MÉTODOS ESTATÍSTICOS
Sumário

APRESENTAÇÃO ......................................................................................................................... 5

CONHECENDO OS PROFESSORES .......................................................................................... 7

UNIDADE 1 – CONCEITOS BÁSICOS ..................................................................................... 9
  Seção 1.1 – Conceitos básicos ............................................................................................... 9
  Seção 1.2 – Aplicação dos Métodos Estatísticos ................................................................. 13

UNIDADE 2 - TABELAS ............................................................................................................. 15
  Seção 2.1 – Representação de dados em tabelas simples ................................................... 15
  Seção 2.2 – Representação de dados em tabelas cruzadas ................................................. 19
  Seção 2.3 – Representação de dados em tabela de distribuição de freqüência .............. 20

UNIDADE 3 - GRÁFICOS .......................................................................................................... 29
  Seção 3.1 – Gráfico para representar variáveis qualitativas e quantitativas discretas ....... 30
    3.1.1 Gráfico de setores ....................................................................................................... 30
    3.1.2 Gráfico de colunas simples ...................................................................................... 31
    3.1.3 Gráfico de barras simples ....................................................................................... 32
    3.1.4 Gráfico de colunas agrupadas ................................................................................ 32
    3.1.5 Gráfico de barras agrupadas ................................................................................... 33
  Seção 3.2 – Gráfico para representar uma série temporal .................................................... 33
  Seção 3.3 – Gráfico para representar duas variáveis quantitativas .................................... 34
  Seção 3.4 – Gráfico para representar uma distribuição de freqüência ................................ 35
    3.4.1 Histograma .............................................................................................................. 35
    3.4.2 Polígono de freqüência ........................................................................................... 36
Seção 3.5 – Outros tipos de gráficos .................................................................37
  3.5.1 Pirâmide etária .......................................................................................37
  3.5.2 Cartograma ............................................................................................38

UNIDADE 4 – MEDIDAS DESCRITIVAS .............................................................39
Seção 4.1 – Medidas de posição .................................................................42
  4.1.1 Média aritmética ..................................................................................42
  4.1.2 Média ponderada ..................................................................................43
  4.1.3 Mediana ...............................................................................................44
  4.1.4 Moda ....................................................................................................47
Seção 4.2 – Medidas de dispersão .................................................................48
  4.2.1 Variância ...............................................................................................48
  4.2.2 Desvio-padrão ....................................................................................49
  4.2.3 Coeficiente de variação ........................................................................50

UNIDADE 5 – FERRAMENTAS DE ANÁLISE ESTATÍSTICA NO EXCEL ..........53
Seção 5.1 – Elaborando um banco de dados no EXCEL ..................................55
Seção 5.2 – Utilizando as técnicas estatísticas para analisar os dados no EXCEL ..56
Seção 5.3 – Medidas descritivas para as variáveis quantitativas ......................57
Seção 5.4 – Tabelas simples para as variáveis qualitativas
  e para as variáveis quantitativas com pouca variabilidade .........................59
Seção 5.5 – Tabelas cruzadas .........................................................................66
Seção 5.6 – Gráficos de setores, colunas e barras ..........................................68

REFERÊNCIAS .................................................................................................79
Apresentação

O componente curricular *Métodos Estatísticos* visa a fornecer embasamento teórico e prático para tratar as informações obtidas em uma pesquisa de forma correta, fornecendo o suporte necessário para auxiliar a gestão pública e a compreensão de fatos sociais e econômicos.

Iniciamos com a abordagem dos conceitos básicos, na sequência são apresentadas as técnicas para organização e análise de dados em tabelas e gráficos, seguidas das medidas descritivas e finalizamos com a aplicação computacional na planilha eletrônica Excel.

Na expectativa de que o presente componente curricular se constitua em importante subsídio para os alunos e para a futura atuação profissional, fornecendo embasamento para tratar e analisar corretamente as informações, ficamos à disposição para esclarecimentos ou sugestões.

_Iara Denise Endruweit Battisti_
ulaire.battisti@unijui.edu.br

_Gerson Battisti_
battisti@unijui.edu.br

Atua como professora de Estatística no magistério superior desde agosto de 1998 na Unijuí e nos cursos de especialização da mesma Universidade. Também atua como pesquisadora em estatística aplicada.


Atua como professor de Informática e Ciência da Computação na Unijuí desde agosto de 1994.
Conceitos Básicos

Objetivos Desta Unidade

• Compreender o que é Estatística e seus conceitos básicos.

• Conhecer onde se aplica os Métodos Estatísticos.

As Seções Desta Unidade

Seção 1.1 – Conceitos básicos

Seção 1.2 – Aplicação dos Métodos Estatísticos

Seção 1.1

Conceitos Básicos

Você já ouviu a palavra “Estatística”? O que você lembra quando escuta esta palavra?
O que você lembra quando ouve a palavra probabilidade? Se você respondeu números, gráficos, tabelas, contas, então acertou em parte, pois a Estatística é muito mais que isso. Além de fornecer métodos para gerar tabelas e gráficos e produzir resultados por meio de contas, a Estatística também dispõe de métodos para coletar dados e auxiliar na análise destes.

Lembra das pesquisas eleitorais que frequentemente são feitas antes das eleições para prefeito, vereadores, presidente e deputados? Nestas são aplicados os métodos estatísticos, isto é, primeiramente a pesquisa é planejada: quantos indivíduos serão entrevistados, como
serão selecionados, onde e quando ocorrerá a pesquisa? Depois os dados são processados, gerando tabelas, gráficos e medidas descritivas. No final estes dados são analisados por meio de testes estatísticos e então divulgados.

Então, a **Estatística** fornece métodos para a coleta, organização, análise e interpretação de dados. Apresentaremos neste componente curricular uma introdução aos métodos estatísticos necessários para apoiar as tomadas de decisão na gestão pública.

Realizamos uma pesquisa quando temos uma questão para responder, que chamamos de "questão de pesquisa", por exemplo: qual é proporção de mulheres que exercem atividade remunerada no município de Santa Rosa no ano de 2008?

Utilizaremos essa questão de pesquisa como exemplo para apresentar alguns conceitos básicos necessários na compreensão e aplicação dos métodos estatísticos.

**População:** são todos os elementos que têm uma ou mais características em comum definidas antes de se iniciar qualquer pesquisa. Os elementos podem ser pessoas, animais, plantas, objetos, etc. Para o exemplo dado, a população é definida por todas as mulheres acima de 14 anos que residem no município de Santa Rosa.

A letra "N" indica o tamanho da população.

O estudo de todos os elementos de uma população é chamado de censo. Realizar um censo necessita de tempo e tem custo elevado, assim, frequentemente, utilizamos uma amostra da população.
**Amostra**: é uma parte representativa da população. O tamanho da amostra é definido antes de se iniciar a pesquisa. A amostra deve ser aleatória, isto é, selecionada por sorteio. Para o exemplo dado, a amostra é definida por uma parte da população de mulheres acima de 14 anos que residem no município de Santa Rosa.

A letra “n” indica o tamanho da amostra.

**Variável aleatória**: são as características de uma população ou uma amostra. Para o exemplo dado, as variáveis aleatórias são as questões que o instrumento de coleta de dados (também chamado de questionário) contem, por exemplo: idade, estado civil, escolaridade, número de filhos, qual atividade exerce, tempo que exerce a atividade, quantas horas trabalha por semana, se é autônoma ou empregada, e muitas outras.

Classificamos as variáveis aleatórias em qualitativas e quantitativas. As variáveis qualitativas têm seus valores (respostas para cada questão do questionário) não numéricos, como sexo, estado civil, nível de escolaridade, bairro, profissão, nível de satisfação. As variáveis quantitativas têm seus valores numéricos, tais como: idade, peso, salário, tempo de serviço, número de filhos.

As variáveis qualitativas são subdivididas em nominais e ordinais. Quando as diferentes categorias (respostas) não têm relação entre si, ou seja, são independentes, classificamos a variável como qualitativa nominal, por exemplo, sexo, estado civil, curso de graduação e bairro. Por outro lado, quando as categorias têm uma relação entre si, geralmente atribuindo níveis, como o nível de escolaridade e o grau de satisfação do cliente, são denominadas qualitativas ordinais.

As variáveis quantitativas são subdivididas em discretas e contínuas. As primeiras assumem somente valores numéricos inteiros como: número de filhos, número de alunos, número de computadores. Já as variáveis quantitativas contínuas podem assumir qualquer valor numérico, resultado de uma medida, como: peso, idade e salário. Nesta última classificação os valores geralmente são registrados até a precisão da medida utilizada, por exemplo, a idade de uma pessoa pode ser registrada em anos (25 anos), meses (310 meses) ou ainda em anos e meses (25 anos e 10 meses).
A Figura 1 apresenta um esquema de classificação das variáveis aleatórias:

![Diagrama de classificação das variáveis aleatórias](image)

Figura 1 – Classificação das variáveis aleatórias  
Fonte: Construção dos autores.

A classificação das variáveis é muito importante, pois diferentes tipos de variáveis exigem tratamentos estatísticos específicos, por exemplo: qual é a idade média das mulheres que exercem atividade remunerada? Qual é a proporção (percentual) de mulheres que trabalham fora o dia todo?

**Estatística descritiva e inferência estatística:** A estatística descritiva é o conjunto de métodos para descrever e sistematizar os dados de uma amostra ou uma população, por meio de tabelas, gráficos e medidas descritivas, sendo que estes métodos serão descritos nas próximas seções. A inferência estatística é o conjunto de métodos para projetar os resultados de uma amostra para uma população. Veja que a figura 2 representa a estatística descritiva e a inferência estatística.

![Diagrama de estatística descritiva e inferência estatística](image)

Figura 2 – Estatística descritiva e inferência estatística  
Fonte: Construção dos autores.
**Dados primários e dados secundários:** os dados primários estão disponíveis na sociedade (idade, sexo, estado civil...) e os secundários estão organizados de alguma forma, geralmente nos meios de comunicação e publicações científicas (tabelas, gráficos...).

**Regra do arredondamento:** na apresentação dos resultados dos cálculos usamos a regra do arredondamento para “quebrar” o número na quantidade de casas decimais desejadas ou padronizadas, da seguinte forma:

- se o algarismo a ser excluído for 0, 1, 2, 3, 4 o algarismo que permanece não é alterado;
- se o algarismo a ser excluído for 5, 6, 7, 8, 9 o algarismo que permanece aumenta de uma unidade.

