

Conceitos básicos para o entendimento da estatística pela sociedade

Jean Michel Lattaro

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Jean Michel Lattaro

Conceitos básicos para o entendimento da estatística pela sociedade

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Mestrado Profissional em Matemática em Rede Nacional. *VERSÃO REVISADA*

Área de Concentração: Mestrado Profissional em Matemática em Rede Nacional

Orientadora: Profa. Dra. Juliana Cobre

USP – São Carlos
Julho de 2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

L351c Lattaro, Jean Michel
 Conceitos básicos para o entendimento da
estatística pela sociedade / Jean Michel Lattaro;
orientadora Juliana Cobre. -- São Carlos, 2022.
 65 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Mestrado Profissional em Matemática em Rede
Nacional) -- Instituto de Ciências Matemáticas e de
Computação, Universidade de São Paulo, 2022.

1. Estatística. 2. Matemática. 3. Gráficos
estatísticos. 4. Função densidade. I. Cobre, Juliana,
orient. II. Título.

Jean Michel Lattaro

Basic concepts for society's understanding of statistics

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of Mathematics Professional Master's Program. *FINAL VERSION*

Concentration Area: Professional Master Degree Program in Mathematics in National Network

Advisor: Profa. Dra. Juliana Cobre

USP – São Carlos
July 2022

*Dedico esse trabalho ao meu pai Marcos Anthony Morosin Lattaro e à
minha mãe Marli Magda Monte Lattaro.*

AGRADECIMENTOS

Agradeço primeiramente a Deus por me abençoar com discernimento, força e saúde, guiando o meu caminho para concluir esta Dissertação de Mestrado.

Agradeço especialmente à Professora Orientadora Dra. Juliana Cobre pela sua colaboração, atenção e paciência em nossas reuniões.

Agradeço aos Professores do Departamento de Matemática da Universidade de São Paulo do campus de São Carlos, que com tanto conhecimento matemático, me proporcionaram construir ao longo de todos esses anos, o meu crescimento não só profissional, mas também pessoal, tanto na graduação quanto na pós-graduação.

Agradeço a todos os meus colegas de turma, por toda colaboração e ajuda.

Agradeço a todos os meus amigos e à minha família, em especial ao meu pai Marcos Lattaro por todo o tempo que passamos juntos fazendo esta dissertação e pelos ensinamentos que ele me proporcionou e que me levaram a ter equilíbrio e a chegar até aqui, à minha mãe Marli Lattaro pelo carinho e dedicação comigo durante a minha vida inteira que me levaram a ter fé, confiança e esperança.

Consagre ao Senhor tudo o que você faz, e os seus planos serão bem-sucedidos.

(Provérbios 16:3)

RESUMO

LATTARO, J. M. **Conceitos básicos para o entendimento da estatística pela sociedade**. 2022. Dissertação (Mestrado em Ciências – Matemática em Rede Nacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

As figuras gráficas estão presentes nos veículos de comunicação diariamente e, nem sempre, de forma correta. Sendo assim, é bastante relevante o entendimento de diversos tipos de gráficos. O objetivo deste trabalho é provocar nos leitores uma visão mais ampla e até mesmo questionadora em relação aos gráficos estatísticos apresentados pelas mídias sociais em nosso cotidiano, para um pleno exercício da cidadania. O conteúdo deste trabalho também pode ser utilizado como material de apoio em uma aula de Estatística cujo objetivo seja abordar a leitura crítica dos gráficos apresentados pelas mídias sociais. Apresentamos uma revisão das principais estruturas e conceitos iniciais da Estatística importantes para a construção e compreensão de certos gráficos. Apresentamos também algumas análises de gráficos estatísticos divulgados pela mídia, com destaque para o tão mencionado "achatamento da curva" durante a pandemia do Covid-19. A pesquisa trouxe, como resultado, a constatação de que o conhecimento estatístico é extremamente importante para a formação do indivíduo, visto que permite que o cidadão seja mais crítico frente às informações com as quais se depara.

Palavras-chave: Gráficos estatísticos. Estatística. Matemática.

ABSTRACT

LATTARO, J. M. **Basic concepts for society's understanding of statistics**. 2022. Dissertação (Mestrado em Ciências – Matemática em Rede Nacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

The graphic figures are present in the communication vehicles daily and, not always, correctly. Therefore, it is very important to understand different types of graphs. The objective of this work is to provoke in the readers a broader and even questioning view in relation to the statistical graphs presented by social media in our daily lives, for a full exercise of citizenship. The content of this work can also be used as support material in a Statistics class whose objective is to approach the critical reading of the graphs presented by social media. We present a review of the main structures and initial concepts of Statistics important for the construction and understanding of certain graphs. We also present some analysis of statistical graphs released by the media, highlighting the much-mentioned "flattening of the curve" during the Covid-19 pandemic. The research brought, as a result, the finding that statistical knowledge is extremely important for the formation of the individual, since it allows the citizen to be more critical in the face of the information with which he is faced.

Key-words: Statistical graphs. Statistic. Math.

LISTA DE ILUSTRAÇÕES

Figura 1 - Classificação de uma variável.....	25
Figura 2 - Histograma relativo ao peso dos professores do PROFMAT em números absolutos.....	40
Figura 3 - Histograma relativo ao peso dos professores do PROFMAT em frequência relativa.....	41
Figura 4 - Histograma de densidade de frequência relativo ao peso dos professores do PROFMAT. E a curva de densidade de probabilidade estimada da variável aleatória em questão.....	42
Figura 5 - Polígono de frequência relativo ao peso dos professores do PROFMAT.....	43
Figura 6 - Gráfico de barras relativo ao número de escolas.....	46
Figura 7 - Gráfico de barras relativo ao nível preferido.....	46
Figura 8 - Gráfico de barras da Globo News da Inflação do Brasil.....	48
Figura 9 - Gráfico de barras da Inflação com destaque para um dos erros.....	49
Figura 10 - Gráfico de barras analisando eixo vertical, escala e o zero.....	50
Figura 11 - Gráfico de barras da Inflação retificado pela emissora.....	52
Figura 12 - Comparação entre os gráficos publicados pela emissora.....	52
Figura 13 - Comparação entre o gráficos da Pnad e da Receita Federal.....	55
Figura 14 - Massa apurada pela Pnad e Massa ajustada com a Receita.....	56
Figura 15 - Gráfico da Diferença de Renda em porcentagem.....	57
Figura 16 - Gráfico de comparação das curvas.....	58
Figura 17 - Gráfico da curva alta em "bloquinhos": Ruim.....	59
Figura 18 - Gráfico da curva alta: Ruim.....	60
Figura 19 - Gráfico da curva baixa em "bloquinhos": Boa.....	61
Figura 20 - Gráfico da curva baixa: "Achatamento da Curva".....	62

LISTA DE TABELAS

Tabela 1 - Pesos de uma amostra de 40 professores.....	32
Tabela 2 - Frequência dos pesos de uma amostra de 40 professores.....	36
Tabela 3 - Tabela de frequências dos pesos de uma amostra de 40 professores.....	39
Tabela 4 - Números de escolas.	45
Tabela 5 - Nível de escolaridade preferido.....	45

LISTA DE QUADROS

Quadro 1 - Pesos em quilogramas de uma amostra de 40 professores.....	30
---	----

SUMÁRIO

1. INTRODUÇÃO	21
2. REFERENCIAL TEÓRICO	23
2.1. Conceitos Iniciais	23
2.1.1. Dados e Variáveis	24
2.1.2. População	25
2.1.3. Amostra.....	25
2.1.4. Censo.....	25
2.1.5. Amostragem e Técnicas de Amostragem:	27
2.2. Apresentação de dados.....	30
2.2.1. Rol.....	30
2.2.2. Tabela	31
2.2.3. Distribuição de Frequência ou Tabela de Frequência:	33
2.2.4. Gráficos.....	39
3. ESTUDOS DE CASO	47
3.1. Gráfico 1: Inflação brasileira de 2013	47
3.2. Gráfico 2: Classes de rendimento: A, B, C, D e E.....	53
3.3. Gráfico 3: “Achatamento da curva” do coronavírus.....	58
4. CONSIDERAÇÕES FINAIS	63
REFERÊNCIAS	64

1. INTRODUÇÃO

Na sociedade moderna somos bombardeados por uma enorme quantidade de informação sobre os mais variados assuntos, tais como: testes ligados ao desempenho escolar, pesquisas eleitorais, estudos financeiros, controle de qualidade, análises de crescimento de doenças, taxas populacionais, índices de desenvolvimento, índices de desemprego, modelagem de fenômenos da natureza, etc. Por esta razão, torna-se essencial que a sociedade seja capaz de administrar, assimilar e interpretar a maior quantidade de dados e informações sobre os seus assuntos de interesse.

Uma das ferramentas mais utilizadas hoje em dia em todas as áreas do conhecimento é a Estatística, uma ciência que descreve os dados observados e desenvolve metodologia para auxiliar a tomada de decisão na presença de incerteza. Para fins didáticos podemos dividi-la em três partes: a) estatística descritiva; b) probabilidade; c) estatística inferencial.

A seguir focaremos em gráficos. Logo, vale destacar a Estatística Descritiva, cujo objetivo consiste em descrever os dados, buscando resumir e organizá-los de modo a esclarecer as características de interesse existentes nos mesmos. Para isso, utiliza instrumentos como tabelas, gráficos, medidas de tendência central ou dispersão.

Por sua natureza e complexidade, a informação muitas das vezes não é facilmente compreendida, motivo pelo qual deve passar por tratamentos que venham permitir uma melhor apresentação e assimilação. A Estatística surge justamente para preencher esse requisito, e a inclusão da mesma nos currículos do ensino básico vem se tornando uma realidade, tendo em vista as necessidades de seus conhecimentos em nosso cotidiano. Tal postura no Brasil é defendida pelos Parâmetros Curriculares Nacionais (PCNs), que, a fim de atender a tal necessidade, criou um bloco com diretrizes nos PCNs denominado Tratamento da Informação. Sobre a criação desse bloco, é esclarecido que:

A demanda social é que leva a destacar este tema como um bloco de conteúdo, embora pudesse ser incorporado aos anteriores. A finalidade do destaque é evidenciar sua importância, em função de seu uso atual na sociedade... Com relação à Estatística, a finalidade é fazer com que o aluno venha a construir procedimentos para coletar,

organizar, comunicar dados, utilizando tabelas, gráficos e representações que aparecem frequentemente em seu dia-a-dia. Além disso, calcular algumas medidas estatísticas como média, mediana e moda com o objetivo de fornecer novos elementos para interpretar dados estatísticos. (BRASIL, 1998, p. 52).

Sobre o tema, os PCNs também advertem que:

Ao trabalhar com o Tratamento da Informação é fundamental ainda que, ao ler e interpretar gráficos, os alunos se habituem a observar alguns aspectos que permitam confiar ou não nos resultados apresentados: por exemplo, observar a presença da frequência relativa e se as escalas utilizadas são convenientes. Costuma ser frequente nos resumos estatísticos a manipulação dos dados, que são apresentados em gráficos inadequados, o que leva a erros de julgamento. Esses erros também poderão ser evitados se os alunos forem habituados, em seus trabalhos de pesquisa, a identificar informações que não foram levantadas, bem como informações complementares, a comprovar erros que são cometidos ao recolher os dados, a verificar informações para chegar a uma conclusão. Assim, eles terão oportunidade de desenvolver conhecimentos para poder compreender, analisar e apreciar as estatísticas apresentadas pelos meios de comunicação e para um melhor reconhecimento das informações confiáveis. (BRASIL, 1998, p. 136).

Portanto, visando esse contexto, o nosso trabalho se concentra na parte de gráficos, indispensável em nossa sociedade contemporânea de informação. Para tanto ele encontra-se estruturado em duas partes. O segundo Capítulo apresenta a teoria necessária à compreensão e construção de um apanhado de conhecimentos, permitindo assim ao leitor a apropriação de saberes estatísticos. Já o terceiro Capítulo, busca, de forma prática, mostrar situações problema referentes à análise de gráficos veiculados pela mídia em diversos meios de comunicação.

