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

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**.

No módulo passado...

Aula 1:

- Revisão: Estatística Básica
- Parte Teórica: Probabilidade
- Parte Prática: Exemplos, Apresentação dos Dados

Aula 2:

- Parte Teórica: Probabilidade e Variáveis Aleatórias
- Parte Prática: Exemplos, Exploração dos Dados

Aula 3:

- Parte Teórica: Variáveis Aleatórias e Introdução a Distribuições
- Parte Prática: Exemplos, Exploração dos Dados

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.

Revisão

Distribuição Cumulativa

Funções de distribuição e Funções de probabilidade:

Distribuição Cumulativa

Funções de distribuição e Funções de probabilidade:

Função de Distribuição Cumulativa:

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.

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.

Definição:

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.

Definição:

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

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

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.

Definição:

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

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

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

Variável Discreta

Função de Probabilidade:

Variável Discreta

Função de Probabilidade:

Definição:

Variável Discreta

Função de Probabilidade:

Definição:

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

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

Variável Discreta

Função de Probabilidade:

Definição:

- X é **discreta** se receber valores contáveis $\{x_1, x_2, ...\}$. 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 $\sum_i f_X(x_i) = 1$.

Variável Discreta

Função de Probabilidade:

Definição:

- X é **discreta** se receber valores contáveis $\{x_1, x_2, ...\}$. 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

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

Variável Contínua

Função Densidade de Probabilidade:

Variável Contínua

Função Densidade de Probabilidade:

Definição:

Variável Contínua

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 $a \leq b$,

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

Variável Contínua

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,

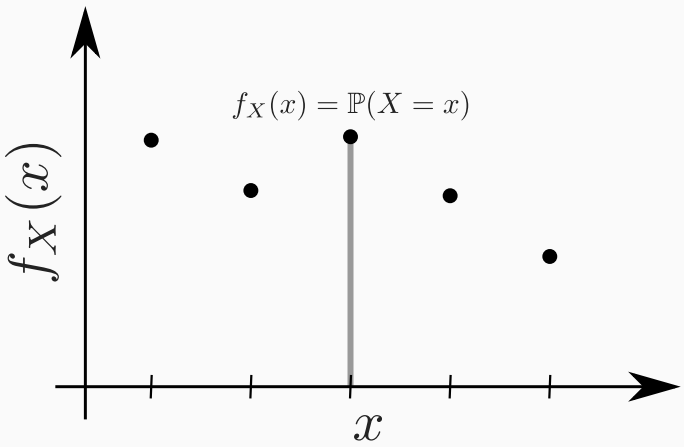
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

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

Variável Discreta × Variável Contínua

Variável Discreta

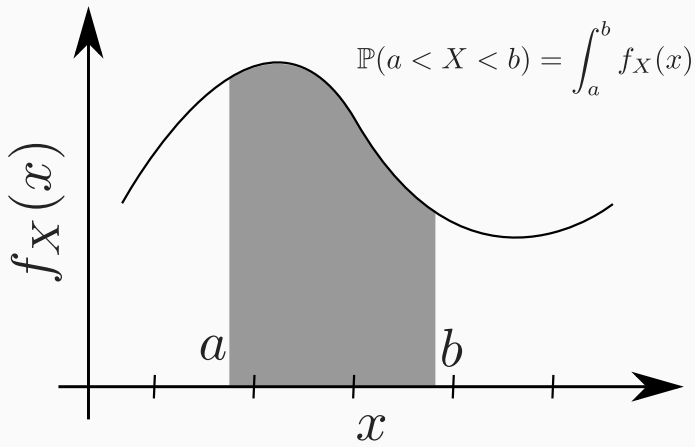
Função de Probabilidade:



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

Variável Contínua

Função Densidade de Probabilidade:



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

Distribuições Discretas

Distribuição Uniforme:

- Seja $k > 1$ um dado inteiro.

