



Quem se prepara, não para.

Business Intelligence

4º período

Professora: Michelle Hanne

Visão Geral Arquitetura BI

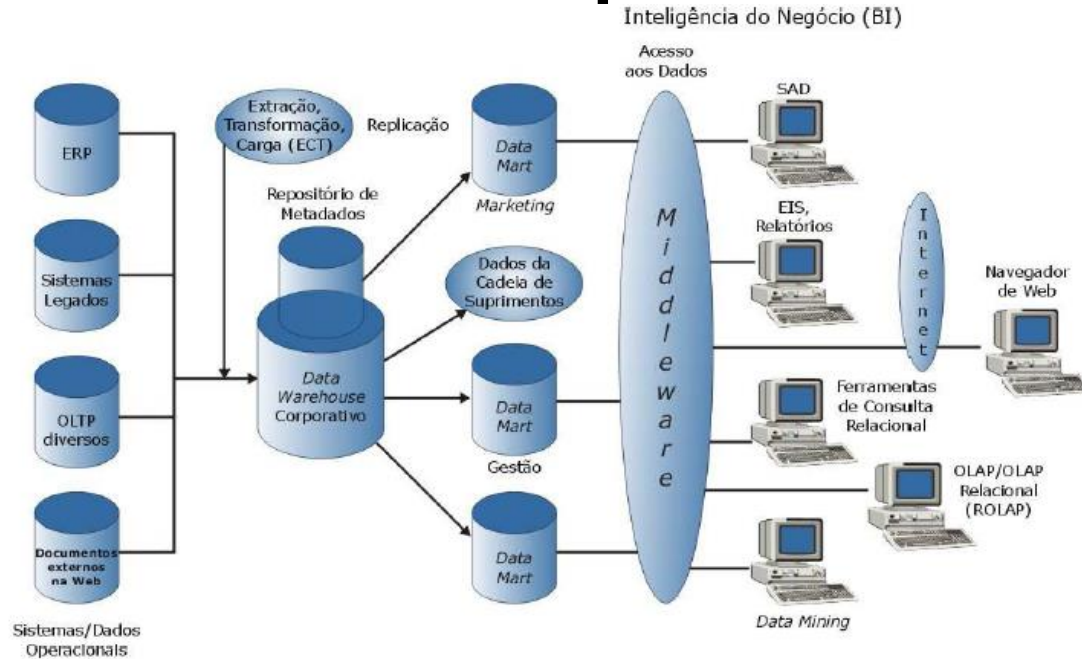


Figura 1 - Estrutura e visões do *Data Warehouse*.

Fonte: Turban *et al* (2004).

Aplicações de BI

NA ÁREA DE TELECOMUNICAÇÕES

- Análises comuns na área de telecomunicações são:
- Comparação de performance entre regionais;
- Análise do perfil de clientes;
- Comparação de performance de vendas em lojas de uma mesma rede;
- Aceitação dos produtos e serviços por cliente e região;
- Índice de **Churn** (taxa de perda de clientes no segmento de telecomunicações);
- Análise de informações para **fidelização** e **retenção** de clientes.

Aplicações de BI

NA ÁREA FINANCEIRA

- Análises comuns na área financeira são:
- Análise de lucratividade de produtos bancários mais vendidos;
- Receitas e despesas em faixas de BI;
- Análise de Risco de Crédito;
- Rentabilidade.

Aplicações de BI

NA ÁREA COMERCIAL

- Análises comuns na área comercial são:
- Análise de índice de atendimento dos pedidos, avaliando problemas que geram ruptura dos mesmos, ocasionando perda de faturamento;
- Análise de pedidos efetuados por clientes que foram atendidos em plenitude e entregues nas datas desejadas;
- Desempenho de equipes de vendas por meio de análise de indicadores de metas vs. Venda realizada demonstrada por meio de painéis de informação.

Aplicações de BI

NA ÁREA DE MARKETING

- Segmentação de clientes;
- Resultado de campanhas fornecendo dados para que equipes possam medir e comparar resultados;
- Área de canais;
- Análise de informações por meio de dados fornecidos pelo SAC (Sistema de Atendimento ao Consumidor) e pela Ouvidoria. Empresas que utilizam esse benefício por meio de Business Intelligence conseguem entender facilmente a performance de seus produtos e aceitação pelos compradores, conseguindo mais e mais informações para melhoria contínua de seus produtos.

Aplicações de BI

NA ÁREA INDUSTRIAL

- Análise quantitativa de produtos produzidos por planta, por regiões específicas ou mesmo por insumos diferentes permitindo a comparação de resultado de venda entre eles;
- Produtividade de equipes e pessoas;
- Equipamentos indisponíveis por falta de manutenção ou parados por falta de operadores qualificados;
- Comparação de horas trabalhadas na produção por equipes ou plantas.

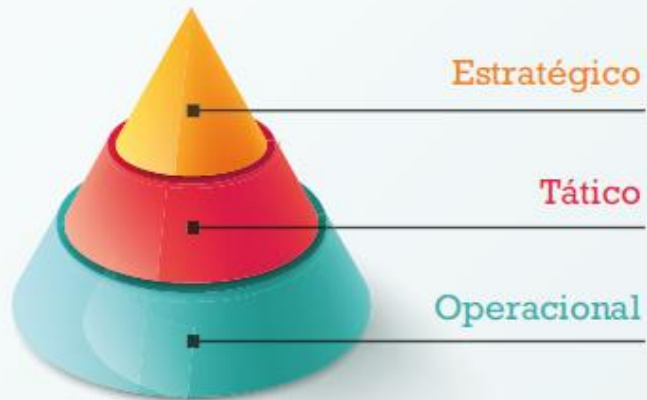
Aplicações de BI

NA ÁREA MÉDICA

- Convênios médicos analisam dados relativos à medicação ministrada versus resultado de tratamentos;
- Comparação de medicamentos receitados por vários médicos para a mesma patologia;
- Análise de resultados de exames de laboratórios, criando base histórica para futuras ações de machine learning e big data em casos de câncer;
- Ocupação de equipamentos de alta tecnologia;
- Previsão de ocupação de leitos;
- Tipo de atendimento em pronto-socorro por sazonalidade.