Esperamos que a utilização das situações reais abordadas chame a atenção tanto do cidadão discente quanto do cidadão comum para as informações presentes intencionalmente, ou não, nas representações gráficas estatísticas, contribuindo assim para despertar no leitor o senso crítico e o interesse por tal conhecimento.

2. REFERENCIAL TEÓRICO

2.1. Conceitos Iniciais

Vamos iniciar o nosso estudo apresentando alguns conceitos preliminares acerca da Estatística.

Segundo Downing e Clark (2003, p. 2) “a palavra *estatística* tem dois significados: (1) um conjunto de dados numéricos e (2) um ramo da matemática, abrangendo a estatística descritiva e a inferência estatística, que se ocupa da análise de dados estatísticos”.

Para Walpole *et al.* (2009, p. 3), “a disciplina de probabilidade, assim, fornece a transição entre a estatística descritiva e os métodos inferenciais”. Logo, a estatística é uma área que utiliza teorias de probabilidade para explicação de eventos, estudos e experimentos. Essa teoria tem por objetivo obter, organizar e analisar dados, determinar suas correlações, tirar delas suas consequências para descrição, explicar o que passou e fazer previsões. Portanto, de acordo com as definições colocadas anteriormente, para entendermos melhor a Estatística podemos dividi-la em três frentes: Estatística Descritiva, Probabilidade e Inferência Estatística, que estão apresentadas a seguir.

Estatística Descritiva:

É a etapa inicial, em que vamos sumarizar uma série de valores. Em função da disponibilidade de uma grande quantidade de dados, torna complexa a simplificação e a percepção de certas informações e características de interesse. Essa etapa compreende a organização, resume e descreve os dados, sobretudo através de tabelas, gráficos e medidas de tendência central ou dispersão.

Probabilidade:

Em nosso cotidiano, muitos fenômenos são de natureza aleatória. Por exemplo, não podemos saber, de antemão, se vamos obter cara ou coroa no lançamento de uma moeda. Esse é um exemplo de fenômeno aleatório, pois o resultado é incerto. A probabilidade é um ramo da matemática que estuda esse tipo de fenômeno, ou seja, analisa situações que envolvam acaso, como jogos de dados, número de carros que passam em certo cruzamento em um período de tempo, número de infectados por um vírus, tempo até atendimento em um banco.

Estatística Inferencial:

A Inferência Estatística, por sua vez, visa analisar e interpretar dados amostrais, ou seja, tem por objetivo entrever respostas sobre um determinado conjunto a partir de amostras/subconjuntos desse conjunto.

Neste trabalho, o nosso foco se voltará para a primeira frente, que se trata da Estatística Descritiva. Buscando uma revisão teórica, começaremos com as definições.

2.1.1. Dados e Variáveis

“Dado” é o conjunto de valores medidos, mensurados ou observados (STEVENSON, 1981). Já o termo “variável” é descrito como possível resultado de um fenômeno (CRESPO, 1997).

Conforme Morettin e Bussab (2010, p. 9):

Algumas variáveis, como sexo, educação, estado civil, apresentam como possíveis realizações uma qualidade (ou atributo) do indivíduo pesquisado, ao passo que outras, como número de filhos, salário, idade, apresentam como possíveis realizações números resultantes de uma contagem ou mensuração. As variáveis do primeiro tipo são chamadas *qualitativas*, e as do segundo tipo, *quantitativas*.

De acordo com Morettin e Bussab (2010, p. 10):

Dentre as variáveis qualitativas, ainda podemos fazer uma distinção entre dois tipos: variável qualitativa *nominal*, para a qual não existe nenhuma ordenação nas possíveis realizações, e variável qualitativa *ordinal*, para a qual existe uma ordem nos seus resultados.

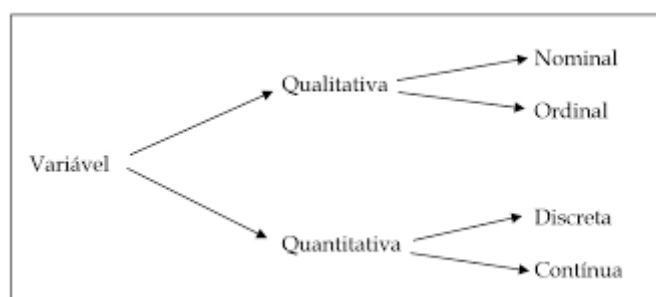
Por exemplo, a região de procedência, que pode ser classificada como interior, capital ou outra, numa pesquisa sobre alguns aspectos socioeconômicos dos empregados da seção de orçamentos de uma companhia, é um caso de variável nominal. Por outro lado, grau de instrução e classe social são exemplos de variável ordinal.

De modo análogo, as variáveis quantitativas podem ser classificadas entre dois tipos:

- (a) variáveis quantitativas *discretas*, cujos possíveis valores formam um conjunto finito ou enumerável de números, e que resultam, frequentemente, de uma contagem, como por exemplo número de filhos (0, 1, 2, ...);
- (b) variáveis quantitativas *contínuas*, cujos possíveis valores pertencem a um intervalo de números reais e que resultam de uma mensuração, como por exemplo estatura e peso (melhor seria dizer massa) de um indivíduo. (MORETTIN; BUSSAB, 2010, p. 10).

A Figura 1 representa a classificação de uma variável.

Figura 1 - Classificação de uma variável.



Fonte: Morettin e Bussab (2010).

2.1.2. População

Para Downing e Clark (2003, p. 2) “o termo *população* se refere a todos os indivíduos ou a todos os objetos do grupo em que estamos interessados”.

2.1.3. Amostra

Segundo Downing e Clark (2003, p. 2) “uma *amostra* é um conjunto de elementos extraídos da população”.

2.1.4. Censo

Trata-se de um levantamento que visa coletar dados populacionais, mais claramente, dados de toda uma população.

Para entendermos melhor a diferença entre os conceitos de amostra e de censo, e a necessidade da existência deles, pensemos o seguinte: uma empresa de *e-commerce* deseja lançar nacionalmente um novo produto. Então, possui basicamente três opções:

Opção nº 01: A empresa pode lançar o produto, realizando investimentos nas áreas de produção, logística, marketing e demais necessidades, aguardando a reação do mercado. Se a aceitação for positiva, então a estratégia foi bem-sucedida. Porém, se a aceitação for negativa, a empresa sofrerá prejuízos, o que não é nada interessante.

Opção nº 02: Antes de lançar o produto, a empresa pode encomendar uma pesquisa com toda a população, no caso, o país todo, buscando assim, encontrar os parâmetros de aceitação do mesmo e, se economicamente for viável, realizar então o seu lançamento. Porém, apesar de identificar a viabilidade econômica do produto, a realização de tal pesquisa pode se tornar um novo problema, devido ao valor financeiro investido para realizar um censo em um país do tamanho do Brasil, que, de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE) no dia 21 de janeiro de 2022, possuía um total de 214.133.097 habitantes.

Para se ter uma noção, seguem os dados do último censo, realizado em 2010.

Percorrer por inteiro um país como o Brasil, de dimensões continentais, com cerca de 8 milhões de km² de um território heterogêneo e, muitas vezes, de difícil acesso, é uma tarefa que envolve grandes números e uma logística de guerra. Veja, a seguir, os números que mostram as dimensões do Censo 2010, segundo o IBGE.

Universo a ser recenseado: todo o Território Nacional brasileiro

Número de municípios: 5.565 municípios

Número de domicílios: aproximadamente 58 milhões

Números de setores censitários: 314.018 setores censitários

Pessoal contratado e treinado: cerca de 240 mil pessoas (coleta, supervisão, apoio e administrativo)

Orçamento previsto: R\$ 1,4 bilhão

Tecnologia: centenas de computadores em rede nacional, rede de comunicação em banda larga e 220 mil computadores de mão equipados com receptores de GPS.

Unidades executoras: 27 unidades estaduais, 7 mil postos de coleta informatizados e 1.283 Coordenações de Subárea.

Levando-se em consideração os números astronômicos acima, é possível observar que tal pesquisa pode ser inviável do ponto de vista econômico, para o lançamento do produto. Sendo assim, isso nos remete à opção número três.

Opção nº 03: Antes de lançar o produto, a empresa pode encomendar uma pesquisa, a partir de uma amostra da população. Com esse método, constata-se a viabilidade do produto em questão, mediante a um investimento financeiro aceitável. Logo, a última opção, ou seja, uma pesquisa estatística, parece ser o caminho mais viável.

Para que aconteça tal pesquisa, a utilização de uma amostra da população é essencial, ou seja, necessitamos de um conjunto de elementos extraídos da população. Esses dados são tão relevantes que, conforme Downing e Clark (2003, p. 3), “no caso de escolhermos uma amostra que não represente bem o todo, iremos fazer previsões inexatas ao tentar estimar as características da população com base na amostra”. Um erro muito comum de várias pesquisas é não coletar uma amostra segundo critérios de amostragem. Aqui cabem dois novos conceitos: parâmetro e estatística.

Parâmetro

A partir de Morettin e Bussab (2010, p. 272) “um *parâmetro* é uma medida usada para descrever uma característica da população”.

Estatística

Estatística é uma característica da amostra de uma população (MORETTIN; BUSSAB, 2010).

2.1.5. *Amostragem e Técnicas de Amostragem:*

Amostragem:

A amostragem é uma técnica que garante o acaso na coleta de itens de uma população, formando estes uma amostra (CRESPO, 1997).

Uma amostra será tendenciosa caso não seja selecionada aleatoriamente, fazendo com que os dados não representem corretamente a população.

Abaixo descreveremos algumas técnicas de amostragem.

Amostragem probabilística:

Neste tipo de amostragem, cada elemento de uma população, para ser incluído na amostra, deve ter a sua probabilidade conhecida e diferente de zero.

Amostragem aleatória:

Na amostragem aleatória, se temos uma população de tamanho N , cada único elemento dela deve ter a mesma probabilidade de ser escolhido para integrar a amostra. Esta é a melhor técnica para escolha de uma amostra (DOWNING; CLARK, 2003).

Amostragem aleatória simples:

Na amostragem aleatória simples, os elementos da população que farão parte da amostra serão selecionados ao acaso. Apresenta como propriedade o fato de que, uma vez selecionada uma amostra de n elementos da população, qualquer outra amostra de n elementos possui a mesma probabilidade de ser selecionada. Por exemplo, um professor coloca os nomes dos alunos em uma caixa e seleciona alguns sem olhar, para obter uma amostra de alunos.

Amostragem aleatória sistemática:

Neste tipo de amostragem, um ponto inicial é escolhido aleatoriamente. A partir disso, seleciona-se um intervalo, por exemplo, de tamanho 5. Com esse intervalo conhecido, a cada sequência de 5 elementos da população, inclui-se o último elemento desse intervalo na amostra. Por exemplo, o professor de uma sala usa uma lista com os nomes dos alunos em ordem alfabética e escolhe um ponto inicial aleatório. Suponha que este ponto inicial aleatório resulte o número 18, então o 18º da lista é selecionado e de 5 em 5, selecionamos os demais elementos da amostra. Assim, seriam incluídos, o 18º, o 23º, o 28º, e assim por diante.

Amostragem aleatória estratificada:

Na amostragem estratificada, primeiramente, a população é dividida em subgrupos, ou seja, estratos. A amostra geral consiste em alguns membros de cada um desses grupos. Os membros de cada grupo são escolhidos ao acaso. Por exemplo, uma escola técnica entrevista 200 alunos obtendo amostras aleatórias

de 50 do primeiro ano, 50 alunos do segundo ano, 50 alunos do terceiro ano, e 50 alunos do último ano.

