INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA
PROGRAMA DE MESTRADO PROFISSIONAL EM MATEMÁTICA

Recursos Computacionais Voltados ao Ensino de Estatística e Probabilidade

Autor: Mauricio Gonçalo de Carvalho
Orientador: Paulo Cezar Pinto Carvalho

Rio de Janeiro, 22 de fevereiro de 2018

Conteúdo

Resumo **4**
Abstract **5**
Agradecimentos **6**
Introdução **7**
O que é Estatística **8**
A Estatística no Ensino Médio **9**
Análise de livros didáticos *9*
Livro 1: Matemática – Volume único (Iezzi, Dolce, Degenszajn, Périgo) 10
Livro 2: Matemática – Uma nova abordagem – volume 2 (Giovanni, Bonjorno) 11
Livro 3: Matemática – Volume único (Dante) 12
Como a Estatística é abordada na nova Base Nacional Comum Curricular (BNCC) **13**
Material teórico para professores **15**
Termos importantes em Estatística *15*
População 15
Amostra 15
Variável 15
Tabela de frequência, frequência absoluta, relativa e acumulada 16
Gráficos e tabelas *16*
Gráfico de setores (ou de pizza) 17
Gráfico de barras verticais ou horizontais 18
Gráfico de linhas (de segmentos ou poligonal) 20
Gráfico de dispersão 20
Histograma 21
Medidas de posição e suas aplicações *21*
Média aritmética 21
Média aritmética ponderada 22
Mediana 23
Moda 24

Outras médias: média geométrica, média harmônica e média quadrática 25
Desigualdade das médias 26
Medidas de dispersão e suas aplicações *27*
Amplitude 27
Desvio absoluto médio 27
Variância 28
Desvio padrão 28
Sugestões de atividades práticas 30
Atividade prática 1: A desigualdade das médias (Geogebra) *30*
Atividade prática 2: Lançando moedas (Excel) *33*
Atividade prática 3: Estimando o valor de π (Excel) *36*
Atividade prática 4: Quantos peixes há no lago? (Excel) *40*
Atividade prática 5: Pesquisa eleitoral (Excel) *42*
Conclusão 45
Anexo 1: cursos da UNICAMP com disciplinas obrigatórias na área de Estatística 46
Anexo 2: cursos da UERJ com disciplinas obrigatórias na área de Estatística 48
Bibliografia 50

Resumo

Enquanto que o tema Probabilidade é amplamente estudado no Ensino Médio brasileiro, o ensino da Estatística se resume às medidas de posição mais comuns (média aritmética, mediana e moda) e às medidas de dispersão (variância e desvio padrão). Por vezes, o aluno ingressa no Ensino Superior sem saber a aplicabilidade de cada um desses termos. Neste trabalho será sugerida uma série de atividades para que esse estudo seja mais proveitoso para os alunos.

Palavras-chave: probabilidade, estatística, ensino, atividade

Abstract

While the topic Probability is widely studied in Brazilian High School, the teaching of Statistics is reduced to the following topics: mean, median, mode, variance and standard deviation. Hence, many students start their undergraduate studies without knowing the applicability of such terms. In this paper, a series of activities will be suggested to make this study more useful for the students.

Key words: probability, statistics, teaching, activity.

Agradecimentos

Agradeço ao meu pai e minha mãe, Mauricio e Sandra, que me educaram com todo o amor e carinho, e que me deram a oportunidade de estudar.

Aos meus professores de matemática do Ensino Fundamental e Médio: Márcia, Nilza e Ricardo.

À minha esposa, amiga e companheira Juliana, que além de ser uma mulher inspiradora, sempre esteve do meu lado me apoiando.

Aos meus amigos do Mestrado, pelos diversos momentos de estudo que se intercalaram com os momentos de descontração na cafeteria. Um agradecimento especial ao Thomas, Flávio e Diego, os quais tive o privilégio de conhecer nessa caminhada, e que espero manter contato por muito tempo.

Aos meus amigos de Santos-SP, que mesmo estando longe posso senti-los de perto.

A todos os professores do IMPA que nos passaram um pouco de seu vasto conhecimento.

Ao professor Paulo Cezar, que dispôs de seu tempo e paciência para orientar-me neste trabalho.

Introdução

No Ensino Médio, costuma-se abordar o tema Probabilidade de maneira detalhada. Entretanto, isso já não ocorre quando se fala em Estatística.

Influenciada pelos tópicos mais cobrado em vestibulares, o ensino da Estatística no Ensino Médio se resume aos conceitos de média aritmética, mediana, moda, variância e desvio padrão.

Os livros didáticos não costumam trazer explicações detalhadas sobre a aplicabilidade de cada um desses temas, fazendo com que o aluno simplesmente "decore" as fórmulas, visando sua aprovação na disciplina.

O mais surpreendente é que, apesar de ser visto de maneira superficial no Ensino Médio, Estatística é disciplina obrigatória em diversos cursos de nível superior, não apenas na área de exatas. Na Universidade Estadual de Campinas (UNICAMP), por exemplo, todos os cursos da área de Biológicas (exceto Medicina e Fonoaudiologia) possuem alguma disciplina obrigatória na área de Estatística (anexo 1) – inclusive, o curso de Farmácia compartilha uma disciplina com a Engenharia (Estatística para Experimentalistas). O mesmo acontece na Universidade do Estado do Rio de Janeiro (UERJ), onde todos os cursos da área de Biológicas (exceto Medicina e Odontologia) possuem Estatística no currículo obrigatório (anexo 2). Já na PUC Rio, o curso de Ciências Biológicas conta com 3 disciplinas de Estatística Aplicada.

Neste trabalho, serão propostas atividades para que os alunos saiam do Ensino Médio entendendo (principalmente de forma qualitativa):

- Como a Probabilidade pode ser usada na para tomar decisões (Inferência Estatística);
- Por que estudos estatísticos bem feitos (por exemplo, uma pesquisa eleitoral) costumam oferecer resultados bem satisfatórios;
- A importância da amostra em um estudo estatístico.

O que é Estatística

Segundo MAGALHÃES e DE LIMA (2004), Estatística é o conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento. HOLANDA e MUNIZ NETO (2017) afirmam que Estatística é a ciência que coleta, organiza e analisa dados visando responder certas questões cotidianas utilizando um método científico. CORDANI (2006) ressalta também que a Estatística desenvolve metodologia para tomada de decisão em presença de incerteza.

MAGALHÃES e DE LIMA (2004) dividem a Estatística em:

- Estatística Descritiva: conjunto de técnicas destinadas a descrever e resumir os dados, a fim de que possamos tirar conclusões a respeito de características de interesse;

- Probabilidade: teoria matemática utilizada para se estudar a incerteza oriunda de fenômenos de caráter aleatório;

- Inferência Estatística: estudo das técnicas que possibilitam a extrapolação, a um grande conjunto de dados, das informações e conclusões obtidas em uma amostra, geralmente de dimensão muito menor.

Vale ressaltar que alguns autores não consideram a Probabilidade como uma área da Estatística. Entretanto, essa classificação será irrelevante para este trabalho.

A Estatística no Ensino Médio

Como dito anteriormente, Estatística é disciplina obrigatória em diversos cursos de Nível Superior. Entretanto, como já observado por CORDANI (2006), a Estatística é vista timidamente no Ensino Básico (com algumas exceções, é claro). Ela também afirma que:

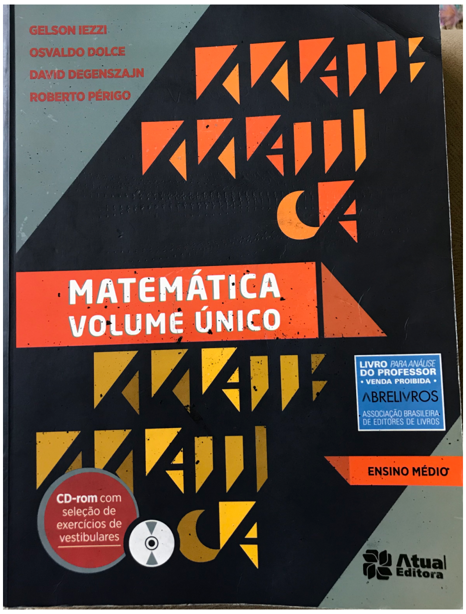
> “por razões históricas, todo o crescimento da área não foi acompanhado da inserção dos elementos básicos de Estatística no currículo escolar (pré-universitário), o que foi um favor decisivo para o despreparo de nossos alunos e professores em relação ao tema. Podemos mesmo dizer que o despreparo é da população como um todo, pois somente os que alcançaram a Universidade (e sabemos a ínfima parcela da população aí incluída) é que tiveram os elementos básicos (ou mais avançados, conforme o caso), da área”. (2006, p. 3)

Isso se ratifica quando são analisados livros didáticos e apostilas: percebe-se que o capítulo de Estatística é dedicado apenas à apresentação de tipos de gráfico, medidas de posição (média aritmética, mediana, moda) e medidas de dispersão (variância e desvio padrão). A respeito das medidas de posição, não se costuma explicar a aplicabilidade de cada uma. A respeito das medidas de dispersão, não se explica para que cada uma serve (os alunos acabam apenas aprendendo as fórmulas). Não se costuma abordar o tema Amostragem e Inferência Estatística.

Análise de livros didáticos

Nos 3 livros analisados, os autores enfocam a Estatística Descritiva e Probabilidade. A parte de Amostragem e Inferência não são abordadas.

