**Tipos de variáveis**

Variável é a característica de interesse que é medida em cada elemento da amostra ou população. Como o nome diz, seus valores variam de elemento para elemento. As variáveis podem ter valores numéricos ou não numéricos.

Variáveis podem ser classificadas da seguinte forma:

**1- Variáveis Quantitativas:** são as características que podem ser medidas em uma escala quantitativa, ou seja, apresentam valores numéricos que fazem sentido. Podem ser discretas ou contínuas.

**a) Variáveis discretas:** características mensuráveis que podem assumir apenas um número finito ou infinito contável de valores e, assim, somente fazem sentido valores inteiros. Geralmente, são o resultado de contagens. Exemplos: número de filhos, número de bactérias por litro de leite, número de cigarros fumados por dia.

**b) Variáveis contínuas**, características mensuráveis que assumem valores em uma escala contínua (na reta real), para as quais valores fracionais fazem sentido. Usualmente devem ser medidas através de algum instrumento. Exemplos: peso (balança), altura (régua), tempo (relógio), pressão arterial, idade.

**2- Variáveis Qualitativas (ou categóricas)**: são as características que não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos. Podem ser nominais ou ordinais.

**a) Variáveis nominais**: não existe ordenação dentre as categorias. Exemplos: sexo, cor dos olhos, fumante/não fumante, doente/sadio.

**b) Variáveis ordinais**: existe uma ordenação entre as categorias. Exemplos: escolaridade (1º, 2º, 3º graus), estágio da doença (inicial, intermediário, terminal), mês de observação (janeiro, fevereiro,..., dezembro).

As distinções são menos rígidas do que a descrição acima insinua.

Uma variável originalmente quantitativa pode ser coletada de forma qualitativa.   
Por exemplo, a variável idade, medida em anos completos, é quantitativa (contínua); mas, se for informada apenas a faixa etária (0 a 5 anos, 6 a 10 anos, etc...), é qualitativa (ordinal). Outro exemplo é o peso dos lutadores de boxe, uma variável quantitativa (contínua) se trabalhamos com o valor obtido na balança, mas qualitativa (ordinal) se o classificarmos nas categorias do boxe (peso-pena, peso-leve, peso-pesado, etc.).

Outro ponto importante é que nem sempre uma variável representada por números é quantitativa.

O número do telefone de uma pessoa, o número da casa, o número de sua identidade. Às vezes o sexo do indivíduo é registrado na planilha de dados como 1 se macho e 2 se fêmea, por exemplo. Isto não significa que a variável sexo passou a ser quantitativa.

**Fases do Método Estatístico**

1) Coleta de dados: após cuidadoso planejamento e a devida determinação das características mensuráveis (o que dá para se medir) do fenômeno que desejamos pesquisar, damos início à coleta de dados.

Os dados podem ser escolhidos, por exemplo, através de:

- Questionários;

- Observação;

- Experimentação;

- Pesquisa bibliográfica.

Essa coleta pode ser:

Direta: Feita sobre elementos informativos de registro obrigatório (nascimentos, óbitos, compra e venda de mercadorias), ou, ainda quando os dados são coletados pelo próprio pesquisador através de questionários (censo demográfico). A coleta de dados direta pode ainda, ser classificada quanto ao fator em tempo em:

• Contínua – realizada continuamente (ex.: frequência de alunos às aulas). Dia a dia.  
• Periódica – realizada em intervalos constantes de tempos (ex: censo demográfico). Uma vez por mês, de dois em dois meses.

• Ocasional – realizada a fim de atender uma conjuntura ou a uma emergência (ex: epidemias, pesquisa de opinião pública a respeito do lançamento de um novo produto). Uma situação emergencial.

Indireta: quando é feita com base em elementos já pesquisados (revista, jornal, livros etc.).