Este tipo de amostragem garante a representação na amostra dos membros de cada grupo, logo, este método é interessante quando queremos incluir alguns membros de cada grupo.

Amostragem aleatória por conglomerado:

Na amostragem por conglomerado, primeiramente, a população é dividida em grupos. A amostra geral é formada por todos os membros de alguns dos grupos. Os grupos são selecionados ao acaso. Por exemplo, uma companhia rodoviária decide entrevistar seus clientes em um dia, então, eles escolhem, aleatoriamente, 5 viagens de ônibus desse dia e entrevistam todos os passageiros das viagens selecionadas.

Este tipo de amostragem envolve todos os membros de alguns dos grupos, logo este método é interessante quando a população como um todo é refletida por cada grupo.

Amostragem não probabilística:

Ao contrário da amostragem probabilística, as probabilidades são desconhecidas, logo não é possível obter por exemplo a margem de erro.

Amostragem por conveniência:

É um método não probabilístico. Neste, o pesquisador seleciona uma amostra de fácil acesso de alguma maneira não aleatória.

Por exemplo, um pesquisador quer saber a opinião dos eleitores do Estado de São Paulo sobre a realização das eleições mediante a utilização das urnas eletrônicas. Para obter uma amostragem probabilística, seria necessário possuir um acesso à população dos eleitores paulistas para que pudéssemos selecionar uma amostra aleatoriamente, e então realizar a pesquisa. Já na amostragem por conveniência, poderíamos escolher algumas cidades próximas e realizar a pesquisa com os eleitores desses locais ou dessas cidades.

Amostragem de resposta voluntária ou auto-seleção:

É um tipo de amostragem na qual os membros de uma população decidem se participam ou não da pesquisa. Um exemplo são as pesquisas de sites na internet, programas de TV, etc, que realizam uma ou mais perguntas, cabendo ao indivíduo decidir se responde ou não. Essa amostra não é representativa, pois os membros que participam da pesquisa respondendo às questões tendem a apresentar opiniões semelhantes, em comparação ao restante da população.

É válido ressaltar a importância de saber se uma pesquisa divulgada nas mídias sociais foi feita seguindo um critério de amostragem, pois uma amostra pode ser escolhida de diversas maneiras, dependendo do foco da pesquisa, ou seja, o essencial é que a amostra atenda às expectativas feitas a partir dela para uma determinada análise estatística. Para que isso seja possível, inúmeras características devem ser devidamente observadas como aleatoriedade; ter um tamanho adequado; não ser tendenciosa (escolher de maneira metódica um certo grupo em uma população); e outros.

Até o devido momento, abrangemos alguns conceitos Estatísticos, assim como relatamos sobre dados e como coletá-los. O que nos leva à próxima parte do nosso trabalho, ou seja, resumir e organizar estes dados.

2.2. Apresentação de dados

Os dados podem ser apresentados de diversas formas, sendo que algumas delas facilitam a extração de certas informações.

Para descrever cada item dessa Subseção, consideramos os dados provenientes de uma amostra hipotética de 40 professores do PROFMAT, em quilogramas, apresentados no Quadro 1.

Quadro 1 - Pesos em quilogramas de uma amostra de 40 professores.

71	62	68	68	66	68	63	82	66	71
76	74	80	61	78	65	63	65	70	64
73	62	72	78	74	65	70	82	66	75
68	70	73	78	71	65	70	74	77	75

Fonte: O autor.

2.2.1. Rol

Quando a variável é quantitativa, utiliza-se o rol, que é uma forma relativamente simples para organização desses dados brutos. Um rol pode ser crescente ou decrescente. Abaixo, faremos um rol na sua forma crescente, relativo aos dados dos pesos de uma amostra de 40 professores.

61, 62, 62, 63, 63, 64, 65, 65, 65, 65, 65, 65, 66, 68, 68, 68, 68, 70, 70, 70, 70, 71, 71, 71, 71, 72, 73, 73, 74, 74, 74, 75, 75, 76, 77, 78, 78, 78, 80, 82, 82.

É possível perceber que algumas características das variáveis se tornam mais evidentes após a criação do rol, como por exemplo: o menor peso; o maior peso; a amplitude dos dados ($82 - 61 = 21$ kg); o valor que mais surge (o 65 kg, com 6 ocorrências); o fato de que a frequência de alunos com 80 kg ou mais é bem pequena; entre outros. Para as variáveis qualitativas, é possível construir um rol em ordem temporal ou alfabética, por exemplo.

2.2.2. Tabela

Uma tabela pode ser definida como uma forma de se agrupar sistematicamente dados numéricos em linhas e colunas, possibilitando dessa maneira sua comparação.

Cada linha (ou coluna) de uma tabela representa um indivíduo (ou objeto observado) e cada coluna (ou linha) da tabela representa uma informação do conjunto de dados. Vale destacar que tanto a elaboração de tabelas, quanto a elaboração de quadros, seguem normas da Associação Brasileira de Normas Técnicas (ABNT) e do IBGE.

Como exemplo, os dados do Quadro 1 podem ser apresentados na Tabela 1.

Tabela 1 - Pesos de uma amostra de 40 professores.

Peso (kg)	Nomes (S)
71	A
62	B
68	C
68	D
66	E
68	F
63	G
82	H
66	I
71	J
76	K
74	L
80	M
61	N
78	O
65	P
63	Q
65	R
70	S
64	T
73	U
62	V
72	W
78	X
74	Y
65	Z
70	AA
82	BB
66	CC
75	DD
68	EE
70	FF
73	GG
78	HH
71	II
65	JJ
70	KK
74	LL
77	MM
75	NN

$\Sigma S = 40$

Fonte: O autor.

2.2.3. Distribuição de Frequência ou Tabela de Frequência:

O rol torna-se insuficiente quando contém vários elementos e muitas repetições de dados, pois a organização desses é inviável. Assim, torna-se eficaz organizarmos a distribuição de frequências em tabelas. Estas podem representar tanto valores agrupados em classes quanto valores pontuais. A seguir construiremos uma distribuição de frequências do conjunto de dados da Tabela 1. Porém, definiremos antes algumas de suas partes:

Classes

Na estatística, quando uma variável possui diversos valores, precisamos agrupá-los em intervalos de números reais, a esses intervalos chamamos de *classes*.

Número de classes:

O número de classes varia em função de diversos fatores. Dentre eles podemos citar os objetivos, a natureza e a unidade utilizada para expressar as variáveis. Porém, é importante o número adequado de classes. Triola (2014, p. 40), sugere que “o número de classes deve estar entre 5 e 20, e o número que você escolher deve ser influenciado pela conveniência de se usar números redondos”. Isso acontece, pois se utilizarmos um número muito pequeno de classes, é perdido um número considerável de informações. Ao contrário, se utilizarmos muitas classes, pelo menos alguma dessas apresentará uma frequência nula ou muito pequena, apontando assim uma distribuição irregular e prejudicial à análise. Já Crespo (1997) afirma que podemos utilizar uma regra denominada de “Regra de Sturges”, em que o número de classes i é dado pela fórmula:

$$i \cong 1 + 3,3 \cdot \log n,$$

Sendo n o número de observações no conjunto de dados, ou seja, o tamanho da amostra. Utilizando a Regra de Sturges no nosso exemplo temos:

$$\begin{aligned} i &\cong 1 + 3,3 \cdot \log 40 \\ i &\cong 6,2 \end{aligned}$$

Como a Regra de Sturges indica, o nosso resultado foi de 6,2. Decidimos então permanecer com 6 classes. Com essa escolha, perdemos um pouco em detalhes, mas temos um ganho em simplicidade.

Limites de classe:

São os extremos de cada classe. O menor valor é o limite inferior da classe (l_k) e o maior é o limite superior da classe (L_k).

Sendo que a letra k designa qual a classe que está sendo considerada. Por exemplo, o limite inferior da classe 4 é designado por l_4 , assim como $k = 4$ designa a quarta classe.

Amplitude total da distribuição e Amplitude amostral:

A diferença entre o limite superior da última classe e o limite inferior da primeira classe nos dá a chamada amplitude total da distribuição, que pode ser designada por AT :

$$AT = L_k - l_1$$

Ao passo que a amplitude amostral é obtida através da diferença entre o valor máximo e o mínimo da amostra, que pode ser designada como AA :

$$AA = x(\text{máx.}) - x(\text{mín.})$$

Logo, no Rol apresentado anteriormente, possuímos:

$$\begin{aligned} x(\text{máx.}) &= 82 \\ x(\text{mín.}) &= 61 \\ AA &= 82 - 61 = 21 \end{aligned}$$

Amplitude de classe:

O cálculo da amplitude de classe pode ser realizado pela fórmula a seguir, onde h = amplitude de classe e i = número de classes.

$$h \cong \frac{AA}{i}$$

$$h \cong \frac{x(\text{máx.}) - x(\text{mín.})}{i}$$

$$h \cong \frac{82 - 61}{6} = 3,5$$

Utilizaremos $h = 4$, pois o nosso resultado é um valor não inteiro, ou seja, $h = 3,5$, logo, podemos arredondar esse valor para o número natural acima por pura conveniência. Então, já que temos o número de classes (6) e a amplitude de classe (4), estamos aptos a construir a tabela de frequência.

Na primeira classe colocaremos $l_1 = 60$, e, somando a 60 a amplitude da classe ($60 + 4$) obtemos $L_1 = 64$. Assim, na primeira classe, teremos o intervalo 60 - 64, o que significa que serão inseridos todos os valores de 60 inclusive até 64 exclusive; isto é, intervalo fechado em 60 e aberto em 64, outras formas também são possíveis para essa representação de intervalos. Assim sendo, contabilizaremos para essa classe todos os valores maiores ou iguais a 60 kg e menores que 64 kg. Ocorre o mesmo nos limites superiores e inferiores de todas as classes.

Frequência simples ou absoluta de uma classe:

É o número de elementos pertencentes a uma classe, ou seja, é o número de vezes em que cada elemento se repete na amostra.

Usaremos como símbolo para frequência absoluta f_k , no qual o índice k indica a classe. Por exemplo: f_3 indica a frequência da terceira classe, que vai de 68 inclusive a 72 exclusive, e apresenta os seguintes valores: 68; 68; 68; 68; 70; 70; 70; 70; 71; 71; 71. Logo, $f_3 = 11$.

A soma das frequências absolutas é igual ao número total de observações:

$$\sum f_k = 40$$

E, finalmente, chegamos no nosso objetivo com a organização dos dados em classes, apresentado na Tabela 2.

Tabela 2 - Frequência dos pesos de uma amostra de 40 professores.

Peso (kg)	Frequência
60 - 64	5
64 - 68	8
68 - 72	11
72 - 76	8
76 - 80	5
80 - 84	3
$\Sigma f_k = 40$	

Fonte: O autor.

Ponto médio de uma classe:

O ponto médio de uma classe divide o intervalo da classe em duas partes iguais, ou seja, refere-se à média aritmética dos limites da mesma. Podemos designá-lo por x_k .

$$x_k = \frac{Lk + lk}{2}$$

Exemplo:

O ponto médio da segunda classe:

$$x_2 = \frac{68 + 64}{2}$$

$$x_2 = \frac{132}{2}$$

$$x_2 = 66$$

Tipos de Frequências:**Frequência simples ou absoluta de uma classe:**

São os valores absolutos que representam a quantidade de dados em uma classe, conforme já demonstrado.

Na segunda classe temos: $f_2 = 8$.

Frequência relativa de uma classe:

É a razão entre a frequência absoluta e a frequência total. Utilizaremos, o símbolo f_{rk} :

$$f_{rk} = \frac{f_k}{\sum f_k}$$

Pode-se expressar esse resultado em termos percentuais multiplicando a frequência relativa por 100.