Livro 1: Matemática – Volume único (Iezzi, Dolce, Degenszajn, Périgo)



Estruturação do capítulo de Estatística:

- Introdução: Os autores utilizam uma situação real para definir os conceitos de população e amostra. Comentam as etapas de um estudo estatístico (coleta de dados, organização de dados e inferência), e ressaltam que a estatística está presente em várias situações do cotidiano.

- Variável: Explicam o que é uma variável qualitativa e o que é uma variável quantitativa.

- Tabelas de frequência: Frequência absoluta e frequência relativa

- Representações gráficas: Gráfico de setores, de barras, histograma e gráfico de linhas.

- Medidas de centralidade: Média aritmética, média aritmética ponderada, mediana, moda

- Medidas de dispersão: Variância e desvio padrão

- Medidas de centralidade e dispersão para dados agrupados: média, desvio padrão, classe modal, mediana

Comentários: Os autores exploram temas importantes. As medidas de posição principais estão bem explicadas, embora o conceito de moda não tenha exemplo de aplicação. Sente-se falta de outros tipos de média, principalmente a geométrica. Quanto as medidas de dispersão, sente-se falta de mais medidas, como desvio absoluto, desvio absoluto médio, amplitude, e a aplicação de variância e desvio padrão.

Livro 2: Matemática – Uma nova abordagem – volume 2 (Giovanni, Bonjorno)

Estruturação do capítulo de Estatística:

- Introdução: Explica-se o que é a Estatística, como ela é observada no dia-a-dia, e exploram-se conceitos básicos como população, amostra e variável.

- Frequência absoluta

- Frequência relativa

- Representação gráfica da distribuição de frequências: Gráfico de barras, setores, de linhas, pictogramas, análise e interpretação de gráficos

- Distribuição de frequências com dados agrupados: Histograma de frequências e polígono de frequências.

- Medidas de tendência central: Média aritmética, média aritmética ponderada, mediana, moda

- Desvio médio

- Variância e desvio padrão

Comentários: não se comenta sobre a aplicabilidade de cada medida de posição. Também não se explica os motivos de se estudar a variância e o desvio padrão, em detrimento de outras medidas de dispersão.

Livro 3: Matemática – Volume único (Dante)

Estruturação do capítulo de Estatística:

- Introdução: Através de exemplos reais, o autor mostra a presença da Estatística

- Termos de uma pesquisa estatística: São explorados os conceitos de população, amostra, indivíduo, objeto, variável (qualitativa e quantitativa), frequência absoluta e frequência relativa e tabela de frequência.

- Representação gráfica: Gráfico de segmentos, de barras, de setores e histograma.

- Medidas de tendência central: Média aritmética, moda e mediana

- Medidas de dispersão: Variância e desvio padrão

- Estatística e probabilidade: O autor tece comentários sobre como a estatística pode ser usada para se estimar uma probabilidade desconhecida.

Comentários: o destaque vai para o último tópico, que vai ao encontro com um dos objetivos desse trabalho.

Como a Estatística é abordada na nova Base Nacional Comum Curricular (BNCC)

De acordo com o MEC:

> “A Base Nacional Comum Curricular (BNCC) é um documento de caráter normativo que define o conjunto orgânico e progressivo de aprendizagens essenciais que todos os alunos devem desenvolver ao longo das etapas e modalidades da Educação Básica. Conforme definido na Lei de Diretrizes e Bases da Educação Nacional (LDB, Lei nº 9.394/1996), a Base deve nortear os currículos dos sistemas e redes de ensino das Unidades Federativas, como também as propostas pedagógicas de todas as escolas públicas e privadas de Educação Infantil, Ensino Fundamental e Ensino Médio, em todo o Brasil.
>
> A Base estabelece conhecimentos, competências e habilidades que se espera que todos os estudantes desenvolvam ao longo da escolaridade básica. Orientada pelos princípios éticos, políticos e estéticos traçados pelas Diretrizes Curriculares Nacionais da Educação Básica, a Base soma-se aos propósitos que direcionam a educação brasileira para a formação humana integral e para a construção de uma sociedade justa, democrática e inclusiva”.

Em 20 dezembro de 2017, o ministro da Educação homologou a BNCC para Educação Infantil e Ensino Fundamental. Até janeiro de 2018, não havia uma versão definitiva para a BNCC para Ensino Médio. Assim sendo, foi utilizada a prévia de abril de 2016.

A respeito da unidade temática Probabilidade e Estatística no Ensino Fundamental, a BNCC prevê que:

> “Ela propõe a abordagem de conceitos, fatos e procedimentos presentes em muitas situações-problema da vida cotidiana, das ciências e da tecnologia. Assim, todos os cidadãos precisam desenvolver habilidades para coletar, organizar, representar, interpretar e analisar dados em uma variedade de contextos, de maneira a fazer julgamentos bem fundamentados e tomar as decisões adequadas. Isso inclui raciocinar e utilizar conceitos, representações e índices estatísticos para descrever, explicar e predizer fenômenos”.

A mesma BNCC também deixa clara a necessidade do uso tecnologias (como planilhas eletrônicas, um dos objetivos desse trabalho), e que os alunos dos Anos Finais do Ensino Fundamental realizem experimentos aleatórios para confrontar os resultados obtidos com a probabilidade teórica (probabilidade frequentista).

Já no Ensino Médio, a versão preliminar (2016) da BNCC diz que:

> “Ao final da Educação Básica, espera-se que os conhecimentos estatísticos, desenvolvidos desde os anos iniciais, tornem os estudantes aptos para analisar criticamente o que se produz e divulga usando as ferramentas e representações típicas dessa área do conhecimento, muitas vezes de forma imprópria. É comum estarmos diante de generalizações equivocadas de resultados de pesquisa, que não fazem uso adequado da amostragem ou não divulgam como os dados foram obtidos. Outras vezes, observa-se o uso de gráficos inadequados (ou adequados para esconder fatos), ou com problemas de escala, ou de proporcionalidade entre as partes. Assim, a Estatística, nessa etapa, deve estar ainda mais voltada para a discussão e investigação, aumentando-se o rigor das análises de resultados de pesquisas, tanto as realizadas pelos estudantes quanto as encontradas nas diversas mídias, o que é fundamental para o exercício de uma cidadania consciente e ativa”.

Nessa versão preliminar, também é deixado claro que o importante é a capacidade de interpretação do significado de cada medida de posição ou dispersão, e não o cálculo delas, fazendo com que as planilhas eletrônicas sejam ferramentas muito úteis, inclusive na construção e comparação de gráficos.

Material teórico para professores

A ideia é que essa seção seja um manual de consulta para que os professores de Ensino Médio possam ter sugestões para a abordagem de diversos tópicos da Estatística.

Termos importantes em Estatística

População

É o conjunto de objetos, pessoas, etc, que têm em comum a característica ou grandeza que será estudada. Exemplos:

- Em uma pesquisa eleitoral para prefeito(a) de uma cidade, a população é o conjunto formado por todos os eleitores daquela cidade;
- Em uma pesquisa eleitoral para presidente de um país, a população é o conjunto formado por todos os eleitores daquele país;
- Em um estudo sobre a altura média dos homens de um estado, a população é formada por todos os homens daquele estado;
- Em um estudo sobre o gosto musical dos professores de uma escola, a população é formada por todos os professores dessa escola.

Quase sempre é inviável coletar as informações de todos os elementos da população (nos exemplos anteriores, a exceção seria o último deles). Portanto, faz-se necessária a análise de uma amostra dessa população.

Amostra

É o subconjunto da população que será analisado. A partir da análise desse grupo menor, serão tomadas decisões para a população toda (ou serão concluídas informações sobre a população toda). Essa é uma etapa delicada do estudo estatístico. Por exemplo, se uma pesquisa eleitoral para presidente for feita apenas com as pessoas que vivem no Leblon (Rio de Janeiro), o resultado certamente não refletirá as intenções de voto da população brasileira inteira.

Variável

É o objeto de estudo da pesquisa. Por exemplo, quando se estuda a cor dos olhos dos alunos do Ensino Médio de um colégio, a variável “cor dos olhos” pode assumir valores como verde, castanho claro, castanho escuro, azul, preto, etc. Já quando se estuda a quantidade de filhos nas famílias de Manaus, a variável "número de filhos" pode assumir apenas números naturais: 0, 1, 2, 3, ...

No primeiro exemplo, temos uma variável qualitativa, e no segundo, uma variável quantitativa.

Tabela de frequência, frequência absoluta, relativa e acumulada

Em muitos estudos estatísticos, os dados são armazenados em tabelas de frequência, possibilitando a rápida leitura dos resultados obtidos.

