

# Análise de dados: uma leitura crítica das informações

Análise de Dados: Apresentação e  
Ferramentas Básicas

1  
Módulo

**Fundação Escola Nacional de Administração Pública**

**Diretoria de Desenvolvimento Profissional**

**Conteudista/s**

Ricardo Alexandre Amaral (conteudista, 2022);  
Diretoria de Desenvolvimento Profissional.



Enap, 2022

Fundação Escola Nacional de Administração Pública

Diretoria de Desenvolvimento Profissional

SAIS - Área 2-A - 70610-900 — Brasília, DF

# Sumário

<b>Unidade 1: Noções básicas de análise de dados .....</b>	<b>5</b>
1.1 Dados: definição, tipos e a tomada de decisão .....	5
1.2 Análise qualitativa, quantitativa e complementaridade.....	7
Referências .....	11
<b>Unidade 2: Por Que Estatística? .....</b>	<b>12</b>
2.1 O domínio de atuação da estatística.....	12
2.2 Aplicações do tratamento estatístico de informações .....	14
Referências .....	16
<b>Unidade 3: O Poder das Lentes Estatísticas.....</b>	<b>17</b>
3.1 A leitura de dados sociais com base na análise de dados.....	17
3.2 Exame de dados sociais brasileiros.....	21
Referências .....	23

## Apresentação e Boas-vindas

Seja bem-vindo e bem-vinda ao curso Análise de Dados: uma Leitura Crítica das Informações.

Neste curso as noções de estatística serão as lentes fundamentais que nortearão o seu estudo. Mas não se preocupe, as ferramentas estatísticas necessárias para o entendimento do conteúdo serão definidas e retomadas ao longo de todo o curso para ajudar você a alcançar o objetivo geral desta capacitação. Ao final de seus estudos é esperado que você saiba utilizar os fundamentos estatísticos para ler os dados de maneira objetiva e crítica. Dessa forma, o objetivo central do curso se volta à necessidade de entender, na atual era da informação, o conceito de dados e de como tecer uma leitura crítica e objetiva destes.

A estatística é entendida como a ciência dos dados, visto que tem ligação direta com o tratamento desses dados em vários campos do saber. Especialmente neste curso, você poderá analisar o alinhamento da estatística com o desenvolvimento da sociedade, via produção de pesquisa social e por meio de indicadores que permitem mensurar desde características regionais de um país, até suas demandas de desenvolvimento socioeconômico.

O propósito é que este curso sirva para aprimorar um olhar crítico sobre o tema análise de dados e sobre outras áreas de atuação adjacentes ao tema como, por exemplo: gestão, mercado, políticas públicas e a cultura do cidadão pleno. Dessa forma, uma vez que o indivíduo estabeleça vínculo com as mídias informativas espera-se que ele esteja apto à leitura e interpretação de dados.

O conteúdo ao longo do curso está distribuído em três módulos. No primeiro módulo você irá reconhecer qual a importância da análise de dados, irá compreender a conexão entre estatística e tratamento de dados, além de examinar sistemas sob o prisma estatístico.

No segundo módulo você irá reconhecer os principais formatos de representação de dados, analisar correlações não-lineares, além de examinar os ajustes linear e quadrático.

Por fim, no terceiro módulo, irá identificar a contribuição computacional associada ao tratamento de dados, reconhecerá as conexões entre análise de dados e pensamento estratégico, além de reconhecer a importância da análise de dados para a leitura crítica das informações.

Então é hora de começar!

# 1 Análise de Dados: Apresentação e Ferramentas Básicas

## Unidade 1: Noções básicas de análise de dados

### Objetivo de aprendizagem

*Ao final desta unidade, você será capaz de reconhecer a relevância da análise de dados.*

---

Por meio da construção de um contexto que exemplifique as definições básicas relativas à análise de dados e suas aplicações diretas você irá desenvolver um arcabouço teórico básico aplicado ao tema.

### 1.1 Dados: definição, tipos e a tomada de decisão

Para iniciar seus estudos sobre o tema “dados” você pode começar pensando sobre duas perguntas:

- Como é possível descrever um sistema complexo e intrincado como o de um grande país como o Brasil?
- Neste contexto, como é possível representar temas como renda, educação, trabalho, economia, em âmbito nacional, de uma forma mais simples?