Na segunda classe temos:

$$f_2 = \frac{8}{40} = 0,2 = 20\%$$

Frequência absoluta acumulada:

É a soma das frequências absolutas até uma considerada classe, inclusive a frequência dessa classe. Utilizaremos o símbolo F_k . Portanto, temos:

$$F_k = f_1 + f_2 + f_3 + \dots + f_k$$

Logo, a frequência acumulada na segunda classe será:

$$\begin{aligned} F_2 &= f_1 + f_2 \\ F_2 &= 5 + 8 \\ F_2 &= 13 \end{aligned}$$

Frequência relativa acumulada:

É a razão entre a frequência acumulada de uma classe e a frequência total da distribuição. Utilizaremos, como símbolo F_{rk} :

$$F_{rk} = \frac{F_k}{\sum f_k}$$

A frequência acumulada relativa na segunda classe:

$$F_{r2} = \frac{F_2}{\sum f_2} = \frac{13}{40}$$

$$F_{r2} = 0,325 = 32,5\%$$

Densidade de frequência:

A densidade de frequência de cada classe é obtida dividindo-se a frequência relativa da classe por sua amplitude. Ideal quando as amplitudes das classes são diferentes. Utilizaremos, o símbolo d_{fk} :

$$d_{fk} = \frac{f_{rk}}{h}$$

Na segunda classe temos:

$$d_{f2} = \frac{0,2}{4} = 0,05$$

Construímos, então, a Tabela 3 a partir desses resultados.

Tabela 3 - Tabela de frequências dos pesos de uma amostra de 40 professores.

Classe	Peso (kg)	f_k	F_k	f_{rk}	F_{rk}	d_{fk}
1	60 F 64	5	5	0,125	0,125	0,03125
2	64 F 68	8	13	0,200	0,325	0,05000
3	68 F 72	11	24	0,275	0,600	0,06875
4	72 F 76	8	32	0,200	0,800	0,05000
5	76 F 80	5	37	0,125	0,925	0,03125
6	80 F 84	3	40	0,075	1,000	0,01875
		$\Sigma f_k = 40$		$\Sigma f_{rk} = 1$		

Fonte: O autor.

Ao analisarmos a Tabela 3, observam-se certas características dos dados que antes não ficavam explícitas, tal como a estimativa de que 32,5% dos professores têm menos de 70 kg.

2.2.4. Gráficos

Faremos agora uma breve apresentação sobre gráficos, sendo apenas apresentados os gráficos essenciais para entendimento dos nossos Estudos de Casos, ou seja, para o bom entendimento desse trabalho. Vale destacar que o conteúdo apresentado nesta Subseção contém os gráficos mais comumente utilizados pelas mídias sociais.

Segundo Reis (2008, p. 26):

A representação gráfica dos dados estatísticos tem por finalidade dar uma ideia, a mais imediata possível dos resultados obtidos permitindo chegar em conclusões rápidas sobre a relação entre os diferentes valores apresentados. Para que tal seja conseguido, quando se constrói um gráfico deverá ter-se em conta os elementos: simplicidade, clareza e veracidade.

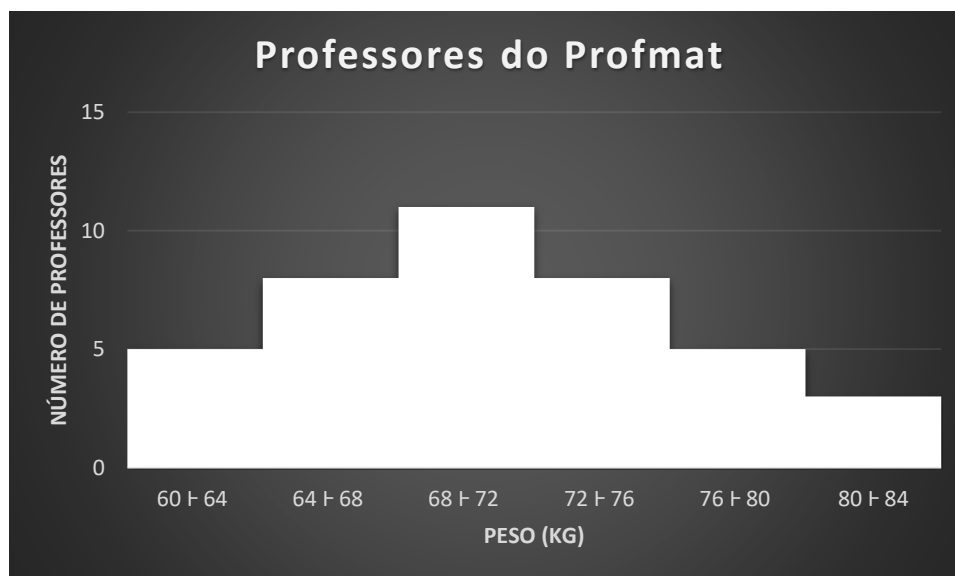
Gráficos para Variáveis Quantitativas Contínuas

Histograma:

Conforme Morettin e Bussab (2010, p. 18) “o histograma é um gráfico de barras contíguas, com as bases proporcionais aos intervalos das classes e a área de cada retângulo proporcional à respectiva frequência”. Pode-se usar tanto a frequência absoluta f_k , como a relativa f_{rk} ou a densidade de frequência.

O histograma é a ferramenta gráfica mais adequada quando temos um conjunto com um número elevado de dados quantitativos, sendo que a distribuição de frequências é agrupada em classes para representar as variáveis quantitativas. Por exemplo, no gráfico da Figura 2, temos que $f_2 = 8$ e $\sum f_k = 40$. Em termos de porcentagem $\frac{8}{40} = 20\%$; logo, a área da coluna que representa a segunda classe é 20% da área total do gráfico.

Figura 2 - Histograma relativo ao peso dos professores do PROFMAT em números absolutos.

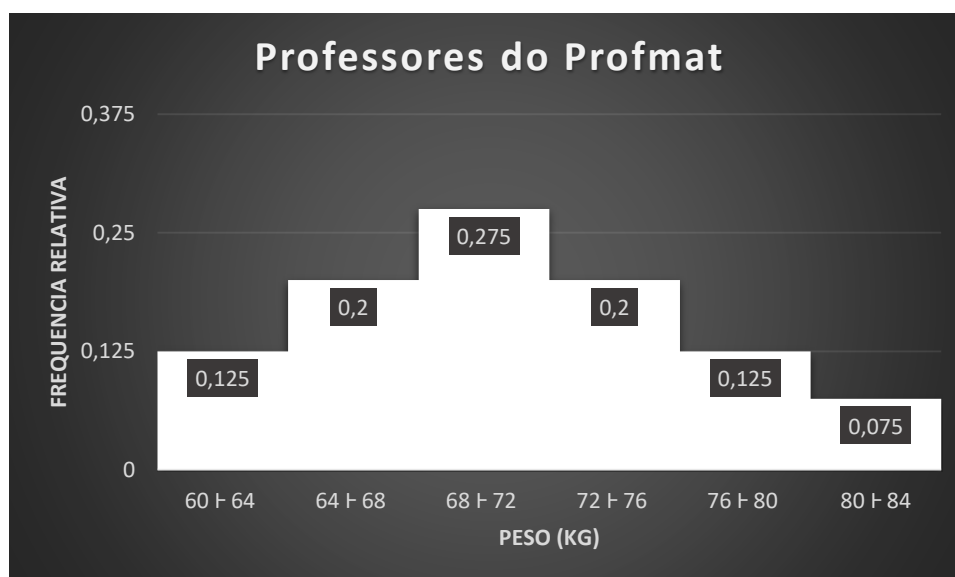


Fonte: O autor.

Histograma de frequência relativa:

É um tipo de histograma análogo ao da Figura 2, que se distingue por trazer frequências relativas em sua escala vertical ao invés das frequências absolutas. O gráfico da Figura 3 ilustra um histograma de frequência relativa.

Figura 3 - Histograma relativo ao peso dos professores do PROFMAT em frequência relativa.

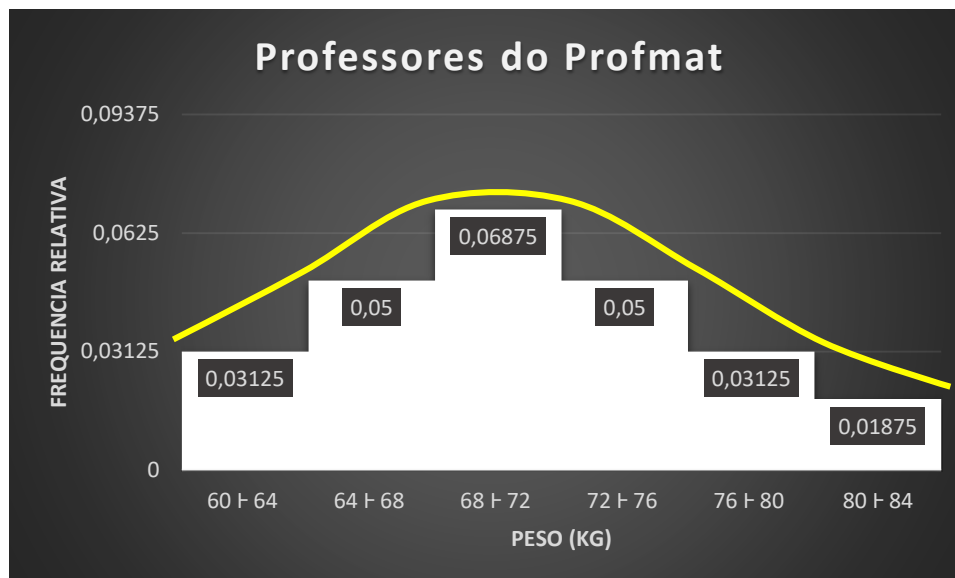


Fonte: O autor.

Histograma de Densidade de Frequência:

Os histogramas de densidade de frequência, assim como os histogramas de frequência relativa, aproximam-se de uma curva contínua para um número adequado de classes. Segundo (MORETTIN; BUSSAB, 2010), as alturas das classes no histograma de densidade de frequência, diferentemente dos histogramas de frequência relativa, não se aproximam de zero. Isso significa que, à medida que o número de classes aumenta, as alturas se mantêm relativamente constantes. Logo, é possível observar que conforme o número de classes se expande, a frequência relativa de cada classe diminui, conseqüentemente a amplitude da classe diminui proporcionalmente. Deste modo, a densidade de frequência acaba por tender a um valor constante. Dito isso, através do gráfico da Figura 4, é possível constatarmos o histograma de densidade de frequência. Uma característica importante desse gráfico é que a sua área é sempre igual a 1. Isso permite compararmos histogramas de diferentes variáveis ou diferentes conjuntos de dados.

Figura 4 - Histograma de densidade de frequência relativo ao peso dos professores do PROFMAT. E a curva de densidade de probabilidade estimada da variável aleatória em questão.



Fonte: O autor.

