

# Introdução à Estatística

No desenvolvimento científico e em nosso próprio dia-a-dia, estamos sempre fazendo observações de fenômenos, gerando *dados*:

- O meteorologista informa que a probabilidade de chover hoje é de 30%;
- O governo informa que a renda média de uma família de quatro pessoas aumentou 5% de um ano para cá;

Todos nós, ao lermos jornais e revistas ou acessando *sites* na internet, estamos vendo resultados estatísticos provenientes do censo, de pesquisas eleitorais, etc.

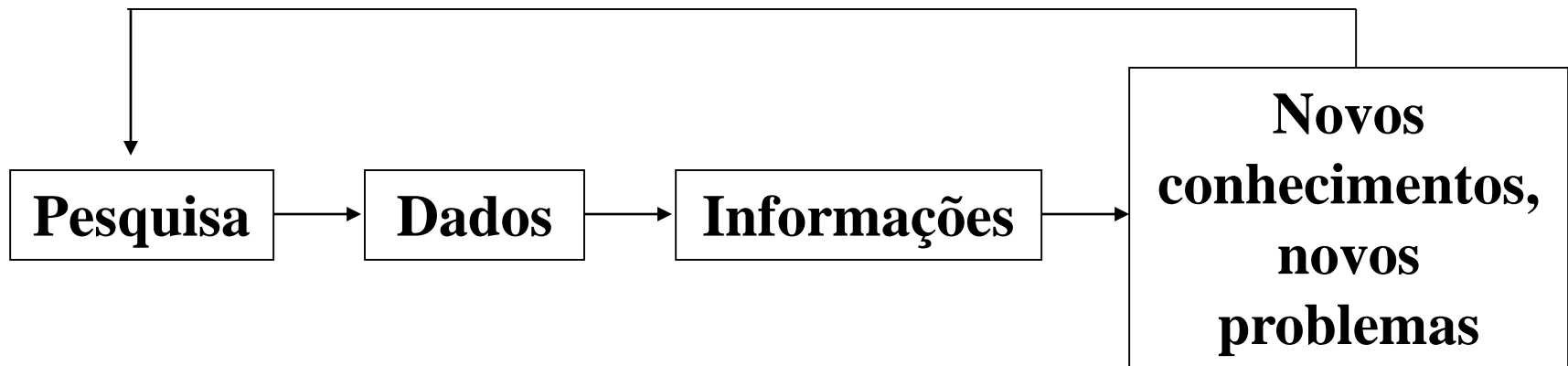
# Objetivo e método da Estatística

A *estatística* envolve técnicas para coletar, organizar, descrever, analisar e interpretar *dados* provenientes de experimentos ou vindos de outros estudos observacionais.

