

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Mestrado Profissional em Matemática em Rede Nacional

Marcus Vinicius Araujo

Gráficos Estatísticos: uma postura crítica

Juiz de Fora

2016

Marcus Vinicius Araujo

Gráficos Estatísticos: uma postura crítica

Dissertação apresentada ao Mestrado Profissional em Matemática em Rede Nacional da Universidade Federal de Juiz de Fora, na área de Ensino de Matemática, como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. Luís Fernando Crocco Afonso

Juiz de Fora

2016

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Araujo, Marcus Vinicius.
Gráficos Estatísticos : uma postura crítica / Marcus Vinicius Araujo.
– 2016.
71 f. : il.

Orientador: Prof. Dr. Luís Fernando Crocco Afonso
Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto
de Ciências Exatas. Mestrado Profissional em Matemática em Rede Nacional
, 2016.

1. Gráficos Estatísticos. 2. Estatística. I. Afonso, Luís Fernando Crocco,
orient. II. Título.

Marcus Vinicius Araujo

Gráficos Estatísticos: uma postura crítica

Dissertação apresentada ao Mestrado Profissional em Matemática em Rede Nacional da Universidade Federal de Juiz de Fora, na área de Ensino de Matemática, como requisito parcial para obtenção do título de Mestre em Matemática.

Aprovada em: 18 de agosto de 2016

BANCA EXAMINADORA

Prof. Dr. Luís Fernando Crocco Afonso - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Olímpio Hiroshi Miyagaki
Universidade Federal de Juiz de Fora

Prof. Dr. Francinildo Nobre Ferreira
Universidade Federal de São João Del-Rei

AGRADECIMENTOS

Agradeço...

A Deus, por sempre guiar os meus passos e me sustentar, permitindo assim que eu chegasse até aqui.

À minha mãe, Neide, pela vida, por se debruçar sobre a máquina de costura, sobre a máquina de escrever, sobre todo tipo de serviço digno, lutando com todas as forças para prover o sustento de nossa família. Guerreira e mãe, obrigado.

À minha irmã Rachel, minha fofa, sempre me amando e torcendo por mim.

Ao José Antônio, meu “paidraastro”, por ter feito da nossa a sua família.

Aos meus professores do mestrado, que com paciência fizeram com que eu crescesse na maior das aventuras, a do saber. Agradeço em especial ao meu orientador, Luis Fernando Crocco, um professor com o qual aprendi várias lições não só de matemática, mas também de humanidade.

Aos meus familiares, por todo o enorme carinho e amor incrível que me dedicaram durante toda uma vida, construindo assim, em parte, o que sou hoje. À Kátia, por ter acreditado em mim, me abraçado, amado e me apoiado em um processo longo de reconstrução.

Aos meus amigos, pessoas que fazem com que minha vida tenha muito mais cor e sabor, e que, sem alguns deles, talvez eu realmente nem estivesse mais por aqui.

Aos meus companheiros de caminhada no mestrado que, pacientemente, me ensinaram muitas vezes aquilo que para mim ainda era inalcançável. Provavelmente sem esse apoio eu não estaria escrevendo esse texto. Muito obrigado.

À CAPES, pelo incentivo financeiro que tornou possível este trabalho.

À minha esposa Aline, a quem eu amo. Companheira extraordinária que me apoiou e realizou essa caminhada ao meu lado; que chegou em minha vida e a fez florir, e que conduziu até a mim uma flor de valor incomensurável que está me ensinando um tipo de amor antes desconhecido: meu filho Rafael.

Agradeço ao Rafael por causar, uma nova forma de olhar, a de um pai, e assim, me fazer retornar ao passado e visitar o breve mas mágico espaço de tempo em que tive o meu pai comigo, e rever esses momentos através de outra perspectiva. Você foi maravilhoso, obrigado pai, onde quer que você esteja!

“Alguns usam a estatística como os bêbados usam postes: mais para apoio do que para
iluminação.” Andrew Lang

RESUMO

O presente trabalho tem por objetivo instigar nos leitores que tenham acesso a ele, um olhar mais crítico e questionador em relação aos gráficos estatísticos que surgem diariamente na mídia, e também servir como material de apoio em uma aula de Estatística, esclarecendo aos discentes a importância da compreensão de tal conteúdo para um pleno exercer da cidadania. Para atingir a tais fins, o mesmo foi dividido em duas partes. A primeira, um referencial teórico, traz uma revisão sobre os conceitos iniciais da Estatística, mais especificamente sobre as estruturas necessárias para compreensão e construção de gráficos. A segunda, um estudo de cinco casos (cinco gráficos estatísticos divulgados pela mídia) e as respectivas análises sobre cada um deles, pretende assim levar o leitor a uma postura mais amadurecida e cautelosa perante os mesmos.

Palavras-chave: Gráficos estatísticos. Estatística.

ABSTRACT

This study aims to entice in readers who have access to it, a look more critical and questioning in relation to the statistical graphs that appear daily in the media, and also serve as support material in a Statistics class, explaining to students the importance of such content understanding to a full exercise of citizenship. To achieve such purposes, it was divided into two parts. The first, a theoretical referential, provides a review of the initial concepts of Statistics, more specifically on the structures necessary for understanding and building graphics. The second one, a study of five cases (five statistical charts released by the media) and the analysis of each of them, plans to take the reader to become aware of such material to a more mature and cautious stance towards them.

Key-words: Statistical graphs. Statistic.

LISTA DE ILUSTRAÇÕES

Figura 1 – Histograma relativo à estatura dos alunos do segundo ano do Colégio Profmat em números absolutos.	27
Figura 2 – Histograma relativo à estatura dos alunos do segundo ano do Colégio Profmat em frequência relativa.	28
Figura 3 – Polígono de frequência relativo à estatura dos alunos do segundo ano do Colégio Profmat.	29
Figura 4 – Polígono de frequência acumulada relativo à estatura dos alunos do segundo ano do Colégio Profmat.	30
Figura 5 – Gráfico de barras relativo ao número de moradores na residência.	31
Figura 6 – Gráfico de barras relativo à disciplina de preferência.	32
Figura 7 – Gráfico de setores relativo à disciplina de preferência.	33
Figura 8 – Pesquisa referente aos motivos de dificuldades nas avaliações.	34
Figura 9 – Gráfico de Pareto relativo aos motivos de dificuldades nas avaliações.	35
Figura 10 – Exemplo de Pictograma.	36
Figura 11 – Participação das classes na massa de renda segundo a PNAD.	38
Figura 12 – Participação das classes no total de número de famílias segundo a PNAD.	39
Figura 13 – Comparação da participação das classes	41
Figura 14 – Gráfico de setores comparativo da renda per capita, por classes no Brasmil.	43
Figura 15 – Gráfico de barras comparativo das rendas per capita aproximadas do Brasmil, segundo a PNAD e do Brasmil, segundo a PNAD ajustada.	44
Figura 16 – Gráfico de barras Globo News: Taxa de Desemprego.	45
Figura 17 – Mensurando as barras do gráfico relativo à taxa de desemprego.	46
Figura 18 – Coluna ajustada.	47
Figura 19 – Possível gráfico sem os cortes.	49
Figura 20 – A riqueza do estado cresce e a renda das pessoas cai.	50
Figura 21 – A riqueza do estado cresce e a renda das pessoas cai mensurado no Geogebra.	51
Figura 22 – Os elefantes e os segmentos vertical e horizontal.	52
Figura 23 – R1 E R2.	53
Figura 24 – Mensurando as relações existentes entre os elefantes.	54
Figura 25 – Comparando os gastos do governo com suas respectivas representações gráficas.	56
Figura 26 – Recorte no rendimento médio dos trabalhadores.	57
Figura 27 – Os círculos e a elipse.	57
Figura 28 – rendimentos médios e rendimentos segundo as áreas.	59
Figura 29 – Gráfico de barras Globo News: Inflação do Brasil.	60
Figura 30 – Inflação do Brasil mensurada no GeoGebra	62

Figura 31 – Um possível gráfico da inflação do Brasil sem o corte	64
Figura 32 – Gráfico Globo News: Desemprego.	65
Figura 33 – O desemprego e a referência horizontal.	66
Figura 34 – As colunas mensuradas no GeoGebra.	67
Figura 35 – Desemprego: um possível gráfico ajustado.	68

LISTA DE TABELAS

Tabela 1 – Estatura de uma amostra de 50 alunos.	20
Tabela 2 – Estatura dos alunos do segundo ano do colégio Profmat I.	24
Tabela 3 – Estatura dos alunos do segundo ano do colégio Profmat II.	25
Tabela 4 – Números de moradores da residência.	31
Tabela 5 – Disciplina preferida.	31
Tabela 6 – Motivos de dificuldades nas avaliações	34
Tabela 7 – TABELA PNAD AJUSTADA “BRASMIL”	42
Tabela 8 – Razão de semelhança entre o segundo e o primeiro elefante.	54
Tabela 9 – Razão de semelhança entre o terceiro e o segundo elefante.	54
Tabela 10 – O aumento na área e o aumento correto.	55
Tabela 11 – As áreas dos círculos	57
Tabela 12 – A área da elipse.	58
Tabela 13 – As áreas e as porcentagens existentes entre c e d.	58
Tabela 14 – As áreas e as porcentagens existentes entre d e e.	58

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO INTRODUTÓRIO	14
2.1	CONCEITOS INICIAIS	14
2.1.1	Dados e variáveis	14
2.1.2	População	15
2.1.3	Amostra	15
2.1.4	Censo	15
2.1.5	Amostragem e Técnicas de Amostragem:	17
2.2	APRESENTAÇÃO DOS DADOS	20
2.2.1	Tabelas e Similares	20
2.2.1.1	Tabela:	20
2.2.1.2	Rol:	20
2.2.1.3	Distribuição de Frequência ou Tabela de frequência:	21
2.2.2	Gráficos	26
3	ESTUDOS DE CASO	37
3.1	GRÁFICO 1: CLASSES DE RENDIMENTO: A, B, C, D e E	37
3.2	GRÁFICO 2: EXPANDINDO O DESEMPREGO TRIMESTRAL	44
3.3	GRÁFICO 3: O GRÁFICO DA VEJA E O SITE	49
3.4	GRÁFICO 4: INFLAÇÃO BRASILEIRA DE 2013	60
3.5	GRÁFICO 5: DESEMPREGO EM VÁRIOS PAÍSES	64
4	CONSIDERAÇÕES FINAIS	69
	REFERÊNCIAS	70

1 INTRODUÇÃO

O homem, em sua caminhada neste planeta, teve que aprender a resolver os problemas e desafios que se interpunham entre ele e seus objetivos. O domínio da tecnologia do fogo, a criação da roda, os problemas de contagem para não perder animais nos campos, bem como o formato das velas nos navios para navegar contra o vento, são alguns exemplos de soluções criadas para problemas que limitavam ou muitas das vezes ameaçavam sua própria existência em algum momento histórico. Naturalmente, com o passar do tempo, seus problemas foram ficando cada vez mais complexos; concomitantemente a isso, as ferramentas de que dispunha para atacá-los foram ficando mais sofisticadas, o que de certa forma lhe garantia o continuar da sua existência.

Na sociedade moderna somos bombardeados por uma enorme quantidade de informação sobre os mais variados assuntos, tais como: tempo, investimentos, taxas de diversos tipos, comportamento de moedas, gastos pessoais e de instituições, crescimento ou decréscimo de lucros, endividamento, dentre outros. Com isso, se torna imperativo na tomada de decisões que o indivíduo participante dessa sociedade tenha a capacidade de gerenciar, absorver e interpretar a maior quantidade possível de dados e informações sobre os seus assuntos de interesse, correndo o risco de, se assim não o fizer, realizar escolhas prejudiciais a ele e àqueles que dependam de suas decisões.

Novamente o homem foi confrontado por um problema que se colocou entre ele e seus objetivos, e necessário se fez resolvê-lo, criar estratégias para sua solução. Ele, dessa vez, respondeu a tal problemática com uma ferramenta, um conjunto de conhecimentos matemáticos acumulados por ele em sua caminhada, a qual foi denominada de Estatística.

A Estatística pode ser vista como uma ciência de aprendizagem, um conjunto de métodos adequados à coleta, organização, resumo, descrição e apresentação de dados, auxiliando assim na compreensão e tomada de decisões nos mais variados problemas onde a aleatoriedade, a incerteza se apresentem. Para fins didáticos podemos dividi-la em três partes: a) estatística descritiva; b) probabilidade; c) estatística inferencial.

Nas páginas que se seguem focaremos nossa atenção na Estatística Descritiva, que, como o próprio nome já diz, tem como objetivo a descrição dos dados, buscando resumir-los e organizá-los de modo que as características de interesse existentes nos mesmos fiquem claras. Para tanto, lança mão de ferramentas como tabelas, gráficos, medidas de tendência central ou dispersão.

Temos que a informação é um direito de todos, sendo um fator indispensável ao exercício pleno da cidadania. Porém, muitas das vezes, por sua natureza e complexidade, não é facilmente compreendida, motivo pelo qual deve passar por tratamentos que venham permitir uma melhor apresentação e, conseqüentemente, melhor assimilação. A Estatística surge justamente para preencher essa lacuna.

Sendo assim, nossas estruturas educacionais devem ter como objetivo amparar o cidadão na descoberta e construção desse conjunto de conhecimentos, de forma que ele seja capaz de assimilar e compreender a informação estatística, assim como ser crítico em relação a ela e capaz de produzi-la.

Tal postura no Brasil é defendida pelos PCNs – Parâmetros Curriculares Nacionais, que, na edição referente ao ensino básico, dos terceiro e quartos ciclos do fundamental, aponta como objetivos do ensino fundamental (dentre outros) os destacados abaixo:

(...) saber utilizar diferentes fontes de informação e recursos tecnológicos para adquirir e construir conhecimentos; questionar a realidade formulando-se problemas e tratando de resolvê-los, utilizando para isso o pensamento lógico, a criatividade, a intuição, a capacidade de análise crítica, selecionando procedimentos e verificando sua adequação. [1, p.8]

A fim de atender a tal necessidade, foi criado um bloco com diretrizes nos PCNs denominado Tratamento da informação. “Os conteúdos matemáticos estabelecidos no bloco tratamento da Informação fornecem instrumentos necessários para obter e organizar as informações, interpretá-las, fazer cálculos e desse modo produzir argumentos para fundamentar conclusões sobre elas” [1, p.29]. Ainda sobre a criação desse bloco, ele esclarece 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 freqüentemente 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.[1, 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 freqüência relativa e se as escalas utilizadas são convenientes. Costuma ser freqüente 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. [1, p.136]

Tendo em vista esse contexto – e a fim de estabelecer um recorte no nosso objeto de estudo – o presente trabalho se concentrará na parte de gráficos, tão importante nos dias atuais de nossa sociedade de informação para o exercer de uma cidadania plena, justamente por amparar na compreensão de dados e fenômenos, e, conseqüentemente, na tomada de decisões.

Para tanto ele encontra-se estruturado em duas partes.

O segundo capítulo, de cunho mais conceitual, traz a teoria necessária à compreensão e construção de um sólido corpo de conhecimentos permitindo assim ao leitor a apropriação de saberes estatísticos.

O terceiro capítulo, por sua vez, busca apresentar de forma prática algumas situações problema a partir da análise de gráficos veiculados pela mídia em distintos meios de comunicação.

Esperamos que tal abordagem – a utilização de situações reais – chame a atenção do cidadão discente ou não para tais embustes presentes intencionalmente ou não nas representações gráficas estatísticas, contribuindo assim para o despertar no mesmo da chama da criticidade e o interesse por tal conhecimento.

2 REFERENCIAL TEÓRICO INTRODUTÓRIO

2.1 CONCEITOS INICIAIS

Iniciaremos nosso trabalho apresentando alguns conceitos preliminares acerca da Estatística.

Segundo Crespo em [5, p.13], “a Estatística é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões”.

Já Triola, em [13, p.4], define a estatística como “a ciência do planejamento de estudos e experimentos, da obtenção de dados e, em seguida, da organização, resumo, apresentação, análise, interpretação e elaboração de conclusões com base nos dados”.

Para entendermos melhor a Estatística podemos dividi-la em três partes, a saber: Estatística Descritiva, Probabilidade e Inferência Estatística, que estão apresentadas abaixo.

Estatística Descritiva:

É o começo de tudo, o momento inicial, em que vamos sintetizar uma série de valores (dados) que, em função da quantidade, têm a percepção de certas características de interesse dificultadas. Essa estatística, resume e descreve os dados, principalmente através de tabelas, gráficos e medidas de tendência central ou dispersão.

Probabilidade:

Em nosso mundo, muitos fenômenos são de natureza aleatória. Por exemplo, ao lançarmos um dado, de antemão, não sabemos qual número surgirá na face voltada para cima. Esse é o tipo de fenômeno chamado de aleatório, pois é inerente a ele a incerteza do resultado. A probabilidade é a área da matemática que estuda esse tipo de fenômeno.

Estatística Inferencial:

A Inferência Estatística, por vez, tem por objetivo vislumbrar respostas a perguntas feitas sobre um determinado conjunto a partir de amostras/subconjuntos retiradas(os) desse conjunto. No presente trabalho, o nosso foco se dará na primeira parte acima, ou seja, a Estatística Descritiva.

Começemos com as definições, visando uma breve revisão teórica.

2.1.1 Dados e variáveis

Dados são coleções de observações (por exemplo, medidas, gênero, respostas de pesquisas).[13, p.4]

Os dados, basicamente podem ser de dois tipos: quantitativos ou qualitativos.

Os dados quantitativos são expressos por números que representam medidas, mensurações. Já os qualitativos não são números, são atributos, qualidades, rótulos dos dados. Esse tipo de dado não pode ser ordenado, a não ser em casos especiais como, por exemplo, no caso dos ordinais de mensuração. Um exemplo desses são os conceitos aplicados em um curso onde o indivíduo pode receber uma nota A, B, C ou D, decrescentes nessa ordem. Mas nesse caso, cálculos como adição e subtração não são aplicáveis.

