# ESTATÍSTICA BÁSICA

Fonte: http://www.portalaction.com.br/estatistica-basica

A Estatística (ou ciência Estatística) é um conjunto de técnicas e métodos de pesquisa que entre outros tópicos envolve o planejamento do experimento a ser realizado, a coleta qualificada dos dados, a inferência, o processamento, a análise e a disseminação das informações.

Na estatística trabalhamos com dados, nos quais podem ser obtidos por meio de uma amostra da população em estudo. A seguir, definimos esses conceitos básicos:

**População:** conjunto de elementos que tem pelo menos uma característica em comum. Esta característica deve delimitar corretamente quais são os elementos da população.

**Amostra:** subconjunto de elementos de uma população, que são representativos para estudar a característica de interesse da população. A seleção dos elementos que irão compor a amostra pode ser feita de várias maneiras e irá depender do conhecimento que se tem da população e da quantidade de recursos disponíveis.

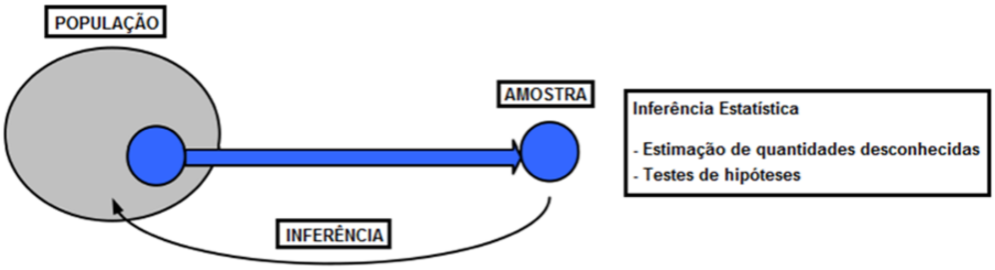



Figura 1. Conceitos básicos de estatística.

Vamos apresentar os elementos básicos da análise de dados. Veremos as Estatísticas Descritivas utilizadas para organizar, resumir e descrever os aspectos importantes do comportamento dos dados. A descrição dos dados também pode identificar anomalias, até mesmo resultantes do registro incorreto de valores e valores extremos (aqueles que não seguem a tendência geral do restante do conjunto de dados). As ferramentas descritivas são os muitos tipos de gráficos e tabelas, bem como as medidas de síntese: medidas de posição e medidas de dispersão.

Sempre que resumimos um conjunto de dados, perdemos informação sobre o mesmo, pois condensamos as observações originais. Entretanto, esta perda de informação é pequena se comparada ao ganho que se tem com a clareza da interpretação proporcionada.

# 1 - NOÇÕES BÁSICAS DE ESTATÍSTICA

A Estatística durante séculos foi usada inconscientemente pelos povos como um caráter meramente descritivo e de registro de ocorrências. As primeiras atividades foram por volta de 2000 a.C. e foram usados no recenseamento das populações agrícolas chinesas. No início do século XIX, os grandes matemáticos entraram em cena, como exemplo, o francês Simon Laplace e o alemão Carl Friedrich Gauss (1777 –1855), este último surge com aplicações da "distribuição normal" para modelagem de erros de medição. A teoria da distribuição normal foi usada pelo astrônomo e matemático belga Adolphe Quételet (1796 –1874), no estudo estatístico de diversas características das populações humanas: altura, peso, natalidade, mortalidade, renda mensal etc. Ronald Aylmer Fisher (1890 – 1962), estatístico britânico, foi o gênio que criou a moderna teoria da estatística. Na Estatística trabalhou com ajustes de curvas de frequências, com coeficientes de correlação, os chamados coeficientes de Fisher, na análise de variância (ANOVA) e nas técnicas de estimação dos parâmetros. Influenciado pelos trabalhos de Karl Pearson, outro importante estatístico britânico. Fisher utilizou os resultados que obteve na Estatística como ferramentas para aplicação nos seus estudos de genética, sendo hoje considerado um dos maiores nomes na Teoria de Estatística e na Estatística aplicada à Biologia.

## 1.1 - Introdução

Aqui, vamos apresentar os elementos básicos da análise de dados. Veremos as Estatísticas Descritivas utilizadas para organizar, resumir e descrever o comportamento dos dados. A descrição dos dados também pode identificar anomalias, até mesmo resultantes do registro incorreto de valores e valores extremos (aqueles que não seguem a tendência geral do restante do conjunto de dados). As ferramentas descritivas são os muitos tipos de gráficos e tabelas, bem como as medidas de síntese, como as medidas de posição, medidas de dispersão, medidas de assimetria e medidas de curtose.

Sempre que resumimos um conjunto de dados, perdemos informação sobre o mesmo, pois condensamos as observações originais. Entretanto, esta perda de informação é pequena se comparada ao ganho que se tem com a clareza da interpretação proporcionada.

A Estatística utiliza a variabilidade presente nos dados para obter tal informação para obter informação sobre o comportamento de processos e produtos. A variabilidade está presente em todo lugar. Por exemplo, a posição de um carro estacionado em uma garagem não é a mesma ao longo dos dias. Neste caso, a posição do carro apresenta uma variação. Nossa estratégia consiste em avaliar as variações e obter informações através dela.

Em geral, a aplicação de técnicas estatísticas envolve várias etapas, conforme Figura 2. Um ponto importante na análise estatística é a definição do problema de gestão que estamos tratando. Por exemplo, avaliar a capacidade do processo, prever demanda de serviços, entre outros. Na sequência, planejamos e executamos a coleta dos dados. A partir dos dados, aplicamos técnicas estatísticas para extrair informação sobre o problema de gestão e, consequentemente sua solução.

Figura 2. Aplicação de técnicas estatísticas

## 1.2. Coleta de dados

Diversos problemas podem ocorrer durante o processo de coleta de dados, os quais podem comprometer seriamente as soluções propostas no final do processo, ou seja, a qualidade da solução do problema de gestão está diretamente relacionada com a qualidade dos dados obtidos. Podemos evitar que alguns problemas ocorram observando fatos como:

- Não se deve coletar dados sem que antes se tenha definido claramente o problema ou situação a ser enfrentada, bem como os objetivos com relação aos mesmos;
- Os sistemas de medição (instrumento, operadores, método, meio) que serão utilizados devem ser avaliados e ter capacidade de medição suficiente (ver, Análise de sistema de medição);
- Os cálculos e leituras devem ser feitos com muita atenção para evitar distorções;
- Devem ser utilizados métodos adequados para coleta de dados de acordo com o problema estudado.

Para excetuarmos a coleta de dados adequadamente devemos conhecer os conceitos básicos de População e Amostra. Uma população é um agregado de elementos (finitos ou não) do qual desejamos obter informações sobre algumas de suas características. Duas populações são consideradas distintas se em uma delas existir um elemento que não está contido na outra população. Como exemplo de população temos a produção diária de uma empresa, medição de hastes de aço realizada com um micrômetro, entre outras. Uma amostra é uma parcela de uma população que pode conter informações sobre esta população, ver Figura 3.

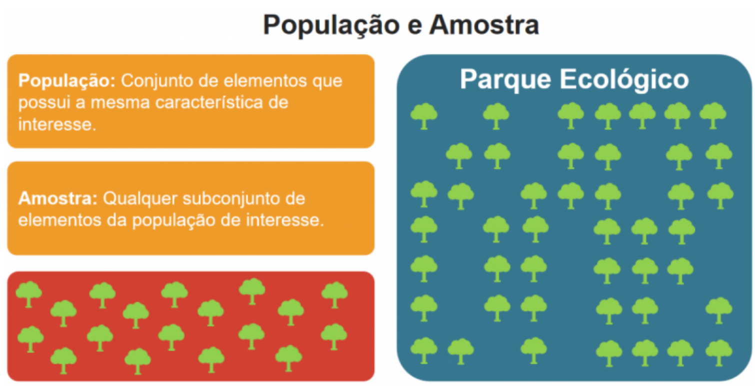


Figura 3. População e amostra.

Outra definição importante para a escolha da técnica estatística e das interpretações dos resultados é a classificação dos dados ou das variáveis relacionadas ao problema de gestão, conforme Figura 4. Uma variável é uma característica específica da população, tal como comprimento de uma peça usinada, massa de um comprimido, taxa Selic, idade, sexo ou preferência partidária. Cada variável consiste de um conjunto de categorias que descrevem a natureza e o tipo de variação associados com a característica. Algumas variáveis podem ter inúmeras categorias de resposta, dependendo do objetivo e foco do problema de gestão. As variáveis têm propriedades de medida distintas associadas às suas categorias. Estas categorias são referidas como níveis de medição ou escalas de medição. As escalas podem ser classificadas como: quantitativas ou qualitativas.

A escala quantitativa está relacionada com características que podem ser medidas ou contadas. Em ambos os casos são características numéricas. Por exemplo, a massa de um comprimido, o preço de um ativo no mercado financeiro, o número de defeitos em um carro, a quantidade de úlcera por pressão em um paciente de um hospital. Por outro lado, a escala qualitativa está relacionada com a descrição da característica. Neste caso, a variável não pode ser medida mas pode ser observada. Por exemplo, os tipos de defeitos em um produto (televisor) no final da linha de montagem, o nível educacional (ensino fundamental, ensino médio, graduação e pós-graduação).

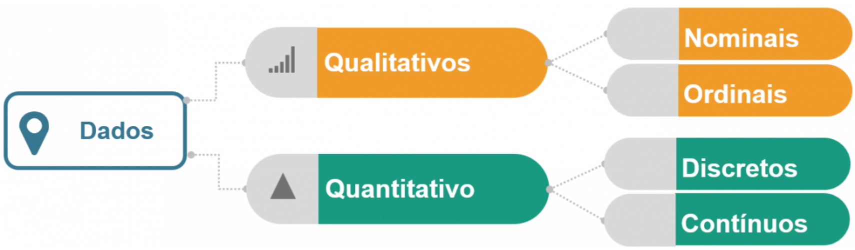



Figura 4. Escala de dados.

No caso de **dados qualitativos nominais** temos apenas a identificação ou rotulação das observações que constituem os dados. Neste caso, os dados são alocados em categorias e contados apenas com relação à frequência de ocorrência. Nenhuma ordenação está implícita. Uma variável como ``sexo" pode ser categorizada em duas respostas possíveis: masculino ou feminino. Tais categorias servem apenas para contar o número de indivíduos que indicam as respectivas categorias.

A **escala qualitativa ordinal** classifica as categorias da variável em termos do grau que possui cada categoria. Neste caso, temos informação sobre a ordenação de categorias, mas não temos indicação da magnitude das diferenças entre essas categorias. Por exemplo, o nível educacional de participantes de uma corrida (ensino fundamental, ensino médio, graduação e pós-graduação) é uma variável qualitativa ordinal.

Na **escala quantitativa discreta** temos a contagem de um evento de interesse. Neste caso, conseguimos determinar a magnitude da diferença entre as categorias da variável. Por exemplo, a contagem de defeitos em uma peça pintada ou o número de úlcera por pressão em um paciente.

Por outro lado, na **escala quantitativa contínua**, as categorias são definidas através da escala de sistemas ou equipamentos de medição. Em geral, a variável assume valores em um intervalo de números. Por exemplo, renda familiar, altura de um indivíduo, idade, distância e temperatura estão associadas com medições realizadas por sistemas de medição que apresentam indicações do valor de cada categoria e as diferenças entre elas.

### 1.2.1. Planejando a coleta de dados

Para estudarmos adequadamente uma população através de uma amostra, devemos planejar a coleta de dados (Figura 5). Com este objetivo, formulamos algumas perguntas:

- Com que frequência ocorrem os problemas?
- Quais são as causas potenciais do problema?

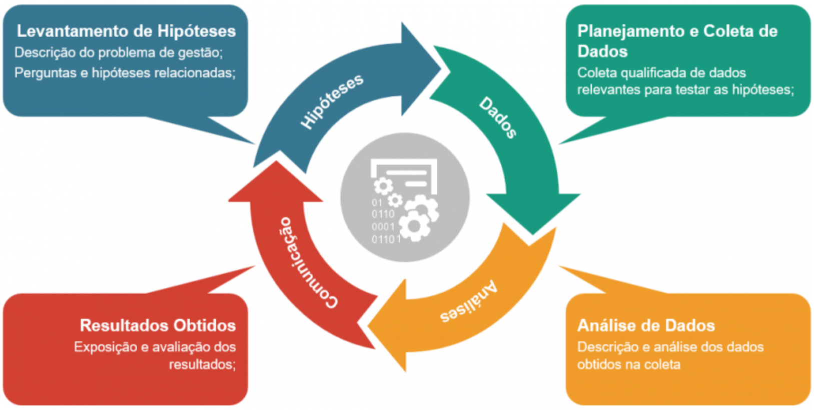



Figura 5. Planejamento da coleta de dados.

Um bom planejamento para coleta de dados deve considerar as seguintes perguntas:

- Qual a pergunta a ser respondida?
- Como comunicar a resposta obtida?
- Qual ferramenta de análise pretendemos usar e como utilizar os resultados?
- Qual tipo de dado é necessário para utilizar as ferramentas desejadas e responder a pergunta?
- Como coletar esses dados com o mínimo de esforço e erro?
- Onde acessar estes dados?
- Quem pode nos fornecer os dados?
- Qual o período em que os dados serão coletados?

Tendo as respostas para estas perguntas, devemos:

- Construir uma metodologia para nos certificar de que todas as informações estão definidas;
- Coletar os dados de forma consistente e honesta;
- Certificar-se de que existe tempo suficiente para a coleta de dados;
- Definir quais informações adicionais serão necessárias para estudos futuros, referências ou reconhecimento.

## 1.3. Exposição dos Dados

Antes da exposição dos dados coletados é necessário que se faça um trabalho de revisão e correção nos dados coletados na tentativa de eliminar possíveis enganos na elaboração do relatório. Inicialmente, os dados podem ser classificados como "qualitativos" ou "quantitativos". Através desta classificação, vamos definir algumas técnicas para resumir o conjunto de dados.