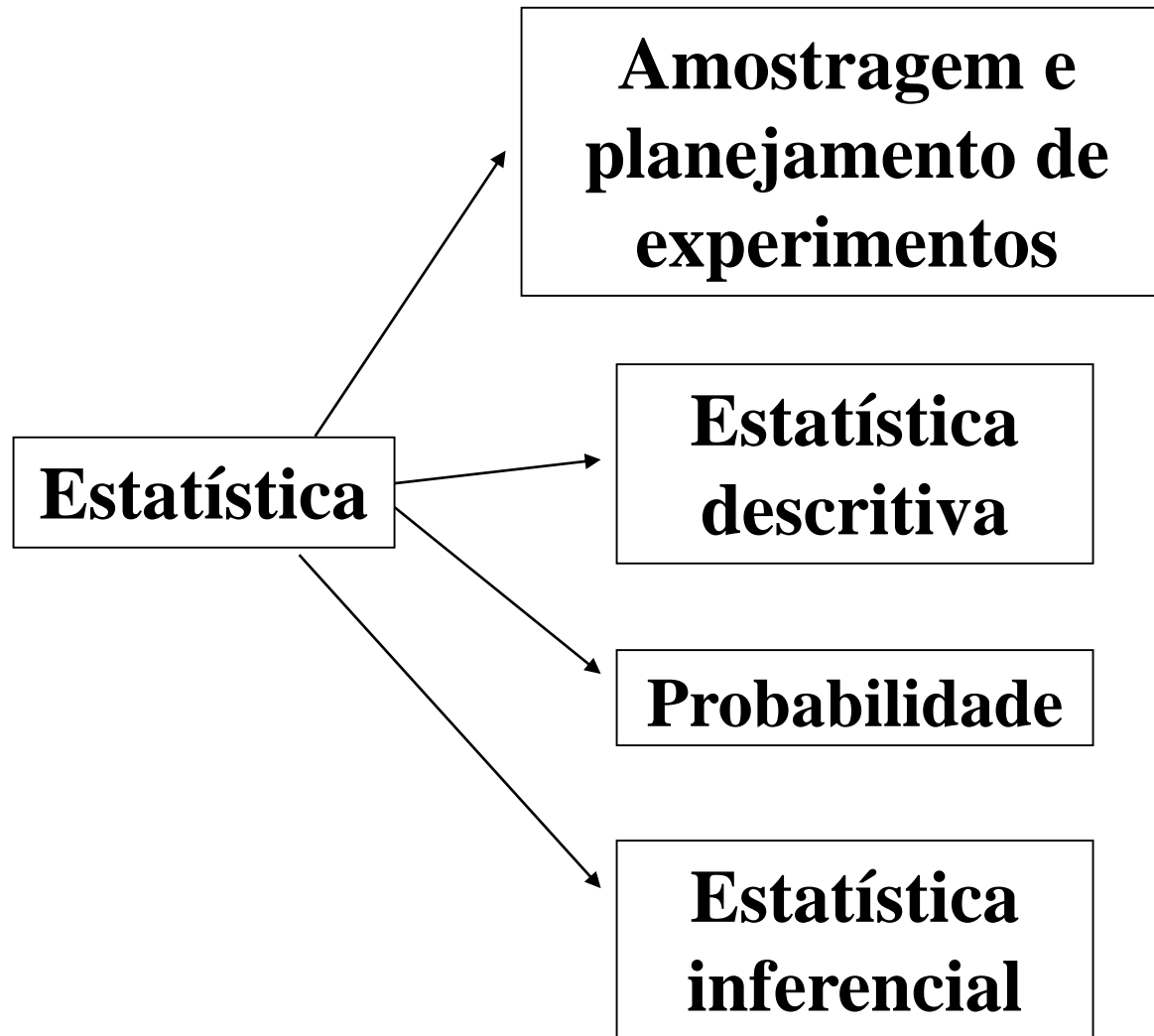
- *Estudo observacional*: É aquele em que as características de uma população são levantadas (observadas ou medidas), mas sem manipulação. É o caso de um censo demográfico, pesquisas eleitorais, pesquisas de mercado, inspeção da qualidade, etc. Em todos esses casos, quer-se ter a idéia de certa população tal qual ela é na natureza ou no processo.

• *Estudo experimental*: É aquele em que grupos de indivíduos (ou animais, ou objetos) são manipulados para se avaliar o efeito de diferentes tratamentos. É o caso de se verificar o rendimento de um processo químico para diferentes temperaturas de reação, que são manipuladas de acordo com o interesse prático.

A análise estatística de dados geralmente tem por objetivo tomadas de decisões, resoluções de problemas ou produção de conhecimento. Mas novos conhecimentos normalmente geram novos problemas de pesquisa, resultando em um processo iterativo:



# Grandes áreas da Estatística



## ✓ **Amostragem e planejamento de experimentos**

- É o processo de escolha da amostra.
- É a parte inicial de qualquer estudo estatístico. Consiste na escolha criteriosa dos elementos a serem submetidos ao estudo.
- Podemos citar alguns métodos de amostragem: aleatória simples, sistemática, estratificada, por conglomerado, entre outras.

## Situação – Pesquisas sobre tendências de votação:

“Em épocas de eleição, é comum a realização de pesquisas com o objetivo de se conhecerem as tendências do eleitorado. Para que os resultados sejam, de fato, representativos, toma-se o cuidado de entrevistar um conjunto de pessoas com características sócio-econômicas, culturais, religiosas, etc. tão próximas quanto possível da população à qual os resultados da pesquisa serão estendidos. A escolha da amostra, a redação do questionário, a entrevista, a codificação dos dados, a apuração dos resultados são etapas fundamentais deste tipo de pesquisa.”