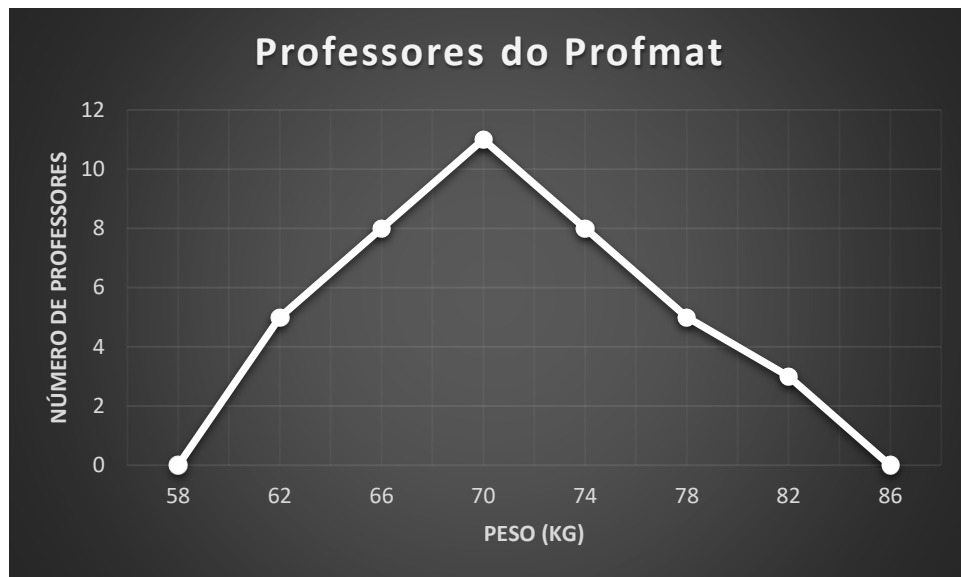
Polígono de Frequência:

O polígono de frequência é um gráfico em linha que, segundo Downing e Clark (2003, p. 23), “em lugar de uma barra para representar cada classe, colocamos um ponto acima do ponto médio da classe, e unimos esses pontos por segmentos de reta”. O objetivo de um polígono de frequência é esboçar a distribuição dos dados, sendo que a distribuição descreve matematicamente o comportamento dos dados.

É possível observar na Figura 5 que, para completarmos o polígono de frequência, é necessário ligarmos os extremos da linha obtida aos pontos médios da classe anterior à primeira e posterior à última, que em verdade são duas classes “virtuais”, e considerar em ambas que a frequência seja nula.

O polígono de frequência também pode ser considerado como sendo uma aproximação (grosseira) para a curva de densidade de probabilidade da variável aleatória em questão.

Figura 5 - Polígono de frequência relativo ao peso dos professores do PROFMAT.



Fonte: O autor.

Curva de densidade de probabilidade

A curva de densidade de probabilidade é o gráfico da função de densidade de probabilidade (fdp) de uma variável aleatória contínua. Para que uma função $f(x)$ seja uma fdp, $f(x)$ deve satisfazer (i) a (iii)

(i) $f(x) \geq 0,$

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$ e

(iii) $P(a \leq X \leq b) = \int_a^b f(x)dx = \text{área abaixo de } f(x) \text{ entre } a \text{ e } b.$

Podemos dizer que a fdp descreve matematicamente uma v.a. contínua e uma de suas importantes características é nos fornecer a probabilidade de uma v.a. contínua em certo intervalo da reta (ou reunião de intervalos). A curva (ou o gráfico) dessa função nos mostra como é o comportamento dessa probabilidade ao longo da reta, permitindo observarmos intervalos de menor ou maior probabilidade.

Vale ressaltar que não é possível obtermos a fdp exata de uma v.a. através de uma amostra, por isso costumamos estimar esta função. Na Figura 4 observamos a fdp aproximada (ou estimada) da variável aleatória peso dos professores.

Gráficos para Variáveis Qualitativas

Segundo Morettin e Bussab (2010, p. 15) “existem vários tipos de gráficos para representar variáveis qualitativas. Vários são versões diferentes do mesmo princípio”. Logo, nos limitaremos a apresentar apenas um deles, ou seja, gráfico em barras.

Gráfico de Barras:

O gráfico de barras pode ser utilizado em variáveis qualitativas ou quantitativas discretas.

Consiste em construir barras retangulares, em que um dos eixos do gráfico em questão é proporcional à frequência da variável em análise, enquanto a outra, é arbitrária, porém, igual para todos os retângulos. As barras são paralelas entre si, podendo ser apresentadas tanto horizontalmente quanto verticalmente. No caso de barras horizontais, costumamos chamá-lo de gráfico de colunas. Por exemplo, imaginemos que fossem realizadas duas perguntas para a amostra dos professores do PROFMAT:

- Qual o número de escolas em que você trabalha?
- Dentre os níveis de escolaridade, fundamental, médio ou técnico, qual o seu preferido para trabalhar? (Nota: Somente uma pode ser escolhida).

Após serem organizadas as respostas do questionário, foram obtidas as Tabelas 4 e 5 e, conseqüentemente, os gráficos das Figuras 6 e 7.

Tabela 4 - Números de escolas.

Nº de escolas	Nº de professores
1	2
2	11
3	16
4	8
5	3
Total	40

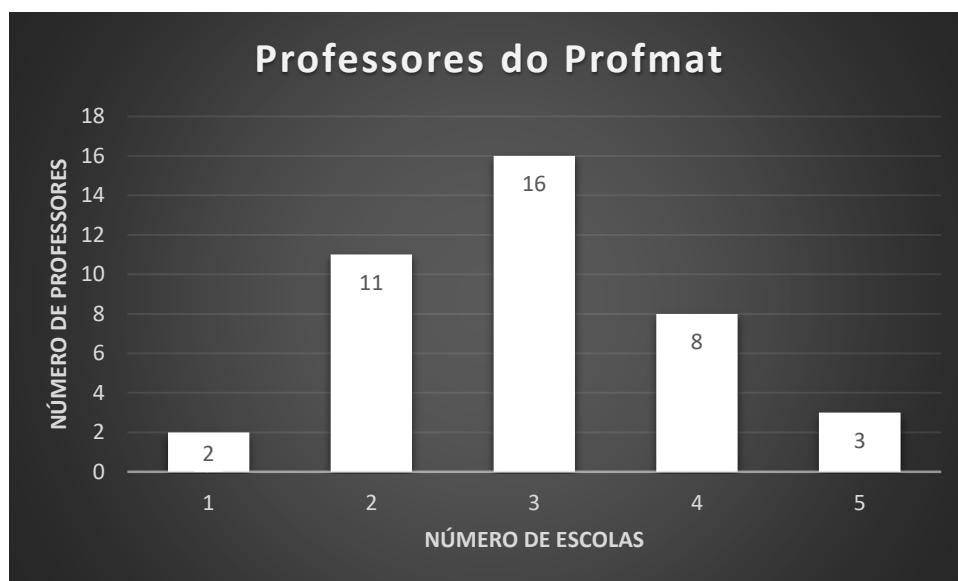
Fonte: O autor.

Tabela 5 - Nível de escolaridade preferido.

Nível Preferido	Nº de professores
Fundamental	16
Médio	13
Técnico	11
Total	40

Fonte: O autor.

Figura 6 - Gráfico de barras relativo ao número de escolas.



Fonte: O autor.

Figura 7 - Gráfico de barras relativo ao nível preferido.



Fonte: O autor.

Finalizamos a parte da revisão teórica. E, a partir de agora, no próximo Capítulo 3, iniciaremos os Estudos de Caso.

3. ESTUDOS DE CASO

Depois de apresentarmos alguns conceitos iniciais da Estatística e expormos a teoria referente à apresentação dos dados, enfim chegamos no foco do nosso trabalho: gráficos.

A seguir, discutiremos exemplos de gráficos estatísticos retirados das mídias de comunicação. Nesses gráficos não analisaremos, por exemplo, possíveis casos de falsificação, o que na prática seria de difícil averiguação. O nosso foco será em torno de distorções na apresentação, elaboração e problemas na construção dos gráficos.

A partir deste momento, relataremos um material que acreditamos ser útil tanto em uma aula de estatística, assim como para qualquer pessoa que tenha o interesse ou a necessidade de tal conhecimento, assumindo assim uma postura dinâmica e crítica perante os mesmos. Abordaremos três gráficos procurando verificar características, tais como: erros, interpretação adequada de dados, etc.

Dos Reis *et al.* (2020) afirmam que, ao ficar exposto a informações Estatísticas veiculadas pela mídia, o cidadão, desprovido de conhecimento na área, pode consumi-las sem uma filtragem. Isso, conseqüentemente, torna-o vulnerável a interpretações que não fazem jus à realidade.

Iniciemos nossa análise com um exemplo recorrente nas mídias sociais, independente da época, que é a inflação. Gráficos deste tipo devem ser muito bem elaborados, ou seja, evitando por exemplo distorções, para que dessa maneira não haja nenhuma dúvida quanto ao aumento ou queda da inflação.

Prosseguiremos agora com os gráficos analisados.

3.1. Gráfico 1: Inflação brasileira de 2013

No início de 2014, segundo matéria do blog Brasil Acadêmico “Gráficos que mentem: Manipulando gráficos e dados”, foi mostrado pelo canal Globo News o gráfico de barras apresentado na Figura 8, que pretendia mostrar a inflação brasileira referente aos últimos 5 anos.

Figura 8 - Gráfico de barras da Globo News da Inflação do Brasil.



Fonte: <http://blog.brasilacademico.com/2014/04/graficos-que-mentem-manipulando.html>
Acessado em: 19/03/21.

Analisando o gráfico da Figura 8, podemos observar que os valores indicados no topo das barras não correspondem ao tamanho do comprimento das mesmas. Nota-se, por exemplo, que a barra do ano de 2013, cujo valor de 5,91% é maior que a barra do ano de 2011, que está com o valor 6,50%, e maior ainda que a barra de 2010, que tem o valor de 5,92%.

Além disso, também é válido destacar que o comprimento da segunda barra, que representa o valor de 5,92%, é praticamente mais que o dobro do tamanho do comprimento da primeira barra que representa o valor de 4,31%, sendo a diferença entre esses valores relativamente pequena. O objetivo da comparação entre a medida do comprimento das barras é enfatizar as diferenças entre os valores numéricos (porcentagens) representados por cada ano pesquisado. De acordo com o gráfico da Figura 9, torna-se fácil constatar, visualmente, que essa diferença se acentua ainda mais, por consequência da incompatibilidade entre os rótulos mostrados e os elementos gráficos.

Figura 9 - Gráfico de barras da Inflação com destaque para um dos erros.



Fonte: <http://blog.brasilacademico.com/2014/04/graficos-que-mentem-manipulando.html> modificada pelo autor.
Acessado em: 19/03/21.

Dessa forma, também é possível destacar a ausência do eixo vertical, da escala, bem como a omissão do zero. A falta destes elementos em determinados gráficos de barras impossibilita a elaboração de parâmetros sistemáticos de comparação entre as barras, podendo ocasionar assim uma distorção no entendimento da informação, conforme ilustrado na Figura 10.

Figura 10 - Gráfico de barras analisando eixo vertical, escala e o zero.



Fonte: <http://blog.brasilacademico.com/2014/04/graficos-que-mentem-manipulando.html>
 modificada pelo autor.
 Acessado em: 19/03/21.

Dessa maneira, é interessante afirmar que existe uma suspeita de que o eixo das ordenadas não se inicia no valor zero.

Ao começar as barras fora do valor zero, o gráfico impõe uma comparação apenas entre os topos das colunas. A partir de (GAL, 2002), é possível evidenciar que o autor foi incapaz de avaliar as proporções do comprimento de cada barra em relação aos rótulos dos dados. Note, ainda, que a imagem do gráfico está em perspectiva, ou seja, quanto mais à direita uma coluna, menor a sua dimensão, causando assim uma distorção. Desse modo, a observação da proporcionalidade entre as colunas se torna dificultada. Logo, no contexto dos gráficos de barras, a interrupção na escala do eixo vertical altera a informação, trazendo ao leitor uma apresentação totalmente distorcida dos dados, como mostrado no gráfico da Figura 10.

No âmbito dessa discussão, Toledo e Ovalle (1981, p. 80) ainda elucidam que:

Um gráfico [de colunas ou de barras], construído para mostrar grandezas absolutas, deverá ter uma linha zero claramente definida e uma escala de quantidades ininterrupta, caso contrário a leitura e a interpretação do gráfico poderão ficar distorcidas.

A respeito dessa questão, Crespo (2002, p. 40) esclarece:

O zero, de modo geral, deverá ser indicado sempre que possível, especialmente no eixo vertical. Se, por alguma razão, for impossível tal indicação e se essa omissão puder levar o observador a conclusões errôneas, é prudente chamar a atenção para a omissão.