O número de vezes que uma determinada variável aparece recebe o nome de frequência absoluta. Se dividirmos esse número pelo número total de dados, obtemos a frequência relativa (um número entre 0 e 1, ou 0% e 100%). Já para variáveis quantitativas, pode fazer sentido o conceito de frequência acumulada (que pode ser absoluta ou relativa), quando quer se contabilizar o número de observações que são menores (ou maiores) que um certo valor. Observe a seguinte tabela de frequência, que mostra o número de aparelhos de televisão (n) nos 50 apartamentos de um edifício:

Número de aparelhos de televisão	**Frequência absoluta**	**Frequência relativa**	**Frequência acumulada absoluta**	**Frequência acumulada relativa**
0	0	$\frac{0}{50} = 0 = 0\%$	0	0%
1	2	$\frac{2}{50} = 0{,}04 = 4\%$	$0 + 2 = 2$	$0\% + 4\% = 4\%$
2	5	$\frac{5}{50} = 0{,}1 = 10\%$	$2 + 5 = 7$	$4\% + 10\% = 14\%$
3	15	$\frac{15}{50} = 0{,}3 = 30\%$	$7 + 15 = 22$	$14\% + 30\% = 44\%$
4	25	$\frac{25}{50} = 0{,}5 = 50\%$	$22 + 25 = 47$	$44\% + 50\% = 94\%$
5	3	$\frac{3}{50} = 0{,}06 = 6\%$	$47 + 3 = 50$	$94\% + 6\% = 100\%$

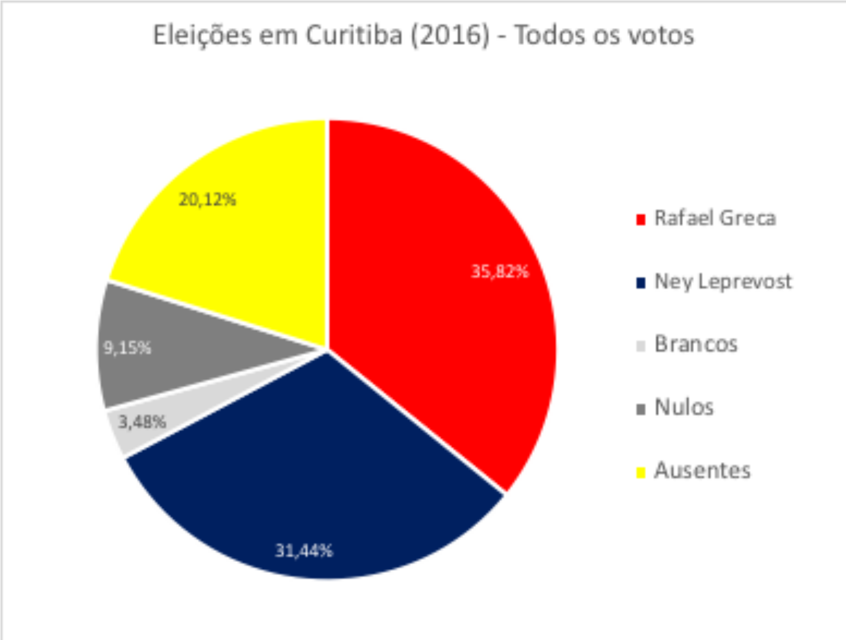
Nesse exemplo, o valor de 44% encontrado na coluna de frequência acumulada relativa indica que 44% dos apartamentos possuem até 3 televisores.

Gráficos e tabelas

São vários os tipos de gráfico. Aqui, serão comentados os principais deles.

Gráfico de setores (ou de pizza)

Apesar de não ser uma regra, costuma ser usado quando o objetivo é analisar a frequência relativa da variável estudada. Por exemplo, em 2016, o resultado do segundo turno das eleições para prefeito de Curitiba foram:



Nesse caso, vence a eleição o candidato que obteve mais votos. Assim, o gráfico anterior é suficiente para saber que o candidato Rafael Greca foi eleito prefeito com os votos de 35,82% dos eleitores da cidade.

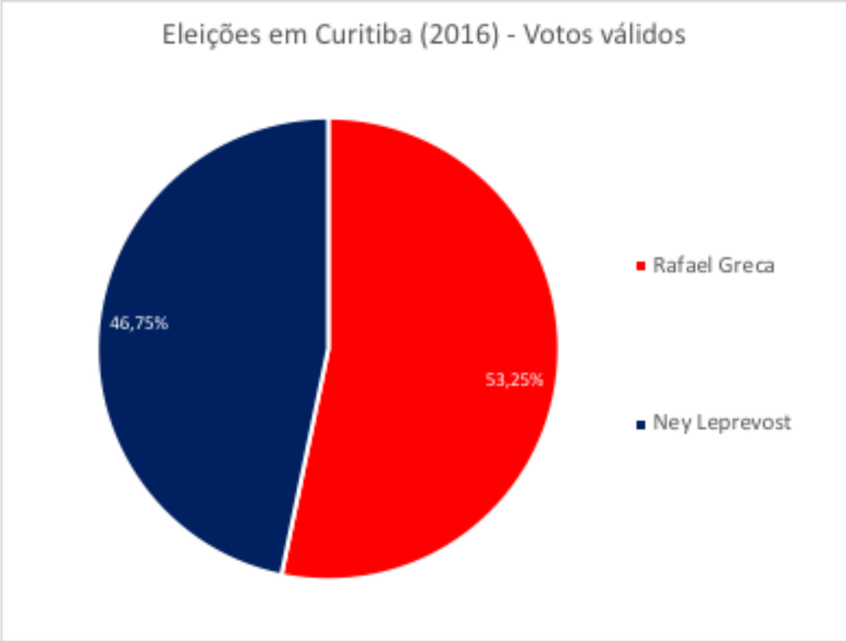
Observação 1: o ângulo central de cada setor é proporcional à frequência variável (ou absoluta) correspondente.

Observação 2: é evidente que o mesmo gráfico poderia ter sido construído com as frequências absolutas. Entretanto, é mais comum observarmos as frequências relativas nesse tipo de gráfico.

Observação 3: analisando-se apenas os votos válidos, o candidato Rafael Greca obteve uma porcentagem de votos igual a

$$\frac{35{,}82\%}{35{,}82\% + 31{,}44\%} = \frac{35{,}82}{67{,}26} \approx 0{,}5325 = 53{,}25\%$$

Logo, o gráfico de setores que representa os votos válidos fica:



Observação 4: o gráfico de setores também pode vir na forma de setores de coroa circular, por exemplo. Esse gráfico também é chamado de “gráfico de rosca”. Observe:

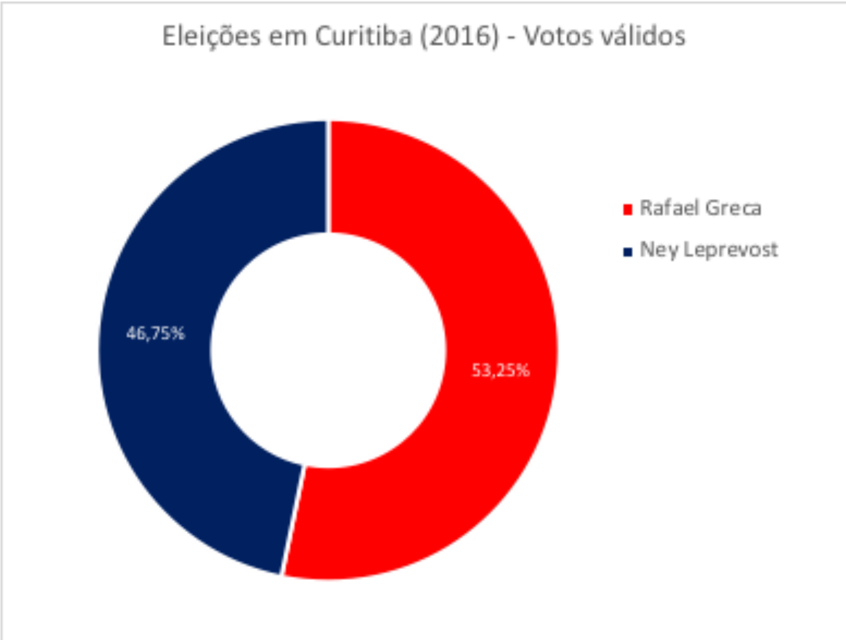
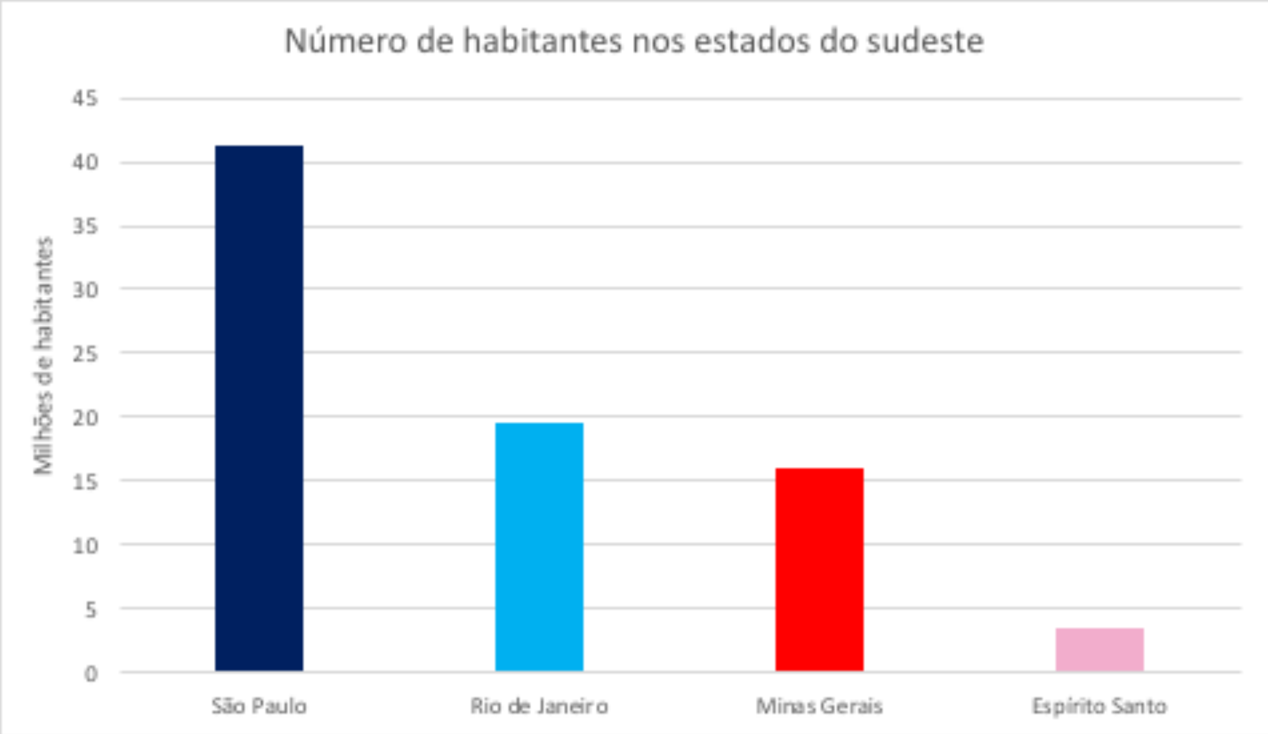
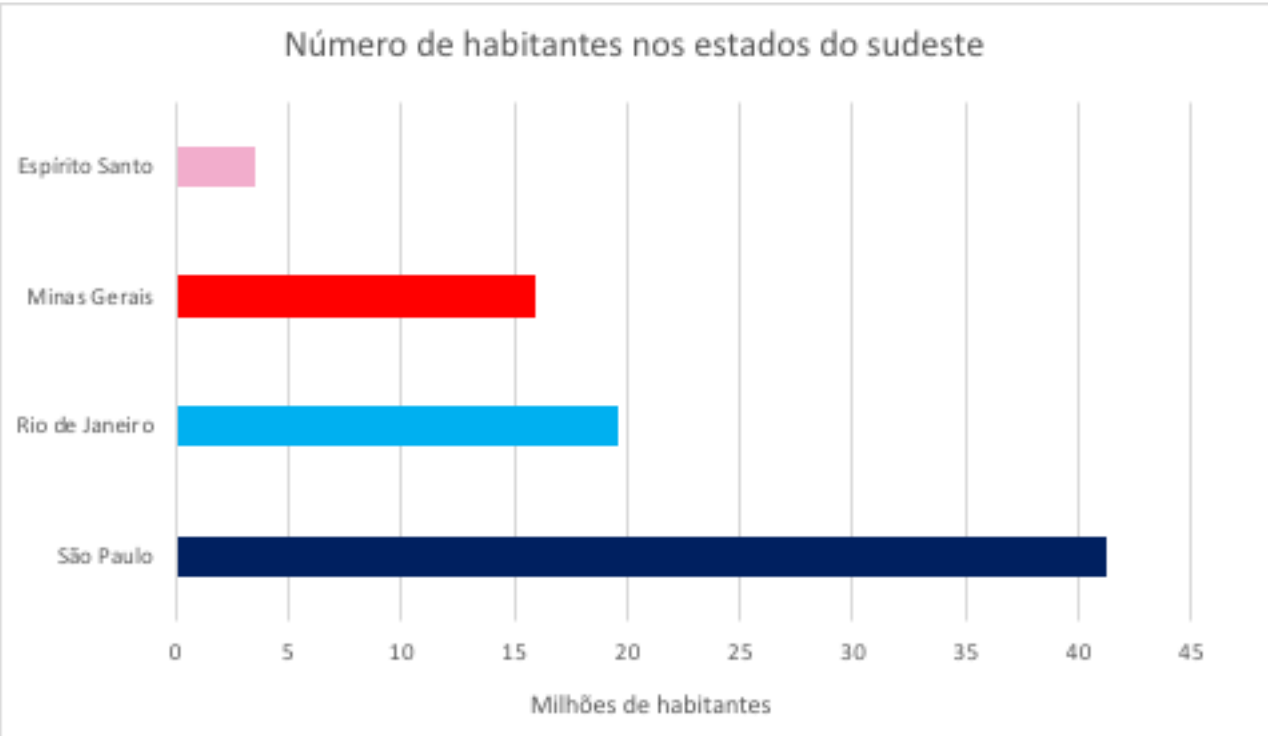