2) Crítica de dados: é a conferência dos dados coletados; se houver erros, pode ser por motivos externos, ou seja, erros por parte do informante, ou por motivos internos, os quais ocorrem por parte do entrevistador ou da equipe de pesquisa. É a correção dos dados coletados.

3) Apuração dos dados: é a soma e o processamento dos dados obtidos e a disposição mediante critérios de classificação. Depois de realizada as etapas anteriores são feita a contagem do que foi obtido.

4) Exposição ou apresentação dos dados: é como vai ser apresentado o trabalho para o melhor entendimento da parte interessada de maneira mais clara possível, para que todos os interessados possam compreender..

Há três formas de apresentação que não se excluem mutuamente:

- apresentação por tabelas;

- apresentação por gráficos;

- relatórios.

Estas formas de apresentar dados permitem sintetizar grandes quantidades de dados, tornando mais fácil a compreensão do atributo em estudo e permitindo uma futura análise.

5) Análise dos resultados: são as conclusões sobre o trabalho realizado; análise e interpretação dos dados obtidos.

**População e amostra**

População: é o todo e pode ser finita ou infinita. Finita – possui um número determinado de elementos. Exemplo: número de alunos da classe, ou seja, tem um fim. Infinita – possui um grande número de elementos. Exemplo: a população da cidade de São Paulo. São dados de difícil contagem e que geram muitos custos às empresas de pesquisa.

Amostra: é um subconjunto da população, ou seja, uma parte dela. Quando há um número muito grande de elementos, fica difícil a observação dos aspectos a serem estudados de cada um dos elementos devido ao alto custo, ao intenso trabalho e ao tempo despendido para concluir uma exaustiva observação de todos os elementos da população. Neste caso, fazemos a seleção de uma amostra (cerca de 10% da população a ser estudada), e, por intermédio dessa observação, estaremos aptos a analisar os resultados de um todo, da mesma forma que se estudássemos toda a população.

Por exemplo: numa pesquisa de intenção de votos numa cidade de 656.000 habitantes, ao invés de entrevistar todos esses 656.000 habitantes poderíamos coletar a intenção de votos de 4.500 eleitores.

**Técnicas de amostragem**

**1. Casual ou aleatória simples**: equivalente a um sorteio, é utilizada quando a população encontra-se desordenada, sendo que por essa técnica, qualquer elemento tem a mesma chance de ser sorteado.

- Quando a população é relativamente pequena (até 30, por exemplo), pode-se numerar os elementos e, em seguida, realizar o sorteio.

**2. Sistemática:** nessa amostragem os elementos da população já se encontram ordenados e, nesses casos, não é necessário se construir um sistema de referência. É feita a retirada de elementos da população de acordo com critérios estabelecidos pelo próprio pesquisador.

Exemplos de populações ordenadas: fichas individuais de empregados (alfabética), casas de uma rua (número), notas fiscais (data), etc.

**3.** **Estratificada proporcional**: utilizada quando a população encontra-se dividida em estratos (ou camadas, faixas, intervalos etc). Determina-se a quantidade de indivíduos de cada estrato que serão selecionados.

Exemplos de populações divididas em estratos

-sexo (homem e mulher);

- idade (criança, adolescente, adulto e idoso);

- setores de uma empresa (administração, vendas, tesouraria, serviços gerais etc)

-Cursos de uma faculdade (Ciências Contábeis, Administração, Direito, Enfermagem etc);

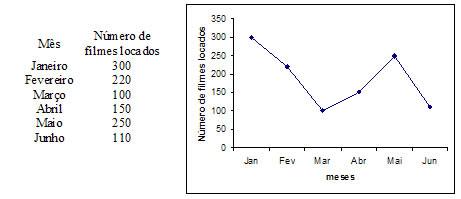
-Faixa salarial (até 1 SM, de 1 a 2 SM, de 2 a 4 SM, acima de 4 salários-mínimos).