Por exemplo:

- o número 23,578 fica 23,6 considerando uma casa decimal
- o número 23,538 fica 23,5 considerando uma casa decimal
- o número 23,98 fica 24,0 considerando uma casa decimal

**Seção 1.2 Aplicação dos Métodos Estatísticos**

A Estatística é aplicada em diversas áreas do conhecimento. Aqueles que se valerem dos métodos estatísticos estarão mais preparados para organizar e analisar os dados em sua atuação profissional, como também nas informações que são divulgadas dia-a-dia nos meios de comunicação.

Na área das Ciências Humanas e Sociais podemos fazer uma pesquisa de opinião para conhecer a satisfação dos clientes de uma empresa em relação aos serviços prestados ou a opinião dos habitantes de um município sobre as prioridades de investimento público ou ainda para acompanhar o valor da cesta básica.
No domínio das Ciências da Saúde, fazemos pesquisa para saber a prevalência de diabetes na população ou para saber que tratamento é mais eficaz para a hipertensão. Nas Ciências Biológicas podemos fazer pesquisa para projetar a população de determinadas espécies em determinado local.

No âmbito das Ciências Agrárias podemos fazer pesquisa para determinar a quantidade adequada de adubo em uma plantação. Na Engenharia podemos realizar uma pesquisa para saber a resistência de uma coluna de concreto. Esses são alguns exemplos da grande aplicação que a Estatística tem em todas as áreas do conhecimento.

Nosso interesse pode ser analisar dados secundários, como os dados que o IBGE (Instituto Brasileiro de Geografia e Estatística – <www.ibge.gov.br>) ou os dados que o Datasus (<www.datasus.gov.br>) disponibilizam.

Muitas vezes utilizamos o conhecimento dos métodos estatísticos para compreender melhor os dados que são disponibilizados nos meios de comunicação, como a TMI (Taxa de Mortalidade Infantil), INPC (Índice Nacional de Preços ao Consumidor) e tantos outros.

**SÍNTESE DA UNIDADE 1**

Nesta Unidade explicamos o que é Estatística, seus conceitos básicos e onde aplicamos os métodos estatísticos.
Tabelas

Objetivos Desta Unidade

• Verificar como apresentamos os dados em uma tabela.

• Interpretar os resultados de uma tabela.

As Seções Desta Unidade

Seção 2.1 – Representação de dados em tabelas simples

Seção 2.2 – Representação de dados em tabelas cruzadas

Seção 2.3 – Representação de dados em tabela de distribuição de frequência

Quando realizamos uma pesquisa para coletar dados primários, primeiramente organizamos os dados em tabelas e gráficos para termos uma primeira ideia da distribuição destes dados. Existem diferentes tipos de tabelas e gráficos, como você verá neste componente curricular. Também muitos dados secundários são apresentados em tabelas e gráficos e assim precisamos saber como interpretá-los. Estudaremos nesta unidade as tabelas, e para tanto ela está dividida em três seções. Os gráficos serão estudados na próxima unidade.

Seção 2.1

Representação de Dados em Tabelas Simples

Na tabela podemos apresentar vários dados de forma organizada, facilitando o entendimento. Geralmente apresentamos um texto logo após a tabela com a análise dos dados. Para sintetizar os dados numa tabela seguimos algumas normas, como apresentamos a seguir.
A tabela tem três partes principais: título, corpo da tabela e fonte.

Quando elaboramos um título de uma tabela tentamos responder a três perguntas: O quê? Onde? Quando?

Na fonte apresentamos a origem dos dados; se estes forem originários de fonte secundária é obrigatório apresentar. O tamanho de letra da fonte é menor do que a do corpo da tabela. Na linha abaixo da fonte pode ser apresentada uma nota de esclarecimento sobre a tabela. As laterais não possuem borda (traço), somente uma borda horizontal superior e uma horizontal inferior para delimitar o corpo da tabela, e mais uma borda para separar o cabeçalho do resto da tabela. Se a tabela possui uma linha do total então devemos colocar uma borda acima e abaixo desta linha. Se existir mais de uma tabela estas são numeradas em ordem sequencial.

Na tabela simples são apresentados os valores referentes às categorias (respostas) de uma variável aleatória, possuindo geralmente três colunas, como representado a seguir:

<table>
<thead>
<tr>
<th>Título</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nome da variável</td>
</tr>
<tr>
<td>Categoria 1</td>
</tr>
<tr>
<td>Categoria 2</td>
</tr>
<tr>
<td>Categoria 3</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Fonte
A primeira linha do corpo da tabela é o seu cabeçalho, e é usada para identificar o que cada coluna contém. Em uma tabela simples, na primeira coluna apresentamos as categorias da variável aleatória, na segunda apresentamos a quantidade de vezes (freqüência absoluta) que cada categoria aparece e na terceira coluna apresentamos o percentual (freqüência relativa) para cada categoria.

Utilizamos a freqüência absoluta para ordenar as categorias de uma variável qualitativa nominal na tabela (Tabela 1). Nas variáveis qualitativas ordinais e quantitativas, ordenamos as categorias por ordem crescente (ou decrescente) das próprias categorias (Tabela 2). A seguir alguns exemplos de tabelas simples:

| Tabela 1: Distribuição por sexo, Rio Grande do Sul, 2007 |
|-------------|---------|--------|
| Sexo        | n¹      | %      |
| Feminino    | 3.150.909 | 50,44  |
| Masculino   | 3.095.615 | 49,56  |
| Total       | 6.246.524 | 100,00 |

¹ População estimada incluindo os domicílios fechados é de 6.273.345.

Observamos na Tabela 1 que o percentual de mulheres é maior que o percentual de homens no Rio Grande do Sul em 2007, por isso apresentamos a categoria feminino em primeiro lugar, pois estamos analisando uma variável qualitativa nominal.

| Tabela 2: Nível de satisfação dos clientes da empresa XY, abril/2008 |
|---------------------|-------|-----|
| Nível de satisfação | n     | %   |
| Muito satisfeito    | 57    | 24,8|
| Satisfeito          | 73    | 44,5|
| Mais ou menos       | 10    | 6,1 |
| Insatisfeito        | 21    | 12,8|
| Muito insatisfeito  | 3     | 1,8 |
| Total               | 164   | 100,00 |

Fonte: Pesquisa por amostragem realizada na empresa XY.
Observamos na Tabela 2 que a maioria (24,8% + 44,5%) dos clientes está muito satisfeito ou satisfeita. As frequências são apresentadas na ordem decrescente das categorias (iniciando em muito satisfeito e finalizando em muito insatisfeito) por estarmos analisando uma variável qualitativa ordinal.

Para calcular o percentual (%) de cada categoria, dividimos a frequência absoluta da categoria pelo total e multiplicamos por 100. Um exemplo de como calcular o percentual (%) na Tabela 2:

Muito satisfeito = \( \frac{57}{164} \times 100 = 24,8\% \)

**Tabela de série temporal**

Quando temos dados registrados em diferentes momentos no tempo, mas igualmente espaçados (mês a mês, ano a ano, hora a hora...) referente a mesma variável aleatória chamamos de uma série temporal.

Apresentamos o tempo sempre na primeira coluna da tabela e na segunda coluna apresentamos os valores que a variável assume nos diferentes momentos do tempo. Observe o exemplo a seguir.

Tabela 3: Rendimento médio nominal do trabalho principal, habitualmente recebido por mês, pelas pessoas de 10 anos ou mais de idade na região metropolitana de Porto Alegre, janeiro a dezembro/2007.

<table>
<thead>
<tr>
<th>Mês</th>
<th>Valor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janeiro</td>
<td>1008,7</td>
</tr>
<tr>
<td>Fevereiro</td>
<td>1040,6</td>
</tr>
<tr>
<td>Março</td>
<td>1057,3</td>
</tr>
<tr>
<td>Abril</td>
<td>1055,7</td>
</tr>
<tr>
<td>Maio</td>
<td>1057,2</td>
</tr>
<tr>
<td>Junho</td>
<td>1068,0</td>
</tr>
<tr>
<td>Julho</td>
<td>1080,1</td>
</tr>
<tr>
<td>Agosto</td>
<td>1074,6</td>
</tr>
<tr>
<td>Setembro</td>
<td>1095,8</td>
</tr>
<tr>
<td>Outubro</td>
<td>1090,0</td>
</tr>
<tr>
<td>Novembro</td>
<td>1108,8</td>
</tr>
<tr>
<td>Dezembro</td>
<td>1114,4</td>
</tr>
</tbody>
</table>


Seção 2.2

Representação de Dados em Tabelas Cruzadas

Nas tabelas cruzadas apresentamos duas ou mais variáveis. Geralmente utilizamos uma tabela cruzada para representar a relação entre duas variáveis aleatórias, que neste caso também pode ser chamada de tabela de dupla entrada.

O formato de uma tabela cruzada, com duas variáveis aleatórias, é apresentado a seguir:

<table>
<thead>
<tr>
<th>Nome da variável 1</th>
<th>Categoria 1 da variável 2</th>
<th>Categoria 2 da variável 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Título</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nome da variável 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categoría 1 da variável 1</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Categoría 2 da variável 1</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Categoría 3 da variável 1</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Total</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
</tbody>
</table>

Fonte

Temos três formas de apresentar o percentual (%) numa tabela cruzada: por linha, por coluna e por total. Isto é, quando queremos comparar as categorias da variável 1, utilizamos o percentual por linha. Quando o intuito é comparar as categorias da variável 2, utilizamos o percentual por coluna. O percentual total considera todos os elementos pesquisados. Vamos utilizar o exemplo a seguir para facilitar a compreensão:
Tabela 4: Nível de satisfação dos clientes da empresa XY em relação ao sexo, abril/2008

<table>
<thead>
<tr>
<th>Nível de satisfação</th>
<th>Sexo</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Feminino</td>
<td>Masculino</td>
<td>Total</td>
</tr>
<tr>
<td>Muito satisfeito</td>
<td>43 (44,3%)</td>
<td>14 (20,9%)</td>
<td>57 (24,8%)</td>
</tr>
<tr>
<td>Satisfeito</td>
<td>44 (45,4%)</td>
<td>29 (43,3%)</td>
<td>73 (44,5%)</td>
</tr>
<tr>
<td>Mais ou menos</td>
<td>4 (4,1%)</td>
<td>6 (9,0%)</td>
<td>10 (6,1%)</td>
</tr>
<tr>
<td>Insatisfeito</td>
<td>5 (5,2%)</td>
<td>16 (23,8%)</td>
<td>21 (12,8%)</td>
</tr>
<tr>
<td>Muito insatisfeito</td>
<td>1 (1,0%)</td>
<td>2 (3,0%)</td>
<td>3 (1,8%)</td>
</tr>
<tr>
<td>Total</td>
<td>97 (100,0%)</td>
<td>67 (100,0%)</td>
<td>164 (100,0%)</td>
</tr>
</tbody>
</table>

Fonte: Pesquisa por amostragem realizada na empresa XY.