Aplicações de BI

Conceito - BI deve atuar apenas na camada estratégica



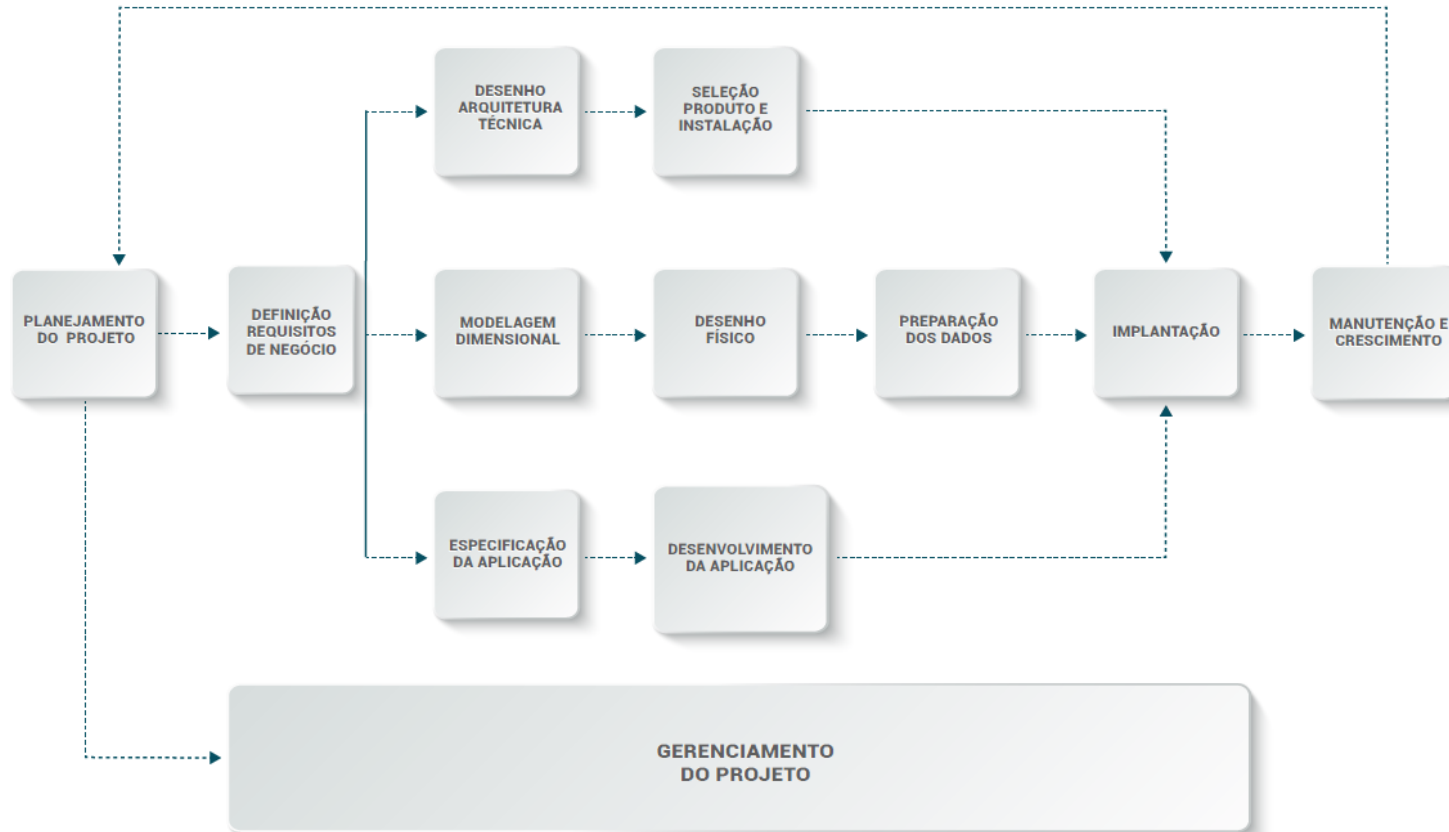
Prática - Maioria dos projetos de BI estão focados na base operacional

IE> Content Hub © 2019

Na prática, por volta de 70% dos projetos de Business Intelligence são utilizados apenas para integração dos vários sistemas operacionais presentes nas organizações.

Perceba então projetos de Business Intelligence utilizados apenas como fábrica de relatórios.

Ciclo de Vida do BI



Ciclo de Vida do BI

1.1 PLANEJAMENTO DO PROJETO

- Aqui se deve descrever os objetivos do projeto, identificar as tarefas, programar as tarefas dos colaboradores e elaborar todo o plano que existirá no projeto.

1.2 DEFINIÇÃO DE REQUISITOS DE NEGÓCIO

- Neste processo devem ser realizadas as entrevistas com as pessoas do negócio. A intenção é de se conhecer os objetivos estratégicos, táticos e operacionais do negócio conhecendo detalhadamente os processos e entregáveis. Com essas entrevistas é possível obter os temas analíticos.

Ciclo de Vida do BI

1.3 DADOS: MODELAGEM DIMENSIONAL

- O próximo passo é desenhar modelos de dados para dar suporte a essas análises com uma abordagem diferente daquela usada para modelagem de sistemas transacionais. Esse modelo identifica a tabela de fato, as dimensões associadas, os atributos e caminhos e fatos hierárquicos.
- Esse modelo identifica a tabela de fato, as dimensões associadas, os atributos e caminhos e fatos hierárquicos. O desenho do banco de dados lógico é concluído com estruturas de tabela apropriadas e relacionamentos de chave primária/estrangeira.

Ciclo de Vida do BI

1.4. DADOS: FÍSICO

- O próximo passo é o desenho do banco de dados físico que se concentra na definição das estruturas físicas necessárias para apoiar o desenho de dados lógico. Estratégias preliminares de indexação e particionamento também são determinadas nessa fase.

1.5 DADOS: PREPARAÇÃO DOS DADOS

- Este passo de preparação de dados possui três etapas principais: extração, transformação e carga (ETL).

Ciclo de Vida do BI

1.6 TECNOLÓGICA: DESENHO DE ARQUITETURA TÉCNICA

- É necessário um ambiente para suportar o data warehouse, o que requer integrações com inúmeras tecnologias existentes. Esse desenho deve estabelecer um framework que considere três fatores – o negócio que deve ser suportado, tecnologias atualizadas e estratégia para crescimento.

1.7 TECNOLÓGICA: SELEÇÃO DE PRODUTO E INSTALAÇÃO

- Com o framework, uma arquitetura desenhada, você deve acoplar componentes arquitetônicos específicos, como a plataforma de hardware e o sistema de gerenciamento de banco de dados às ferramentas de ETL. Relatórios e dashboards precisam ser avaliados e selecionados.

Ciclo de Vida do BI

1.8 APLICAÇÃO: ESPECIFICAÇÃO DA APLICAÇÃO

- As especificações do aplicativo descrevem o modelo de relatório, os parâmetros conduzidos pelo usuário e os cálculos necessários. Essas especificações garantem que a equipe de desenvolvimento e os usuários comerciais tenham um entendimento comum dos aplicativos a serem entregues.

1.9 APLICAÇÃO: DESENVOLVIMENTO DA APLICAÇÃO

- Com as especificações do ETL, relatórios e dashboard, você deve iniciar o desenvolvimento de aplicativos do usuário final. Isso envolve a configuração dos metadados da ferramenta e a construção dos relatórios especificados.

Ciclo de Vida do BI

1.10 IMPLANTAÇÃO

- Agora, com o desenvolvimento entregue, você deve implantar: a convergência de aplicativos, tecnologia e dados para acesso final aos usuários.