**Gráficos para representar dados não agrupados**

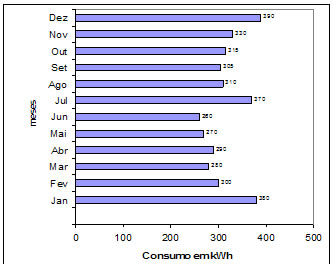
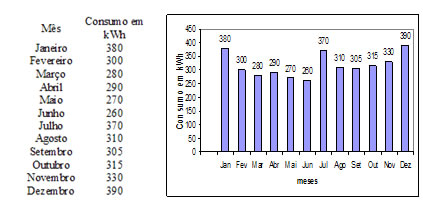
Quando temos uma tabela de variável qualitativa, um tipo de gráfico adequado para apresentar os resultados corresponde ao gráfico de setores, também popularmente conhecido como gráfico tipo pizza. Sua construção é simples: sabe-se que o ângulo de 360º equivale a 100% da área da circunferência; assim, para obter-se o ângulo do setor cuja área representa uma determinada frequência, basta resolver uma regra de três simples, como a apresentada a seguir:



**O gráfico de Segmento ou gráfico de linhas**é utilizado usualmente para verificar o comportamento de uma determinada variável ao longo do tempo. O eixo x representa anos, meses, semestres, semestres, entre outros.



No **Gráfico em Colunas ou Barras** os dados são representados através de retângulos dispostos verticalmente (colunas) ou horizontalmente (barras).



**Cartograma**

****

**Pictograma**

****

**Distribuição de frequência**

A simples organização dos dados em um **rol** (é a mais simples organização numérica; é a ordenação dos dados em ordem crescente ou decrescente)aumenta muito a capacidade de informação destes. Na Tabela 2 (rol), você pode verificar que o menor tempo observado foi 82 minutos, e o maior, 210 minutos, o que nos fornece uma **amplitude total** (corresponde à diferença entre o maior e o menor valor observado em um conjunto de dados - notaremos por **A**) de variação da ordem de 128 minutos.

Outra informação que podemos obter nos dados por meio da Tabela 2 (organizada em rol crescente) é que alguns tempos, como 122 min, 132 min, 138 min e 142 min, foram os mais frequentes, ou seja, os mais citados na pesquisa.



Uma maneira de organizar um conjunto de dados para você melhor representá-lo é por meio de uma **tabela de distribuição de frequências** (tabela onde são apresentadas as frequências de cada uma das classes).

Distribuindo-se os dados observados em **classes** (Intervalos nos quais os valores da variável analisada são agrupados.e contando-se o número de observações contidas em cada classe, obtém-se a **frequência de classe**. Veja que a disposição tabular dos dados agrupados em classes, juntamente com as frequências correspondentes, se denomina distribuição de frequências.

**Amplitude Total (H)**: É a diferença entre o maior valor e o menor valor observado da variável em estudo.

H = Xmáx –Xmín

**Limites de Classe**: São os números extremos de cada intervalo. Sendo assim, temos um limite inferior (li) e um superior (Li). Se a primeira classe tiver um intervalo de notas de 0 até 10, o 0 será o limite inferior enquanto que o 10 será o limite superior desta classe.

**Classe**: É cada um dos intervalos em que os dados são agrupados. Existem várias maneiras de apresentarmos o intervalo de classes: iguais ou diferentes entre si. Porém, sempre que possível, deveremos optar por intervalos iguais, o que facilitará os cálculos posteriores.

**Ponto médio das classes (Xi)**: É a média aritmética entre o limite superior e o limite inferior da classe.

**Número de Classes**: Existem vários critérios que podem ser utilizados a fim de determinar o número de classes, porém tais critérios servirão apenas como indicação e nunca como regra fixa, pois caberá sempre ao pesquisador estabelecer o melhor número, levando-se em conta o intervalo de classe e a facilidade para os posteriores cálculos numéricos. Temos:

- Fórmula de Sturges, que estabelece que o número de classes K é calculado por: K = 1 + 3,3 log n onde n = número de elementos observados.