É comum vermos o termo variável em vez de dados, no estudo de um fenômeno. Crespo em [5, p.17], por exemplo, descreve as variáveis – possíveis resultados de um fenômeno – da forma que se segue:

a) qualitativa: quando seus valores são expressos por atributos: sexo (masculino – feminino), cor da pele (branca, preta, amarela, vermelha, parda) etc.;

b) quantitativa – quando seus valores são expressos em números (salários dos operários, idade dos alunos de uma escola etc.). Uma variável quantitativa que pode assumir, teoricamente, qualquer valor entre dois limites recebe o nome de variável contínua; uma variável que só pode assumir valores pertencentes a um conjunto enumerável recebe o nome de variável discreta. [5, p.17]

Já um exemplo de variável discreta é o número total de alunos em uma sala de aula. Um exemplo de variável contínua é a altura de um aluno, que tanto pode ser 180 cm quanto 180,4 cm.

2.1.2 População

Uma população é a coleção completa de todos indivíduos (escores, pessoas, medidas e outros) a serem estudados. A coleção é completa no sentido que inclui todos os sujeitos a serem estudados. [13, p.4]

2.1.3 Amostra

Uma amostra é uma subcoleção dos dados obtidos de todos os membros da população. [13, p.4]

2.1.4 Censo

Um censo é a coleção dos dados obtidos de todos os membros da população. [13, p.4]

Cabe aqui um comentário muito importante sobre a diferença entre parâmetro e estatística. Um parâmetro é um número que indica, descreve algo sobre toda uma

população, e estatística é um número que indica/descreve algo sobre uma amostra de uma população.

Já para entendermos melhor a diferença entre os conceitos de amostra e de censo, e a necessidade da existência deles, pensemos o seguinte:

Você é dono de uma grande empresa de comida congelada e quer lançar nacionalmente um novo produto. Então, tem basicamente três opções:

Primeira opção: Lança o produto, investindo em marketing, produção, transporte e outras condições, e aguarda para ver se haverá aceitação no mercado. Se sim, tudo deu certo, ok. Mas, se não der certo, você arcou com todos os custos, e, no final, perdeu dinheiro, o que não é nada desejável.

Segunda opção: Antes do lançamento do produto, encomenda uma pesquisa com toda a população, tendo como foco identificar o nível de aceitação que o produto terá, e, se percebida a existência de uma viabilidade econômica, lança o produto. Nesse caso, apesar de talvez ter percebido a viabilidade econômica do produto, pode esbarrar em um outro tipo de problema de viabilidade, que é a das realizações de tal pesquisa. Imagine o investimento necessário para tornar isso real em um país do tamanho do Brasil, que, de acordo com o IBGE – Instituto Brasileiro de Geografia e Estatística – no dia 1 de julho de 2013, possuía um total de 201.032.714 habitantes.

Para se ter ideia coloquemos abaixo os dados do último censo, que é uma pesquisa com toda a população. Sua realização se deu no ano de 2010.

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

Universo recenseado: todo o Território Nacional

Número de municípios: 5.565 municípios

Número de domicílios: 67.569.688 de domicílios

Número de setores censitários: 314.018 setores censitários

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

Orçamento: aproximadamente 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.

[9]

Os números gigantescos acima indicam uma possível inviabilidade econômica de tal pesquisa, tendo em vista o lançamento do produto. Sendo assim, isso nos leva à terceira possibilidade.

Terceira opção: Antes de lançar o produto, faz uma pesquisa, utilizando-se de uma amostra da população. Esse procedimento pode levar à conclusão tão desejada sobre a viabilidade do produto em questão, sendo necessário para tanto um investimento financeiro aceitável. Assim sendo, a terceira opção – uma pesquisa estatística – parece ser a escolha mais acertada.

Mas perceba que, para que aconteça tal pesquisa, é necessária a obtenção de uma amostra da população, ou seja, necessitamos de uma subcoleção de dados que são obtidos a partir da população. Esses dados são tão importantes que Triola em [13, p.20] sobre eles chega a afirmar que:

Se os dados amostrais não forem coletados de maneira apropriada, eles podem ser de tal modo inúteis que nenhuma manipulação Estatística poderá salvá-los. [13, p.20]

O que nos leva, naturalmente, ao nosso próximo item.

2.1.5 Amostragem e Técnicas de Amostragem:

Amostragem:

A amostragem, segundo Crespo em [5, p.20], é “uma técnica especial – amostragem - para recolher amostras, que garante, tanto quanto possível, o acaso na escolha”.

A escolha da amostra pode ser feita de várias formas tendo em vista o objetivo da pesquisa. O mais importante é que a amostra seja representativa, adequada; isto é, útil para as análises estatísticas que serão feitas a partir dela. Para que uma amostra seja representativa, vários fatores devem ser observados, como por exemplo: não ser viesada; ter um tamanho adequado; seguir em sua construção princípios de aleatoriedade; e outros.

Para fins elucidativos, podemos definir uma amostra enviesada como sendo aquela que possui uma tendência sistemática de selecionar um certo grupo em uma população.

Abaixo descreveremos algumas técnicas de amostragem.

Amostragem probabilística:

Na amostragem probabilística, a probabilidade de cada elemento ou membro de uma população ser incluído na amostra deve ser conhecida e diferente de zero.

Amostragem aleatória:

Neste tipo de amostragem, se temos uma população de tamanho n , cada único elemento dela deve ter a mesma chance ou probabilidade de ser escolhido para compor a amostra.

Amostragem aleatória simples ou casual:

Na amostragem aleatória simples utilizamos o acaso ao selecionarmos os dados, isto é, os membros ou elementos da população que estarão na amostra. Tem como característica o fato de que uma vez selecionada uma amostra de k elementos, qualquer outra amostra de k elementos tem a mesma probabilidade de ser selecionada.

Amostragem sistemática:

Neste tipo de amostragem estabelece-se um ponto ou um elemento como sendo o inicial, e escolhe-se um intervalo que pode ser, por exemplo, de tamanho 20. Com esse intervalo escolhido, procede-se da seguinte forma: a cada sequência de 20 elementos da população, inclui-se o último elemento desse intervalo na amostra.

Segue abaixo um esquema.

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, **20**, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, **40**, 41, 42, 43, 44, 45, 46,...

Assim, seriam incluídos, o 20º, o 40º, o 60º, e assim por diante

Amostragem por conveniência:

É um método não probabilístico, ou seja, os indivíduos participantes da pesquisa são escolhidos, selecionados em função de serem de fácil acesso, e não através de um critério estatístico. Essa conveniência redundava em uma facilidade operacional e um baixo investimento financeiro, mas tem como consequência a incapacidade de se fazer afirmações gerais com rigor estatístico sobre a população. Um ponto forte dessa técnica de amostragem é a sua utilização em um momento exploratório da pesquisa, auxiliando assim na construção de hipóteses e obtenção de *insights*. Por exemplo, suponhamos que você quer saber a opinião dos estudantes do ensino médio sobre a realização das Olimpíadas no Brasil. Para se fazer uma amostragem probabilística, seria necessário possuir um acesso à população dos estudantes brasileiros do Ensino médio para que pudessemos selecionar uma amostra aleatoriamente, e então realizar a pesquisa. Já na amostragem por conveniência, poderíamos escolher algumas escolas próximas e realizar a pesquisa com os estudantes dessas escolas.

Amostragem Estratificada:

É comum dividirmos uma população em subpopulações, subconjuntos, subgrupos, isto é, em estratos (ex.: homens e mulheres; estado civil: solteiro ou casado; e outros, podendo vir a ser designados mais de dois estratos) buscando viabilizar uma melhor observação de uma determinada variável de interesse. É comumente utilizada quando

são percebidas diferenças relevantes entre os subgrupos populacionais que objetivamos estudar. É feita de forma a garantir que os estratos ou subgrupos estejam representados proporcionalmente ao seu peso na população de estudo.

Por exemplo, se for percebida entre a população urbana e não urbana uma considerável diferença em relação à vivência religiosa, e se essa questão em nossa pesquisa for relevante, então, é importante que a nossa amostra inclua um número de homens e de mulheres residentes em áreas urbanas e não urbanas que seja proporcionalmente igual aos existentes na população em estudo.

Amostragem por conglomerado:

A amostra por conglomerados baseia-se na existência de clusters, grupos em uma população. Tais grupos devem ser representativos em relação à população total, ou seja, úteis nas análises estatísticas que pretendemos. Atendida a essa exigência, são selecionados um ou mais desses conglomerados para a realização do estudo e todos os elementos desses conglomerados são utilizados.

Um exemplo comum dessa técnica é a escolha dos conglomerados de forma geográfica. Pode-se, por exemplo, de forma aleatória, serem escolhidas algumas regiões. Após isso, dentre essas regiões, escolher-se alguma(s) subregião(ões) e realizar-se a pesquisa com todos os elementos pertencentes a essa(s) subregião(ões) escolhida(o)s.

Amostragem de resposta voluntária ou autoseleção:

É um tipo de amostragem onde o indivíduo decide se participa ou não da pesquisa, ou seja, ele mesmo se seleciona. Um bom exemplo disso são as pesquisas de sites na internet, onde são feitas uma ou mais perguntas e o internauta decide ou não se responde. Apesar de parecer algo bem prático, essa amostra não é representativa. Isso se deve ao fato de que pesquisas desse tipo atraem pessoas especificamente interessadas pelo assunto em questão, e com opinião forte o bastante para tomar a decisão de dispor de seu tempo para fornecer as repostas da pesquisa. Dessa forma, podemos dizer que a amostra assim obtida é enviesada, e como tal, não é útil para as análises estatísticas sobre a mesma. Interessante também observar que o próprio tipo de público que acessa certos sites também é bem específico, o que por si só já leva ao enviesamento.

Até agora, fizemos uma breve apresentação sobre os conceitos iniciais da Estatística e algumas de suas partes; discursamos sobre dados e como coletá-los. Uma vez vencida essa etapa, necessário se faz que esses sejam resumidos e organizados. O que nos leva à próxima parte de nosso trabalho.

2.2 APRESENTAÇÃO DOS DADOS

Após a obtenção dos dados, é importante dispô-los, organizá-los de forma que seja facilitada a observação e compreensão desses. Visando atingir a esse objetivo, dispomos, em estatística, de um conjunto de ferramentas. Algumas delas serão apresentadas a seguir.

2.2.1 Tabelas e Similares

2.2.1.1 Tabela:

Podemos definir uma tabela como uma forma de se arranjar sistematicamente dados numéricos em linhas e colunas, objetivando sua comparação.

Tal estruturação – tabela – facilita a exposição e leitura dos dados. Segue abaixo um exemplo, que denominaremos de Tabela 1.

Estatura de uma amostra de 50 alunos dos segundos anos do ensino médio da Rede de Ensino PROFMAT, em centímetros.

Tabela 1 – Estatura de uma amostra de 50 alunos.

171	162	168	168	166	168	163	182	166	171
176	174	180	161	178	165	163	165	170	164
173	162	172	178	174	165	170	182	166	175
168	170	173	178	171	165	170	174	177	175
178	177	170	173	180	172	161	171	170	164

2.2.1.2 Rol:

É uma forma relativamente simples de se organizar variáveis quantitativas, como por exemplo: altura; renda mensal; quantidade de telefonemas de uma determinada população ou amostra; e outros. Criar um Rol consiste em simplesmente dispormos os dados em uma sequência crescente ou decrescente. Abaixo, faremos um rol na sua forma crescente, relativo aos dados da Tabela 1.

161, 161, 162, 162, 163, 163, 164, 164, 165, 165, 165, 165, 166, 166, 166, 168, 168, 168, 168, 170, 170, 170, 170, 170, 170, 171, 171, 171, 171, 172, 172, 173, 173, 173, 174, 174, 174, 175, 175, 176, 177, 177, 178, 178, 178, 179, 180, 180, 182, 182.

Perceba que, ao criarmos o Rol, algumas características dos dados já ficam mais claras, como por exemplo: a menor estatura; a maior estatura; a amplitude dos dados ($182 - 161 = 21$ cm); o valor que mais surge (o 170 cm, com 6 ocorrências); o fato de que a frequência de alunos com 180 cm ou mais é bem pequena; entre outros.

2.2.1.3 Distribuição de Frequência ou Tabela de frequência:

Ao agruparmos dados na forma de uma tabela, contendo classes, e contendo a frequência dos dados nas respectivas classes, criamos uma distribuição de frequência ou tabela de frequência. Abaixo construiremos uma distribuição de frequências a partir da tabela acima. Para isso, antes, definiremos algumas de suas partes:

Classe de frequência:

“Classes de frequência ou, simplesmente, classes, são intervalos de variação da variável.” [5, p.57]

O número de classes varia em função de fatores que veremos abaixo no texto.

Número de classes:

Para a determinação do número de classes, o mais importante é o bom senso, que deve ser norteado pelos objetivos que desejamos alcançar. Triola em [13, p.40], por exemplo, 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”. Já Crespo em [5, p.61], afirma que podemos utilizar uma regra denominada de “Regra de Sturges”, onde o número de classes é dado pela fórmula:

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

Mas comenta também que essa fórmula não é decisiva, participando do processo de determinação do número de classes outros fatores, como a natureza dos dados, objetivos e a unidade utilizada para expressar os dados; ou seja, o bom senso mais uma vez é requisitado.

Na fórmula acima temos que o número de classes é representado por i , e o número total de dados por n . Utilizando Sturges teremos:

$$\begin{aligned} i &\cong 1 + 3,3 \cdot \log 50 \\ i &\cong 6,6 \end{aligned}$$

Apesar de Sturges nos dizer 6,6 e, ao arredondarmos termos 7 classes, decidimos permanecer com 6 classes. Com essa escolha, perdemos um pouco em detalhes, mas temos um ganho em simplicidade.

Limite inferior da classe:

É o menor número/valor pertencente a uma classe. Pode ser designado por l_i .

Limite superior da classe:

É o maior número/valor pertencente a uma classe. Pode ser designado por L_i .

Nota: a letra i designa qual a classe que está sendo considerada. Por exemplo: o limite inferior da classe 3 é designado por l_3 , assim como $i = 3$ designa a terceira classe.

Amplitude total da distribuição e Amplitude amostral:

A amplitude total da distribuição é obtida pela diferença entre o limite superior da última classe e o limite inferior da primeira classe, enquanto que a amplitude amostral é obtida através da diferença entre o valor máximo e o mínimo da amostra.

A amplitude total pode ser designada por AT :

$$AT = L(\text{máx.}) - l(\text{mín.})$$

Já a amplitude amostral pode ser designada como AA :

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

No Rol apresentado anteriormente, temos:

$$\begin{aligned} x(\text{máx.}) &= 182 \\ x(\text{mín.}) &= 161 \\ AA &= 182 - 161 = 21 \end{aligned}$$

Amplitude de classe:

Uma vez superada a decisão relativa ao número de classes, torna-se fácil o cálculo da amplitude de classe, que pode ser feito pela fórmula:

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

$$h \cong \frac{x(\text{máx.}) - x(\text{mín.})}{i} \cong \frac{182 - 161}{6} = 3,5$$

Onde h = amplitude de classe; e i = número de classes. Se por ventura obtivermos um h não inteiro como por exemplo $h = 3,5$, devemos arredondar esse valor para o número natural imediatamente acima. No nosso caso, utilizaríamos $h = 4$.

Mas, se acharmos melhor, podemos com tranqüilidade alterar o número de classes para obter uma amplitude de classe mais conveniente.

Assim, como temos o número de classes (que são 6), e, como já obtivemos 4 como amplitude de classe, estamos aptos a construir a tabela de freqüência.

Na primeira classe colocaremos $l_1 = 160$, e, somando a 160 a amplitude da classe ($160 + 4$) obtemos $L_1 = 164$. Assim, na primeira classe, teremos o intervalo $160 \vdash 164$, o que significa que serão inseridos todos os valores de 160 inclusive até 164 exclusive; isto é, intervalo fechado em 160 e aberto em 164. Veja que apesar do valor 164 ser o limite

superior da primeira classe, ele não é contabilizado nessa classe, e sim na segunda. Assim sendo, contabilizaremos para essa classe todos os valores maiores ou iguais a 160 cm e menores que 164 cm. Ocorre o mesmo nos limites superiores e inferiores de todas as classes.

Na segunda classe, por vez, colocaremos $l_2 = 164$, e, somando a 164 a amplitude da classe ($164 + 4$), obtemos $L_2 = 168$. Assim, teremos como intervalo de classe $164 \vdash 168$.

Repetindo esse processo nas demais classes, chegaremos finalmente à sexta e última classe, que apresentará um intervalo de $l_6 = 180$ e $L_6 = 184$.

Frequência simples ou absoluta:

Segundo Crespo em [5, p.59], “frequência simples ou frequência absoluta ou, simplesmente, frequência de uma classe ou de um valor individual é o número de observações correspondentes a essa classe ou a esse valor.”

Usaremos como símbolo para frequência simples f_i , onde, como anteriormente, o índice i indica a classe. Por exemplo:

f_2 indica a frequência da segunda classe, que vai de 164 inclusive a 168 exclusive, e apresenta os seguintes valores: 164; 164; 165; 165; 165; 165; 166; 166; 166. Logo, $f_2 = 9$.

f_6 indica a frequência da sexta classe; assim, temos $f_6 = 4$.

O símbolo de somatório:

$$\sum_{i=1}^k f_i,$$

Indica o somatório da frequência da primeira classe (pois $i = 1$) até a classe K . Se essa for a última classe, esse somatório será igual à frequência total da tabela ou distribuição de frequência.

E ainda, se não houver a possibilidade do equívoco, para o somatório de todas as frequências podemos usar simplesmente:

$$\sum f_i$$

Na nossa tabela teremos como somatório de todas as seis frequências:

$$\sum_{i=1}^6 f_i = 50 \text{ ou simplesmente } \sum f_i = 50$$

E, finalmente, chegamos ao objetivo.

Tabela 2 – Estatura dos alunos do segundo ano do colégio Profmat I.