Na Tabela 4 estamos comparando o nível de satisfação entre os sexos, assim calculamos o percentual (%) na coluna, por exemplo:

Muito satisfeito para o sexo feminino: \( \frac{43}{97} \times 100 = 44,3\% \)

Muito satisfeito para o sexo masculino: \( \frac{14}{67} \times 100 = 20,9\% \)

Observamos que 89,7% (44,3% + 45,4%) das mulheres estão muito satisfeitas ou satisfeitas, enquanto 65,2% (20,9% + 43,3%) dos homens estão satisfeitos ou muito satisfeitos.

Seção 2.3

Representação de Dados em Tabela de Distribuição de Freqüências

Uma tabela de distribuição de freqüência é utilizada para representar respostas de uma variável aleatória quantitativa quando o tamanho do conjunto de dados é maior ou igual a 20 (\( n \geq 20 \)).

Tabela 5: Rendimento médio nominal do trabalho principal, habitualmente recebido por mês, pelas pessoas de 10 anos ou mais de idade, ocupadas na semana de referência, em Porto Alegre, janeiro/2005 a dezembro/2007

<table>
<thead>
<tr>
<th>Mês</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janeiro</td>
<td>903,2</td>
<td>953,0</td>
<td>1008,7</td>
</tr>
<tr>
<td>Fevereiro</td>
<td>940,3</td>
<td>971,3</td>
<td>1040,6</td>
</tr>
<tr>
<td>Março</td>
<td>915,0</td>
<td>981,2</td>
<td>1057,3</td>
</tr>
<tr>
<td>Abril</td>
<td>913,7</td>
<td>969,6</td>
<td>1055,7</td>
</tr>
<tr>
<td>Maio</td>
<td>922,9</td>
<td>988,7</td>
<td>1057,2</td>
</tr>
<tr>
<td>Junho</td>
<td>930,0</td>
<td>972,1</td>
<td>1068,0</td>
</tr>
<tr>
<td>Julho</td>
<td>940,8</td>
<td>996,1</td>
<td>1080,1</td>
</tr>
<tr>
<td>Agosto</td>
<td>950,8</td>
<td>1004,9</td>
<td>1074,6</td>
</tr>
<tr>
<td>Setembro</td>
<td>956,2</td>
<td>1016,2</td>
<td>1095,8</td>
</tr>
<tr>
<td>Outubro</td>
<td>967,2</td>
<td>1019,5</td>
<td>1090,0</td>
</tr>
<tr>
<td>Novembro</td>
<td>941,4</td>
<td>1035,6</td>
<td>1108,8</td>
</tr>
<tr>
<td>Dezembro</td>
<td>957,2</td>
<td>1023,4</td>
<td>1114,4</td>
</tr>
</tbody>
</table>


Nota: Exclusive trabalhadores não remunerados e trabalhadores que receberam somente em benefícios.

Para visualizar os dados brutos da Tabela 5 acesse o site do IBGE no seguinte link: <http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pme_nova/default.shtm>.

As primeiras informações que podemos obter desta tabela é o número de dados, o valor mínimo, o valor máximo e a amplitude total, como segue:

\[ n = \text{número de dados}, \quad \text{para este exemplo} \quad n = 36 \quad (\text{valores do rendimento nominal em 36 meses}) \]

\[ \text{Li} = \text{limite inferior ou valor mínimo}, \quad \text{para este exemplo} \quad \text{Li} = 903,2 \]

\[ \text{Ls} = \text{limite superior ou valor máximo}, \quad \text{para este exemplo} \quad \text{Ls} = 1114,4 \]

\[ \text{At} = \text{Ls} – \text{Li} = \text{amplitude total} \quad \text{(variação entre o maior e o menor valor)}, \quad \text{para este exemplo} \quad \text{At} = 1114,4 - 903,2 = 211,2 \]
Assim, podemos concluir que nos 36 meses analisados, o rendimento médio mínimo das pessoas residentes em Porto Alegre foi de R$ 903,20 e o rendimento médio máximo foi de R$ 1.114,40, nos meses de janeiro/2005 e dezembro/2007 respectivamente. Da mesma forma, pode-se afirmar que a variação entre o menor e maior valor foi de R$ 211,20, isto é, a variação entre o rendimento médio mínimo e o rendimento médio máximo.

A tabela de distribuição de frequências é construída tendo como base um conjunto de classe ou intervalos. Os intervalos são divisões de valores que permitem agrupar ou contabilizar todos os dados observados. Cada valor observado só pode estar em um dos intervalos definidos.

O número de intervalos que serão utilizados para construir a tabela pode ser definido de forma aleatória, porém em determinados casos o uso de poucos ou muitos intervalos pode esconder informações relevantes sobre os dados. Para evitar estes problemas existe uma pequena fórmula para calcular a amplitude dos intervalos de acordo com a quantidade de dados existentes e amplitude total dos dados. A fórmula é a seguinte:

\[ hi = \frac{At}{\sqrt{n}} \]

Em que:

- \( hi \): tamanho do intervalo
- \( \sqrt{n} \): número de intervalos
- \( At \): amplitude total

Relembrando, o \( n \) representa o número de dados que estamos analisando. Neste exemplo, são os valores que representam o rendimento médio do trabalhador nos anos de 2005, 2006 e 2007, de acordo com a Tabela 5. A amplitude total é a diferença do maior valor pelo menor valor presente nos dados (211,2).

A nossa fórmula, portanto, fica assim:

\[ hi = \frac{211,2}{\sqrt{36}} = 35,2 \]
Quando o valor de $hi$ gerado é um valor com casas decimais, podemos aproximá-lo para um valor próximo inteiro, de acordo com a precisão desejada. Neste caso aproximamos para cima, logo, $hi = 36$.

Uma vez definido o $hi$, podemos iniciar a construção da tabela de distribuição de frequência. O primeiro intervalo tem como limite inferior o menor valor encontrado nos dados ou o menor valor inteiro mais próximo do valor inferior. O limite superior é a soma do menor valor com o valor do $hi$ ($903 \rightarrow 939$).

Os próximos intervalos são gerados tendo como o limite inferior o limite superior do intervalo anterior e o limite superior como a soma do limite inferior desse intervalo com o $hi$, e assim sucessivamente até definir o intervalo que englobe o maior valor presente nos dados.

Exemplo:

Como o menor valor de todo o conjunto de dados é 903,2, decidiu-se pegar o valor inteiro menor mais próximo a este, isto é, o número 903. Este, somado ao $hi=36$, resulta no limite superior deste intervalo: 903 + 36 = 939. Então o primeiro intervalo da tabela tem o limite inferior = 903 e o limite superior = 939, como segue: 903 \rightarrow 939.

O segundo intervalo da tabela tem o limite inferior igual ao limite superior do intervalo anterior, como segue: 939 \rightarrow 975.

O número 975 é obtido somando 939 + 36, lembrando que $hi=36$ e assim sucessivamente.

Uma vez definidos os intervalos, a tabela de distribuição de frequências é construída contabilizando a quantidade de valores dos dados que estão dentro de cada intervalo. De acordo com os valores da Tabela 5, temos a seguinte distribuição dos valores:
No intervalo 903 |— 939 temos 5 valores, no intervalo 939 |— 975 temos 11 valores e assim por diante.

A tabela de distribuição de frequência é complementada com um conjunto de valores gerados a partir da distribuição calculada anteriormente. A tabela completa é apresentada a seguir, permitindo obter mais informações sobre o comportamento dos dados que foram coletados.

### Exemplo de tabela de distribuição de frequência

Tabela 6: Rendimento médio nominal do trabalho principal, habitualmente recebido por mês, pelas pessoas de 10 anos ou mais de idades ocupadas na semana de referência, em Porto Alegre, janeiro/2005 a dezembro/2007

<table>
<thead>
<tr>
<th>Intervalos</th>
<th>Contagem</th>
<th>Freqüência (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>903</td>
<td>— 939</td>
<td>// // // // //</td>
</tr>
<tr>
<td>975</td>
<td>— 1011</td>
<td>// // // // //</td>
</tr>
<tr>
<td>1011</td>
<td>— 1047</td>
<td>// // // // //</td>
</tr>
<tr>
<td>1047</td>
<td>— 1083</td>
<td>// // // // //</td>
</tr>
<tr>
<td>1083</td>
<td>— 1119</td>
<td>// // // // //</td>
</tr>
</tbody>
</table>


Nota: Exclusive trabalhadores não remunerados e trabalhadores que receberam somente em benefícios.
Os valores adicionados à tabela são:

- $f_i = \text{freqüência absoluta simples}$. Representa o número de valores existentes em cada intervalo. Para o primeiro intervalo $f_i = 5$. Também pode ser representado por “n”.

- $f_a = \text{freqüência absoluta simples acumulada}$. Representa o número de dados até o limite superior do intervalo considerado. Por exemplo, a $f_a$ do segundo intervalo é igual a $f_i$ do primeiro intervalo mais $f_i$ do segundo intervalo ($5 + 11 = 16$). O $f_a$ do terceiro intervalo é $f_i$ do primeiro intervalo + $f_i$ do segundo intervalo + $f_i$ do terceiro intervalo ($5 + 11 + 5$).

- $f_r = \text{freqüência relativa}$. É a $f_i$ do intervalo considerado dividida pelo total da $f_i$ (total da $f_i = 36$). Para o $f_i$ do primeiro intervalo é $5/36 = 0,1389$.

- $f_r\% = \text{freqüência relativa percentual}$. É a $f_r$ do intervalo multiplicado por 100. Para o primeiro intervalo $0,1389 \times 100 = 13,89$.

- $frac\% = \text{freqüência relativa acumulada percentual}$. Representa o valor, em percentual, até o limite superior do intervalo considerado. A $frac\%$ do segundo intervalo, por exemplo, é igual a $f_r\%$ do primeiro intervalo mais $f_r\%$ do segundo intervalo ($13,89 + 30,56 = 44,45$).

- $x_i = \text{ponto médio do intervalo}$. É o valor médio entre o limite inferior e o limite superior do intervalo. Por exemplo: $(903 + 939)/2 = 921$ é o ponto médio do primeiro intervalo.

Interpretação da tabela: temos 11 valores que estão entre 939 e 975, isto é, em 11 (30,56%) meses o rendimento médio foi de R$ 939 a R$ 975. Em 21 (58,34%) meses o rendimento foi no máximo de R$ 1011.