Gráfico de barras verticais ou horizontais

São muito utilizados quando se deseja analisar as frequências absolutas (e não as relativas) dos valores que uma certa variável assume. Por exemplo, o gráfico de barras verticais a seguir ilustra o número de habitantes (por estado) da região sudeste brasileira:



O mesmo gráfico pode vir em barras horizontais:



Nos dois exemplos, o tamanho de cada barra é proporcional ao valor correspondente.

Observação: valores retirados de http://mundoeducacao.bol.uol.com.br/geografia/a-populacao-regiao-sudeste.htm (acesso em 29/01/2018)

Gráfico de linhas (de segmentos ou poligonal)

É muito utilizado quando se quer analisar como o valor de uma certa variável variou com o tempo. Por exemplo, as intenções de voto para Presidente da República no primeiro turno (até o dia 30/09/2014) evoluíram da seguinte maneira:

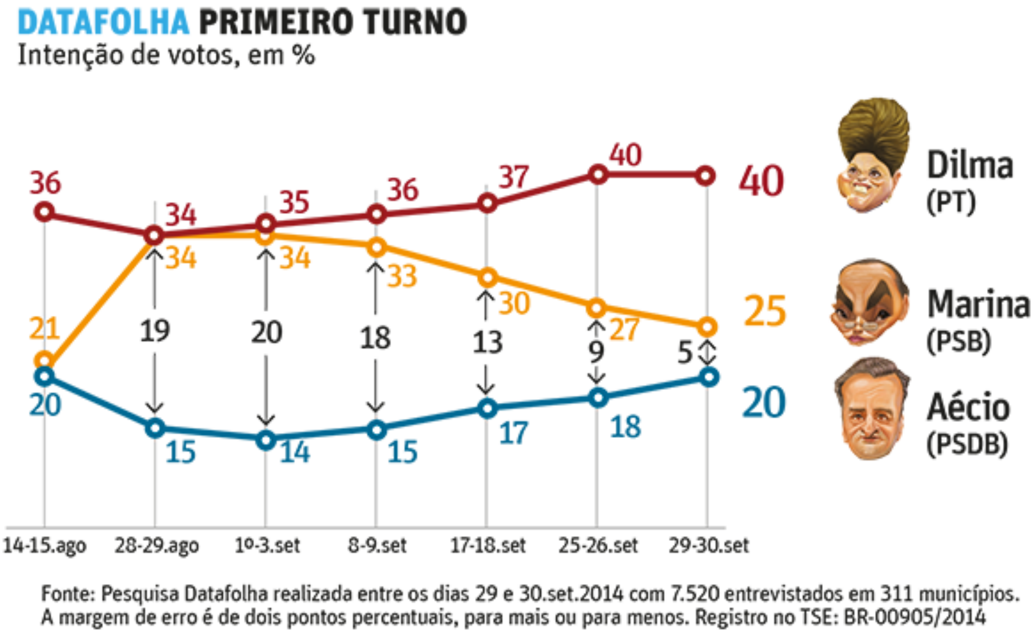
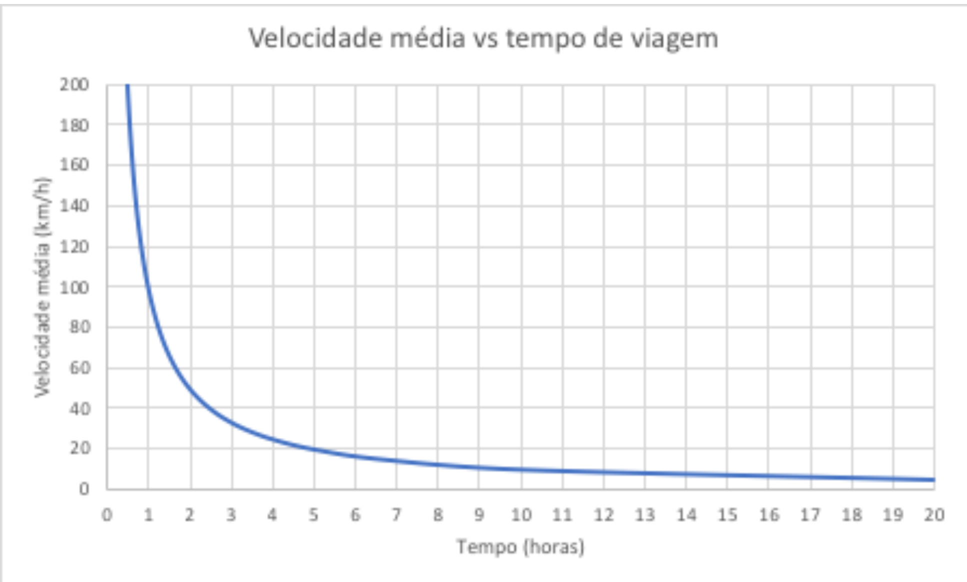


Gráfico de dispersão

É utilizado para relacionar duas variáveis. Por exemplo: em uma viagem de 100 km, a velocidade média e tempo de viagem se relacionam da seguinte forma:



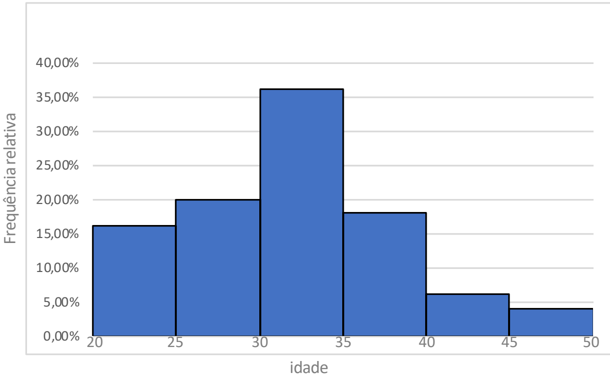
Observação 1: Note que, para uma velocidade média de 20 km/h, o tempo de viagem será de 5 horas.