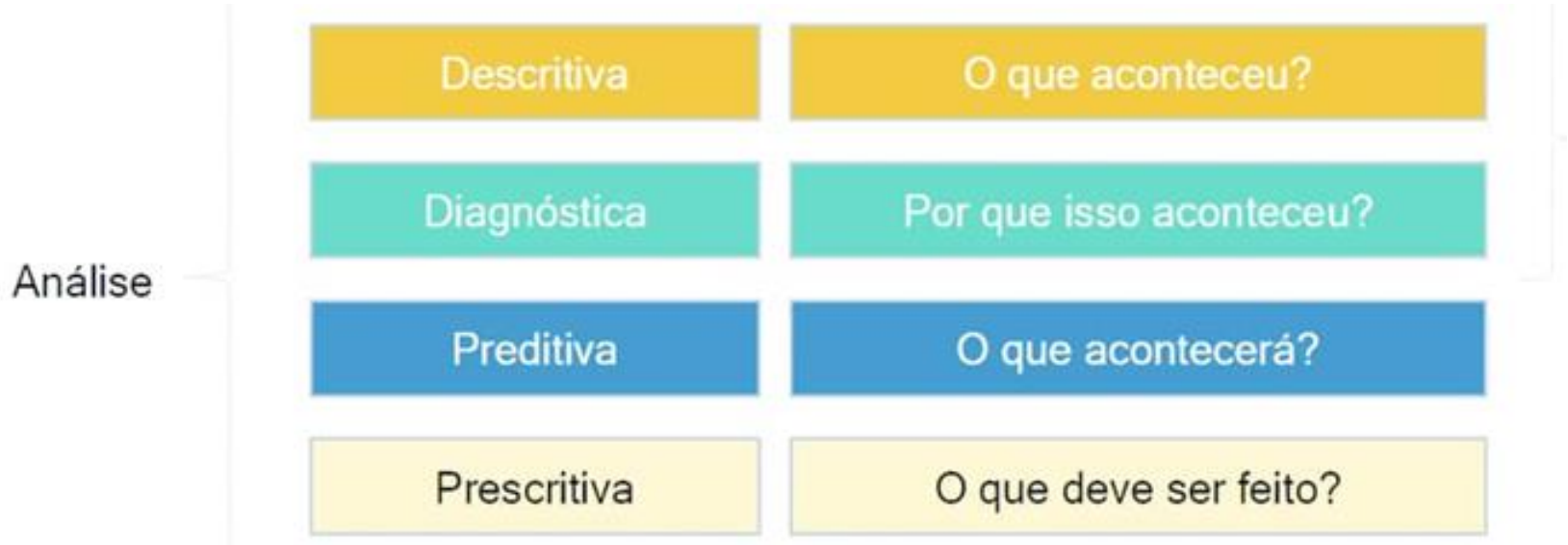
1.11 MANUTENÇÃO E CRESCIMENTO

- Atenção nas possíveis alterações de negócio, garantindo que os processos e procedimentos estejam em vigor para uma operação contínua e eficaz. As métricas de aceitação e desempenho do data warehouse devem ser medidas ao longo do tempo

1.12 GERENCIAMENTO DE PROJETO

- O gerenciamento de projetos garante que as atividades do Ciclo de Vida Dimensional dos Negócios permaneçam nos trilhos.

Análise Dados: Estatística x BI x IA



Análise Univariada

A análise estatísticas univariada analisa a distribuição e dispersão dos dados: a análise da distribuição de frequências e a análise das medidas de localização, dispersão, assimetria e curtose dos dados.

A análise das estatística descritivas inclui a análise dos valores da média, moda, mediana, quartis e percentis, limites da amostra, amplitude, intervalos interquartis, variância e desvio padrão, assimetria e curtose.

Resumo dos Dados

Descrever numericamente, e de forma sucinta, características importantes dos dados.

Resumos incluem basicamente medidas de posição, separação e dispersão.

Medidas de Posição

Indicam alguma tendência central ou comportamento esperado, extraídos de uma amostra qualquer.

Dentre as principais medidas, destacam-se: **média, mediana e moda.**

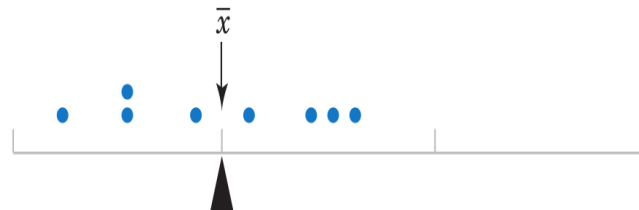
Medidas de Posição

Sejam x_1, x_2, \dots, x_n n observações de um fenômeno aleatório qualquer. Denominamos média aritmética da amostra.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}.$$

Pode ser interpretada como o centro de massa (baricentro das observações).



Exemplo

Uma pesquisa avaliou a idade dos 25 alunos de uma turma de engenharia. Os dados foram organizados na seguinte tabela de frequências:

| Idade | freq. |
|--------------|-----------|
| 19 | 3 |
| 20 | 5 |
| 21 | 1 |
| 22 | 8 |
| 23 | 4 |
| 24 | 1 |
| 25 | 0 |
| 26 | 3 |
| Total | 25 |

A média aritmética pode ser obtida diretamente do dispositivo (1):

$$\bar{x} = \frac{\sum_{i=1}^8 x_i f_i}{\sum_{i=1}^8 f_i} = \frac{548}{25} = 21,92.$$

Exemplo

Os dados a seguir correspondem ao tempo de vida, em anos, de cada elemento de uma amostra de lâmpadas produzidas por uma fábrica qualquer.

| t | freq. |
|--------------|-----------|
| 0 ┤ 0,5 | 3 |
| 0,5 ┤ 1,0 | 12 |
| 1,0 ┤ 1,5 | 20 |
| 1,5 ┤ 2,0 | 5 |
| 2,0 ┤ | 0 |
| Total | 40 |

Tomando o ponto médio dos quatro primeiros intervalos e o extremo esquerdo do último intervalo, obtemos:

$$\bar{x} = \frac{\sum_{i=1}^5 x_i f_i}{\sum_{i=1}^5 f_i} = \frac{43,5}{40} = 1,0875 \text{ anos.}$$

Quando buscamos uma tendência central e os dados estão agrupados em intervalos, devemos selecionar um elemento específico como representante de cada classe a fim de efetuarmos os cálculos necessários. Algumas das escolhas comuns são: centro ou extremos do intervalo. Neste caso, dizemos que a média foi **estimada** e não calculada.

Sensibilidade da Média

Médias aritméticas são bastante sensíveis a valores extremos.

Pode levar a conclusões equivocadas.

Suponhamos, por exemplo, que os salários dos funcionários de uma empresa de pequeno porte são: R\$2500,00, R\$2500,00, R\$1000,00, R\$1800,00 e R\$20000.

A média salarial da empresa será

$$\bar{x} = \frac{2500 + 2500 + 1000 + 1800 + 20000}{5} = 5560$$

Moda (Mo)

A moda é o valor que mais se repete na amostra. É pouco sensível a valores atípicos.

Pode ser aplicada a variáveis nominais.

A moda apresenta **interpretação mais razoável**, comparada a média, quando desejamos estudar variáveis ordinais.

Moda (Mo)

Ex 1: A série 7, 8, 9, 10, 10, 10, 11, 12, 13, 15 tem $Mo=10$.

Ex 2: 3, 5, 8, 10, 12, 13 é amodal (não tem moda).

Ex 3: 2, 3, 4, 4, 4, 5, 6, 7, 7, 7, 8, 9 é bimodal (tem duas modas).

