# Simulação monte Carlo aplicada ao Transporte Aéreo

*Prof. Marcelo Xavier Guterres*

*Fonte: Modelagem e Simulação de Eventos Discretos: Teoria e Aplicações - Leonardo Chwif*

# Coleta e Modelagem dos Dados de Entrada

*Este material é disponibilizado para uso* ***exclusivo*** *de docentes que adotam o livro* ***Modelagem e Simulação de Eventos Discretos*** *em suas disciplinas. O material pode (e deve) ser editado pelo professor.*

*Pedimos apenas que seja sempre* ***citada*** *a fonte original de consulta.*

CHWIF, Leonardo; MEDINA, Afonso Celso. Modelagem e simulação de eventos discretos. Afonso C. Medina, 2006.

# Três Etapas

- ***Coleta***
- ***Tratamento***
- ***Inferência***

# Coleta dos Dados

1. Escolha adequada da variável de estudo

2. O tamanho da amostra deve estar entre 100 e 200 observações. Amostras com menos de 100 observações podem comprometer a identificação do melhor modelo probabilístico, e amostras com mais de 200 observações não trazem ganhos significativos ao estudo;

# Coleta dos Dados

3. Coletar e anotar as observações **na mesma ordem** em que o fenômeno está ocorrendo, para permitir a análise de correlação ;

4. Se existe alguma suspeita de que os dados mudam em função do horário ou do dia da coleta, a **coleta deve ser refeita** para outros horários e dias. Na modelagem de dados, vale a regra: **toda suspeita deve ser comprovada ou descartada estatisticamente.**

# Exemplo: Check-in

Um gerente de cia aérea está preocupado com as filas formadas nos balcões de pagamento durante um dos turnos de operação. Quais seriam as variáveis de estudo para coleta de dados?

Os tempos de atendimento nos balcões

Os tempos entre chegadas sucessivas de clientes nos caixas de pagamento

*O número de clientes em fila*

***É resultado!!***

# Exemplo: Coleta de Dados

Intervalo entre chegadas de pessoas nos balcões de atendimento (100 medidas). Tempos em minutos:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 11 | 5 | 2 | 0 | 9 | 9 | 1 | 5 | 5 | 1 |
| 1 | 3 | 3 | 3 | 7 | 4 | 12 | 8 | 7 | 5 |
| 5 | 2 | 6 | 1 | 11 | 1 | 2 | 4 | 4 | 2 |
| 2 | 1 | 3 | 9 | 0 | 10 | 3 | 3 | 4 | 5 |
| 1 | 5 | 18 | 4 | 22 | 8 | 3 | 0 | 4 | 4 |
| 8 | 9 | 2 | 3 | 12 | 1 | 3 | 1 | 11 | 9 |
| 7 | 5 | 14 | 7 | 7 | 28 | 1 | 3 | 3 | 4 |
| 2 | 11 | 13 | 2 | 0 | 1 | 6 | 12 | 8 | 12 |
| 15 | 0 | 6 | 7 | 19 | 1 | 1 | 9 | 12 | 4 |
| 1 | 5 | 3 | 17 | 10 | 15 | 43 | 2 | 9 | 11 |
| 6 | 1 | 13 | 13 | 19 | 10 | 9 | 20 | 17 | 24 |
| 19 | 2 | 27 | 5 | 20 | 5 | 10 | 8 | 728 | 8 |
| 2 | 3 | 1 | 1 | 4 | 3 | 6 | 13 | 12 | 12 |
| 10 | 9 | 1 | 1 | 3 | 9 | 9 | 4 | 6 | 3 |
| 0 | 3 | 6 | 3 | 27 | 3 | 18 | 4 | 4 | 7 |
| 6 | 0 | 2 | 2 | 8 | 4 | 5 | 1 | 3 | 1 |
| 4 | 18 | 1 | 0 | 16 | 20 | 2 | 2 | 9 | 3 |
| 2 | 12 | 28 | 0 | 7 | 3 | 18 | 12 | 2 | 1 |
| 3 | 2 | 8 | 3 | 19 | 12 | 5 | 4 | 0 | 3 |
| 6 | 0 | 5 | 0 | 3 | 7 | 0 | 8 | 5 | 8 |

# Exemplo: Medidas de Posição e Dispersão

| | | |
|---|---|---|
| ***Medidas de posição*** | *Média* | *10,44* |
| | *Mediana* | *5* |
| | *Moda* | *3* |
| | *Mínimo* | *0* |
| | ***Máximo*** | ***728*** |
| ***Medidas de dispersão*** | ***Amplitude*** | ***728*** |
| | *Desvio padrão* | *51,42* |
| | ***Variância da amostra*** | ***2.643,81*** |
| | ***Coeficiente de Variação*** | ***493%*** |
| | *Coeficiente Assimetria* | *13,80* |

**O 728 é um outlier?**

# Exemplo: Outlier

Intervalo entre chegadas de pessoas nos balcões de atendimento (100 medidas). Tempos em minutos

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 11 | 5 | 2 | 0 | 9 | 9 | 1 | 5 | 5 | 1 |
| 1 | 3 | 3 | 3 | 7 | 4 | 12 | 8 | 7 | 5 |
| 5 | 2 | 6 | 1 | 11 | 1 | 2 | 4 | 4 | 2 |
| 2 | 1 | 3 | 9 | 0 | 10 | 3 | 3 | 4 | 5 |
| 1 | 5 | 18 | 4 | 22 | 8 | 3 | 0 | 4 | 4 |
| 8 | 9 | 2 | 3 | 12 | 1 | 3 | 1 | 11 | 9 |
| 7 | 5 | 14 | 7 | 7 | 28 | 1 | 3 | 3 | 4 |
| 2 | 11 | 13 | 2 | 0 | 1 | 6 | 12 | 8 | 12 |
| 15 | 0 | 6 | 7 | 19 | 1 | 1 | 9 | 12 | 4 |
| 1 | 5 | 3 | 17 | 10 | 15 | 43 | 2 | 9 | 11 |
| 6 | 1 | 13 | 13 | 19 | 10 | 9 | 20 | 17 | 24 |
| 19 | 2 | 27 | 5 | 20 | 5 | 10 | 8 | 728 | 8 |
| 2 | 3 | 1 | 1 | 4 | 3 | 6 | 13 | 12 | 12 |
| 10 | 9 | 1 | 1 | 3 | 9 | 9 | 4 | 6 | 3 |
| 0 | 3 | 6 | 3 | 27 | 3 | 18 | 4 | 4 | 7 |
| 6 | 0 | 2 | 2 | 8 | 4 | 5 | 1 | 3 | 1 |
| 4 | 18 | 1 | 0 | 16 | 20 | 2 | 2 | 9 | 3 |
| 2 | 12 | 28 | 0 | 7 | 3 | 18 | 12 | 2 | 1 |
| 3 | 2 | 8 | 3 | 19 | 12 | 5 | 4 | 0 | 3 |
| 6 | 0 | 5 | 0 | 3 | 7 | 0 | 8 | 5 | 8 |

# *Outliers* ou Valores Discrepantes

- **Erro na coleta de dados.** Este tipo de outlier é o mais comum, principalmente quando o levantamento de dados é feito por meio manual.