**Dados qualitativos:** Os dados qualitativos representam uma característica da qualidade (ou atributo) associado ao item pesquisado. Por exemplo, podemos classificar um produto em: bom, razoável ou ruim. Os dados qualitativos podem ser divididos em dois tipos:

**Dado qualitativo nominal** - Para o qual não existe nenhuma ordenação nas possíveis realizações;

**Dado qualitativo ordinal** - Para o qual existe uma ordem em seus resultados.

**Exemplo 1.3.1:**

Uma indústria de calculadoras eletrônicas, preocupada com vários defeitos que um de seus produtos vem apresentando, fez um levantamento e constatou os seguintes problemas:

A: Defeito na cobertura plástica;
B: Defeito no teclado;
C: Defeito na fonte de energia;
D: Soldas soltas;
E: Defeito na placa da unidade de processamento;
F: Defeito no visor;
G: Outros.

Este é um típico exemplo de dados qualitativos nominais. Nesta situação, para cada item inspecionado, existe uma variável T que representa o tipo de defeito encontrado. Portanto, essa variável T pode assumir os valores: A,B,C,D,E,F ou G. Logo, para uma calculadora com defeito na cobertura plástica, temos que T = A. A seguir, temos uma Tabela 1 com os valores observados da variável T em um dia de inspeção.

**Tabela 1. Valores observados da variável T.**

| *Tipo de Problemas (T)* | *Frequência* |
|---|---|
| *A* | 10 |
| *B* | 20 |
| *C* | 55 |
| *D* | 80 |
| *E* | 25 |
| *F* | 3 |
| *G* | 7 |

Neste exemplo, todos os defeitos apresentam o mesmo nível de severidade e portanto, não apresentamos uma ordem entre os atributos (defeitos). Neste caso, temos um exemplo de dado qualitativo nominal.

**Exemplo 1.3.2:**

Em um concurso público foram contabilizados os números de pessoas inscritas segundo os níveis de escolaridade: fundamental completo, médio completo, superior completo e pós-graduação completa. Segue abaixo a Tabela 2 com os valores observados.

**Tabela 2. Inscritos de acordo com a escolaridade.**

| Nível de escolaridade | Inscritos |
|---|---|
| Fundamental completo | 451 |
| Médio completo | 627 |
| Superior completo | 292 |
| Pós-graduação completa | 95 |

Neste exemplo, temos uma ordem natural entre os atributos (nível de escolaridade) e consequentemente, temos um exemplo de dados qualitativos ordinais.

**Dados quantitativos**: Neste caso, a característica observada assume valores numéricos que podem ser classificados em "discretos" ou "contínuos".

**Dados quantitativos discretos**: Os dados quantitativos discretos assumem valores dentro de um conjunto com os números especificados. Por exemplo, o número de produtos produzidos por uma máquina em um determinado período de tempo pode ser 0,1,2,3,4,... Neste caso, os dados observados formam um conjunto finito (ou enumerável) de números. Geralmente, quando contamos defeitos, temos dados quantitativos discretos.

**Exemplo 1.3.3:**

Em um hospital, foram contabilizados o número de pessoas com diabetes em 20 grupos de 1000 pessoas cada. Neste caso, obtemos os seguintes dados: 10, 12, 9, 11, 10, 8, 9, 10, 7, 10, 8, 9, 9, 10, 10, 11, 9, 11, 10, 10. Um possível resumo dos dados é desenvolvido na Tabela 3.

**Tabela 3. Pessoas com diabetes.**

| Pessoas com diabetes | Apuração dos grupos | Nº de grupos |
|---|---|---|
| 7 | / | 1 |
| 8 | / / | 2 |
| 9 | / / / / / | 5 |
| 10 | / / / / / / / / | 8 |
| 11 | / / / | 3 |
| 12 | / | 1 |

Portanto, a variável "número de pessoas com diabetes" assume valores discretos, isto é, inteiros: ...,7,8,9,... .

**Dados quantitativos contínuos**: Os dados quantitativos contínuos assumem valores em um intervalo contínuo de números. Em geral, este tipo de dado é proveniente de medições de uma característica da qualidade de uma peça ou produto. Os possíveis valores incluem "todos" os números do intervalo de variação da característica medida. Por exemplo, ao medirmos os diâmetros dos eixos de determinados motores com uma célula eletrônica, obtemos dados quantitativos contínuos.

**Exemplo 1.3.4:**

Numa fábrica de motores elétricos, o gerente de produção precisa avaliar o problema de ruído excessivo do motor. Uma das possíveis causas está associada com variações no diâmetro do eixo. Assim, o gerente de produção mediu o diâmetro do eixo de 200 motores e o resultado está apresentado na Tabela a seguir. Os valores estão em milésimos de milímetros.

Podemos fazer a apuração considerando intervalos de medidas, como apresentado na Tabela 4.

**Tabela 4. Diâmetro do Eixo de 100 motores**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4,8 | 4,2 | 5,1 | 5,2 | 4,8 | 4,7 | 4,9 | 4,5 | 4,9 | 4,5 |
| 4,9 | 5,1 | 4,8 | 4,9 | 4,8 | 5 | 5,3 | 4,9 | 5,5 | 5,2 |
| 5,1 | 4,6 | 4,9 | 4,8 | 5,1 | 4,6 | 4,3 | 4,9 | 4,7 | 5,2 |
| 4,8 | 4,4 | 5,6 | 5 | 5 | 5 | 4,8 | 5,2 | 4,5 | 5,1 |
| 5,1 | 4,9 | 4,8 | 4,8 | 5 | 4,8 | 5,1 | 5,4 | 4,2 | 5,1 |
| 4,9 | 4,6 | 5,4 | 4,9 | 4,3 | 4,6 | 4,7 | 4,7 | 5,3 | 4,4 |
| 4,7 | 4,8 | 5,2 | 4,5 | 5,1 | 4,6 | 5,8 | 4,9 | 5,2 | 4,8 |
| 4,9 | 4,9 | 4,4 | 4,7 | 4,8 | 5,1 | 5,4 | 5 | 4,4 | 5,1 |
| 4,9 | 4,9 | 5,1 | 5,2 | 4,7 | 4,8 | 4,6 | 5,2 | 5,5 | 5,2 |
| 4,2 | 4,9 | 4,9 | 4,8 | 4,2 | 5,2 | 4,7 | 4,8 | 4,6 | 5,2 |

Um possível resumo dos dados (Tabela 5):

**Tabela 5. Diâmetro do eixo de 100 motores (com apuração)**

| Diâmetro | Apuração | Nº de motores apurados |
|---|---|---|
| [4,2;4,4) | ////// | 6 |
| [4,4;4,6) | //////// | 8 |
| [4,6;4,8) | /////////////// | 15 |
| [4,8;5,0) | //////////...////////// | 33 |
| [5,0;5,2) | ////////////////// | 18 |
| [5,2;5,4) | ///////////// | 13 |
| [5,4;5,6) | ///// | 5 |
| [5,6;5,8) | // | 2 |

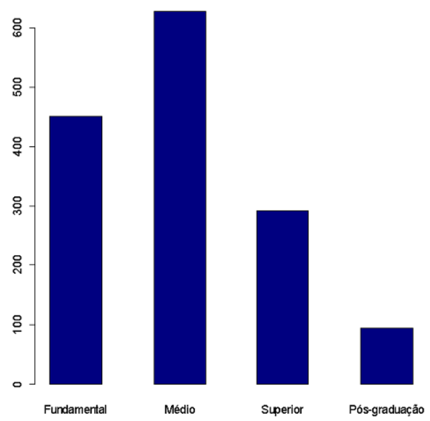
Ao estabelecermos intervalos de classes, estamos admitindo que o eixo pode assumir qualquer valor entre o limite inferior (inclusive) e o limite superior (exclusive).

**Exemplo 1.4.1:**

Considere os dados do Exemplo 1.3.2. Construa o gráfico de barras correspondente.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

O gráfico de barras apresenta dados categorizados em barras retangulares nos quais os retângulos correspondentes a cada categoria são proporcionais ao número de observações na respectiva categoria. O gráfico de barras é utilizado para realizar comparações entre as categorias de uma variável qualitativa ou quantitativa discreta. Este gráfico pode ser utilizado na vertical ou horizontal. No exemplo 1.4.1, a altura de cada retângulo corresponde ao número de inscritos com o respectivo grau de escolaridade. Podemos utilizar o Actio Stat para executarmos o gráfico de Barras.




Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## 1.6. Histograma

## Distribuição de frequências

A distribuição de frequências é um agrupamento de dados em classes, de tal forma que contabilizamos o número de ocorrências em cada classe. O número de ocorrências de uma determinada classe recebe o nome de frequência absoluta. O objetivo é apresentar os dados de uma maneira mais concisa e que nos permita extrair informação sobre seu comportamento. A seguir, apresentamos algumas definições necessárias à construção da distribuição de frequências.

**Frequência absoluta ($f$i)**: É o número de observações correspondente a cada classe. A frequência absoluta é, geralmente, chamada apenas de frequência.

**Frequência relativa ($f$ri)**: É o quociente entre a frequência absoluta da classe correspondente e a soma das frequências (total observado), isto é, $f_{ri} = \dfrac{f_i}{\sum_j f_j}$ onde n representa o número total de observações.

**Frequência percentual (pi):** É obtida multiplicando a frequência relativa por 100%.

**Frequência acumulada**: É o total acumulado (soma) de todas as classes anteriores até a classe atual. Pode ser: frequência acumulada absoluta (Fi), frequência acumulada relativa (Fri), ou frequência acumulada percentual (Pi).

## Distribuição de frequência pontual: dados quantitativos discretos

A construção de uma tabela de distribuição de frequência pontual é equivalente à construção de uma tabela simples, onde se listam os diferentes valores observados da variável com suas frequências absolutas, denotadas por (ƒi) (o índice i corresponde ao número de linhas da Tabela) como é mostrado na Tabela 6. Utilizamos a distribuição de frequência pontual quando se trabalha com dados discretos. Um gráfico utilizado para representar este tipo de distribuição de frequência é o Gráfico de Barras.

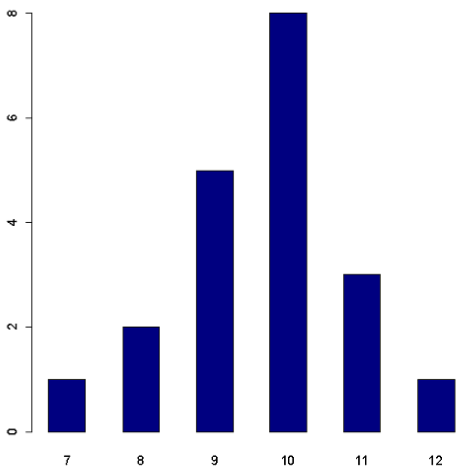
**Exemplo 1.6.1:**

Considere os dados do Exemplo 1.3.3. Construa a distribuição de frequências para este conjunto de dados e o gráfico de barras.

**clique aqui para efetuar o download dos dados utilizados nesse exemplo**

**Tabela 6. Tabela de distribuição de frequência pontual.**

| Número de pessoas com diabetes | Frequência (ƒi) | Frequência relativa (ƒri) | Frequência percentual | Frequência acumulada |
|---|---|---|---|---|
| 7 | 1 | 0,05 | 5 | 5 |
| 8 | 2 | 0,1 | 10 | 15 |
| 9 | 5 | 0,25 | 25 | 40 |
| 10 | 8 | 0,4 | 40 | 80 |
| 11 | 3 | 0,15 | 15 | 95 |
| 12 | 1 | 0,05 | 5 | 100 |




Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## Distribuição de frequência em intervalos de classes: Dados quantitativos contínuos

Para dados quantitativos contínuos, geralmente resultantes de medições de características da qualidade de peças ou produtos, dividimos a faixa de variação dos dados em intervalos de classes. O menor valor da classe é denominado limite inferior (li) e o maior valor da classe é denominado limite superior (Li).

O intervalo ou classe pode ser representado das seguintes maneiras:

1. (li)$\vdash$(Li), onde o limite inferior da classe é incluído na contagem da frequência absoluta, mas o superior não;

2. (li)$\dashv$(Li) , onde o limite superior da classe é incluído na contagem, mas o inferior não.

Podemos escolher qualquer uma destas opções, mas é importante que deixemos claro no texto ou na tabela qual delas está sendo usada. Embora não seja necessário, os intervalos são frequentemente construídos de modo que todos tenham larguras iguais, o que facilita as comparações entre as classes.

Na tabela de distribuição de frequência, acrescentamos uma coluna com os pontos médios de cada intervalo de classe, denotada por xi. Esta é definida como a média dos limites da classe $x_i = \dfrac{l_i + L_i}{2}$. Estes valores são utilizados na construção de gráficos.

Algumas indicações na construção de distribuição de frequências são:

- Na medida do possível, as classes deverão ter amplitudes iguais.
- Escolher os limites dos intervalos entre duas possíveis observações.
- O número de intervalos não deve ultrapassar 20.
- Escolher limites que facilitem o agrupamento.
- Marcar os pontos médios dos intervalos.
- Ao construir o histograma, cada retângulo deverá ter área proporcional à frequência relativa (ou à frequência absoluta, o que dá no mesmo) correspondente.

Um ponto importante na construção da distribuição de frequência é o numero de intervalos de classes. No Action Stat utilizamos a regra de Sturges para obter o número de intervalos de classes $k = 1 + 3,3\ log_{10}(n)$, no qual $n$ é o tamanho do conjunto de dados.