Estatura (cm)	Frequência
160 † 164	6
164 † 168	9
168 † 172	14
172 † 176	10
176 † 180	7
180 † 184	4
	$\sum f_i = 50$

Ponto médio de uma classe:

O ponto médio de uma classe divide o intervalo da classe em duas partes iguais. Pode ser calculado através da média aritmética entre o limite inferior e o limite superior da classe. Podemos designá-lo por x_i .

$$x_i = \frac{L_i + l_i}{2}$$

Exemplo:

O ponto médio da terceira classe:

$$x_3 = \frac{172+168}{2} = \frac{340}{2} \Rightarrow x_3 = 170$$

Tipos de Frequências:

As frequências das classes ou a frequência total podem ser mostradas de várias formas, que, em essência, vão nos ajudar a perceber características dos dados como centralidade, variação, valores discrepantes (*outliers*), distribuição e outros. Descreveremos abaixo alguns tipos.

Frequência simples de uma classe:

Como já visto acima, são os valores absolutos que representam a quantidade de dados em uma classe.

Na terceira classe temos:

$$f_3 = 14$$

Frequência relativa de uma classe:

É a razão/quociente entre a frequência simples da classe e a frequência total. Utilizaremos o símbolo f_{r_i} .

$$f_{r_i} = \frac{f_i}{\sum f_i}$$

Na terceira classe temos:

$$f_3 = \frac{14}{50} = 0,28 = 28\%$$

Frequência acumulada:

É o somatório de todas as frequências até uma considerada classe, inclusive a frequência dessa classe. Utilizaremos o símbolo F_i . Assim sendo, abaixo temos a frequência acumulada até a classe k .

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

Sendo assim, a frequência acumulada na terceira classe será:

$$F_3 = f_1 + f_2 + f_3$$

$$F_3 = 6 + 9 + 14$$

$$F_3 = 29$$

Frequência relativa acumulada:

É a razão/quociente entre a frequência acumulada de uma classe e a frequência total da distribuição ou conjunto de dados. Utilizaremos como símbolo F_{r_i} .

$$F_{r_i} = \frac{F_i}{\sum f_i}$$

A frequência acumulada relativa na terceira classe:

$$F_{r_3} = \frac{F_3}{\sum f_3} = \frac{29}{50}$$

$$F_{r_3} = 0,58 = 58\%$$

Agora, construiremos a tabela com esses dados adicionais.

Tabela 3 – Estatura dos alunos dos segundo ano do colégio Profmat II.

Classe	Estaturas(cm)	f_i	F_i	f_{r_i}	F_{r_i}
1	160 – 164	6	6	0,12	0,12
2	164 – 168	9	15	0,18	0,30
3	168 – 172	14	29	0,28	0,58
4	172 – 176	10	39	0,20	0,78
5	176 – 180	7	46	0,14	0,92
6	180 – 184	4	50	0,08	1,00
		$\sum f_i = 50$		$\sum f_{r_i} = 1$	

Vejamos que, com o auxílio dessa tabela, podemos perceber características dos dados que antes não ficavam evidenciadas, como por exemplo, a estimativa de que 78% dos alunos têm menos que 1,78 m.

Para corroborar com essa tarefa, a de evidenciar características dos dados, podemos lançar mão de um recurso muito poderoso: os gráficos. Faremos uma singela apresentação desses nas linhas que se seguem.

2.2.2 Gráficos

A partir de agora, falaremos sobre este outro recurso de resumo e apresentação dos dados: os gráficos.

Histograma:

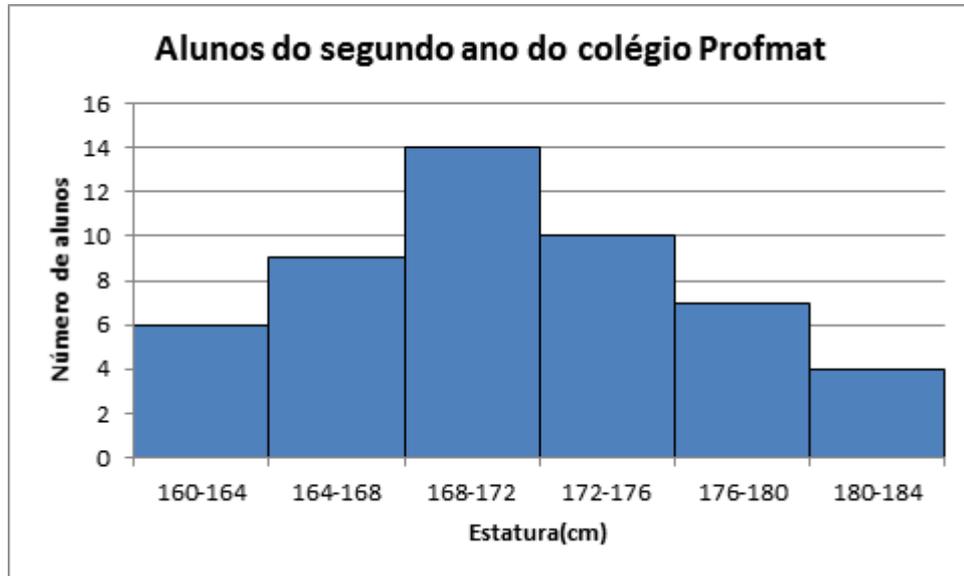
Segundo Iezzi [11, p. 250]:

O histograma é um gráfico formado por retângulos contíguos, isto é, que estão em contato entre si (os retângulos se “encostam”). A base de cada retângulo corresponde a um segmento cuja extremidade são os limites de cada classe de intervalo, e a altura de cada retângulo é proporcional à frequência (absoluta, relativa ou em porcentagem) da classe correspondente. [11, p. 250]

Como dito acima, os histogramas possuem uma característica muito interessante, que é o fato de que a área de um retângulo ou coluna, em relação à área total do histograma (a área ocupada pelos retângulos) guarda a mesma proporção que a da frequência dessa coluna em relação à frequência total da distribuição de frequência referente ao gráfico. Por exemplo: no gráfico abaixo, temos que $f_3 = 14$ e $\sum f_i = 50$. Em termos de porcentagem $\frac{14}{50} = 28\%$; logo, a área da coluna que representa a terceira classe é 28% da área total do gráfico. Assim, no caso de serem usadas as frequências relativas no eixo vertical, a área total encontrada no histograma será unitária. Se tivermos como objetivo comparar duas ou mais distribuições de frequência que possuem quantidade distinta de elementos, isso pode ser facilmente feito utilizando-se de histogramas de frequências relativas, pois esse faz com que os gráficos de duas amostras ou populações de número distinto de elementos possuam mesma área total unitária.

O Histograma da estatura dos alunos dos segundos anos do colégio PROFMAT, em centímetros, será feito na Figura 1.

Figura 1 – Histograma relativo à estatura dos alunos do segundo ano do Colégio Profmat em números absolutos.

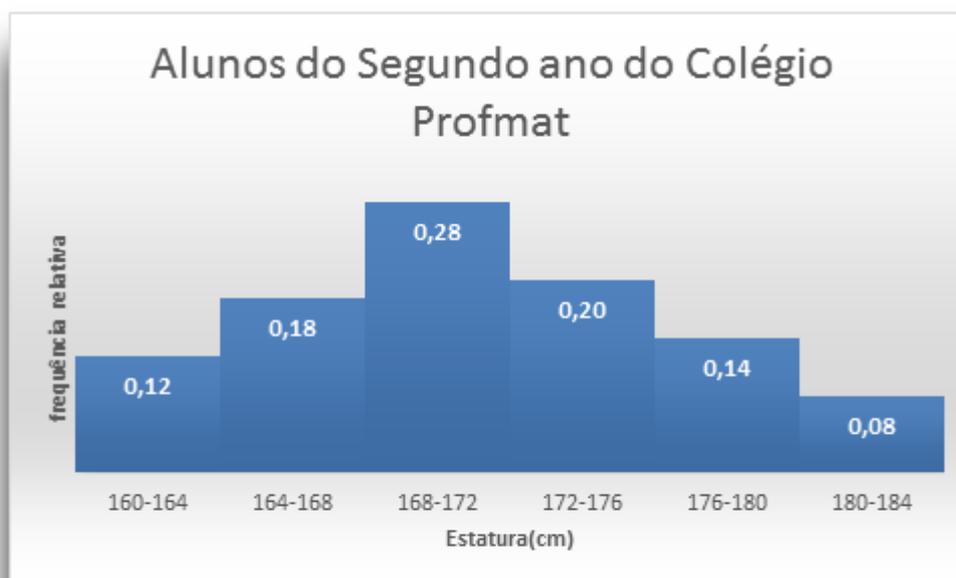


Fonte: o autor

Histograma de frequência relativa:

O Histograma de frequência relativa é análogo ao histograma acima, diferenciando-se dele por trazer em sua escala vertical as frequências relativas em vez das frequências absolutas.

Figura 2 – Histograma relativo à estatura dos alunos do segundo ano do Colégio Profmat em frequência relativa.



Fonte: o autor

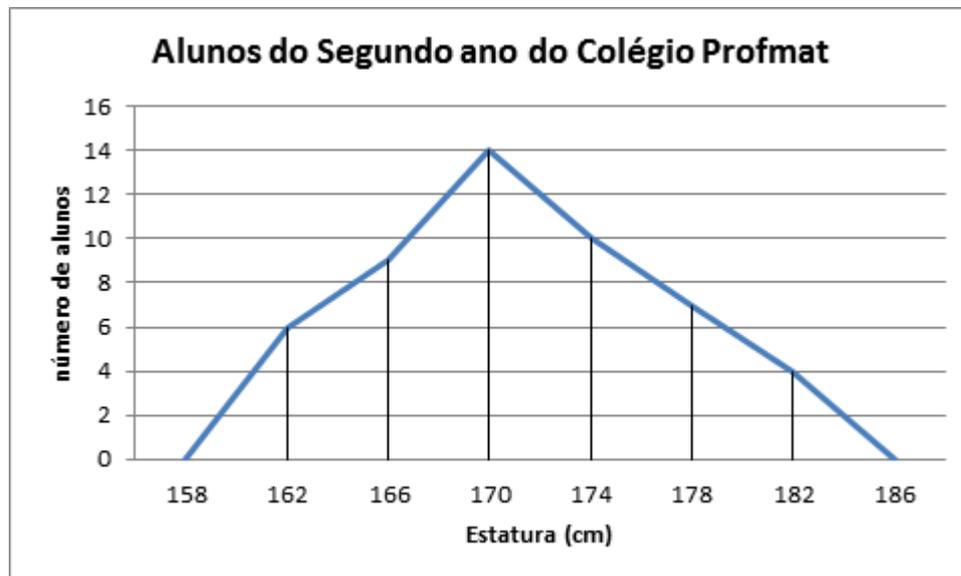
Assim como no Histograma, todos os gráficos a seguir podem ser feitos utilizando-se a frequência relativa. Não o faremos a título de tornar mais breve o presente texto.

Polígono de Frequência:

Segundo Crespo [5, p.70], “o polígono de frequência é um gráfico em linha, sendo as frequências marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classes”.

Mas, para “fecharmos” o polígono, devemos ligar os extremos da linha obtida (como descrito acima) 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.

Figura 3 – Polígono de frequência relativo à estatura dos alunos do segundo ano do Colégio Profmat.



Fonte: o autor

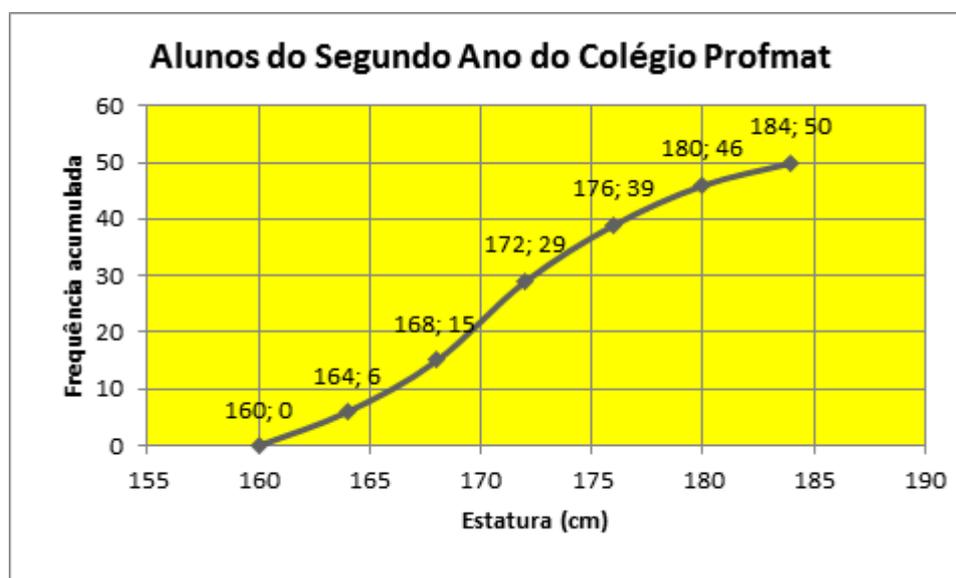
Polígono de frequência acumulada:

Este gráfico se inicia no limite inferior da primeira classe de frequência e termina no limite superior da última classe. Muito útil para responder rapidamente a perguntas como: Quantos indivíduos têm sua estatura abaixo de 1,72 metros ou 172 cm?

Resposta: 29 indivíduos.

No modelo abaixo, cada ponto destacado apresenta não apenas os limites das classes como traz também a frequência acumulada até ele (exclusive).

Figura 4 – Polígono de frequência acumulada relativo à estatura dos alunos do segundo ano do Colégio Profmat.



Fonte: o autor

Gráfico de Barras:

É um gráfico utilizado em variáveis qualitativas ou discretas.

Pensando em plano cartesiano, as variáveis (numéricas ou textos) ficarão no “eixo das abscissas”, e as frequências ficarão no “eixo das ordenadas”. Para cada variável, construímos uma barra de altura ou comprimento proporcional a sua frequência.

Exemplo:

Foram feitas, dentre outras, duas perguntas para a amostra dos alunos do segundo ano do Colégio Profmat que estão transcritas abaixo:

- Qual o número de moradores da sua residência?
- Dentre as disciplinas: Matemática, Português, Física, Química ou Biologia, qual a sua preferida? (Nota: Somente uma pode ser escolhida).

Organizadas as respostas, foram obtidas as Tabelas 4 e 5:

Tabela 4 – Números de moradores da residência.

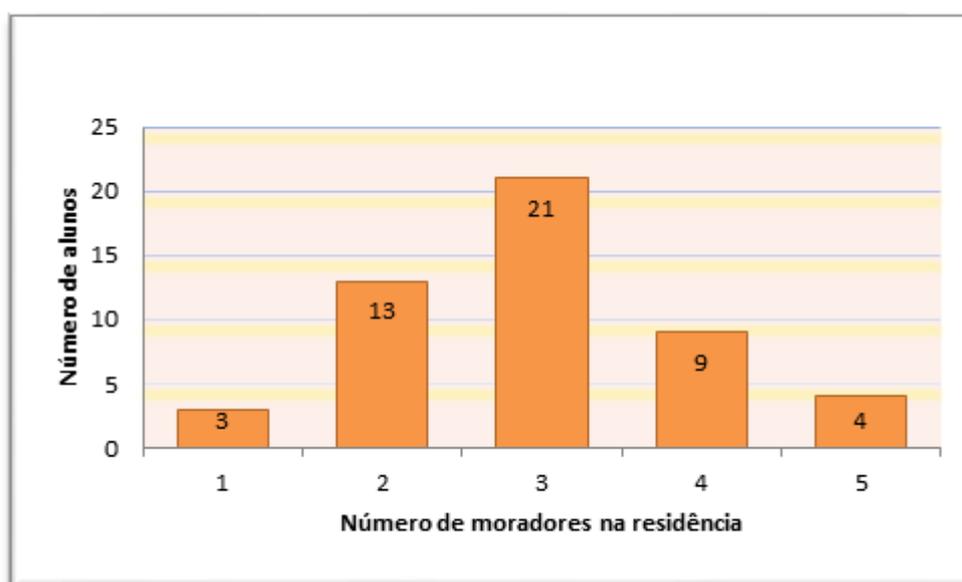
<i>Nº de moradores</i>	<i>Nº de alunos</i>
1	3
2	13
3	21
4	9
5	4
Total	50

Tabela 5 – Disciplina preferida.

Disciplina Preferida	<i>Nº de alunos</i>
Matemática	16
Português	7
Física	9
Química	12
Biologia	6
Total	50

E obtidos os gráficos 5 e 6:

Figura 5 – Gráfico de barras relativo ao número de moradores na residência.

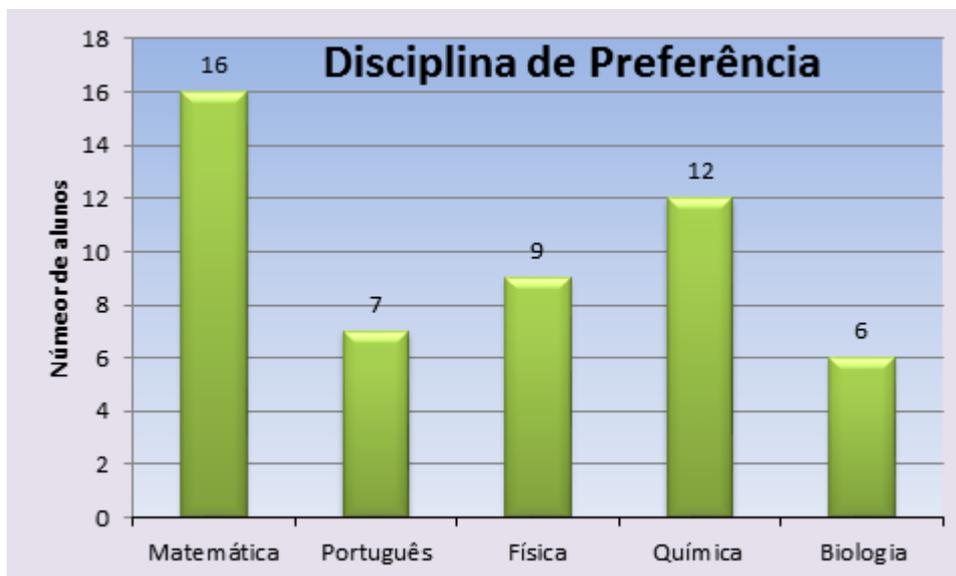


Fonte: o autor

No gráfico 5, temos uma variável quantitativa discreta, e, através dele, rapidamente podemos observar as respostas dos alunos.