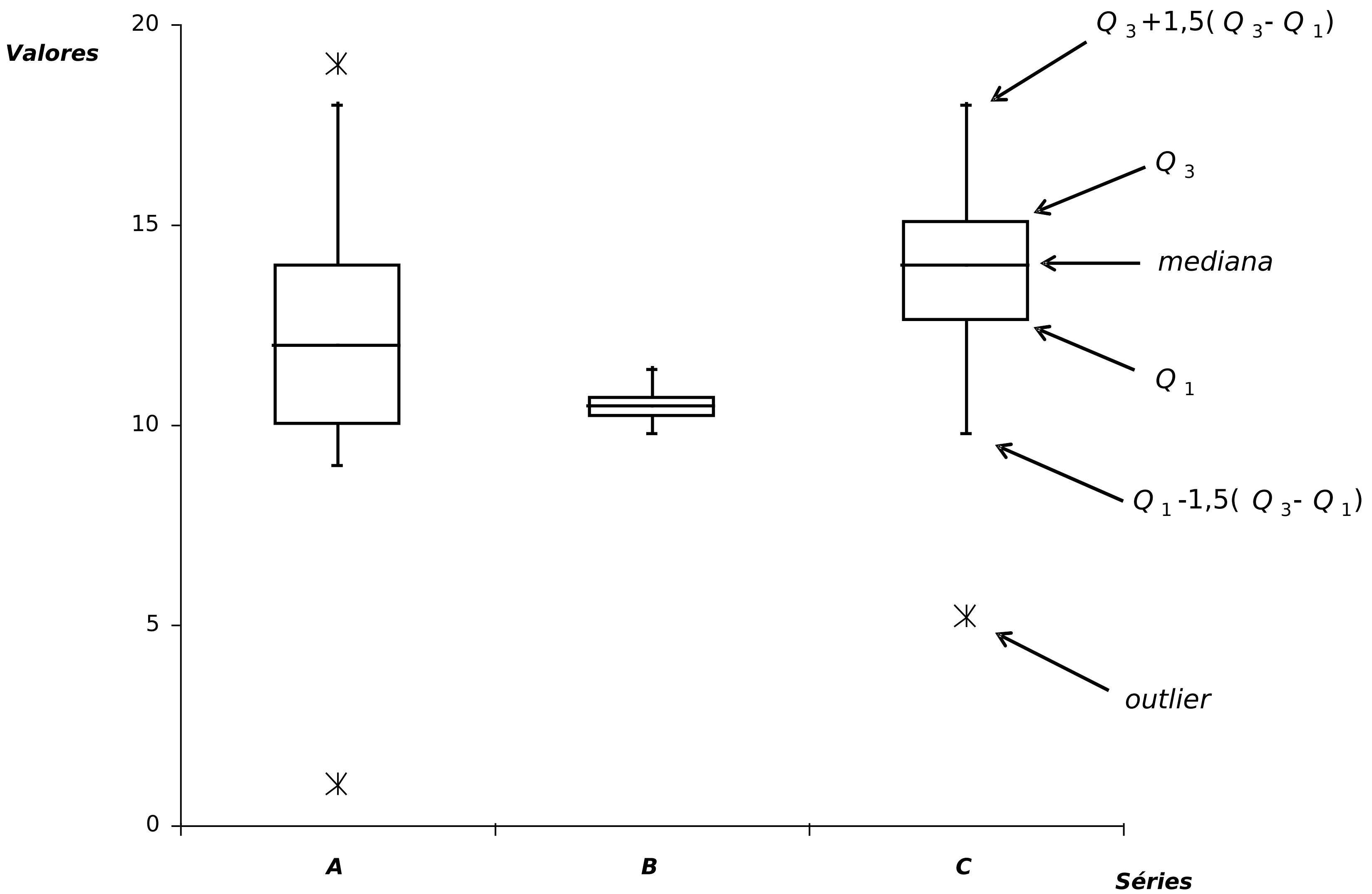
- **Eventos Raros.** Nada impede que situações totalmente atípicas ocorram na nossa coleta de dados. Alguns exemplos:

  - ✓ Um dia de temperatura negativa no verão da cidade do Rio de Janeiro;
  - ✓ Um tempo de execução de um operador ser muito curto em relação aos melhores desempenhos obtidos naquela tarefa;
  - ✓ Um tempo de viagem de um caminhão de entregas na cidade de São Paulo, durante o horário de rush, ser muito menor do que fora deste horário.

# Exemplo: Outlier (valor discrepante)

| | Dados | |
|---|---|---|
| | com o outlier | sem o outlier |
| **Média** | 10,44 | 6,83 |
| **Mediana** | 5 | 5 |
| **Variância da amostra** | **2.643,81** | **43,60** |

# Identificação de Outliers: Box-plot



# Análise de Correlação

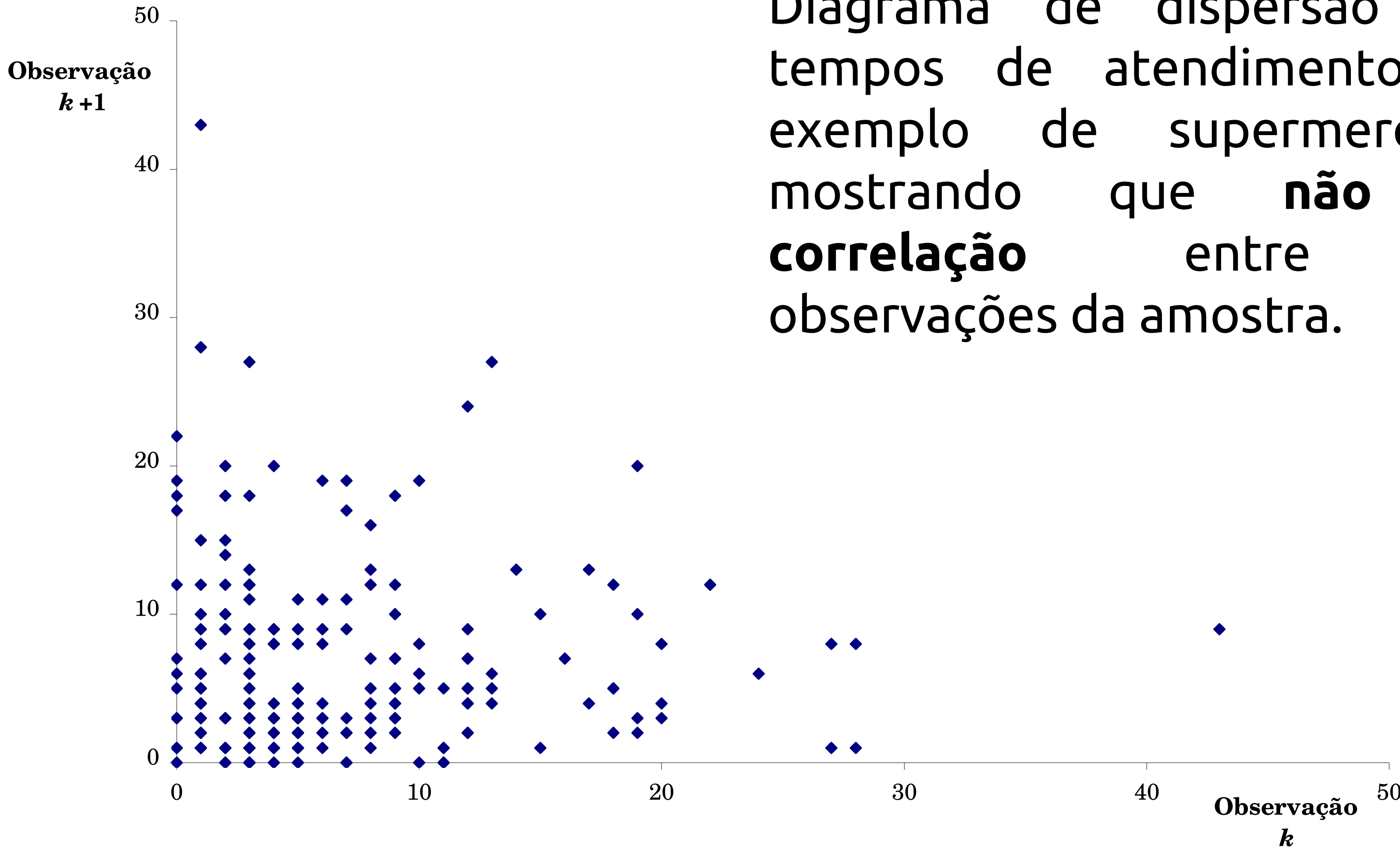



Diagrama de dispersão dos tempos de atendimento do exemplo de supermercado, mostrando que **não há correlação** entre as observações da amostra.