# Histograma

Histograma é uma representação gráfica (um gráfico de barras verticais ou barras horizontais) da distribuição de frequências de um conjunto de dados quantitativos contínuos. O histograma pode ser um gráfico por valores absolutos ou frequência relativa ou densidade. No caso de densidade, a frequência relativa do intervalo i, (fri), é representada pela área de um retângulo que é colocado acima do ponto médio da classe i. Consequentemente, a área total do histograma (igual a soma das áreas de todos os retângulos) será igual a 1. Assim, ao construir o histograma, cada retângulo deverá ter área proporcional à frequência relativa (ou à frequência absoluta, o que é indiferente) correspondente. No caso em que os intervalos são de tamanhos (amplitudes) iguais, as alturas dos retângulos serão iguais às frequências relativas (ou iguais às frequências absolutas) dos intervalos correspondentes.

**Exemplo 1.6.2:**

Considerando os dados do Exemplo 1.3.4, monte a distribuição de frequências e construa o histograma correspondente.

**clique aqui para efetuar o download dos dados utilizados nesse exemplo**

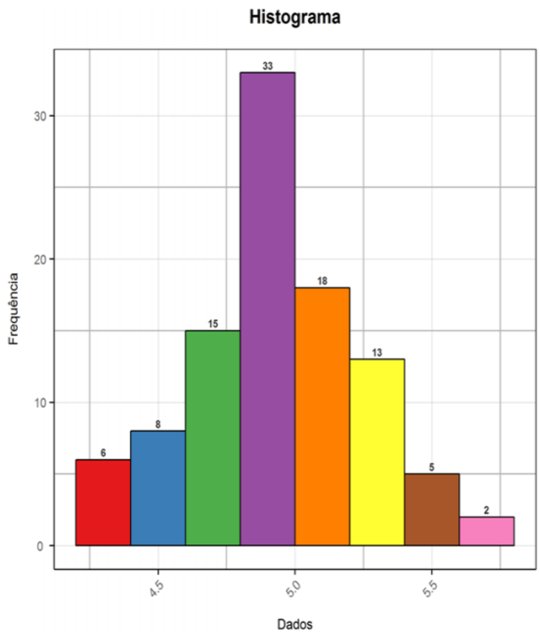
Como temos dados quantitativos contínuos, para construir a distribuição de frequências, vamos separar os dados em classes. Ao aplicarmos a regra de Sturges obtemos $k = 1 + 3,3\ log_{10}(100)\ = 7,6.$

Assim, dividimos os dados em 8 classes de tamanhos iguais. A distribuição de frequências então é a seguinte

**Tabela 7. Tabela de Frequências**

| Classe | Frequência | Freq. Rel. | Freq. Perc. | Freq. Acum. | Densidades | Ponto Médio |
|---|---|---|---|---|---|---|
| [4,2 ; 4,4) | 6 | 0,06 | 6 | 6 | 0,3 | 4,3 |
| [4,4 ; 4,6) | 8 | 0,08 | 8 | 14 | 0,4 | 4,5 |
| [4,6 ; 4,8) | 15 | 0,15 | 15 | 29 | 0,75 | 4,7 |
| [4,8 ; 5) | 33 | 0,33 | 33 | 62 | 1,65 | 4,9 |
| [5 ; 5,2) | 18 | 0,18 | 18 | 80 | 0,9 | 5,1 |
| [5,2 ; 5,4) | 13 | 0,13 | 13 | 93 | 0,65 | 5,3 |
| [5,4 ; 5,6) | 5 | 0,05 | 5 | 98 | 0,25 | 5,5 |
| [5,6 ; 5,8) | 2 | 0,02 | 2 | 100 | 0,1 | 5,7 |

A seguir, apresentamos o histograma obtido com o software Action Stat.




Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

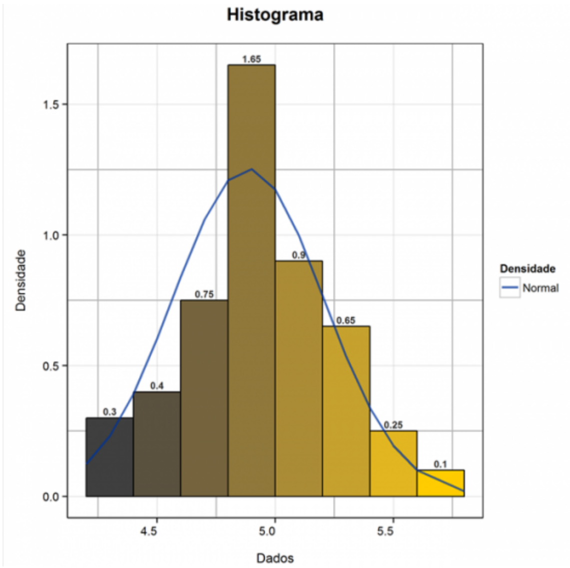
Muitas vezes, queremos representar a curva de uma distribuição de probabilidade (por exemplo, normal) junto com o histograma. Esta é uma forma visual de avaliar o ajuste dos dados pela referida distribuição de probabilidade. Entretanto, como a área total da distribuição de probabilidade é igual a $1$, não faz sentido utilizarmos a altura do retângulo como a frequência absoluta ou a frequência relativa. Neste caso, comparamos elementos em escalas distintas. Para contornar este problema, sugerimos utilizar a área de cada retângulo como a frequência relativa cuja a soma é igual a 1. Neste caso, a altura de cada retângulo é dada pela densidade (D) que corresponde frequência relativa dividida pelo tamanho do intervalo de classe.

**Exemplo 1.6.3:**

Considerando os dados do Exemplo 1.3.4, construa o histograma de densidades correspondente

**clique aqui para efetuar o download dos dados utilizados nesse exemplo**

Para construir o histograma de densidades, basta que os retângulos tenham altura do tamanho da densidade de cada classe e largura do tamanho da classe. Neste caso, o histograma ficaria da seguinte forma:




Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## 1.7. Gráfico de Pizza

O gráfico de pizza, também conhecido como gráfico de setores ou gráfico circular é um diagrama circular onde os valores de cada categoria estatística representada são proporcionais às respectivas frequências. Este gráfico pode vir acompanhado de porcentagens. É utilizado para dados qualitativos nominais. Para construir um gráfico tipo pizza é necessário determinar o ângulo dos setores circulares correspondentes à contribuição percentual de cada valor no total.

## Exemplo 1.7.1:

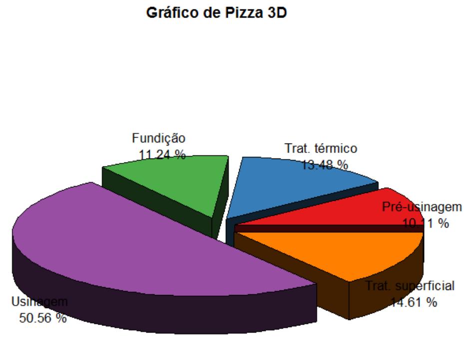
Uma empresa da área automobilística acompanha o número de defeitos encontrados nos equipamentos enviados para a calibração. Na Tabela 8 apresentamos os dados referentes a um mês de acompanhamento dos defeitos encontrados nos equipamentos das diversas áreas.

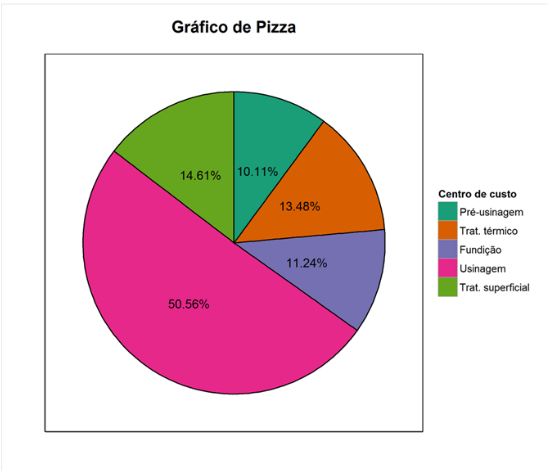
**Tabela 8. Acompanhamento dos defeitos encontrados.**

| Centro de custo | Número de defeitos |
|---|---|
| Pré-usinagem | 9 |
| Tratamento térmico | 12 |
| Fundição | 10 |
| Usinagem | 45 |
| Tratamento superficial | 13 |
| Total | 89 |

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Como temos um total de 89 defeitos, o setor circular de 360º será equivalente a 89. Calculando as proporções, encontramos os ângulos correspondentes aos números de defeitos de cada centro de custo. Com isso, construímos o seguinte gráfico de pizza.



| Frequências Relativas | | | |
|---|---|---|---|
| Categorias | Frequências | Freq. Perc. | Freq. Acum. |
| Pré-usinagem | 9 | 10,11 | 10,11 |
| Trat. térmico | 12 | 13,48 | 23,59 |
| Fundição | 10 | 11,24 | 34,83 |
| Usinagem | 45 | 50,56 | 85,39 |
| Trat. superficial | 13 | 14,61 | 100 |
| Total | 89 | 100 | 100 |

Tal exemplo pode ser realizado utilizando o software Action.

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

# 2 - ESTATÍSTICAS DESCRITIVAS

As estatísticas descritivas são números que resumem e descrevem o conjunto de dados. As estatísticas descritivas apenas "descrevem" os dados, elas não representam generalizações da amostra para a população. A técnica utilizada para estender conclusões da amostra para a população é a inferência. A seguir, apresentamos as medidas básicas de uma análise descritiva dos dados: as medidas de posição, medidas de dispersão, quartis, coeficiente de assimetria, coeficiente de curtose e o esquema dos cinco números.

São as estatísticas que representam uma série de dados orientando-nos quanto à posição da distribuição em relação ao eixo horizontal do gráfico da curva de frequência. As medidas de posições mais importantes são média aritmética, mediana e moda. Usaremos as seguintes notações:

x: valor de cada indivíduo da amostra.

$\overline{x}$: média amostral.

n: tamanho amostral.

## Média populacional

A média populacional é calculada somando-se todos os valores da população e dividindo o resultado pelo total de elementos da população. Numa população de $N$ elementos, a média populacional é dada por $\mu = \dfrac{x_1 + \ldots + x_N}{N}$.

## Média Amostral

A média amostral, aritmética, ou simplesmente média, é calculada somando-se os valores das observações da amostra e dividindo-se o resultado pelo número de valores. Assim, a média amostral é dada por $\overline{x} = \dfrac{x_1 + \ldots + x_n}{n}$

**Exemplo 2.1.1:**

Uma amostra de 5 barras de aço foi retirada da linha de produção e seus comprimentos foram medidos. Os valores foram: 4,5; 4,6; 4,5; 4,4; 4,5.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Desta forma, a média é dada por $\overline{x} = \dfrac{4,5 + 4,6 + 4,5 + 4,4 + 4,5}{5} = 4,5$.

Podemos utilizar o Action para resolver este problema. O resultado obtido foi:

| | | | |
|---|---|---|---|
| **SUMÁRIO BÁSICO E POR FATOR** | | | |
| | | | |
| ***DADOS DO PROCESSO*** | | | |
| | | | |
| ***Média*** | | | |
| 4,5 | | | |

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

**Exemplo 2.1.2:**

Foram medidos os comprimentos de 5 leitos hospitalares e os valores (em metros) obtidos foram: 2,26; 2,30; 2,31; 2,28; 2,32.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

A média é então, dada por $\overline{X} = \frac{2,26 + 2,30 + 2,31 + 2,28 + 2,32}{5} = \frac{11,47}{5} = 2,294.$

Utilizando o Action, temos que

| SUMÁRIO BÁSICO E POR FATOR | | | |
|---|---|---|---|
| | | | |
| *DADOS DO PROCESSO* | | | |
| | | | |
| *Média* | | | |
| 2,294 | | | |

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## Mediana

Para calcular a mediana devemos, em primeiro lugar, ordenar os dados do menor para o maior valor. Se o número de observações for ímpar, a mediana será a observação central. Se o número de observações for par, a mediana será a média aritmética das duas observações centrais. Notação: $\tilde{X}$.

## Exemplo 2.1.3:

Consideremos os seguintes dados correspondentes aos comprimentos de 8 rolos de fio de aço: 65, 72, 70, 72, 60, 67, 69, 68.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Ordenando os valores temos: 60, 65, 67, 68, 69, 70, 72, 72. Como o número de observações é par, a mediana é dada pela média dos dois valores centrais que são 68 e 69, isto é, $\tilde{X} = \frac{68 + 69}{2} = 68,5$

Também resolvemos o problema utilizando o Action. Os resultados são mostrados a seguir

| Informação | Valor |
|---|---|
| Mediana | 68,5 |

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## Moda

A moda de um conjunto de valores é o valor que apresenta a maior frequência.

## Exemplo 2.1.4:

Considerando os dados do Exemplo 2.1.3 temos que sua moda é 72, pois este é o valor do conjunto de dados que aparece com maior frequência.

## Exemplo 2.1.5:

Em um hospital, foram contabilizados o número de pessoas atendidas pela ortopedia durante os 30 dias de um mês. Os valores observados estão apresentados na Tabela 9:

Tabela 9. Número de pessoas atendidas pela ortopedia

| | | | | |
|---|---|---|---|---|
| 119 | 118 | 125 | 115 | 107 |
| 128 | 133 | 133 | 121 | 101 |
| 118 | 143 | 126 | 117 | 141 |
| 109 | 135 | 115 | 115 | 119 |
| 131 | 116 | 115 | 124 | 134 |
| 140 | 129 | 129 | 115 | 119 |

A média dos dados é dada por: $\overline{X} = \frac{119 + 118 + \cdots + 115 + 119}{30} \approx 123$

Portanto, temos em média aproximadamente 123 pessoas atendidas pela ortopedia, diariamente.

