

## Fases do Método Estatístico

### Coleta de dados

São inúmeras as maneiras de coletar dados, e a amostragem é a maneira mais frequente. Embora a amostragem seja um conceito simples, muitas vezes inúmeras e complexas questões sobre a população precisam ser respondidas, fazendo com que o processo seja conseqüentemente de mesma complexidade.

Quando se fala em coleta de dados, estamos nos referindo à obtenção e reunião de registros sistemáticos de dados. E como um primeiro ponto, é necessário fazer uma distinção nos dados estatísticos quanto à sua **origem**, que podem ser:

- **Dados primários ou**
- **Dados secundários.**

Os **dados primários** são aqueles publicados ou informados por quem os recolheu, pessoa ou organização.

Os **dados secundários** são aqueles publicados ou fornecidos por terceiros.

Por exemplo: as tabelas de Censo Demográficos fornecidas pelo IBGE são dados primários. Quando um jornal faz publicações de dados estatísticos oriundos de várias fontes e fazendo relações com outros setores e áreas, esses dados são secundários.

É importante entender que, sempre que possível, deve-se buscar as fontes primárias por *possuírem informações mais detalhadas dos dados, por conterem as definições dos termos usados e das unidades contidas nos valores referenciados; por relatarem, em algumas situações, informações dos procedimentos adotados na pesquisa, o tipo da amostra e seu tamanho; além de diminuir os riscos de erros de digitação.*

Os dados também podem ser classificados de acordo com a maneira como eles podem ser encontrados e obtidos:

- **Enumerados;**
- **Mensurados;**
- **Avaliados.**

Os **dados enumerados** podem ser contados e ordenados, são contabilizados como, por exemplo, contagem física de pessoas e objetos, levantamento de estoque etc.

Os **dados mensurados** são obtidos por meio de medições, utilizando qualquer instrumento de medida. São dados relativos a altura, peso, volume etc.

Os **dados avaliados** NÃO podem ser medidos ou enumerados. São dados avaliados de forma empírica como, por exemplo, a estimativa do número de pessoas contidas em uma sala.

A coleta de dados pode ser **direta** ou **indireta**. Ela é considerada direta quando é feita na população por intermédio de levantamento de registros ou coleta de informações por meio de questionários etc.

A **coleta direta**, quanto ao fator tempo, pode ser classificada como:

- **Coleta contínua;**
- **Coleta periódica;**
- **Coleta ocasional.**

As coletas **contínuas** são feitas de **forma continuada e sem interrupções** como, por exemplo, a estatística de nascimentos em um determinado hospital.

As coletas **periódicas** são feitas *em intervalos constantes de tempo*, como por exemplo, levantamentos semestrais de renovações de matrículas dos alunos.

As coletas **ocasionais** são realizadas *para atender a necessidades momentâneas* como, por exemplo, levantamento estatístico de epidemias.

A **coleta indireta** é tomada de informações e elementos já conhecidos, *oriundos de uma coleta direta feita anteriormente*. Mas também pode ser feita por conhecimento de fenômenos anteriores relacionados com a situação em análise, por deduções ou suposições. Assim sendo, consideramos a **coleta indireta** de dados podendo ser realizada por:

- Coleta por analogia;
- Coleta por proporcionalização;
- Coleta por indícios;
- Coleta por avaliação.

A **coleta por analogia** é feita quando dois fenômenos possuem uma relação de casualidade e podemos tirar o conhecimento induzido a partir do outro fenômeno anterior. Essa relação é fácil de ver nos exemplos de estimativas de quantidades como, por exemplo: *no ano novo passado um milhão de pessoas estiveram nas praias de Copacabana e espera-se o mesmo número de pessoas este ano*.

A **coleta por proporcionalização** é feita utilizando uma regra de três simples, normalmente quando o conhecimento de um fenômeno se induz a partir do conhecimento de uma porção dele. Pode-se considerar como uma porcentagem do tipo: *estiveram presentes três mil pessoas no show onde havia seis mil lugares, agora temos quatro mil lugares e estima-se que duas mil pessoas estejam presentes*.