Exemplo

Dados agrupados sem intervalo de classe

| Nº DE MENINOS | f_i |
|------------------|-------|
| 0 | 2 |
| 1 | 6 |
| 2 | 10 |
| 3 | 12 |
| 4 | 4 |
| $\Sigma = 34$ | |

$Mo = 3,$

(pois 3 tem frequência 12)

Exemplo

Dados agrupados sem intervalo de classe

| Nº DE MENINOS | f_i |
|------------------|-------|
| 0 | 2 |
| 1 | 6 |
| 2 | 10 |
| 3 | 12 |
| 4 | 4 |
| $\Sigma = 34$ | |

$Mo = 3,$

(pois 3 tem frequência 12)

Exemplo

Moda de dados agrupados com intervalo de classe:

A classe de maior frequência é chamada classe modal. A moda será, então, o ponto médio desta classe.

$$Mo = \frac{l^* + L^*}{2}$$

| i | ESTATURAS (cm) | f _i |
|---|-------------------|----------------|
| 1 | 150 – 154 | 4 |
| 2 | 154 – 158 | 9 |
| 3 | 158 – 162 | 11 ← |
| 4 | 162 – 166 | 8 |
| 5 | 166 – 170 | 5 |
| 6 | 170 – 174 | 3 |
| | | Σ = 40 |

$$Mo = \frac{l^* + L^*}{2}$$

$$Mo = \frac{158 + 162}{2} = \frac{320}{2} = 160$$

$$Mo = 160 \text{ cm}$$

Mediana

A mediana corresponde ao valor que ocupa a posição central dos dados após sua ordenação.

Desta forma, a mediana desconsidera os valores observados em si, já que basta analisar a posição das observações perante sua ordenação.

Pode apresentar uma medida de tendência mais honesta na presença de valores atípicos.

Mediana

Para dados discretos ou contínuos sem agrupamento:

Basta observar a paridade do número n de observações:

se n é ímpar, então a mediana é exatamente o valor central das observações.

$$x_{\left(\frac{n+1}{2}\right)}$$

caso contrario, a mediana é dada pela media aritmética dos dois valores centrais

$$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Exemplo

- No exemplo dos salários, temos a seguinte ordenação: R\$1000,00, R\$1800,00, R\$2500,00, R\$2500,00, R\$20000.
- Como o número de observações é ímpar, temos que a mediana é dada pelo elemento:
 - **central $x(3) = 2500$.**

Exemplo

No exemplo das classes sociais, temos a seguinte ordenação:

B, B, B, B, B, M, M, M, M, M, M, M, M, M, M, A, A, A, A, A, A, A, A, A, A, A, A, A, A, A, A.

Como o número de observações é par, segue que a mediana é dada pelo ponto médio de $x(15)$ e $x(16)$.

A mediana então ficaria entre os valores M e A.

Medidas Separatrizes

Medidas separatrizes dividem a sequência ordenada dos dados em partes que contêm a mesma quantidade de elementos: Quartis, Decis, Percentis...

Medidas Separatrizes

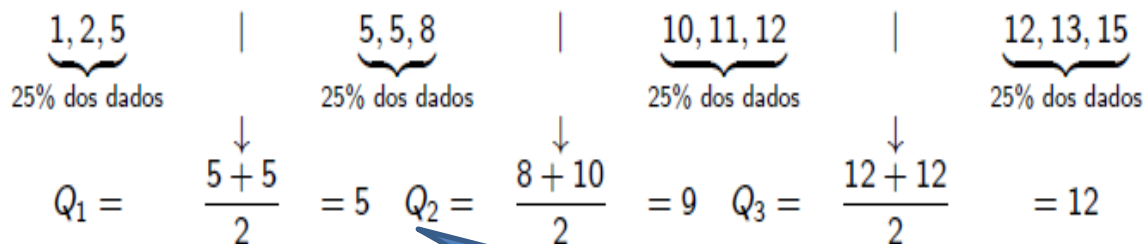
A mediana trata de um caso particular pois separa a amostra em duas porções as quais contêm 50% da informação cada. Considerando, por exemplo, o conjunto de dados 1, 2, 5, 5, 5, 8, 10, 11, 12, 12, 13, 15:

$$\begin{array}{ccc} \underbrace{1, 2, 5, 5, 5, 8}_{50\% \text{ das observações}} & | & \underbrace{10, 11, 12, 12, 13, 15}_{50\% \text{ das observações}} \\ & \downarrow & \\ \text{mediana: } \frac{8 + 10}{2} & = & 9 \end{array}$$

Quartis

Ao dividir os dados **ordenados em 4 partes** de mesmo tamanho, cada um resumirá 25% da informação.

Os elementos Q1, Q2 e Q3 que separam tais grupos são denominados os quartis amostrais.


$$\begin{array}{ccccccc} \underbrace{1, 2, 5}_{25\% \text{ dos dados}} & | & \underbrace{5, 5, 8}_{25\% \text{ dos dados}} & | & \underbrace{10, 11, 12}_{25\% \text{ dos dados}} & | & \underbrace{12, 13, 15}_{25\% \text{ dos dados}} \\ & \downarrow & & \downarrow & & \downarrow & \\ Q_1 = \frac{5 + 5}{2} = 5 & Q_2 = \frac{8 + 10}{2} = 9 & Q_3 = \frac{12 + 12}{2} = 12 & & & & \end{array}$$

Corresponde a Mediana

Quartis- Exemplo

Dados não Agrupados

3, 4, 5, 5, 6, 8, 9, 11, 15, 21

$$Q_2 = 7$$

$$Q_1 = 5$$

$$Q_3 = 11$$

Quartis- Exemplo

Calcule os quartis da série: { 5, 2, 6, 9, 10, 13, 15 }

Ordenação (crescente ou decrescente) dos valores: { 2, 5, 6, 9, 10, 13, 15 }

O valor que divide a série acima em duas partes **iguais é o nº 9**, logo a **Md** = 9 e **Q2**=9.

Quartis- Exemplo

Temos: **{2, 5, 6 }** e **{10, 13, 15 }** outros dois grupos de valores iguais proporcionais pela mediana (quartil 2).

Em { 2, 5, 6 } a mediana é = 5. Ou seja: será o Q1;

Em {10, 13, 15 } a mediana é =13. Ou seja: será o Q3;

Quartis- Exemplo

Calcule os quartis da série: { 1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13 }

Ordenar e Calcular o Quartil 2 = Md = $(5+6)/2 = 5,5$

O quartil 1 será a mediana da série à esquerda de Md: { 1, 1, 2, 3, 5, 5 }

$$\rightarrow Q1 = (2+3)/2 = 2,5$$

O quartil 3 será a mediana da série à direita de Md: { 6, 7, 9, 9, 10, 13 }

$$\rightarrow Q3 = (9+9)/2 = 9$$

Box-Plot (Diagrama de caixa)

O **boxplot** capta importantes aspectos de um conjunto de dados através do seu resumo dos cinco números (**valor mínimo, primeiro quartil, segundo quartil, terceiro quartil e valor máximo**).

Bem como, o centro, dispersão, desvio da simetria e identificação das observações que estão longe do centro dos dados (*outliers*).

Box-Plot (Diagrama de caixa)

O gráfico é formado por uma caixa construída paralelamente ao eixo da escala dos dados (pode ser horizontal ou vertical).