Para o calculo da mediana, inicialmente ordenamos os dados:

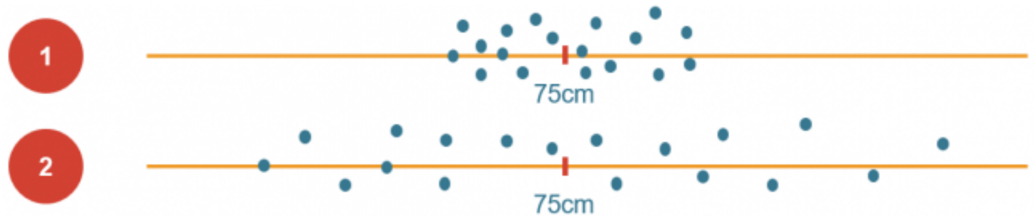
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 101 | 107 | 109 | 115 | 115 | 115 | 115 | 115 | 116 | 117 |
| 118 | 118 | 119 | 119 | 119 | 121 | 124 | 125 | 126 | 128 |
| 129 | 129 | 131 | 133 | 133 | 134 | 135 | 140 | 141 | 143 |

Como n=30 é par, selecionamos as duas observações centrais e calculamos a média $\tilde{X} = \frac{119 + 121}{2} = 120$.

Portanto, metade das informações estão localizadas abaixo de 120 e a outra metade acima de 120.

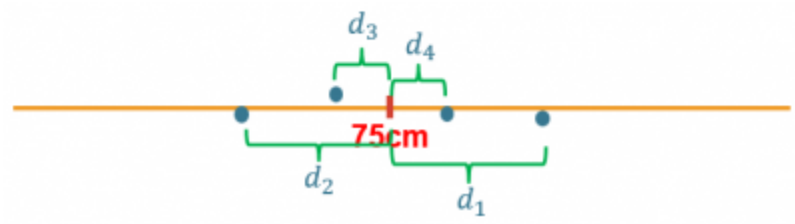
## 2.2. Medidas de Dispersão

Considere o exemplo de duas linhas de produção de uma peça. A medida média do comprimento da peça é de 75cm e ambas as linhas estão produzindo peças com médias próximas desse valor. Podemos considerar que as peças produzidas por ambas as linhas são adequadas?




É claro que as peças produzidas pela primeira linha de produção são melhores que a segunda. Isso ocorre porque a dispersão dos elementos em torno da média é menor, ou seja, os elementos estão mais concentrados em torno da média na primeira linha de produção.

Como queremos avaliar a dispersão dos dados em torno da média, esse valor estará relacionado com a distância dos dados em relação à média. Essa distância será chamada de desvio, $d_i$.$d_i = X_i - \overline{X}$




No exemplo da imagem acima, temos $d_1 + d_2 + d_3 + d_4 = 0$

O qual nos levaria à conclusão errada de que não existe variação entre os dados. Desta forma, precisamos de algumas medidas estatísticas para poder estudar a dispersão dos dados de forma correta.

Dispersão é sinônimo de variação ou variabilidade. Para medir a dispersão, duas medidas são usadas mais frequentemente: a **amplitude** e o **desvio padrão.**

## Amplitude

A amplitude é definida como sendo a diferença entre o maior e o menor valor do conjunto de dados. Denotaremos a amplitude por R, portanto, consideremos o conjunto de dados ordenado $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \cdots \leq X_{(n-1)} \leq X_{(n)}$

A amplitude R dos dados é dada por: $R = X_{(n)} - X_{(1)}$

## Exemplo 2.2.1:

Considere o Exemplo 2.1.3. Qual a amplitude deste conjunto de dados?

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Como o valor máximo do conjunto é 72 e o valor mínimo é 60, temos que a amplitude é: R = 72 - 60 = 12.

Utilizando o Action, temos o seguinte resultado:

| Informação | Valor |
|---|---|
| Amplitude | 12 |

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

Para definirmos desvio padrão é necessário definir variância. A notação mais comumente usada é:

**s2** - variância amostral.

**σ2** - variância populacional.

**s** - desvio padrão amostral.

**σ** - desvio padrão populacional.

## Variância populacional

A variância de uma população {x1,...,xN} de N elementos é a medida de dispersão definida como a média do quadrado dos desvios dos elementos em relação à média populacional μ. Ou seja, a variância populacional é dada por: $\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}$

## Variância amostral

A variância de uma amostra {x1,...,xn} de n elementos é definida como a soma ao quadrado dos desvios dos elementos em relação à sua média $\overline{x}$ dividido por (n-1). Ou seja, a variância amostral é dada por: $s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$

Ao utilizarmos a média amostral como estimador de m para calcularmos a variância amostral, perdemos 1 grau de liberdade em relação à variância populacional.

## Desvio padrão populacional

O desvio padrão populacional de um conjunto de dados é igual à raiz quadrada da variância populacional. Desta forma, o desvio padrão populacional é dado por: $\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}}$

## Desvio padrão amostral

O desvio padrão amostral de um conjunto de dados é igual à raiz quadrada da variância amostral. Desta forma, o desvio padrão amostral é dado por: $s = \sqrt{s^2} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}}$

## Exemplo 2.2.2:

Considere novamente os dados do Exemplo 2.1.3. Calcule o desvio padrão dos dados.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Para calcularmos o desvio padrão devemos primeiramente calcular a média $\overline{x}$, isto é:

$$\overline{x} = \frac{65+72+70+72+60+67+69+68}{8} = 67,875$$.

Agora vamos subtrair $\overline{x}$ de cada valor, elevar os resultados ao quadrado e somá-los. Então dividimos o total dos quadrados pelo número de valores menos 1, ou seja, por (n-1) e extraímos a raiz quadrada:

| $(x - \overline{x})$ | $(x - \overline{x})^2$ |
|---|---|
| 65-67,875 = -2,875 | (-2,875)2 = 8,265625 |
| 72-67,875 = 4,125 | (4,125)2 = 17,015625 |
| 70-67,875 = 2,125 | (2,125)2 = 4,515625 |
| 72-67,875 = 4,125 | (4,125)2 = 17,015625 |
| 60-67,875 = -7,875 | (-7,875)2 = 62,015625 |
| 67-67,875 = -0,875 | (-0,875)2 = 0,765625 |
| 69-67,875 = 1,125 | (1,125)2 = 1,265625 |
| 68-67,875 = 0,125 | (0,125)2 = 0,015625 |
| | **Total** = 110,875 |

$$\frac{110,875}{7} = 15,83929 \Rightarrow s = \sqrt{15,83929} \Rightarrow s = 3,97986$$.

Portanto, o desvio padrão é 3,97986.

Utilizando o Action, temos o seguinte resultado:

| Informação | Valor |
|---|---|
| Desvio-padrão | 3,97986 |

## Exemplo 2.2.3:

Consideremos o Exemplo 2.1.5, em que foram contabilizados o número de pessoas atendidas pela ortopedia durante os 30 dias de um mês. Os valores observados estão apresentados na Tabela 10.

Tabela 10. Número de pessoas atendidas pela ortopedia

| | | | | |
|---|---|---|---|---|
| 119 | 118 | 125 | 115 | 107 |
| 128 | 133 | 133 | 121 | 101 |
| 118 | 143 | 126 | 117 | 141 |
| 109 | 135 | 115 | 115 | 119 |
| 131 | 116 | 115 | 124 | 134 |
| 140 | 129 | 129 | 115 | 119 |

Vimos que $\overline{X} = 123$

Calculando a variância, temos:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n}(X_i - 123)^2}{30-1} = \frac{(119-123)^2 + (118-123)^2 + \cdots + (119-123)^2}{29} = 106,7586$$

O desvio padrão é dado por $S = \sqrt{S^2} = \sqrt{106,7586} = 10,3324$

Observamos que o desvio-padrão representa pouco menos de 10% do valor da média.

O calculo da amplitude é dado por $R = X_{(30)} - X_{(1)} = 143 - 101 = 42$.

Portanto, o tamanho do intervalo em que os dados estão inseridos é de 42.

Para entender como executar essa função do Software Action, você pode consultar o manual do usuário.

## 2.3. QUARTIS

Uma análise das estatísticas descritivas da amostra é fundamental para resumirmos algumas informações sobre a população. Estas informações são utilizadas para tomada de decisão e formação de modelos estatísticos paramétricos. Definiremos como:

**Mínimo**: menor elemento da amostra;

**Máximo**: maior elemento da amostra;

**Quartis** (Q1, Q2 e Q3): São valores dados a partir do conjunto de observações ordenado em ordem crescente, que dividem a distribuição em quatro partes iguais. O primeiro quartil, Q1, é o número que deixa 25% das observações abaixo e 75% acima, enquanto que o terceiro quartil, Q3, deixa 75% das observações abaixo e 25% acima. Já Q2 é a mediana, deixa 50% das observações abaixo e 50% das observações acima.

Seja n o número total de elementos da amostra e calcule j(n+1)/4, para j=1,2 e 3. Desta forma Qj será um elemento entre Xk e Xk+1, onde k é o maior inteiro menor ou igual a j(n+1)/4 e será calculado da seguinte forma

$$Q_j = X_k + \left(\frac{j(n+1)}{4} - k\right)(X_{k+1} - X_k).$$

Podemos observar que quando $k$ é um valor inteiro, o quantil será o próprio $X_k$, isto é, $Q_j = X_k$, onde $k = \frac{j(n+1)}{4}, j = 1, 2, 3$.

### Exemplo 2.3.1:

Considere uma amostra de 6 elementos com os seguintes valores: 7,1; 7,4; 7,5; 7,7; 7,8; 7,9.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Deste modo temos que (n+1)/4 = 7/4 = 1,75 e com isso k = 1, logo

$$Q_1 = X_1 + \left(\frac{n+1}{4} - k\right)(X_{k+1} - X_k) = 7,1 + (1,75 - 1)(7,4 - 7,1) = 7,1 + 0,75(0,3) = 7,325.$$

Também temos que 2(n+1)/4 = 14/4 = 3,5, com isso k = 3, logo

$$Q_2 = X_3 + \left(\frac{2(n+1)}{4} - k\right)(X_{k+1} - X_k) = 7,5 + (3,5 - 3)(7,7 - 7,5) = 7,5 + 0,5(0,2) = 7,6.$$

E, temos que 3(n+1)/4= 21/4 = 5,25, com isso k = 5, logo

$$Q_3 = X_5 + \left(\frac{3(n+1)}{4} - k\right)(X_{k+1} - X_k) = 7,8 + (5,25 - 5)(7,9 - 7,8) = 7,8 + 0,25(0,1) = 7,825.$$

Podemos utilizar o Action para o cálculo dos quartis. Os resultados são dados abaixo

SUMÁRIO BÁSICO

*DADOS DO PROCESSO*

| Informação | Valor |
|---|---|
| 1º Quartil | 7,325 |
| Mediana | 7,6 |
| 3º Quartil | 7,825 |

Para entender como executar essa função do **Software Action**, v

você pode consultar o **manual do usuário.**

## Exemplo 2.3.2:

Considere o Exemplo 2.1.3, calcule os quartis dos dados.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Primeiramente ordenamos os dados, 60, 65, 67, 68, 69, 70, 72, 72.

Deste modo temos que (n+1)/4 = 9/4 = 2,25 e com isso k = 2, logo

$$Q_1 = X_2 + \left(\frac{n+1}{4} - k\right)(X_{k+1} - X_k) = 65 + (2,25 - 2)(67 - 65) = 65 + 0,25(2) = 65,5.$$

Também temos que 2(n+1)/4 = 18/4 = 4,5, com isso k = 4, logo

$$Q_2 = X_4 + \left(\frac{2(n+1)}{4} - k\right)(X_{k+1} - X_k) = 68 + (4,5 - 4)(69 - 68) = 68 + 0,5(1) = 68,5.$$

E, temos que 3(n+1)/4= 27/4 = 6,75, com isso k = 6, logo

$$Q_3 = X_6 + \left(\frac{3(n+1)}{4} - k\right)(X_{k+1} - X_k) = 70 + (6,75 - 6)(72 - 70) = 70 + 0,75(2) = 71,5.$$

Os mesmos podem ser calculados no Software Action

| SUMÁRIO BÁSICO E POR FATOR | | |
|---|---|---|
| | | |
| *DADOS DO PROCESSO* | | |
| | | |
| ***1º Quartil*** | ***Mediana*** | ***3º Quartil*** |
| 65,5 | 68,5 | 71,5 |

Para entender como executar essa função do **Software Action**,

você pode consultar o **manual do usuário.**

## Exemplo 2.3.3:

Suponha que uma amostra dos comprimentos de 11 rolos de fio de aço cujos valores foram 72, 70, 77, 60, 67, 69, 68, 66, 65, 71, 69.

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Os dados ordenados de forma crescente são: 60, 65, 66, 67, 68, 68, 69, 70, 71, 72, 77. Então temos que:

Mínimo = 60.

Máximo = 77.

Posição do Q1 = $\frac{11+1}{4} = 3 \Rightarrow Q_1 = 66$

Logo, 25% das observações estão abaixo de 66 e 75% das observações estão acima de 66.

Posição do Q3 = $3 \times \left(\frac{11+1}{4}\right) = 9 \Rightarrow Q_3 = 71$

Portanto, 75% das observações estão abaixo de 71 e 25% das observações estão acima de 71.