## ✓ Estatística descritiva

- É a parte mais conhecida.
- Compreende a organização, o resumo e, em geral, a simplificação de informações que podem ser muito complexas.
- Quem vê o noticiário na televisão ou nos jornais ou *sites* de informação, sabe quão frequente é o uso de médias, índices e gráficos nas notícias.



## ✓ Probabilidade

- É a teoria matemática utilizada para se estudar a *incerteza* oriunda de fenômenos de caráter *aleatório*.
- Fornece a base para os métodos que utilizamos quando fazemos generalizações a partir de dados observados.

## ✓ Estatística inferencial

- A tomada de decisões sobre a população, com base em estudos feitos sobre os dados da amostra, constitui o problema central da *inferência estatística*.
- As decisões sobre a população estão sempre associadas a um grau de incerteza e, conseqüentemente, uma probabilidade de erro.
- A generalização da amostra para a população é feita por meio do modelo estatísticos que melhor se ajusta a situação em estudo.

## Situação – Testes sobre medicamentos

“Um experimento médico testa um novo analgésico para ver se é melhor do que o produto padrão correspondente. Dez pessoas selecionadas aleatoriamente tomam o novo medicamento, e as outras tomam o remédio padrão. O experimento é do tipo *duplo-cego*, isto é, nem o paciente nem o médico sabem qual dos dois remédios está sendo administrado. Esta informação é do conhecimento apenas do estatístico e de outros que vão analisar os dados. Tais processos garantem a validade dos resultados. Suponhamos que os resultados sejam os seguintes:

| Remédio | Quantos relataram diminuição da dor |
|---------|-------------------------------------|
| Novo    | 8                                   |
| Padrão  | 5                                   |

É lícito concluir que o novo remédio é melhor só porque mais pessoas no grupo das que o tomaram, relataram diminuição da dor? Trata-se de um problema da maior relevância. É preciso saber se a margem de 8 para 5 é real, ou se os dois remédios são igualmente eficientes, tendo a diferença decorrido apenas de uma variação aleatória.

# Conceitos básicos

- População: É o conjunto de *todos* os resultados, respostas, medidas ou contagens a serem estudados.
- Censo: É o conjunto de *dados* relativos a *todos* os elementos de uma população.
- Dados: Consistem em informações provenientes de observações, contagens, medidas ou respostas.
- Amostra: É um subconjunto de elementos de uma população.
- Amostragem: É o processo de seleção da amostra.

## Importante

A menos que uma população seja pequena, geralmente é difícil obter todos os seus dados. Na maior parte dos estudos, a informação deve ser obtida a partir de uma amostra.

- Parâmetro: É uma medida que descreve numericamente uma característica da *população*.
- Estatística: É uma medida que descreve numericamente uma característica da *amostra*.

# Exercício 1: Identificando a população e a amostra

Para cada situação abaixo identifique a população e a amostra.

A) Em um estudo recente, perguntou-se a 3002 adultos no Brasil se eles liam diariamente notícias na Internet. Um mil e seiscentos adultos responderam que sim.

Resposta: A população é composta por todos os adultos brasileiros e a amostra são os 3002 adultos questionados na pesquisa.

B) O Departamento de Energia Americano conduz levantamentos semanais em 800 postos de gasolina para determinar o preço médio por galão da gasolina comum. No último levantamento realizado o preço médio era de U\$2,97 por galão.

Resposta: A população é composta por todos os postos de gasolina registrados nos EUA e a amostra são os 800 postos utilizados na pesquisa.

## Exercício 2: Distinguindo entre parâmetro e estatística

Decida se, nas situações abaixo, o valor numérico descreve um parâmetro ou uma estatística:

A) Em uma pesquisa realizada em todos os postos de saúde de uma determinada capital brasileira observou-se que 24% deles possuíam computador.

