

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...

Revisão: Distribuições

- ▶ Variável Discreta
- ▶ Variável Contínua

Parte Teórica: Distribuições Discretas

- ▶ Distribuição Uniforme
- ▶ Distribuição de Bernoulli
- ▶ Distribuição Binomial
- ▶ Distribuição de Poisson
- ▶ Distribuição Geométrica
- ▶ Distribuição Hipergeométrica

Parte Prática:

- ▶ Exploração dos Dados da PAM
- ▶ Cálculo de algumas distribuições
- ▶ Apresentação dos dados do PIB

Distribuições Contínuas

Distribuição Uniforme:

- ▶ X segue uma distribuição uniforme, escrita como

$$X \sim \text{Uniforme}(a, b)$$

se

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } x \in [a, b] \\ 0 & \text{caso contrário} \end{cases}$$

Distribuição Uniforme:

- ▶ X segue uma distribuição uniforme, escrita como

$$X \sim \text{Uniforme}(a, b)$$

se

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } x \in [a, b] \\ 0 & \text{caso contrário} \end{cases}$$

- ▶ e a distribuição cumulativa é

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

Distribuição Exponencial:

- ▶ X segue uma distribuição Exponencial com parâmetro β , escrita com

$$X \sim \text{Exp}(\beta),$$

se

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$$

onde $\beta > 0$.

Distribuição Exponencial:

- ▶ X segue uma distribuição Exponencial com parâmetro β , escrita com

$$X \sim \text{Exp}(\beta),$$

se

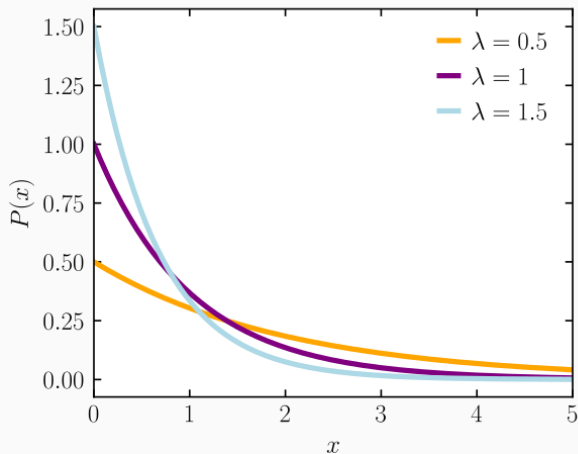
$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$$

onde $\beta > 0$.

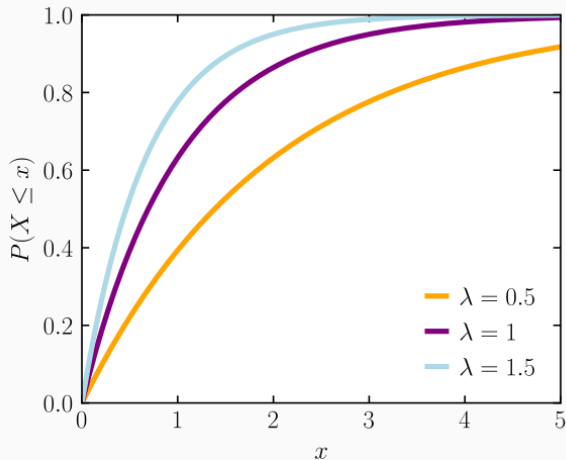
- ▶ A distribuição exponencial é usada para modelar os tempos de vida dos componentes eletrônicos e os tempos de espera entre eventos raros.

Distribuições Contínuas

Distribuição Exponencial:

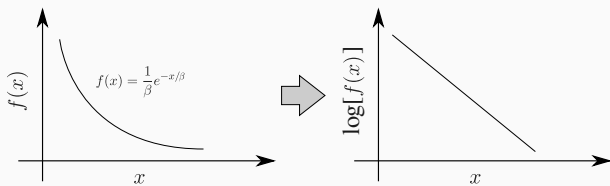


Distribuição Exponencial:



Distribuições Contínuas

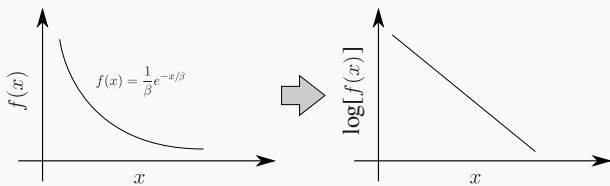
Distribuição Exponencial:



Veja que:

Distribuições Contínuas

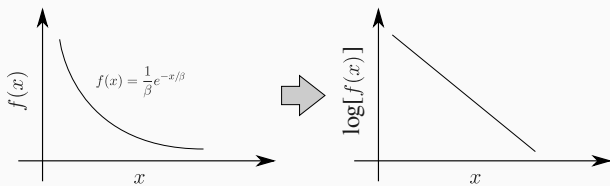
Distribuição Exponencial:



