

Pré-Modelagem em Ciência de Dados

Prof. Rilder S. Pires

MBA em Ciência de Dados

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

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**.

Na aula passada...

Parte Teórica: Distribuições Contínuas

- ▶ Distribuição Uniforme
- ▶ Distribuição Exponencial
- ▶ Distribuição Normal ou Gaussiana
- ▶ Cálculo de Probabilidade em Distribuições Normais
- ▶ Funções Lineares de Distribuições Normais

Parte Prática:

- ▶ Cálculo de algumas distribuições

Distribuições de Probabilidade

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.
- ▶ O princípio de Pareto **não** se aplica sempre, mas serve como um “lembrete” de que a relação entre os inputs e os outputs não é balanceada.

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.
- ▶ O princípio de Pareto **não** se aplica sempre, mas serve como um “lembrete” de que a relação entre os inputs e os outputs não é balanceada.
- ▶ **Exemplos:**

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.
- ▶ O princípio de Pareto **não** se aplica sempre, mas serve como um “lembrete” de que a relação entre os inputs e os outputs não é balanceada.
- ▶ **Exemplos:**
 - ▶ 20% das pessoas concentram 80% da renda.

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.
- ▶ O princípio de Pareto **não** se aplica sempre, mas serve como um “lembrete” de que a relação entre os inputs e os outputs não é balanceada.
- ▶ **Exemplos:**
 - ▶ 20% das pessoas concentram 80% da renda.
 - ▶ 20% dos consumidores representam 80% das vendas.

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.
- ▶ O princípio de Pareto **não** se aplica sempre, mas serve como um “lembrete” de que a relação entre os inputs e os outputs não é balanceada.
- ▶ **Exemplos:**
 - ▶ 20% das pessoas concentram 80% da renda.
 - ▶ 20% dos consumidores representam 80% das vendas.
 - ▶ 20% dos empregados correspondem a 80% do total da produção.

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.
- ▶ O princípio de Pareto **não** se aplica sempre, mas serve como um “lembrete” de que a relação entre os inputs e os outputs não é balanceada.
- ▶ **Exemplos:**
 - ▶ 20% das pessoas concentram 80% da renda.
 - ▶ 20% dos consumidores representam 80% das vendas.
 - ▶ 20% dos empregados correspondem a 80% do total da produção.
- ▶ O fator 80/20 é apenas um exemplo, o princípio de Pareto pode se aplicar a situações com outras proporções (85/15 por exemplo).

Distribuições de Probabilidade

Princípio de Pareto:

- ▶ O princípio ou regra de Pareto afirma que 80% das consequências vêm de 20% das causas.
- ▶ Por isso, o princípio de Pareto também é conhecido como **regra 80/20**.
- ▶ O princípio de Pareto **não** se aplica sempre, mas serve como um “lembrete” de que a relação entre os inputs e os outputs não é balanceada.
- ▶ **Exemplos:**
 - ▶ 20% das pessoas concentram 80% da renda.
 - ▶ 20% dos consumidores representam 80% das vendas.
 - ▶ 20% dos empregados correspondem a 80% do total da produção.
- ▶ O fator 80/20 é apenas um exemplo, o princípio de Pareto pode se aplicar a situações com outras proporções (85/15 por exemplo).
- ▶ A regra de Pareto é aproximadamente descrita por uma **Lei de Potência**.

Distribuição de Pareto:

- ▶ A distribuição de Pareto é uma distribuição de probabilidade contínua dada por:

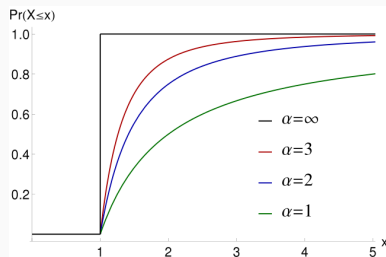
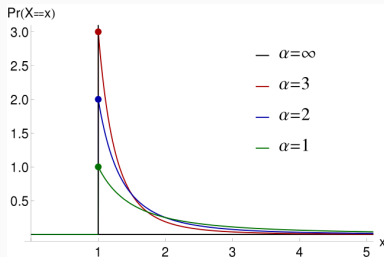
$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}$$

Distribuições de Probabilidade

Distribuição de Pareto:

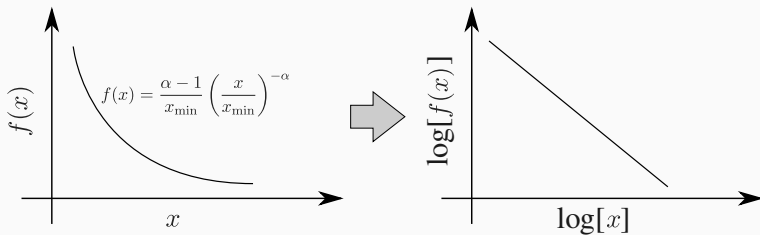
- ▶ A distribuição de Pareto é uma distribuição de probabilidade contínua dada por:

$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}$$



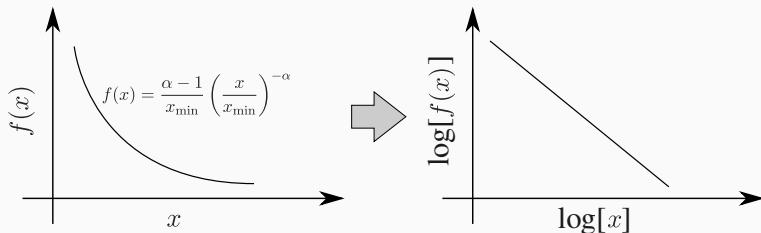
Distribuições de Probabilidade

Distribuição de Pareto:



Distribuições de Probabilidade

Distribuição de Pareto:



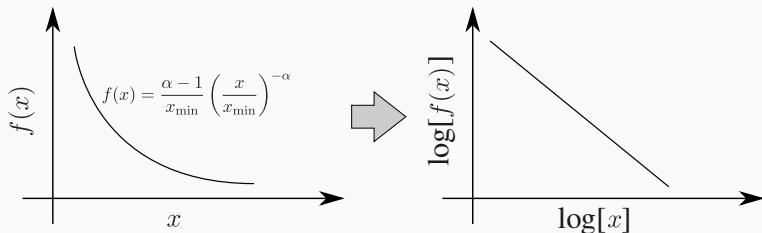
Veja que:

$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \Rightarrow \log[f(x)] = \log \left[\frac{\alpha x_m^\alpha}{x^{\alpha+1}} \right]$$

$$\log[f(x)] = \log [\alpha x_m^\alpha] - \log [x^{\alpha+1}]$$

Distribuições de Probabilidade

Distribuição de Pareto:



Veja que:

$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \Rightarrow \log[f(x)] = \log \left[\frac{\alpha x_m^\alpha}{x^{\alpha+1}} \right]$$

$$\log[f(x)] = \log [\alpha x_m^\alpha] - \log [x^{\alpha+1}]$$

então

$$\log[f(x)] = -(\alpha + 1) \log [x] + \log [\alpha x_m^\alpha]$$

Distribuições Bivariadas:

Distribuições Bivariadas:

- ▶ Dado um par de variáveis aleatórias discretas X e Y , definimos a **função distribuição de probabilidade conjunta** por $f(x, y) = \mathbb{P}(X = x, Y = y)$.

Distribuições de Probabilidade

Distribuições Bivariadas:

- ▶ Dado um par de variáveis aleatórias discretas X e Y , definimos a **função distribuição de probabilidade conjunta** por $f(x, y) = \mathbb{P}(X = x, Y = y)$.

Distribuições Multivariadas:

- ▶ Seja $X = (X_1, \dots, X_n)$ onde X_1, \dots, X_n são variáveis aleatórias. Chamamos X de **vetor aleatório**.

Distribuições de Probabilidade

Distribuições Bivariadas:

- ▶ Dado um par de variáveis aleatórias discretas X e Y , definimos a **função distribuição de probabilidade conjunta** por $f(x, y) = \mathbb{P}(X = x, Y = y)$.

Distribuições Multivariadas:

- ▶ Seja $X = (X_1, \dots, X_n)$ onde X_1, \dots, X_n são variáveis aleatórias. Chamamos X **de vetor aleatório**.
- ▶ Seja $f(X_1, \dots, X_n)$ a função distribuição de probabilidade. Dizemos que X_1, \dots, X_n são independentes se, para todo A_1, \dots, A_n ,

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

Distribuições de Probabilidade

Distribuições Bivariadas:

- ▶ Dado um par de variáveis aleatórias discretas X e Y , definimos a **função distribuição de probabilidade conjunta** por $f(x, y) = \mathbb{P}(X = x, Y = y)$.

Distribuições Multivariadas:

- ▶ Seja $X = (X_1, \dots, X_n)$ onde X_1, \dots, X_n são variáveis aleatórias. Chamamos X **de vetor aleatório**.
- ▶ Seja $f(X_1, \dots, X_n)$ a função distribuição de probabilidade. Dizemos que X_1, \dots, X_n são independentes se, para todo A_1, \dots, A_n ,

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

- ▶ É suficiente verificar que $f(X_1, \dots, X_n) = \prod_{i=1}^n f(x_i)$

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!