A coleta por indícios é feita quando são escolhidos fenômenos sintomáticos para analisar e discutir um aspecto geral da vida social. *Um exemplo típico é a reunião de elementos como prova de um ato ilícito para que se possa descobrir culpados de um crime ou orientar as investigações.*

A coleta por avaliação é feita quando se presume o estado quantitativo de um determinado fenômeno, baseado em informações confiáveis ou estimativas que foram cadastradas anteriormente com base em fatos ocorridos. Esta coleta é muito utilizada em exemplos do tipo: *em função das informações cadastradas, é possível supor que existam 90 pessoas nesta sala.*

## Crítica dos dados

Após a obtenção dos dados, é necessário que haja uma crítica cuidadosa e detalhada da maneira como eles foram obtidos, a fim de que erros críticos não influenciem nos resultados. Essa busca de possíveis falhas e imperfeições na obtenção dos dados deve ser acompanhada de uma análise com relação à adequação ao tipo de método empregado para coleta das informações.

As críticas podem ser dos tipos:

- Crítica interna ou
- Crítica externa.

A crítica interna é feita diretamente sobre os dados obtidos na coleta, verificando se as informações estão coerentes como, por exemplo, a soma de números etc.

A crítica externa é feita buscando a causa dos erros nos dados informados, muitas vezes causados por perguntas mal formuladas, distração por parte dos avaliados ou respostas com duplo sentido.

## Apuração e apresentação dos dados

Após a coleta e antes da análise dos dados, é necessário fazer algum tratamento prévio, a fim de entre outras coisas deixá-los mais fácil de serem visualizados. Consideraremos esta como sendo uma etapa de apuração com o objetivo de resumir os dados, ordenando e agrupando de forma a facilitar a sua visualização. Os dados que chegam de forma desordenada e muitas vezes, pela grande quantidade de informações, torna impossível a tarefa de análise e interpretação desses resultados.

Nesta fase de apuração é possível condensar os dados de maneira a alcançar uma forma compacta, permitindo uma melhor análise dos resultados como um todo. Essa facilidade é obtida em troca da perda de detalhes ocorrida pelo processo de sintetização das informações.

Esta apuração se torna necessária, a fim de que se possa fazer uma apresentação dos dados de forma adequada e de fácil entendimento. Esta apresentação pode ser:

- **Apresentação tabular ou**
- **Apresentação gráfica.**

A **apresentação tabular** é a disposição numérica dos dados na forma de tabela, que pode ser apenas por ordenamento, como também, distribuído na forma de classes, indicando a frequência dos dados em cada classe. As tabelas possuem a vantagem de, em um só lugar e de forma sintética, conseguirem apresentar grande quantidade de resultados sobre determinado assunto. Existem normas para a apresentação de tabelas ditadas pelo Conselho Nacional de Estatística (CNE) e pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que devem ser seguidas.

A **apresentação gráfica** é a disposição geométrica e comparativa dos dados divididos em classes ou algum tipo de separação. A representação gráfica dos dados permite uma análise de forma fácil, rápida e clara dos resultados obtidos, do fenômeno como um todo, bem como a sua variação ao longo do gráfico. Esta apresentação não deve substituir totalmente a apresentação tabular, por ser de grande importância e conter um maior detalhamento de informações. Os gráficos podem ser na forma:

- Gráficos em linha;
- Gráficos em colunas;
- Gráficos em barras;
- Gráficos em colunas e barras múltiplas agrupadas;
- Gráficos em setores, entre outros.

## **Análise e interpretação dos dados**

Nesta fase são feitos os cálculos das medidas necessárias para descrever o fenômeno, expressando os dados obtidos por números que irão resumir as características do conjunto. A análise dos dados é função direta desses valores obtidos nos cálculos das diversas medidas estatísticas, que deverão ser cuidadosamente interpretados.

É nesta fase que se faz a análise das amostras e uma generalização para as características da população, levando um grau de incerteza que é intrínseco da natureza dos levantamentos estatísticos.