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## Exemplo 2.3.4:

Considere as medidas das alturas de 11 pacientes, dadas abaixo

| Altura dos pacientes | | | |
|---|---|---|---|
| 1,59 | 1,79 | 1,68 | 1,80 |
| 1,58 | 1,60 | 1,69 | 1,73 |
| 1,87 | 1,68 | 1,85 | |

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Ordenando os valores, temos que

$$1,58 \leq 1,59 \leq 1,60 \leq 1,68 \leq 1,68 \leq 1,69 \leq 1,73 \leq 1,79 \leq 1,80 \leq 1,85 \leq 1,87.$$

Desta forma, temos que o valor mínimo é 1,58, o valor máximo é 1,87. Dado que temos 11 observações, o calculo do primeiro quartil é:

$$\frac{1(n+1)}{4} = \frac{1(11+1)}{4} = 3 \Longrightarrow k = 3$$

$$Q_1 = X_3 + \left(\frac{n+1}{4} - 3\right)(X_4 - X_3) = X_3 = 1,60$$

Para o segundo quartil temos:

$$\frac{2(n+1)}{4} = \frac{(2(11+1)}{4} = 6 \Longrightarrow k = 6$$

$$Q_2 = X_6 + \left(\frac{2(n+1)}{4} - 6\right)(X_7 - X_6) = X_6 nbsp; = 1,69$$

O terceiro quartil é dado por

$$\frac{3(n+1)}{4} = \frac{(3(11+1)}{4} = 9 \Longrightarrow k = 9$$

$$Q_3 = X_9 + \left(\frac{3(n+1)}{4} - 9\right)(X_{10} - X_9) = X_9 nbsp; = 1,80$$

Utilizando o software Action, temos os seguinte resultados.

| | |
|---|---|
| *Mínimo* | 1,58 |
| *1º Quartil* | 1,6 |
| *Tri-Média* | 1,714545 |
| *3º Quartil* | 1,8 |
| *Máximo* | 1,87 |
| *Assimetria* | 0,111765 |
| *Kurtosis* | -1,569809 |
| *Amplitude* | 0,29 |

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## Exemplo 2.3.5:

Considerando os valores diários de um retorno do índice S&P 500, ano de 2010. Encontre os coeficientes de assimetria e curtose.

clique aqui para efetuar o download dos dados

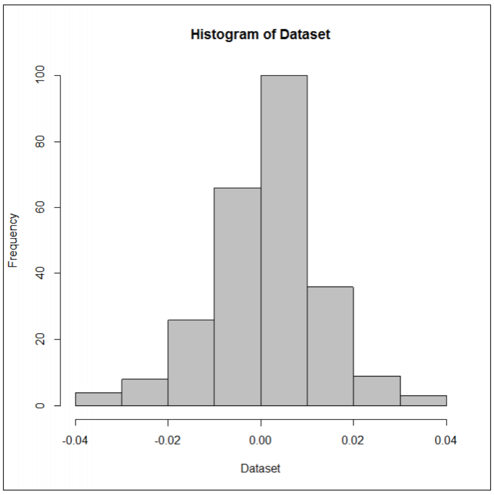
Utilizando o Software Action temos os seguintes resultados

**SUMÁRIO BÁSICO E POR FATOR**

***DADOS DO PROCESSO***

| Mínimo | 1º Quartil | Média | Mediana | 3º Quartil | Máximo | Desvio Padrão da Média | Desvio Padrão | Assimetria | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|
| -0,038172 | -0,004556 | 0,000879 | 0,001081 | 0,007333 | 0,03927 | 0,000734 | 0,011646 | -0,280502 | 1,347751 |

Observamos uma curtose alta de 1,35 e assimetria negativa de -0,28. Este é um fenômeno típico de dados financeiros, que são assimétricos e possuem caudas pesadas.



# 3. GRÁFICOS

Gráficos estatísticos são formas de apresentação dos dados estatísticos, cujo objetivo é o de produzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo. A representação gráfica de um fenômeno deve obedecer a certos requisitos fundamentais para ser realmente útil:

Simplicidade: O gráfico deve ser destituído de detalhes de importância secundária, assim como de traços desnecessários que possam levar o observador a uma análise com erros.

Clareza: O gráfico deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.

Veracidade: O gráfico deve expressar a verdade sobre o fenômeno em estudo.

Alguns exemplos de gráficos já foram vistos acima, como o histograma e o diagrama de Pareto. A seguir apresentamos o *boxplot* e o *dotplot*.

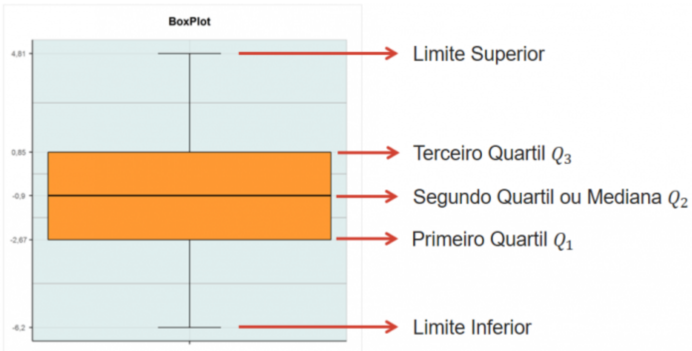
## 3.1. Boxplot

O boxplot (gráfico de caixa) é um gráfico utilizado para avaliar a distribuição empírica dos dados. O boxplot é formado pelo primeiro e terceiro quartil e pela mediana. As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior ao limite inferior e do quartil superior até o maior valor não superior ao limite superior. Os limites são calculados da forma abaixo

Limite inferior: $\max\{\min(\text{dados}); Q_1 - 1,5(Q_3 - Q_1)\}$.

Limite superior: $\min\{\max(\text{dados}); Q_3 + 1,5(Q_3 - Q_1)\}$.

Para este caso, os pontos fora destes limites são considerados valores discrepantes (outliers) e são denotados por asterisco (*). A Figura a seguir apresenta um exemplo do formato de um boxplot.




O boxplot pode ainda ser utilizado para uma comparação visual entre dois ou mais grupos. Por exemplo, duas ou mais caixas são colocadas lado a lado e se compara a variabilidade entre elas, a mediana e assim por diante. Outro ponto importante é a diferença entre os quartis $(Q_3 - Q_1)$ que é uma medida da variabilidade dos dados.

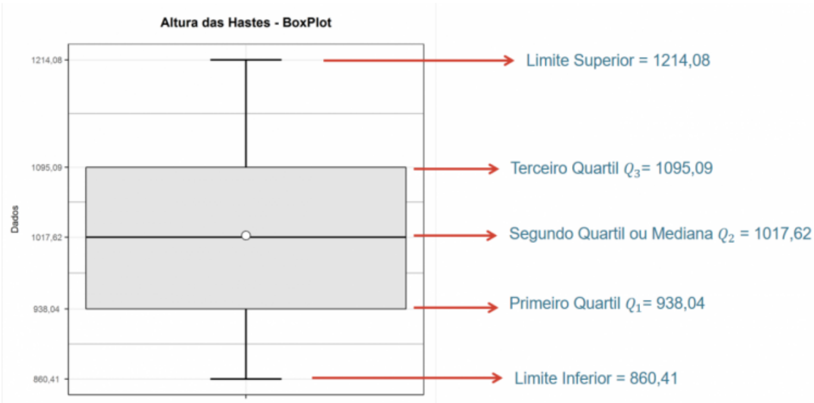
### Exemplo 3.1.1:

Na Tabela a seguir temos as medidas da altura de 20 hastes. Faça o boxplot correspondente.

| 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 |

clique aqui para efetuar o download dos dados utilizados nesse exemplo

O boxplot é dado por



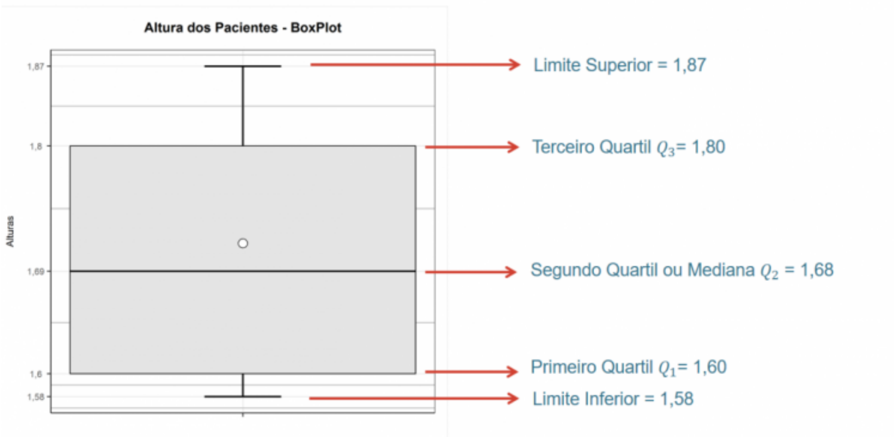

## Exemplo 3.1.2:

Utilizando os dados do Exemplo 2.3.4, temos

clique aqui para efetuar o download dos dados utilizados nesse exemplo

| | |
|---|---|
| Mínimo | 1,58 |
| 1° Quartil | 1,6 |
| Tri-Média | 1,714545 |
| 3° Quartil | 1,8 |
| Máximo | 1,87 |
| Assimetria | 0,111765 |
| Curtose | -1,569809 |
| Amplitude | 0,29 |

Assim, obtemos o seguinte boxplot



Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

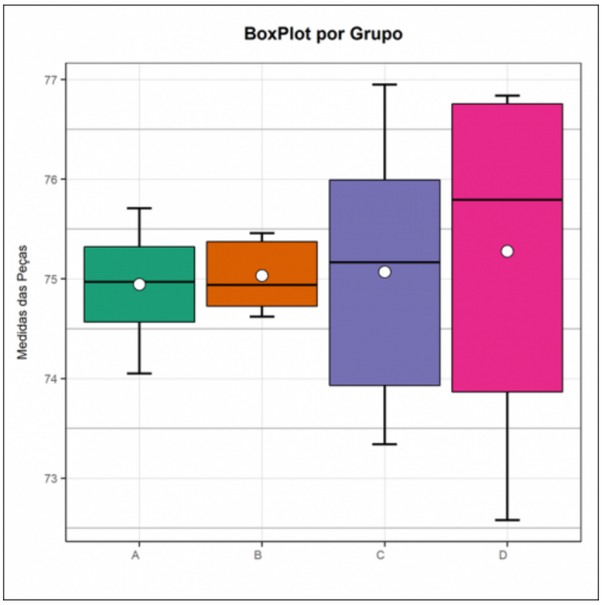
Podemos utilizar o Boxplot para comparar dados estratificados e comparar diferenças nas distribuições empíricas dos estratos

## Exemplo 3.1.3

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. 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 |

clique aqui para efetuar o download dos dados utilizados nesse exemplo




Analisando 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.

**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.

## 3.2. Dotplot

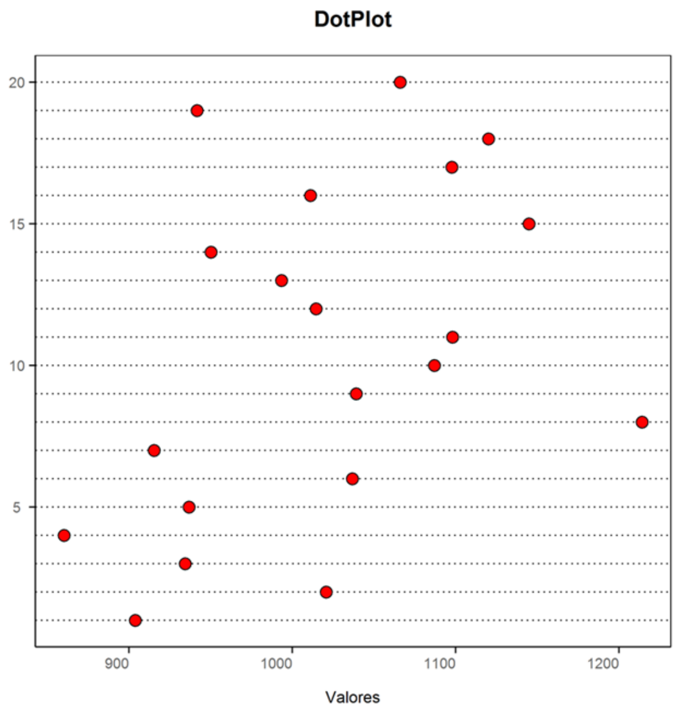
O gráfico Dotplot (gráfico de pontos) representa cada observação obtida em uma escala horizontal, permitindo visualizar a distribuição dos dados ao longo deste eixo. No eixo horizontal, dividimos a escala dos valores em pequenos intervalos, sendo marcado um ponto por observação.

O Dotplot, ou gráfico de pontos, é muito útil para visualizar estratificações. A estratificação é uma técnica que agrupa dados em subgrupos, de acordo com determinados critérios, aumentando o poder da análise.

### Exemplo 3.2.1:

Considerando novamente os dados do Exemplo 3.1.1, construímos o gráfico Dotplot utilizando o Action Stat:

clique aqui para efetuar o download dos dados utilizados nesse exemplo




Para entender como executar essa função do **Software Action**,

você pode consultar o **manual do usuário.**

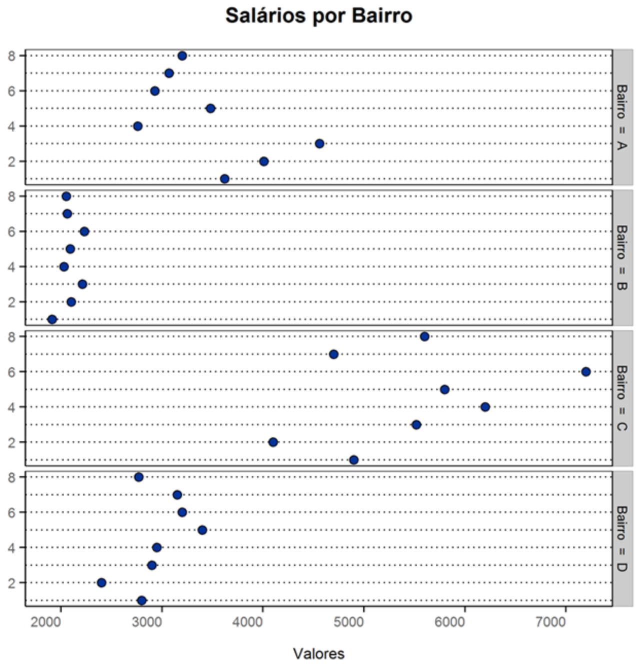
Eventualmente, é interessante analisar dados estratificados, ou seja, realizar um dotplot em grupos. Podemos analisar, graficamente, qual grupo (estrato) possui dispersão maior e ter uma ideia da localização da média de cada grupo.

