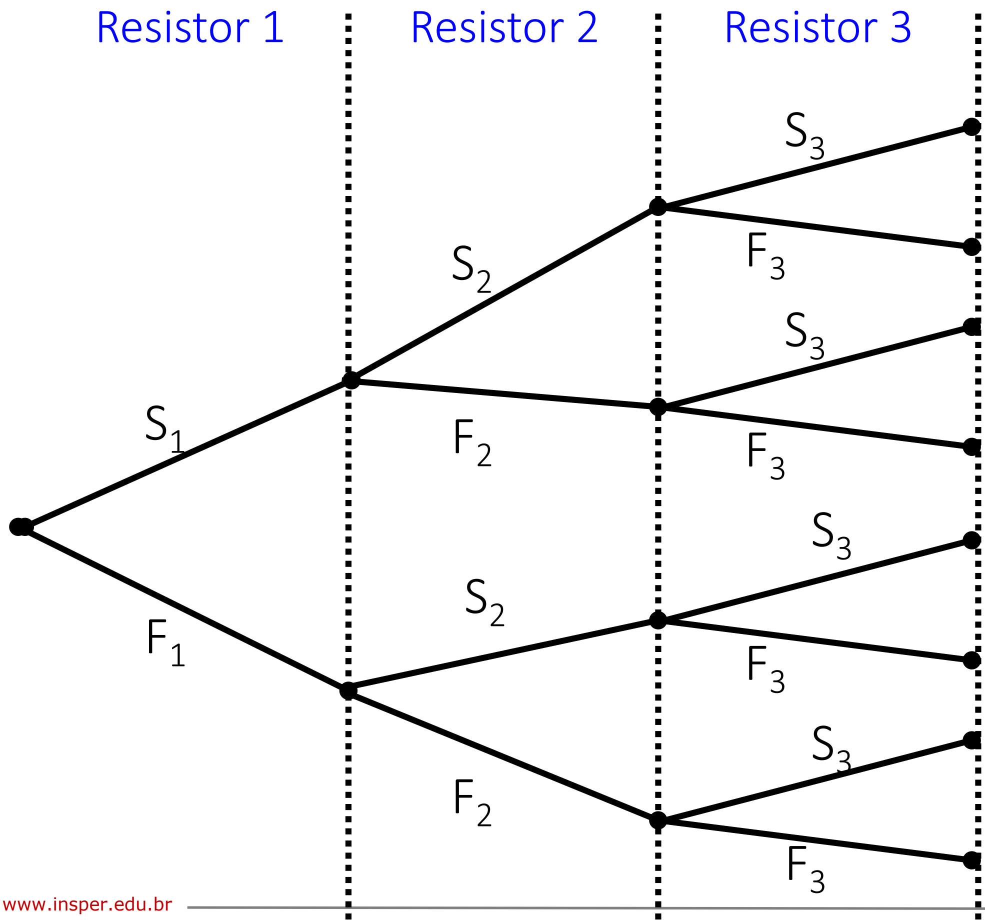
Árvore de probabilidades



Árvore de probabilidades

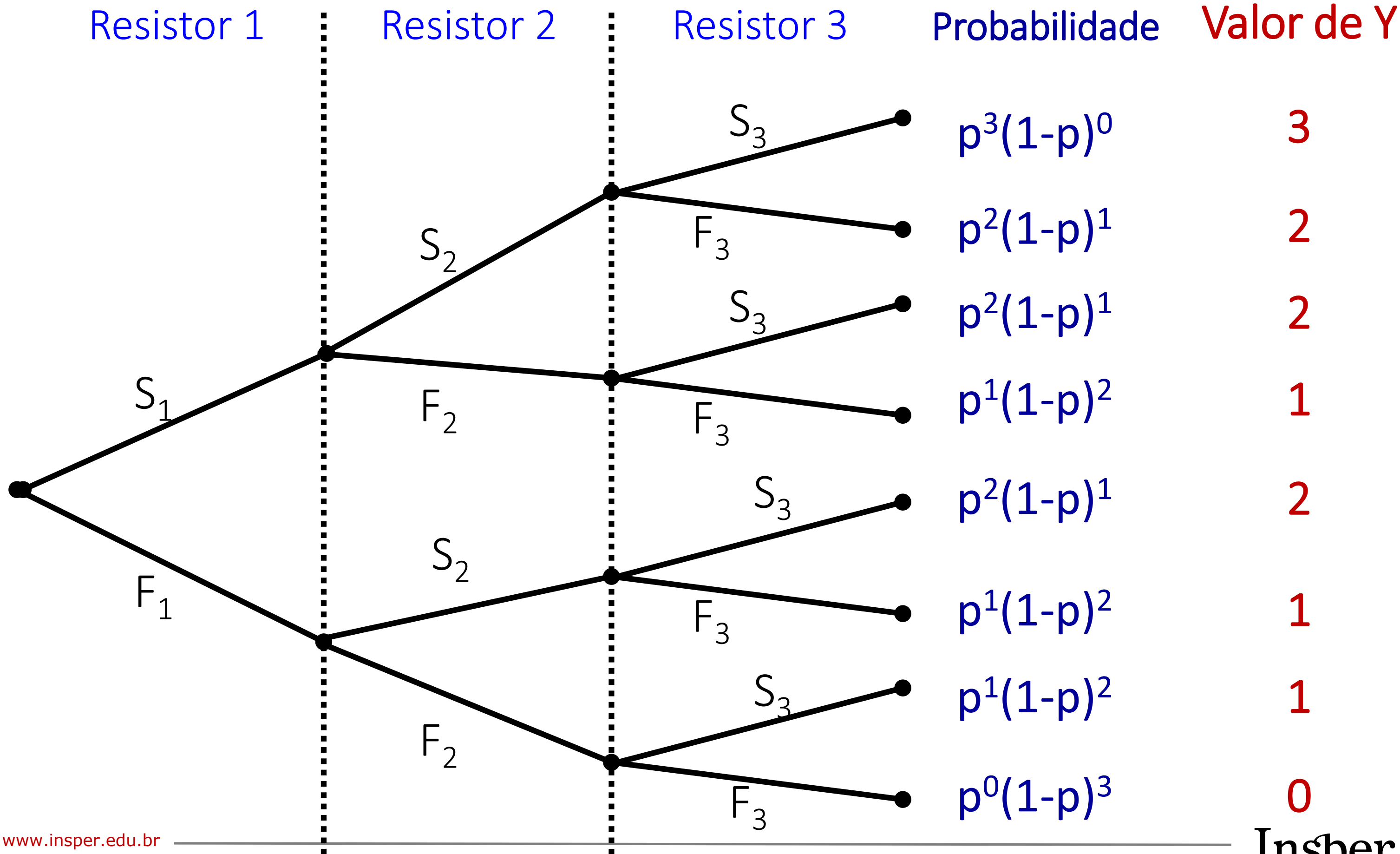


Árvore de probabilidades



Resultados	Probabilidade
(S_1, S_2, S_3)	$p^3(1-p)^0$
(S_1, S_2, F_3)	$p^2(1-p)^1$
(S_1, F_2, S_3)	$p^2(1-p)^1$
(S_1, F_2, F_3)	$p^1(1-p)^2$
(F_1, S_2, S_3)	$p^2(1-p)^1$
(F_1, S_2, F_3)	$p^1(1-p)^2$
(F_1, F_2, S_3)	$p^1(1-p)^2$
(F_1, F_2, F_3)	$p^0(1-p)^3$

Árvore de probabilidades



Probabilidade para sequências com 2 sucessos

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

Resultado	Probabilidade	
$S_1S_2F_3$	$pp(1\text{-}p)$	$p^2(1\text{-}p)$
$S_1F_2S_3$	$p(1\text{-}p)p$	$p^2(1\text{-}p)$
$F_1S_2S_3$	$(1\text{-}p)pp$	$p^2(1\text{-}p)$

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

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

Resultado	*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
	soma	**1,000**

Distribuição de probabilidades de Y

y	*p(Y=y)*
0	0,512
1	0,384
2	0,096
3	0,008
soma	**1,000**

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
- $\binom{n}{y} = \dfrac{n!}{y!(n-y)!}$

Esperança e Variância da Binomial

Quando **Y ~ Bin(n;p)**, então

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

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

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: $\mathbf{Y \sim Binomial(\boldsymbol{n}, \boldsymbol{p})}$

In [ ]: from scipy import stats

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

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

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

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

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

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

Exemplo

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

`Aula11_Exemplo_....ipynb`

Atividade

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

`Aula11_Atividade_....ipynb`

Exercício

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

`Aula11_Exercicio_....ipynb`

Atenção: Tem APS6 para esse exercício.