# Análise de Correlação

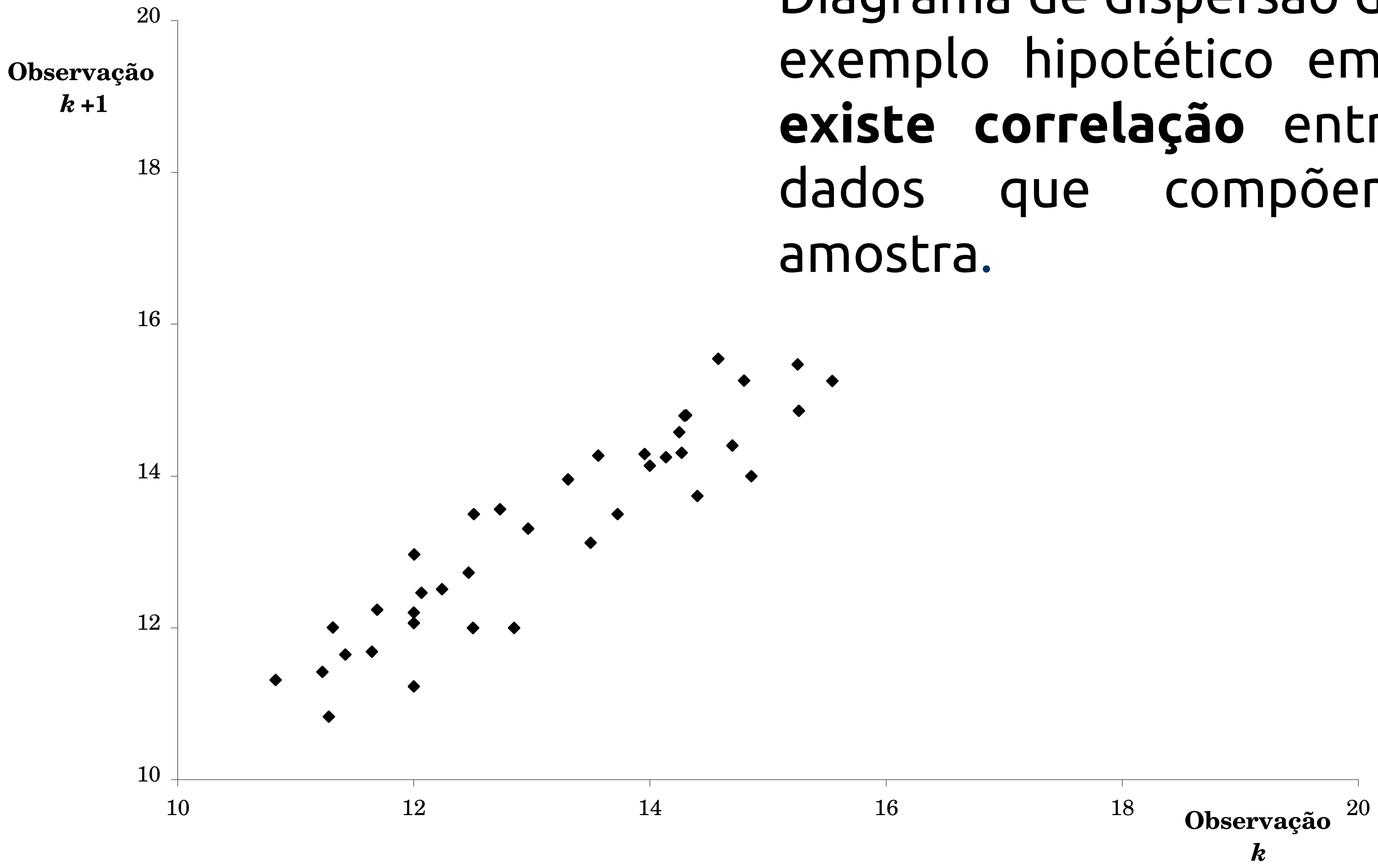



Diagrama de dispersão de um exemplo hipotético em que **existe correlação** entre os dados que compõem a amostra.

# Exemplo: Construção do Histograma

O **histograma** é utilizado para identificar qual a distribuição a ser ajustada aos dados coletados ou é utilizado diretamente dentro do modelo de simulação.

**1. Definir o número de classes:**

$$K = 1 + 3{,}3\log_{10} n$$

$$K = \sqrt{n}$$

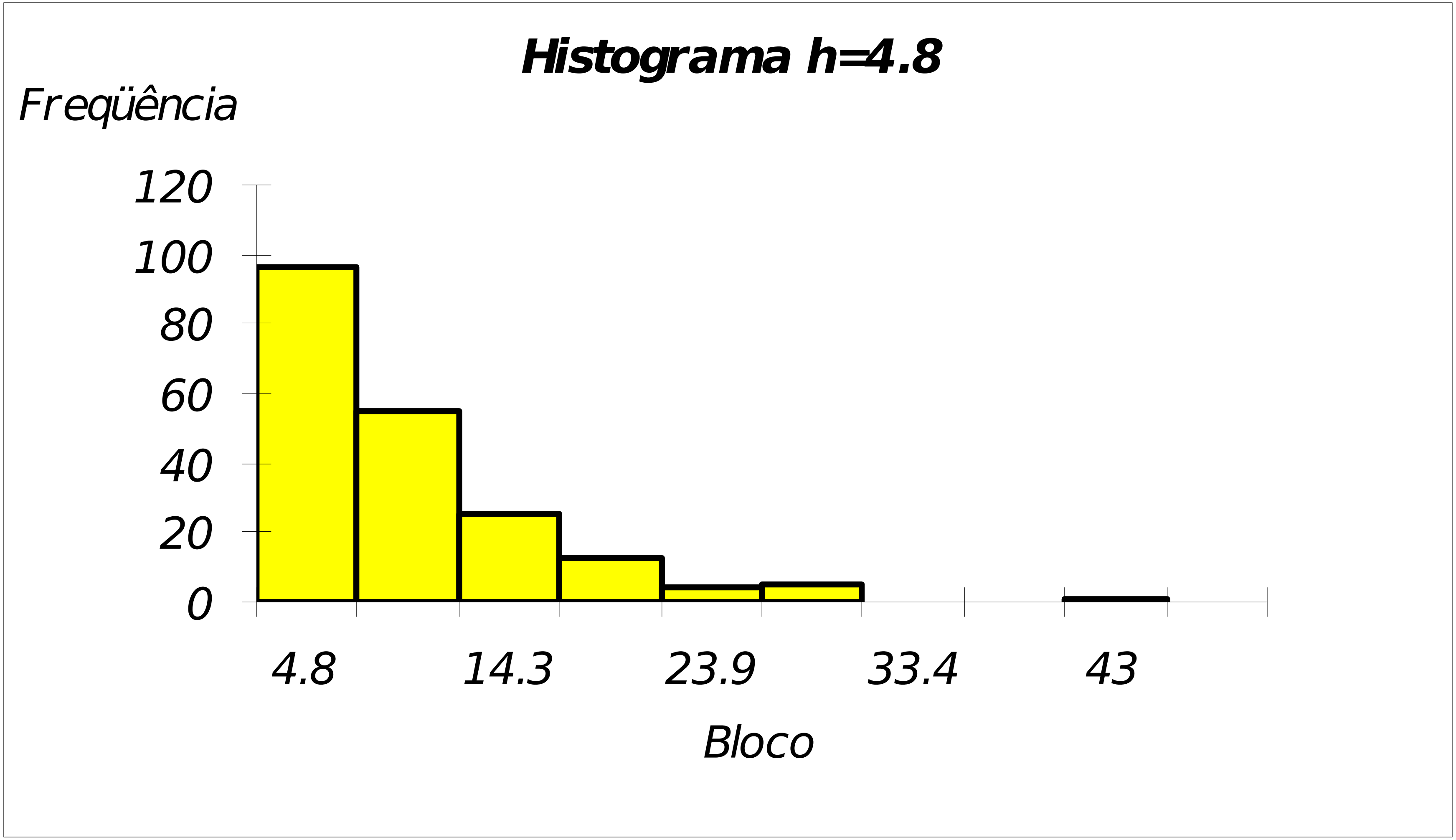
**2. Definir o tamanho do intervalo:**

$$h = \frac{Amplitude}{K}$$

**3. Construir a tabela de frequências**

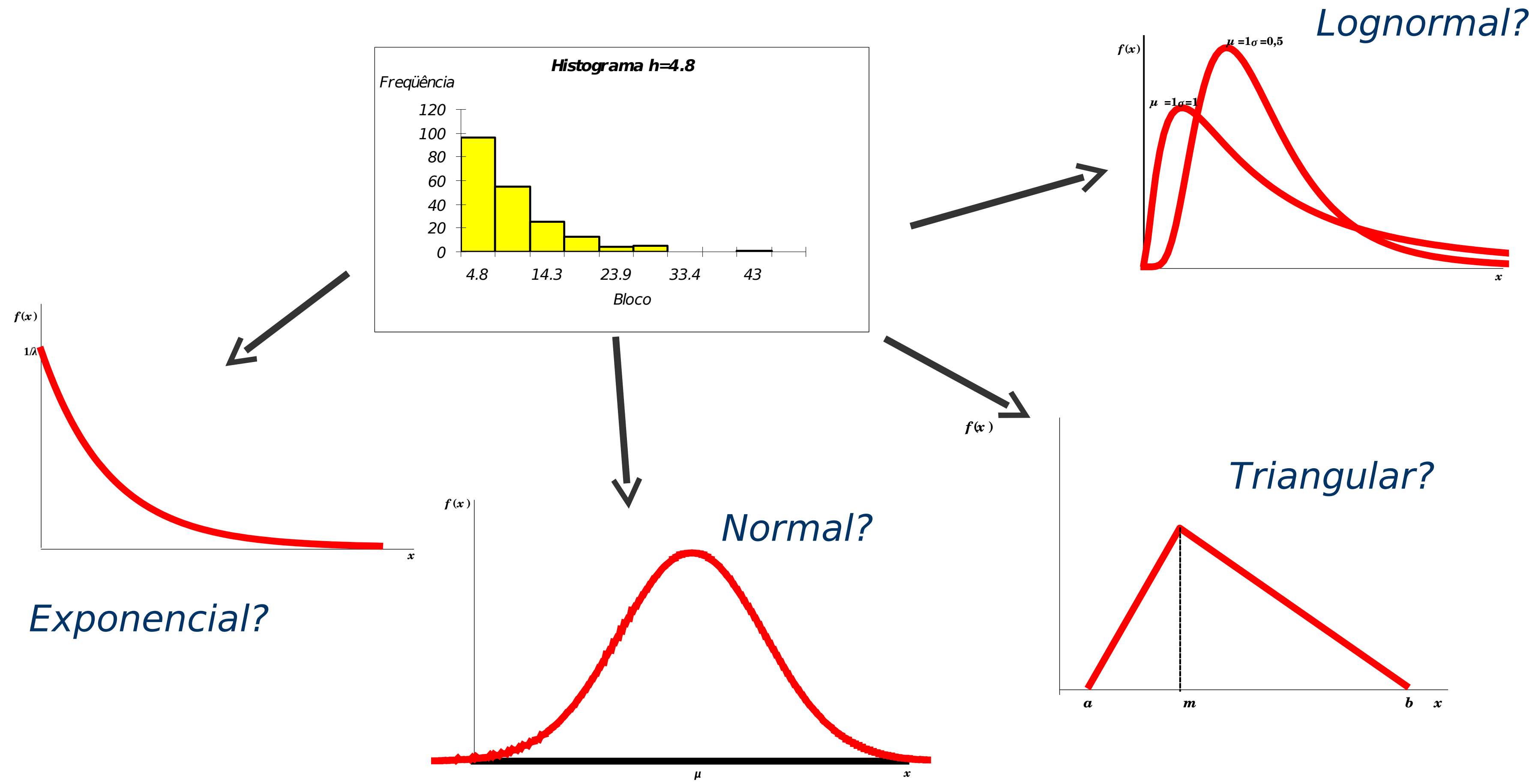
**4. Construir o histograma**

# Exemplo: Histograma



# Exemplo: Inferência

Qual o melhor modelo probabilístico ou distribuição estatística que pode representar a amostra coletada?