### Exemplo 3.2.2:

Uma pequena pesquisa realizada em 4 bairros (A, B, C e D) de uma cidade perguntava sobre o salário líquido de 32 famílias (8 famílias de cada bairro). Os resultados obtidos (dados em reais) estão na tabela abaixo clique aqui para efetuar o download dos dados utilizados nesse exemplo

| A | B | C | D |
|---|---|---|---|
| 3620 | 1910 | 4900 | 2800 |
| 4010 | 2100 | 4100 | 2400 |
| 4560 | 2210 | 5520 | 2900 |
| 2760 | 2030 | 6200 | 2950 |
| 3480 | 2090 | 5800 | 3400 |
| 2930 | 2230 | 7200 | 3200 |
| 3070 | 2060 | 4700 | 3150 |
| 3200 | 2050 | 5600 | 2770 |

Usando o software Action Stat, geramos o seguinte gráfico:




Analisando o gráfico, podemos observar que:

- O bairro mais pobre, em média é o bairro B. Entretanto, é o que apresenta a menor variabilidade entre os salários.
- O segundo bairro mais pobre, em média é o bairro D, seguido do bairro A. A variação entre os salários desses dois bairros é parecida.
- O bairro mais rico é o bairro C, porém é o que apresenta maior variabilidade entre os salários das famílias.

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

## 3.3. Gráfico de Linhas

Gráficos de linhas ou pontos são normalmente usados para controlar alterações ao longo do tempo e para facilitar a identificação de tendências ou de anomalias.

### Exemplo 3.3.1:

O número de manutenções por dia em equipamentos foi acompanhado durante um período de 5 semanas, em uma empresa automobilística. Na tabela a seguir temos o número de manutenções por dia.

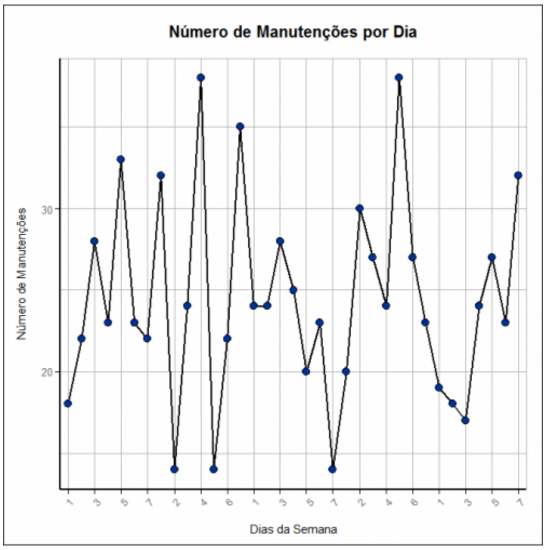
| Nº de manutenções | Dia | Semana |
|---|---|---|
| 18 | 1 | Semana 1 |
| 22 | 2 | Semana 1 |
| 28 | 3 | Semana 1 |
| 23 | 4 | Semana 1 |
| 33 | 5 | Semana 1 |
| 23 | 6 | Semana 1 |
| 22 | 7 | Semana 1 |
| 32 | 1 | Semana 2 |
| 14 | 2 | Semana 2 |
| 24 | 3 | Semana 2 |
| 38 | 4 | Semana 2 |
| 14 | 5 | Semana 2 |
| 22 | 6 | Semana 2 |
| 35 | 7 | Semana 2 |
| 24 | 1 | Semana 3 |
| 24 | 2 | Semana 3 |
| 28 | 3 | Semana 3 |
| 25 | 4 | Semana 3 |
| 20 | 5 | Semana 3 |
| 23 | 6 | Semana 3 |
| 14 | 7 | Semana 3 |
| 20 | 1 | Semana 4 |
| 30 | 2 | Semana 4 |
| 27 | 3 | Semana 4 |
| 24 | 4 | Semana 4 |
| 38 | 5 | Semana 4 |
| 27 | 6 | Semana 4 |
| 23 | 7 | Semana 4 |
| 19 | 1 | Semana 5 |
| 18 | 2 | Semana 5 |
| 17 | 3 | Semana 5 |
| 24 | 4 | Semana 5 |
| 27 | 5 | Semana 5 |
| 23 | 6 | Semana 5 |
| 32 | 7 | Semana 5 |

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Podemos construir o gráfico de duas formas: com ou sem meta.

### Sem meta

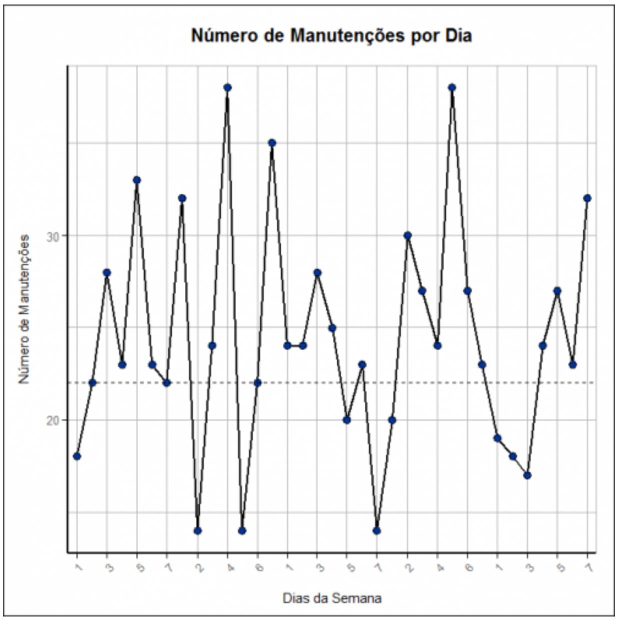
Neste caso, montamos um gráfico em duas dimensões onde, colocamos um ponto no par ordenado correspondente ao dia da semana e o valor observado, em seguida, unimos cada ponto ao seu sucessor por retas. O gráfico é dado a seguir:




## Com meta

Na análise com meta, basta determinarmos um valor (meta) para a comparação dos dados e, então inserimos uma reta horizontal na altura deste valor. Com isso, todos os pontos que estiverem acima do valor dado, atingiram a meta, enquanto os que estiverem abaixo não atingiram a meta.

Utilizaremos os mesmos dados acima, estipulando uma meta de 22. O gráfico é dado a seguir:




Os gráficos acima foram obtidos utilizando o software Action.

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

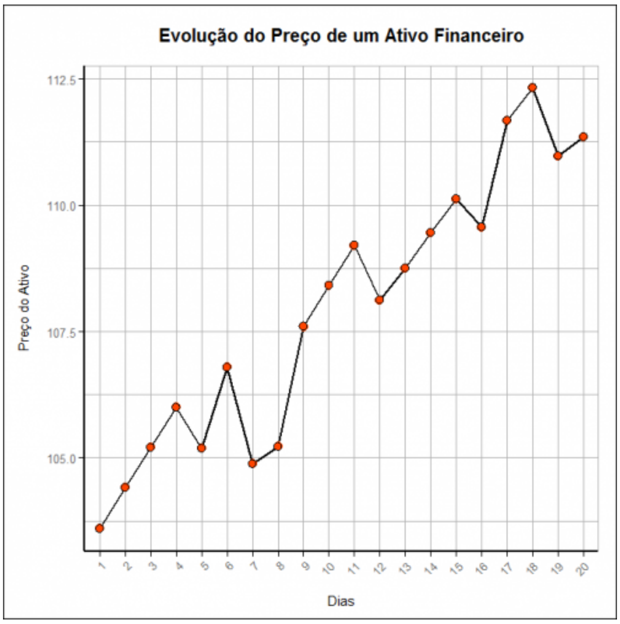
## Exemplo 3.3.2:

Na tabela abaixo, temos os dados referentes a evolução do preço de um ativo financeiro ao longo de 20 dias.

| Evolução do preço de um ativo financeiro ao longo de 20 dias | | | |
|---|---|---|---|
| Dia 1 | 103,60 | Dia 11 | 109,21 |
| Dia 2 | 104,41 | Dia 12 | 108,13 |
| Dia 3 | 105,21 | Dia 13 | 108,76 |
| Dia 4 | 106,00 | Dia 14 | 109,45 |
| Dia 5 | 105,18 | Dia 15 | 110,12 |
| Dia 6 | 106,80 | Dia 16 | 109,56 |
| Dia 7 | 104,89 | Dia 17 | 111,67 |
| Dia 8 | 105,23 | Dia 18 | 112,32 |
| Dia 9 | 107,60 | Dia 19 | 110,97 |
| Dia 10 | 108,41 | Dia 20 | 111,34 |

clique aqui para efetuar o download dos dados utilizados nesse exemplo

A seguir, temos a construção do gráfico de linhas e pontos para estes dados.



## 4 - EXERCÍCIOS

**4.1 -** As taxas de juros recebidas por 10 ações durante um certo período foram (medidas em porcentagens) 2,59; 2,64; 2,60; 2,62; 2,57; 2,61; 2,50; 2,63; 2,64. Calcule a média, a mediana, o desvio padrão, a variância, valor mínimo, valor máximo, os quartis $Q_1$ e $Q_3$, assimetria e curtose.

clique aqui para efetuar o download dos dados

**4.2 -** Para facilitar um projeto de ampliação da rede de esgoto de uma certa região de uma cidade, as autoridades tomaram uma amostra de tamanho 50 dos 270 quarteirões que compõem a região e foram encontrados os seguintes números de casas por quarteirão:

| Casas | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 14 | 18 | 22 | 26 | 32 | 45 | 59 | 66 | 80 |
| 2 | 15 | 18 | 23 | 27 | 36 | 46 | 61 | 66 | 89 |
| 3 | 15 | 20 | 24 | 29 | 42 | 48 | 61 | 68 | 90 |
| 10 | 16 | 21 | 25 | 29 | 44 | 52 | 61 | 75 | 92 |
| 13 | 16 | 22 | 25 | 30 | 45 | 58 | 65 | 78 | 97 |

**a)** Construa uma tabela de frequências;
**b)** Determine as medidas de posição e as medidas de dispersão;

clique aqui para efetuar o download dos dados

**4.3 -** Numa pesquisa realizada com 100 famílias, levantaram-se as seguintes informações:

| Número de filhos | 0 | 1 | 2 | 3 | 4 | 5 | Mais de 5 |
|---|---|---|---|---|---|---|---|
| **Famílias** | 17 | 20 | 28 | 19 | 7 | 4 | 5 |

**a)** Qual a mediana do número de filhos?
**b)** E a moda?
**c)** Que problemas enfrentaríamos no cálculo da média de filhos?

**4.4** - Considere as informações contidas na tabela abaixo acerca dos salários de 20 funcionários de um hospital. Faça uma distribuição de frequências e o gráfico histograma correspondente.

| Salários (x salários mínimos) | | | |
|---|---|---|---|
| 5,25 | 7,39 | 8,74 | 9,80 |
| 5,73 | 7,44 | 8,95 | 10,53 |
| 6,26 | 7,59 | 9,13 | 10,76 |
| 6,66 | 8,12 | 9,35 | 11,06 |
| 6,86 | 8,46 | 9,77 | 11,59 |

clique aqui para efetuar o download dos dados

**4.5 -** Uma pesquisa de opinião perguntou a 124 pessoas qual o antitérmico preferido entre as marcas A, B, C, D e E. Os resultados estão representados na tabela abaixo

| Marcas | Nº pessoas |
|---|---|
| A | 45 |
| B | 32 |
| C | 23 |
| D | 15 |
| E | 9 |

clique aqui para efetuar o download dos dados

Construa o gráfico de barras correspondente.

**4.6 -** Em um levantamento realizado em um hospital, 95 pessoas responderam a seguinte questão: Como você classifica o atendimento recebido? As possíveis respostas estavam classificadas em Péssimo, Ruim, Razoável, Bom ou Excelente. O número de pessoas em cada uma dessas classes pode ser encontrado na tabela abaixo.

| Qualidade | Nº pessoas |
|---|---|
| Péssimo | 10 |
| Ruim | 23 |
| Razoável | 38 |
| Bom | 20 |
| Excelente | 4 |

clique aqui para efetuar o download dos dados

Construa o diagrama de Pareto correspondente.

**4.7 -** Foram contabilizados o número de pessoas que foram atendidas em um dia em diversas áreas de um hospital. Os resultados obtidos podem ser encontrados na tabela abaixo.

| Áreas | Atendimentos |
|---|---|
| Pronto-Socorro | 364 |
| Pediatria | 286 |
| Psicologia | 127 |
| Neurologia | 86 |
| Ginecologia | 176 |
| **Total** | **1039** |

clique aqui para efetuar o download dos dados

Construa um gráfico de barras correspondente.

**4.8 -** Um hospital tem o interesse em determinar a altura média dos pacientes de uma determinada área e relacioná-la com a incidência de determinada anomalia ortopédica. Foram selecionados 80 pacientes e as alturas (em m) podem ser encontradas na tabela abaixo.

| Altura dos pacientes | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1,72 | 1,78 | 1,87 | 1,86 | 1,79 | 1,79 | 1,83 | 1,74 | 1,64 | 1,62 |
| 1,75 | 1,65 | 1,75 | 1,58 | 1,63 | 1,77 | 1,64 | 1,68 | 1,66 | 1,82 |
| 1,68 | 1,80 | 1,74 | 1,76 | 1,74 | 1,72 | 1,75 | 1,89 | 1,73 | 1,76 |
| 1,72 | 1,71 | 1,63 | 1,81 | 1,65 | 1,58 | 1,63 | 1,70 | 1,73 | 1,57 |
| 1,75 | 1,64 | 1,73 | 1,70 | 1,75 | 1,56 | 1,70 | 1,68 | 1,68 | 1,79 |
| 1,75 | 1,71 | 1,62 | 1,83 | 1,72 | 1,76 | 1,67 | 1,82 | 1,67 | 1,60 |
| 1,67 | 1,61 | 1,61 | 1,67 | 1,75 | 1,80 | 1,70 | 1,77 | 1,73 | 1,77 |
| 1,64 | 1,66 | 1,74 | 1,66 | 1,66 | 1,79 | 1,68 | 1,79 | 1,69 | 1,80 |

clique aqui para efetuar o download dos dados

Construa a tabela de distribuição de frequências por intervalos de classes e o histograma correspondente.