Em resposta a estes questionamentos iniciais surge o campo de pesquisa da Estatística, que é definida como um conjunto de técnicas para o tratamento de dados.

Para exemplificar como a estatística trabalha com o tratamento de dados, a figura a seguir mostra um conjunto de índices relacionados ao Brasil. Nela estão associados resultados estatísticos construídos através da interpretação de dados que tentam traduzir, de forma resumida, como é o país em diferentes áreas de interesse. Neste caso os dados estão organizados nas categorias de escolarização, analfabetismo, fecundidade, mortalidade infantil, renda e desocupação.



**Escolarização** de 6 a 14 anos

99,7% [2019]

AZ

**Analfabetismo** + 15 anos

6,6% [2019]



**Fecundidade**

1,76 filhos por mulher [2020]



**Mortalidade infantil**

11,56 óbitos por mil nascidos vivos [2020]



**PIB per capita**

35.161,70 R\$ [2019]



**Desocupação**

12,60% [3º trimestre 2021]

#### **Conjunto de índices relacionados ao Brasil.**

Fonte: IBGE – Indicadores Brasileiros. Elaboração: CEPED/UFSC (2022).

É preciso ter em mente que a análise de dados utiliza fragmentos estatísticos para tratar e organizar dados com a intenção de desvendar informações e caracterizar sistemas que sem a utilização de recursos estatísticos seriam praticamente de impossível compreensão, devido à imensa quantidade de informação intrinsecamente atrelada a cada uma de suas esferas.

Órgãos governamentais, entidades e empresas como as Organizações das Nações Unidas (ONU) e a Google tem um pensamento fortemente calcado em dados. Isso pode ser percebido em seus relatórios ou em suas ferramentas de tecnologia como o Google Analytics, que é um serviço automatizado de fluxo de acesso a páginas da internet.

No cenário brasileiro, um exemplo típico da análise de dados são as pesquisas por amostragem da população. A Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua), por exemplo, é realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e, ao invés de entrevistar cada brasileiro, constrói uma base de dados relativos a grupos de indivíduos e tece um panorama global das características da população brasileira sobre aspectos como renda, escolarização, moradia, constituição familiar etc.



## SAIBA MAIS

Para saber mais sobre a Pesquisa Nacional por Amostra de Domicílios Contínua, clique [aqui](#) e acesse o site.

A análise de dados é vista hoje como uma ferramenta fundamental para tratar dados e extrair informações de modo hábil e dinâmico nas mais diversas áreas do conhecimento. Elementos da análise de dados são percebidos a todo momento no contato com veículos de informação, que mostram taxas, índices e gráficos nos mais diversos ramos do conhecimento.

### 1.2 Análise qualitativa, quantitativa e complementaridade

É interessante que no estudo sobre análise de dados haja o conhecimento de uma base mínima de termos estatísticos. Para isso, veja a definição de alguns conceitos básicos que serão norteadores ao longo deste conteúdo:

- **Elementos:** partes, objetos que tecem um conjunto. Exemplo: Distrito Federal, Acre, Alagoas, Sergipe e Tocantins, como elementos do conjunto de estados brasileiros.
- **População:** é um conjunto de elementos, é o rol a ser estudado. Exemplo: estados e Distrito Federal, compõem a população de “estados brasileiros”.
- **Amostra:** é uma parte, um subconjunto da população. Exemplo: a seleção dos 5 estados com maior participação industrial no Brasil.

Deste modo, em estudos acerca de dados econômicos da indústria de tecnologia brasileira (população) é praticamente impossível analisar todas as informações (elementos), uma vez que se trata de uma população exaustivamente grande, ao invés disso são tecidos estudos em amostras para recriar o entendimento acerca de todo o sistema (população).

Da mesma forma, no exame quanto às características da constituição familiar brasileira é igualmente inviável entrevistar todos os brasileiros (população) sobre as particularidades de sua família (elementos), ao passo que a alternativa é tecer e analisar conjuntos de elementos, subconjuntos da população (amostras) formados por uma parcela de todas as famílias, cuja finalidade é retratar a população familiar brasileira inteira.