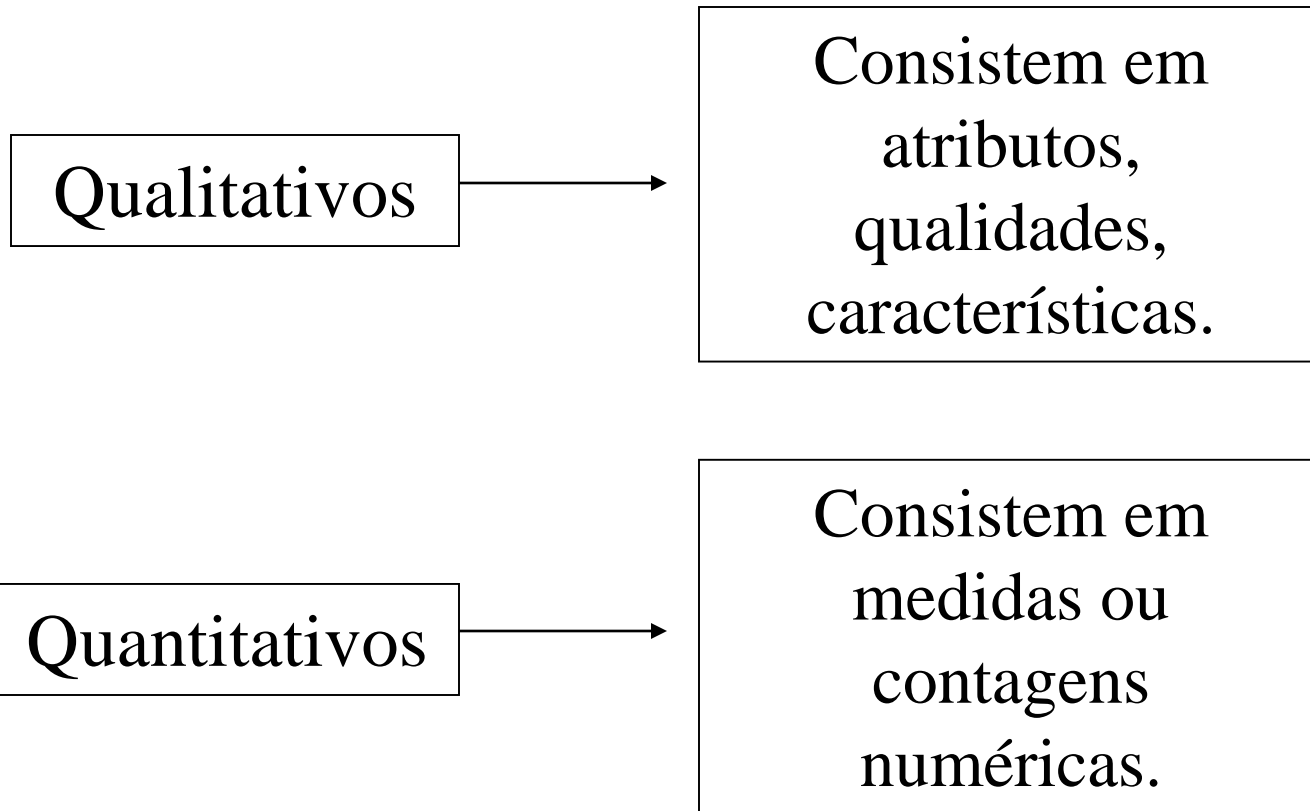
Resposta: Trata-se de um parâmetro já que foi obtido por uma pesquisa realizada com TODOS os postos de saúde de uma capital, ou seja, a partir de um censo.