Mas essa não é a única questão digna de nota na análise do gráfico da Figura 10. Gal (2002) aborda esta questão afirmando que:

Os criadores de mensagens podem ter diversos objetivos em termos dos fatos presumidos, imagens, ou conclusões que visam criar ou incutir na mente do leitor. Algumas mensagens podem ser criadas para convencer o leitor ou ouvinte a adotar um ponto de vista específico ou rejeitar outro, e, portanto, usar argumentos unilaterais ou apresentar informação seletiva. (GAL, 2002, p. 7).

O autor ainda chama a atenção para o fato de que “mensagens dirigidas aos cidadãos em geral podem ser moldadas por agendas políticas, comerciais ou outras” (GAL, 2002, p. 15).

Concluimos, então, que estamos abordando uma distorção muito grave, pois exagera as diferenças entre as barras do gráfico, fazendo-nos acreditar que a inflação subiu demais, dando verdadeiros saltos nos anos que se seguem.

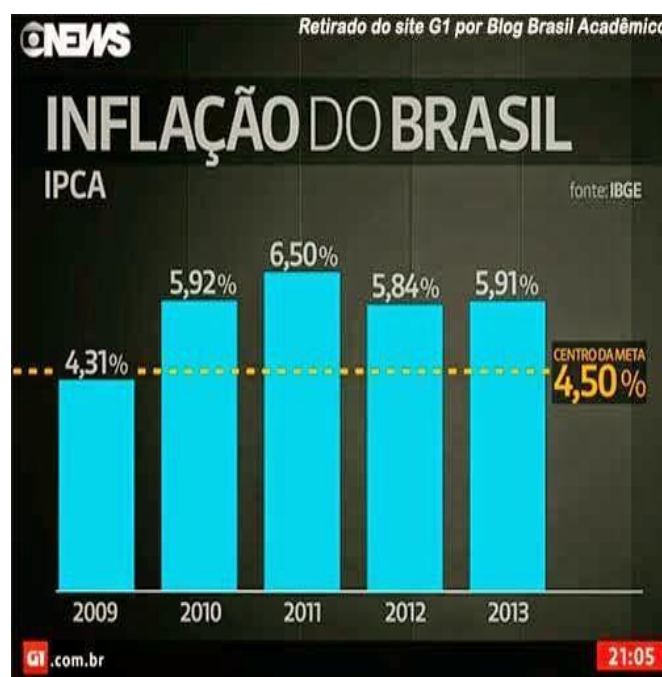
Na Figura 11, temos a representação do gráfico retificado dias depois pela emissora. A Figura 12 apresenta os dois gráficos juntos, ou seja, o gráfico primeiramente publicado pela reportagem e o corrigido. Comparando o gráfico da Figura 8 com o gráfico da Figura 11, fica explícita a diferença entre o gráfico tendencioso (errado) e o correto.

Figura 11 - Gráfico de barras da Inflação retificado pela emissora.



Fonte: <http://blog.brasilacademico.com/2014/04/graficos-que-mentem-manipulando.html>
Acessado em: 19/03/21.

Figura 12 - Comparação entre os gráficos publicados pela emissora.



Fonte: <http://blog.brasilacademico.com/2014/04/graficos-que-mentem-manipulando.html>
Acessado em: 19/03/21.

3.2. Gráfico 2: Classes de rendimento: A, B, C, D e E

No Brasil, o IBGE é o órgão responsável por realizar a Pesquisa Nacional por Amostragem de Domicílio (PNAD) que, segundo o próprio órgão:

É uma pesquisa por amostra probabilística de domicílios, de abrangência nacional, planejada para atender a diversos propósitos. Visa produzir informações básicas para o estudo do desenvolvimento socioeconômico do País e permitir a investigação contínua de indicadores sobre trabalho e rendimento. A Pnad Contínua segue um esquema de rotação de domicílios. Isso significa que cada domicílio selecionado será entrevistado cinco vezes, uma vez a cada trimestre, durante cinco trimestres consecutivos.

Principais Indicadores que serão produzidos com base na Pnad Contínua:

- * População residente segundo o sexo e os grupos de idade;
- * Taxa de desocupação;
- * Taxa de atividade;
- * Nível da ocupação;
- * Taxa de analfabetismo segundo os grupos de idade e o sexo;
- * Pessoas de 14 anos ou mais segundo a condição de ocupação;
- * Pessoas ocupadas na semana de referência segundo o sexo e os grupos de anos de estudo;
- * População residente segundo a naturalidade em relação à Unidade da Federação e ao município de residência;
- * Rendimento médio mensal per capita dos domicílios.

Diante dessa explicação sobre a Pnad, podemos dar sequência ao assunto.

No final de 2019, a GloboNews e o g1 publicaram uma reportagem denominada: “Estudo mostra que concentração de renda no Brasil pode ser quase o dobro da apurada pelos dados oficiais”, em que se utilizava a apuração feita pela Pnad (pesquisa com base em entrevistas pessoais feitas em domicílios) para analisar a desigualdade social no Brasil mediante a diferença da renda entre as classes sociais, onde:

- A classe A, mais alta, que possui rendimentos mensais superiores a R\$ 18462,00.

- A classe B, que representa famílias brasileiras com renda mensal entre R\$ 5929,00 e R\$ 18468,00.

- A classe C, que representa famílias com uma renda mensal que varia de R\$ 2459,00 até R\$ 5929,00.

- E as classes D e E, que representam famílias com renda mensal no valor de até R\$ 2459,00.

O responsável no Brasil pelos cálculos dos dados oficiais de renda e desigualdade é o IBGE, com base nos números da Pnad, que coleta as informações por meio de entrevistas pessoais feitas em domicílios. A partir deste levantamento, o IBGE estima que, da classe A para a classe D/E, existe uma enorme diferença de renda, cerca de 21,4 vezes, ou seja, os mais ricos ganham 21,4 vezes mais do que recebem os mais pobres. Sabemos que o IBGE é um órgão confiável. Porém, observando os valores mostrados anteriormente e sabendo que a Pnad tem como base entrevistas pessoais, torna-se relevante considerarmos um ponto muito importante, que é talvez o entrevistado não declarar todos os ganhos adquiridos, principalmente os entrevistados que têm a maior renda, ou seja, os mais ricos, que também podem receber, por exemplo, bônus, renda de aluguel e dividendos.

Triola (2014) faz comentários a respeito das escolhas em um processo estatístico. Ele denomina um desses comentários como “resultado informado”, que aparentemente pode elucidar os questionamentos aqui abordados:

Ao coletar dados de pessoas, é melhor que você mesmo faça as medidas, em vez de pedir que os sujeitos as relatem. Pergunte as pessoas quanto elas pesam e, provavelmente você receberá as medidas desejadas, não as medidas reais. Se você realmente deseja pesos precisos, use uma balança e pese as pessoas. (TRIOLA, 2014, p. 16).

Em entrevista à GloboNews e g1, a economista e sócia da consultoria Tendências, Alessandra Ribeiro, diz que “a Pnad subestima muito a renda das faixas mais altas”. A economista ainda afirma que “quando é feito o ajuste da massa de renda

com dados da Receita, o que se observa é que a diferença da classe A para a classe D e E é muito maior do que os dados do IBGE mostram”.

No recorte por faixas de renda, a consultoria Tendências utiliza os dados da Pnad para quem ganha até cinco salários mínimos, pois a maioria dessa parte da população é isenta da declaração do Imposto de Renda (IR). Já, para rendas acima de cinco salários mínimos, são utilizados dados recebidos pela Receita Federal, ou seja, declarados no IR. A partir dessa metodologia, é possível observar que a diferença de renda entre a classe A e a D/E é quase o dobro da obtida pela Pnad, ou seja, sobe de 21,4 para 38,8 vezes. Logo, fica clara a subestimação da renda pela Pnad, como demonstrada no gráfico da Figura 13.

Figura 13 - Comparação entre o gráficos da Pnad e da Receita Federal.



Fonte: <https://g1.globo.com/economia/noticia/2019/10/30/estudo-mostra-que-concentracao-de-renda-no-brasil-pode-ser-quase-o-dobro-da-apurada-pelos-dados-oficiais.ghtml>

Acessado em: 15/04/21.

A mesma discrepância ocorre quando olhamos para a distribuição feita por faixas de renda, pois nesta distribuição, os dados realizados pelo IBGE mostram, por exemplo, que a massa de renda, ou seja, todas as rendas somadas, é de R\$ 14,671 bilhões entre os que ganham acima de 160 salários mínimos. Já, de acordo com a pesquisa da consultoria Tendências, baseada nos dados da Receita Federal, esse valor aumenta para R\$ 397,920 bilhões, mostrando dessa maneira uma diferença de 2.612,4%.

Logo, entre os que têm renda de 80 a 160 salários mínimos, a diferença entre os dois levantamentos, Pnad e Pnad ajustada (que utiliza os dados da Receita Federal) chega a 1.454,8%. Ainda é possível observar na Figura 14, que a massa total do país de acordo com os dados da Receita Federal, ajustados para as classes mais altas, possui um valor maior que 50% tendo em vista o levantamento da Pnad.

Figura 14 - Massa apurada pela Pnad e Massa ajustada com a Receita.

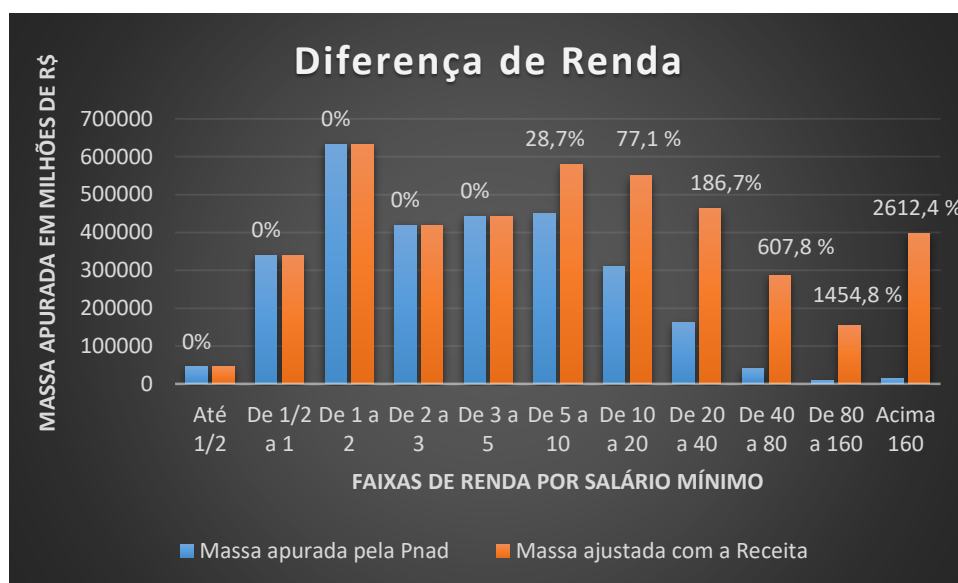
Faixas de renda por salário mínimo	Massa apurada pela Pnad (em milhões de R\$)	Massa ajustada com a Receita (em milhões de R\$)	Diferença
Até ½	44.778	44.778	0%
De ½ a 1	338.521	338.521	0%
De 1 a 2	633.012	633.012	0%
De 2 a 3	418.980	418.980	0%
De 3 a 5	442.967	442.967	0%
De 5 a 10	449.309	578.320	28,7%
De 10 a 20	310.878	550.443	77,1%
De 20 a 40	161.541	463.067	186,7%
De 40 a 80	40.651	287.731	607,8%
De 80 a 160	9.887	153.731	1.454,8%
Acima de 160	14.671	397.920	2.612,4%
Total	2.865.195	4.309.47	50,4%

Fonte: <https://g1.globo.com/economia/noticia/2019/10/30/estudo-mostra-que-concentracao-de-renda-no-brasil-pode-ser-quase-o-dobro-da-apurada-pelos-dados-oficiais.ghtml>

Acessado em: 15/04/21.