Distribuição Uniforme:

- Seja $k > 1$ um dado inteiro.
- Suponha que X tenha a função de massa de probabilidade dada por

$$f(x) = \begin{cases} 1/k & \text{para } x = 1, \ldots, k \\ 0 & \text{caso contrário} \end{cases}$$

Distribuições Discretas

Distribuição Uniforme:

- Seja $k > 1$ um dado inteiro.
- Suponha que X tenha a função de massa de probabilidade dada por

$$f(x) = \begin{cases} 1/k & \text{para } x = 1, \ldots, k \\ 0 & \text{caso contrário} \end{cases}$$

- Dizemos que X segue um distribuição de uniforme em $\{1, \ldots, k\}$;

Distribuições Discretas

Distribuição de Bernoulli:

- Seja X a representação de um lançamento de uma moeda.

Distribuições Discretas

Distribuição de Bernoulli:

- Seja X a representação de um lançamento de uma moeda.
- Onde

$$\mathbb{P}(X = 1) = p$$

e

$$\mathbb{P}(X = 0) = 1 - p$$

para algum $p \in [0, 1]$.

Distribuições Discretas

Distribuição de Bernoulli:

- Seja X a representação de um lançamento de uma moeda.
- Onde

$$\mathbb{P}(X = 1) = p$$

e

$$\mathbb{P}(X = 0) = 1 - p$$

para algum $p \in [0, 1]$.

- Dizemos que X segue um distribuição de Bernoulli

$$X \sim \text{Bernoulli}(p)$$

Distribuições Discretas

Distribuição de Bernoulli:

- Seja X a representação de um lançamento de uma moeda.
- Onde

$$\mathbb{P}(X = 1) = p$$

e

$$\mathbb{P}(X = 0) = 1 - p$$

para algum $p \in [0, 1]$.

- Dizemos que X segue um distribuição de Bernoulli

$$X \sim \text{Bernoulli}(p)$$

- A função de probabilidade nesse caso é dada por

$$f(x) = p^x (1 - p)^{1-x}$$

para

$$x \in \{0, 1\}$$

.

Distribuições Discretas

Distribuição Binomial:

- Suponha que temos uma moeda que cai cara com probabilidade p para $0 \leq p \leq 1$. **Jogue a moeda n vezes e deixe X ser o número de caras.**

Distribuições Discretas

Distribuição Binomial:

- Suponha que temos uma moeda que cai cara com probabilidade p para $0 \leq p \leq 1$. **Jogue a moeda n vezes e deixe X ser o número de caras.**
- Supondo que os lançamentos sejam independentes e que

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

seja a função de massa. Pode ser mostrado que

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, \ldots, n \\ 0 & \text{caso contrário} \end{cases}$$

Distribuições Discretas

Distribuição Binomial:

- Suponha que temos uma moeda que cai cara com probabilidade p para $0 \leq p \leq 1$. **Jogue a moeda n vezes e deixe X ser o número de caras.**
- Supondo que os lançamentos sejam independentes e que

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

seja a função de massa. Pode ser mostrado que

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, \ldots, n \\ 0 & \text{caso contrário} \end{cases}$$

- Uma variável aleatória com esta função de massa é chamada de variável aleatória binomial e escrevemos

$$X \sim \text{Binomial}(n, p).$$

Distribuições Discretas

Distribuição Binomial:

- Suponha que temos uma moeda que cai cara com probabilidade p para $0 \leq p \leq 1$. **Jogue a moeda n vezes e deixe X ser o número de caras.**
- Supondo que os lançamentos sejam independentes e que

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

seja a função de massa. Pode ser mostrado que

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, \ldots, n \\ 0 & \text{caso contrário} \end{cases}$$