- Regra da Potência de 2 onde k = menor valor interiro tal que 2k ≥ n

- Regra da Raiz Quadrada: √n

**Amplitude das Classes (h)**: h = Li - li

**Frequência absoluta ou simples simples (fi)**: É o número de observações que se encontra presente em uma classe ou intervalo especifico.

**Frequência acumulada (Fi)**: Corresponde à soma das frequências de determinada classe com as anteriores.

**Frequência relativa (fri)**: Corresponde ao quociente entre a frequência absoluta da classe e o total de elementos.

**Frequência acumulada relativa (Fri)**: Corresponde ao quociente entre a frequência acumulada e o total de elementos.

**Exercício**

Considere as notas de um teste de inteligência aplicado a alunos de um estabelecimento de ensino:



Construa a distribuição de frequências com intervalo de classes e responda:

a) Quantos alunos participaram do teste?

b) Qual a menor nota registrada?

c) Qual é o número de classes?

d) Qual é a terceira classe?

e) Qual é a amplitude total da amostra?

f) Qual é a amplitude total da distribuição?

g) Qual é o limite superior da quarta classe?

h) Qual é o limite inferior da sétima classe?

i) Qual é a amplitude da quinta classe?

**Gráficos Estatísticos para dados agrupados**

Os gráficos constituem uma forma clara e objetiva de apresentar dados estatísticos. A intenção é a de proporcionar aos leitores em geral a compreensão e a veracidade dos fatos. De acordo com a característica da informação precisamos escolher o gráfico correto.

**Histogramas** são constituídos por um conjunto de retângulos, com as bases assentadas sobre um eixo horizontal, tendo o centro da mesma no ponto médio da classe que representa, e cuja altura é proporcional à frequência da classe. São gráficos utilizados para representar tabelas intervalares.



**Polígono de frequências** é um gráfico de análise no qual as frequências das classes são localizadas sobre perpendiculares levantadas nos pontos médios das classes. O polígono de frequência, você pode obter pela simples união dos pontos médios dos topos dos retângulos de um histograma. Completa-se o polígono unindo as extremidades da linha que ligam os pontos representativos das frequências de classe aos pontos médios das classes, imediatamente, anterior e posterior às classes extremas, que têm frequência nula.



Os gráficos chamados de **ogivas** correspondem a um polígono de frequências acumuladas, nas quais estas frequências são localizadas sobre perpendiculares levantadas nos limites inferiores ou superiores das classes, dependendo se a ogiva representar as frequências acumuladas.



**Medidas de posição**

As medidas de posição ou de tendência central constituem uma forma mais sintética de apresentar os resultados contidos nos dados observados, pois representam um valor central, em torno do qual os dados se concentram. As medidas de tendência central mais empregadas são a média, a mediana e a moda. A **média aritmética** é a mais usada das três medidas de posição mencionadas, por ser a mais comum e compreensível delas, bem como pela relativa simplicidade do seu cálculo, além de prestar-se bem ao tratamento algébrico. De uma forma geral, podemos interpretar a média como sendo um valor típico do conjunto de dados. Pode ser um valor que não pertence ao conjunto de dados.

**Dados não agrupados – Média, mediana e moda**

Tomem-se como dados as seguintes classificações de um aluno no final do1º período do 11º ano:

12, 13, 13, 13, 14, 15, 15, 16, 16, 17, 17   
**Média:** calcula-se somando as classificações, dividindo este total pelo seu número:

(12 + 13 + 13 + 13 + 14 + 15 + 15 + 16 + 16 + 17 + 17 ) / 11 = 14,6  
**Moda:** é o valor com maior frequência. (13)