Esse *box* vai desde o primeiro quartil até o terceiro quartil e nela traça-se uma linha na posição da mediana.

Exemplo

| | | | | | | | | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|
| Ordem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| Valor | 3,0 | 3,5 | 4,5 | 5,0 | 5,0 | 5,5 | 6,5 | 6,5 | 6,5 | 7,5 | 7,6 | 7,9 | 8,0 | 8,0 | 9,0 | 9,5 | 10,0 | 15,0 |

A mediana divide o conjunto em duas partes, cada uma com 9 observações. A mediana será, então, a média dos dois valores centrais:

$$Q2 = \frac{6,5 + 7,5}{2} = 7,0$$

| | | | | | | | | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|
| Ordem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| Valor | 3,0 | 3,5 | 4,5 | 5,0 | 5,0 | 5,5 | 6,5 | 6,5 | 6,5 | 7,5 | 7,6 | 7,9 | 8,0 | 8,0 | 9,0 | 9,5 | 10,0 | 15,0 |

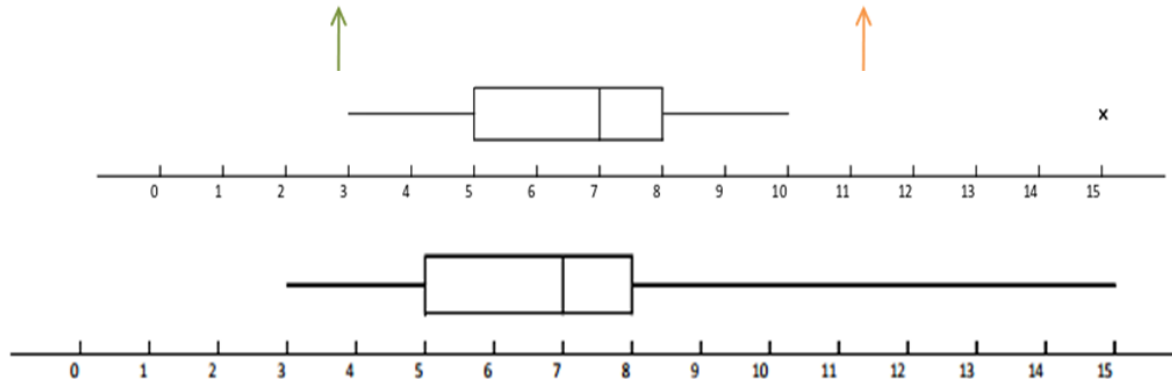
Exemplo

- O cálculo do primeiro e do terceiro quartis:
 - Calcular as medianas das duas metades – o primeiro quartil é a mediana da metade inferior e o terceiro quartil é a mediana da metade superior.

$$Q1 = 5,0$$

$$Q3 = 8,0$$

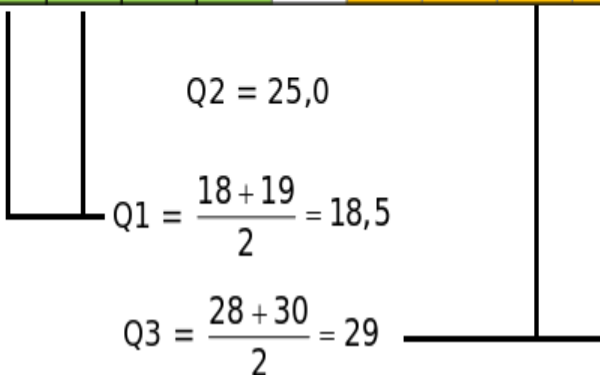
| | | | | | | | | | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|
| Ordem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| Valor | 3,0 | 3,5 | 4,5 | 5,0 | 5,0 | 5,5 | 6,5 | 6,5 | 6,5 | 7,5 | 7,6 | 7,9 | 8,0 | 8,0 | 9,0 | 9,5 | 10,0 | 15,0 |



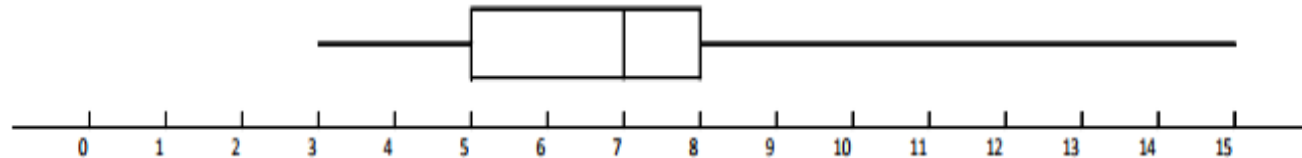
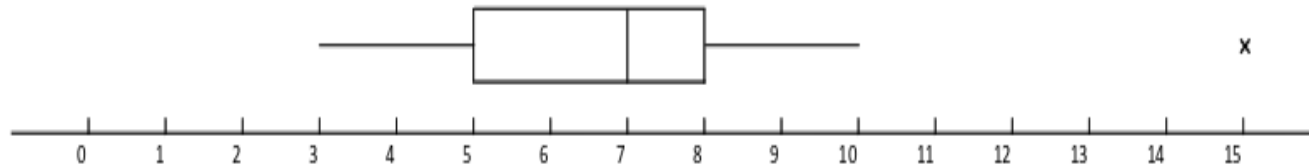
Exemplo

| | | | | | | | | | | | | | |
|-------|----|----|----|----|----|----|------|----|----|----|----|----|----|
| Ordem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Valor | 15 | 17 | 18 | 19 | 19 | 20 | 25,0 | 26 | 26 | 28 | 30 | 32 | 42 |

| | | | | | | | | | | | | | |
|-------|----|----|----|----|----|----|------|----|----|----|----|----|----|
| Ordem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Valor | 15 | 17 | 18 | 19 | 19 | 20 | 25,0 | 26 | 26 | 28 | 30 | 32 | 42 |