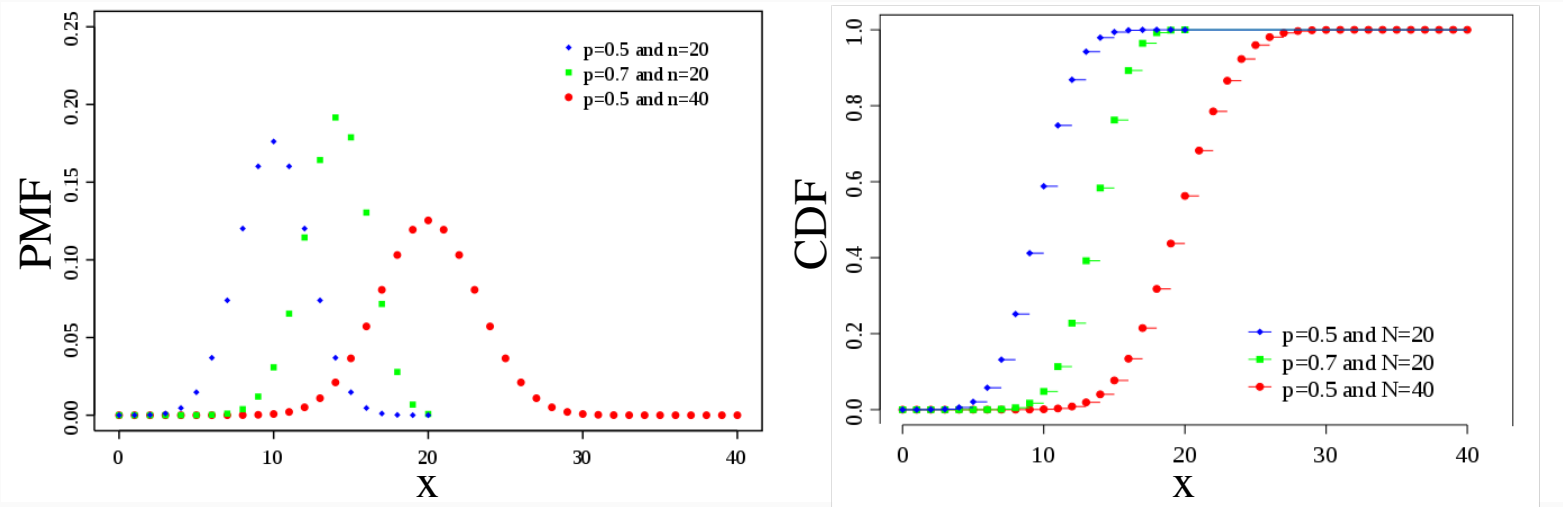
- Uma variável aleatória com esta função de massa é chamada de variável aleatória binomial e escrevemos

$$X \sim \text{Binomial}(n, p).$$

- Se $X_1 \sim \text{Binomial}(n_1, p)$ e $X_2 \sim \text{Binomial}(n_2, p)$ então $X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p)$.

Distribuições Discretas

Distribuição de Binomial:



Distribuição de Poisson:

- X tem uma distribuição de Poisson com parâmetro λ, escrita como

$$X \sim \text{Poisson}(\lambda)$$

se

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad x \geq 0$$

Distribuição de Poisson:

- X tem uma distribuição de Poisson com parâmetro λ, escrita como

$$X \sim \text{Poisson}(\lambda)$$

se

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad x \geq 0$$

- A distribuição de Poisson é frequentemente usada como modelo para contagens de eventos raros, como decaimento radioativo e acidentes de trânsito.

Distribuição de Poisson:

- X tem uma distribuição de Poisson com parâmetro λ, escrita como

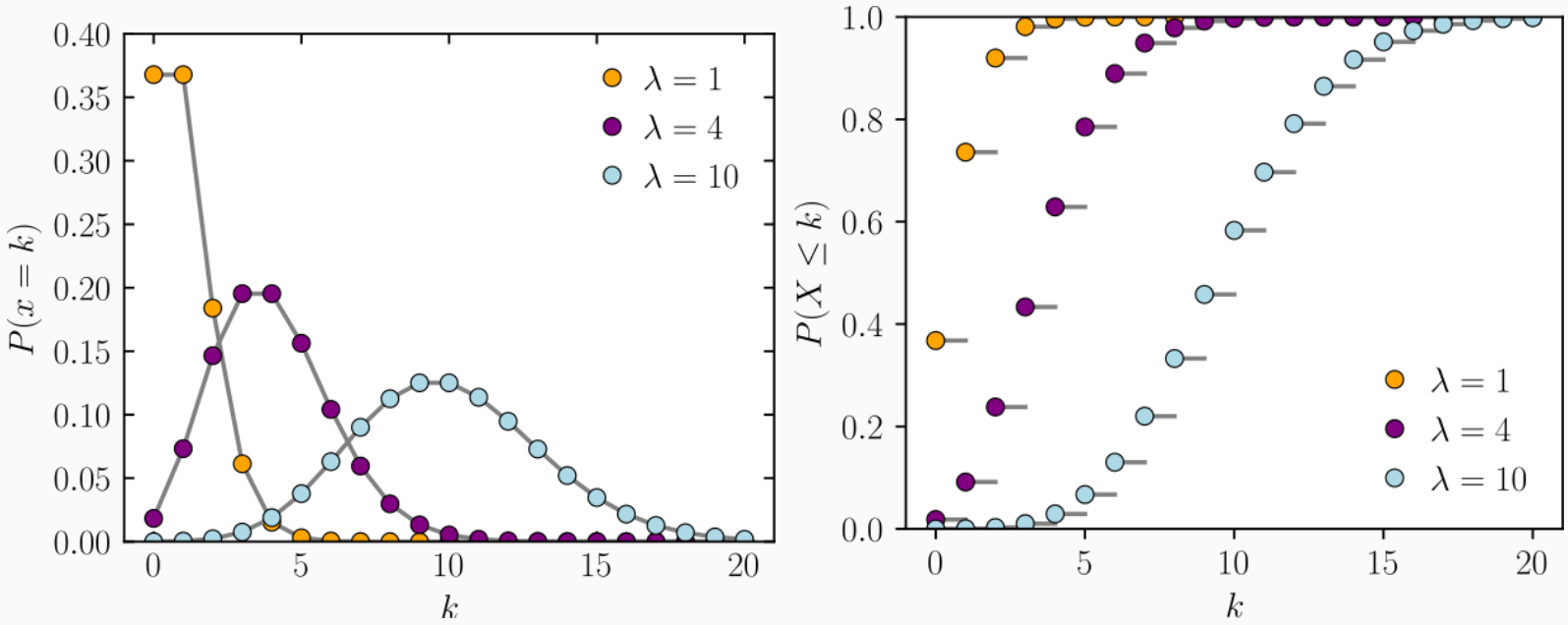
$$X \sim \text{Poisson}(\lambda)$$

se

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad x \geq 0$$

- A distribuição de Poisson é frequentemente usada como modelo para contagens de eventos raros, como decaimento radioativo e acidentes de trânsito.
- Se $X_1 \sim \text{Poisson}(\lambda_1)$ e $X_2 \sim \text{Poisson}(\lambda_2)$ então $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Distribuição de Poisson:



Distribuição Geométrica:

- X segue uma distribuição Geométrica com parâmetro $p \in (0,1)$, escrita como $X \sim \text{Geom}(p)$ se

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

Distribuições Discretas

Distribuição Geométrica:

- X segue uma distribuição Geométrica com parâmetro $p \in (0,1)$, escrita como $X \sim \text{Geom}(p)$ se

