FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA
ENSINANDO E APRENDENDO

Pré-Modelagem em Ciência de Dados

Prof. Rilder S. Pires

MBA em Ciência de Dados

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?

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.

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.

Encontros:

Pré-Modelagem em Ciência de Dados

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021
- Módulo 2: 13, 14 e 15 de janeiro de 2022

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021
- Módulo 2: 13, 14 e 15 de janeiro de 2022
- Módulo 3: 27, 28 e 29 de janeiro de 2022

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021
- Módulo 2: 13, 14 e 15 de janeiro de 2022
- Módulo 3: 27, 28 e 29 de janeiro de 2022

Observações:

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021
- Módulo 2: 13, 14 e 15 de janeiro de 2022
- Módulo 3: 27, 28 e 29 de janeiro de 2022

Observações:

- Aulas remotas, híbridas ou presenciais

Pré-Modelagem em Ciência de Dados

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021
- Módulo 2: 13, 14 e 15 de janeiro de 2022
- Módulo 3: 27, 28 e 29 de janeiro de 2022

Observações:

- Aulas remotas, híbridas ou presenciais
- Aulas teóricas e práticas

Pré-Modelagem em Ciência de Dados

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021
- Módulo 2: 13, 14 e 15 de janeiro de 2022
- Módulo 3: 27, 28 e 29 de janeiro de 2022

Observações:

- Aulas remotas, híbridas ou presenciais
- Aulas teóricas e práticas
- Conclusão da disciplina: **Projeto Final**

Encontros:

- Módulo 1: 09, 10 e 11 de dezembro de 2021
- Módulo 2: 13, 14 e 15 de janeiro de 2022
- Módulo 3: 27, 28 e 29 de janeiro de 2022

Observações:

- Aulas remotas, híbridas ou presenciais
- Aulas teóricas e práticas
- Conclusão da disciplina: **Projeto Final**
- Linguagem utilizada: **Python**

Projeto Final

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

- Quão diferente são as Mesoregiões Cearenses?

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

- Quão diferente são as Mesoregiões Cearenses?

Observações:

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

- Quão diferente são as Mesoregiões Cearenses?

Observações:

- Dados da Plataforma SIDRA-IBGE

Pré-Modelagem em Ciência de Dados

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

- Quão diferente são as Mesoregiões Cearenses?

Observações:

- Dados da Plataforma SIDRA-IBGE
- Produção Agrícola Municipal (https://sidra.ibge.gov.br/tabela/5457)

Pré-Modelagem em Ciência de Dados

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

- Quão diferente são as Mesoregiões Cearenses?

Observações:

- Dados da Plataforma SIDRA-IBGE
- Produção Agrícola Municipal (https://sidra.ibge.gov.br/tabela/5457)
- Produto Interno Bruto dos Municípios (https://sidra.ibge.gov.br/tabela/5938)

Pré-Modelagem em Ciência de Dados

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

- Quão diferente são as Mesoregiões Cearenses?

Observações:

- Dados da Plataforma SIDRA-IBGE
- Produção Agrícola Municipal (https://sidra.ibge.gov.br/tabela/5457)
- Produto Interno Bruto dos Municípios (https://sidra.ibge.gov.br/tabela/5938)
- Estimativas de População: (https://sidra.ibge.gov.br/tabela/6579)

Projeto Final

- Análise de Dados Sócio-Econômicos das Mesoregiões Cearenses

Pergunta Norteadora:

- Quão diferente são as Mesoregiões Cearenses?

Observações:

- Dados da Plataforma SIDRA-IBGE
- Produção Agrícola Municipal (https://sidra.ibge.gov.br/tabela/5457)
- Produto Interno Bruto dos Municípios (https://sidra.ibge.gov.br/tabela/5938)
- Estimativas de População: (https://sidra.ibge.gov.br/tabela/6579)
- Entregar os **notebooks com códigos e explicações**.