Apesar de seu significado ser simples, a moda nem sempre é única. Quando no conjunto existirem poucas observações, muito frequentemente não há valores repetidos, com o que nenhum deles satisfaz a condição de moda. Se o peso (em Kg) correspondente a nove pessoas são: 82; 65; 59; 74; 60; 67; 71 e 73 estes nove dados não possuem uma moda, sendo um conjunto amodal. Por outro lado, se a distribuição de peso de 15 pessoas for: 63; 67; 70; 69; 81; 57; 63; 73; 68; 63; 71; 71; 71 e 83, possui duas modas (63 e 71 Kg). Neste caso a distribuição diz-se bimodal. Será unimodal no caso de apresentar uma só moda e assim por diante

**Mediana:** é a observação que separa 50% dos valores mais baixos dos 50% mais altos.

Tendo as classificações ordenadas, verifica-se facilmente que o 15 separa 50% das classificações mais baixas (12, 13, 13, 13 e 14) de 50% das classificações mais altas (15, 16, 16, 17 e 17).

Observação: Como o número de observações era ímpar (11) havia um valor central. Nos casos em que o número de observações é par temos dois valores centrais. Então, para obter a mediana, calcula-se a sua média destes.

**Exercícios**  
1. Considera os seguintes dados:

10, 11, 17, 15, 14, 13, 12, 12, 12, 14   
Calcular:

a) a média

b) a moda

c) a mediana

2. Considera os seguintes dados:

8, 9, 18, 16, 14, 13, 11, 7, 12   
Calcular:

a) a média b) a moda c) a mediana

.**Dados agrupados – Média, mediana e moda**

**Média:** Em algumas situações estatísticas, os dados são apresentados em intervalos agrupados. Dessa forma, o cálculo da média aritmética é realizado de forma mais complexa. Nesse caso, temos que determinar primeiramente o ponto médio de cada intervalo multiplicando o resultado pela frequência absoluta (frequência simples) do intervalo. O somatório desses produtos deverá ser dividido pelo somatório da frequência absoluta, constituindo a média dos valores agrupados em intervalos. Observe o seguinte exemplo:

A tabela a seguir mostra a massa (em quilograma) de um grupo de pessoas. Os dados foram informados em intervalos. Veja:

|  |  |  |  |
| --- | --- | --- | --- |
| **Massa (Kg)** | **fi** | **xi** | **fi .xi** |
| 50 ─ 54 | 3 |  |  |
| 54 ─ 58 | 6 |  |  |
| 58 ─ 62 | 10 |  |  |
| 62 ─ 66 | 15 |  |  |
| 66 ─ 70 | 20 |  |  |
| 70 ─ 74 | 12 |  |  |
| 74 ─ 78 | 8 |  |  |
| 78 ─ 82 | 7 |  |  |
| 82 ─ 86 | 5 |  |  |
| 86 ─ 90 | 2 |  |  |
| **Total** | 88 |  |  |

Calcule a média aritmética.

**Moda:** No caso de dados agrupados em tabelas de frequências, o cálculo é feito por:

[http://www.infoescola.com/wp-content/uploads/2008/09/moda.jpg](http://www.infoescola.com/wp-content/uploads/2008/09/moda.jpg)

Onde:

* I = limite inferior da classe modal;
* f1 = frequência da classe modal
* f0 = frequência da classe que precede a classe modal
* f2 = frequência da classe que sucede a classe modal
* h = tamanho do intervalo de classe

Calcular a moda para a tabela anterior.

**Mediana:**



Calcule a mediana para a tabela anterior.

**Exercício**

Calcule a média, a moda e a mediana para a tabela abaixo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estatura (cm)** | **fi** | **xi** | **fi .xi** | **Fi** |
| 150 ─ 155 | 5 |  |  |  |
| 155 ─ 160 | 11 |  |  |  |
| 160 ─ 165 | 16 |  |  |  |
| 165 ─ 170 | 13 |  |  |  |
| 170 ─ 175 | 12 |  |  |  |
| 175 ─ 180 | 9 |  |  |  |
| 180 ─ 185 | 8 |  |  |  |
| 185 ─ 190 | 6 |  |  |  |
| **Total** | 80 |  |  |  |