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

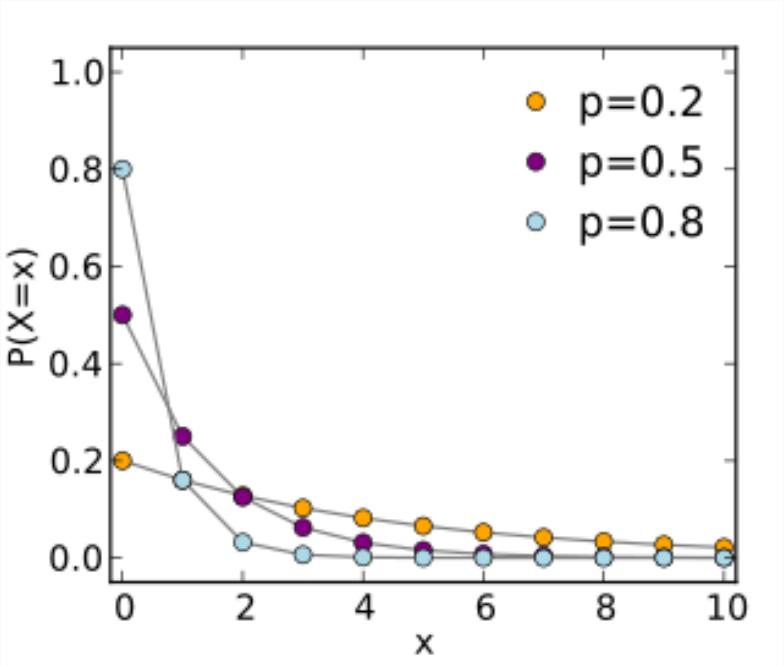
- Exemplo: X é o número de jogadas necessárias para obter a primeira cara quando jogamos uma moeda.

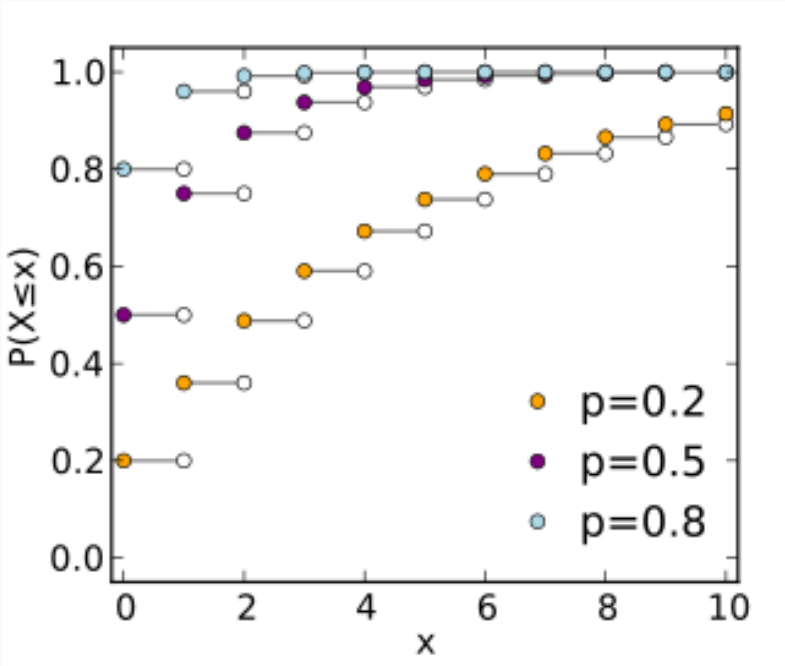
Distribuição Geométrica:

- X segue uma distribuição Geométrica com parâmetro $p \in (0,1)$, escrita como $X \sim \text{Geom}(p)$ se

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

- Exemplo: X é o número de jogadas necessárias para obter a primeira cara quando jogamos uma moeda.





Distribuições Discretas

Distribuição Hipergeométrica:

- X segue uma distribuição Hipergeométrica se

$$\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \geq 1.$$

Distribuições Discretas

Distribuição Hipergeométrica:

- X segue uma distribuição Hipergeométrica se

$$\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \geq 1.$$

- Representa a probabilidade de k sucessos em n jogadas, sem reposição, de uma população N que contem exatamente K objetos com a caracteristica desejada.