**4.9 -** Em um hospital foram coletados os salários (em salários mínimos) de 36 funcionários. Os resultados estão dispostos na tabela abaixo. Construa a distribuição de frequências em intervalos de amplitude 2 e o histograma correspondente.

| Salários de 36 funcionários (x salário mínimo) | | | | | |
|---|---|---|---|---|---|
| 4,00 | 6,86 | 8,74 | 10,53 | 13,23 | 16,22 |
| 4,56 | 7,39 | 8,95 | 10,76 | 13,60 | 16,61 |
| 5,25 | 7,59 | 9,13 | 11,06 | 13,85 | 17,26 |
| 5,73 | 7,44 | 9,35 | 11,59 | 14,69 | 18,75 |
| 6,26 | 8,12 | 9,77 | 12,00 | 14,71 | 19,40 |
| 6,66 | 8,46 | 9,80 | 12,79 | 15,99 | 24,00 |

clique aqui para efetuar o download dos dados

# REFERÊNCIAS BIBLIOGRÁFICAS

[1] Douglas C. Montgomery - Introduction to Statistical Quality Control, John Wiley and Sons, 1985.

[2] Fundamentos do Controle Estatístico do Processo - Manual de Referência, IQA.

[3] Schmidt, S. R. and Launsby, R. G. - Understanding Industrial Designed Experiments, Air Academic Press, Colorado Springs, CO, (1997).

[4] Forrest W. Breyfogle (1999) - Implementing Six Sigma: Smarter Solution Using Statistical Methods, John Wiley & Sons, INC.

[5] M. N. Magalhães e A. C. P. De Lima (2001) Noções de Probabilidade e Estatística, 3 edição, Editora USP.

[6] P. L. Meyer (1983) - Probabilidade: Aplicações à Estatística, segunda edição, Livros técnicos e Científicos Editora.

[7] W. O. Bussab e P. A. Morettin (1987) - Estatística Básica - 4 Edição, Atual Editora.

[8] Breyfogle, Forrest W. - Implementing Six Sigma: Smarter Solution Using Statistical Methods, John Wiley & Sons, INC, (1999).

[9] Fundamentos do Controle Estatístico do Processo - Manual de Referência, IQA.

[10] R. L. Mason, R. F. Gunst, J. L. Hess (1989) Statistical Design and Analysis of Experiments: with applications to Engineering and Science, John Wiley & Sons.

[11] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, Guide to the Expression of Uncertainty in Measurement, 1995.

[12] NBR ISO 10012-1, Requisitos de Garantia da Qualidade para Equipamentos de Medição, 1993.

[13] MSA, Análise dos Sistemas de medição, terceira edição.

[14] Versão Brasileira do documento de referência EA-4/02: Expressão da incerteza de medição na calibração, INMETRO.

[15] Montgomery, Douglas C., Design and Analysis of Experiments - Fourth Edition - Ed. John Wiley & Sons, 1997.

# 1. Análise dos Sistemas de Medição

Dados de medição têm sido utilizados nas mais diferentes maneiras. A decisão de ajustar ou não um processo de fabricação é baseado em dados de medição. Dados de medição ou dados derivados destes são comparados aos limites de Controle Estatístico do Processo. Caso esta comparação indicar que o processo está fora do controle estatístico, algum tipo de ajuste deverá ser feito. Caso contrário, o processo poderá prosseguir sem ajustes.

Outra importante utilização de dados de medição é em planejamento de experimentos. Um planejamento de experimentos permite conhecer o efeito de diferentes fatores que podem variar dentro de um processo, como por exemplo: matéria prima, condições de operação, tipos de ajustes de máquinas, entre outros. Neste caso, a análise do efeito destes fatores depende de dados de medições de uma peça, por exemplo. Em geral, estudos que exploram esta relação são denominados pelo Dr. W. E. Deming de estudos analíticos. O estudo analítico é aquele que aumenta o conhecimento sobre o sistema de causas que afetam o processo. Estes estudos estão entre as mais importantes aplicações de dados de medição, visto que recentemente eles têm conduzido ao melhor entendimento de produtos e processos.

Os benefícios obtidos com a utilização de procedimentos baseados em dados são diretamente determinados pela qualidade dos dados de medição utilizados. Se a qualidade for baixa, o benefício do procedimento provavelmente será baixo. De maneira similar, se a qualidade for alta, o benefício deverá ser alto também. Desta forma, devem ser tomados os devidos cuidados na obtenção dos dados, pois o benefício decorrente da utilização destes dados deve superar os custos de sua obtenção.

A qualidade dos dados de medição está relacionada com as propriedades estatísticas de medições múltiplas obtidas pelo sistema de medição. Suponhamos que um sistema de medição que está operando sob condições estabilizadas seja utilizado para coleta de várias medições de uma certa característica de uma peça. Se as medidas obtidas estiverem todas “próximas” do valor de referência, então a qualidade dos dados será dita “alta”. Da mesma forma, se algumas ou todas as medidas estiverem “afastadas” do valor de referência, então a qualidade dos dados será considerada baixa.

As propriedades estatísticas mais utilizadas para caracterizar a qualidade dos dados são tendências e variâncias. A propriedade chamada tendência refere-se à localização dos dados com relação ao valor de referência, a propriedade chamada de variância refere-se à dispersão dos dados. Porém, outras propriedades estatísticas, como a taxa de classificação incorretas, poderão ser apropriadas em alguns casos, como os sistemas de medição por atributo.

Uma das razões mais comuns que gera dados de baixa qualidade é a variação muito grande dos dados. Grande parte da variação em um conjunto de medições é devido à interação entre o sistema de medição e o seu meio. Se esta interação gerar variação muita alta, a qualidade dos dados poderá ser tão baixa que eles não terão utilidade.

Boa parte do trabalho de gestão de sistemas de medição visa monitorar e controlar variações. Dentre outras coisas, isto significa que devemos conhecer a forma como o sistema de medição interage com seu ambiente, de modo que sejam gerados somente dados de qualidade aceitável.

A seguir encontramos os tópicos sobre Análise dos sistemas de Medição:

# 1.1 - Sistema de Medição

O principal ponto para análise consiste em interpretarmos o sistema de medição como um processo. Desta forma, é importante ressaltarmos que não estamos avaliando simplesmente os equipamentos, mas o processo no qual utilizamos os equipamentos, o método e as pessoas para obtermos o resultado da medição.

*Sistema de Medição*: É a coleção de instrumentos ou dispositivos de medição, padrões, operações, métodos, dispositivos de fixação, software, pessoal, ambiente e premissas utilizadas para quantificar a unidade de medição ou corrigir a avaliação de uma característica sendo medida; o processo completo para obter medições

Medição é o conjunto de operações com objetivo de determinar o valor de uma grandeza. Estas operações podem ser realizadas automaticamente. (ISO GUM, 2008).

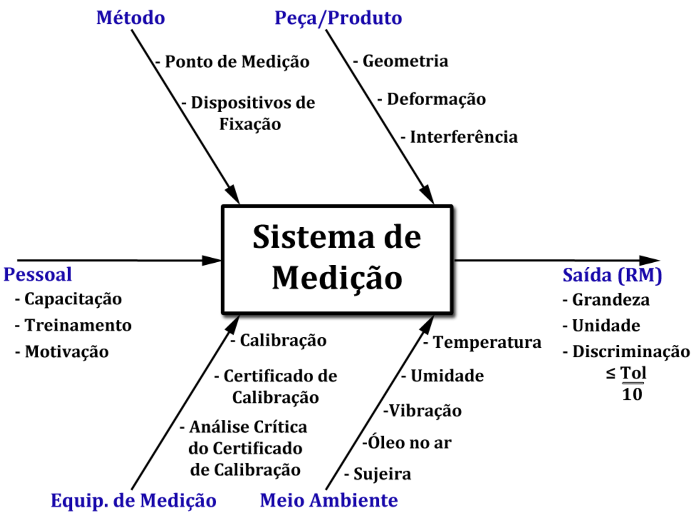



**Figura 1.1.1:** Sistema de medição.

O objetivo de uma medição é determinar o valor de uma grandeza a ser medida. Esta medição começa com uma apropriada especificação da grandeza, do método e procedimento de medição.

**Exemplo 1.1.1:**

Considere um sistema de medição para medir o diâmetro de um conector de torneira com tolerância de +/- 0,5 mm.

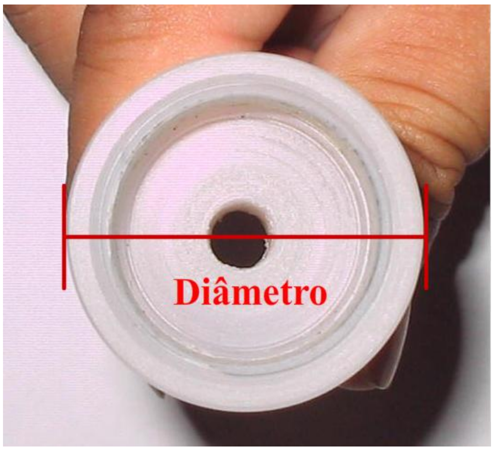



**Figura 1.1.2:** Medição de um conector de torneiras

Antes de qualquer análise estatística devemos obter uma boa definição do sistema de medição. Abaixo, apresentamos de forma simplificada o sistema de medição para medir o diâmetro do conector.

Definição do sistema de medição:

Equipamento de medição: paquímetro digital de resolução 0,01mm;

Observe que o equipamento de medição (paquímetro) apresenta uma resolução adequada para a característica que vamos medir, pois temos uma tolerância de +/- 0,5 mm, o que corresponde a uma faixa de 1 mm. Ao dividirmos a tolerância por 10, obtemos que a exatidão mínima requerida é de 0,1 mm. Como o paquímero digital tem resolução de 0,01 mm, concluímos que este é adequado para realizar tal medição.

Método de medição:

Posicionar o paquímetro no centro do conector;

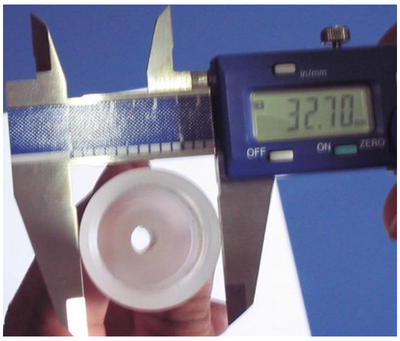



Figura 1.1.3: método de medição

Executar a medida

**Erro de Medição**

Toda medição tem imperfeições que dão origem a erros no resultado da medição. Tradicionalmente, um erro é visto como tendo dois componentes, a saber, um componente aleatório e um componente sistemático.

Um sistema de medição ideal produziria somente medições “corretas” a cada vez que fosse utilizado. No entanto, sistemas de medição com tal propriedade não existem. O erro de medição é definido por:

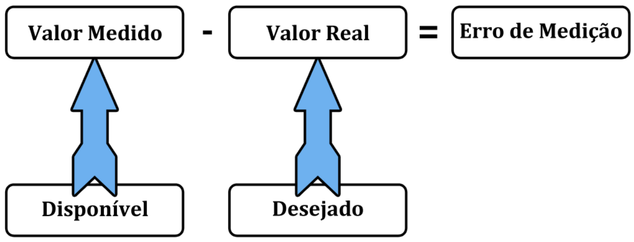



**Figura 1.1.4:** Erro de medição

Erro é um conceito idealizado e os erros não podem ser conhecidos exatamente. Na prática, associamos uma variável aleatória (por exemplo, a distribuição normal) para representar o erro de medição.

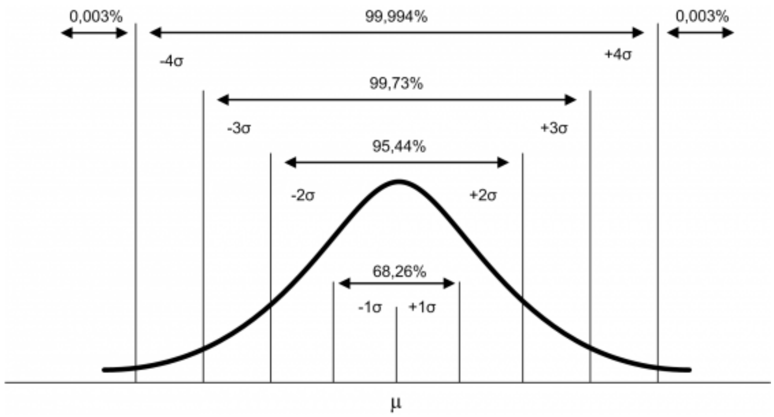



**Figura 1.1.5:** Erro de medição

Incerteza: **σ** ;

RR: **σ** .

Não confundir **σ** com erro!

Em geral, existe uma certa confusão entre o significado de RR e a incerteza de medição. A incerteza de medição é um termo utilizado internacionalmente para descrever a qualidade de uma medida. Até alguns anos atrás, este termo era frequentemente associado com a qualidade de equipamentos de medição. Entretanto, devido a importância de alguns ensaios, o conceito de incerteza de medição foi estendido para estabelecer a qualidade dos ensaios.

A incerteza de medição é um valor associado ao resultado da medição que descreve uma faixa no qual esperamos conter o valor verdadeiro da medida, com um determinado nível de confiança.

No processo de calibração, no qual avaliamos um equipamento de medição, a incerteza de medição corresponde ao desvio padrão (ou, múltiplo dele) associado às medições do equipamento de medição obtidas sob condições ideais de medição (calibração). Na calibração, o equipamento é comparado com respeito a um padrão de referência em um laboratório com condições ambientais controladas. Além disso, utilizamos um técnico devidamente capacitado para realizar tal comparação.

Em um ensaio, devemos considerar o sistema de medição como um processo, no qual temos os equipamentos, métodos, software, pessoal, tudo envolvido na medição. O RR tem como objetivo quantificar a variabilidade associada às medições do sistema de medição (equipamentos, método e pessoal) obtidas sob condições reais de utilização do sistema de medição.