## DESTAQUE

Neste contexto, a estatística é a ciência responsável pela coleta, organização, tratamento, interpretação de dados e pela extrapolação dos resultados desses conceitos, lembrando que as definições estatísticas dependem da interpretação do problema concretamente estudado.

Existem muitas classificações de dados, dentre as quais é possível destacar:

- **Qualitativa:** subjetiva, não numérica, associada a qualidades. Exemplo: entre os estados brasileiros, São Paulo é o mais representativo com relação ao PIB nacional (IBGE, 2021);
- **Quantitativa:** é objetiva, numérica. Exemplo: o estado de São Paulo representa 31,8% da soma de todos os bens e serviços finais produzidos pelo Brasil (PIB) segundo o IBGE em 2021;
- **Quali-Quantitativas:** híbridas, utilizam-se de uma análise qualitativa e quantitativa. Exemplo: Embora o PIB total do Brasil seja da ordem de R\$ 7,4 trilhões, distribuídos por 26 unidades, o que resulta em R\$ 270 milhões por estado (média), o valor médio exposto não caracteriza suficientemente bem o PIB de cada estado, uma vez que os dados são muito heterogêneos (análise feita com base nos dados do IBGE, 2021).

A seguir, um quadro relacionado ao PIB dos estados brasileiros ajuda a exemplificar como poderiam ser apresentadas as variáveis quantitativas do PIB de cada estado, ou seu percentual. O quadro mostra elementos do PIB de cada estado brasileiro, que por sua vez tecem uma população de dados quantitativos com relação a este indicador econômico.



<b>PIB médio dos estados</b>		
<b>Estado</b>	<b>PIB (de milhões de R\$)</b>	<b>%</b>
<b>Acre</b>	15.331	0,22
<b>Alagoas</b>	54.413	0,78
<b>Amapá</b>	16.795	0,24
<b>Amazonas</b>	100.109	1,73
<b>Bahia</b>	286.240	4,09
<b>Ceará</b>	155.904	2,23
<b>Distrito Federal</b>	254.817	3,64
<b>Espírito Santo</b>	137.020	1,96
<b>Goiás</b>	195.682	2,79
<b>Maranhão</b>	98.179	1,40
<b>Mato Grosso</b>	137.443	1,96
<b>Mato Grosso do Sul</b>	106.969	1,53
<b>Minas Gerais</b>	614.876	8,78
<b>Paraná</b>	440.029	6,28
<b>Paraíba</b>	64.374	0,92
<b>Pará</b>	161.350	2,30
<b>Pernambuco</b>	186.352	2,66
<b>Piauí</b>	50.378	0,72
<b>Rio de Janeiro</b>	758.859	10,83
<b>Rio Grande do Norte</b>	66.970	0,96
<b>Rio Grande do Sul</b>	457.294	6,53
<b>Rondônia</b>	44.914	0,64
<b>Roraima</b>	13.370	0,19
<b>Santa Catarina</b>	298.227	4,26
<b>Sergipe</b>	42.018	0,60
<b>São Paulo</b>	2.210.562	31,56
<b>Tocantins</b>	35.666	0,51
	7.004.141	100,00
<b>PIB Médio dos Estados (milhões de R\$)</b>	269.390,04	

**PIB médio dos estados.**

Fonte: IBGE. Elaboração: CEPED/UFSC (2022).

Ao chegar até aqui você finalizou o estudo desta unidade! Parabéns! Caso surjam dúvidas, revise o conteúdo para aprofundar seu entendimento sobre o que foi apresentado.

## Referências

ALVES, Isabel Fraga. Data Science, Big Data e um novo olhar sobre a Estatística. **Boletim SPE: O Tema Central da Estatística - um novo olhar**, Lisboa, v. 12, n. 2, p. 29-31, 2017. Semestral.

GIBBS, Graham. **Análise de dados qualitativos**. Porto Alegre, Artmed; 2009.

HURWITZ, Judith et al. **Big Data para leigos**. Rio de Janeiro, Alta Books Editora, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA (IBGE). **O que é o PIB**. Rio de Janeiro, IBGE, 2021. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 12 nov. 2021.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. Sebastopol (USA) O'Reilly Media, Inc., 2013.