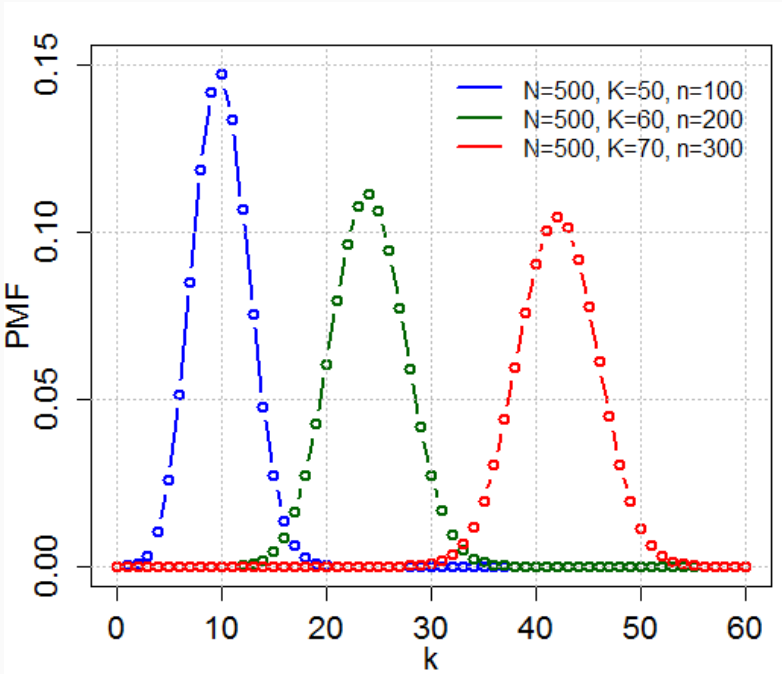
Distribuições Discretas

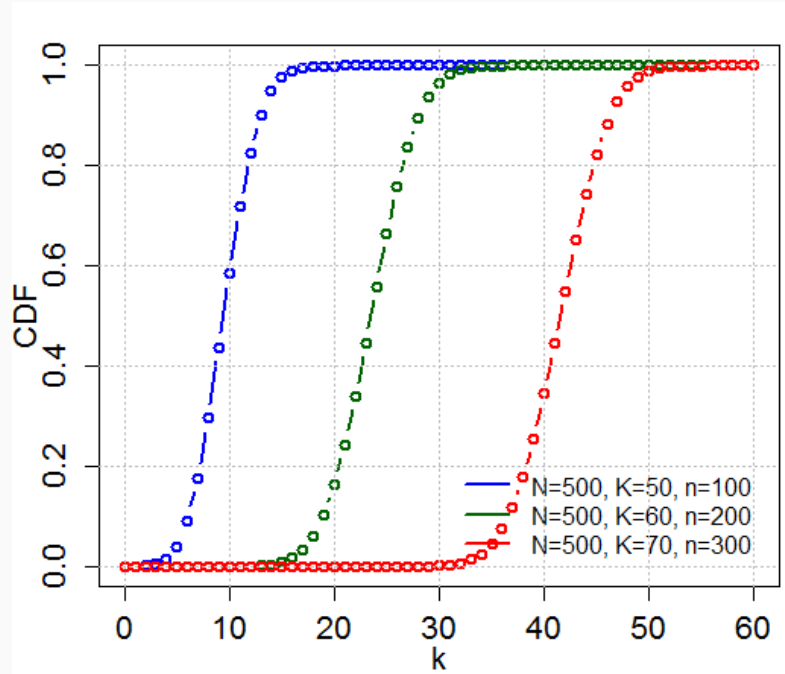
Distribuição Hipergeométrica:

- X segue uma distribuição Hipergeométrica se

$$\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \geq 1.$$

- Representa a probabilidade de k sucessos em n jogadas, sem reposição, de uma população N que contem exatamente K objetos com a caracteristica desejada.





Projeto Final:

Projeto Final

Projeto Final:

Perguntas

Projeto Final

Projeto Final:

Perguntas

1. Qual a distribuição da “diversidade” dos 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?

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?

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

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!