Para entendermos melhor a situação descrita acima e representada pela tabela da Figura 14, façamos um exercício, ou melhor, uma construção de gráficos, possibilitando dessa maneira uma análise e uma visualização ainda melhor e mais interessante da discrepância envolvida e retratada pela reportagem, como destacada na Figura 15.

Figura 15 - Gráfico da Diferença de Renda em porcentagem.



Fonte: O autor.

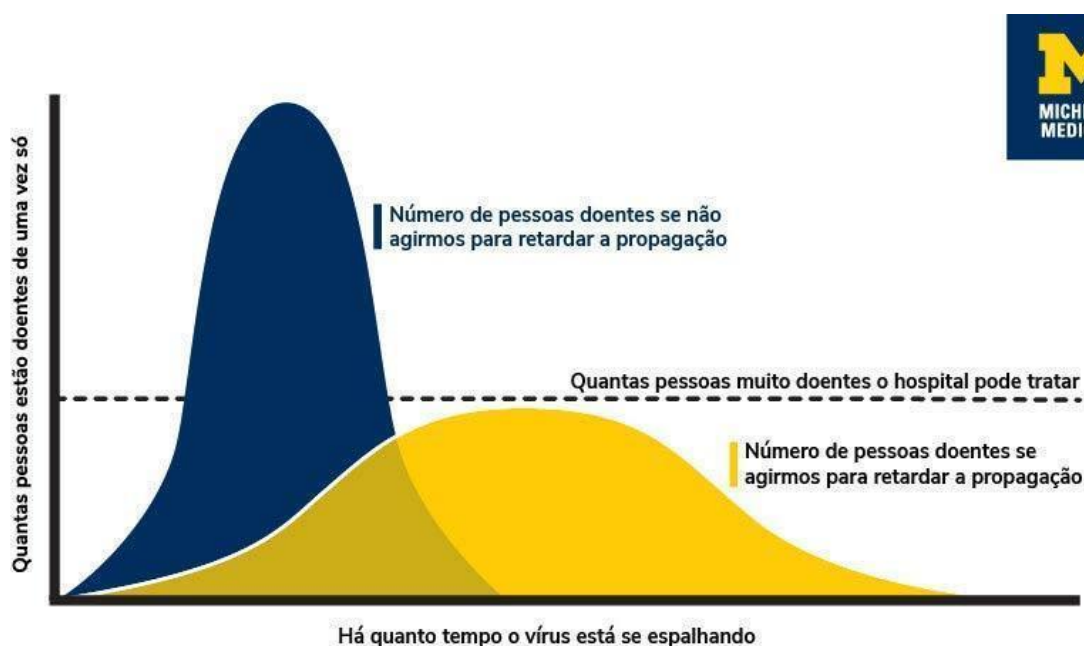
Note que o problema não é necessariamente a forma como se amostram as observações, e sim a fonte da informação. Dessa forma, os dados podem ser distorcidos, influenciando assim nos valores que fazem parte da construção do gráfico, bem como na conclusão por parte do leitor, fato esse observado nos gráficos tanto a respeito da desigualdade quanto em relação ao da massa de renda. Enfim, a partir de todo esse olhar crítico, é possível sugerir algo muito alarmante, ou seja, o abismo retratado entre a classe A em comparação com as demais classes, que já era apontado na Pnad como extremamente significativo, torna-se ainda maior na Pnad ajustada.

3.3. Gráfico 3: “Achatamento da curva” do coronavírus.

A partir de 2020, devido à pandemia da covid-19, a população mundial tem vivido situações adversas. Viagens canceladas, hotéis fechados, universidades e escolas trabalhando em *home office*, ou seja, tudo que envolva uma grande concentração de pessoas em um determinado local foi cancelado ou adiado como, por exemplo ainda, shows e eventos esportivos.

Durante esse período, uma expressão vem se destacando entre as mídias sociais a respeito da contenção do coronavírus, que é o chamado “achatoamento da curva”. Segundo matéria publicada no site Olhar Digital (por Renato Santino), no dia 25 de março de 2020, o termo tem a ver com reduzir ao máximo o ritmo de transmissão do vírus, fazendo com que o número de casos ativos, que necessitem de hospitalizações, não supere o número de leitos hospitalares disponíveis, garantindo que todos que precisem tenham o atendimento adequado, minimizando o número de mortes. O gráfico da Figura 16, publicado pela Michigan Medicine, permite-nos retratar e aprofundar essa questão.

Figura 16 - Gráfico de comparação das curvas.

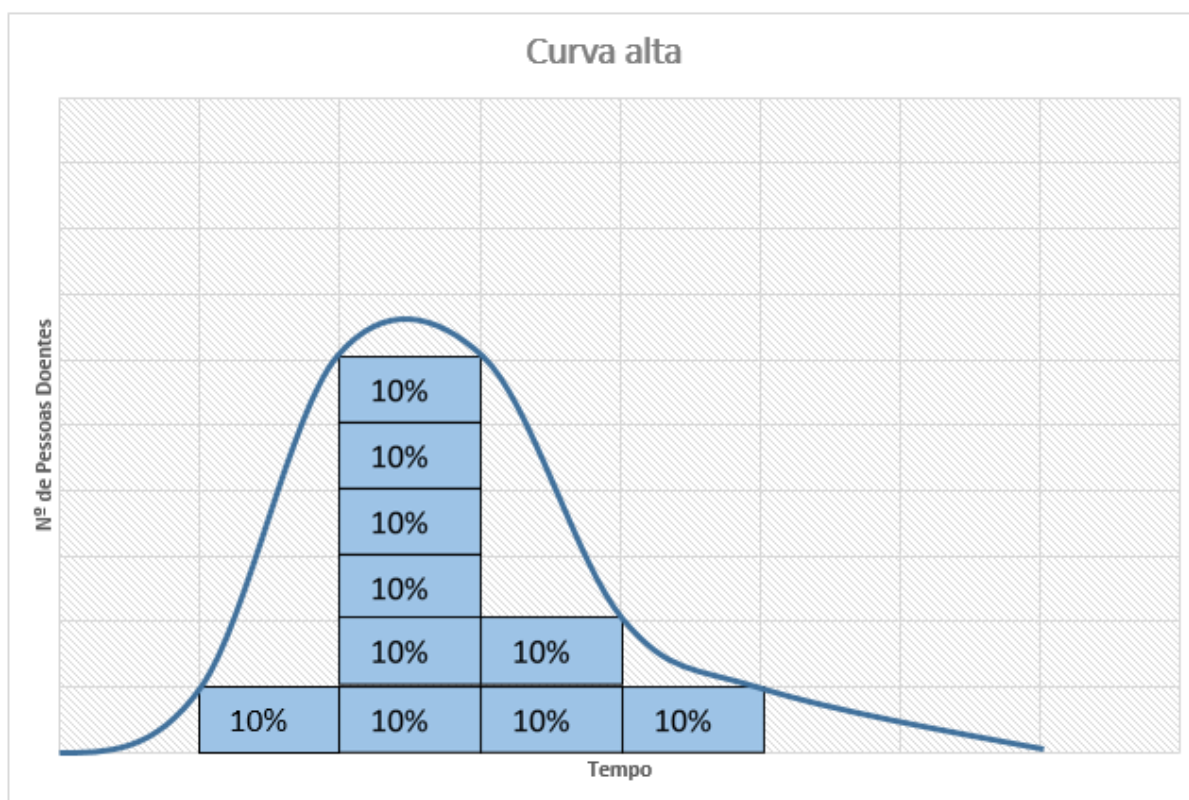


Adapted from the CDC

Fonte: <https://news.umich.edu/pt-br/achatando-a-curva-do-covid-19-o-que-significa-e-como-voce-pode-ajudar/> Acessado em: 29/05/21.

Para entender o que seria o “achamento da curva”, vamos lembrar que a área total abaixo da curva da fdp é igual a 1 (100%). Então poderíamos ter um histograma composto por 10 “bloquinhos” de 0,1 de área que fosse aproximado por esta curva. Se a curva é mais “alta”, significa que em um intervalo de tempo ao redor do ponto máximo dessa curva há mais bloquinhos do que em outros intervalos. A Figura 17 representa esta situação.

Figura 17 - Gráfico da curva alta em "bloquinhos": Ruim.

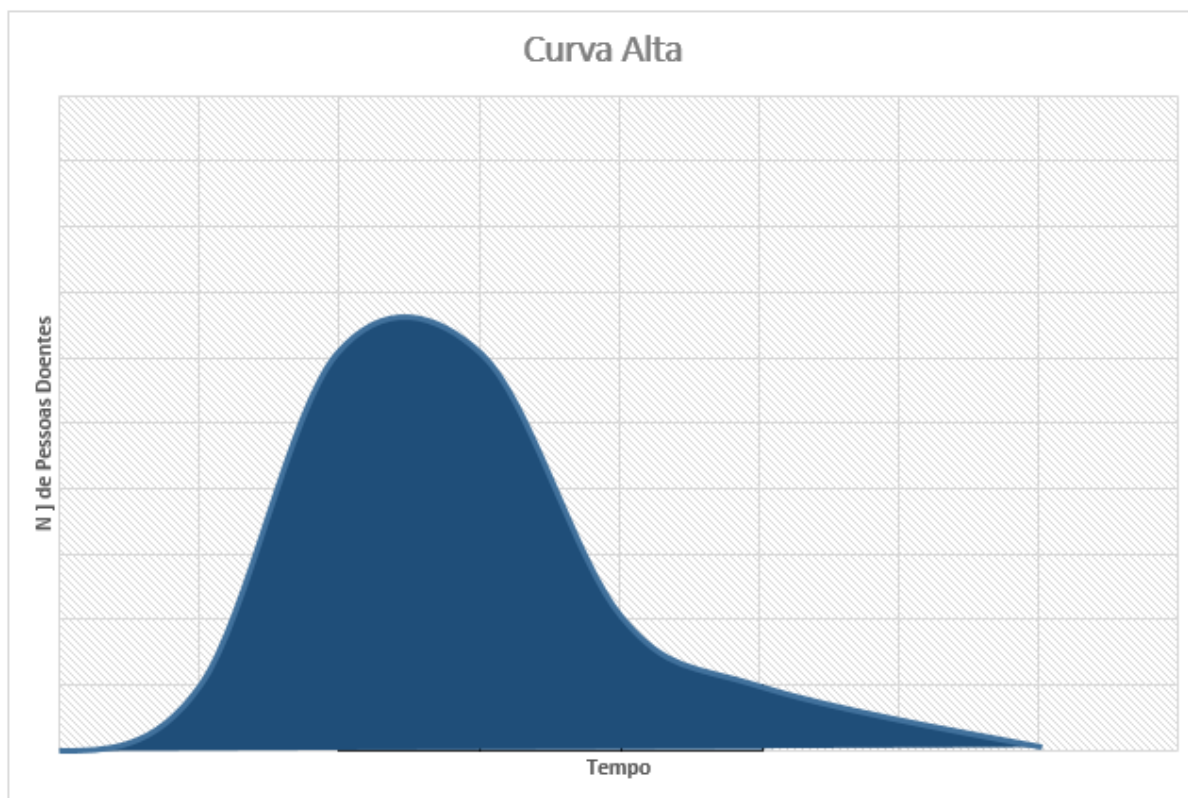


Fonte: O autor.

Se estamos analisando a proporção (nº de pessoas doentes) de pessoas com covid-19 ao longo do tempo, a Figura 17 nos permite observar que a curva alta é ruim, pois significa que muitas pessoas ficam doentes de uma só vez, em um curto período de tempo, o que acaba por sobrecarregar o sistema de saúde, que em dias normais já possui um elevado nível de atendimento, ou seja, leitos, respiradores, medicamentos e mão de obra hospitalar são insuficientes para atender a tal demanda.

