

# Pré-Modelagem em Ciência de Dados

Prof. Rilder S. Pires

MBA em Ciência de Dados

Na aula passada...

# Na aula passada...

## Parte Teórica:

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional
- ▶ Variáveis Aleatórias:

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional
- ▶ Variáveis Aleatórias:
  - ▶ Definição



# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional
- ▶ Variáveis Aleatórias:
  - ▶ Definição
  - ▶ Exemplos

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional
- ▶ Variáveis Aleatórias:
  - ▶ Definição
  - ▶ Exemplos

## Parte Prática:

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional
- ▶ Variáveis Aleatórias:
  - ▶ Definição
  - ▶ Exemplos

## Parte Prática:

- ▶ Projeto Final:

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional
- ▶ Variáveis Aleatórias:
  - ▶ Definição
  - ▶ Exemplos

## Parte Prática:

- ▶ Projeto Final:
- ▶ Quais produtos a sua região produz?

# Na aula passada...

## Parte Teórica:

- ▶ Probabilidade:
  - ▶ Eventos Independentes
  - ▶ Probabilidade Condicional
- ▶ Variáveis Aleatórias:
  - ▶ Definição
  - ▶ Exemplos

## Parte Prática:

- ▶ Projeto Final:
- ▶ Quais produtos a sua região produz?
- ▶ Quais os produtos mais produzidos (em valor) pelos municípios da sua região?

## Variável Aleatória

# Variáveis Aleatórias

## Variável Aleatória

Exemplo:

# Variáveis Aleatórias

## Variável Aleatória

### Exemplo:

Jogue uma moeda duas vezes e seja  $X$  o número de caras.



## Variável Aleatória

### Exemplo:

Jogue uma moeda duas vezes e seja  $X$  o número de caras.

Então,  $\mathbb{P}(X = 0) = \mathbb{P}(\{TT\}) = 1/4$ ,

$\mathbb{P}(X = 1) = \mathbb{P}(\{HT, TH\}) = 1/2$  e

$\mathbb{P}(X = 2) = \mathbb{P}(\{HH\}) = 1/4$ .

# Variáveis Aleatórias

## Variável Aleatória

### Exemplo:

Jogue uma moeda duas vezes e seja  $X$  o número de caras.

Então,  $\mathbb{P}(X = 0) = \mathbb{P}(\{TT\}) = 1/4$ ,

$\mathbb{P}(X = 1) = \mathbb{P}(\{HT, TH\}) = 1/2$  e

$\mathbb{P}(X = 2) = \mathbb{P}(\{HH\}) = 1/4$ .

A variável aleatória e sua distribuição podem ser resumidas da seguinte forma:

$\omega$	$\mathbb{P}(\{\omega\})$	$X(\omega)$
$TT$	$1/4$	0
$TH$	$1/4$	1
$HT$	$1/4$	1
$HH$	$1/4$	2

$x$	$\mathbb{P}(X = x)$
0	$1/4$
1	$1/2$
2	$1/4$

**Funções de distribuição e Funções de probabilidade:**

# Variável Aleatória

Funções de distribuição e Funções de probabilidade:

Função de Distribuição Cumulativa:

## Funções de distribuição e Funções de probabilidade:

### Função de Distribuição Cumulativa:

- ▶ Dada uma variável aleatória  $X$ , definimos a função de distribuição cumulativa (ou função de distribuição) da seguinte forma.

## Funções de distribuição e Funções de probabilidade:

### Função de Distribuição Cumulativa:

- ▶ Dada uma variável aleatória  $X$ , definimos a função de distribuição cumulativa (ou função de distribuição) da seguinte forma.

### Definição:

## Funções de distribuição e Funções de probabilidade:

### Função de Distribuição Cumulativa:

- ▶ Dada uma variável aleatória  $X$ , definimos a função de distribuição cumulativa (ou função de distribuição) da seguinte forma.

### Definição:

- ▶ A **função de distribuição cumulativa**, ou *CDF* (cumulative distribution function), é a função  $F_X : \mathbb{R} \rightarrow [0, 1]$  definida por.

$$F_X(x) = \mathbb{P}(X \leq x)$$

## Funções de distribuição e Funções de probabilidade:

### Função de Distribuição Cumulativa:

- ▶ Dada uma variável aleatória  $X$ , definimos a função de distribuição cumulativa (ou função de distribuição) da seguinte forma.

### Definição:

- ▶ A **função de distribuição cumulativa**, ou *CDF* (cumulative distribution function), é a função  $F_X : \mathbb{R} \rightarrow [0, 1]$  definida por.

$$F_X(x) = \mathbb{P}(X \leq x)$$

- ▶ A função de distribuição cumulativa contém efetivamente toda a informação sobre a variável aleatória.