SILVESTRE, António. **Análise de dados e estatística descritiva**. Forte da Casa, Escolar Editora, 2007.

VUOLO, José Henrique. **Fundamentos da teoria de erros**. São Paulo, Editora Blucher, 1996.

## Unidade 2: Por Que Estatística?

### Objetivo de aprendizagem

*Ao final da unidade você será capaz de reconhecer a conexão entre estatística e o tratamento de dados, tendo em vista a construção da informação.*

---

### 2.1 O domínio de atuação da estatística

É muito difícil delimitar a difusão dos recursos estatísticos, visto que a construção de médias, a tabulação de dados, a utilização de uma planilha ou a leitura de um gráfico em um jornal é, essencialmente, praticar análise de dados. A diferença básica entre essas ações é o nível de proficiência exigida em áreas cuja estatística se torna uma verdadeira linguagem como, por exemplo, a Econometria, a Bioestatística, o Controle de Qualidade, a Teoria de Erros etc.

O fato é que a estatística trata dados e contribui de forma significativa para um processo de tomada de decisão. Em conformidade ao crescente volume de dados atualmente tratados, a utilização de recursos computacionais é franca e alavanca as potencialidades das ferramentas estatísticas, o que dá origem a métodos numérico-computacionais para tratamento de informação.

Tais métodos numérico-computacionais utilizam técnicas quantitativas que são tratadas de modo não-manual, sendo executadas por dispositivos eletrônicos como computadores, celulares, entre outros, que se diferenciam, basicamente, pelos diferentes níveis de automação que apresentam.

Um exemplo deste método pode ser observado quando é executada uma soma em uma planilha. A soma e outras operações são feitas de modo automático (ou sem a intervenção humana). Esse é um exemplo típico de um método numérico que possui algum nível de automação.



#### Automação das tabelas.

Fonte: Freepik. Elaboração: CEPED/UFSC (2022).

Amarrado aos métodos numéricos, um termo que está se tornando usual nos dias atuais é o *big data*, que em suma é a enorme capacidade de captação, armazenamento e processamento de informações alavancadas pelos avanços computacionais disponíveis.



## DESTAQUE

***Big data* trata-se da aplicação de sistemas e técnicas de tratamento de dados ajustados para lidar com um grande volume de dados, na maioria das vezes impraticáveis por meio de computadores domésticos.**

Logo, o *big data* é uma extensão dos métodos numéricos de análise de dados que visa tratar de gigantescas quantidades de dados e é comumente caracterizado por um tratamento de dados massivamente automatizado a altura do volume de dados que examina. Um exemplo são as ferramentas que estudam as preferências, interesses e comportamento dos usuários da internet e assim recomendam entretenimento e compras específicas ao perfil do usuário.

## 2.2 Aplicações do tratamento estatístico de informações

Em conformidade com a estatística, as aplicações derivadas dos métodos estatísticos também se mostram incomensuráveis e estão atreladas desde a sugestão de uma música pelo seu serviço de *streaming* de áudio à tomada de decisão em âmbito empresarial.



### DESTAQUE

**Ao contrário de uma demarcação precisa, o propósito deste tópico é elencar as principais aplicações estatísticas de modo contextual e, assim, sustentar o entendimento dos conceitos basilares acerca da análise de dados e a criticidade que emerge dela.**

Na sociedade atual, há um bombardeio de emaranhados de dados de natureza estatística que buscam quantificar e traduzir comportamentos sobre diversos campos de atividades, tais como esportes, política, economia, costumes regionais e sociais. São tabelas, gráficos, índices e muitas outras representações com as quais os indivíduos se deparam em uma simples pesquisa de preços no uso cotidiano de um celular.

Deste modo a análise de dados tem contribuído de forma significativa para o processo de tomada de decisão em diversos ramos, tanto na iniciativa privada como na pública, visto que grande parte do que se executa ou planeja é baseado em métodos quantitativos que extraem resultados a partir de dados brutos.