Observação 2: Caso a variável assuma apenas alguns valores, o gráfico de dispersão será composto de pontos, e não de uma linha.

Histograma

É muito semelhante ao gráfico de barras verticais. Entretanto, ele é usado quando os valores da variável são agrupados em classes. As barras são contíguas, e cada uma das bases corresponde a um segmento cujas extremidades são os extremos do intervalo correspondente. Observe um exemplo de histograma associado a uma tabela:

Idade	Frequência absoluta	Frequência relativa
20 ⊢ 25	80	16%
25 ⊢ 30	100	20%
30 ⊢ 35	180	36%
35 ⊢ 40	90	18%
40 ⊢ 45	30	6%
45 ⊢ 50	20	4%



Observação: o símbolo ⊢ indica que o valor à esquerda é incluído, e o valor à direita excluído.

Medidas de posição e suas aplicações

Mais do que saber como calcular, os alunos devem entender o real significado (e aplicação) de cada medida de posição.

De uma maneira geral, as medidas de posição tentam resumir o comportamento de um conjunto de valores. Entretanto, o conhecimento de apenas uma delas (ou até mais de uma) nem sempre é suficiente para traçarmos o perfil exato desses valores. As principais são:

Média aritmética

É a razão entre a soma de todos os valores e a quantidade desses valores.

Considere n valores: $x_1, x_2, x_3, \dots, x_n$. Sua média aritmética MA é:

$$MA = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Por exemplo, se as notas de 5 alunos foram (4,0; 9,5; 6,0; 4,0; 8,5) a média aritmética dessas notas é:

$$MA = \frac{4 + 9{,}5 + 6 + 4 + 8{,}5}{5} = \frac{32}{5} = 6{,}4$$

Observação 1: também chamamos esse valor como média aritmética simples.

Observação 2: o que representa esse valor de 6,4? LIMA et al. (2016) explicam que uma média é um valor que pode substituir todos os valores de uma lista sem alterar uma determinada característica dessa lista. Caso essa característica seja a soma, temos a chamada média aritmética. Observe:

$$4 + 9{,}5 + 6 + 4 + 8{,}5 = 6{,}4 + 6{,}4 + 6{,}4 + 6{,}4 + 6{,}4 = 32$$

Observação 3: esse exemplo é suficiente para verificarmos que a média aritmética pode ser diferente de todos os valores da lista.

Observação 4: caso os valores estejam agrupados em classes, costuma-se utilizar o ponto médio de cada classe no cálculo de qualquer medida de posição. Por exemplo, a média aritmética das alturas dos 16 alunos da classe representada pela tabela a seguir é:

Altura (m)	Quantidade
1,50 ⊢ 1,60	1
1,60 ⊢ 1,70	4
1,70 ⊢ 1,80	7
1,80 ⊢ 1,90	3
1,90 ⊢ 2,00	1

$$MA = \frac{1 \cdot 1{,}55 + 4 \cdot 1{,}65 + 7 \cdot 1{,}75 + 3 \cdot 1{,}85 + 1 \cdot 1{,}95}{16} = \frac{27{,}9}{16} \approx 1{,}74\ m$$

(este pode ser considerado um caso de média aritmética ponderada, a ser explicada a seguir)

Média aritmética ponderada

Quando se deseja que determinados valores tenham uma maior (ou menor) importância que outros, podemos usar a média aritmética ponderada. Considere n valores: $x_1, x_2, x_3, \ldots, x_n$, com pesos respectivamente $p_1, p_2, p_3, \ldots, p_n$. A média aritmética ponderada MAP desses valores é:

$$MAP = \frac{p_1x_1 + p_2x_2 + p_3x_3 + \cdots + p_nx_n}{p_1 + p_2 + p_3 + \cdots + p_n}$$

Por exemplo, se as notas bimestrais de um aluno foram 10 (peso 1), 8 (peso 2), 6 (peso 3) e 4 (peso 4), a média aritmética ponderada desses valores será:

$$MAP = \frac{1 \cdot 10 + 2 \cdot 8 + 3 \cdot 6 + 4 \cdot 4}{1 + 2 + 3 + 4} = \frac{60}{10} = 6$$

Observação 1: caso tivéssemos feito a média aritmética simples, teríamos encontrado um valor diferente: $\frac{10+8+6+4}{4} = \frac{28}{4} = 7$

Observação 2: a média aritmética simples é um caso especial no qual todos os valores possuem o mesmo peso

$$MAP = \frac{px_1 + px_2 + px_3 + \cdots + px_n}{p + p + p + \cdots + p} = \frac{p(x_1 + x_2 + x_3 + \cdots + x_n)}{pn} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = MA$$

Observação 3: no exemplo dado, poderia ter sido pensado da seguinte forma:

- Mantém-se a nota 10 com peso 1;
- A nota 8 com peso 2 pode ser substituída por 2 notas iguais a 8, com peso 1;
- A nota 6 com peso 3 pode ser substituída por 3 notas iguais a 6, com peso 1;
- A nota 4 com peso 4 pode ser substituída por 4 notas iguais a 4, com peso 1.

Nessa situação hipotética, o aluno fez 1 + 2 + 3 + 4 = 10 provas, cuja média aritmética simples é:

$$MA = \frac{10 + 8 + 8 + 6 + 6 + 6 + 4 + 4 + 4 + 4}{10} = \frac{60}{10} = 6$$

Mediana

Antes de qualquer coisa, é necessário colocar os valores em ordem crescente (ou decrescente). Daí, são duas possibilidades:

- Caso a quantidade de valores seja ímpar, a mediana (Md) será o termo central dessa lista ordenada;
- Caso a quantidade de valores seja par, a mediana (Md) será a média aritmética dos dois termos centrais dessa lista.

Exemplo 1: Calcule a mediana de 5, -4, 8, 1, 10.

Ordenando os valores: -4, 1, 5, 8, 10. Assim, $Md = 5$.

Exemplo 2: Calcule a mediana de 8, 12, 7, 5, 0, 5.

Ordenando os valores: 0, 5, 5, 7, 8, 12. Assim, $Md = \frac{5+7}{2} = \frac{12}{2} = 6$.

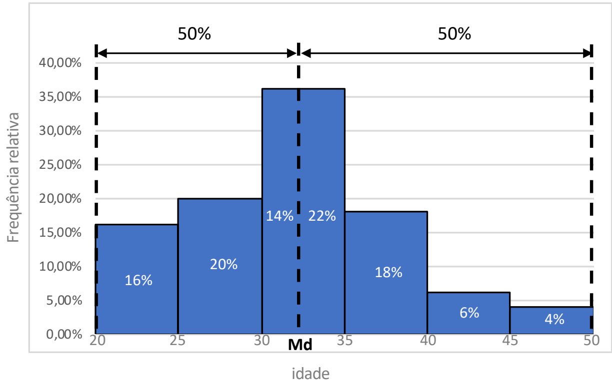
Observação 1: dizemos que a mediana divide os valores da amostra ao meio. No exemplo 2, metade dos valores são menores ou iguais a 6 (0, 5, 5) e a outra metade dos valores são maiores ou iguais a 6 (7, 8, 12). Já no exemplo 1, aos retirarmos a mediana, metade dos valores que sobram são menores ou iguais a ela (-4, 1) e a outra metade dos valores são maiores ou iguais a ela (9, 10).

Observação 2: a mediana é uma medida que é pouco afetada por valores discrepantes (*outliers*). Se no exemplo 1 substituirmos o valor 10 por 1.000.000.000, a mediana continuará sendo a mesma, enquanto que a média mudará muito.

Observação 3: quando se deseja, então, eliminar o efeito dos *outliers*, deve-se utilizar a mediana. Entretanto, devido a necessidade de organizar os dados, nem sempre isso é viável.

Observação 4: o cálculo da mediana para valores agrupados é um pouco diferente. Precisamos encontrar qual é o valor que divide a amostra exatamente ao meio. Para isso, costuma-se usar o conceito de proporções. Veja:

Idade	Frequência absoluta	Frequência relativa
20 ⊢ 25	80	16%
25 ⊢ 30	100	20%
30 ⊢ 35	180	36%
35 ⊢ 40	90	18%
40 ⊢ 45	30	6%
45 ⊢ 50	20	4%



$$\frac{Md - 30}{14\%} = \frac{35 - 30}{36\%} \therefore Md \approx 31{,}94$$

Moda

É o valor que mais aparece em uma amostra. Por exemplo:

- Moda de 9, 12, 15, 10, 12, 8: $Mo = 12$
- Moda de 5, 11, 5, 9, 9, 11, 9: $Mo = 9$

É possível que haja mais de uma moda. Por exemplo:

- Moda de 7, 8, 13, 4, 7, 8: $Mo_1 = 7$ e $Mo_2 = 8$

Caso nenhum elemento da amostra se destaque, não há moda. Por exemplo:

- Não há moda em 1, 3, 2, 6, 5, 7
- Não há moda em 2, 3, 2, 4, 3, 4

Observação 1: uma aplicação do conceito de moda seria o segundo turno de uma eleição, onde o candidato ou candidata que obter mais votos válidos vence.

Observação 2: caso os valores estejam agrupados em classes, o conceito de moda pode se transformar em classe modal (a classe que possui mais valores).

Idade	Frequência absoluta	Frequência relativa
20 ⊢ 25	80	16%
25 ⊢ 30	100	20%
30 ⊢ 35	180	36%
35 ⊢ 40	90	18%
40 ⊢ 45	30	6%
45 ⊢ 50	20	4%

A classe modal é **30 ⊢ 35.**

Outras médias: média geométrica, média harmônica e média quadrática

Considere n valores positivos[1] $x_1, x_2, x_3, \dots, x_n$.