# Variáveis Aleatórias

**Função de Distribuição Cumulativa:**

# Variáveis Aleatórias

**Função de Distribuição Cumulativa:**  
**Exemplo:**

# Variáveis Aleatórias

## Função de Distribuição Cumulativa:

### Exemplo:

Jogue uma moeda duas vezes e seja  $X$  o número de caras.

$\omega$	$\mathbb{P}(\{\omega\})$	$X(\omega)$
$TT$	$1/4$	0
$TH$	$1/4$	1
$HT$	$1/4$	1
$HH$	$1/4$	2

$x$	$\mathbb{P}(X = x)$
0	$1/4$
1	$1/2$
2	$1/4$

# Variáveis Aleatórias

## Função de Distribuição Cumulativa:

### Exemplo:

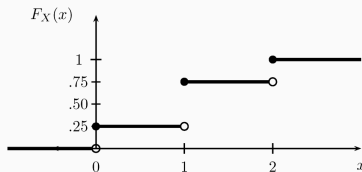
Jogue uma moeda duas vezes e seja  $X$  o número de caras.

$\omega$	$\mathbb{P}(\{\omega\})$	$X(\omega)$
$TT$	$1/4$	0
$TH$	$1/4$	1
$HT$	$1/4$	1
$HH$	$1/4$	2

$x$	$\mathbb{P}(X = x)$
0	$1/4$
1	$1/2$
2	$1/4$

A função de distribuição é

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$



# Variáveis Aleatórias

## Função de Distribuição Cumulativa:

### Exemplo:

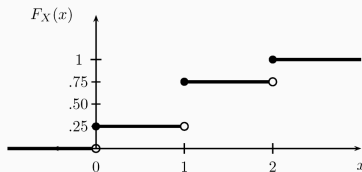
Jogue uma moeda duas vezes e seja  $X$  o número de caras.

$\omega$	$\mathbb{P}(\{\omega\})$	$X(\omega)$
$TT$	$1/4$	0
$TH$	$1/4$	1
$HT$	$1/4$	1
$HH$	$1/4$	2

$x$	$\mathbb{P}(X = x)$
0	$1/4$
1	$1/2$
2	$1/4$

A função de distribuição é

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$



► Observe que a função é não decrescente.

# Variáveis Aleatórias

## Função de Distribuição Cumulativa:

### Exemplo:

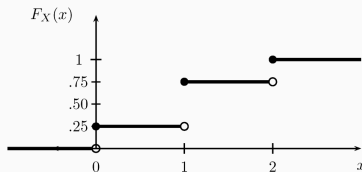
Jogue uma moeda duas vezes e seja  $X$  o número de caras.

$\omega$	$\mathbb{P}(\{\omega\})$	$X(\omega)$
$TT$	$1/4$	0
$TH$	$1/4$	1
$HT$	$1/4$	1
$HH$	$1/4$	2

$x$	$\mathbb{P}(X = x)$
0	$1/4$
1	$1/2$
2	$1/4$

A função de distribuição é

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$



- Observe que a função é não decrescente.
- É definida para todo  $x$ , mesmo que  $X$  aceite apenas 0, 1 e 2.

## Função de Probabilidade:

**Função de Probabilidade:**

**Definição:**



## Função de Probabilidade:

### Definição:

- ▶  $X$  é discreta se receber valores contáveis  $\{x_1, x_2, \dots\}$ . Definimos a **função de probabilidade** para  $X$  por

$$f_X(x) = \mathbb{P}(X = x)$$

## Função de Probabilidade:

### Definição:

- ▶  $X$  é discreta se receber valores contáveis  $\{x_1, x_2, \dots\}$ . Definimos a **função de probabilidade** para  $X$  por

$$f_X(x) = \mathbb{P}(X = x)$$

- ▶ Assim,  $f_X(x) \geq 0$  para todos os  $x \in \mathbb{R}$  e  $\sum_i f_X(x_i) = 1$ .

## Função de Probabilidade:

### Definição:

- ▶  $X$  é discreta se receber valores contáveis  $\{x_1, x_2, \dots\}$ . Definimos a **função de probabilidade** para  $X$  por

$$f_X(x) = \mathbb{P}(X = x)$$

- ▶ Assim,  $f_X(x) \geq 0$  para todos os  $x \in \mathbb{R}$  e  $\sum_i f_X(x_i) = 1$ .
- ▶ A função de distribuição cumulativa  $X$  é relacionada com  $f_X$  por

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

## Função Densidade de Probabilidade:

## Função Densidade de Probabilidade: Definição:

## Função Densidade de Probabilidade:

### Definição:

- ▶ Uma variável aleatória  $X$  é **contínua** se houver uma função  $f_X$  de modo que  $f_X(x) \geq 0$  para todo  $x$ ,  $\int_{-\infty}^{\infty} f_X(x)dx = 1$  e para todo  $a \leq b$ ,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx.$$