Abaixo apresentamos um gráfico de barras para variáveis qualitativas.

Figura 6 – Gráfico de barras relativo à disciplina de preferência.



Fonte: o autor

Onde podemos ver de forma clara a preferência pela disciplina matemática, pelo menos nesse gráfico.

Gráfico de Setores ou de Pizza:

Este tipo de gráfico se associa muito bem a variáveis qualitativas. Segundo Bussab e Morettin em [4, p.15] o gráfico de setores “consiste num círculo unitário de raio arbitrário, representando o todo, dividido em setores, que correspondem às partes de maneira proporcional”.

A partir da tabela acima, de disciplina preferida, podemos fazer os cálculos tanto da porcentagem ocupada por um setor, assim como de seu ângulo central.

Cálculo da porcentagem para a variável qualitativa matemática:

$$100 \cdot f_{r_1} = 100 \cdot \frac{16}{50} = 100 \cdot 0,32 = 32\%$$

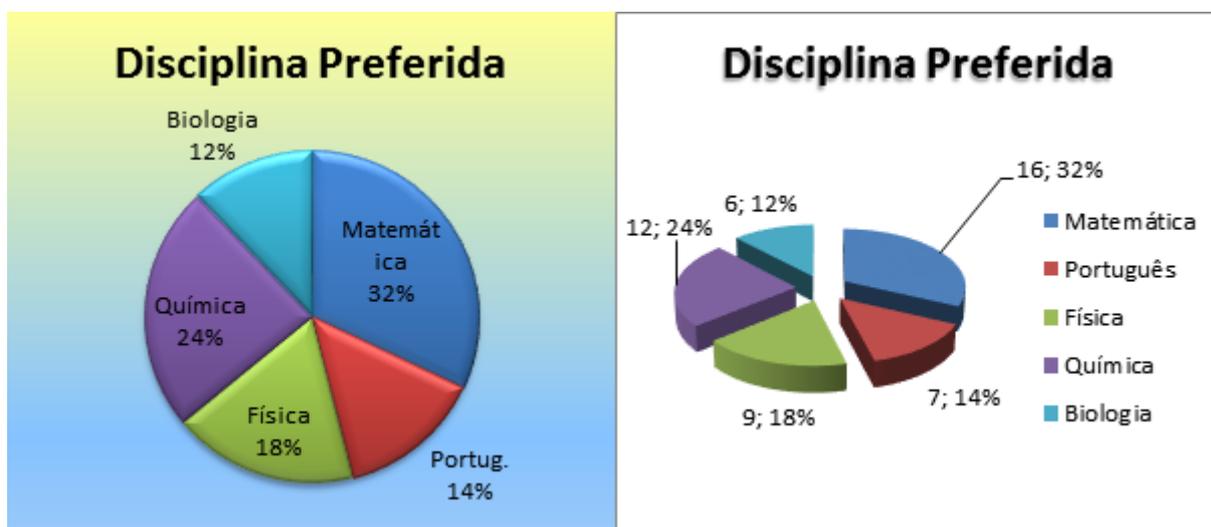
Cálculo do ângulo central para a variável matemática abaixo denominada de x_1 :

$$\alpha_1 = f_{r_i} \cdot 360^\circ = 0,32 \cdot 360^\circ = 115,2^\circ$$

De maneira análoga construiríamos estas duas componentes do gráfico para as demais variáveis.

A seguir, faremos dois gráficos de setores, ou pizza. O primeiro será um gráfico plano, unidimensional. O segundo será em 3D, ou seja, tridimensional.

Figura 7 – Gráfico de setores relativo à disciplina de preferência.



Fonte: o autor

Os gráficos de setores são bem confiáveis por trabalhar basicamente com proporcionalidade, o que, para efeito de comparação, é muito bom. Muitos problemas em gráficos surgem quando esse quesito é desrespeitado. Além disso, nele, ao sairmos de uma dimensão, como o fizemos acima, a fidelidade do recurso gráfico à proporcionalidade nos dados permanece. Abordaremos mais ostensivamente essa questão de fidelidade no capítulo 3.

Gráfico ou Diagrama de Pareto:

É um gráfico de barras muito útil por possibilitar uma fácil visualização e classificação das classes ou variáveis de maior frequência, pois estas são ordenadas de forma decrescente em relação à frequência. O Diagrama de Pareto é uma técnica estatística que ampara na tomada de decisão, pois auxilia a identificação de prioridades. O Princípio do Diagrama de Pareto diz que aproximadamente 80% das perdas são causados por 20% dos defeitos ou comportamentos (nesse caso variáveis). Interpretando de maneira mais flexível o princípio, o que ele nos diz é que a maior parte dos problemas vem de um pequeno número de causas, e, se pudermos identificar essas causas, poderemos com um esforço relativamente pequeno sanar grande parte, na verdade, a maior parte dos problemas. Logo, parece ser muito útil identificar esses “20%” para que sejam feitos procedimentos visando a erradicação ou ajustes nesses, e, assim, seja alcançada uma otimização no todo.

Um exemplo: Em feita uma pesquisa com as 25 turmas que totalizam 1215 alunos

do ensino médio da rede de ensino Profmat, e umas das perguntas está abaixo:

Figura 8 – Pesquisa referente aos motivos de dificuldades nas avaliações.

Na sua opinião, qual o principal motivo de suas dificuldades nas avaliações? Marque apenas uma opção. No caso de escolher “outros”, por favor liste no espaço adequado abaixo o motivo.

() Pouco tempo de estudo.

() Prova de curta duração.

() Grande dificuldade na disciplina.

() Problemas familiares.

() Didática do professor.

() Uso excessivo de redes sociais.

() Grau de dificuldade da prova alto.

() Outros.

Outros: _____

Fonte: o autor

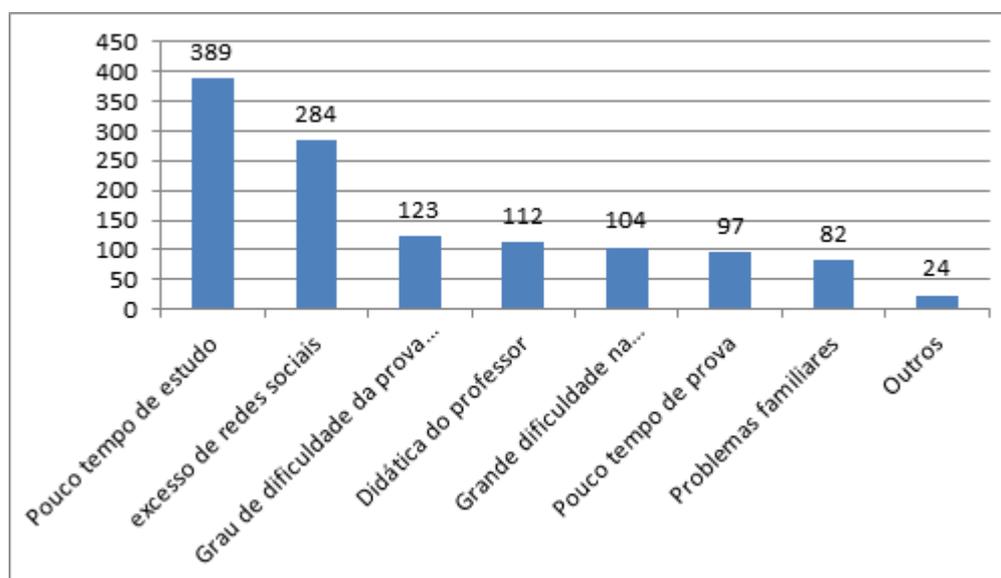
Após tabelados os dados, tem-se os seguintes resultados:

Tabela 6 – Motivos de dificuldades nas avaliações

Motivos de dificuldades nas avaliações	Frequência
Pouco tempo de estudo	389
Grande dificuldade na disciplina	104
Problemas familiares	82
Uso excessivo de redes sociais	284
Pouco tempo de prova	97
Didática do professor	112
Grau de dificuldade da prova alto	123
Outros	24
	$\sum f_i = 1215$

E, finalmente, temos o seguinte gráfico de Pareto:

Figura 9 – Gráfico de Pareto relativo aos motivos de dificuldades nas avaliações.



Fonte: o autor

Que facilmente nos mostra que mais da metade dos problemas são causados por duas categorias: “pouco de tempo de estudo” e “excesso de redes sociais”, que juntas dão mais de 50% das respostas. Perceba que de certa forma, o excesso de uso de redes sociais leva ao pouco tempo de estudo. Intencionalmente, o pesquisador criou a categoria “excesso de redes sociais” no intuito de chamar a atenção dos pesquisados para esse comportamento, e o quão prejudicial ele pode ser. Também é bem plausível que boa parte da categoria: “grande dificuldade na disciplina” seja em função do pouco tempo de estudo. Assim, se a direção geral buscar atacar primeiramente a questão do tempo de estudo, possivelmente estará cuidando de mais de 60% dos problemas relacionados às dificuldades nas avaliações. Posteriormente, poderá fazer intervenções nos demais aspectos.

Pictogramas:

Segundo Crespo em [5, p.48], “o pictograma constitui um dos processos gráficos que melhor fala ao público, pela sua forma ao mesmo tempo atraente e sugestiva. A representação gráfica consta de figuras”.

Já Triola [13, p.55], ele cita alguns princípios importantes que segundo ele são sugeridos por Tufte ¹ em seu livro; “The Visual Display of Quantitative Information”, segunda edição.

Transcreveremos abaixo:

¹ Tufte, Edward R., and P. R. Graves-Morris. The visual display of quantitative information. Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

- Um gráfico de dados deve fazer o observador concentrar-se na verdadeira natureza dos dados, e não em outros elementos, tais como características de desenho que chamam a atenção, mas distraem.
- Não distorça os dados; construa um gráfico para revelar a verdadeira natureza dos dados.
- Quase toda tinta em um gráfico deve ser usada para os dados, não para outros elementos do desenho.
- Não use área ou volumes para dados que são, na verdade unidimensionais por natureza. (Por exemplo, não use desenhos de nota de dólares para representar quantidades do orçamento para diferentes anos. [13, p.55-56]

Bom, independentemente das opiniões, realmente deve-se tomar um certo cuidado na utilização dos pictogramas, uma vez que eles expressam ideias e essas podem ser influenciadas pelas imagens, cores, intensidade das cores utilizadas e outros. Ainda há a necessidade da proporcionalidade entre os dados ser respeitada nas figuras, o que nem sempre acontece².

Abaixo segue um exemplo de pictograma:

Figura 10 – Exemplo de Pictograma.



Fonte: Crespo em [5, p.50]

Terminamos aqui a parte da revisão teórica. Passaremos, no próximo capítulo, para os Estudos de Caso.

² Como veremos nos exemplos do capítulo 3

3 ESTUDOS DE CASO

Nas páginas anteriores discorreremos sobre os conceitos iniciais da Estatística e sobre a teoria referente à apresentação dos dados na forma tabelas e similares, e, após essa caminhada preparatória, alcançamos o assunto central de nosso trabalho: gráficos.

No presente capítulo, apresentaremos um material que tange mais especificamente a esses e que acreditamos ser capaz de enriquecer e muito uma aula de estatística, assim como alertar a qualquer pessoa que dele se intere da necessidade desse saber. Serão feitas as análises de cinco gráficos procurando verificar vários aspectos nos mesmos. A matemática utilizada nas análises seguintes será a mais simples possível. Lançaremos mão de recursos como regras de três, taxas de porcentagens e réguas. Em substituição a essa última, nos utilizaremos do programa GeoGebra – um software de matemática dinâmica que apresenta tanto recursos de geometria quanto recursos de álgebra e cálculo – para realizar medições. Sua utilização se justifica pela necessidade de uma maior exatidão uma vez que o presente trabalho consiste em um escrito acadêmico.

Escolhemos esse caminho – o da simplicidade – para deixar claro que qualquer indivíduo com algum conhecimento em Estatística e com recursos simples de cálculo e mensuração é plenamente capaz de, com um pouco de criatividade e curiosidade, formular hipóteses sobre os gráficos que estão presentes no seu dia-a-dia e verificá-las, assumindo assim uma postura dinâmica e crítica perante os mesmos. Passemos agora, então, para os gráficos analisados.

3.1 GRÁFICO 1: CLASSES DE RENDIMENTO: A, B, C, D e E

No Brasil, o IBGE – Instituto Brasileiro de Geografia e Estatística – é o órgão responsável por realizar a PNAD – Pesquisa Nacional por amostra de Domicílios – 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:

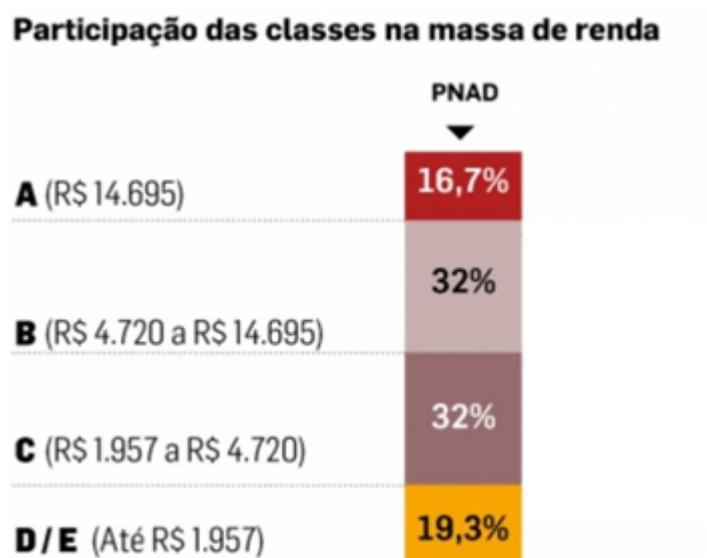
- * População residente segundo o sexo e os grupos de idade;
- * Taxa de desocupação;
- * Taxa de atividade;

- * 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. [10]

Após esse breve esclarecimento sobre a PNAD, podemos dar desenvolvimento ao assunto.

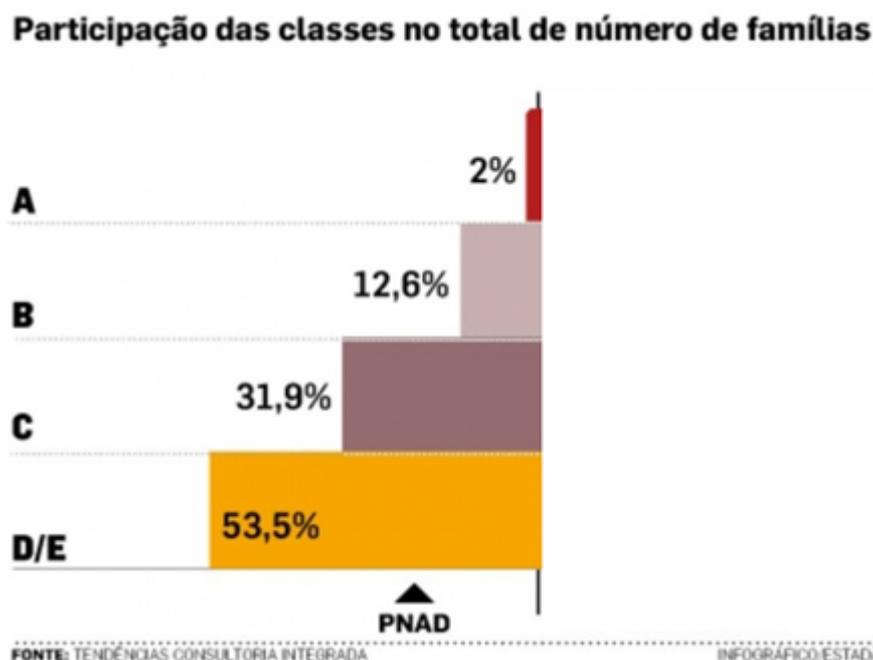
No início de 2016, o jornal o Estadão publicou uma reportagem denominada: “Classe A tem maior fatia da renda no País” , em que se utilizava a PNAD para fazer, dentre outros, comentários sobre as classes sociais no Brasil e divisões de renda, assim como base para a construção de alguns gráficos sobre esses assuntos. Partes de dois dos gráficos estão abaixo. Neles, são mostradas a participação das classes na massa de renda brasileira e a participação das classes no total de número de famílias.

Figura 11 – Participação das classes na massa de renda segundo a PNAD.



Fonte: <http://economia.estadao.com.br/noticias/geral,classe-a-tem-maior-fatia-da-renda-do-pais,10000007285> Acessado em: 09/04/16

Figura 12 – Participação das classes no total de número de famílias segundo a PNAD.



Fonte: <http://economia.estadao.com.br/noticias/geral,classe-a-tem-maior-fatia-da-renda-do-pais,10000007285> Acessado em: 09/04/16

Esses gráficos produzidos a partir dos dados coletados pelo IBGE, através da PNAD, demonstram uma profunda desigualdade na divisão de renda no Brasil, onde:

- A classe A, mais alta, que possui rendimentos maiores ou iguais a R\$ 14695,00, e que representa 2% das famílias brasileiras, fica com 16,7% da massa de renda.
- A classe B, que representa 12,6% das famílias brasileiras, fica com 32% da massa de renda.
- A classe C, que representa 31,9% das famílias, fica também com 32% da massa de renda.
- E as classes D e E, que representam 53,5%, ou seja, mais da metade da população brasileira, dividem 19,3%; ou seja, menos de 1/5 da massa de renda.

Como podemos perceber, uma situação profunda de desigualdade é descrita, onde mais da metade da população fica com 19,3% da massa de renda, e apenas 2% da população ficam com quase 17% da mesma massa. Mas será que esses números são confiáveis? Lembremo-nos que essa é uma pesquisa feita pelo IBGE, um órgão brasileiro muito confiável. Sendo assim, será que essa desigualdade pode ser ainda maior do que a demonstrada pela PNAD? E se sim, qual seria um possível motivo para esse fenômeno?