Atualmente o volume de dados produzidos pelo mundo moderno é muito grande, de modo que a automatização dos processos, como a utilização de recursos computacionais, é notória, todavia as bases matemáticas são comumente mantidas. Em outras palavras, a consistência da análise de dados é a mesma, apenas os algoritmos de cálculo são executados por máquinas para conseguir cumprir demandas cada vez maiores, o que significa que um cálculo de média que antes era tecido manualmente, agora é computacional.



**Volume de dados produzidos pelo mundo moderno.**

Fonte: Freepik. Elaboração: CEPED/UFSC (2022).

É difícil perceber uma área do conhecimento humano alheia às técnicas de análise de dados, muito comuns e que tangem ao planejamento, coleta, análise, organização e interpretação de dados de pesquisas envolvendo censos ou estudos por amostragem, associados a um estudo de toda a população ou elementos aleatórios desta.

Assim é notório frisar que as estratégias de tratamento de dados não se limitam a nichos específicos ou órgãos estatais como o IBGE. Essas estratégias fazem parte de diversas áreas e podem ser empregadas, por exemplo, para gerir uma empresa que precise determinar a associação entre a renda média e o perfil de produtos demandado pelo público que quer atingir. Para isso, essa empresa poderia utilizar a análise de informações e resultados que são fruto das estratégias de tratamento de dados.

De uma forma cada vez mais cotidiana, o contato com jornais, sites de notícias, ou mesmo uma procura de preços via internet, requer uma frequente leitura de tabelas de dados, gráficos, médias, tendências etc. Atualmente, praticamente toda a informação que é fornecida ao público utiliza ferramentas de análise de dados.

Ao chegar até aqui você finalizou o estudo desta unidade! Caso surjam dúvidas, reveja o conteúdo. Bons estudos!

## Referências

ALVES, Isabel Fraga. Data Science, Big Data e um novo olhar sobre a Estatística. **Boletim SPE: O Tema Central da Estatística - um novo olhar**, Lisboa, v. 12, n. 2, p. 29-31, 2017. Semestral.

CARVALHO, Marília Sá; SOUZA-SANTOS, Reinaldo. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 21, p. 361-378, 2005.

COMARELA, Giovanni et al. Introdução à Ciência de Dados: Uma Visão Pragmática utilizando Python, Aplicações e Oportunidades em Redes de Computadores. SCHAEFFER FILHO, Alberto Egon; CORDEIRO, Weverton Luis da Costa; CAMPISTA, Miguel Elias Mitre (ed.). **Minicursos do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, Porto Alegre, Sociedade Brasileira de Computação, p. 246-295, 2019.

GIBBS, Graham. **Análise de dados qualitativos**. Porto Alegre, Artmed; 2009.

HURWITZ, Judith et al. **Big Data para leigos**. Rio de Janeiro, Alta Books Editora, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA (IBGE). **O que é o PIB**. Rio de Janeiro, IBGE, 2021. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 12 nov. 2021.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. Sebastopol (USA) O'Reilly Media, Inc., 2013.

SILVESTRE, António. **Análise de dados e estatística descritiva**. Forte da Casa, Escolar Editora, 2007.

SOUZA, Emanuel Fernando Maia de; PETERNELLI, Luiz Alexandre; MELLO, Márcio Pupin de. **Software Livre R: aplicação estatística**. 2014. Universidade Federal da Paraíba.

VITALI, Marieli Mezari. Estatística sem matemática para psicologia. **Revista Brasileira de Psicodrama**, São Paulo, v. 27, n. 1, p. 139-144, 2019.

VUOLO, José Henrique. **Fundamentos da teoria de erros**. São Paulo, Editora Blucher, 1996.



## Unidade 3: O Poder das Lentes Estatísticas

### Objetivo de aprendizagem

*Ao final desta unidade você será capaz de examinar sistemas sob o prisma estatístico.*

---

### 3.1 A leitura de dados sociais com base na análise de dados

Um campo notório da análise de dados é a construção e compreensão de parâmetros estatísticos para ilustrar características de um grande volume de dados. A partir da análise de dados é possível compreender, por exemplo, sistemas complexos como a sociedade brasileira e aspectos de seus indicadores sociais. Neste tipo de tarefa são utilizados métodos estatísticos, frequentemente quantitativos, para ilustrar e comunicar um conjunto de fenômenos complexos de uma forma simples.