## Função Densidade de Probabilidade:

### Definição:

- ▶ Uma variável aleatória  $X$  é **contínua** se houver uma função  $f_X$  de modo que  $f_X(x) \geq 0$  para todo  $x$ ,  $\int_{-\infty}^{\infty} f_X(x)dx = 1$  e para todo  $a \leq b$ ,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx.$$

- ▶ A função  $f_X$  é chamada de **função densidade de probabilidade**. Além disso,

$$F_X(x) = \int_{-\infty}^x f_X(t)dt.$$

e  $f_X(x) = F'_X(x)$  em todos os pontos  $x$  nos quais  $F_X$  é diferenciável.

# Pré-Modelagem em Ciência de Dados

## Ementa:



## Ementa:

- ▶ Conceitos de Axiomas da Probabilidade

## **Ementa:**

- ▶ Conceitos de Axiomas da Probabilidade
- ▶ Atribuições das Probabilidades

# Pré-Modelagem em Ciência de Dados

## Ementa:

- ▶ Conceitos de Axiomas da Probabilidade
- ▶ Atribuições das Probabilidades
- ▶ O que é uma variável aleatória?

# Pré-Modelagem em Ciência de Dados

## Ementa:

- ▶ Conceitos de Axiomas da Probabilidade
- ▶ Atribuições das Probabilidades
- ▶ O que é uma variável aleatória?
- ▶ Distribuição de Probabilidade Discretas:
  - ▶ Distribuição de Bernoulli,
  - ▶ Distribuição Binomial,
  - ▶ Distribuição de Poisson,
  - ▶ Distribuição Geométrica e Hipergeométrica

# Pré-Modelagem em Ciência de Dados

## Ementa:

- ▶ Conceitos de Axiomas da Probabilidade
- ▶ Atribuições das Probabilidades
- ▶ O que é uma variável aleatória?
- ▶ Distribuição de Probabilidade Discretas:
  - ▶ Distribuição de Bernoulli,
  - ▶ Distribuição Binomial,
  - ▶ Distribuição de Poisson,
  - ▶ Distribuição Geométrica e Hipergeométrica
- ▶ Distribuições Contínuas:
  - ▶ Distribuição Uniforme,
  - ▶ Distribuição Exponencial,
  - ▶ Distribuição Normal ou Gaussiana,
  - ▶ Cálculo de Probabilidade em Distribuições Normais e Funções lineares de Distribuições Normais.

# Pré-Modelagem em Ciência de Dados

## Ementa:

- ▶ Conceitos de Axiomas da Probabilidade
- ▶ Atribuições das Probabilidades
- ▶ O que é uma variável aleatória?
- ▶ Distribuição de Probabilidade Discretas:
  - ▶ Distribuição de Bernoulli,
  - ▶ Distribuição Binomial,
  - ▶ Distribuição de Poisson,
  - ▶ Distribuição Geométrica e Hipergeométrica
- ▶ Distribuições Contínuas:
  - ▶ Distribuição Uniforme,
  - ▶ Distribuição Exponencial,
  - ▶ Distribuição Normal ou Gaussiana,
  - ▶ Cálculo de Probabilidade em Distribuições Normais e Funções lineares de Distribuições Normais.
- ▶ Inferência Estatística: Noções de amostragem e estimação.

# Projeto Final

## Projeto Final:

# Projeto Final

Projeto Final:

Perguntas



## Projeto Final:

## Perguntas

1. Qual a distribuição da “diversidade” dos municípios da sua região?

## Projeto Final:

### Perguntas

1. Qual a distribuição da “diversidade” dos municípios da sua região?
2. Qual a distribuição dos valores de produção agrícola dos municípios da sua região?

## Projeto Final:

### Perguntas

1. Qual a distribuição da “diversidade” dos municípios da sua região?
2. Qual a distribuição dos valores de produção agrícola dos municípios da sua região?
3. Qual a distribuição dos valores de produção do principal produto para municípios da sua região?

# Projeto Final

## Projeto Final:

### Perguntas

1. Qual a distribuição da “diversidade” dos municípios da sua região?
2. Qual a distribuição dos valores de produção agrícola dos municípios da sua região?
3. Qual a distribuição dos valores de produção do principal produto para municípios da sua região?
4. e para o Ceará?

## Projeto Final:

### Perguntas

1. Qual a distribuição da “diversidade” dos municípios da sua região?
2. Qual a distribuição dos valores de produção agrícola dos municípios da sua região?
3. Qual a distribuição dos valores de produção do principal produto para municípios da sua região?
4. e para o Ceará?
5. Quais outras variáveis podemos considerar?

# Fim

*Obrigado pela atenção!*