Em [13, p.16] Triola reserva uma parte para tecer comentários sobre alguns escolhos

do processo estatístico. Dentre eles, encontra-se um que ele denomina como “resultado informado”, que parece trazer um pouco de luz para os questionamentos acima:

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. [13, p.16]

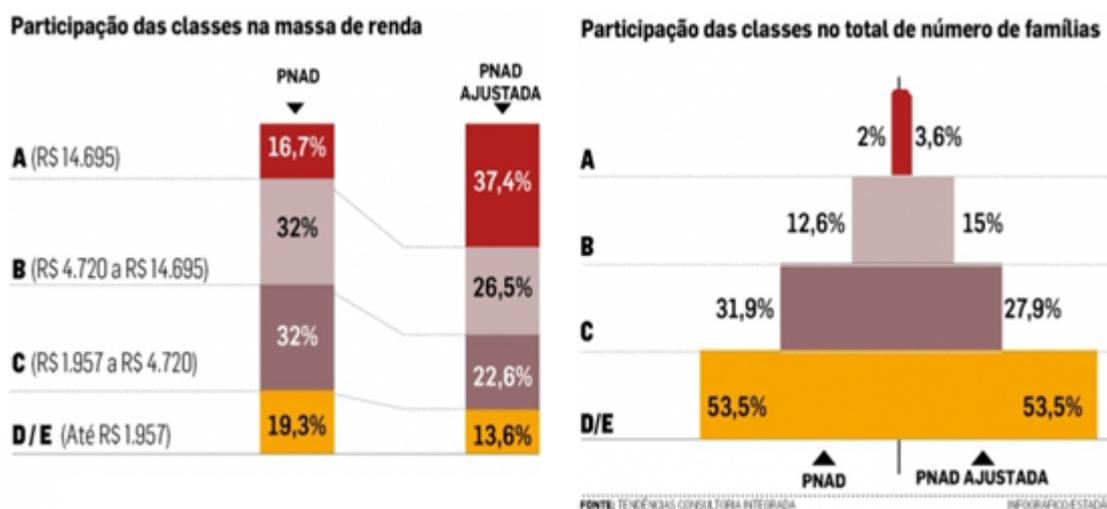
E esse é justamente um ponto crucial sobre o quesito renda na PNAD. Como a comprovação da renda que é declarada pelo indivíduo entrevistado mediante pergunta não é exigida, muitas das vezes ela não é a real.

Em entrevista para o jornal “Estadão”, o economista Adriano Pitoli, integrante do quadro da Tendências Consultoria Integrada, afirma que “as pesquisas declaratórias (como a PNAD) são ineficientes para capturar a renda de aplicações financeiras, aluguéis e ganhos de capital. Na verdade, ninguém tem esses números de cabeça.”[6]

Temos, então, possivelmente um viés. Como dito na parte teórica do presente trabalho, um viés no momento da coleta de dados pode comprometer toda uma pesquisa, pois todas as estatísticas serão construídas sobre eles – os dados –, e podem ser comprometidas. Surge então um objetivo: buscar ferramentas para perceber se o problema ocorreu, e, em caso positivo, buscar estratégias para driblar esse problema e chegar a valores mais fidedignos à realidade, libertos – o máximo possível – desse provável viés.

Percebendo o problema, os economistas Ernesto Guedes, Adriano Pitoli e Camila Saito, da “Tendências Consultoria Integrada”, que doravante no texto será tratada como “Tendências”, tiveram uma ideia interessante que foi publicada na reportagem do Estadão mencionada no início desse texto. Eles se utilizaram dos dados da receita federal para sua análise sobre a distribuição de renda, buscando contornar, pelo menos em parte, esta situação indesejável que surge às vezes ao se utilizar resultados informados. Assim, obtiveram, segundo eles, novos números sobre a distribuição de renda, tendo por base o que os mesmos denominaram de PNAD ajustada. Nos cálculos, foram utilizados os dados da PNAD para as famílias com renda de até R\$ 4.400,00 (ou seja, cinco salários mínimos de R\$880,00); já para aquelas cuja renda superou esse valor, foram utilizados os dados das declarações do Imposto de Renda. Através dessa estrutura eles chegaram aos resultados que começaremos a apresentar agora.

Figura 13 – Comparação da participação das classes na massa de renda segundo a PNAD e segundo a PNAD ajustada e comparação da participação das classes no total de número de famílias segundo a PNAD e segundo a PNAD ajustada.



Fonte: <http://economia.estadao.com.br/noticias/geral,classe-a-tem-maior-fatia-da-renda-do-pais,10000007285> Acessado em: 09/04/16

Nos gráficos acima, que foram retirados da reportagem do Estadão, são comparados os dados da PNAD com os resultados da PNAD ajustada feita pela “Tendências” a partir do imposto de renda.

Segundo eles, a classe A, que representa 3,6% da população, fica com 37,4% da massa de renda brasileira; A classe B, que representa 15% da população, fica com 26,5% da massa de renda; a classe C, que representa 27,9% da população, fica com 22,6%; e, finalmente, a classe D/E, que representa 53,5% da população brasileira, fica com 13,6%, que é menos de 1/7 da massa de renda; ou seja, uma vez dividida a massa de renda em sete partes, pegue uma quantidade menor que apenas uma dessas sete partes e a distribua para mais da metade das famílias brasileiras. Situação ainda pior que a mencionada anteriormente, que citava menos de 1/5 da massa de renda para o mesmo grupo.

Apesar de serem preocupantes esses novos números, Fernando Brito (jornalista e fundador do Blog Tijolaço) comenta em um artigo do Blog denominado “Tijolaço” que essa correção feita pela “Tendências” pode ser ainda incompleta. Segue texto abaixo:

Correção, talvez, ainda incompleta, porque a sonegação é maior entre os mais ricos, seu padrão de deduções é igualmente maior e a agregação por núcleo familiar esconde o fato de que as famílias mais pobres em geral têm mais integrantes que as famílias mais ricas. [3]

Para clarearmos um pouco mais as porcentagens encontradas na PNAD ajustada,

vamos agora usar uma ideia muito interessante, que o escritor do Blog acima usou na matéria que escreveu sobre a mesma, que é a de criar um local fictício e aplicar nele as porcentagens das classes. Não nos utilizaremos dos mesmos valores utilizados na matéria, posto que ele se utiliza de aproximações nas quais não desejamos nos aventurar, e, tampouco também nos utilizaremos das mesmas denominações e estrutura. Somente aproveitaremos a ideia.

Primeiramente vamos criar o “Brasmil”, um Brasil imaginário com 1000 moradores e com massa de renda mensal de um mil reais. Assim, dividindo percentualmente os moradores e a renda de “Brasmil” nas classes A, B, C e D/E, teremos a tabela abaixo:

Tabela 7 – TABELA PNAD AJUSTADA “BRASMIL”

Classe	Número de pessoas	Massa de renda (R\$)*	Renda per capita aproximada(R\$)
A	36	374	10,40
B	150	265	1,77
C	279	226	0,81
E	535	136	0,25

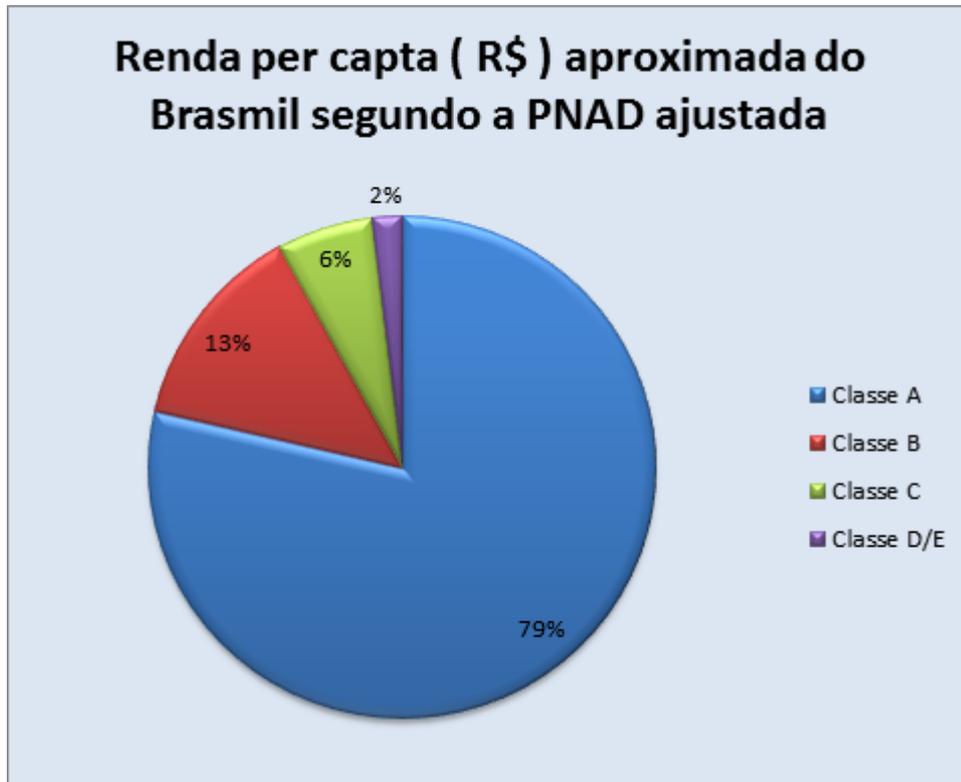
* o somatório da renda tem como resultado R\$1001 devido a aproximações de cálculo

Para fins de visualização, construiremos um gráfico de setores e um gráfico de colunas:

1) O gráfico de setores:

Na PNAD ajustada, se somarmos os ganhos de uma pessoa de cada uma das classes teremos um total de R\$ 13,23. Agora, se compararmos qual a porcentagem que um indivíduo de cada classe recebe desses R\$ 13,23, teremos o gráfico abaixo.

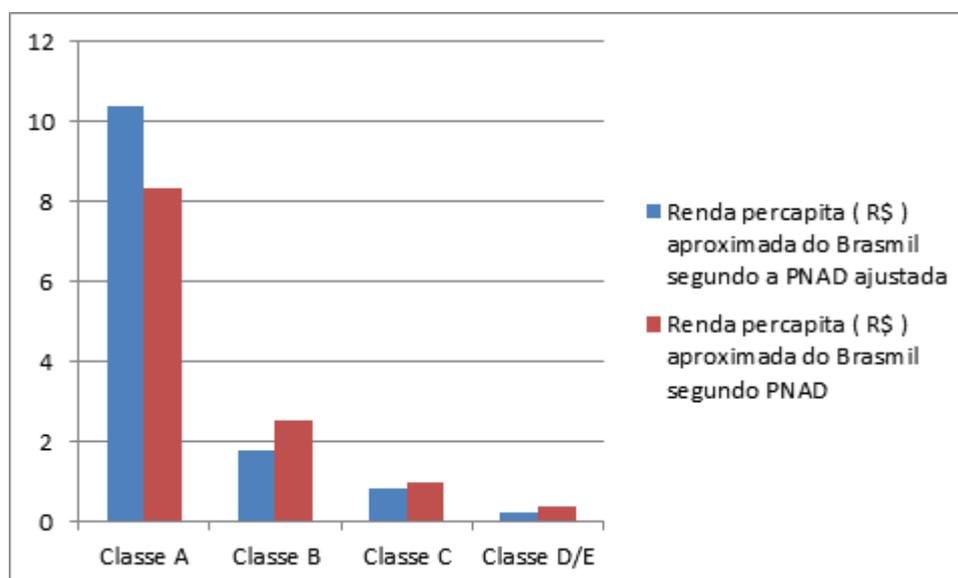
Figura 14 – Gráfico de setores comparativo da renda per capita, por classes no Brasmil.



Que já é por si só, bastante elucidativo.

2) O gráfico de colunas:

Figura 15 – Gráfico de barras comparativo das rendas per capita aproximadas do Brasmil, segundo a PNAD e do Brasmil, segundo a PNAD ajustada.



Fonte: o autor

Comparando as rendas per capita da PNAD e da PNAD ajustada do nosso “Brasmil”, no gráfico acima, fica muito perceptível o abismo que separa as classes sociais na PNAD – em vermelho – e que esse abismo é ainda maior na PNAD ajustada, em azul. Fica ainda claro, ao compararmos as nossas duas PNADs, que todas as classes sociais perdem significativo poder aquisitivo na PNAD ajustada, exceto a classe A, que nesse quesito tem um aumento considerável. Enfim, essas análises sugerem algo muito preocupante, que é o fato de que, possivelmente, o abismo existente entre a classe A brasileira e as demais classes, que já era percebido como algo de grande magnitude, pode ser ainda muito maior do que se imaginava.

3.2 GRÁFICO 2: EXPANDINDO O DESEMPREGO TRIMESTRAL

No site “G1” da Emissora Globo, está alocada uma vídeo reportagem que começa com o seguinte texto dito pela apresentadora, em tom lúgubre e com ênfases em alguns pontos:

Incerteza política; disparada da inflação; falta de confiança na economia. Essa é uma combinação de fatores que tende a agravar o lado mais perverso da crise, que é o desemprego, em um ano quase três milhões de brasileiros perderam o lugar no mercado de trabalho. A taxa de desemprego atingiu em janeiro 9,5%, esse é o maior nível desde o início da série histórica desde que começou a ser medido em 2012.[8]

No decorrer da reportagem (que no total dura 2 minutos e 49 segundos), surge de repente, aos 56 segundos de duração, o seguinte o gráfico de barras, a fim de, supostamente, ilustrar e auxiliar o telespectador na compreensão da situação do desemprego.

Figura 16 – Gráfico de barras Globo News: Taxa de Desemprego.



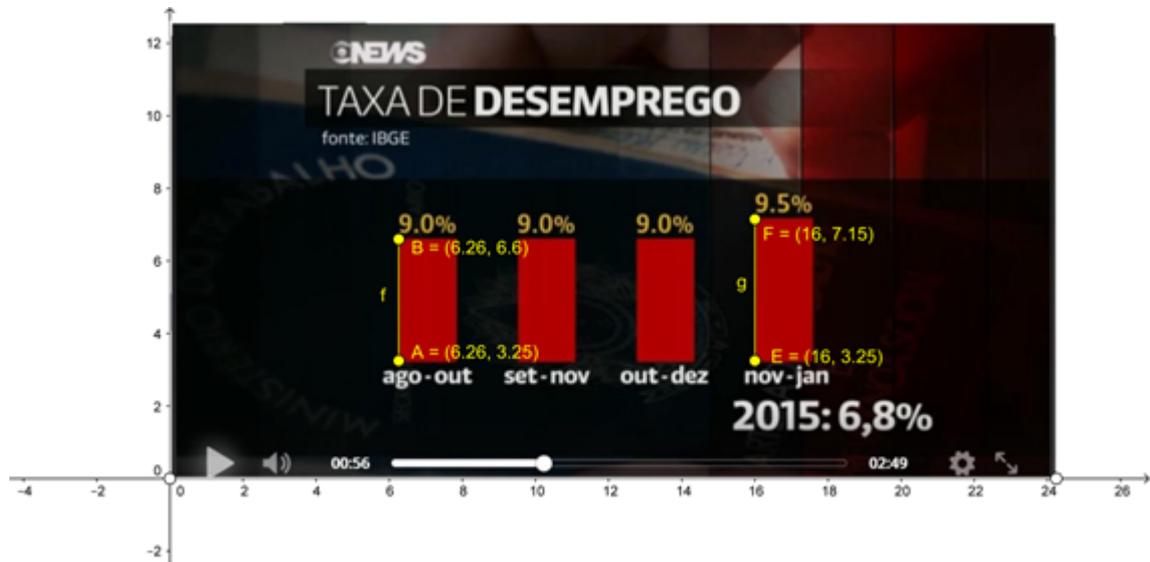
Fonte: <http://g1.globo.com/globo-news/jornal-das-dez/videos/v/taxa-de-desemprego-atinge-95-em-janeiro/4909364/> Acessado em:04/04/16

Vejam os bem:

O gráfico traz abaixo e à direita, e sem maiores explicações, a taxa de 6,8% associada a uma referência relativa ao ano de 2015. Esse valor diz respeito à taxa de desemprego anual medida pelo IBGE nas cidades do Rio de Janeiro, São Paulo, Belo Horizonte, Recife, Salvador e Porto Alegre; ou seja, nas maiores regiões metropolitanas do país através da PME (Pesquisa Mensal de Emprego realizada pelo IBGE), e parece um pouco descontextualizado se levarmos em conta o assunto tratado nas colunas, que são as taxas trimestrais medidas pela PNAD.

Como não há explicações sobre o valor de 6,8%, pode ser feita naturalmente pelo telespectador uma comparação entre esse dado (6,8%) e os dados das colunas, o que, por sua vez, traz uma impressão errônea de extremo aumento. Algo descabido em função da natureza totalmente distinta dos dados. Ainda, segundo o gráfico, a taxa de desemprego vinha se mantendo estável nas três medições anteriores e se altera no período que vai de novembro de 2015 a janeiro de 2016, momento no qual ela dá um salto de 0,5%. Aqui, utilizamo-nos da palavra “salto” pois nos parece ser o que aconteceu com o gráfico. Faremos todavia, a seguir, algumas medições para verificarmos se as alturas das colunas do gráfico refletem proporcionalmente o aumento de 0,5% no desemprego, ou se houve realmente um “salto”, uma distorção na representação gráfica dessa variável.

Figura 17 – Mensurando as barras do gráfico relativo à taxa de desemprego.