Se ocorrer um intervalo com $f_i = 0$, isto é, um intervalo que não contém valores e no intervalo seguinte ocorrer somente um ou dois valores então suspeita-se que esse ou esses valores são outliers.
Banco (planilha) de dados

Banco de Dados ou planilha de dados: num banco de dados apresentamos as respostas de cada variável (questão do questionário) numa coluna e cada indivíduo numa linha. É a forma de apresentarmos e organizarmos os dados brutos da pesquisa para depois elaborarmos tabelas e gráficos.

Agora, para finalizar esta unidade, vamos exercitar o conteúdo analisando os resultados de uma pesquisa de satisfação realizada, por amostragem, com clientes da empresa Costa (empresa fictícia), em maio de 2008 para construir uma tabela simples da variável sexo, uma tabela cruzada das variáveis sexo e satisfação com o atendimento.

**Banco (planilha) de dados** da pesquisa de satisfação na empresa Costa

<table>
<thead>
<tr>
<th>Cliente</th>
<th>Sexo</th>
<th>Satisfação com atendimento</th>
<th>Satisfação com preço</th>
<th>Nota geral</th>
<th>Idade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>MS</td>
<td>MS</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>S</td>
<td>+-</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>+</td>
<td>+-</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>+</td>
<td>S</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>S</td>
<td>MS</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>+-</td>
<td>+-</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>+-</td>
<td>S</td>
<td>7</td>
<td>37</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>S</td>
<td>I</td>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>I</td>
<td>I</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>S</td>
<td>+-</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>MI</td>
<td>+-</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>14</td>
<td>F</td>
<td>S</td>
<td>+-</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>S</td>
<td>I</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>16</td>
<td>M</td>
<td>MS</td>
<td>S</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>17</td>
<td>F</td>
<td>S</td>
<td>+-</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>51</td>
</tr>
<tr>
<td>19</td>
<td>F</td>
<td>S</td>
<td>S</td>
<td>9</td>
<td>55</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>I</td>
<td>+-</td>
<td>7</td>
<td>45</td>
</tr>
<tr>
<td>21</td>
<td>F</td>
<td>+</td>
<td>S</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td>22</td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>33</td>
</tr>
<tr>
<td>23</td>
<td>M</td>
<td>MS</td>
<td>S</td>
<td>9</td>
<td>34</td>
</tr>
</tbody>
</table>

F = feminino, M = masculino, MS = Muito satisfeito, S = satisfeito, +- = mais ou menos satisfeito, I = insatisfeito, MI = muito insatisfeito.

Para construir uma tabela simples para a variável sexo precisamos contabilizar quantos “F” e quantos “M” ocorreram na pesquisa. Observamos que existem 16 “F” correspondendo ao sexo feminino e 7 “M” correspondendo ao sexo masculino. Após, apresentamos esses valores na tabela simples seguindo as normas na seção 2.1, como segue:
• Para elaborar o título, responder as 3 questões:

• O que: sexo dos clientes

• Onde: empresa Costa

• Quando: maio de 2008

Então o título e o corpo da tabela ficam assim:


<table>
<thead>
<tr>
<th>Sexo</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feminino</td>
<td>16</td>
<td>69,57</td>
</tr>
<tr>
<td>Masculino</td>
<td>7</td>
<td>30,43</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>100,00</td>
</tr>
</tbody>
</table>

Para calcular o percentual do sexo feminino: \(\frac{16}{23} \times 100 = 69,57\%\), e para o sexo masculino: \(\frac{7}{23} \times 100 = 30,43\%\), lembrando que o total do percentual sempre deve ser = 100%.

Para construir uma tabela cruzada da satisfação com atendimento em relação ao sexo precisamos contabilizar quantos “F”, e quantos “M” temos como muito satisfeito (MS) e da mesma forma precisamos contabilizar quantos “F” e quantos “M” temos como satisfeito (S), e assim sucessivamente.

Observando a planilha de dados percebemos que temos dois “F” com “MS”, temos cinco “M” com “S”, e assim por diante. Após apresentamos esses valores na tabela cruzada seguindo as normas da seção 2.2, como segue:

• Para elaborar o título, responder as 3 questões:

• O que: satisfação com o atendimento em relação ao sexo dos clientes

• Onde: empresa Costa

• Quando: maio de 2008
Então o título e o corpo da tabela ficam assim:


<table>
<thead>
<tr>
<th>Satisfação</th>
<th>Sexo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>MS</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>S</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>+/-</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MI</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>7</td>
</tr>
</tbody>
</table>

Note que o total de “F” é o número de clientes do sexo feminino e “M” é o número de clientes do sexo masculino, conforme já apresentado na tabela simples anterior.

**SÍNTESE DA UNIDADE 2**

Nesta unidade vimos como organizar os dados numa tabela simples, a qual é usada para apresentar os valores de uma variável aleatória. Também aprendemos como apresentar os valores de duas variáveis aleatórias numa tabela cruzada. E, ainda, quando temos uma variável quantitativa com \( n \geq 20 \) utilizamos uma tabela de distribuição de frequência.
Unidade 3

Gráficos

Objetivos Desta Unidade

- Verificar como apresentamos os dados em um gráfico.
- Verificar como escolhemos o gráfico mais adequado para cada situação.
- Interpretar os resultados de um gráfico.

As Seções Desta Unidade

Seção 3.1 – Gráfico para representar variáveis qualitativas e quantitativas discretas

Seção 3.2 – Gráfico para representar uma série temporal

Seção 3.3 – Gráfico para representar duas variáveis quantitativas

Seção 3.4 – Gráfico para representar uma distribuição de frequência

Seção 3.5 – Outros tipos de gráficos

Os gráficos são figuras que empregamos para apresentar os dados. Eles têm a vantagem de serem mais atrativos que as tabelas e muitas vezes facilitam a visualização do comportamento dos dados. Existem diferentes tipos de gráficos adequados para diferentes situações, como apresentamos em cinco seções.

Para melhor visualização dos gráficos apresentados nesta unidade, consulte a apostila em meio virtual, disponibilizada na biblioteca do Conecta.
Seção 3.1

Gráficos para Representar Variáveis Qualitativas e Quantitativas Discretas

As variáveis qualitativas nominais ou ordinais, as variáveis quantitativas discretas com pouca frequência de diferentes respostas e as variáveis quantitativas continuas com menos de 20 dados podem ser representadas nos tipos de gráficos apresentados nesta seção.

3.1.1 GRÁFICO DE SETORES

É utilizado quando desejamos representar partes do todo de uma variável. Somente podemos utilizá-lo quando temos um totalizador de 100%; é indicado quando a variável é qualitativa nominal e tem no máximo 5 categorias. Cada categoria é representada por uma cor e as frequências são proporcionais ao ângulo da circunferência.

Gráfico 1: Transferências Voluntárias da União para as Regiões Geográficas – 2007, Brasil

Interpretação do gráfico: observamos que as regiões Sudeste e Nordeste obtiveram 2/3 das transferências voluntárias da União.
3.1.2 GRÁFICO DE COLUNAS SIMPLES

Representa uma variável, é indicado para comparação de categorias de variáveis qualitativas ou quantitativas, nesta com menos de 20 categorias. As categorias são apresentadas no eixo horizontal e as freqüências no eixo vertical, e todas as colunas possuem a mesma cor.

É muito importante a definição da escala em um gráfico para que a informação seja repassada com fidelidade.

Gráfico 2: Transferências voluntárias da União para as regiões geográficas por habitante, junho/2008, Brasil


Interpretação do gráfico: constata-se que as regiões Sudeste e Sul receberam o menor valor por habitante no mês de junho de 2008.
3.1.3 GRÁFICO DE BARRAS SIMPLES

É uma variação do gráfico de colunas simples, porém os eixos são invertidos.

Gráfico 3: Transferências voluntárias da União para as regiões geográficas por habitante, junho/2008, Brasil

3.1.4 GRÁFICO DE COLUNAS AGRUPADAS

É usado para representar duas variáveis num mesmo gráfico. As respostas de uma variável são apresentadas no eixo horizontal. Utiliza-se uma legenda para especificar as categorias da outra variável, as quais são apresentadas com cores diferentes.

Gráfico 4: Domicílios com bens duráveis (%), 2005 e 2006, Brasil
Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2005-2006.
Interpretação do gráfico: observamos que mais de 80% dos domicílios possuem fogão, geladeira, rádio e televisão, e que aumentou o nº de domicílios com geladeira e televisão entre 2005 e 2006.

3.1.5 GRÁFICO DE BARRAS AGRUPADAS

É uma variação do gráfico de colunas agrupadas, com inversão dos eixos.

Gráfico 5: Domicílios com bens duráveis (%), 2005 e 2006, Brasil
Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2005-2006.

Seção 3.2

Gráfico para Representar uma Série Temporal – Gráfico de Linha

O gráfico de linha é empregado para representar uma série temporal, ou seja, uma escala de tempo em que o evento será mostrado, sendo que no eixo horizontal sempre é apresentado o tempo, com escala proporcional, e no eixo vertical os valores referentes a cada tempo.
Gráfico 6: Taxa de mortalidade infantil (por mil nascidos vivos), 2000 a 2007, Brasil
Fonte: IBGE, Diretoria de Pesquisas, Coordenação de População e Indicadores Sociais.

Interpretação do gráfico: observamos que a taxa de mortalidade infantil decresceu entre 2000 e 2007.

Seção 3.3
Gráfico para Representar duas Variáveis Quantitativas — Gráfico de Dispersão

Utiliza-se o gráfico de dispersão quando temos duas variáveis quantitativas. Representamos uma variável em cada eixo, como no Gráfico 7. Os valores da variável altura são representados no eixo horizontal e os valores da variável peso são representados no eixo vertical. Com este gráfico é possível visualizar a existência de uma relação entre as duas variáveis.

Gráfico 7: Distribuição do peso em relação a altura dos alunos da turma X, março/2005, Unijuí
Fonte: Construção dos autores.
Interpretação do gráfico: Observamos uma relação linear direta entre a altura e o peso, isto é, quanto maior a altura, em média, também é o peso.

Seção 3.4

Gráfico para Representar uma Distribuição de Frequência

Na seção 2.3 vimos como organizar os dados numa tabela de distribuição de frequência. Se desejamos organizar os mesmos dados em um gráfico precisamos utilizar os gráficos adequados, os quais são apresentados nesta seção: histograma e polígono de frequência.

Da mesma forma, muitas vezes os dados secundários são apresentados em histogramas e polígonos de frequências. Assim precisamos saber interpretá-los, como veremos na sequência.