Veja que:

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \Rightarrow \log[f(x)] = \log \left[\frac{1}{\beta} e^{-x/\beta} \right]$$

Distribuição Exponencial:



Veja que:

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \Rightarrow \log[f(x)] = \log \left[\frac{1}{\beta} e^{-x/\beta} \right]$$

$$\log[f(x)] = \log \left[\frac{1}{\beta} \right] + \log \left[e^{-x/\beta} \right] \Rightarrow \log[f(x)] = -\frac{x}{\beta} + \log \left[\frac{1}{\beta} \right]$$

Distribuição Normal ou Gaussiana:

- ▶ X segue uma distribuição Normal (ou Gaussiana) com parâmetros μ e σ , denotada por

$$X \sim N(\mu, \sigma^2),$$

se

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}, x \in \mathbb{R}$$

onde $\mu \in \mathbb{R}$ e $\sigma > 0$.

Distribuição Normal ou Gaussiana:

- ▶ X segue uma distribuição Normal (ou Gaussiana) com parâmetros μ e σ , denotada por

$$X \sim N(\mu, \sigma^2),$$

se

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}, x \in \mathbb{R}$$

onde $\mu \in \mathbb{R}$ e $\sigma > 0$.

- ▶ O parâmetro μ é a média da distribuição e σ é o desvio padrão.

Distribuição Normal ou Gaussiana:

- ▶ X segue uma distribuição Normal (ou Gaussiana) com parâmetros μ e σ , denotada por

$$X \sim N(\mu, \sigma^2),$$

se

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}, x \in \mathbb{R}$$

onde $\mu \in \mathbb{R}$ e $\sigma > 0$.

- ▶ O parâmetro μ é a média da distribuição e σ é o desvio padrão.
- ▶ O Normal desempenha um papel importante na estatística.

Distribuição Normal ou Gaussiana:

- ▶ X segue uma distribuição Normal (ou Gaussiana) com parâmetros μ e σ , denotada por

$$X \sim N(\mu, \sigma^2),$$

se

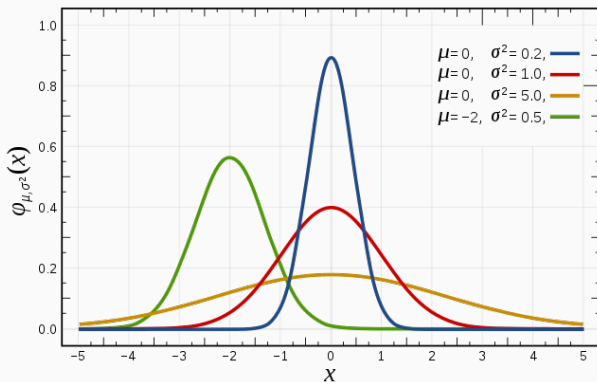
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}, x \in \mathbb{R}$$

onde $\mu \in \mathbb{R}$ e $\sigma > 0$.

- ▶ O parâmetro μ é a média da distribuição e σ é o desvio padrão.
- ▶ O Normal desempenha um papel importante na estatística.
- ▶ Muitos fenômenos na natureza têm distribuições aproximadamente normais.

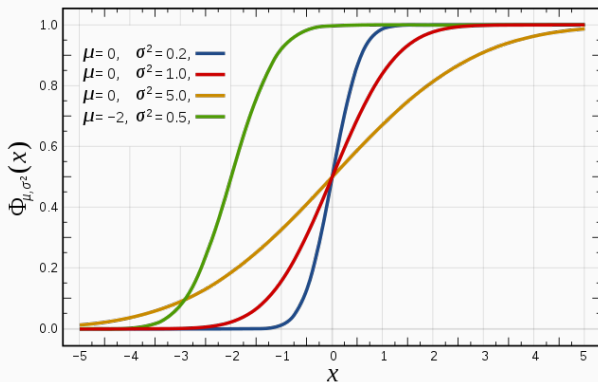
Distribuições Contínuas

Distribuição Normal ou Gaussiana:



Distribuições Contínuas

Distribuição Normal ou Gaussiana:



Cálculo de Probabilidade em Distribuições Normais:

- Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.
- ▶ A tradição dita que uma variável aleatória Normal padrão é denotada por Z .

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.
- ▶ A tradição dita que uma variável aleatória Normal padrão é denotada por Z .
- ▶ O $f(x)$ e o $F(x)$ de uma Normal padrão são indicados por $\phi(z)$ e $\Phi(z)$.

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.
- ▶ A tradição dita que uma variável aleatória Normal padrão é denotada por Z .
- ▶ O $f(x)$ e o $F(x)$ de uma Normal padrão são indicados por $\phi(z)$ e $\Phi(z)$.
- ▶ Não há expressão de forma fechada para Φ .