# Testes de Aderência (não paramétricos)

Testa a **validade ou não** da hipótese de aderência (ou hipótese nula) em confronto com a hipótese alternativa:

- **$H_0$:** o modelo **é adequado** para representar a distribuição da população.
- **$H_a$:** o modelo **não é adequado** para representar a distribuição da população.

Se a um dado nível de significância (alfa)100% rejeitarmos $H_0$, o modelo testado **não é adequado** para representar a distribuição da população. O nível de significância alfa equivale à probabilidade de rejeitarmos a hipótese nula $H_0$, dado que ela está correta. Testes usuais:

- **Qui quadrado**
- **Kolmogorov-Sminov**

# Teste do Qui-quadrado

| ***Limites*** | | ***Freqüências*** | | | |
|---|---|---|---|---|---|
| ***Inf*** | ***Sup*** | ***Exponencial*** | ***Teórica (T)*** | ***Observada (O)*** | ***(O-T)^2/T*** |
| *0* | *4.8* | *0.5022* | *100* | *96* | *0.16* |
| *4.8* | *9.6* | *0.2500* | *50* | *55* | *0.55* |
| *9.6* | *14.3* | *0.1244* | *25* | *25* | *0.00* |
| *14.3* | *19.1* | *0.0620* | *12* | *13* | *0.04* |
| *19.1* | *1.0E+10* | *0.0614* | *12* | *10* | *0.40* |
| | | | | ***E*** | ***1.15*** |

| | |
|---|---|
| ***Confiança*** | *5%* |
| ***Graus de liberdade*** | *3* |

| | |
|---|---|
| ***Valor Teórico*** | ***7.81*** |

| | |
|---|---|
| ***p-value*** | ***0.76*** |

| | | |
|---|---|---|
| ***Portanto,*** | ***não rejeitamos*** | *a hipótese de que os dados aderem ao modelo exponencial* |

# P-value

Parâmetro usual nos softwares de estatística. Para o teste do qui-quadrado no Excel, utilizar:

**=DIST.QUI (valor de E; graus de liberdade)**

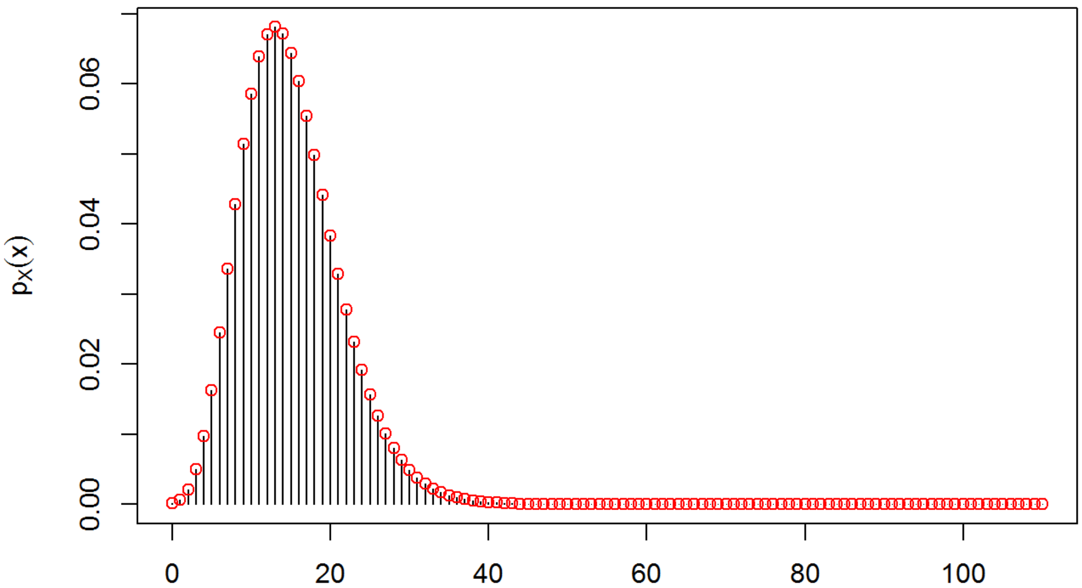
| ***Valor*** | ***Critério*** |
|---|---|
| *p-value<0,01* | *Evidência **forte** contra a hipótese de aderência* |
| *0,01≤p-value<0,05* | *Evidência **moderada** contra a hipótese de aderência* |
| *0,05≤p-value<0,10* | *Evidência **potencial** contra a hipótese de aderência* |
| *0,10≤p-value* | *Evidência **fraca ou inexistente** contra a hipótese de aderência* |

# Distribuições discretas: Binomial

Seja $X$ o número de sucessos obtidos na realização de $n$ ensaios de Bernoulli independentes. Diremos que $X$ tem distribuição binomial com parâmetros $n$ e $p$, em que $p$ é a probabilidade de sucesso em cada ensaio, se sua função de probabilidade for dada por

$$p(x) = \mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Usaremos a notação $X \sim b(n, p)$.

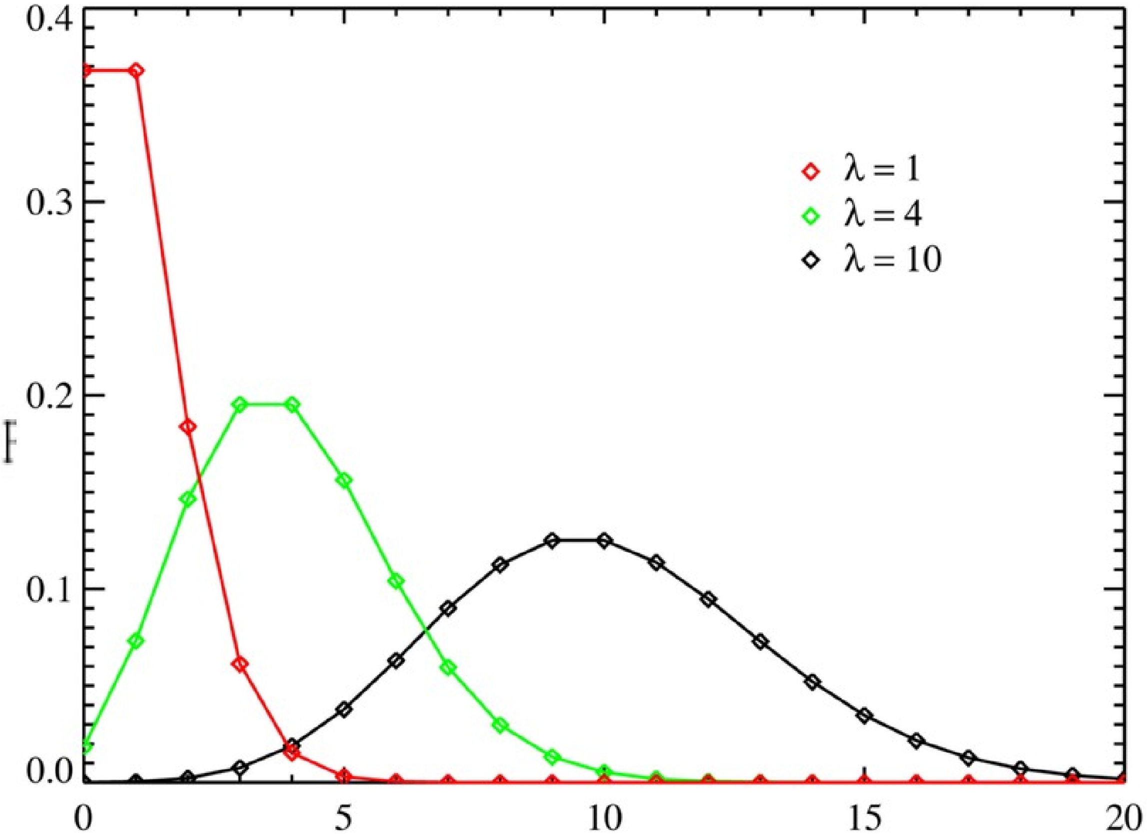



Exemplo de Binomial negativa com r = 10 e p = 0.4

# Distribuições discretas: Poisson

Uma variável aleatória discreta $X$ segue a distribuição de Poisson com parâmetro $\lambda$, $\lambda > 0$, se sua função de probabilidade for dada por

