

# Coleta de dados

[Início](#) / [Estatística Básica](#) / [Noções básicas de estatística](#) / 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 1.

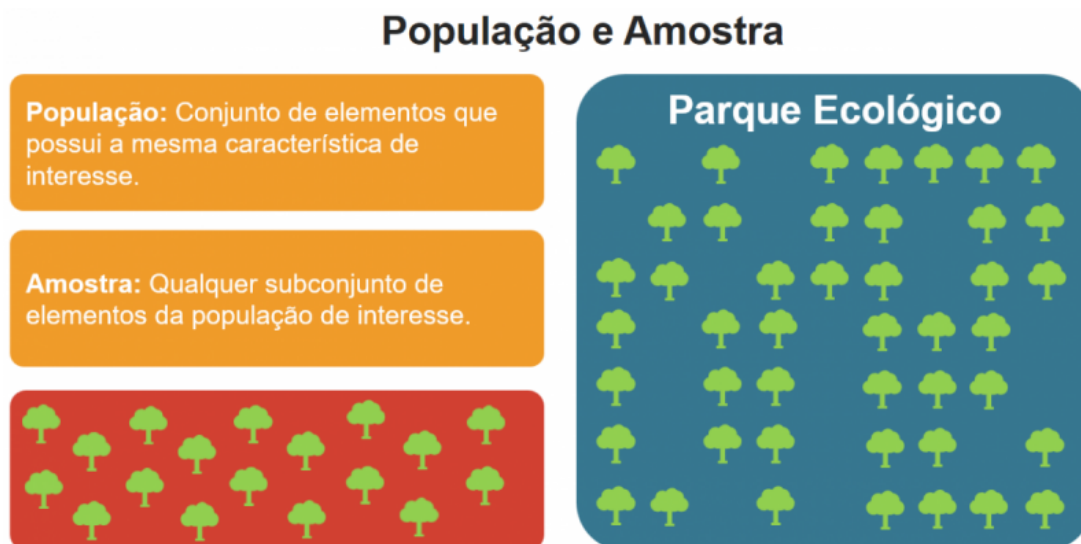


Figura 1. População e amostra.

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 2. 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).

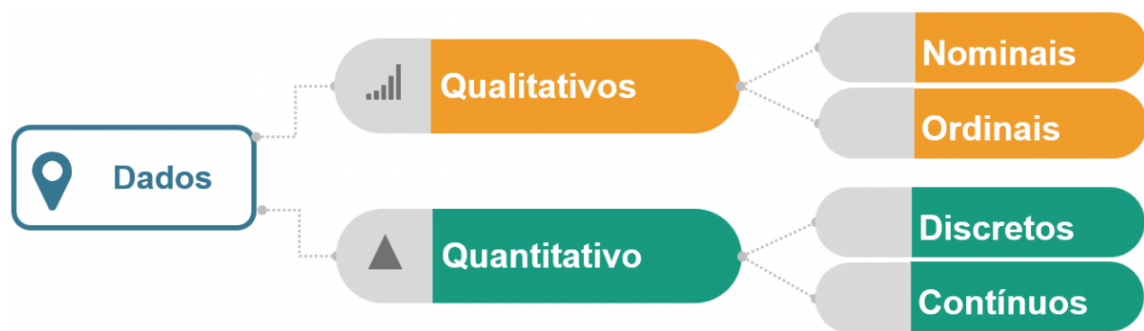


Figura 2. Escala de dados.

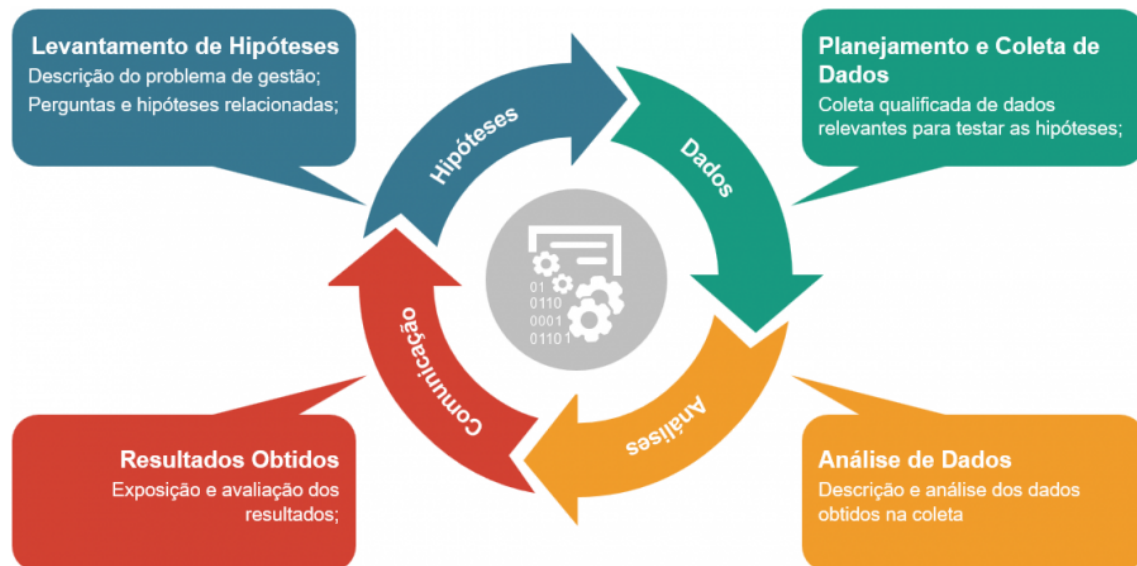
No caso de dados qualitativos nominais temos apenas a identificação ou rotulação das observações que constituem os dados. Neste caso, os dados são alocados em categorias e contados apenas com relação à frequência de ocorrência. Nenhuma ordenação está implícita. Uma variável como "sexo" pode ser categorizada 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.

## Planejando a coleta de dados

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

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



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.