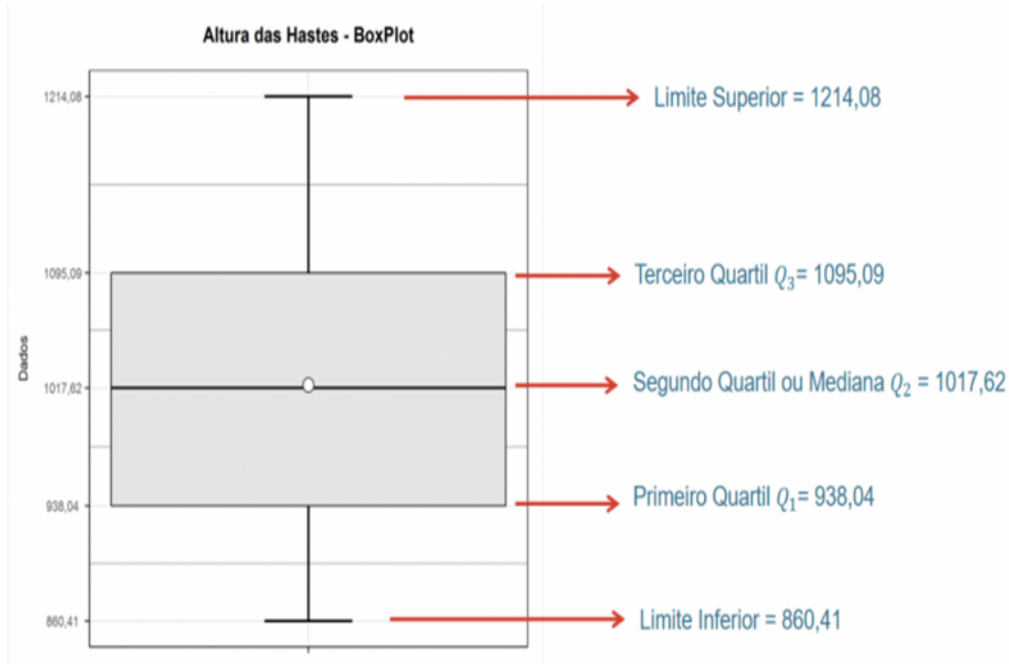

$$Q2 = 25,0$$
$$Q1 = \frac{18 + 19}{2} = 18,5$$
$$Q3 = \frac{28 + 30}{2} = 29$$

Exemplo



Exemplo

- Na Tabela a seguir temos as medidas da altura de 20 hastes.



| Dados da usinagem | | | |
|-------------------|---------|---------|---------|
| 903,88 | 1036,92 | 1098,04 | 1011,26 |
| 1020,70 | 915,38 | 1014,53 | 1097,79 |
| 934,52 | 1214,08 | 993,45 | 1120,19 |
| 860,41 | 1039,19 | 950,38 | 941,83 |
| 936,78 | 1086,98 | 1144,94 | 1066,12 |

Exemplo

Uma indústria produz uma peça automotiva cujo valor de referência é 75cm. Após verificar lotes com peças fora de especificação, enviaram duas equipes de trabalhadores (A e B) para um treinamento.

Exemplo

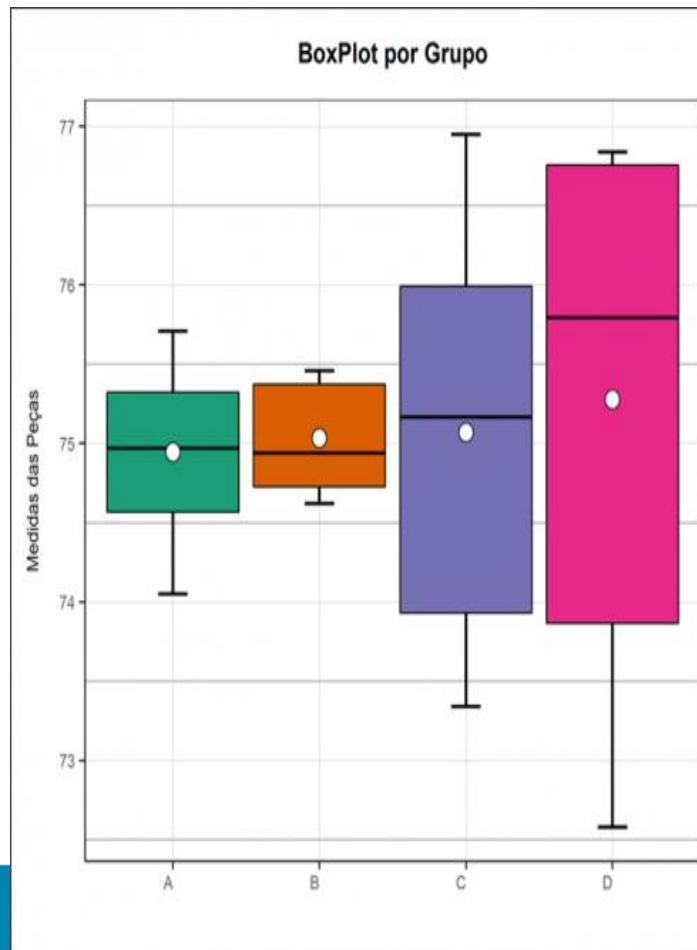
Para verificar a eficiência do treinamento, foram selecionadas 10 peças produzidas pelas equipes A e B e 10 peças produzidas pelas equipes C e D que não participaram do treinamento

| A | | B | | C | | D | |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 75,27 | 74,93 | 74,94 | 74,75 | 75,93 | 73,34 | 75,98 | 76,75 |
| 75,33 | 74,72 | 75,25 | 74,65 | 76,95 | 74,04 | 75,61 | 76,78 |
| 74,58 | 74,53 | 75,44 | 74,94 | 75,47 | 75 | 74,2 | 74,74 |
| 75,01 | 75,32 | 74,62 | 74,92 | 73,6 | 76,18 | 76,44 | 72,58 |
| 75,71 | 74,05 | 75,35 | 75,46 | 74,85 | 75,33 | 76,84 | 72,86 |

Análise Exploratória

Analizando o gráfico podemos observar que:

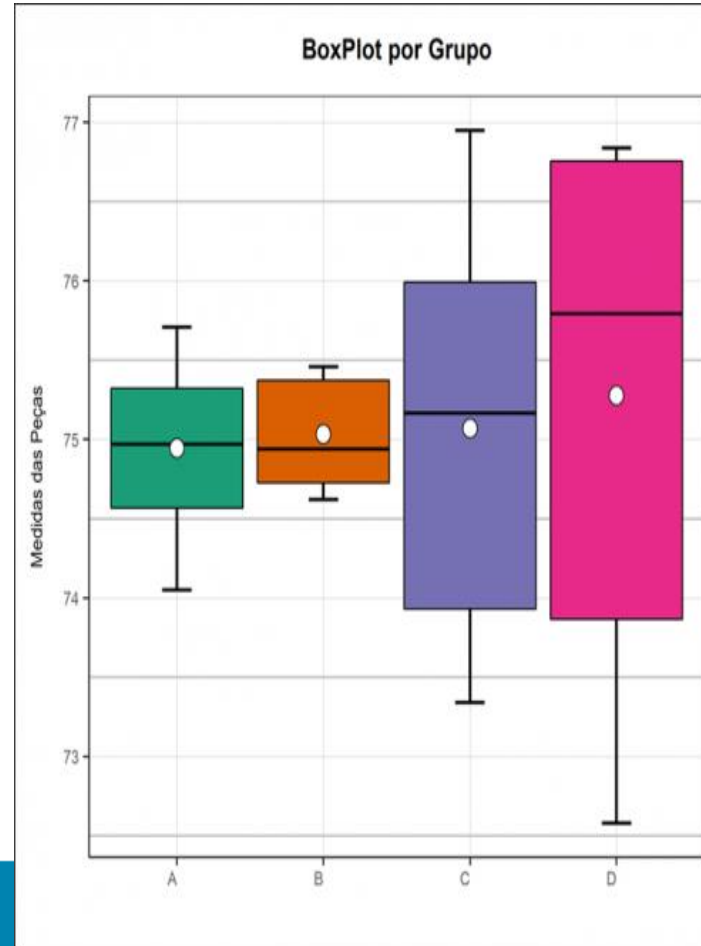
- As equipes **A e B** produzem peças com **menor variabilidade**, indicando que o treinamento teve o efeito desejado;
- A equipe **D** é a que produz peças com **maior variabilidade**;
- A equipe **B** é a que produz peças com **menor variabilidade**.