Durante este tópico será utilizado o exemplo da sociedade brasileira e seus indicadores para que seja mais fácil compreender alguns conceitos complexos que serão apresentados.

Para iniciar essa discussão é preciso que você compreenda o significado do conceito de média. O mecanismo de cálculo de média é comum e aparece cedo nos conteúdos escolares, sendo atrelado ao valor que busca melhor representar um conjunto de medidas.

Veja sua definição:



### DESTAQUE

O algoritmo de cálculo da média (ou média simples) determina que sejam somados todos os termos e divididos pela quantidade de dados envolvidos, e o resultado tem a função de representar os dados que lhe originaram.

**Exemplo:** pense que um conjunto de notas foi atribuído por um professor e são elas as seguintes: 8, 7 e 10. Você poderia determinar uma média utilizando o cálculo abaixo:

$$\text{Média} = (7 + 8 + 10)/3 = 25/3 = 8,33$$

Veja que o resultado é igual a 8,33. Esse fator é o que tende a melhor representar o conjunto de notas analisadas.

Transpondo para o cenário nacional, uma média atrelada a regiões pode recriar indicadores sociais simples, diretos e altamente compactos quanto a informações importantes como escolarização, renda etc. Esses fatores servem, estrategicamente, para identificação de bolsões de vulnerabilidade e pobreza, e a implantação de políticas específicas para sanar essas desigualdades, por exemplo.

A seguir, uma tabela que contextualiza essa discussão com relação à média e sua compacta representação de informações, visto que nela são exibidos, resultados acerca do rendimento médio dos brasileiros no período de 2012 a 2021 com base no IBGE (2021):

Período	Valores (R\$)	Desvio
nov-dez-jan 2021	2.651	75
dez-jan-fev 2021	2.627	51
jan-fev-mar 2021	2.635	59
fev-mar-abr 2021	2.604	28
mar-abr-mai 2021	2.602	26
abr-mai-jun 2021	2.562	-14
mai-jun-jul 2021	2.539	-37
jun-jul-ago 2021	2.503	-73
jul-ago-set 2021	2.459	-117
<b>Média</b>	2.576	

**Dados do rendimento médio dos brasileiros no nos trimestres do ano de 2021 apresentados em tabela.**

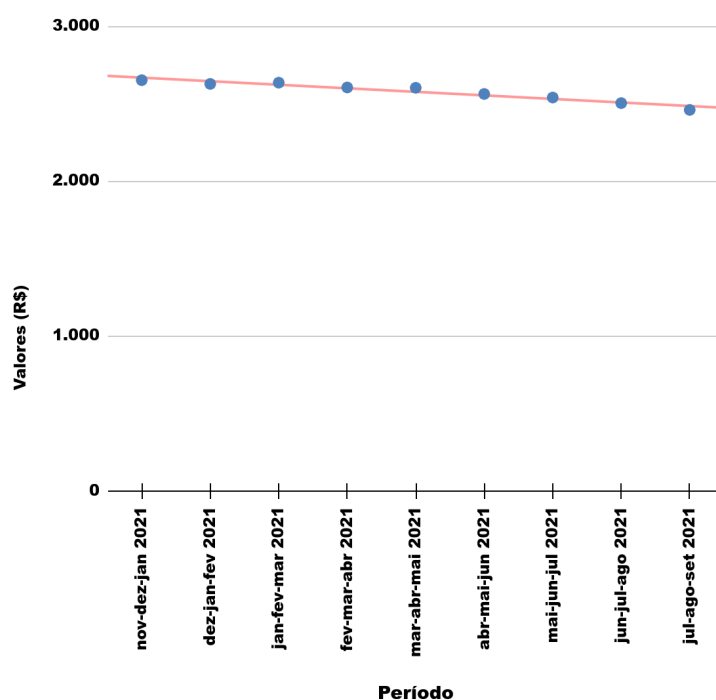
Fonte: IBGE (2021). Elaboração: CEPED/UFSC (2022).