Estatística Básica:

Estatística Básica:

- **Medidas de Tendência Central:**

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana

Revisão

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda
- **Medidas de Variabilidade:**

Revisão

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda
- **Medidas de Variabilidade:**
  - Desvio padrão

Revisão

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda
- **Medidas de Variabilidade:**
  - Desvio padrão
  - Valor mínimo

Revisão

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda
- **Medidas de Variabilidade:**
  - Desvio padrão
  - Valor mínimo
  - Valor máximo

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda
- **Medidas de Variabilidade:**
  - Desvio padrão
  - Valor mínimo
  - Valor máximo
- **Quantils:**

Revisão

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda
- **Medidas de Variabilidade:**
  - Desvio padrão
  - Valor mínimo
  - Valor máximo
- **Quantils:**
  - Quartis

Estatística Básica:

- **Medidas de Tendência Central:**
  - Média
  - Mediana
  - Moda
- **Medidas de Variabilidade:**
  - Desvio padrão
  - Valor mínimo
  - Valor máximo
- **Quantils:**
  - Quartis
  - Percentis

Introdução:

O que é Probabilidade?

Introdução:

O que é Probabilidade?

- **Probabilidade** é uma forma matemática de quantificar a incerteza.

Espaço Amostral e Eventos:

Espaço Amostral e Eventos:

- O **espaço amostral Ω** é o conjunto de todos os resultados possíveis de um experimento.

Espaço Amostral e Eventos:

- O **espaço amostral** Ω é o conjunto de todos os resultados possíveis de um experimento. Pontos ω em Ω são chamados de **resultados**, **realizações** ou **elementos**.

Espaço Amostral e Eventos:

- O **espaço amostral** Ω é o conjunto de todos os resultados possíveis de um experimento. Pontos ω em Ω são chamados de **resultados**, **realizações** ou **elementos**. Subconjuntos de Ω são chamados de **Eventos**.

Probabilidade

Espaço Amostral e Eventos:

- O **espaço amostral** Ω é o conjunto de todos os resultados possíveis de um experimento. Pontos ω em Ω são chamados de **resultados**, **realizações** ou **elementos**. Subconjuntos de Ω são chamados de **Eventos**.
- **Exemplo 1:**
  Se jogarmos uma moeda uma vez, então $\Omega = \{H\ , T\}$. O evento em que o resultado do lance é cara é $A = \{H\}$.

Probabilidade

Espaço Amostral e Eventos:

- O **espaço amostral** Ω é o conjunto de todos os resultados possíveis de um experimento. Pontos ω em Ω são chamados de **resultados**, **realizações** ou **elementos**. Subconjuntos de Ω são chamados de **Eventos**.
- **Exemplo 1:**
  Se jogarmos uma moeda uma vez, então $\Omega = \{H, T\}$. O evento em que o resultado do lance é cara é $A = \{H\}$.
- **Exemplo 2:**
  Se jogarmos uma moeda duas vezes, então
  $\Omega = ?$.
  O evento em que o primeiro lance é cara é
  $A = ?$.

Probabilidade

Espaço Amostral e Eventos:

- O **espaço amostral** Ω é o conjunto de todos os resultados possíveis de um experimento. Pontos ω em Ω são chamados de **resultados**, **realizações** ou **elementos**. Subconjuntos de Ω são chamados de **Eventos**.
- **Exemplo 1:**
  Se jogarmos uma moeda uma vez, então $\Omega = \{H, T\}$. O evento em que o resultado do lance é cara é $A = \{H\}$.
- **Exemplo 2:**
  Se jogarmos uma moeda duas vezes, então
  $\Omega = \{HH, HT, TH, TT\}$.
  O evento em que o primeiro lance é cara é
  $A = ?$.

Espaço Amostral e Eventos:

- O **espaço amostral** Ω é o conjunto de todos os resultados possíveis de um experimento. Pontos ω em Ω são chamados de **resultados**, **realizações** ou **elementos**. Subconjuntos de Ω são chamados de **Eventos**.
- **Exemplo 1:**
  Se jogarmos uma moeda uma vez, então $\Omega = \{H, T\}$. O evento em que o resultado do lance é cara é $A = \{H\}$.
- **Exemplo 2:**
  Se jogarmos uma moeda duas vezes, então
  $\Omega = \{HH, HT, TH, TT\}$.
  O evento em que o primeiro lance é cara é
  $A = \{HH, HT\}$.