Análise Exploratória

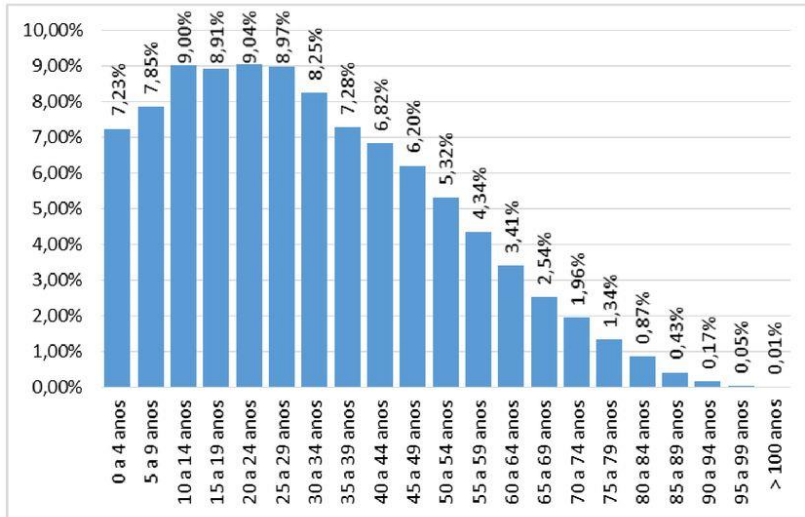
Considerações:

Como as peças das **equipes A e B** tem **menor variabilidade** e com valor **médio próximo do valor de referência**, vale a pena enviar as demais equipes para o treinamento.

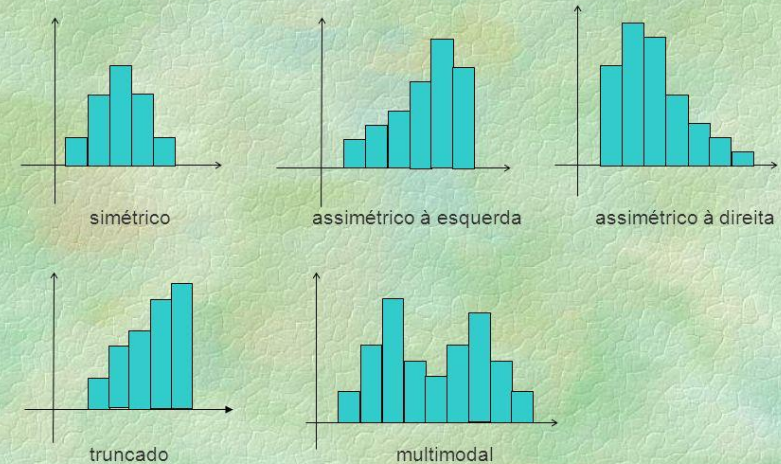


- **Histograma:** ferramenta de análise e representação de dados quantitativos, **agrupados em classes de frequência**.
- Permite identificar se o conjunto de dados é simétrico ou não.
- Permite identificar qual a classe de valores é mais ou menos frequente no conjunto de dados.

Histograma resultante do censo da população brasileira realizado em 2010

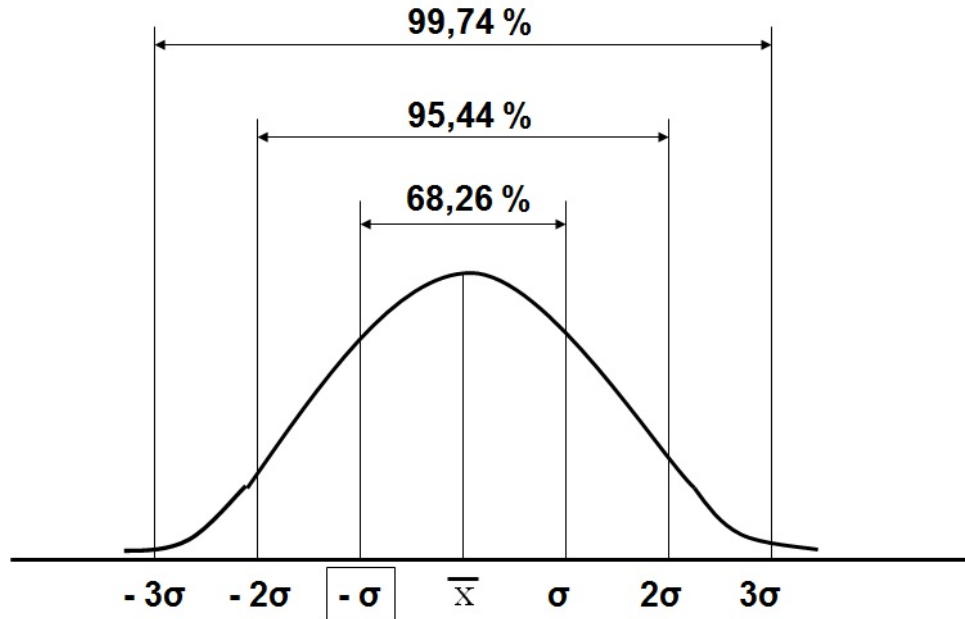


Diferentes formatos de histogramas



Desvio Padrão

- **Regra empírica (regra 68-95-99,7 ou regra de três sigmas):**



Se a distribuição (histograma) da variável for simétrica (forma de um sino)

Desvio Padrão

- **Aplicações do desvio padrão:** Utilizado em problemas que envolvem incerteza.
- **Previsão de vendas:** as vendas de amanhã estarão entre 100 e 400 unidades com 95% de probabilidade.
- **Pesquisa de opinião:** em uma pesquisa de opinião com uma amostra de 2400 pessoas, a margem de erro com 95% de confiança é de 2 pontos percentuais para cima ou para baixo.

Desvio Padrão: exemplo de aplicação

Histórico da venda dos últimos 20 dias, pela manhã.