Definimos:

- Média geométrica:

$$MG = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n}$$

  Exemplo: a média geométrica de 2, 4 e 8 é:

$$MG = \sqrt[3]{2 \cdot 4 \cdot 8} = \sqrt[3]{64} = 4$$

  Quando substituímos todos os valores de uma amostra por sua média geométrica, o produto é preservado:

$$2 \cdot 4 \cdot 8 = 4 \cdot 4 \cdot 4 = 64$$

- Média harmônica:

$$MH = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}}$$

  Exemplo: a média harmônica de 2, 4 e 8 é:

$$MH = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} = \frac{3}{\frac{7}{8}} = \frac{24}{7} \approx 3{,}43$$

- Média quadrática:

$$MQ = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2}{n}}$$

  Exemplo: a média quadrática de 2, 4 e 8 é:

$$MQ = \sqrt{\frac{2^2 + 4^2 + 8^2}{3}} = \sqrt{\frac{84}{3}} \approx 5{,}29$$

Exemplo 1: a inflação em um ano foi de 10%. No ano seguinte, foi de 20%. Pergunta-se: pode-se dizer que, em média, os preços aumentaram 15% ao ano?

---

[1] Os valores devem ser positivos para que os cálculos façam sentido. Por exemplo, se permitíssemos números negativos, a média geométrica de -4 e 4 seria o número complexo 4i. Se permitíssemos que um dos valores fosse nulo, a média harmônica de 0, 1, 2, 3 seria indeterminada.

A resposta é não. Para verificar, suponha um produto que custe $R\$\ 100{,}00$. Caso ele sofra um aumento de 10%, seguido de outro de 20%, seu valor final passará a ser de $100 \cdot 1{,}10 \cdot 1{,}20 = R\$\ 132{,}00$. Caso ele sofra um aumento de 15%, seguido de outro de 15%, seu valor final passará a ser de $100 \cdot 1{,}15 \cdot 1{,}15 = R\$\ 132{,}25$.

Se quiséssemos obter um aumento anual médio nesse período, teríamos que descobrir qual é o fator multiplicativo f, tal que $100 \cdot f \cdot f = 100 \cdot 1{,}10 \cdot 1{,}20$, ou seja, $f = \sqrt{1{,}10 \cdot 1{,}20} \approx 1{,}1489$, o que representa um aumento anual médio de 14,89%. Note que o fator f é a média geométrica dos fatores $1{,}10$ e $1{,}20$.

Exemplo 2: uma pessoa vai de casa para o trabalho com uma velocidade média de 60 km/h. No fim da tarde ela volta para casa, percorrendo o mesmo caminho com uma velocidade média de 40 km/h. Pergunta-se: pode-se dizer que, em média, a velocidade média dessa pessoa no percurso ida-volta foi de 50 km/h?

A resposta também é não. Primeiramente, lembremos que a velocidade média é calculada como a razão entre o deslocamento e o tempo desse deslocamento. Assim, denotando por v a velocidade média, d a distância entre a casa e o trabalho, t_{ida} o tempo de ida e t_{volta} o tempo de volta, tem-se:

$$v = \frac{2d}{t_{ida} + t_{volta}} = \frac{2d}{\dfrac{d}{60} + \dfrac{d}{40}} = \frac{2}{\dfrac{1}{60} + \dfrac{1}{40}} = 48\ km/h$$

Note que a velocidade média nesse caso foi a média harmônica entre 40 e 60.

Exemplo 3: uma aplicação de média quadrática é o desvio padrão de uma amostra. Esse tema será explicado posteriormente.

Desigualdade das médias

Existe uma relação de ordem conhecida entre as 4 médias citadas anteriormente:

$$MQ \geq MA \geq MG \geq MH$$

Observação 1: para que as médias sejam todas iguais, é necessário que todos os valores da amostra sejam iguais.

Observação 2: se duas dessas médias são iguais, então todas as outras são iguais.

Observação 3: a Atividade Prática 1, sugerida neste trabalho, mostra essa desigualdade para o caso de 2 valores.

Observação 4: a demonstração dessa desigualdade para um número qualquer de elementos envolve o uso de indução finita, e não é o objetivo deste trabalho.

Medidas de dispersão e suas aplicações

Ainda que se saiba o valor da média, mediana e moda de uma amostra ou população, nem sempre isso é suficiente para entender o real comportamento das mesmas. Por exemplo, as três salas de aula a seguir têm mesma média, mediana e moda:

- Sala 1: 4,9; 5,0; 5,0; 5,0; 5,0; 5,0; 5,1 ($MA = 5; Md = 5; Mo = 5$)
- Sala 2: 3,0; 4,0; 5,0; 5,0; 5,0; 6,0; 7,0 ($MA = 5; Md = 5; Mo = 5$)
- Sala 3: 0,0; 0,0; 5,0; 5,0; 5,0; 10,0; 10,0 ($MA = 5; Md = 5; Mo = 5$)

Entretanto, percebe-se que a sala 1 foi a mais homogênea (ou a menos heterogênea), ou seja, as notas variaram pouco, enquanto que a sala 3 foi a mais heterogênea (ou a menos homogênea), com as notas variando bastante. Com isso, estudam-se as medidas de dispersão, que visam avaliar o quão afastados da média os valores da população estão.

Amplitude

É a medida de dispersão mais simples. Consiste na diferença entre o maior e o menor valor observado.

- Na sala 1, a amplitude das notas foi 5,1 – 4,9 = 0,2
- Na sala 2, a amplitude das notas foi de 7,0 – 3,0 = 4,0
- Na sala 3, a amplitude das notas foi de 10,0 – 0,0 = 10,0

É importante salientar que esse valor isolado não é suficiente para se tomar muitas conclusões. Por exemplo, as amostras (3, 3, 3, 3, 3, 3, 9) e (3, 4, 5, 6, 7, 8, 9) tem mesma amplitude. É sempre interessante analisar as medidas de posição e de dispersão conjuntamente.

Desvio absoluto médio

Definimos desvio absoluto como o módulo da diferença entre um determinado valor de uma amostra/população e a média aritmética dessa amostra/população. Assim, se os n valores são $x_1, x_2, x_3, \dots, x_n$, com média aritmética MA, o desvio absoluto médio desses valores é:

$$D = \frac{|x_1 - MA| + |x_2 - MA| + |x_3 - MA| + \cdots + |x_n - MA|}{n}$$

Usando os dados anteriores:

- O desvio absoluto médio da sala 1 foi de:

$$D = \frac{|4{,}9 - 5{,}0| + |5{,}0 - 5{,}0| + |5{,}0 - 5{,}0| + |5{,}0 - 5{,}0| + |5{,}0 - 5{,}0| + |5{,}0 - 5{,}0| + |5{,}1 - 5{,}0|}{7} = \frac{0{,}2}{7} \approx 0{,}029$$

- O desvio absoluto médio da sala 2 foi de:

$$D = \frac{|3,0 - 5,0| + |4,0 - 5,0| + |5,0 - 5,0| + |5,0 - 5,0| + |5,0 - 5,0| + |6,0 - 5,0| + |7,0 - 5,0|}{7} = \frac{6}{7} \approx 0,857$$

- O desvio absoluto médio da sala 3 foi de:

$$D = \frac{|0,0 - 5,0| + |0,0 - 5,0| + |5,0 - 5,0| + |5,0 - 5,0| + |5,0 - 5,0| + |10,0 - 5,0| + |10,0 - 5,0|}{7} = \frac{20}{7} \approx 2,857$$

Entretanto, as medidas de dispersão mais utilizadas são a variância e o desvio padrão.

Variância

A variância de uma população costuma ser representada pela letra V ou σ^2. É a média aritmética dos quadrados dos desvios:

$$V = \frac{(x_1 - MA)^2 + (x_2 - MA)^2 + (x_3 - MA)^2 + \cdots + (x_n - MA)^2}{n}$$

É interessante notar desde o primeiro momento que a unidade da variância será o quadrado da unidade original. Por exemplo, se os dados estão em metros, a variância estará em metros ao quadrado. Se os dados estão em metros por segundo ao quadrado, a variância estará em metros ao quadrado por segundo à quarta. E assim por diante.

- A variância da sala 1 foi de:

$$V = \frac{(4,9 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,1 - 5,0)^2}{7} = \frac{0,02}{7} \approx 0,003$$

- A variância da sala 2 foi de:

$$V = \frac{(3,0 - 5,0)^2 + (4,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (6,0 - 5,0)^2 + (7,0 - 5,0)^2}{7} = \frac{10}{7} \approx 1,429$$

- A variância da sala 3 foi de:

$$V = \frac{(0,0 - 5,0)^2 + (0,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (5,0 - 5,0)^2 + (10,0 - 5,0)^2 + (10,0 - 5,0)^2}{7} = \frac{100}{7} \approx 14,286$$

Observação: quanto menor a variância, mais próximos da média aritmética estão os valores da população (ou, como costuma-se dizer, mais homogênea é a população).

Desvio padrão

O desvio padrão, representado por DP ou σ, é, simplesmente, a raiz quadrada da variância. Com isso, a unidade encontrada volta ser a original.