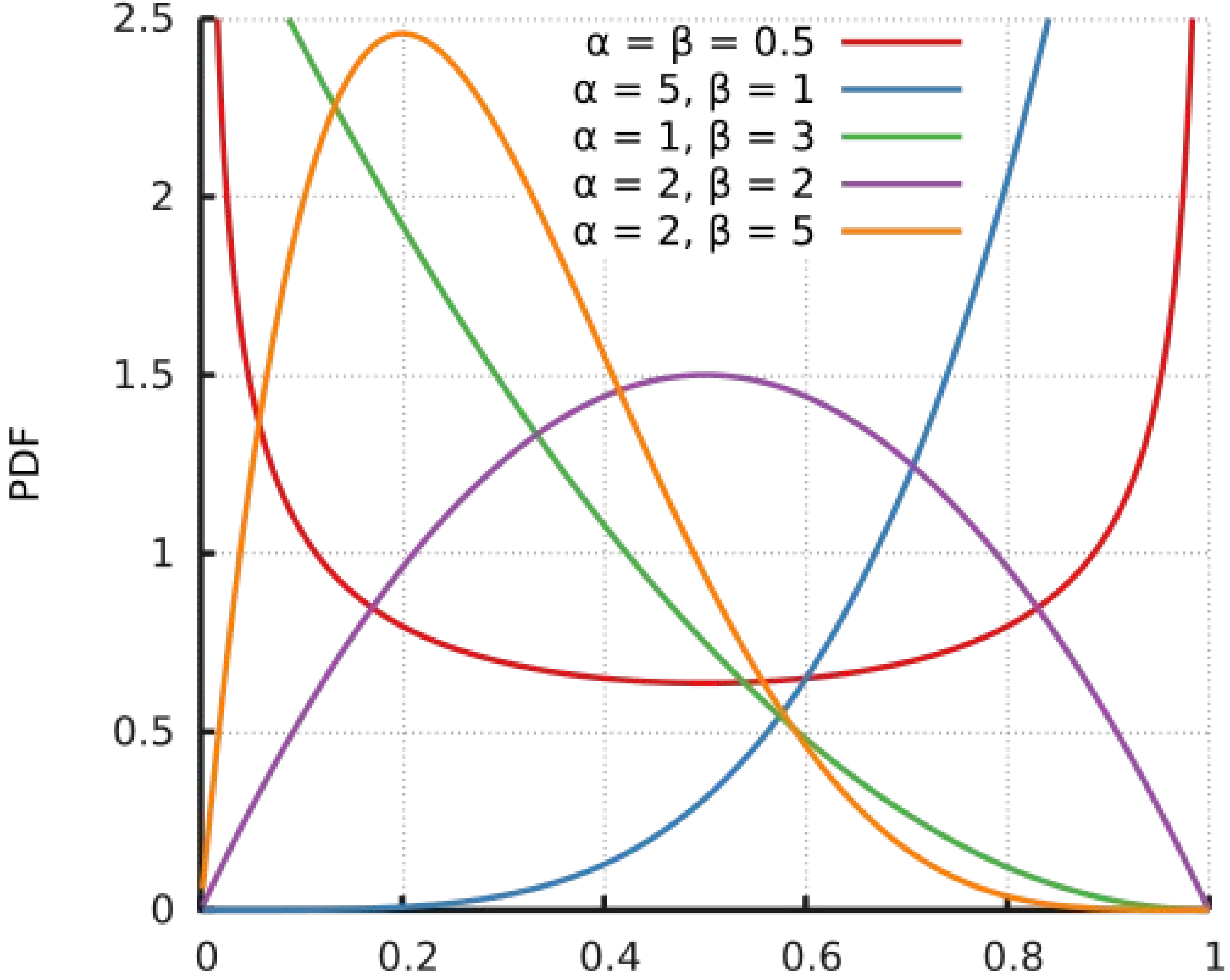
$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$



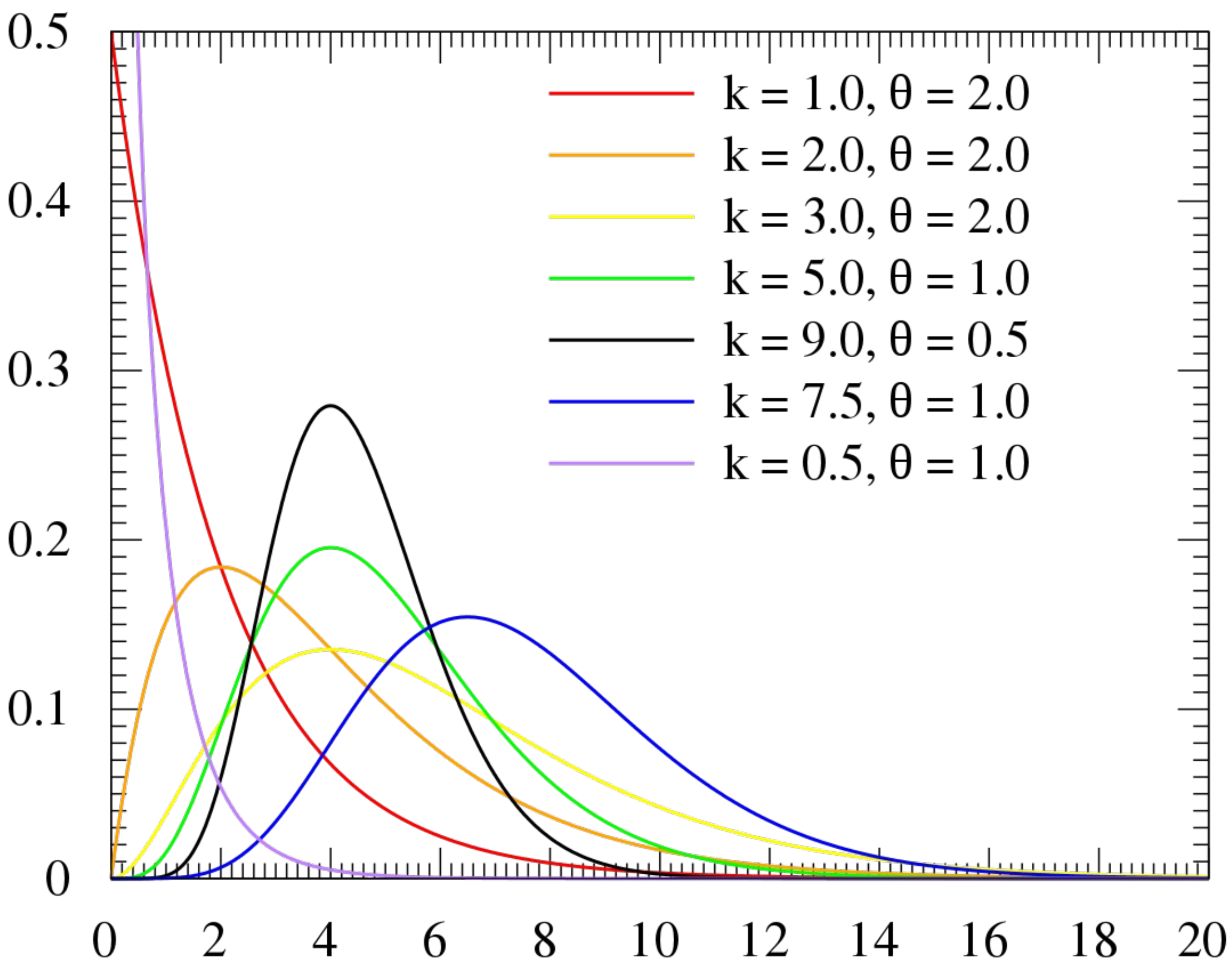

Utilizamos a notação $X \sim$ Poisson$(\lambda)$ ou $X \sim$ P

por unidade medida.