3.4.1 HISTOGRAMA

O histograma é o gráfico adequado para apresentar uma distribuição de frequência, consta de colunas verticais agrupadas apresentando no eixo horizontal os valores (limites dos intervalos) da variável quantitativa e no eixo vertical a fi ou fr ou fr%, conforme o exemplo a seguir (retome à seção 2.3 caso não lembre o que é fi, fr, fr%).

Gráfico 8: Rendimento médio nominal do trabalho principal, habitualmente recebido por mês, pelas pessoas de 10 anos ou mais de idade, ocupadas na semana de referência, em Porto Alegre, janeiro/2005 a dezembro/2007.
Nota: Exclusive trabalhadores não-remunerados e trabalhadores que receberam somente em benefícios.

Interpretação do gráfico: observamos que a maior concentração de valores está no intervalo 939 a 975, isto é, em 11 meses o rendimento médio ficou entre R$ 939,00 e R$ 975,00.

3.4.2 POLÍGONO DE FREQUÊNCIA

Como alternativa existe o polígono de freqüências, em que no eixo horizontal são apresentados os pontos médios do intervalo, conforme exemplo a seguir.

Nota: Exclusive trabalhadores não-remunerados e trabalhadores que receberam somente em benefícios.
Seção 3.5

Outros Tipos de Gráficos

Existem outros tipos de gráficos para situações mais específicas, como a pirâmide etária e o cartograma. Os gráficos com figuras são muito utilizados nos meios de comunicação, como publicidade para chamar a atenção do leitor ou consumidor.

3.5.1 PIRÂMIDE ETÁRIA

A pirâmide etária é empregada para verificar a distribuição do sexo e da faixa etária da população de uma cidade, Estado ou um país. A pirâmide etária a seguir representa a população do Brasil em relação ao sexo e em relação à faixa etária de acordo com o censo demográfico de 2000 do IBGE.

Gráfico 10: Pirâmide etária absoluta – 2000
3.5.2 CARTOGRAMA

Os cartogramas são utilizados quando se tem uma série geográfica (dados coletados em diferentes locais), adotando-se mapas para atrair a atenção ou facilitar a visualização. Na figura a seguir são representados os Estados do Brasil com sua respectiva população. Os Estados com cor mais escura são os mais populosos (acima de 10 milhões de habitantes, conforme legenda do gráfico) e os Estados com cor mais clara os menos populosos (abaixo de 2 milhões). Os demais situam-se entre 2 milhões e 8 milhões de habitantes, representados pelos tons intermediários.

Figura 1: Brasil, População por Estado – 2000

SÍNTESE DA UNIDADE 3

Vimos que existem vários tipos de gráficos para representar as variáveis aleatórias. Alguns são mais indicados para variáveis qualitativas e outros são mais apropriados para variáveis quantitativas. Na Unidade 5, veremos como construir esses gráficos em uma planilha eletrônica.
Medidas Descritivas

Objetivos Desta Unidade

• Compreender o cálculo de medidas descritivas.
• Verificar como escolhemos a medida descritiva mais adequada.
• Interpretar os resultados das medidas descritivas.

As Seções Desta Unidade

Seção 4.1 – Medidas de posição
Seção 4.2 – Medidas de dispersão

Até agora vimos como resumir as informações em tabelas e gráficos para facilitar sua visualização e compreensão. Muitas vezes, porém, quando desejamos estudar variáveis quantitativas, utilizamos também outros métodos estatísticos – medidas descritivas – para complementar a análise dos dados, as quais abordaremos nesta unidade.

As medidas descritivas são utilizadas para resumir as variáveis quantitativas em um ou dois valores. Por exemplo: podemos realizar uma pesquisa para conhecer o salário dos habitantes de um município. Obteremos como resultados dessa pesquisa um valor de salário associado a cada habitante. Alguns valores serão repetidos entre os habitantes e outros não. Uma forma de tratar essas informações é pela tabela de distribuição de frequência, já estudada na Unidade 2. Outra forma, complementar a esta, é por intermédio das medidas descritivas, da seguinte maneira: podemos encontrar um valor – medida de posição, que representa o salário médio dos habitantes desse município, calculado pela média aritmética. Além dessa média, é sempre interessante demonstrar como os valores estão distribuídos em torno desta. Para isso encontramos outro valor – medida de dispersão, calculada por meio do desvio-padrão.
Para facilitar a compreensão vamos utilizar um exemplo: supomos que no município A existem 7 habitantes e que seus salários são:

<table>
<thead>
<tr>
<th>Município A:</th>
</tr>
</thead>
<tbody>
<tr>
<td>R$ 780</td>
</tr>
<tr>
<td>R$ 1.200</td>
</tr>
<tr>
<td>R$ 550</td>
</tr>
<tr>
<td>R$ 600</td>
</tr>
<tr>
<td>R$ 1.500</td>
</tr>
<tr>
<td>R$ 750</td>
</tr>
<tr>
<td>R$ 980</td>
</tr>
</tbody>
</table>

Obtemos o valor R$ 908,57 para representar o salário médio desses habitantes pela média aritmética (veremos a seguir como calcular a média aritmética). Observe que alguns habitantes têm salários menores que a média e outros tem salários maiores que a média. Ainda precisamos de outra medida para representar como os dados estão dispostos em torno do valor médio, isto é, se os valores estão mais distantes ou mais próximos da média. Este valor é R$ 342,85, que representa o desvio-padrão (veremos a seguir como calcular o desvio-padrão), revelando uma grande dispersão entre os valores, pois temos habitantes recebendo desde R$ 550,00 até R$ 1.500,00, este último quase três vezes maior que o primeiro.

Vamos considerar o município B com 7 habitantes cujos salários são:

<table>
<thead>
<tr>
<th>Município B:</th>
</tr>
</thead>
<tbody>
<tr>
<td>R$ 908,57</td>
</tr>
<tr>
<td>R$ 908,57</td>
</tr>
<tr>
<td>R$ 908,57</td>
</tr>
<tr>
<td>R$ 908,57</td>
</tr>
<tr>
<td>R$ 908,57</td>
</tr>
<tr>
<td>R$ 908,57</td>
</tr>
<tr>
<td>R$ 908,57</td>
</tr>
</tbody>
</table>

Obtemos o valor R$ 908,57 como o salário médio desses habitantes e o valor R$ 0,00 como desvio-padrão. Observamos que o salário médio dos municípios A e B são iguais, porém os valores associados a cada habitante diferem bastante entre os dois municípios: no município A, o salário entre os habitantes diverge muito, representado pelo alto valor do desvio-padrão. Já no município B os salários entre os habitantes não diferem, representado pelo valor nulo do desvio-padrão.
Agora consideramos o município C, com 7 habitantes e seus respectivos salários:

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R$ 780</td>
<td>R$ 800</td>
<td>R$ 950</td>
<td>R$ 1.100</td>
<td>R$ 930</td>
<td>R$ 950</td>
<td>R$ 850</td>
</tr>
</tbody>
</table>

Observamos que o salário médio no município C é R$ 908,57, igual aos municípios A e B, e o desvio-padrão é R$ 109,76. Os salários no município C têm menos dispersão que no município A, pois o menor é R$ 780,00 e o maior é R$ 1.100,00, representado pelo baixo valor do desvio-padrão.

Consideremos o município D, com 7 habitantes e seus respectivos salários:

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R$ 1.780</td>
<td>R$ 1.800</td>
<td>R$ 1.950</td>
<td>R$ 2.100</td>
<td>R$ 1.930</td>
<td>R$ 1.950</td>
<td>R$ 1.850</td>
</tr>
</tbody>
</table>

O salário médio dos habitantes do município D é R$ 1.908,57, muito mais alto que nos demais municípios. O desvio-padrão dos salários do município D é R$ 109,76, igual ao desvio-padrão do município C, porém proporcionalmente a dispersão dos valores dos salários em torno da média no município D é menor. Neste caso, é interessante utilizar outra medida de dispersão – coeficiente de variação – para comparar a dispersão dos dados entre os municípios C e D, como veremos adiante.

Com esse exemplo percebemos que é importante apresentar uma medida de posição para representar o centro de um conjunto de dados e uma medida de dispersão para representar a disposição dos valores em torno do valor central. Desta forma, existem dois grupos de medidas descritivas para tratar dados de variáveis quantitativas: medidas de posição e medidas de dispersão, como apresentaremos nesta unidade.
Inicialmente precisamos entender a notação (forma de representação) para representar as variáveis: X, Y, Z (letras maiúsculas) e para representar os valores das variáveis: \( x_i, y_i, z_i \) (respectivas letras minúsculas). O operador somatório \( \sum \) (lê-se sigma) representa uma soma de valores de uma variável e \( i \) representa o índice, iniciando em 1 até \( n \), por exemplo:

\[
\sum_{i=1}^{7} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7
\]

No lado esquerdo da igualdade temos o somatório por extenso que pode ser representado pela notação do lado direito. Supomos que a variável que estamos analisando é o salário dos habitantes do município A, utilizando a letra X para representar a variável salário e \( x_1 \) para representar o salário do primeiro habitante, \( x_2 \) para representar o salário do segundo habitante e, assim por diante, até o \( x_7 \) que representa o salário do sétimo habitante. Utilizamos a notação de somatório para representar de forma resumida a soma dos salários dos 7 habitantes, isto é, iniciando em \( i = 1 \) até \( i = 7 \).

### Seção 4.1

**Medidas de Posição**

As medidas de posição, também chamadas de medidas de tendência central, fornecem um valor que representa a posição central do conjunto de dados, com os demais dados dispostos em torno deste. As medidas de posição são: média aritmética, mediana e moda.

#### 4.1.1 MÉDIA ARITMÉTICA

A média aritmética é a medida de posição mais utilizada. Tem como vantagem a facilidade do seu cálculo e como desvantagem de ser muito afetada por valores extremos (valores outliers). Tem a seguinte forma:
\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]

Em que:

\( x_i \) = valores da variável X

\( n \) = número de dados

Exemplo: vamos calcular a média aritmética para o salário dos habitantes do município A:

\[ \bar{x}_1 = \frac{\sum_{i=1}^{5} x_i}{7} = \frac{780+1200+550+600+1500+750+980}{7} = 908,57 \]

Para calcular a média aritmética do conjunto de dados do exemplo somamos todos os valores e dividimos por 7, pois temos 7 valores (\( n = 7 \)) correspondendo aos salários dos 7 habitantes do município.

Interpretação da média: o salário médio dos habitantes do município A é R$ 908,57.