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.
- ▶ A tradição dita que uma variável aleatória Normal padrão é denotada por Z .
- ▶ O $f(x)$ e o $F(x)$ de uma Normal padrão são indicados por $\phi(z)$ e $\Phi(z)$.
- ▶ Não há expressão de forma fechada para Φ .
- ▶ Aqui estão alguns fatos úteis:

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.
- ▶ A tradição dita que uma variável aleatória Normal padrão é denotada por Z .
- ▶ O $f(x)$ e o $F(x)$ de uma Normal padrão são indicados por $\phi(z)$ e $\Phi(z)$.
- ▶ **Não há expressão de forma fechada para Φ .**
- ▶ Aqui estão alguns fatos úteis:
 - ▶ Se $X \sim N(\mu, \sigma^2)$, então $Z = (X - \mu)/\sigma \sim N(0, 1)$.

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.
- ▶ A tradição dita que uma variável aleatória Normal padrão é denotada por Z .
- ▶ O $f(x)$ e o $F(x)$ de uma Normal padrão são indicados por $\phi(z)$ e $\Phi(z)$.
- ▶ **Não há expressão de forma fechada para Φ .**
- ▶ Aqui estão alguns fatos úteis:
 - ▶ Se $X \sim N(\mu, \sigma^2)$, então $Z = (X - \mu)/\sigma \sim N(0, 1)$.
 - ▶ Se $Z \sim N(0, 1)$, então $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Dizemos que X tem uma **Distribuição Normal Padrão** se $\mu = 0$ e $\sigma = 1$.
- ▶ A tradição dita que uma variável aleatória Normal padrão é denotada por Z .
- ▶ O $f(x)$ e o $F(x)$ de uma Normal padrão são indicados por $\phi(z)$ e $\Phi(z)$.
- ▶ **Não há expressão de forma fechada para Φ .**
- ▶ Aqui estão alguns fatos úteis:
 - ▶ Se $X \sim N(\mu, \sigma^2)$, então $Z = (X - \mu)/\sigma \sim N(0, 1)$.
 - ▶ Se $Z \sim N(0, 1)$, então $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.
 - ▶ Dessa forma:

$$\mathbb{P}(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Exemplo:

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Exemplo:
- ▶ Distribuição normal com $\mu = 50$ e $\sigma = 20$

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Exemplo:
- ▶ Distribuição normal com $\mu = 50$ e $\sigma = 20$
- ▶ Quanto vale $\mathbb{P}(40 < X < 60)$?

Cálculo de Probabilidade em Distribuições Normais:

- ▶ Exemplo:
- ▶ Distribuição normal com $\mu = 50$ e $\sigma = 20$
- ▶ Quanto vale $\mathbb{P}(40 < X < 60)$?
- ▶ Aplicando:

$$\mathbb{P}(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

obtemos:

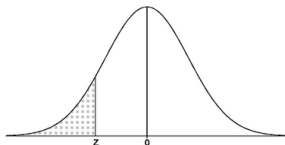
$$\mathbb{P}(40 < X < 60) = \Phi\left(\frac{60 - 50}{20}\right) - \Phi\left(\frac{40 - 50}{20}\right) = \Phi(1/2) - \Phi(-1/2)$$

$$\mathbb{P}(40 < X < 60) = (1.0 - 0.3085) - 0.3085 = 0.383$$

Distribuições Contínuas

Cálculo de Probabilidade em Distribuições Normais:

Cumulative Standard Normal Distribution Table



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.00	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
-0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.20	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.40	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985

Funções Lineares de Distribuições Normais:

- ▶ Seja Y uma função linear de uma variável X distribuída de forma Normal

$$Y = aX + b \quad \Rightarrow \quad Y = a(\mu + \sigma Z) + b$$

Funções Lineares de Distribuições Normais:

- ▶ Seja Y uma função linear de uma variável X distribuída de forma Normal

$$Y = aX + b \quad \Rightarrow \quad Y = a(\mu + \sigma Z) + b$$

- ▶ então

$$Y = (a\mu + b) + (a\sigma)Z \quad \Rightarrow \quad Y = \mu_y + \sigma_y Z$$

Funções Lineares de Distribuições Normais:

- ▶ Seja Y uma função linear de uma variável X distribuída de forma Normal

$$Y = aX + b \quad \Rightarrow \quad Y = a(\mu + \sigma Z) + b$$

- ▶ então

$$Y = (a\mu + b) + (a\sigma)Z \quad \Rightarrow \quad Y = \mu_y + \sigma_y Z$$

- ▶ onde

$$\mu_y = a\mu + b \quad \text{e} \quad \sigma_y = a\sigma$$

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!