Fonte: o autor

Para tanto, inserimos a imagem no programa GeoGebra e fizemos as mensurações, chegando aos seguintes resultados:

Na coluna ago-out, que possui a mesma altura das três primeiras colunas, criamos e medimos o segmento \overline{AB} .

$$med(\overline{AB}) = 6,6 - 3,25 = 3,35$$

Na coluna nov-jan criamos e medimos o segmento \overline{EF} .

$$med(\overline{EF}) = 7,15 - 3,25 = 3,9$$

Agora, ao dividirmos 9% em 18 partes, temos que cada uma dessas partes em porcentagem representará 0,5%. Logo, concluímos que o acréscimo de 0,5% ocorrido no trimestre nov-jan representa um acréscimo de $1/18$ sobre o total percentual do trimestre ago-out, e então, proporcionalmente, a medida do segmento \overline{EF} deve ser $1/18$ maior que a do segmento \overline{AB} . Em comprimento, esse aumento deve ser de:

$$\frac{1}{18} \cdot 3,35 \approx 0,185$$

Agora, o acréscimo ocorrido fica determinado pela diferença entre os comprimentos das colunas referidas acima, que foi de:

$$med(\overline{EF}) - med(\overline{AB}) = 3,9 - 3,35 = 0,55$$

Um valor que vai bastante além de 0,185. Descreveremos abaixo, de forma relativa através de um quociente (uma porcentagem), a dimensão da distorção identificada no gráfico:

AO : aumento ocorrido

AC: aumento correto

$$\frac{AO}{AC} = \frac{0,55}{0,185} = 2,97 = \frac{297}{100} = 297\%$$

Bom, e o que significa esse número? Significa que o aumento que foi feito no gráfico é de aproximadamente o triplo do que deveria ser. Em termos percentuais, o pedaço de gráfico acrescentado é quase 300% do que deveria ser.

Assim, temos então, uma grave inverdade, que leva as pessoas que olham o gráfico despreziosamente, colocando a questão numérica em segundo plano, a considerar que o desemprego foi muito maior do que foi realmente.

Para melhor ilustrar a situação, construímos o gráfico abaixo com uma nova coluna à extrema direita, que apresenta a sua altura proporcional aos dados. Essa deveria ter sido a dimensão da coluna relativa ao trimestre nov-jan apresentada à sociedade.

Figura 18 – Coluna ajustada.



Fonte: o autor

A título de conclusão, a coluna que representa a taxa de 9,5% de desemprego, na verdade, seria, se utilizada de forma correta para uma representação gráfica, quase perfeita para uma taxa de desemprego 10,5%.

Outra forma de pensarmos sobre esse gráfico é assumirmos o valor encontrado como a diferença entre \overline{EF} e \overline{AB} , que é de 0,55, como a real representação de 0,5% no

gráfico. Assim, para acharmos a altura das colunas de 9% e 9,5%, podemos lançar mão das seguintes regras de três:

Altura da coluna de 9%:

0,5% está para 0,55, assim como 9% está para z , onde z é a altura da coluna de 9%. Fazendo os cálculos chegamos ao valor de $z = 9,9$, que é bem distante das 3,35 unidades de comprimento que a referida coluna apresenta.

Já, para a altura da coluna de 9,5%:

0,5% está para 0,55, assim como 9,5% está para t , onde t é a altura da coluna de 9,5%. Fazendo os cálculos chegamos ao valor de $t = 10,45$, que também é bem distante das 3,9 unidades de comprimento que tal coluna apresenta como sua altura na figura original.

Sob esse olhar, pode-se dizer que o gráfico foi cortado, isto é, o eixo vertical foi iniciado fora do zero. Mas, fica então a pergunta: em que altura ele foi iniciado?

Para essa resposta recorreremos mais uma vez à regra de três. Usando como parâmetro a coluna de 9%, temos que a diferença entre 9,9 e 3,35 nos dá em unidades de comprimento o valor 6,65, que é justamente a altura que falta a essa coluna. Esse valor indica o ponto de corte ou início do gráfico. Mas, como as alturas estão em porcentagem, faremos o seguinte cálculo:

0,55 está para 0,5%, assim como 6,65 está para c , onde c é a altura do corte em porcentagem; assim, encontramos para c o valor aproximado de 6,05%. Logo, considerando (como já mencionado) que o valor de 0,55 representa o valor de 0,5%, temos em nossa reportagem um gráfico que tem como início de seu eixo vertical a porcentagem 6,05%.

E, finalmente, assim seria o nosso gráfico antes do corte:

Figura 19 – Possível gráfico sem os cortes.



Fonte: o autor

Bem menos impactante para o telespectador do que o exibido originalmente.

3.3 GRÁFICO 3: O GRÁFICO DA VEJA E O SITE

Ao efetuarmos uma busca no Google com as seguintes palavras chave: prof; andrios; Bemfica e estatística, uma referência que surgirá é o endereço abaixo: <<http://professorandrios.blogspot.com.br/2011/08/representacao-grafica-de-dados.html>>

Ao entrar neste endereço, que tem como assunto a representação gráfica de dados estatísticos, em certo momento se deparará com o que se segue:

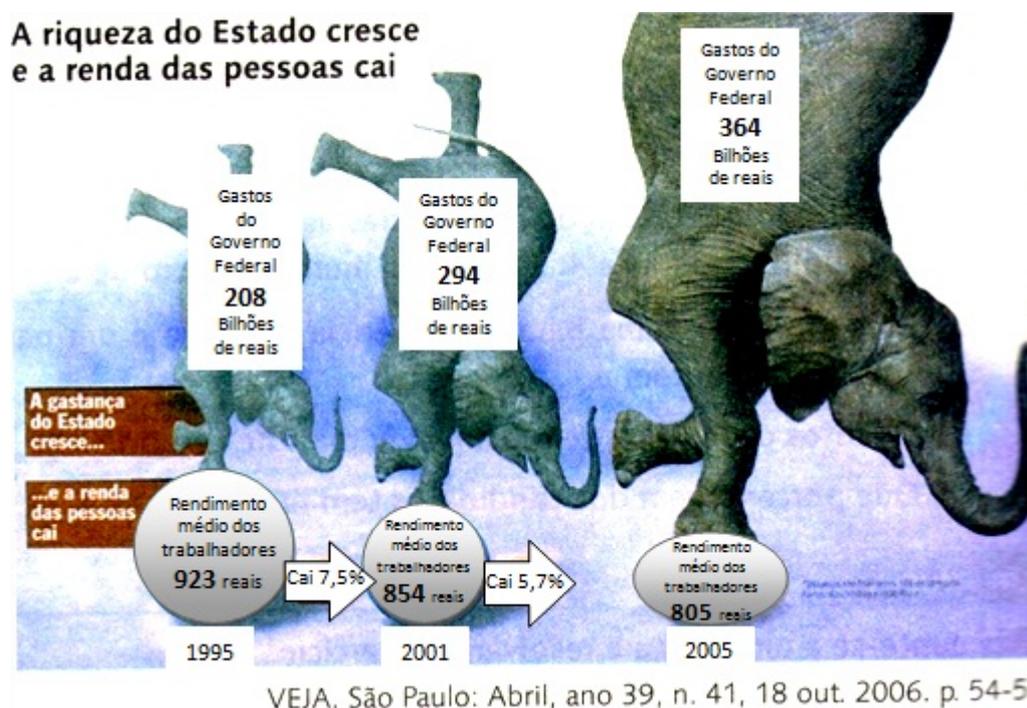
Pictograma

A fim de tornar os gráficos mais atraentes, os meios de comunicação, como revistas, jornais, entre outros, costumam ilustrá-los com imagens relacionadas ao contexto do qual as informações fazem parte. Essa forma de representação é denominada pictograma ou gráfico pictórico.

Nesse tipo de representação, assim como nos gráficos tradicionais, as dimensões das imagens devem ser proporcionais ao dados apresentados.

Veja alguns exemplos de gráficos pictóricos:

Figura 20 – A riqueza do estado cresce e a renda das pessoas cai.



Fonte: <http://professorandrios.blogspot.com.br/2011/08/representacao-grafica-de-dados.html>
 Acessado em: 24/04/16.

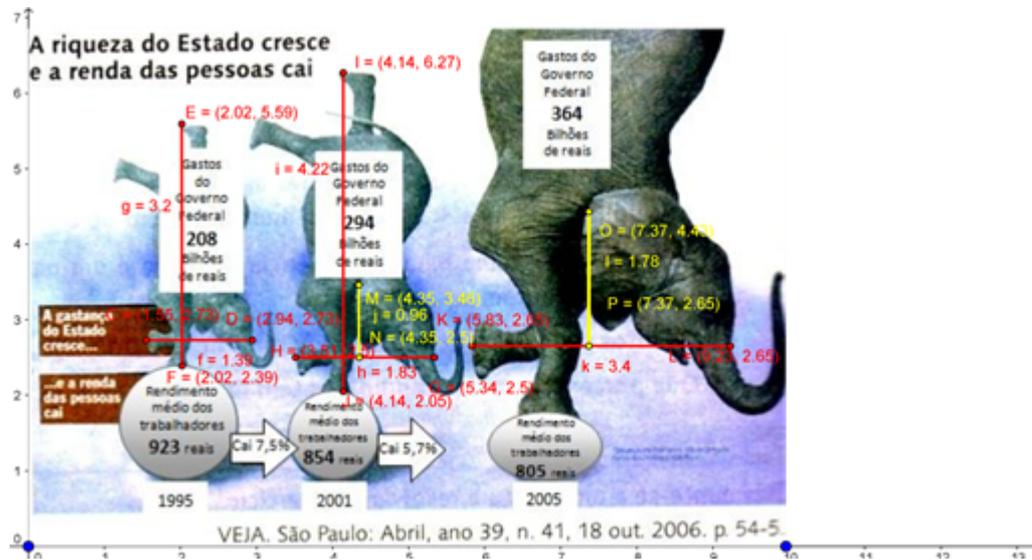
Bom, conforme dito anteriormente, o acima descrito é o encontrado no Blog do Professor de Matemática Andrios Bemfica.

Friso o comentário que é feito por ele: “Nesse tipo de representação, assim como nos gráficos tradicionais, as dimensões das imagens devem ser proporcionais aos dados apresentados.” Professor Andrios Bemfica.

Com base nessa colocação, vamos verificar então se no pictograma acima, existe realmente a proporcionalidade entre os dados e suas respectivas representações gráficas.

Para tal, com a ajuda do programa Geogebra, faremos algumas medições nas imagens (as quais serão apresentadas no decorrer do texto). Mais uma vez salientamos que com o uso de uma boa régua, uma calculadora e conhecimentos básicos de matemática qualquer interessado poderia questionar e verificar as proporcionalidades pretendidas pelo construtor do gráfico.

Figura 21 – A riqueza do estado cresce e a renda das pessoas cai mensurado no Geogebra.

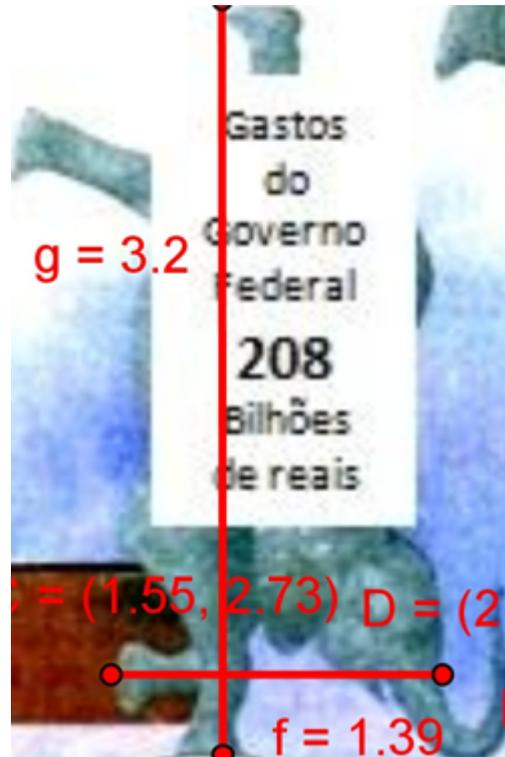


Fonte: o autor

Acima se encontra a representação gráfica analisada no Geogebra. Nela são tratados dois assuntos ao mesmo tempo: o aumento dos gastos do governo federal e o rendimento médio dos trabalhadores em reais. Analisaremos os dois assuntos. Começaremos pelo aumento de gastos do governo federal.

Primeira parte: analisando os gastos do governo federal

Figura 22 – Os elefantes e os segmentos vertical e horizontal.

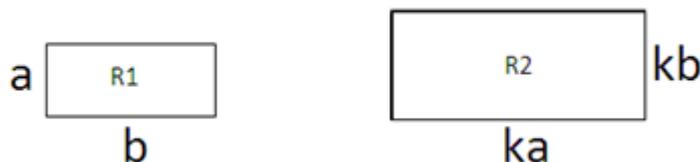


Fonte: o autor

Primeiramente, utilizamos pontos específicos da figura como referência para traçarmos dois segmentos, um vertical e outro horizontal, como mostrado acima. Tais segmentos podem ser interpretados como as dimensões de um retângulo que vai nos servir de referência. Compreendemos que os retângulos criados não absorvem toda a figura dos elefantes, mas, como veremos abaixo, isso não se faz necessário, pois essas dimensões terminam por serem anuladas quando são verificadas as razões entre as áreas. Na verificação das semelhanças entre os retângulos, levamos em consideração o fato de que nas ampliações existem partes dos elefantes que não estão contidas dentro dos retângulos de referência, mas que também obedecem à razão de semelhança verificada entre esses polígonos, pois quando ampliada, a imagem do animal não apresenta distorções, fato esse que acusaria uma falta de proporcionalidade entre as partes internas e externas ao retângulo. Assim, verificaremos se existe a semelhança entre esses retângulos de referência, bem como a semelhança entre as imagens dos elefantes.

Como estamos nos utilizando de retângulos, vamos agora falar sobre a semelhança entre esses polígonos e a relação que existe entre as áreas de retângulos semelhantes. Utilizaremos para tal os retângulos R1 e R2 abaixo.

Figura 23 – R1 E R2.



Possuindo o retângulo R1 as dimensões $a \times b$, temos que ao multiplicarmos seus lados pelo número k , obtemos o retângulo R2 de dimensões $ka \times kb$, que é semelhante a R1 e cuja razão de semelhança entre ele e R1 é de k .

Agora, vamos verificar qual é a razão existente entre suas áreas. Fazendo a razão do maior para o menor, ou seja, a área de R2 sobre a área de R1 temos:

$$\frac{\text{Área de R2}}{\text{Área de R1}} = \frac{ka \cdot kb}{ab} = \frac{k^2 \cdot ab}{ab} = k^2$$

Verificamos assim, que a razão entre as áreas de dois retângulos semelhantes é a razão de semelhança entre suas dimensões elevada ao quadrado.

Em outras palavras: se quisermos determinar a área de R2, basta calcularmos a área de R1 e multiplicá-la pelo quadrado da razão de semelhança de R2 para R1.

Utilizaremos desses recursos de semelhança para investigarmos se existe semelhança entre a figura do primeiro elefante e a do segundo, e entre as figuras do segundo e terceiro elefante. Se forem semelhantes, obteremos a razão de semelhança/ampliação entre esses retângulos, e consecutivamente entre as áreas das imagens dos elefantes, pois, proporcionalmente, os elefantes devem ocupar a mesma porcentagem de área nas figuras.

Dados esses esclarecimentos, estamos aptos para o próximo passo, que é o questionamento das proporções existentes entre os dados dos gastos em reais e suas representações gráficas – os elefantes.

Verificando a proporcionalidade entre os gastos e suas respectivas representações gráficas

No ano de 1995 segundo a revista Veja, o gasto do governo era de 208 bilhões de reais, e, em 2001, o mesmo alcançava a cifra de 294 bilhões de reais. Assim, tivemos um aumento nos gastos de:

$$\frac{294-208}{294} = \frac{86}{294} \approx 0,41346 \approx 41,35\%$$

Pensando nas áreas de retângulos, e dos elefantes, para que esses tivessem fidelidade aos dados em reais, a área do segundo teria que ser 41,35% maior que a área do primeiro.

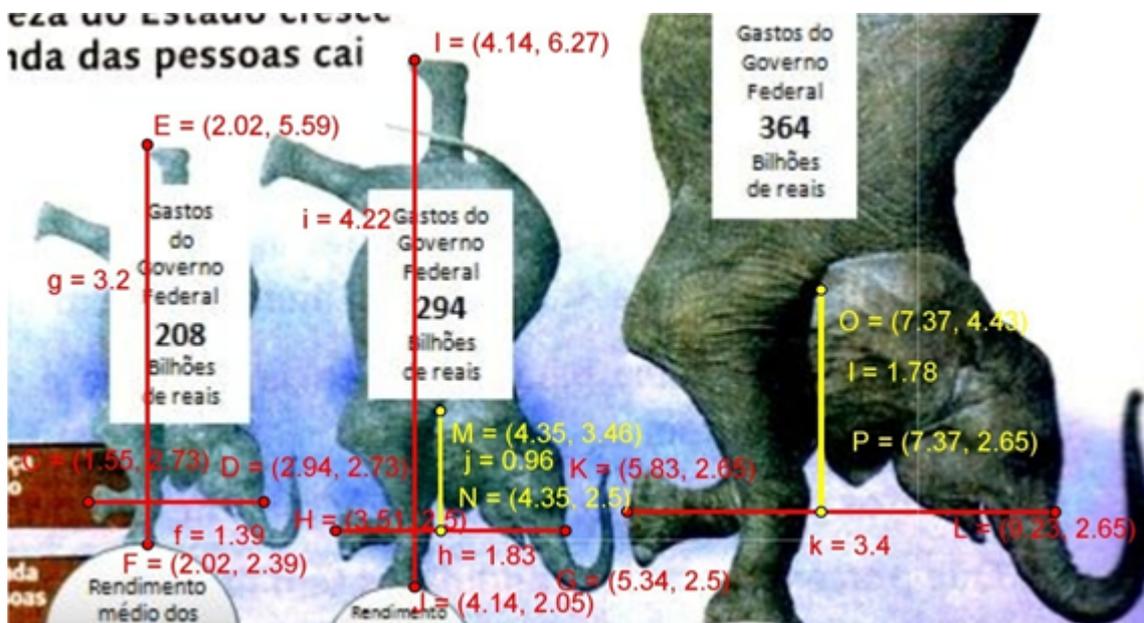
Já de 2001 para 2005, os gastos do governo passaram de 294 para 364 bilhões, tendo assim um aumento de 70 bilhões. Percentualmente temos:

$$\frac{364-294}{294} = \frac{70}{294} \approx 0,2380 \approx 23,80\%$$

Logo, o gráfico de 2005 teria que ser 23,80% maior que o de 2001.