4.1.2 MÉDIA PONDERADA

É a soma dos produtos de cada um dos números por seu peso, dividido pela soma dos pesos, como segue:

\[ \bar{x}_p = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i} \]

Em que:

\( x_i \) = valores da variável X

\( p_i \) = peso associado ao valor \( x_i \)

\( n \) = número de dados
O peso é atribuído pelo pesquisador, de acordo com a contribuição que desejamos que cada valor tenha na média final. A diferença entre a média aritmética e a média ponderada é que na primeira todos os valores contribuem da mesma forma para a média final e na média ponderada atribuímos pesos diferentes que contribuem para a média final.

Exemplo: considere um aluno de Estatística que obteve 9, 6 e 3 nas três avaliações do bimestre, lembrando que cada avaliação tem pontuação máxima de 10 pontos. O peso de cada avaliação é 2, 3 e 5 respectivamente, pois desejamos que a última avaliação contribua com 5 (50%) da nota final do bimestre. Calcule a média ponderada.

\[
\bar{x}_p = \frac{\sum_{i=1}^{3} x_i p_i}{\sum_{i=1}^{3} p_i} = \frac{(9 \times 2)+(6 \times 3)+(3 \times 5)}{2+3+5} = 5,1
\]

Interpretação da média ponderada: a média ponderada das avaliações do bimestre para o aluno é 5,1 pontos.

Caso o aluno tivesse obtido as notas 3, 6 e 9, a média ponderada seria:

\[
\bar{x}_p = \frac{\sum_{i=1}^{3} x_i p_i}{\sum_{i=1}^{3} p_i} = \frac{(3 \times 2)+(6 \times 3)+(9 \times 5)}{2+3+5} = 6,9
\]

Neste caso, observamos que a média ponderada é maior, pois o aluno obteve a nota mais alta na avaliação que contribui com mais peso para a nota média final.

4.1.3 MEDIANA

A mediana é o valor central em relação a um valor mínimo e um valor máximo, precedido e sequido de um mesmo número de dados, isto é, 50% dos dados assumem valores iguais ou menores que o valor da mediana e os outros 50% dos dados assumem valores iguais ou maiores que o valor da mediana, representada por Md.
A divisão do conjunto de dados ordenados pode ser visualizado da seguinte forma:

<table>
<thead>
<tr>
<th>50%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li</td>
<td>Md</td>
</tr>
</tbody>
</table>

Em que:

- **Li** = limite inferior, é o menor valor dos dados
- **Ls** = limite superior, é o maior valor dos dados
- **Md** = mediana

O **procedimento** para calcular a mediana é o seguinte:

1. Ordenar os dados em ordem crescente

2. Calcular a posição da mediana

   \[ P(Md) = \frac{n+1}{2} \]

3. Encontrar o valor localizado nesta posição, que corresponderá à mediana, considerando que:

   - se o \( n \) é ímpar a mediana é o valor central do conjunto de dados ordenados;
   - se o \( n \) é par a mediana é a média dos dois valores centrais do conjunto de dados ordenados.

Exemplo: calcule a mediana para o salário dos habitantes do município A: R$ 780,00, R$ 1.200,00, R$ 550,00, R$ 600,00, R$ 1.500,00, R$ 750,00, R$ 980,00 (número de dados ímpar).

1. Ordenar: 550, 600, 750, 780, 980, 1.200, 1.500

2. Calcular: \[ P(Md) = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4 \], isto é, a quarta posição

3. O valor que ocupa a quarta posição é o 780, que representa a mediana.
Interpretação da mediana: podemos dizer que 50% dos habitantes no município A têm um salário no máximo igual a R$ 780,00 e os outros 50% recebem no mínimo R$ 780,00.

Se o número de dados é par, por exemplo: R$ 780,00, R$ 1.200,00, R$ 550,00, R$ 600,00, R$ 1.500,00, R$ 750,00, R$ 980,00, R$ 950,00.

1. Ordenar: 550, 600, 750, 780, 950, 980, 1.200, 1.500

2. Calcular: $P(Md) = \frac{n+1}{2} = \frac{8+1}{2} = \frac{9}{2} = 4.5$, isto é, a mediana é o valor que está entre a quarta posição e a quinta posição, neste exemplo entre 780 e 950. Desta forma a mediana será calculada pela média aritmética destes dois valores como: $\frac{780 + 950}{2} = 865$.

3. A mediana é igual a R$ 865,00.

A mediana tem a vantagem de não ser afetada por valores outliers. Por exemplo: suponhamos que os habitantes do município E recebem os mesmos salários que os habitantes do município A, com diferença em apenas um salário: no município A = R$ 1.500,00 e no município E = R$ 4.500,00. As medidas para os dois municípios são:

<table>
<thead>
<tr>
<th>Município</th>
<th>Média</th>
<th>Mediana</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>908,57</td>
<td>780</td>
</tr>
<tr>
<td>E</td>
<td>1.337,14</td>
<td>780</td>
</tr>
</tbody>
</table>

Observamos que a diferença ocorre somente na média, pois esta foi afetada pelo alto valor (outlier) de um salário no município E, mas a mediana não diferiu entre os dois municípios, pois para calculá-la somente consideramos a posição dos valores e não a magnitude dos valores. Assim, optamos pela mediana quando temos valor outlier, pois esta irá representar melhor o centro do conjunto de dados do que a média, como podemos observar nas figuras a seguir:
4.1.4 MODA

É o valor que mais aparece no conjunto de dados, isto é, utilizamos para representar o valor típico de um conjunto de dados. É representada por Mo. Utilizamos mais frequentemente em variáveis quantitativas discretas.

Exemplos:

1) Calcular a moda para os salários dos habitantes do município A: R$ 780,00, R$ 1.200,00, R$ 550,00, R$ 600,00, R$ 1.500,00, R$ 750,00, R$ 980,00.

   Moda = não existe, pois nenhum valor se repete.

2) Calcular a moda para os salários dos habitantes do município F: R$ 780,00, R$ 1.200,00, R$ 550,00, R$ 780,00, R$ 1.500,00, R$ 750,00, R$ 980,00.

   Moda = 780, pois é o valor que mais se repete.

3) Calcular a moda para os salários dos habitantes do município G: R$ 780,00, R$ 1.200,00, R$ 550,00, R$ 780,00, R$ 1.200,00, R$ 750,00, R$ 980,00.

   Moda = 780,00 e 1.200,00, pois são os valores que mais se repetem.
4) Calcular a moda para os salários dos habitantes do município G: R$ 780,00, R$ 1.200,00, R$ 550,00, R$ 780,00, R$ 1.200,00, R$ 750,00, R$ 780,00.

Moda = 780, pois é o valor que mais se repete.

**Seção 4.2**

**Medidas de Dispersão**

As medidas de dispersão, também chamadas de variabilidade, fornecem um valor que quantifica a distância dos valores em torno do valor central, ou seja, são utilizadas para verificar se existe grande ou pequena variabilidade de valores no conjunto de dados. As medidas de dispersão são: variância, desvio-padrão e coeficiente de variação.

### 4.2.1 VARIÂNCIA

É a média dos desvios quadráticos de cada valor em relação à média. A variância amostral é dada por:

\[
S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}
\]

ou

\[
S^2 = \frac{\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2}{n - 1}
\]

Em que:

- \(x_i\) = valores de variável \(x\)
- \(n\) = número de dados
- \(\bar{x}\) = média aritmética

"n-1" no denominador quando \(n<30\) e "n" no denominador quando \(n\geq30\).
Utilizaremos como exemplo os salários dos habitantes do município A (R$ 780,00, R$ 1.200,00, R$ 550,00, R$ 600,00, R$ 1.500,00, R$ 750,00, R$ 980,00) para entendermos o cálculo da variância com a primeira fórmula anteriormente apresentada. Lembrando que $x_i$ representa o salário de cada habitante e $\bar{x}$ representa o salário médio.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(780 - 908,57)^2 + (1200 - 908,57)^2 + (550 - 908,57)^2 + (600 - 908,57)^2 + (1500 - 908,57)^2 + (750 - 908,57)^2 + (980 - 908,57)^2}{6} = \frac{16530,24 + 84931,44 + 128572,44 + 95215,44 + 349789,44 + 25144,44 + 5102,24}{6} = \frac{705285,68}{6} = 117547,61$$

Como a diferença $(x_i - \bar{x})$ é elevada ao quadrado, o resultado também será quadrático, dificultando a utilização dessa medida para representar a dispersão dos dados. Assim, utilizamos o desvio-padrão, visualizado a seguir:

### 4.2.2 DESvio-PaDRÃO

É a raiz quadrada da variância.

$$S = \sqrt{S^2}$$

Para o exemplo: $S = \sqrt{117547,61} = 342,85$.

Interpretamos que a média do conjunto de dados é R$ 908,57 com um desvio-padrão para baixo e para cima de R$ 342,85, isto é, aproximadamente 68% dos dados estão entre a média e um desvio padrão para baixo (média – 1 desvio padrão) e a média e um desvio-padrão para cima (média + 1 desvio padrão).

Assim, utilizamos o desvio-padrão acompanhando a média, pois esta fornece um valor central e o desvio-padrão a dispersão em torno desse valor central.
4.2.3 COEFICIENTE DE VARIAÇÃO

É a divisão do desvio-padrão pela média multiplicado por 100. É sempre dado em percentual. O coeficiente de variação fornece a dispersão dos dados em torno da média em percentual, constituindo uma medida alternativa ao desvio-padrão. Quando se deseja comparar a variabilidade entre dois conjuntos de dados, o coeficiente de variação é a medida de dispersão indicada.

\[ CV = \frac{S}{x} \times 100 \]

Para o exemplo dado o coeficiente de variação é calculado como:

\[ CV = \frac{342.85}{908.57} \times 100 = 37.7\% \]

De forma geral, dizemos que o conjunto de dados é homogêneo em torno da média quando o coeficiente de variação é menor que 30% (CV ≤ 30%) e dizemos que o conjunto de dados é heterogêneo em torno da média quando o coeficiente de variação é maior ou igual a 30% (CV ≥ 30%). Em áreas específicas, este ponto de corte pode diferir.

Interpretação do coeficiente de variação: tem-se uma variabilidade média de 37,7% para mais e para menos do salário médio dos habitantes do município A.