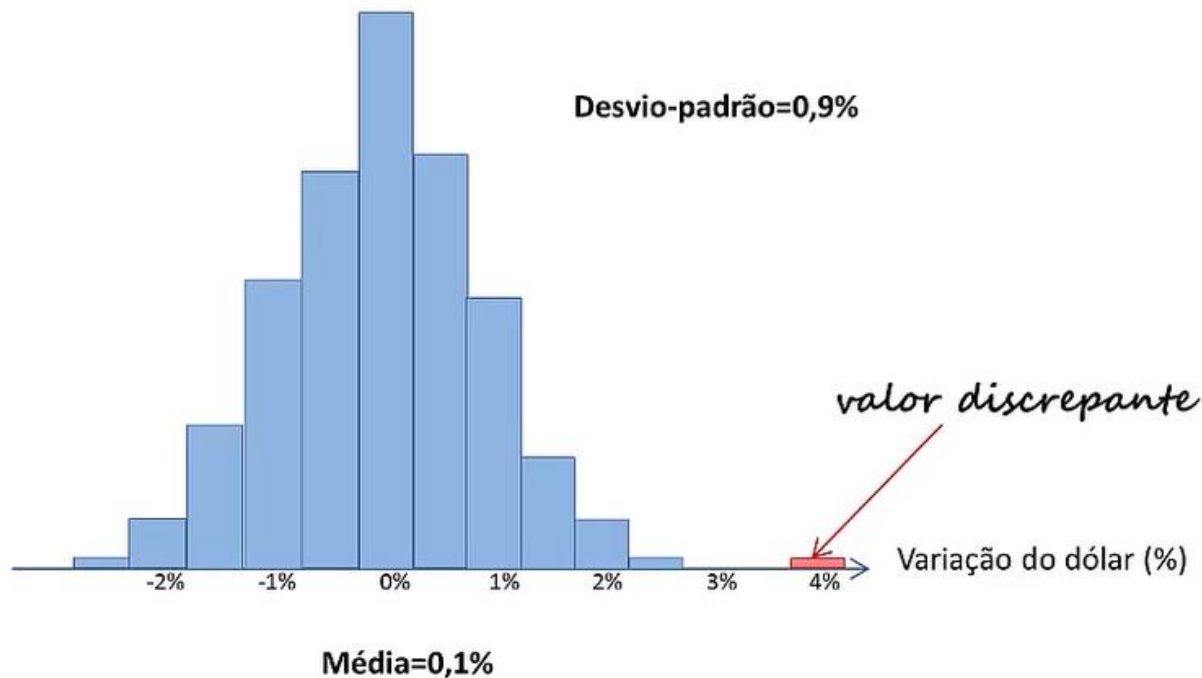
| Dia | Número de pães vendidos |
|---------------|-------------------------|
| 1 | 1570 |
| 2 | 1472 |
| 3 | 1624 |
| 4 | 1728 |
| 5 | 1720 |
| 6 | 1773 |
| 7 | 1382 |
| 8 | 1577 |
| 9 | 1710 |
| 10 | 1491 |
| 11 | 1531 |
| 12 | 1431 |
| 13 | 1415 |
| 14 | 1502 |
| 15 | 1523 |
| 16 | 1388 |
| 17 | 1543 |
| 18 | 1560 |
| 19 | 1613 |
| 20 | 1563 |
| Média | 1556 |
| Desvio-padrão | 114 |



- Quantos pães uma padaria deve produzir no mínimo para que a probabilidade de faltar pão seja de apenas 2,5 % ? R. 1328 pães.
- Qual a quantidade mínima e máxima de pães deve ser produzido para que a padaria consiga vender 95% deles ? R. entre 1328 e 1784

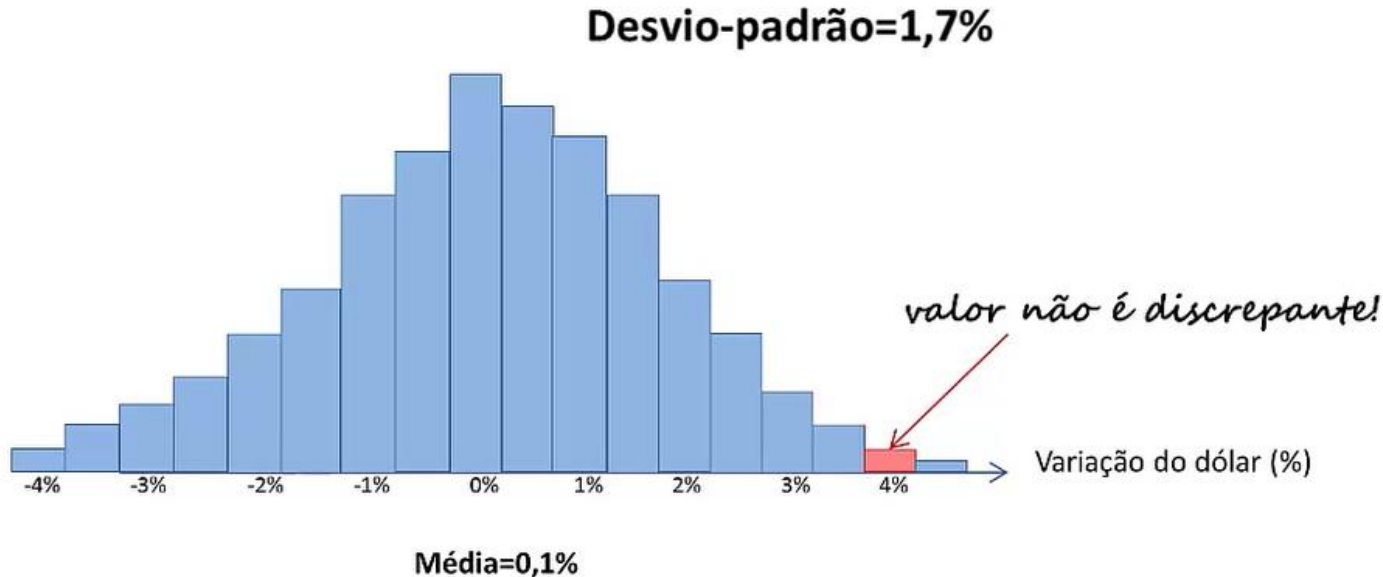
Valores Outliers

- O que é um valor discrepante (outliers)?
 - **Exemplo 01:** 4% é um valor discrepante



Valores Outliers

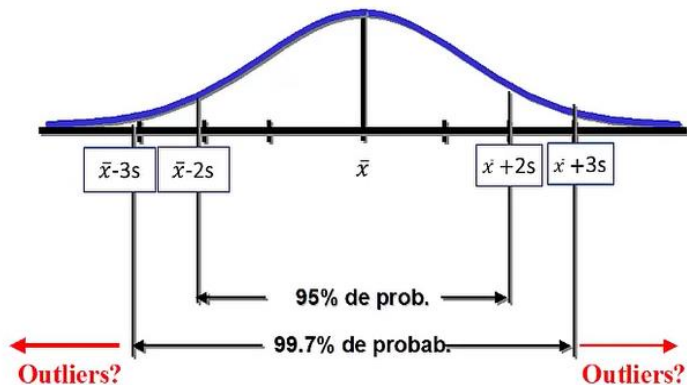
- O que é um valor discrepante (outliers)?
 - **Exemplo 02:** 4% não é um valor discrepante



Valores Outliers

- Como saber se o dado é ou não **muito distante** ?
- **Critério 01:** determinar a distância até a média em termos de desvios padrões.

Se a distribuição da variável for simétrica em forma de sino (normal), podemos usar a regra empírica da estatística:



*Se valor padronizado absoluto for maior que 3 ->
valor é outlier*

Vídeo Recomendado: <https://www.youtube.com/watch?v=UVXXKUrZtRo>

Referências

BRAGHITTONI, Ronaldo. **Business Intelligence**: Implementar do jeito certo e a custo zero. São Paulo: Casa do Código, 2017.

GONÇALVES, Glauber Rogério Barbieri. **Sistemas de informação**. Porto Alegre: SAGAH, 2017.

KIMBALL, Ralph. **Data warehouse toolkit** - Técnicas para Construção de Data Warehouses Dimensionais. São Paulo: Makron Books, 1998.

TURBAN, Efraim et al. **Business Intelligence**: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.