Portanto, o gráfico da Figura 18 mostra uma curva “alta”, ruim para o caso real que representa.

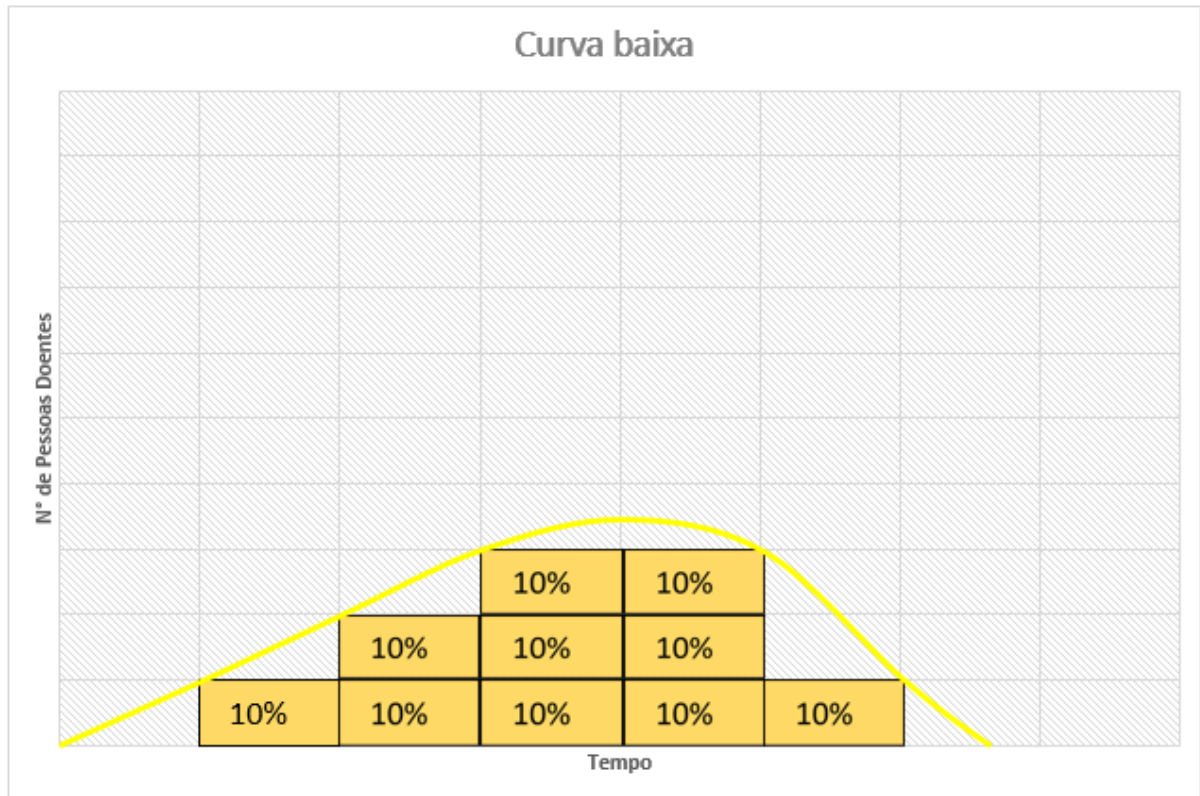
Figura 18 - Gráfico da curva alta: Ruim.



Fonte: O autor.

Já, a curva mais plana e mais baixa está representada pelo gráfico da Figura 19, tendo em vista a mesma ideia de “bloquinhos” descrita anteriormente.

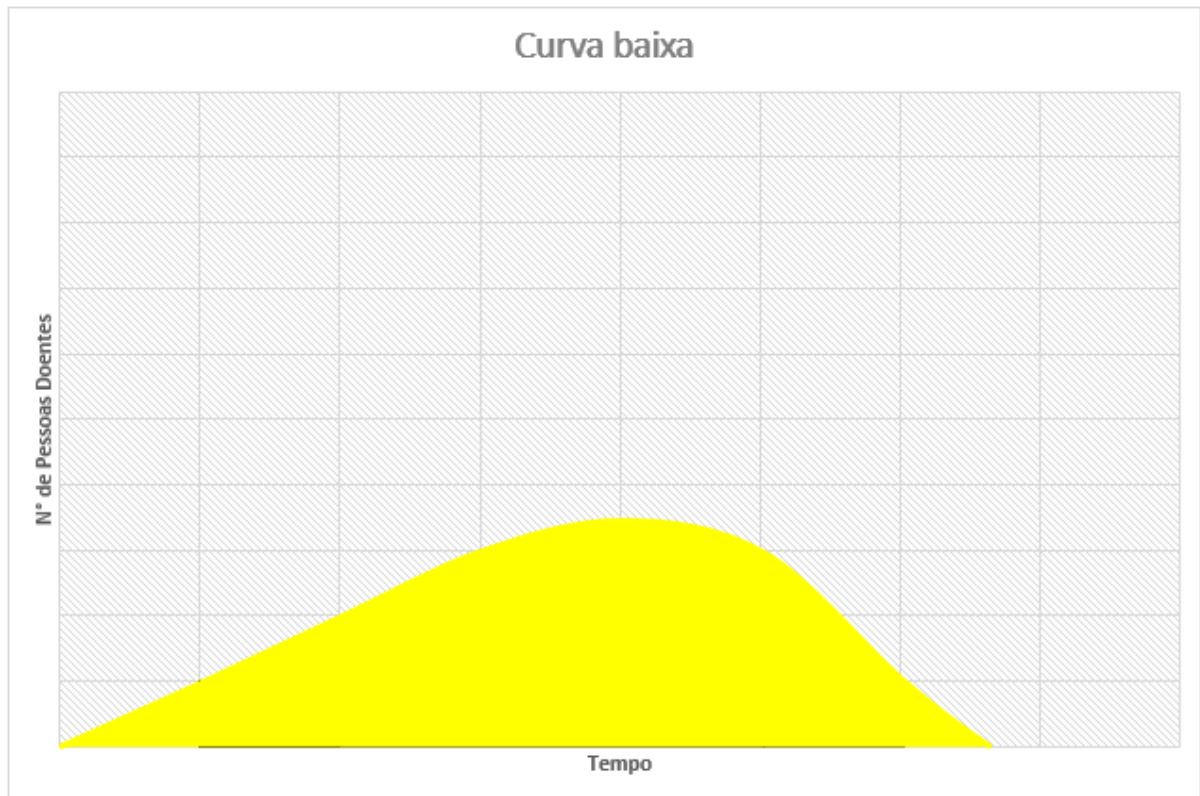
Figura 19 - Gráfico da curva baixa em "bloquinhos": Boa.



Fonte: O autor.

Essa curva representa uma situação melhor, pois significa que menos pessoas ficam doentes de uma só vez, conseqüentemente o tempo é maior para que 100% das pessoas doentes sejam infectadas. Isso se dá pelas ações governamentais de combate à pandemia, vacinação, isolamento social, uso de máscaras, álcool em gel, ou seja, com esses fatores e o empenho de toda a população, essa curva fica mais baixa, isto é, "achatada", como mostra o gráfico da Figura 20.

Figura 20 - Gráfico da curva baixa: "Achatamento da Curva".



Fonte: O autor.

Portanto, torna-se evidente concluirmos que é de fundamental importância o "achamento da curva". Para isso, faz-se necessário o aprendizado básico da estatística pela sociedade, não só para o entendimento dos gráficos, mas também para o dia a dia de cada ser humano.

4. CONSIDERAÇÕES FINAIS

Hoje em dia, os dados estatísticos compreendem muitas das informações apresentadas no nosso cotidiano, sobretudo na forma de gráficos, ferramenta muito utilizada pela mídia, que em várias situações não contêm erros. E, em outros casos, mostram-se com determinados erros, cometidos de forma intencional, ou seja, de maneira tendenciosa, ou de forma não intencional, simplesmente pela falta de conhecimento técnico do indivíduo responsável pela análise e construção, tanto dos dados quanto dos gráficos conseqüentemente (TRIOLA, 2014).

A partir disso, percebe-se a importância do conhecimento estatístico para a formação do indivíduo, que o permitirá ser capaz de entender todo o saber que essa forma de comunicação exige. Isso será possível através do ambiente educacional, o qual fornecerá os saberes necessários para que ele tenha um olhar maduro, crítico e, conseqüentemente, cresça como indivíduo.

Como vimos durante todo o nosso trabalho, principalmente no último capítulo, esse material nos instiga a refletir, podendo assim gerar uma discussão saudável no âmbito escolar entre professores e alunos, que complementarmente o conteúdo denominado tratamento da informação. Dessa maneira, desenvolvendo conhecimentos para não só compreender, analisar e apreciar as estatísticas que são divulgadas pelas mídias sociais, mas também para um mais amplo reconhecimento das informações que são confiáveis (TRIOLA, 2014).

REFERÊNCIAS

- BRASIL ACADÊMICO. **Gráficos que mentem: Manipulando gráficos e dados**. 2014. Disponível em:
<<http://blog.brasilacademico.com/2014/04/graficos-que-mentem-manipulando.html>>.
Acesso em: 02 mar. 2022.
- BRASIL. Secretaria de Educação Fundamental. **Parâmetros curriculares nacionais: Matemática**. Brasília: MEC/SEF, 1998.
- CRESPO, A. A. **Estatística Fácil**. 15. ed. São Paulo: Editora Saraiva, 1997.
- CRESPO, A. A. **Estatística Fácil**. 17. ed. São Paulo: Editora Saraiva, 2002.
- DOS REIS, S. R. *et al.* **Educação Estatística – uma proposta de ensino**. Ciência e Natura, v. 42, p. 17. 2020.
- DOWNING, D.; CLARK, J. **Estatística Aplicada**. 2. ed. Tradução: Alfredo Alves de Farias. São Paulo: Editora Saraiva, 2003.
- GAL, I. **Adults' statistical literacy: meanings, components, responsibilities**. International Statistical Review, v. 70, n. 1, p. 1-25. 2002.
- GLOBO. **Estudo mostra que concentração de renda no Brasil pode ser quase o dobro da apurada pelos dados oficiais**. 2019. Disponível em:
<<https://g1.globo.com/economia/noticia/2019/10/30/estudo-mostra-que-concentracao-de-renda-no-brasil-pode-ser-quase-o-dobro-da-apurada-pelos-dados-oficiais.ghtml>>. Acesso em: 02 mar. 2022.
- IBGE. **PNAD**. 2011. Disponível em:
<http://www.ibge.gov.br/home/estatistica/pesquisas/pesquisaresultados.php?id_pesquisa=149>. Acesso em: 14 jun. 2021.
- IBGE. **População estimada**. 2022. Disponível em:
<<https://www.ibge.gov.br/apps/populacao/projecao/index.html>>. Acesso em: 21 jan. 2022.
- MORETTIN, P. A.; BUSSAB, W. O. **Estatística Básica**. 6. ed. São Paulo: Editora Saraiva, 2010.
- OLHAR DIGITAL. **Achatamento da curva**. 2020. Disponível em:
<<https://olhardigital.com.br/2020/03/25/coronavirus/como-funciona-a-matematica-do-achatamento-da-curva-do-coronavirus/>>. Acesso em: 01 mar. 2022.
- REIS, E. **Estatística Descritiva**. 7. ed. Lisboa: Sílabo, 2008.
- STEVENSON, W. J. **Estatística aplicada à administração**. São Paulo: Harper & Row do Brasil, 1981.

TOLEDO, G. L.; OVALLE, I. I. **Estatística Básica**. São Paulo: Atlas, 1981.

TRIOLA, M. F. **Introdução à estatística**: atualização da tecnologia. Rio de Janeiro: Ltc, 2014.

WALPOLE, R. E. *et al.* **Probabilidade e estatística para engenharia e ciências**. 8. ed. Tradução: Luciane F. Pauleti Vianna. São Paulo: Pearson Prentice Hall, 2009.