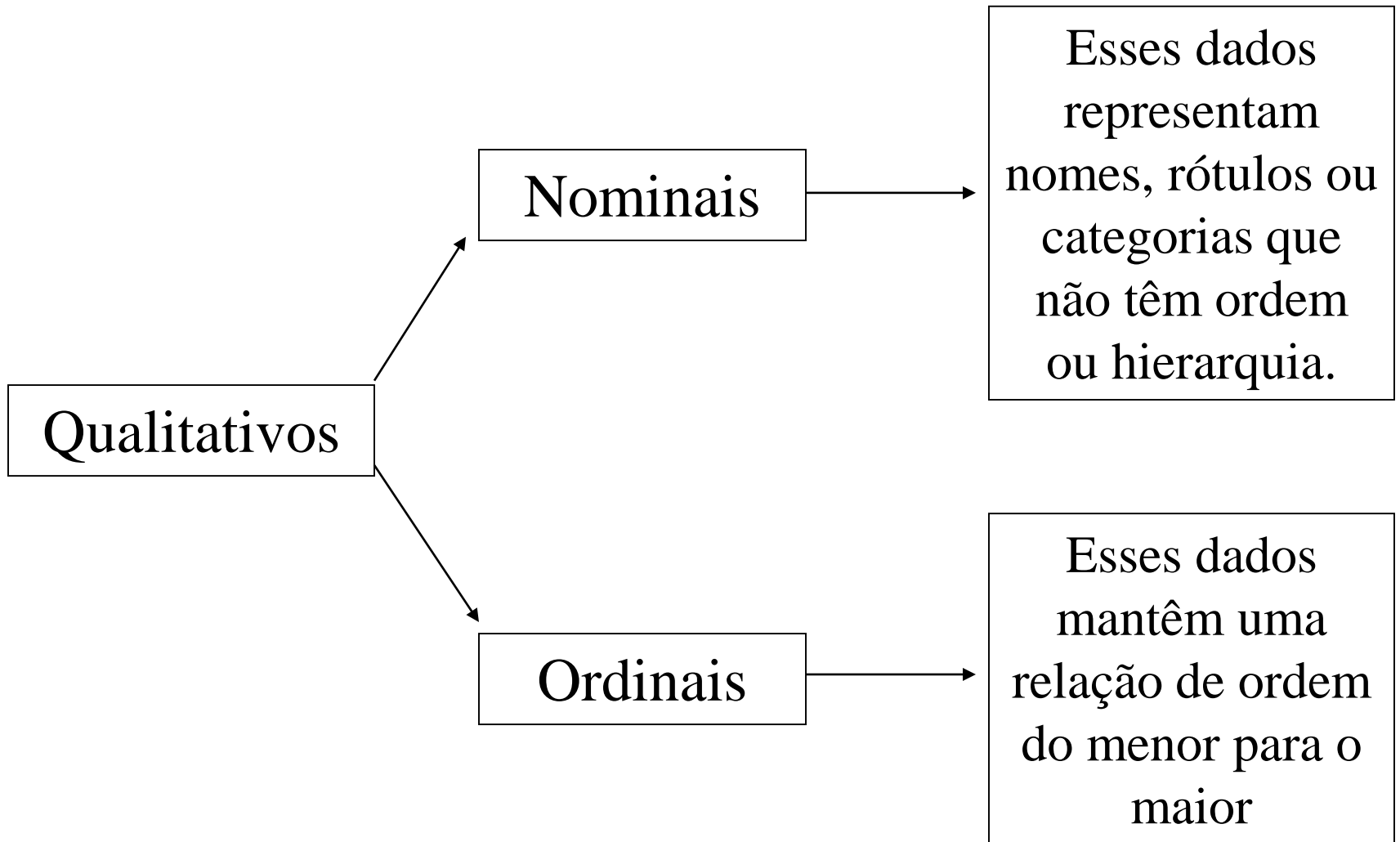
B) Em um levantamento feito com uma amostra de usuários de computador, 10% afirmaram que seus computadores apresentaram algum defeito, no último ano, que necessitou ser consertado por um serviço técnico.