# Distribuições contínuas: Beta



# Distribuições contínuas: Erlang

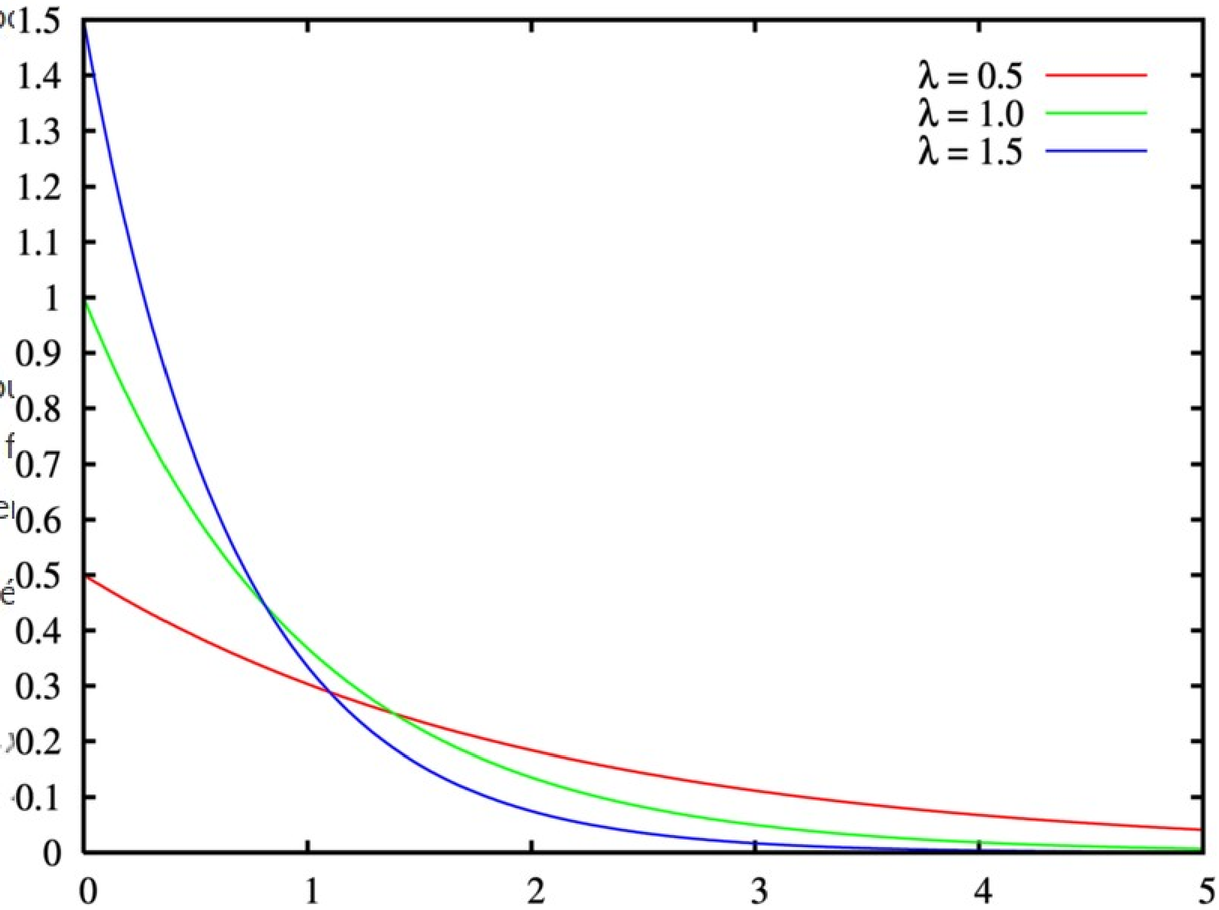


# Distribuições contínuas: Exponencial

A variável aleatória $X$ tem distribuição Exponencial com parâmetro $\lambda$, $\lambda > 0$, se tiver

função densidade de probabilidade dada p

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \text{ se } x \geq 0 \\ 0 \text{ se } x < 0 \end{cases}$$




em que $\lambda$ é o parâmetro de taxa da distribu

o tempo médio de vida e $x$ é um tempo de f

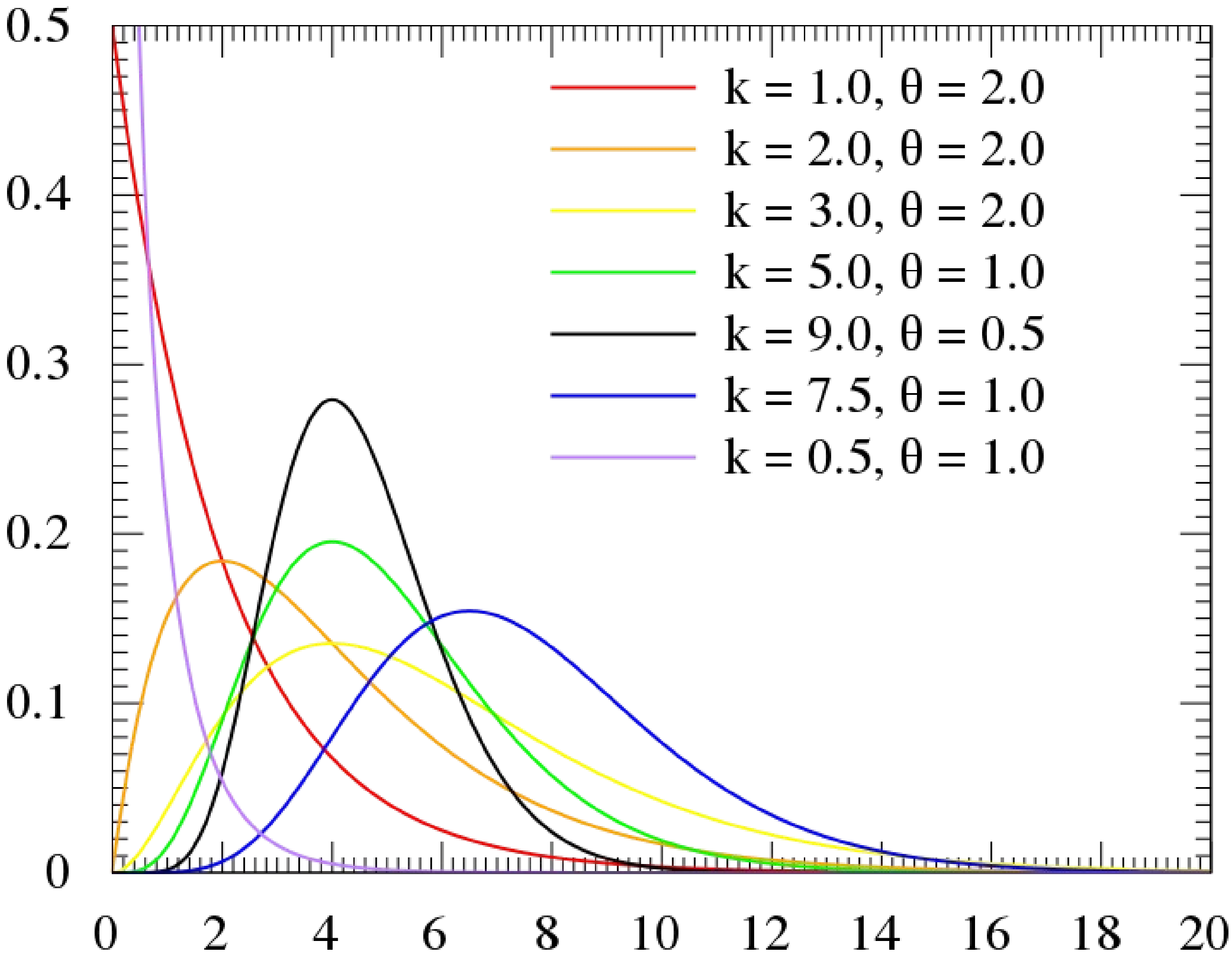
do tempo da falha $x$. Isto é, se $x$ é medido e

A função de distribuição acumulada $F(x)$ é

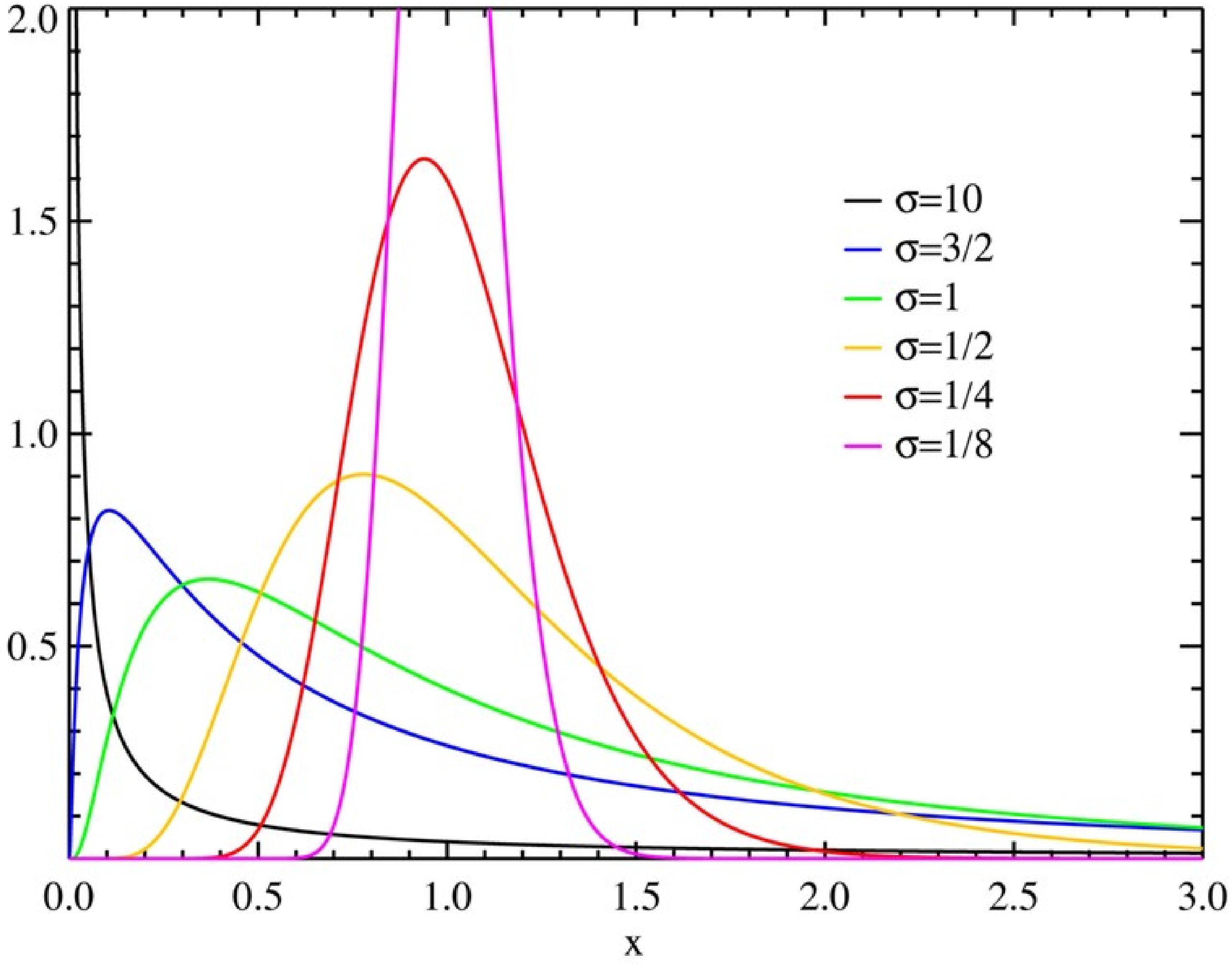
$$F(x) = \int_0^x f(s)ds = \begin{cases} 1 - e^{-\lambda} \\ 0 \text{ se } x \end{cases}$$

Utilizamos a notação $X \sim \text{Exp}(\lambda)$.