Definição:

Definição:

- Uma função $\mathbb{P}$ que atribui um número real $\mathbb{P}(A)$ a cada evento A é uma **distribuição de probabilidade** ou uma **medida de probabilidade** se satisfizer os três axiomas a seguir:

Probabilidade

Definição:

- Uma função $\mathbb{P}$ que atribui um número real $\mathbb{P}(A)$ a cada evento A é uma **distribuição de probabilidade** ou uma **medida de probabilidade** se satisfizer os três axiomas a seguir:

  Axioma 1: $\mathbb{P}(A) \geq 0$ para todo A

Definição:

- Uma função $\mathbb{P}$ que atribui um número real $\mathbb{P}(A)$ a cada evento A é uma **distribuição de probabilidade** ou uma **medida de probabilidade** se satisfizer os três axiomas a seguir:

  Axioma 1: $\mathbb{P}(A) \geq 0$ para todo A

  Axioma 2: $\mathbb{P}(\Omega) = 1$

Probabilidade

Definição:

- Uma função $\mathbb{P}$ que atribui um número real $\mathbb{P}(A)$ a cada evento A é uma **distribuição de probabilidade** ou uma **medida de probabilidade** se satisfizer os três axiomas a seguir:

  Axioma 1: $\mathbb{P}(A) \geq 0$ para todo A

  Axioma 2: $\mathbb{P}(\Omega) = 1$

  Axioma 3: Se $A_1, A_2, \ldots$ são disjuntos então

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Interpretações:

Probabilidade

Interpretações:

- Existem muitas interpretações para $\mathbb{P}(A)$.

Probabilidade

Interpretações:

- Existem muitas interpretações para $\mathbb{P}(A)$.
  - **Interpretação da frequência:** $\mathbb{P}(A)$ é a proporção de longo prazo de vezes que A é verdadeiro nas repetições.

Probabilidade

Interpretações:

- Existem muitas interpretações para $\mathbb{P}(A)$.
  - **Interpretação da frequência:** $\mathbb{P}(A)$ é a proporção de longo prazo de vezes que A é verdadeiro nas repetições.

Na **interpretação da frequência**, se dissermos que a probabilidade de cara é $1/2$, queremos dizer que, se lançarmos a moeda muitas vezes, a proporção de vezes que recebemos cara tende a $1/2$ conforme o número de lançamentos aumenta.

Probabilidade

Interpretações:

- Existem muitas interpretações para $\mathbb{P}(A)$.
  - **Interpretação da frequência:** $\mathbb{P}(A)$ é a proporção de longo prazo de vezes que A é verdadeiro nas repetições.
  - **Interpretação do grau de crença:** $\mathbb{P}(A)$ mede a grau de crença de um observador de que A é verdadeiro.

Na **interpretação da frequência**, se dissermos que a probabilidade de cara é $1/2$, queremos dizer que, se lançarmos a moeda muitas vezes, a proporção de vezes que recebemos cara tende a $1/2$ conforme o número de lançamentos aumenta.

Probabilidade

Interpretações:

- Existem muitas interpretações para $\mathbb{P}(A)$.
  - **Interpretação da frequência:** $\mathbb{P}(A)$ é a proporção de longo prazo de vezes que A é verdadeiro nas repetições.
  - **Interpretação do grau de crença:** $\mathbb{P}(A)$ mede a grau de crença de um observador de que A é verdadeiro.
- Em qualquer das interpretações, os axiomas de 1 a 3 devem ser válidos.

Na **interpretação da frequência**, se dissermos que a probabilidade de cara é $1/2$, queremos dizer que, se lançarmos a moeda muitas vezes, a proporção de vezes que recebemos cara tende a $1/2$ conforme o número de lançamentos aumenta.

Probabilidade

Interpretações:

- Existem muitas interpretações para $\mathbb{P}(A)$.
  - **Interpretação da frequência:** $\mathbb{P}(A)$ é a proporção de longo prazo de vezes que A é verdadeiro nas repetições.
  - **Interpretação do grau de crença:** $\mathbb{P}(A)$ mede a grau de crença de um observador de que A é verdadeiro.
- Em qualquer das interpretações, os axiomas de 1 a 3 devem ser válidos.
- A diferença na interpretação não importará muito até que lidemos com **inferência estatística**. Lá, as diferentes interpretações levam a duas escolas de inferência: a **escola frequentista** e a **escola bayesiana**.

Na **interpretação da frequência**, se dissermos que a probabilidade de cara é 1/2, queremos dizer que, se lançarmos a moeda muitas vezes, a proporção de vezes que recebemos cara tende a 1/2 conforme o número de lançamentos aumenta.