Agora, vamos verificar se os elefantes da figura são ampliações/reduções uns dos outros e consecutivamente semelhantes. Como dito acima fizemos marcações em pontos específicos das três imagens que podem ser vistos abaixo juntamente com as respectivas medições horizontais e verticais feitas no GeoGebra.

Figura 24 – Mensurando as relações existentes entre os elefantes.



Fonte: o autor

Tabela 8 – Razão de semelhança entre o segundo e o primeiro elefante.

	Primeiro Elefante	Segundo Elefante	Razão: segundo/primeiro	Média
Horizontal	1,39	1,83	1,316	1,317
Vertical	3,2	4,22	1,318	

Tabela 9 – Razão de semelhança entre o terceiro e o segundo elefante.

	Segundo Elefante	Terceiro Elefante	Razão: terceiro/segundo	Média
Horizontal	1,83	3,4	1,857	1,855
Vertical	0,96	1,78	1,854	

A partir dos cálculos efetuados (conforme tabelas acima), podemos observar que o segundo elefante é semelhante ao primeiro, com razão de semelhança entre suas dimensões de aproximadamente 1,317; e o terceiro e o segundo também o são, com razão de semelhança nessa ordem de aproximadamente 1,855. A partir desse ponto trabalharemos com essas razões.

Agora, como visto acima a razão de semelhança entre as áreas de duas figuras semelhantes é o quadrado da razão de semelhança, assim, utilizando-nos de duas casas decimais e chamando a razão de semelhança de K , temos:

Tabela 10 – O aumento na área e o aumento correto.

	k^2	(*1) aumento na área	(*2) aumento correto segundo gastos seria de:	aumento indevido na área (*1)-(*2)
segundo/primeiro	1,7345	73,45 %	41,35%	32,1%
terceiro/segundo	3,441	244,1%	23,80%	220,3%

Percentualmente, temos que do primeiro para o segundo elefante o aumento na área foi 77,62% aproximadamente maior que o devido, conforme visto abaixo:

$$\frac{32,1}{41,35} \approx 0,7762 \approx 77,62\%$$

E do segundo para o terceiro elefante o aumento foi de 925,63% maior que o devido:

$$\frac{220,3}{23,80} \approx 9,2563 \approx 925,63\%$$

Isso representa um aumento de quase dez vezes maior que o devido!!!

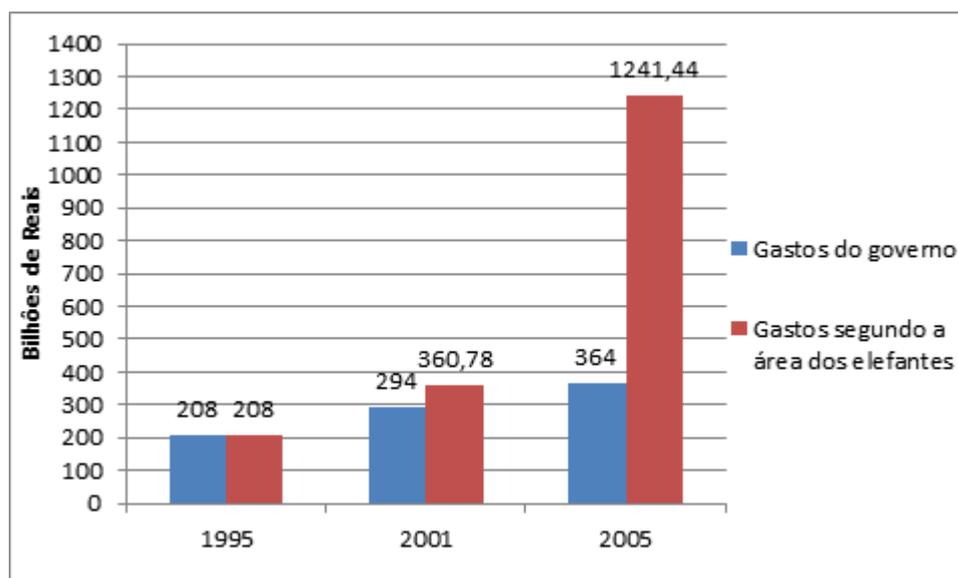
Perceba que esses números se referem apenas às questões de cunho matemático do nosso gráfico. Existem outros fatores que influenciam o leitor sem que talvez ele perceba, como por exemplo:

No elefante menor, as cores são menos vivas, o que faz com que ele chame menos atenção. Além disso, conforme o elefante cresce, as cores vão se tornando mais impressionantes e expressivas, tendo o seu ápice no terceiro, cujas cores são extremamente vivas se comparadas com as dos demais. Isso passa uma sensação de solidez, fazendo com que ele – o terceiro elefante, que está ligado a maiores gastos – se torne mais presente no campo visual do leitor.

Um outro aspecto que merece destaque é a própria representação do governo a partir da figura de um elefante, escolha provavelmente não aleatória. Tal imagem confere ao governo a ideia de algo pesado, talvez lento também, que está esmagando o rendimento médio dos trabalhadores brasileiros.

Agora, vamos fazer um gráfico de barras comparando os gastos e seus aumentos segundo seus dados numéricos, em bilhões, e de quanto seriam os gastos e seus aumentos segundo suas representações gráficas – os elefantes. Para isso aumentaremos os gastos segundo as proporcionalidades existentes nas áreas dos elefantes.

Figura 25 – Comparando os gastos do governo com suas respectivas representações gráficas.



Fonte: o autor

Assim, a representação gráfica utilizada pela revista Veja seria apropriada se o gasto do governo em 2001 tivesse passado para 360,78 bilhões (ao invés de 294 bilhões), e, em 2005, para 1 trilhão 241,44 bilhões de reais (ao invés de 364 bilhões). Bastante diferente da realidade não?

Segunda parte: rendimento médio dos trabalhadores em reais

Na parte inferior do gráfico da Veja encontramos mais informações na forma de um outro pictograma, com dados referentes a uma pequena série histórica dos ganhos médios dos trabalhadores, o qual destacaremos abaixo.

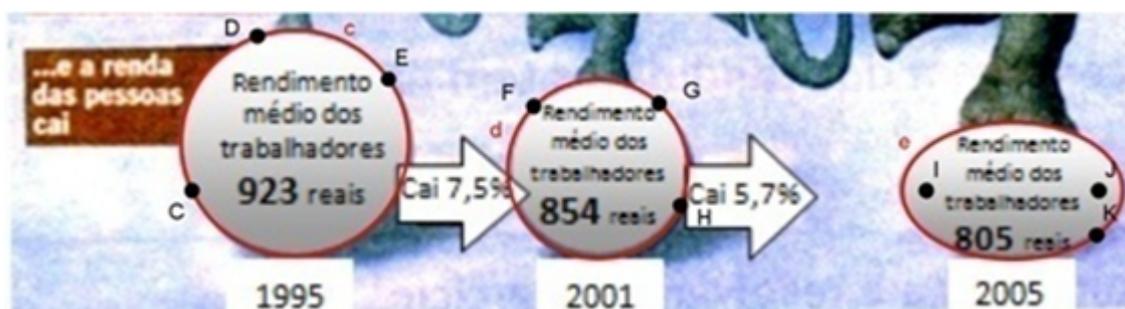
Figura 26 – Recorte no rendimento médio dos trabalhadores.



Fonte: <http://professorandrios.blogspot.com.br/2011/08/representacao-grafica-de-dados.html>
Acessado em: 24/04/16

Após inserir e analisar essa imagem no GeoGebra, temos o seguinte:

Figura 27 – Os círculos e a elipse.



Fonte: o autor

Na representação gráfica do rendimento médio dos trabalhadores de 1995 temos o círculo C; na representação gráfica do rendimento médio dos trabalhadores de 2001 temos o círculo d; e na representação gráfica do rendimento médio dos trabalhadores de 2005 temos a elipse e. Abaixo temos uma tabela com as respectivas equações, elementos e áreas das referidas figuras geométricas.

Tabela 11 – As áreas dos círculos

	Equação	raio de c	área de c
circulo c	$(x - 3,39)^2 + (y - 2,06)^2 = 1,73$	$\sqrt{1,73}$	$1,73\pi$ u.a.
circulo d	$(x - 6,91)^2 + (y - 1,75)^2 = 1,1$	$\sqrt{1,1}$	$1,1\pi$ u.a.

Tabela 12 – A área da elipse.

	Equação	Semi eixo maior	Semi eixo menor	Área de e
Elipse e	$(x - 3,39)^2 + (y - 2,06)^2 = 1,73$	1,28	0,8	$1,24\pi$ u.a.

Agora de posse das áreas, vamos verificar a proporcionalidade das mesmas em relação aos dados.

De 1995 para 2001, segundo a Veja, houve uma queda de 7,5% nos rendimentos médio dos trabalhadores, e, de 2001 para 2005, de 5,7%. Assim, a queda na área que representa esse fenômeno deve ser de igual magnitude. Vamos então verificar se ocorreu dessa forma?

Tabela 13 – As áreas e as porcentagens existentes entre c e d.

	Área de c	Área de d	área de d/ àrea de c	Percentual	Significado
1995 - 2001	$1,73 \pi$	$1,1\pi$	$\frac{1,1\pi}{1,73\pi}$	63,58%	Redução de 36,42%

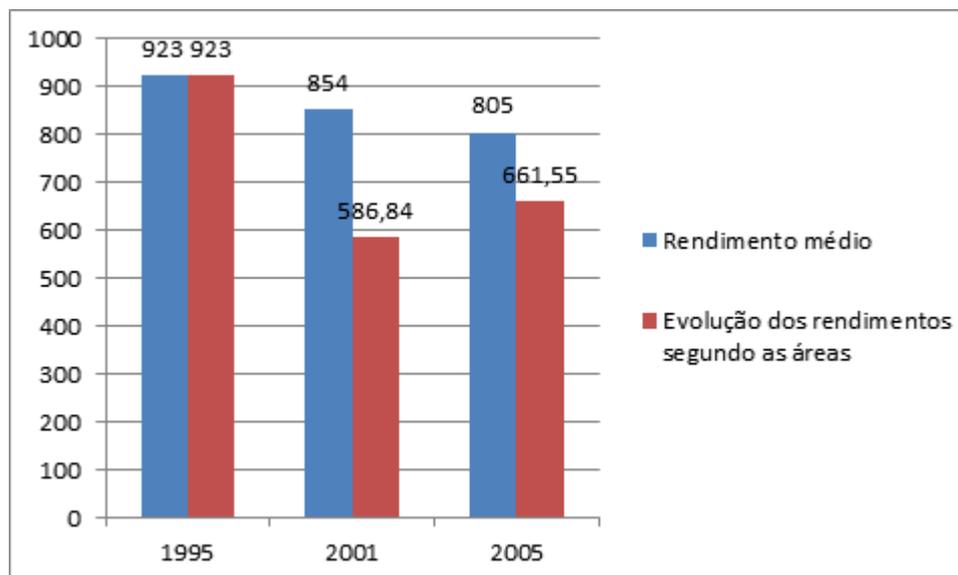
Tabela 14 – As áreas e as porcentagens existentes entre d e e.

	Área de d	Área de e	área de e/ àrea de d	Percentual	Significado
2001 - 2005	$1,1 \pi$	$1,24\pi$	$\frac{1,24\pi}{1,1\pi}$	112,73%	Aumento de 12,73%

A partir dos cálculos acima, temos que de 1995 para 2001, enquanto a representação gráfica deveria sofrer uma queda de 7,5%, sofreu uma queda de 36,42%. Já no período de 2001 a 2005, onde a representação gráfica deveria sofrer uma queda de 5,7%, temos que esta na verdade não sofre sequer uma redução, e sim um aumento, da ordem de 12,73%. Inexplicável.

Vamos agora construir um gráfico de barras comparando a evolução dos rendimentos médios dos trabalhadores em reais e como deveriam estar seus rendimentos médios segundo as relações percentuais das suas respectivas pretendidas representações gráficas.

Figura 28 – rendimentos médios e rendimentos segundo as áreas.



Fonte: o autor

Segundo as áreas analisadas, de 1995 para 2001 o salário vai de R\$ 988 para R\$ 586,04. Já de 2001 para 2005 ele sofre um pequeno aumento, indo de R\$ 586,84 para R\$ 661,55. Ou seja, percebe-se que contrariando todas as proporcionalidades dos dados, ele até sofre um pequeno acréscimo.

Podemos dessa forma frisar que as situações apresentadas acima na construção dos dois pictogramas são extremamente graves, principalmente por se tratar de uma revista de alcance nacional, que influencia a opinião de milhões de brasileiros.

Não menos preocupante o fato desses estarem como exemplo de pictograma em um Blog de um professor de matemática, onde imediatamente abaixo surge a seguinte frase:

Essa forma de representação é denominada pictograma ou gráfico pictórico. Nesse tipo de representação, assim como nos gráficos tradicionais, as dimensões das imagens devem ser proporcionais aos dados apresentados. [12]

O que nos leva a refletir com seriedade na necessidade de se investir em uma educação estatística de mais qualidade em nosso país, que venha a formar cidadãos mais críticos e atentos, capazes de questionar e levantar hipóteses, bem como testá-las com ferramentas adequadas.

3.4 GRÁFICO 4: INFLAÇÃO BRASILEIRA DE 2013

Segundo o site Revista Fórum , no começo de 2014 surgiu o seguinte gráfico de barras construído pela emissora de TV Globo, que pretendia mostrar ao seu telespectador uma série histórica referente à inflação brasileira nos último 5 anos.

Figura 29 – Gráfico de barras Globo News: Inflação do Brasil.



Fonte: <http://www.revistaforum.com.br/blogdorovai/2014/01/11/grafico-da-inflacao-da-globonews-e-o-que-se-pode-chamar-de-vandalismo-jornalístico/> Acessado em: 27/04/16

Ao analisarmos o presente gráfico, vemos que nele surgem duas situações problemáticas:

Primeira: Proporcionalidade

Não se faz necessário recurso extra algum. Olharmos para o gráfico basta para percebermos que a inflação de 2013, que segundo o IBGE foi de 5,91%, supera a de 2010, que foi de 5,92%, e também a de 2011, que foi de 6,50%.

Temos aqui um grave problema de proporcionalidade, pois a representação gráfica dos dados não segue a proporção existente entre eles.

Segunda: O eixo vertical

Note que no gráfico em questão o eixo vertical é iniciado fora do zero. Na verdade

isso quer dizer que este gráfico que foi apresentado pela emissora supracitada é apenas um pedaço de outro gráfico, mais especificamente a parte de cima desse outro gráfico. O intuito dessa manobra (como já mencionado anteriormente) é maximizar, exagerar as diferenças existentes entre as colunas, levando o leitor a conclusões incorretas, direcionando-o de acordo com os interesses dos construtores do gráfico. E isso parece ser justamente o que ocorre acima.

Perceba que não há uma referência, um valor na base das colunas; sendo assim, não nos é informado em que altura esse início ocorre. Mas, se observarmos com atenção, perceberemos que a segunda coluna, que representa o valor de 5,92%, é algumas vezes maior que a primeira, que representa o valor de 4,31%. Mas, como essa pode ser algumas vezes maior que a primeira se 5,92% é um valor ligeiramente maior que 4,31%? Há então neste caso um forte indício de que o eixo vertical, a escala vertical, não se inicia no valor zero, e sim, acima dele. Isso reforça a hipótese de que esse gráfico foi cortado em algum ponto entre 0% e 4,31%.

Perceba que a imagem está em perspectiva, o que causa uma distorção, fazendo com que quanto mais à direita uma coluna, menor a sua dimensão. Assim, fica prejudicada a verificação da proporcionalidade entre as colunas. Como as duas primeiras estão mais à esquerda, sofrem uma menor influência da perspectiva. Assim sendo, vamos tomá-las como referência para a determinação aproximada do ponto onde teria ocorrido o corte no gráfico. As posteriores, como dito acima, sofrem uma maior distorção da perspectiva. Mesmo assim, a razão entre as porcentagens excedentes a 4,31% em cada coluna e a medida em comprimento dessa parte excedente fica em torno de 1, um indício de proporcionalidade entre as colunas, exceto (como dito anteriormente) à referente ao ano de 2013.

Figura 30 – Inflação do Brasil mensurada no GeoGebra



Fonte: o autor

Então, temos como medida do segmento \overline{EF} o valor absoluto 1,64. Como o ponto F possui a mesma ordenada do ponto C, podemos dizer que esse segmento representa em porcentagem a diferença entre 5,92% e 4,31%:

$$5,92\% - 4,31\% = 1,61\%.$$

Mas, se 1,64 representa 1,61%, qual a porcentagem que a altura da barra de inflação de 2009 representa?

Para responder a essa pergunta, vamos nos utilizar da seguinte regra de três:

1,64 está para 1,61%, assim como 0,6 está para x, onde x será a porcentagem desejada. Fazendo os cálculos descobrimos que $x = 0,59\%$. Sendo assim, temos que a coluna referente a 2009 e que está ligada à porcentagem de 4,31% deveria representar em porcentagem um valor bem próximo a 0,59%. Com esse dado descobrimos finalmente que o nosso gráfico original foi “cortado” aproximadamente em um valor que é a diferença entre 4,31% e 0,59%. Calculando:

$$4,31\% - 0,59\% = 3,72\%.$$

Dessa forma, nosso eixo horizontal se inicia em um valor próximo a 3,72%. Perceba que essa é uma distorção gravíssima, pois, como apontado acima, potencializa, exagera as

diferenças entre as barras do gráfico, fazendo parecer que a inflação cresceu demais, dando verdadeiros saltos nos anos que se seguem.

Vale mencionar também que a escala de um gráfico é um fator que maximiza ou minimiza as características, diferenças ou semelhanças existentes nos dados apresentados. Se, ao construirmos um gráfico, colocarmos uma unidade na altura para representarmos 50 unidades da variável ao invés de 5 unidades, as diferenças serão amenizadas; em contrapartida, se colocarmos uma unidade na altura para representar 5 unidades ao invés de 50, as diferenças serão valorizadas. A escala do nosso gráfico apresenta esse recurso, aumentando mais ainda a maximização das diferenças, o que já foi feito pelo eixo vertical não iniciado no zero. Pelas nossas medições, mesmo aproximadas, a relação é de quase um para um, ou seja, uma unidade na nossa altura é utilizada para representar aproximadamente 1%, supervalorizando dessa forma as diferenças entre as colunas. Mas, devido ao fato do gráfico estar cortado, a percepção desse fenômeno fica dificultada para alguém com um olhar menos atento/técnico.