Agora que você já observou os dados da tabela, atente a alguns detalhes:

- Os valores médios em cada período, ou seja, cada linha, recriam de modo compacto um gigantesco número de dados com relação à renda dos brasileiros;

- A tabela em discussão é um recorte, uma vez que o rol completo de tais resultados é muito extenso;
- A última linha da tabela exibe uma média, calculada em relação a todo o período, ou seja, somando todos os fatores da coluna de valores e dividindo pela quantidade de termos e, assim, obtendo o resultado;
- A terceira coluna da tabela, denominada desvio, mostra o quanto os valores se posicionam acima ou abaixo deste valor médio global. Veja o exemplo (primeira linha e terceira coluna), nele o primeiro valor é determinado como a diferença entre o valor no período e a média, assim  $2651-2576= 75$ . Ou seja, o resultado está 75 acima da média.

O gráfico a seguir está atrelado às informações de rendimento médio dos brasileiros no período de 2012-2021 descritas na tabela anterior, e apenas organiza os dados de outra forma. Dentre as nuances deste novo modelo de exposição dos dados é interessante observar que o gráfico é global e compacto, pois consegue acomodar mais facilmente todas as informações, não sendo necessário suprimir linhas como comumente é feito para tabelas extensas, além de explicitar tendências, máximos e mínimos.



**Dados do rendimento médio dos brasileiros no período nos trimestres do ano 2021 apresentados em gráfico.**

Fonte: IBGE. Elaboração: CEPED/UFSC (2022).

No gráfico é possível visualizar as mesmas informações sobre o rendimento médio dos brasileiros que foram vistas na tabela anterior. É importante observar que:

- a linha vermelha mostrada no gráfico está sugestivamente por volta de 2500, indicando que é a média global;
- a renda média em cada período, representada por pontos azuis, é comumente estável ao longo de todo o período de trimestres de 2021, uma vez que se mantém sem grandes aumentos e quedas ao longo de todo o período estudado;
- a renda média em cada período exibe um máximo para o trimestre dos meses de novembro, dezembro e janeiro de 2021 e um mínimo durante o trimestre dos meses de julho, agosto e setembro de 2021.

Outra demonstração da utilidade da análise de dados é a construção de índices que, assim como as médias, auxiliam na construção de medidas estatísticas com capacidade de representar grandes sistemas.



## DESTAQUE

A rigor, um índice apresenta a evolução de um número ao longo do tempo, como uma razão em formato percentual, ou seja, uma divisão direta do novo preço pelo preço antigo, escrita na forma de porcentagem.

**Exemplo:** se o litro de combustível em uma dada semana custa R\$5,00 e, na seguinte, seu valor é R\$5,50, então se divide  $5,5/5$ , o que resulta em 1,1 e é lido em valor percentual como 110%. Portanto, entende-se como uma alta de 10% para o produto no período exemplificado.

O índice de preços ao consumidor (IPC) é um indicador da inflação do país e é determinado com base no preço médio necessário para comprar um conjunto de bens de consumo e serviços, também em comparação a um período anterior.

Deste modo, o IPC pode ser relacionado de maneira análoga ao exemplo anterior, exceto que não faz jus a um único produto mas a uma grande lista de produtos e serviços.

O método de cálculo do IPC é uma comparação direta do preço da cesta de produtos

e serviços entre dois períodos e, assim, é determinado pela razão do preço antigo (PA) e subsequente ou corrente (PC) da mesma cesta. Isso significa dividir tais custos e interpretá-los como percentuais, o que justifica a multiplicação por 100 na fórmula de cálculo. A fórmula do cálculo IPC se organiza como está apresentado a seguir.

$$IPC = (PC)/(PA).100$$

Índices análogos ao IPC são frequentemente encontrados, um dos exemplos são os indicativos de emprego associados à população em idade ativa, que basicamente executam uma razão entre a população ocupada (PO) e a população desocupada (PD), ou seja, taxa de emprego é (PO)/(PD).

Praticamente todos os países, organismos internacionais e empresas líderes de mercado se mostram interessadas em investir na análise de dados, pela qual é possível não só caracterizar o presente, como também estimar e fazer projeções futuras.

### 3.2 Exame de dados sociais brasileiros

O Instituto Brasileiro de Geografia e Estatística (IBGE) é um instituto público da administração federal brasileira, criado em 1934, e cujas atribuições estão intimamente ligadas às geociências e estatísticas sociais, demográficas e econômicas. Caracteriza-se pela realização de censos e tratamento das informações obtidas nessas pesquisas para suprir órgãos governamentais e o público em geral.