Propriedades:

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:

Probabilidade

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:
- $\mathbb{P}(\emptyset) = 0$

Probabilidade

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:
- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

Probabilidade

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:
- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$

Probabilidade

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:
- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Probabilidade

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:
- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Probabilidade

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:
- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Uma propriedade menos óbvia é dada por:

- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$

Probabilidade

Propriedades:

- Pode-se derivar muitas propriedades de $\mathbb{P}$ dos axiomas, tais como:
- $\mathbb{P}(\emptyset) = 0$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
- $0 \leq \mathbb{P}(A) \leq 1$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Uma propriedade menos óbvia é dada por:

- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$
- **Exemplo:** Duas moedas foram lançadas. Seja H_1 o evento em que cara ocorre no lance 1 e seja H_2 o evento em que cara ocorre no lance 2.
  $\mathbb{P}(H_1 \cup H_2) = \mathbb{P}(H_1) + \mathbb{P}(H_2) - \mathbb{P}(H_1 H_2) = 1/2 + 1/2 - 1/4 = 3/4.$

Probabilidade em Espaços Amostrais Finitos:

Probabilidade em Espaços Amostrais Finitos:

- Suponha que o espaço amostral $\Omega = \{\omega_1, \ldots, \omega_n\}$seja finito.

Probabilidade em Espaços Amostrais Finitos:

- Suponha que o espaço amostral $\Omega = \{\omega_1, \ldots, \omega_n\}$seja finito.
- Por exemplo, se lançarmos um dado duas vezes, então Ω terá 36 elementos: $\Omega = \{(i, j); i, j \in \{1, \ldots 6\}\}$.

Probabilidade

Probabilidade em Espaços Amostrais Finitos:

- Suponha que o espaço amostral $\Omega = \{\omega_1, \ldots, \omega_n\}$seja finito.
- Por exemplo, se lançarmos um dado duas vezes, então Ω terá 36 elementos: $\Omega = \{(i, j); i, j \in \{1, \ldots 6\}\}$.
- Se cada resultado for igualmente provável, então $P(A) = |A|/36$ onde $|A|$ denota o número de elementos em A.

Probabilidade

Probabilidade em Espaços Amostrais Finitos:

- Suponha que o espaço amostral $\Omega = \{\omega_1, \dots, \omega_n\}$seja finito.
- Por exemplo, se lançarmos um dado duas vezes, então Ω terá 36 elementos: $\Omega = \{(i, j); i, j \in \{1, \dots 6\}\}$.
- Se cada resultado for igualmente provável, então $P(A) = |A|/36$ onde $|A|$ denota o número de elementos em A.
- A probabilidade de que a soma dos dados seja 11 é $2/36$, uma vez que existem dois resultados que correspondem a este evento.

Probabilidade

Probabilidade em Espaços Amostrais Finitos:

- Suponha que o espaço amostral $\Omega = \{\omega_1, \ldots, \omega_n\}$seja finito.
- Por exemplo, se lançarmos um dado duas vezes, então Ω terá 36 elementos: $\Omega = \{(i, j); i, j \in \{1, \ldots 6\}\}$.
- Se cada resultado for igualmente provável, então $P(A) = |A|/36$ onde $|A|$ denota o número de elementos em A.
- A probabilidade de que a soma dos dados seja 11 é $2/36$, uma vez que existem dois resultados que correspondem a este evento.
- Se Ω for finito e cada resultado for igualmente provável, então

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

que é chamada de **distribuição uniforme** de probabilidade.

Eventos Independentes:

Eventos Independentes:

- Se jogarmos uma moeda justa duas vezes, a probabilidade de duas caras é $\frac{1}{2} \times \frac{1}{2}$.

Probabilidade

Eventos Independentes:

- Se jogarmos uma moeda justa duas vezes, a probabilidade de duas caras é $\frac{1}{2} \times \frac{1}{2}$.
- Multiplicamos porque consideramos as duas jogadas independentes.

Probabilidade

Eventos Independentes:

- Se jogarmos uma moeda justa duas vezes, a probabilidade de duas caras é $\frac{1}{2} \times \frac{1}{2}$.
- Multiplicamos porque consideramos as duas jogadas independentes.
- **Definição:**
  Dois eventos A e B são **independentes** se

  $$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$$

  Um conjunto de eventos $\{A_i : i \in I\}$ é independente se

  $$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

  para cada subconjunto finito J de I.

Fim

Obrigado pela atenção!