Agora, vamos comparar a dispersão dos valores dos salários entre os municípios C e D:

<table>
<thead>
<tr>
<th>Município</th>
<th>Média</th>
<th>Desvio-padrão</th>
<th>Coeficiente de variação</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>R$ 908,57</td>
<td>R$ 109,76</td>
<td>12,08%</td>
</tr>
<tr>
<td>D</td>
<td>R$ 1.908,57</td>
<td>R$ 109,76</td>
<td>5,75%</td>
</tr>
</tbody>
</table>

Observamos que o desvio-padrão dos dois municípios é igual, porém a dispersão em torno da média é diferente entre os dois municípios, como percebemos pelo coeficiente de variação. No município A é igual a 12,08% e no município B é igual a 5,75%, pois neste último a média é bem maior e então o desvio-padrão em percentual representa bem menos.
Salienta-se a importância de apresentar a unidade de medida nos resultados das medidas estatísticas, isto é, se estamos analisando o salário dos habitantes de um município, colocamos sempre a unidade de medida do salário, que no exemplo exposto foi R$. Se estamos trabalhando com a idade dos habitantes de um município, colocamos sempre a unidade de medida da idade, por exemplo, anos ou meses.

Para concluir esta unidade vamos exercitar o cálculo das medidas descritivas utilizando a nota geral da pesquisa de satisfação na empresa Costa apresentada no final da Unidade 2.

**Banco (planilha) de dados da pesquisa de satisfação na empresa Costa**

<table>
<thead>
<tr>
<th>Cliente</th>
<th>Sexo</th>
<th>Satisfação com atendimento</th>
<th>Satisfação com preço</th>
<th>Nota geral</th>
<th>Idade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>MS</td>
<td>MS</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>S</td>
<td>+–</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>+–</td>
<td>S</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>+–</td>
<td>MS</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>+–</td>
<td>S</td>
<td>7</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>S</td>
<td>I</td>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>I</td>
<td>I</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>S</td>
<td>+–</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>M</td>
<td>+–</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>S</td>
<td>+–</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>S</td>
<td>I</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>MS</td>
<td>S</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>S</td>
<td>+–</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>51</td>
</tr>
<tr>
<td>17</td>
<td>M</td>
<td>I</td>
<td>+–</td>
<td>7</td>
<td>45</td>
</tr>
<tr>
<td>18</td>
<td>F</td>
<td>+–</td>
<td>S</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td>19</td>
<td>M</td>
<td>S</td>
<td>S</td>
<td>8</td>
<td>33</td>
</tr>
<tr>
<td>20</td>
<td>M</td>
<td>MS</td>
<td>S</td>
<td>9</td>
<td>34</td>
</tr>
</tbody>
</table>

F = feminino, M = masculino, MS = Muito satisfeito, S = satisfeito, +– = mais ou menos satisfeito, I = insatisfeito, MI = muito insatisfeito.

- Média aritmética:

\[
\bar{x} = \frac{10 + 7 + 6 + 7 + 8 + 9 + 7 + 7 + 5 + 5 + 8 + 8 + 9 + 9 + 7 + 10 + 8 + 8 + 9 + 7 + 7 + 8 + 9}{23}
\]

\[
\frac{178}{23} = 7,74
\]
MÉTODOS ESTATÍSTICOS

- Mediana:

1. Ordenar: 5, 5, 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10

2. Calcular a posição da mediana: \( P(Md) = \frac{n + 1}{2} = \frac{23 + 1}{2} = \frac{24}{2} = 12 \), isto é, a mediana está na 12ª posição dos dados ordenados.

3. O valor que está na 12ª posição é 8, então a mediana = 8.

- Moda: o valor que mais se repete é o número 7, portanto Moda = 7.

- Variância:

\[
S^2 = \frac{(10-7,74)^2 + (7-7,74)^2 + (6-7,74)^2 + \ldots + (9-7,74)^2}{23-1} = 1,8379
\]

- Desvio-padrão: \( S = \sqrt{1,8379} = 1,3557 \)

- Coeficiente de variação: \( CV = \frac{1,3557}{7,74} \times 100 = 17,52\% \)

Interpretação dos resultados: a nota geral média dada pelos clientes da empresa Costa é 7,74 pontos com desvio-padrão de 1,3557 pontos para mais e para menos da média. O conjunto de dados é homogêneo em torno da média, uma vez que o CV é menor que 30%. Metade (50%) dos clientes forneceu uma nota no máximo igual a 8 pontos, dado que a mediana é 8. A nota que mais se repetiu entre os clientes foi 7 pontos, já que a moda é 7.

SÍNTESE DA UNIDADE 4

Nesta unidade vimos como calcular medidas descritivas para variáveis quantitativas.

As medidas descritivas são muito importantes para descrever os dados, fornecendo informações que podem auxiliar no acompanhamento de indicadores socioeconômicos ao longo dos anos num município ou numa organização, auxiliando na administração e na tomada de decisão.

Na próxima unidade veremos como obter essas medidas no Excel.
Unidade 5

Ferramentas de Análise Estatística no Excel

Objetivos Desta Unidade

• Verificar como construir banco de dados no Excel.
• Verificar como elaborar tabelas e gráficos no Excel.
• Verificar como obter medidas descritivas no Excel.

As Seções Desta Unidade

Seção 5.1 – Elaborando um banco de dados no Excel
Seção 5.2 – Utilizando as técnicas estatísticas para analisar os dados no Excel
Seção 5.3 – Medidas descritivas para as variáveis quantitativas
Seção 5.4 – Tabelas simples para as variáveis qualitativas e para as variáveis quantitativas com pouca variabilidade
Seção 5.5 – Tabelas cruzadas
Seção 5.6 – Gráficos de setores, colunas e barras


O MS Excel é um aplicativo do tipo Planilha Eletrônica, ou seja, é um software apropriado para a edição de documentos com características numéricas. Usando uma planilha eletrônica podemos manipular com facilidade a edição de tabelas, efetuar cálculos, operações estatísticas e a geração de gráficos. De forma geral permite armazenar dados e produzir informações a partir destes.

Figura 1: Tela inicial do Excel

Uma planilha é um conjunto de linhas e colunas, como uma grade. No Excel, as colunas são identificadas por letras em ordem alfabética, e as linhas por números em ordem crescente. Uma célula é a menor unidade de trabalho do Excel, identificada pelo cruzamento de uma linha e uma coluna.

Cada célula tem um endereço próprio, formado pela combinação da letra da coluna com o número da linha. Por exemplo, a célula D8 localiza-se no cruzamento da coluna D com a linha 8. Para selecionar uma célula, colocando-a em foco, utiliza-se o cursor, que na planilha aparece como um ponteiro semelhante a uma cruz. Basta clicar na célula desejada.

Um intervalo de células é a representação de uma faixa ou conjunto de células. Pode ser selecionado clicando com o cursor sobre um dos cantos do intervalo e arrastando até o canto oposto, formando um retângulo. Pode também ser apenas uma coluna ou linha. Para selecionar uma coluna inteira deve-se clicar na respectiva letra, no cabeçalho da coluna.

Para selecionar uma linha inteira deve-se clicar no número identificador da linha. O intervalo é identificado por suas coordenadas, sendo que o primeiro elemento das coordenadas é o endereço da célula superior esquerda, e o segundo elemento das coord-
denadas é o endereço da célula inferior direita do intervalo. Os dois elementos são separados por dois pontos (:) Por exemplo, o intervalo C4:E5 compreende as células C4, C5, D4, D5, E4 e E5.

Seção 5.1

Elaborando um Banco de Dados no Excel

Para criar um banco de dados no Excel, destinamos as linhas aos dados de cada elemento da amostra e as colunas a cada uma das variáveis. A Figura 2 apresenta um exemplo de um banco de dados na planilha.

Figura 2: Exemplo de um banco de dados
Seção 5.2

Utilizando as Técnicas Estatísticas para Analisar os Dados no Excel


**No Excel 2003:**

Clique no menu Ferramentas, depois em Suplementos e então marque a opção Ferramentas de Análise e clique no botão ok conforme a Figura 3 (esquerda). De agora em diante, sempre que você desejar utilizar o comando análise de dados, este estará disponível no menu Ferramentas.

**No Excel 2007:**

Clique no botão do Office , depois no botão opções do Excel e no menu. Escolha a opção Suplementos e então, na lista de suplementos de aplicativos inativos, a opção Ferramentas de Análise. Após clique no botão ir

Então aparecerá a janela da Figura 1 (direita). Nesta marque Ferramentas de Análise e clique no botão ok. De agora em diante, sempre que você desejar utilizar o comando análise de dados, este estará disponível no menu Dados.

![Figura 3: Instalação das Ferramentas de Análise no Excel 2003 (esquerda) e no Excel 2007 (direita)](image-url)
Seção 5.3

Medidas Descritivas para as Variáveis Quantitativas

Para obter as medidas descritivas, você clica na opção análise de dados no menu Ferramentas no Excel 2003 ou clica no botão análise de dados no menu Dados no Excel 2007 e então aparecerá a janela da Figura 4.

Figura 4: Ferramentas de Análise

Nesta janela você clica na opção Estatística descritiva. Após, aparecerá a janela da Figura 5. Você deve então seguir os passos:

• selecione os dados da coluna que você deseja analisar, por exemplo, coluna E, que se refere à variável idade;

• marque a opção rótulos na primeira linha, pois foi selecionado junto com os valores o nome (rótulo) da variável;

• marque nova planilha, assim o resultado do comando irá aparecer numa nova planilha da pasta de trabalho;

• marque resumo estatístico;

• clique no botão ok.
Figura 5: Estatística descritiva

Uma nova planilha é criada com os resultados do comando, conforme Figura 6.

Para calcular o coeficiente de variação deve-se incluir uma fórmula da seguinte maneira:

- clique numa célula vazia do Excel; sugestão na célula B16;

- digite =;
• clique no valor do desvio-padrão;

• digite /;

• clique no valor da média;

• digite *;

• digite 100;

• tecle enter.

A fórmula deste exemplo é: \[ \frac{B7/B3*100}{100} \]

Neste exemplo o valor do coeficiente de variação é 29,59%.

Interpretação dos resultados: A idade média dos indivíduos pesquisados é de 33,83 anos, com desvio-padrão de 10 anos. Dos 23 indivíduos avaliados, 50% deles têm idade até 35 anos e os outros 50% dos indivíduos têm idade igual ou maior a 35 anos. A menor idade desta amostra é 19 anos e a maior é 55 anos. Tem-se um conjunto de dados homogêneos, dado que o coeficiente de variação assume valor menor que 30%.

Seção 5.4

Tabelas Simples Para as Variáveis Qualitativas e Para as Variáveis Quantitativas com Pouca Variabilidade

Para elaborar uma tabela simples a partir de um banco de dados utilize o comando relatório de tabela dinâmica, sendo este disponibilizado diferentemente no Excel 2003 e no Excel 2007.