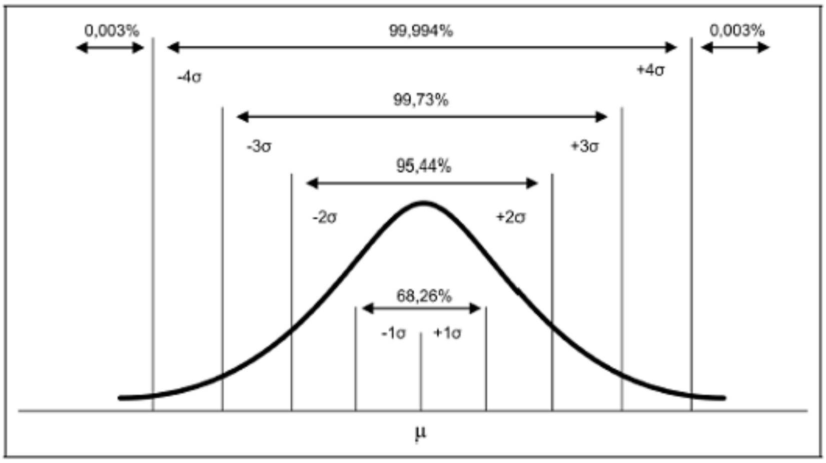
$$DP = \sqrt{\frac{(x_1 - MA)^2 + (x_2 - MA)^2 + (x_3 - MA)^2 + \cdots + (x_n - MA)^2}{n}}$$

Note que o desvio padrão é a média quadrática dos desvios.

- O desvio padrão da sala 1 foi de $\sqrt{\frac{0,02}{7}} \approx 0,053.$
- O desvio padrão da sala 2 foi de $\sqrt{\frac{10}{7}} \approx 1,195.$
- O desvio padrão da sala 3 foi de $\sqrt{\frac{100}{7}} \approx 3,780.$

Uma pergunta que os alunos podem fazer é: para que serve o desvio padrão? Uma maneira de interessante de explicarmos isso de maneira qualitativa seria introduzir o conceito de distribuição normal, comentando sobre a curva em forma de sino, e comentar que:

- Aproximadamente 68% dos valores em uma distribuição normal se encontram em uma vizinhança de 1 desvio padrão da média (ou valor esperado);
- Aproximadamente 95,5% dos valores se encontram em uma vizinhança de 2 desvios padrões da média;
- Aproximadamente 99,7% dos valores se encontram em uma vizinhança de 3 desvios padrões da média.



Fonte: Portal Action

Sugestões de atividades práticas

Aqui serão sugeridas atividades lúdicas para a compreensão de conceitos de probabilidade e estatística, com planilhas eletrônicas para complementar o entendimento dos alunos. Elas estão disponíveis no link a seguir e podem ser reproduzidas, desde que citada a fonte:

https://drive.google.com/drive/folders/1MGOHF3RbpNoMS0h6RBOeFvBziJZTg614?usp=sharing

Atividade prática 1: A desigualdade das médias (Geogebra)

Nessa atividade será provado que, dados a e b números reais positivos, sua média quadrática (MQ), aritmética (MA), geométrica (MG) e harmônica (MH) relacionam-se da seguinte maneira:

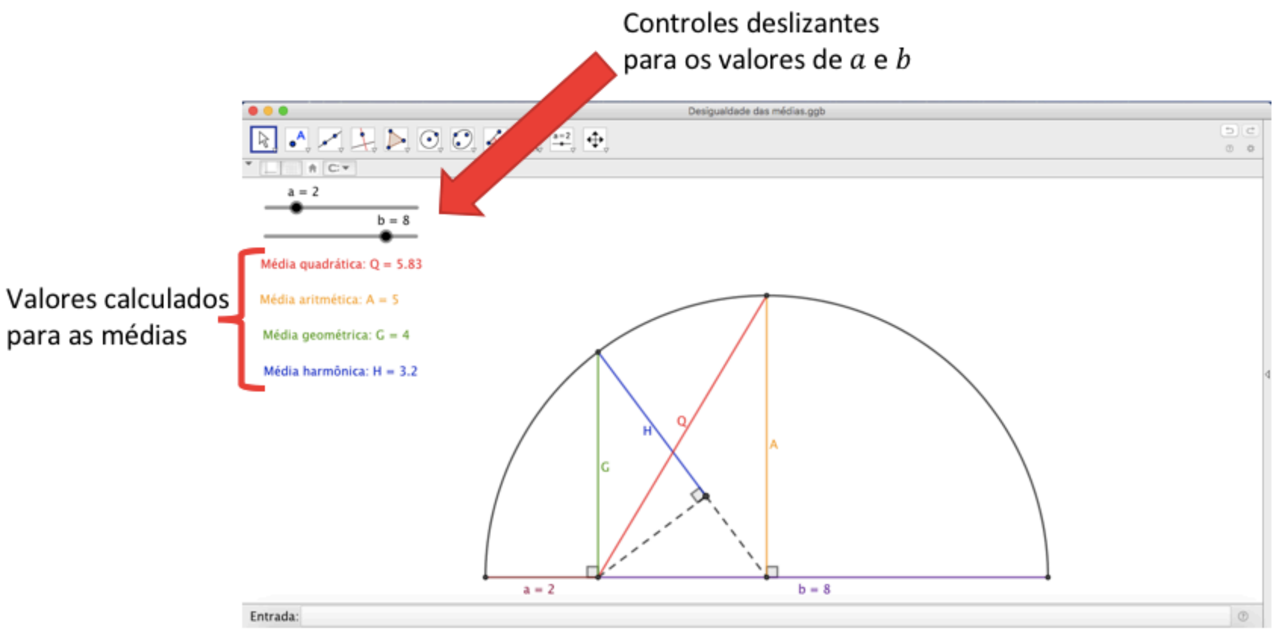
$$MQ \geq MA \geq MG \geq MH$$

Para isso, foi desenvolvida uma atividade no Geogebra, onde essas quatro médias serão visualizadas dentro de uma semicircunferência de diâmetro $a + b$.

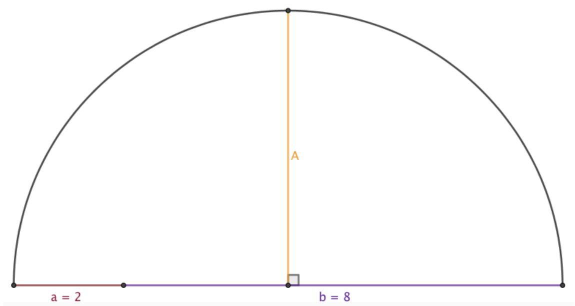
Para o caso $a = 2$ e $b = 8$, temos:

- $MQ = \sqrt{\frac{2^2+8^2}{2}} \approx 5{,}83$
- $MA = \frac{2+8}{2} = 5$
- $MG = \sqrt{2 \cdot 8} = 4$
- $MH = \frac{2}{\frac{1}{2}+\frac{1}{8}} = 3{,}2$

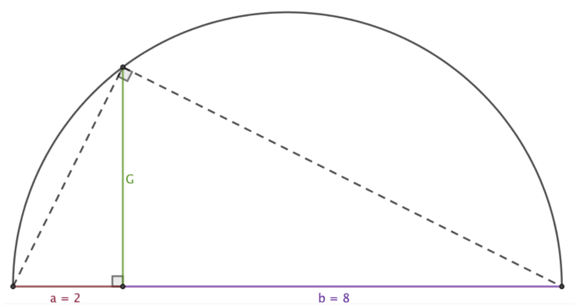
Esse resultado será visualizado como:



Os segmentos, de fato, indicam as quatro médias:



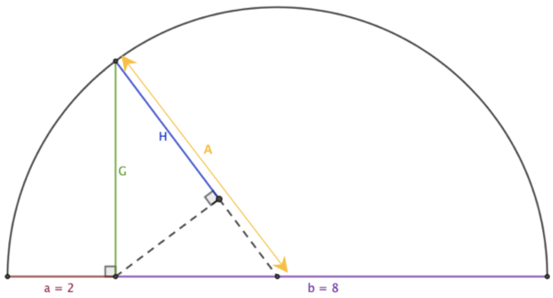
O segmento indicado pela letra A indica um raio da semicircunferência de raio $a + b$. Logo, ele possui medida igual a $\frac{a+b}{2}$, equivalente à média aritmética de a e b.



Traçando o segmento perpendicular ao diâmetro, passando pelo ponto em comum entre os segmentos que representam a e b, determina-se um segmento (indicado por G) com medida equivalente à altura do triângulo retângulo inscrito. Logo, a partir das relações métricas no triângulo retângulo:

$$G^2 = a \cdot b \therefore G = \sqrt{a \cdot b}$$

Desse modo, tem-se a média geométrica de a e b.

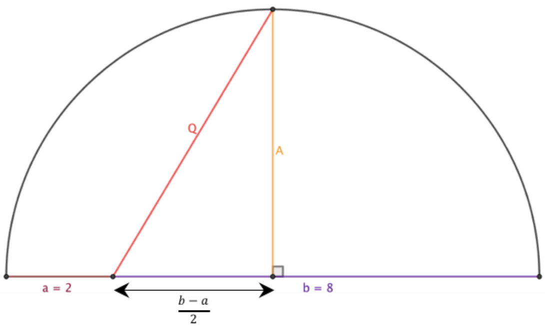


Também das relações métricas no triângulo retângulo, tem-se:

$$A \cdot H = G^2$$

$$H = \frac{G^2}{A} = \frac{a \cdot b}{\frac{a+b}{2}} = \frac{2 \cdot a \cdot b}{a+b}$$

Observe que a última expressão pode ser reescrita como $\frac{2}{\frac{1}{a}+\frac{1}{b}}$ (ou seja, o segmento H tem medida igual à média harmônica entre a e b).