Assim, podemos interpretar a incerteza em uma calibração como a incerteza devido ao equipamento que utilizamos, enquanto que o RR caracteriza as variações devido às interações do sistema de medição com o meio em que está inserido (principalmente, o método, pessoas, meio ambiente, produto, entre outros). De forma geral, a incerteza de medição determina um intervalo de confiança associado ao resultado da medição, enquanto que o RR se preocupa em avaliar as fontes de variação e determinar adequabilidade do sistema de medição para controlar o processo ou avaliar um produto. Quando estudamos um ensaio, muitas vezes utilizamos técnicas como o RR para auxiliar na determinação da incerteza de medição.

# Tipos de Erros:

Dois tipos de erros serão característicos deste estudo:

**Erro Aleatório**

O erro aleatório é aquele que ocorre de forma inesperada e com intensidade que danifica nossas medições. Este erro representa as pequenas variações que ocorrem em medidas repetidas de uma grandeza. Estas variações têm como causa, alterações ambientais ou espaciais, variação devido ao equipamento de medição, interferência elétrica entre outras. Embora não seja possível compensar o erro aleatório, ele pode geralmente ser reduzido se aumentarmos o número de observações ou se melhorarmos a tecnologia do sistema de medição (melhor ambiente, novos equipamentos ou treinamento dos técnicos). Interpretamos o erro aleatório como uma variável aleatória com média zero.

**Erro Sistemático**

O erro sistemático é aquele que ocorre em todas as medições mais ou menos com a mesma intensidade. Assim como o erro aleatório, o erro sistemático não pode ser eliminado, porém ele, frequentemente, pode ser reduzido. Suponha que um erro sistemático se origina de um efeito reconhecido de uma grandeza de influência em um resultado de medição. Se este efeito pode ser quantificado e, se for significativo com relação à exatidão requerida da medição, uma correção ou fator de correção pode ser aplicado para compensar o efeito. Supomos que, após esta correção, a esperança ou valor esperado do erro sistemático seja zero.

Abaixo apresentamos o diagrama de Ishikawa (espinha de peixe) para descrever os principais componentes do erro de medição:

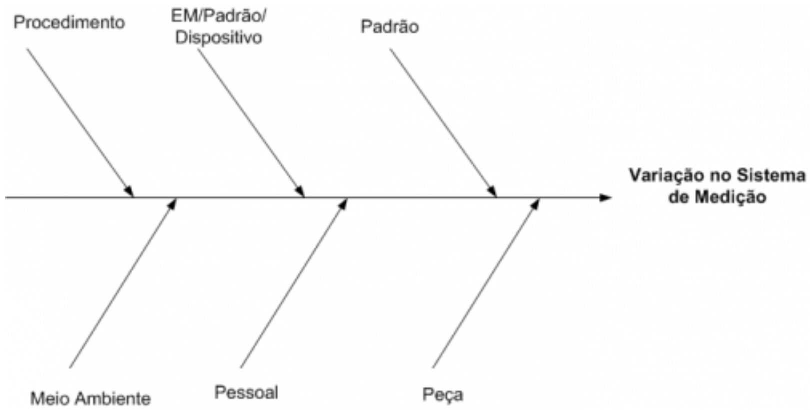



**Figura 1.1.6 :** Diagrama de Ishikawa

**Exemplo 1.1.2:**

Descrição dos principais componentes do erro de medição para o sistema de medição do diâmetro do conector de torneira.

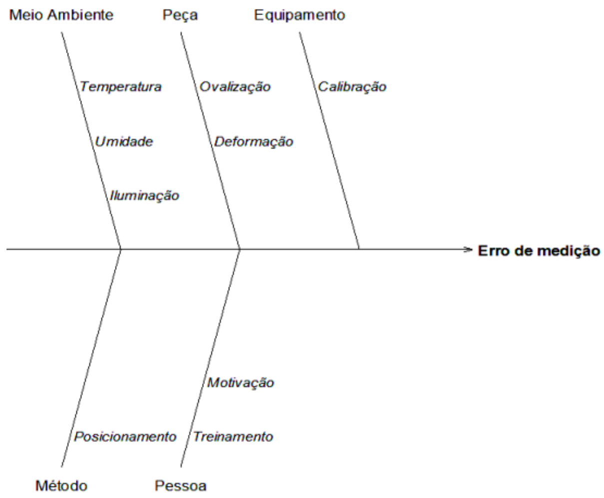



**Figura 1.1.7 :** Diagrama de Ishikawa para o conector de torneira

clique aqui para efetuar o download dos dados utilizados nesse exemplo

Para entender como executar essa função do **Software Action**, você pode consultar:

Para entender como executar essa função do **Software Action**, você pode consultar o **manual do usuário.**

# Requisitos de um sistema de medição

Um sistema de "má qualidade" poderá mascarar a variação real do processo ou produto conduzindo a conclusões erradas:

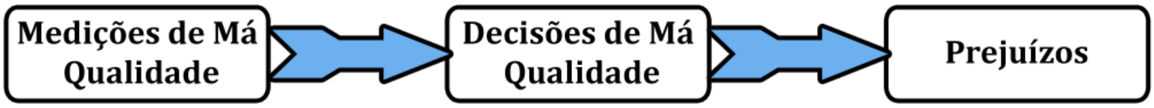



**Figura 1.1.8:** Sistema de "má qualidade"

Existem certas propriedades fundamentais que definem um bom sistema de medição:

- Uma adequada discriminação ou sensibilidade. O incremento de medida deve ser pequeno o suficiente para detectar variações no processo ou nos limites de especificação. A regra comum é conhecida como regra do dez, que consiste em definir a discriminação do sistema de medição dividindo a tolerância (ou variação do processo) em 10 partes.
- O sistema de medição deve estar sob controle estatístico. Isto significa que sob condições de repetitividade, as variações do sistema de medição são devidas à causas comuns e não à causas especiais.

- Para controle de produto, a variabilidade do sistema de medição deve ser pequena comparada com limites de especificação. Comparar a variabilidade do sistema de medição com as tolerâncias do produto.
- Para controle do processo, a variabilidade do sistema de medição deve demonstrar uma resolução efetiva e pequena comparada com a variação do processo de manufatura. Comparar a variabilidade do sistema de medição com 6-sigma da variação do processo e/ou variação total.

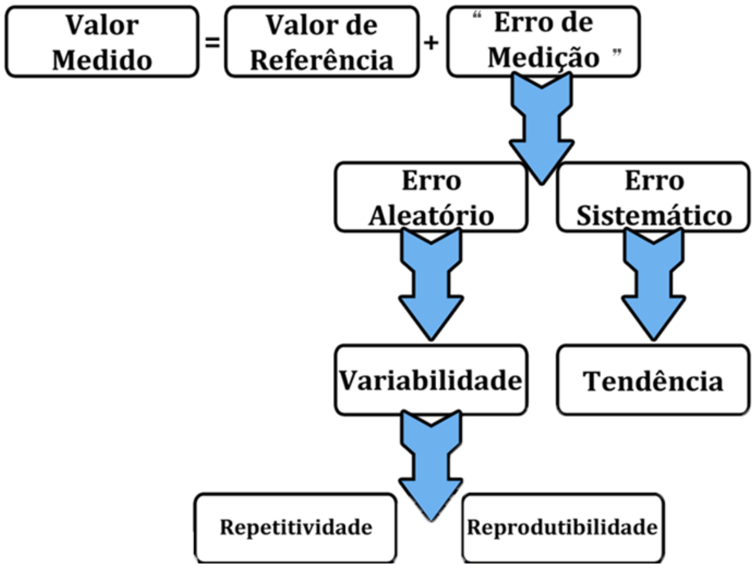


**Figura 1.1.9:** Avaliação do erro de medição.

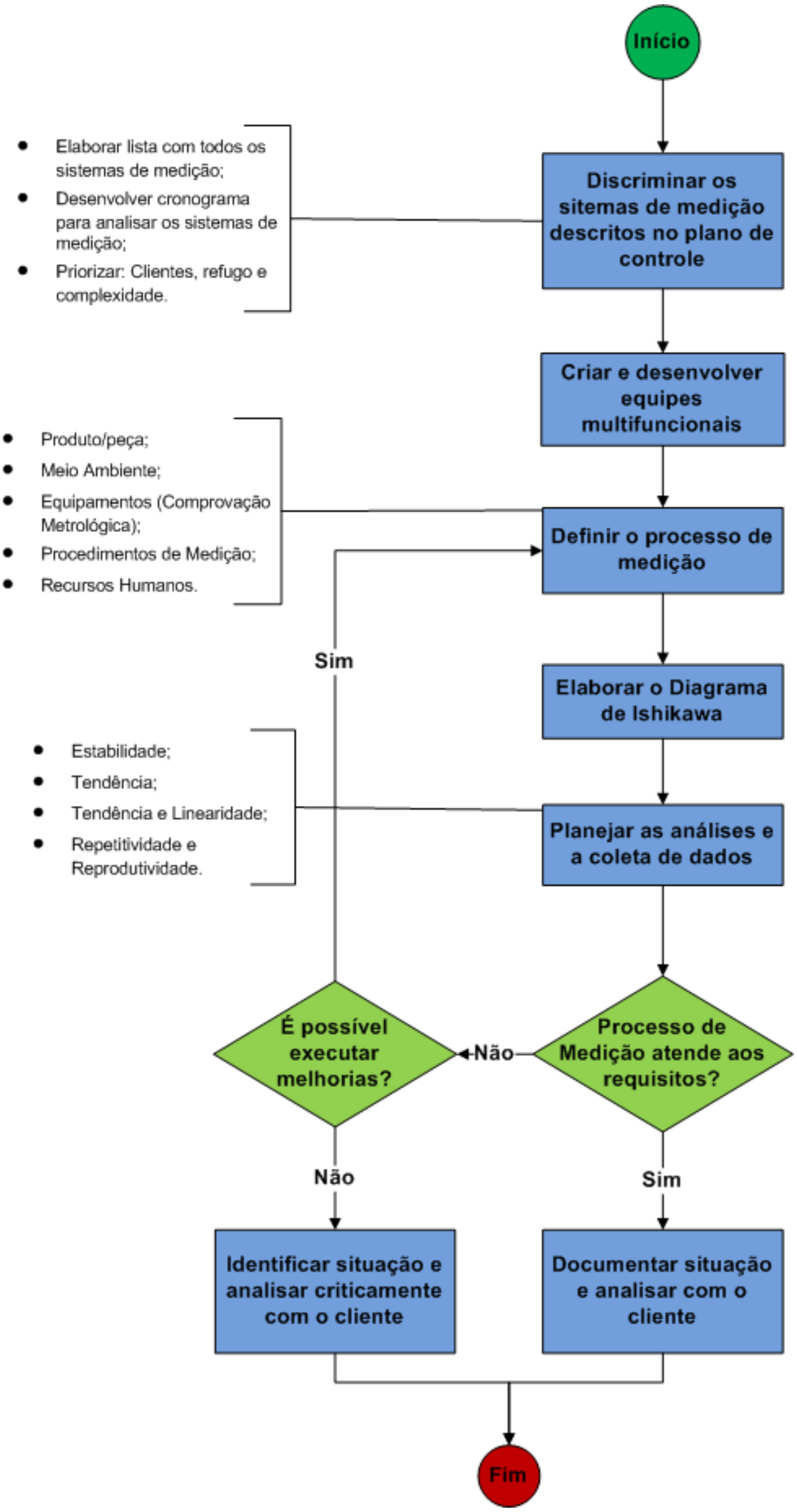
# 1.2 - Planejamento e Estratégia

Nem toda característica do processo ou produto requer uma análise detalhada como a que estamos desenvolvendo. Para sistemas de medição simples, como os sistemas determinados por paquímetros, micrômetros ou calibradores, muitas vezes não requerem uma análise detalhada. A regra básica para escolher o sistema a ser avaliado é se este é identificado no plano de controle ou é importante para determinar a rejeição ou não do processo ou produto. Outro indicativo é o nível de tolerância determinado para a dimensão específica e a criticidade perante ao cliente. Porém, ***o bom senso é o guia em qualquer caso.***

# Diretrizes para análise do sistema de medição

- Discriminar as grandezas relacionadas nos planos de controle;
- Identificar os sistemas de medição
- Definir as prioridades
    - Cliente
    - Refugo
    - Complexidade
- Identificar uma equipe multifuncional
- Para cada sistema de medição priorizado:
    - Desenvolver um fluxograma do processo de medição;
    - Treinar os envolvidos;

- Desenvolver o diagrama de Ishikawa;
- Escolher as ferramentas estatísticas;
- Montar um cronograma de aplicação das ferramentas;
- Documentar as soluções e as correções;
- Institucionalizar a mudança.



# Modelo de Erro de medição

| Fontes de Erro | | Componentes | Fator ou Parâmetro |
|---|---|---|---|
| P | Peças | Peça, Amostra, Mensurando, Unidade sobre Teste, Artefato, Padrão de Variação | Desconhecido |
| I | Instrumento | Equipamento de Medição, Unidade de Medição, Célula de Medição | Meios de Comparação |
| S | Padrão | Escala, Referência, Artefato, Padrão de Verificação, Padrão de Consenso, Material Padrão, Classe, Critério de Aceitação | Valor de Referência ou Critério de Aceitação |
| M | Método | Treinamento *On-the-job*, Instrução de Trabalho, Plano de Controle, Método, Plano de Inspeção, Programa de Teste | Como |
| O | Operador | Instrumentista, Técnico de Teste ou Calibração, Inspetor | Quem |
| E | Meio Ambiente | Temperatura, Umidade, Contaminação, *Housekeeping*, Iluminação, Posição, Vibração, Interferência Eletromagnética, Ruído, Tempo e Ar | Condições de Medição e Ruído |
| A | Concepção | Estatística, Operacional, Calibração, Constantes, Valor de Handbook, Estabilidade Térmica, Elasticidade | Medição Confiável |