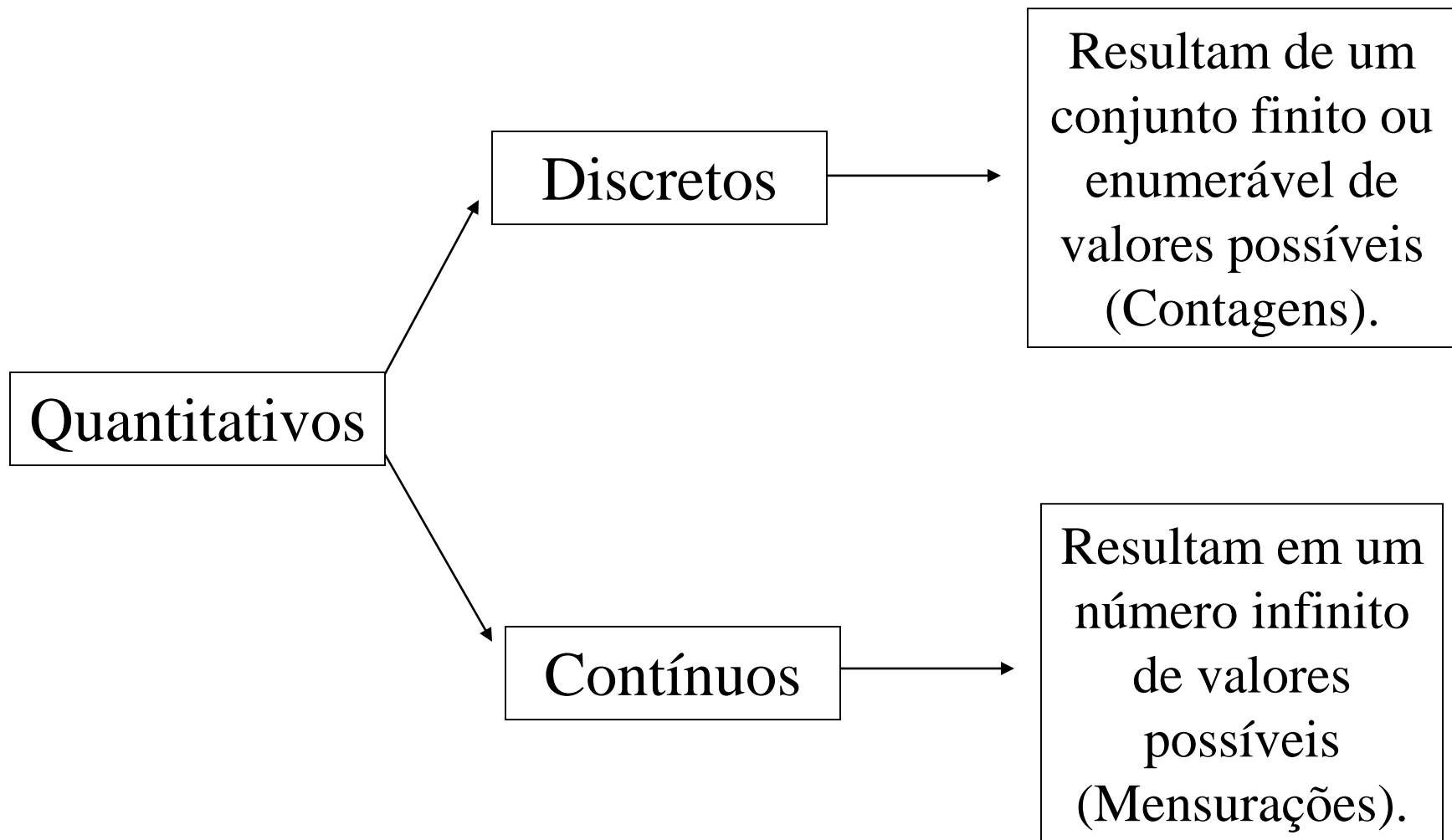
Resposta: Trata-se de uma estatística já que foi obtida por uma pesquisa realizada com uma AMOSTRA de usuários.



# Tipos de dados







## Exercício 3: Classificando os dados

Determine se os dados são nominais, ordinais, discretos ou contínuos:

- a) A temperatura (°C) diária em uma cidade. **Resposta: Quantitativa contínua.** A temperatura medida pode assumir valores positivos, negativos, inteiros ou decimais.
- b) Meio de transporte utilizado para ir ao trabalho. **Resposta: Qualitativa nominal.** Podemos pensar nas possíveis respostas para essa variável: ônibus, metrô, carro particular, à pé, etc. Dessa forma, temos categorias não ordenáveis.
- c) Grau de escolaridade de funcionários de uma fábrica. **Resposta: Qualitativa ordinal.** As possíveis respostas para essa variável são categorias ordenáveis: fundamental, médio, superior, pós-graduado, etc.

### Exercício 3: Classificando os dados - *Continuação*

d) Número de ligações recebidas por uma central telefônica em um minuto qualquer do dia. **Resposta:** Quantitativa discreta. As possíveis respostas para essa variável são contagens de valores inteiros.

e) Cor da pele de estudantes inscritos no ENEM. **Resposta:** Qualitativa nominal. As possíveis respostas para essa variável seriam: branco, pardo, preto, amarelo e indígena. Dessa forma, temos categorias não ordenáveis.

f) Peso (em kg) de indivíduos elegíveis a cirurgia bariátrica. **Resposta:** Quantitativa contínua. As possíveis respostas para essa variável são medidas de valores que podem ser inteiros ou decimais.