No Excel 2003:

Escolha a opção Relatório de tabela e gráfico dinâmico no menu Dados, sendo apresentada a tela da Figura 7.
A opção de tabela dinâmica já está marcada. Caso deseje um gráfico, você deve selecionar a segunda opção, conforme Figura 7.

Após, clique no botão avançar e então a janela da Figura 8 será apresentada.

Caso o banco de dados (informações já digitadas na planilha) não esteja selecionado, selecione com o mouse. Novamente clique no botão avançar e na janela da Figura 9, clique no botão layout.
Neste momento você escolhe o modo de apresentação da tabela (layout), apresentada na Figura 10. Para escolher o local de apresentação das categorias da variável, você deve arrastar o botão referente à variável apresentado à direita da tela até o corpo da tabela sobre LINHA. Depois, arraste novamente o botão da variável para o centro do corpo da tabela para calcular frequência absoluta (n), ou seja, contar as categorias. Deve aparecer no botão: Contar de nome_da_variável. Caso isso não ocorra, clique duplo sobre o botão e escolha a opção ContNúm.

Para calcular a frequência relativa percentual, você deve novamente arrastar o botão da variável até o centro do corpo da tabela e clicar duplo sobre o botão, sendo apresentada a tela da Figura 11. Clique em Opções e no campo Mostrar dados como, selecione a opção % do total e clique no botão ok.
Clique no botão ok e no botão concluir e então aparecerá a tabela simples dinâmica da Figura 12.

Para formatar a tabela clique no botão Formatar Relatório na barra de ferramentas do relatório e escolher o Relatório 7.

No cabeçalho da tabela substituímos contar-de-sexo1 por “n” e contar-de-sexo2 por %, assim obtemos a tabela 1 (Figura 18).

No Excel 2007:

Clique no menu Inserir e na opção Tabela dinâmica e aparecerá a janela da Figura 13. Se o banco de dados (informações já digitadas na planilha) não estiver selecionado você deve selecioná-lo com o mouse e após clicar no botão ok.
Após o ok, aparecerá a janela da Figura 14.

Figura 13: Seleção de banco de dados no Excel 2007

Figura 14: Elaboração de tabela dinâmica no Excel 2007
Marque a variável de interesse no menu à direita, neste exemplo marque a variável sexo e então arraste (clique com o mouse e arraste com o botão clicado) para rótulos de linha. Arraste-a também para valores, duas vezes. Deve aparecer ContNúm em cada botão de valores. Caso não estiver aparecendo clique com o mouse sobre o primeiro botão e escolha a opção configurações do campo valor. Aparecerá a janela da Figura 15.

![Figura 15: Marcando a opção ContNúm](image)

Na janela da 15, marque a opção ContNúm e clique no botão ok, assim aparecerá o valor absoluto na tabela simples (segunda coluna). Clique sobre o segundo botão e escolha a opção configurações do campo valor.

Aparecerá a janela da Figura 16.
Figura 16: Marcando a opção mostrar valores como % do total

Marque a opção ContNúm e clique na ficha mostrar valores como e escolha a opção % do total, conforme 16. Assim aparecerá o valor relativo percentual na tabela simples (terceira coluna). A tabela para este exemplo está apresentada na Figura 17.

Figura 17: Exemplo de tabela simples

<table>
<thead>
<tr>
<th></th>
<th>Valores</th>
<th></th>
<th>Contar de sexo 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>F</td>
<td>16</td>
<td>69.57%</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>7</td>
<td>30.43%</td>
</tr>
<tr>
<td>7</td>
<td>Total geral</td>
<td>23</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Figura 17: Exemplo de tabela simples

No cabeçalho da tabela substituímos contar de sexo por n e contar de sexo 2 por %, assim obtém-se a tabela 1 (Figura 18).

Tabela 1 - Distribuição do sexo dos indivíduos

<table>
<thead>
<tr>
<th>Sexo</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>16</td>
<td>69.57%</td>
</tr>
<tr>
<td>M</td>
<td>7</td>
<td>30.43%</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Figura 18: Exemplo de tabela simples formatada
Interpretação: 69,57% dos indivíduos entrevistados são do sexo feminino e 30,43% dos indivíduos entrevistados são do sexo masculino.

Seção 5.5

**Tabelas Cruzadas**

O processo para a construção de tabelas cruzadas dinâmicas é semelhante à construção de tabelas simples. A seguir apresenta-se o procedimento para o Excel 2003 e Excel 2007.

**No Excel 2003:**

Selezione mais uma variável para o corpo da tabela conforme a Figura 19, colocando-a sobre a COLUNA. Finalizando todo o processo aparecerá a tabela cruzada conforme Figura 20.

![Figura 19: Elaboração da tabela cruzada](image)

![Figura 20: Exemplo de tabela cruzada dinâmica](image)
No Excel 2007:

Selecione mais uma variável conforme a Figura 21 para rótulos de coluna

Finalizando todo o processo aparecerá a tabela cruzada, conforme Figura 22.

Figura 21: Elaboração da tabela cruzada

Figura 22: Exemplo de tabela cruzada dinâmica
Na Figura 23 está apresentada uma tabela cruzada formatada.

<table>
<thead>
<tr>
<th>Satisfação atend</th>
<th>Sexo</th>
<th>F</th>
<th>M</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>---</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>MII</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>9</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>16</td>
<td>7</td>
<td>23</td>
</tr>
</tbody>
</table>

Figura 23: Exemplo de tabela cruzada formatada

### Seção 5.6

**Gráficos de Setores, Colunas e Barras**

Você elabora um gráfico a partir de dados sistematizados (a partir de uma tabela já elaborada) de forma diferente no Excel 2003 e no Excel 2007. Existem variados tipos de gráficos, adequados para cada tipo de variável e representam uma ou duas variáveis. A seguir são apresentados os tipos de gráficos mais usados.

**Gráfico de setores (pizza)**

É usado para representar uma variável aleatória. Suas categorias devem totalizar 100% e é mais indicado quando o número de categorias são no máximo 5. Primeiramente, você seleciona as categorias e os valores absolutos ou relativos que serão apresentados no gráfico, a partir da tabela já elaborada, conforme Figura 24.

Figura 24: Seleccionando os dados para elaborar um gráfico
**No Excel 2003:**


![Assistente de gráfico - etapa 1 do 4 - tipo de gráfico](image)

**Figura 25:** Gráfico de setores – etapa 1

A Figura 26 mostra a segunda etapa, na qual é apresentada uma prévia do gráfico e permite selecionar novamente os dados, se necessário.
Na etapa 3 (Figura 27), clique no campo Título do gráfico e digite o título do gráfico.

Figura 27: Gráfico de setores – etapa 3, título do gráfico
Após, clique na ficha Legenda para alterar o local de posição da legenda ou para excluír legenda (Figura 28).

![Figura 28: Gráfico de setores – etapa 3, legenda do gráfico](image)

Na ficha Rótulos de dados (Figura 29) é possível mostrar o nome das categorias e a percentagem no gráfico. Após, clique no botão Concluir.

![Figura 29: Gráfico de setores – etapa 3, rótulos de dados](image)

Podemos personalizar o gráfico, alterando a cor dos setores, alterando a cor do fundo, alterando o tamanho, tipo e cor da fonte.
Para alterar a cor dos setores, você clica uma vez no gráfico, após clica mais uma vez no setor (fatia) cuja cor deseja alterar e após clica duplo para aparecer o menu de cores. Então, selecione uma cor ou no botão Efeitos de Preenchimento para utilizar as opções gradiente ou textura. Na Figura 30 está apresentado um exemplo de gráfico formatado.

Figura 30: Gráfico de setores – exemplo

Utilizamos um gráfico de colunas simples ou barras simples para representar uma variável aleatória. O procedimento é o mesmo utilizado no gráfico de setores, sendo que na etapa 3 (Figura 31) digitamos, além do título do gráfico, o título do eixo x (horizontal) e o título do eixo y (vertical).
O gráfico de colunas finalizado é apresentado na Figura 32.

![Gráfico de colunas - exemplo](image1)

Figura 32: Gráfico de colunas – exemplo

De forma similar podemos construir um gráfico de barras, como o apresentado na Figura 33.

![Gráfico de barras - exemplo](image2)

Figura 33: Gráfico de barras – exemplo

**No Excel 2007:**

Para construir um gráfico de setores (pizza) no Excel 2007 utilize o menu Inserir e escolha gráfico de pizza, conforme a Figura 34.
Figura 34 – Elaboração de gráfico

No menu do gráfico de pizza escolha um tipo e o gráfico será apresentado como na Figura 35.

Figura 35: Exemplo de gráfico de setores

Clique com o mouse na área do gráfico (parte branca dentro do retângulo) e clique na opção Layout 1 no menu Design, conforme Figura 36.

Figura 36: Formatando o gráfico de setores
O gráfico final ficará conforme a Figura 37.

Você pode clicar com o mouse sobre o Título do Gráfico para editá-lo e então digitar o título para o seu gráfico. Também pode mudar o estilo do gráfico clicando no menu Design e então em Estilo (Figura 38).
Para elaborar um gráfico de colunas o procedimento é semelhante: primeiro selecione os dados na tabela (categorias e valores absolutos ou percentuais) e depois clique no menu Inserir e escolha Colunas, aparecerá o gráfico da Figura 39.

![Figura 39: Exemplo de gráfico de colunas](image)

Você pode clicar na área do gráfico (parte branca dentro do retângulo) e usar as opção do menu Design para alterar definições do gráfico: em Layout do gráfico você pode incluir o título do gráfico e os valores. Clicando com o mouse no título é possível editar e colocar o título para o seu gráfico, conforme Figura 40.

![Figura 40: Utilizando a opção design do gráfico para gráfico de colunas](image)
Para o desenvolvimento de um gráfico de barras o procedimento a ser seguido é o mesmo descrito para o gráfico de colunas, mudando apenas a escolha do tipo de gráfico no início do procedimento.

**SÍNTESE DA UNIDADE 5**

Nesta unidade utilizamos os recursos estatísticos do Excel para tratar as informações.

A aplicação dos métodos estatísticos é favorecida pela informática, no que diz respeito a hardware e software, uma vez que comumente necessitamos trabalhar com grande quantidade de dados.

Assim, o conhecimento da utilização de um software estatístico ou um software que fornece técnicas estatísticas é essencial para alunos ou profissionais que estão iniciando seu aprendizado em Métodos Estatísticos.

A planilha eletrônica Excel é uma boa alternativa para alunos que estão iniciando seus estudos de Métodos Estatísticos, por ser de fácil acesso, uma vez que grande parte dos usuários domésticos e organizações possuem o pacote Microsoft Office.
Referências