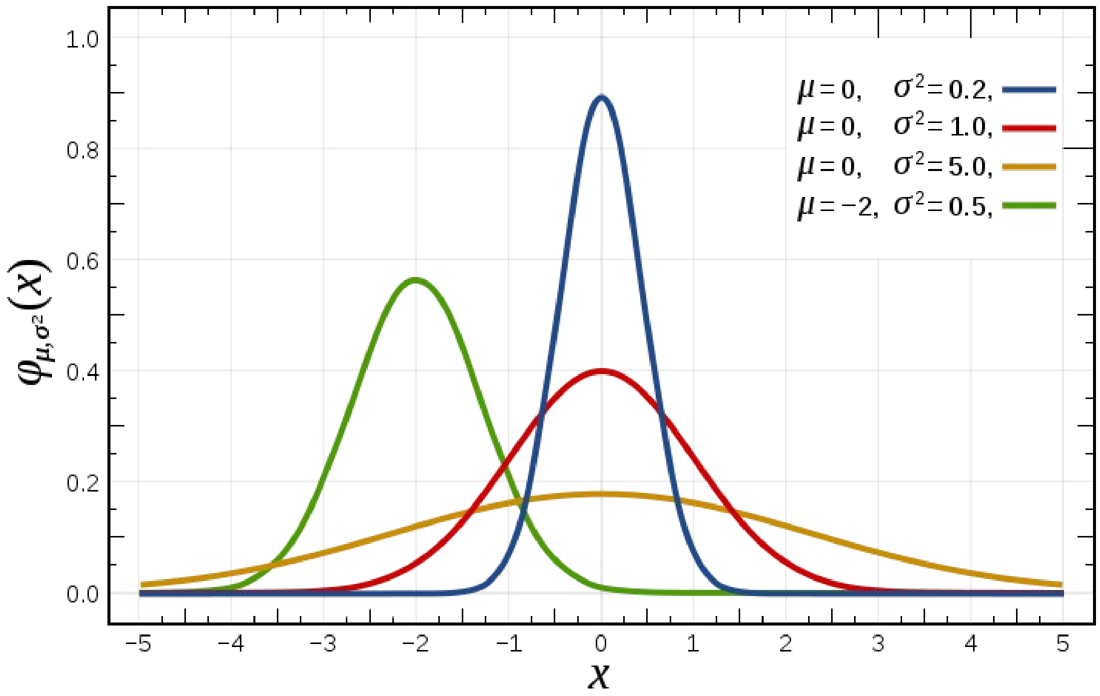
# Distribuições contínuas: Gama

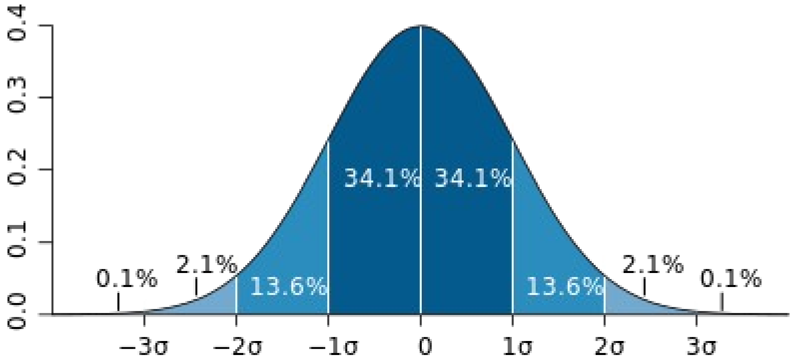


# Distribuições contínuas: Lognormal

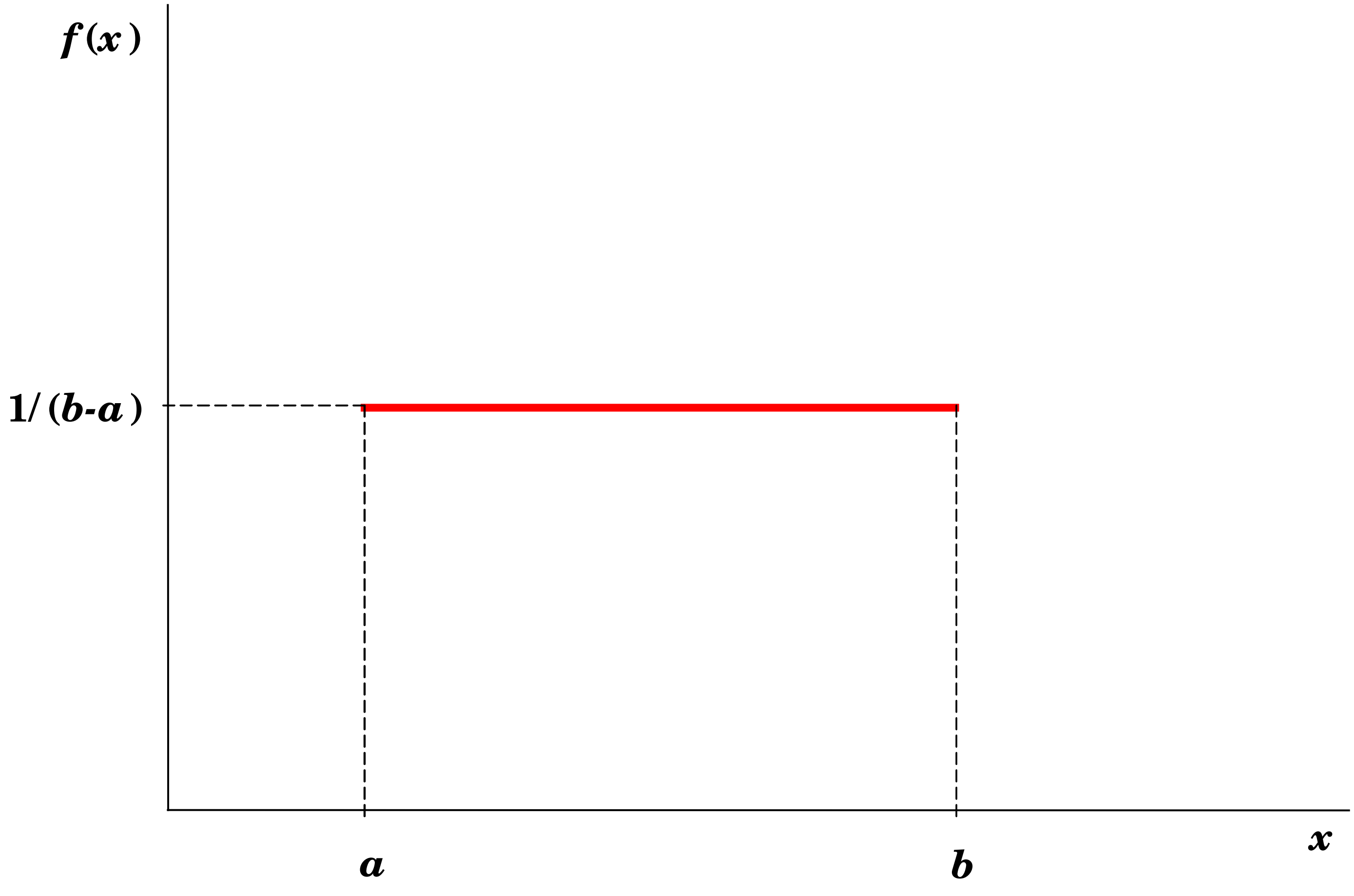


# Distribuições contínuas: Normal

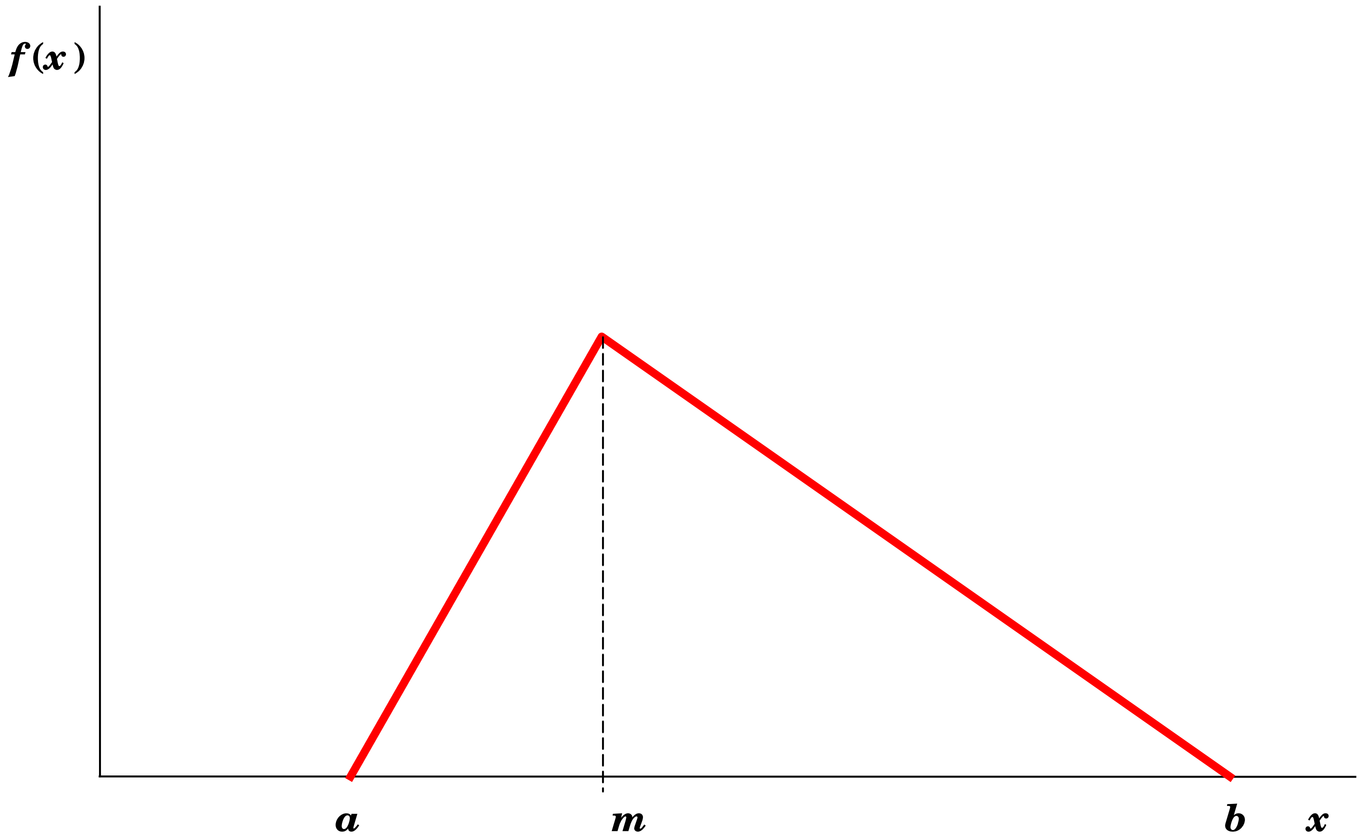


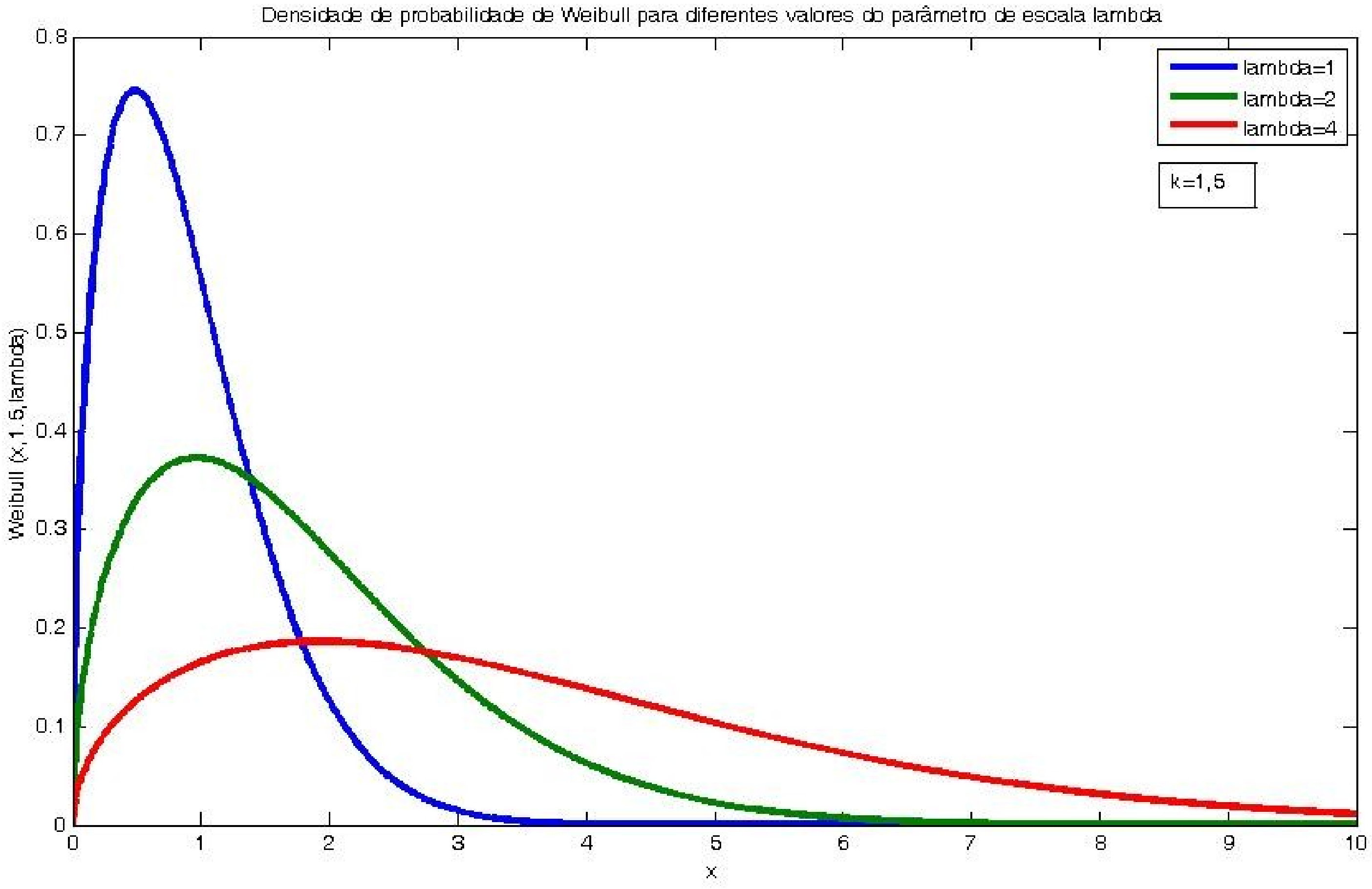
# Distribuições contínuas: Uniforme



# Distribuições contínuas: Triangular



# Distribuições contínuas: Weibull



# Modelagem de dados... Sem dados!

| *Distribuição* | *Parâmetros* | *Características* | *Aplicabilidade* |
|---|---|---|---|
| *Exponencial* | *Média* | ▪*Variância alta*<br>▪*Cauda para direita* | ▪*Grande variabilidade dos valores*<br>▪*Independência entre um valor e outro*<br>▪*Muitos valores baixos e poucos valores altos*<br>▪*Utilizada para representar o tempo entre chegadas sucessivas e o tempo entre falhas sucessivas* |
| *Triangular* | *Menor valor, moda e maior valor* | ▪*Simétrica ou não* | ▪*Quando se conhece ou se tem um bom "chute" sobre a moda (valor que mais ocorre), o menor valor e o maior valor que podem ocorrer* |
| *Normal* | *Média e desvio-padrão* | ▪*Simétrica*<br>▪*Forma de sino*<br>▪*Variabilidade controlada pelo desvio-padrão* | ▪*Quando a probabilidade de ocorrência de valores acima da média é a mesma que valores abaixo da média*<br>▪*Quando o tempo de um processo pode ser considerado a soma de diversos tempos de sub-processos*<br>▪*Processos manuais* |
| *Uniforme* | *Maior valor e menor valor* | ▪*Todos os valores no intervalo são igualmente prováveis de ocorrer* | ▪*Quando não se tem nenhuma informação sobre o processo ou apenas os valores limites (simulação do pior caso)* |
| *Discreta* | *Valores e probabilidade de ocorrência destes valores* | ▪*Apenas assume os valores fornecidos pelo analista* | ▪*Utilizada para a escolha de parâmetros das entidades (por exemplo: em uma certa loja, 30% dos clientes realizam suas compras no balcão e 70% nas prateleiras)*<br>▪*Quando se conhecem apenas "valores intermediários" da distribuição ou a porcentagem de ocorrência de alguns valores discretos* |