Como um todo, os dados coletados e tratados pelo IBGE são instrumentos de caracterização da população brasileira. Um exemplo é o censo de 1970, um marco com relação à organização, à riqueza de detalhes e à confiabilidade dos resultados.



**Instituto Brasileiro de Geografia e Estatística.**

Fonte: IBGE. Elaboração: CEPED/UFSC (2022).

É interessante constatar que o IBGE faz e promove pesquisa de cunho fortemente social, visto que é o maior provedor de informações do país, as quais planificam inúmeras ações, sobretudo no campo de políticas públicas. Isso inclui ações de distribuição de renda, ou a implantação de escolas técnicas federais em um município que detém baixos indicadores socioeconômicos com vista no desenvolvimento regional, por exemplo.



## DESTAQUE

É relevante deixar claro que a representação promovida pela análise de dados, embora importante na caracterização de um sistema ou da dinâmica social de um país, é uma abstração, um esboço estatístico, o qual não é perfeito. É o análogo da organização de um rol de pistas, em um processo de investigação.

Dentre as alternativas para minimizar as imperfeições inerentes à análise de dados é interessante utilizar diferentes métodos de investigação, como métodos qualitativos, quantitativos e híbridos (ambos em conjunto), visto que tais vieses não são hierarquizados, mas complementares.

A Pesquisa Nacional por Amostra de Domicílios (PNAD) e suas reformulações como a PNAD Contínua, é um estudo aplicado desde 1967 pelo IBGE que se utiliza desses métodos. Esse estudo capta informações para produzir resultados estatísticos estratégicos para o Brasil e ao longo do tempo foi aperfeiçoando fatores como periodicidade, métodos de cálculo, captação de amostras etc. Sua ideia central gira em torno de apurar resultados e mapear características gerais sobre população, educação, trabalho, rendimento e habitação.

Assim são captados dados com relação à posse e uso de utensílios, escolaridade, renda, naturalidade, formas de obtenção de informação, etc.

## Referências

ALVES, Isabel Fraga. Data Science, Big Data e um novo olhar sobre a Estatística. **Boletim SPE: O Tema Central da Estatística - um novo olhar**, Lisboa, v. 12, n. 2, p. 29-31, 2017. Semestral.

CARVALHO, Marília Sá; SOUZA-SANTOS, Reinaldo. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 21, p. 361-378, 2005.

COMARELA, Giovanni et al. Introdução à Ciência de Dados: Uma Visão Pragmática utilizando Python, Aplicações e Oportunidades em Redes de Computadores. SCHAEFFER FILHO, Alberto Egon; CORDEIRO, Weverton Luis da Costa; CAMPISTA, Miguel Elias Mitre (ed.). **Minicursos do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, Porto Alegre, Sociedade Brasileira de Computação, p. 246-295, 2019.

GIBBS, Graham. **Análise de dados qualitativos**. Porto Alegre, Artmed; 2009.

HURWITZ, Judith et al. Big Data para leigos. Rio de Janeiro, Alta Books Editora, 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA (IBGE). **O que é o PIB**. Rio de Janeiro, IBGE, 2021. Disponível em: <https://www.ibge.gov.br/explica/pib.php>. Acesso em: 12 nov. 2021.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business: What you need to know about data mining and data-analytic thinking**. Sebastopol (USA) O'Reilly Media, Inc., 2013.

SILVESTRE, António. **Análise de dados e estatística descritiva**. Forte da Casa, Escolar Editora, 2007.

SOUZA, Emanuel Fernando Maia de; PETERNELLI, Luiz Alexandre; MELLO, Márcio Pupin de. Software Livre R: aplicação estatística. 2014. Universidade Federal da Paraíba.

VITALI, Marieli Mezari. Estatística sem matemática para psicologia. **Revista Brasileira de Psicodrama**, São Paulo, v. 27, n. 1, p. 139-144, 2019.

VUOLO, José Henrique. **Fundamentos da teoria de erros**. São Paulo, Editora Blucher, 1996.