Finalizando esse exemplo de uso ruim de um recurso gráfico, que sofreu dois tipos de distorção, lembremo-nos da última barra, referente a 2013, desproporcionalmente maior que todas as anteriores, o que contribui significativamente para a construção de uma mensagem que tem como tema central o descontrole inflacionário.

Abaixo encontra-se uma representação de como seria aproximadamente a imagem do gráfico apresentada no início desse caso. Reconstruída a partir do analisado, podemos perceber que mesmo com o problema de perspectiva a mensagem transmitida é totalmente distinta daquela que adentrou nos lares brasileiros.

Figura 31 – Um possível gráfico da inflação do Brasil sem o corte

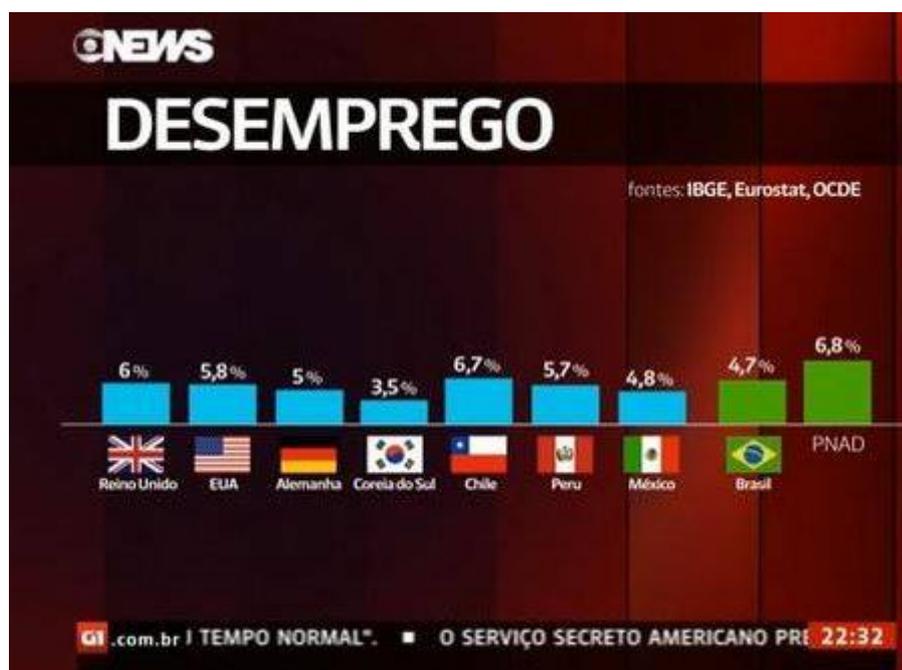


Fonte: o autor

3.5 GRÁFICO 5: DESEMPREGO EM VÁRIOS PAÍSES

De acordo com o jornal GGN [14], a emissora GloboNews, em novembro de 2014, exibiu o seguinte gráfico de barras em um de seus programas:

Figura 32 – Gráfico Globo News: Desemprego.



Fonte: <http://jornalggn.com.br/noticia/globo-news-manipula-graficos-contrataxa-de-desemprego-brasileira/> Acessado em: 05/05/16

O material acima tinha como objetivo comparar a taxa de desemprego do mês de outubro de 2014 entre um grupo de países, dentre eles o Brasil.

A taxa de 6,8 % que surge à direita é referente ao índice de desemprego brasileiro no terceiro trimestre de 2014 segundo a PNAD (Pesquisa Nacional por Amostra de Domicílios), e parece estar descontextualizada considerando-se o assunto tratado no restante do gráfico, que, como dito acima, é o desemprego do mês de outubro.

Parece haver também um jogo de cores que cria uma associação entre essas duas taxas brasileiras – a de outubro de 2014 e a do terceiro trimestre de 2014, ambas as colunas em verde. Em função disso, talvez, um indivíduo que olhe esse gráfico de uma forma um pouco menos observadora possa vir a ser influenciado de forma negativa em relação ao Brasil de duas possíveis formas distintas:

Na primeira, um indivíduo ao tomar contato com o gráfico, vê a coluna verde que se refere ao desemprego do terceiro trimestre, com o número 6,8% associado a ela, valor este maior que todas as outras colunas em número e em altura, e não compreendendo o seu real significado, pode vir a concluir que o desemprego brasileiro é maior que o de todos os outros países.

A outra se encontra em um conjunto de sutilezas. Observe o distanciamento entre as colunas. É sempre igual, a não ser justamente entre a brasileira e a mexicana. Isso,

associado à presença das duas colunas verdes de considerável tamanho no canto direito da tela, forma como que uma massa verde que se destaca no gráfico, massa essa que está associada à ideia de desemprego no Brasil. Mais uma vez, pode-se questionar o impacto de tal estrutura no telespectador.

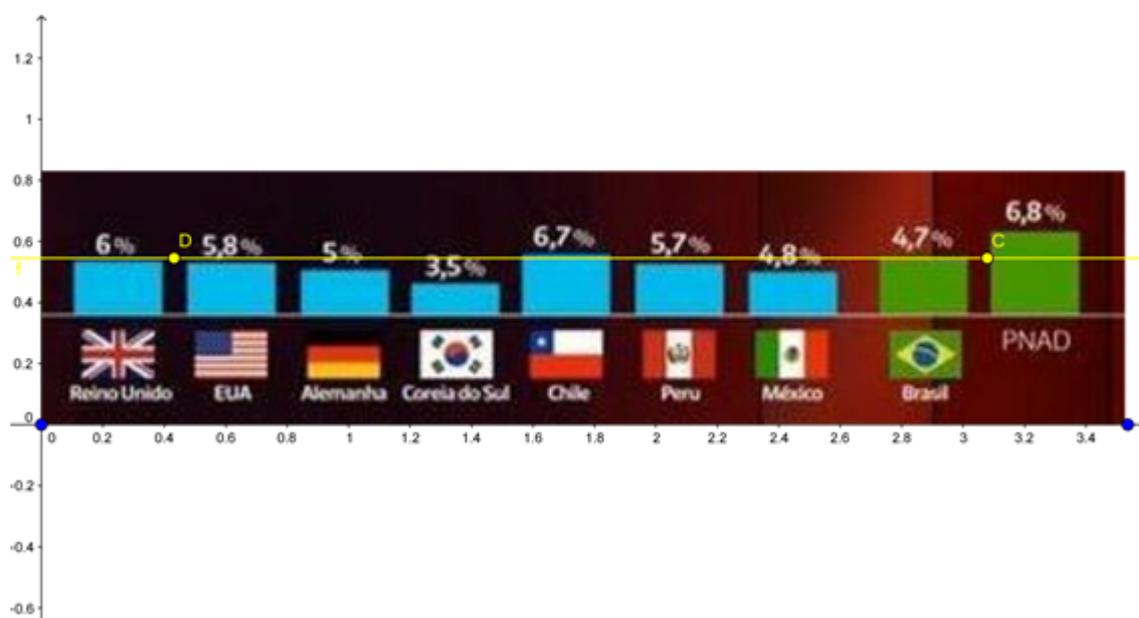
Mas, vamos a partir de agora analisar um pouco esse gráfico através de outro critério, o da proporcionalidade.

Mais uma vez tivemos um problema de desrespeito a essa relação que deve haver entre os dados e seus representantes gráficos.

Bom, a olho nu, ou com o auxílio de uma simples régua, já é possível perceber que a coluna que representa o Brasil no gráfico é uma das mais altas, sendo superada apenas pela coluna que representa o Chile, e por uma diferença tão pequena que chega a existir uma dúvida sobre qual das duas é a maior, ao utilizarmos apenas uma régua para o confronto de ambas.

Assim, para resultados mais expressivos, vamos analisar esse gráfico com o auxílio do programa GeoGebra através da seguinte estratégia: criaremos uma linha paralela ao eixo horizontal por toda a imagem, de mesma altura da coluna referente ao Brasil, e faremos um recorte no gráfico privilegiando sua parte central, buscando assim uma melhor visualização das colunas.

Figura 33 – O desemprego e a referência horizontal.



Fonte: o autor

Acreditamos ser dispensável qualquer comentário sobre a total inexistência da proporcionalidade da barra relativa ao desemprego no Brasil em relação às demais barras,

em azul, que entre si, segundo nossos cálculos, apresentam essa característica. Para compreender isso, basta observar a coluna imediatamente à esquerda da brasileira, a do México: seus 4,8 % de desemprego ficam abaixo – e bastante – dos 4,7 % de desemprego no Brasil.

Também as colunas, em ordem decrescente, do Chile (6,7%), do Reino Unido (6%), EUA (5,8%), Peru (5,7%) e da Alemanha (5%), que deveriam estar todas mais altas que a brasileira, se encontram abaixo dela. Essa desproporção faz parecer, nesse gráfico, que a taxa de desemprego no país é uma das maiores dentre as nações supracitadas, quando, na verdade, ela é menor que as mesmas, sendo maior apenas que a da Coreia do Sul.

Analisando um pouco mais e estabelecendo a coluna de 6% do Reino unido (cuja medida é de 0,47) como padrão, teremos os seguintes resultados:

Figura 34 – As colunas mensuradas no GeoGebra.

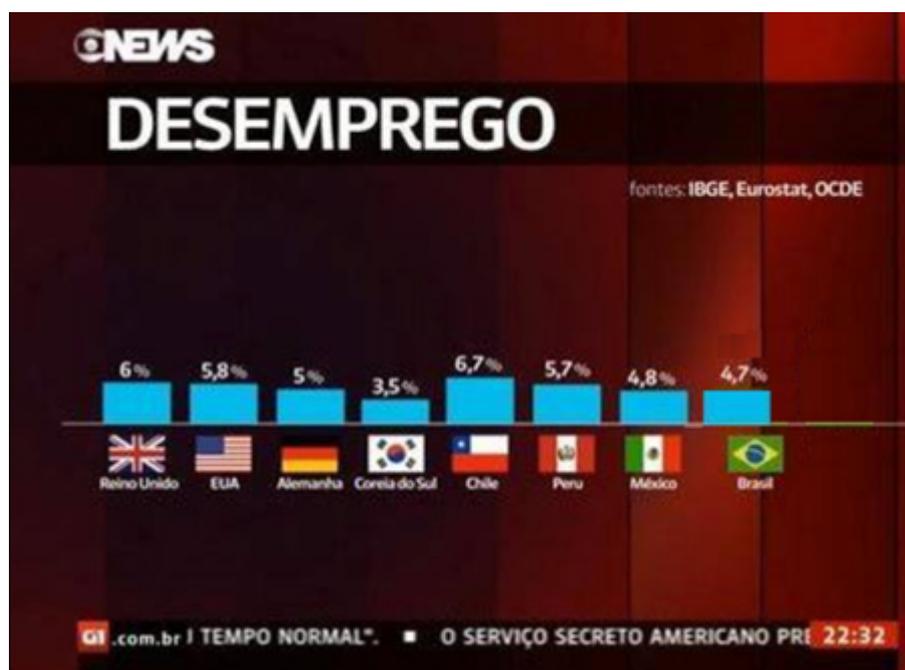


Fonte: o autor

A altura da coluna da inflação de outubro no Brasil, que está ligada ao índice de 4,7%, seria ideal para uma inflação de aproximadamente 6,25%, e a do PNAD, que traz a inflação do terceiro trimestre de 2014, seria ideal para uma inflação de aproximadamente 9,1%.

Construiremos agora, a partir do gráfico original, outro com as proporções corretas, sem a coluna do PNAD e com a coluna do Brasil em azul.

Figura 35 – Desemprego: um possível gráfico ajustado.



Fonte: o autor

Vejam os que nessa nova imagem – com as proporções ajustadas – temos outro olhar sobre a questão da inflação e da taxa de 4,7% no referido período, mais positiva e que comunga com a visão de vários veículos de comunicação, que na época demonstravam positivismo em relação à taxa de 4,7%, conforme os dois exemplos abaixo:

Portal Brasil [2] “Desemprego no Brasil cai ao menor nível para outubro desde 2002, diz IBGE”.

Uol Economia: “Taxa de desemprego no Brasil cai a 4,7%, menor nível para outubro, e renda cresce”. [7]

Dessa forma, temos a dizer como observação final, que o gráfico conforme foi apresentado constrói uma imagem negativa do país perante a opinião pública, gerando assim um provável mal estar em relação à economia e em relação ao governo Brasileiro.

4 CONSIDERAÇÕES FINAIS

A sociedade atual, como já mencionado, é uma sociedade que produz e disponibiliza um volume enorme de informações. Muitas dessas informações – os dados estatísticos – são apresentadas na forma de gráficos estatísticos. Infelizmente, na criação desses, ocorrem erros que podem ser oriundos às vezes de uma má intenção, do desejo de apresentar as informações de uma forma tendenciosa ou simplesmente porque os indivíduos que os construíram não possuíam conhecimento técnico suficiente para fazê-los. Triola em [13, p.55] comenta que “hoje, os gráficos encontrados em jornais, revistas e televisão são, muitas vezes criados por repórteres com conhecimento de jornalismo e comunicação, mas com pouca ou nenhuma experiência no trabalho com dados”. Ainda existem casos onde, tecnicamente falando, os gráficos não contêm erros, mas são enganosos . Também é fato que os meios de comunicação se utilizam dessa linguagem de forma ostensiva e diária.

Tendo em vista essa realidade atual, se torna urgente, como parte do processo que busca a construção do cidadão, o letramento estatístico. Isso significa torná-lo apto a decodificar e atribuir significados corretos para essa forma de comunicação visual. A partir desse momento lhe será possível o lançar de um olhar mais crítico sobre esse tipo de material. Essa tarefa, em parte, cabe aos ambientes educacionais.

Acreditamos fortemente que gráficos e análises como os apresentados no último capítulo do presente trabalho podem prestar um enorme serviço à sociedade ao serem utilizados em tais estabelecimentos como material de discussão entre docentes e discentes, fomentando de forma natural e viva o interesse e a compreensão da importância de um conteúdo tão essencial, o tratamento da informação. “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.” [1, p.136]

REFERÊNCIAS

- [1] BRASIL, Secretaria de Educação Fundamental. **Parametros curriculares nacionais: Matemática**. Brasília: Mec/sef, 1998. 148 p.
- [2] BRASIL, Portal. **Desemprego no Brasil cai ao menor nível para outubro desde 2002, diz IBGE**. 2014. Disponível em: <www.brasil.gov.br/economia-e-emprego/2014/11/desemprego-no-brasil-cai-ao-menor-nivel-para-outubro-desde-2002-diz-ibge>. Acesso em: 23 jun. 2015.
- [3] BRITO, Fernando. **Algum gênio sabe reduzir desigualdade sem impostos?** 2016. Disponível em: <[//www.tijolaco.com.br/blog/algum-genio-sabe-reduzir-desigualdade-sem-impostos/](http://www.tijolaco.com.br/blog/algum-genio-sabe-reduzir-desigualdade-sem-impostos/)>. Acesso em: 05 abr. 2016.
- [4] BUSSAB, Wilson de O.; MORETTIN, Pedro A.. **Estatística básica**. 5. ed. São Paulo: Saraiva, 2004.
- [5] CRESPO, Antonio Arnot. **Estatística fácil**. 15. ed. São Paulo: Saraiva, 1997.
- [6] GERBELLI, Luiz Guilherme. **Classe A tem maior fatia de renda do país**. 2016. Disponível em: <<http://economia.estadao.com.br/noticias/geral,classe-a-tem-maior-fatia-da-renda-do-pais,10000007285>>. Acesso em: 05 mar. 2016.
- [7] GAIER, Rodrigo Viga. **Taxa de desemprego no Brasil cai a 4,7%, menor nível para outubro, e renda cresce**. 2014. Disponível em: <<http://economia.uol.com.br/noticias/reuters/2014/11/19/taxa-de-desemprego-no-brasil-cai-a-47-menor-nivel-para-outubro-diz-ibge.htm>>. Acesso em: 23 jun. 2014.
- [8] GLOBO. **Taxa de desemprego atingiu 9,5% em janeiro**. S.i: Globo Comunicações e Participações S.a., 2016. Disponível em: <<http://g1.globo.com/globo-news/jornal-das-dez/videos/v/taxa-de-desemprego-atinge-95-em-janeiro/4909364/>>. Acesso em: 04 abr. 2016.
- [9] IBGE. **Dimensões do censo 2010**. 2011. Disponível em: <<http://censo2010.ibge.gov.br/sobre-censo/dimensoes-do-censo-2010.html>>. Acesso em: 10 jan. 2016.
- [10] IBGE. **PNAD**. 2011. Disponível em: <<http://www.ibge.gov.br/home/estatistica/pesquisas/pesquisaresultados.php?idpesquisa=149>>. Acesso em: 09 abr. 2016.
- [11] IEZZI, Gelson ; Dolce, Oswaldo; Degenszajn, David; Périgo, Roberto; Almeida, Nilze de. **Matemática: ciência e aplicações, 3: ensino médio**. 5. ed. São Paulo: Atual, 2010.
- [12] SANTOS, Andrios Bemfica dos. **Representação gráfica de dados estatísticos**. 2011. Disponível em: <<http://professorandrios.blogspot.com.br/2011/08/representacao-grafica-de-dados.html>>. Acesso em: 24 abr. 2016.
- [13] TRIOLA, Mario F.. **Introdução à estatística: atualização da tecnologia**. Rio de Janeiro: Ltc, 2014.

- [14] VILLAÇA, Pablo. **Globo News manipula gráficos contra taxa de desemprego brasileira.** 2014. Disponível em: <<http://jornalgnn.com.br/noticia/globo-news-manipula-graficos-contr-taxa-de-desemprego-brasileira>>. Acesso em: 05 maio 2016.