Pelo Teorema de Pitágoras:

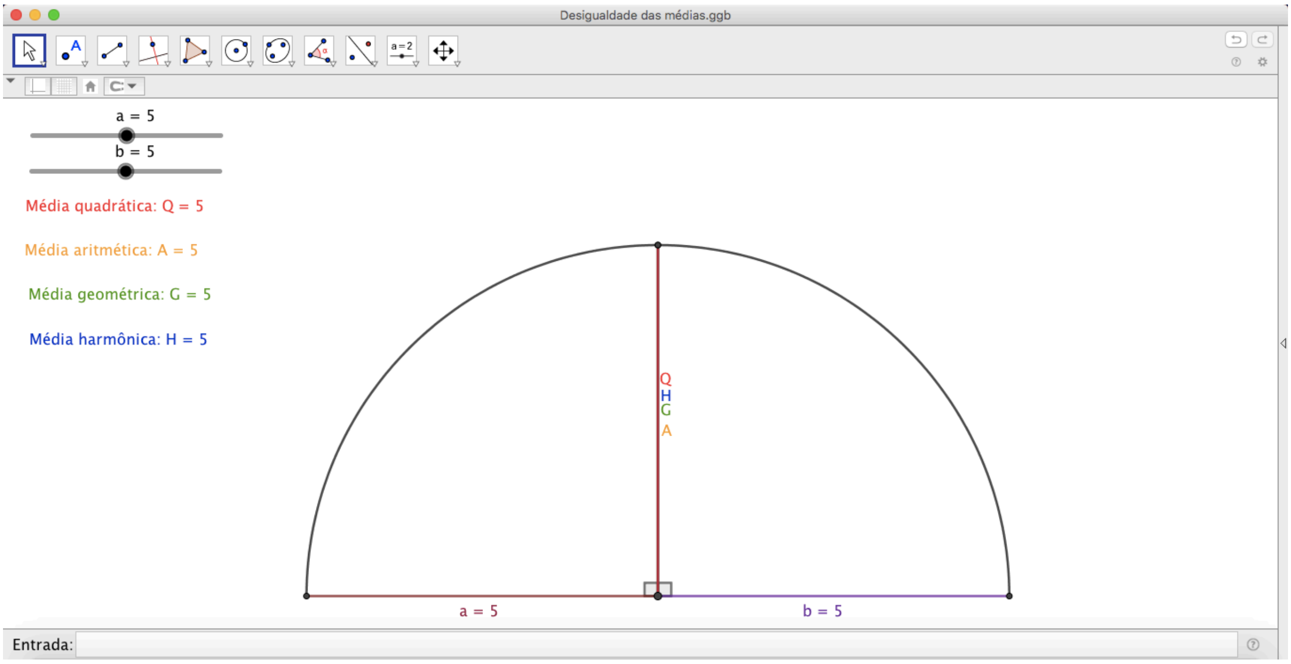
$$Q^2 = A^2 + \left(\frac{b-a}{2}\right)^2$$

$$Q^2 = \left(\frac{a+b}{2}\right)^2 + \left(\frac{b-a}{2}\right)^2$$

$$Q^2 = \frac{a^2 + 2ab + b^2 + b^2 - 2ab + a^2}{4}$$

$Q = \sqrt{\frac{a^2+b^2}{2}}$ (ou seja, o segmento Q tem medida igual à média quadrática entre a e b)

Sugestões: varie os controles deslizantes, e mostre que a desigualdade sempre será verdadeira. Mostre também que as médias se igualarão apenas quando os valores de a e b forem iguais. Por exemplo, caso $a = b = 5$, os quatro segmentos coincidirão, ficando perpendiculares em relação ao diâmetro da semicircunferência:



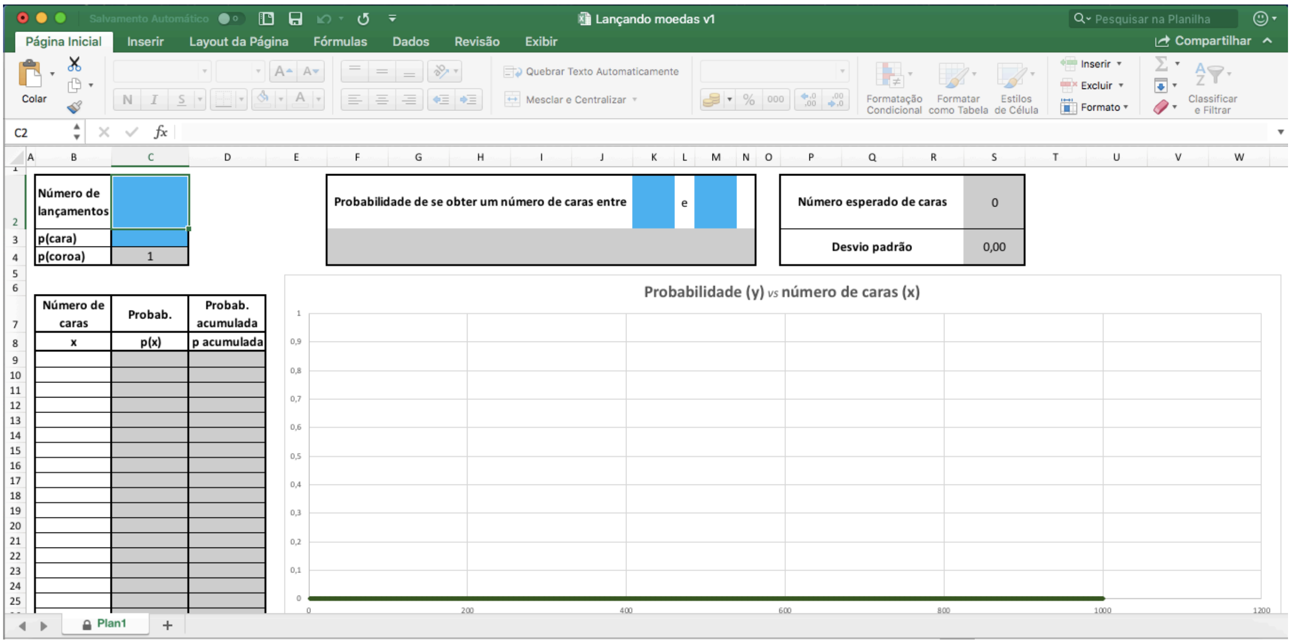
Atividade prática 2: Lançando moedas (Excel)

Ao lançarmos n vezes uma moeda, a probabilidade de se obter x caras é:

$$p(x) = \frac{n!}{x!\,(n-x)!} \cdot (p_{cara})^x \cdot (1 - p_{cara})^{n-x}$$

onde $0 \leq x \leq n$ $(x \in \mathbb{N})$ e p_{cara} representa a probabilidade de se obter a face cara em um lançamento da moeda.

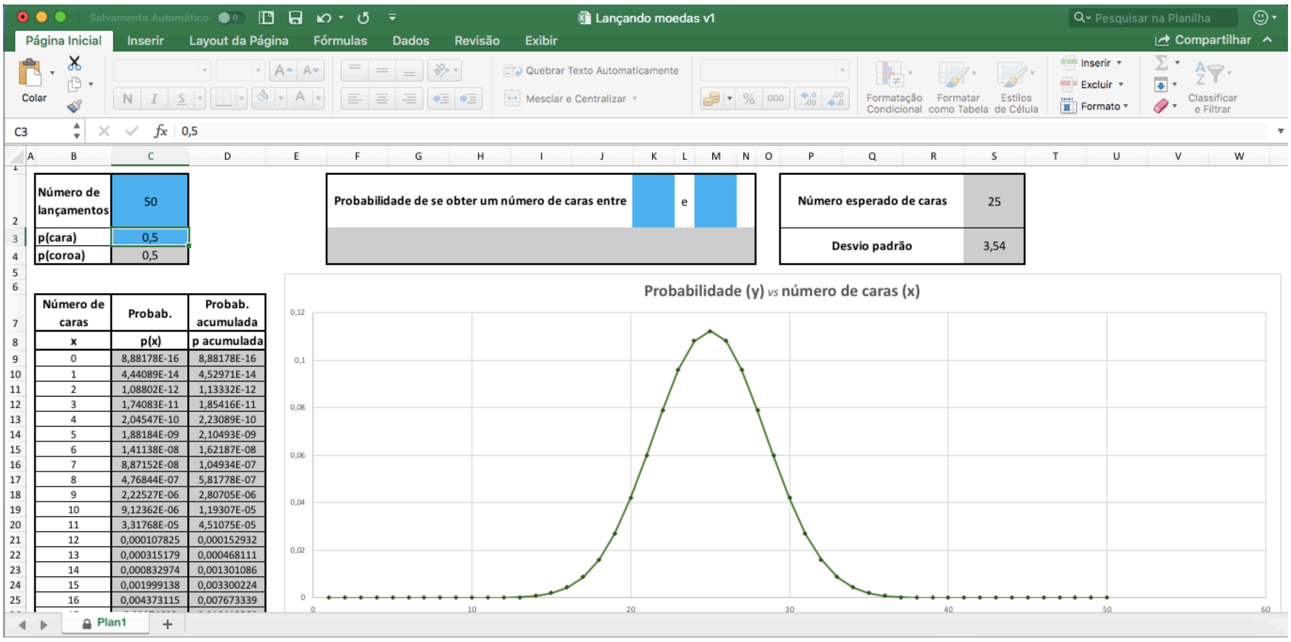
Esse é um problema que pode ser abordado na sala de aula nas aulas de probabilidade. Entretanto, valores muito altos de n tornam a execução das contas inviável. O que a planilha faz é automatizar todas as contas:



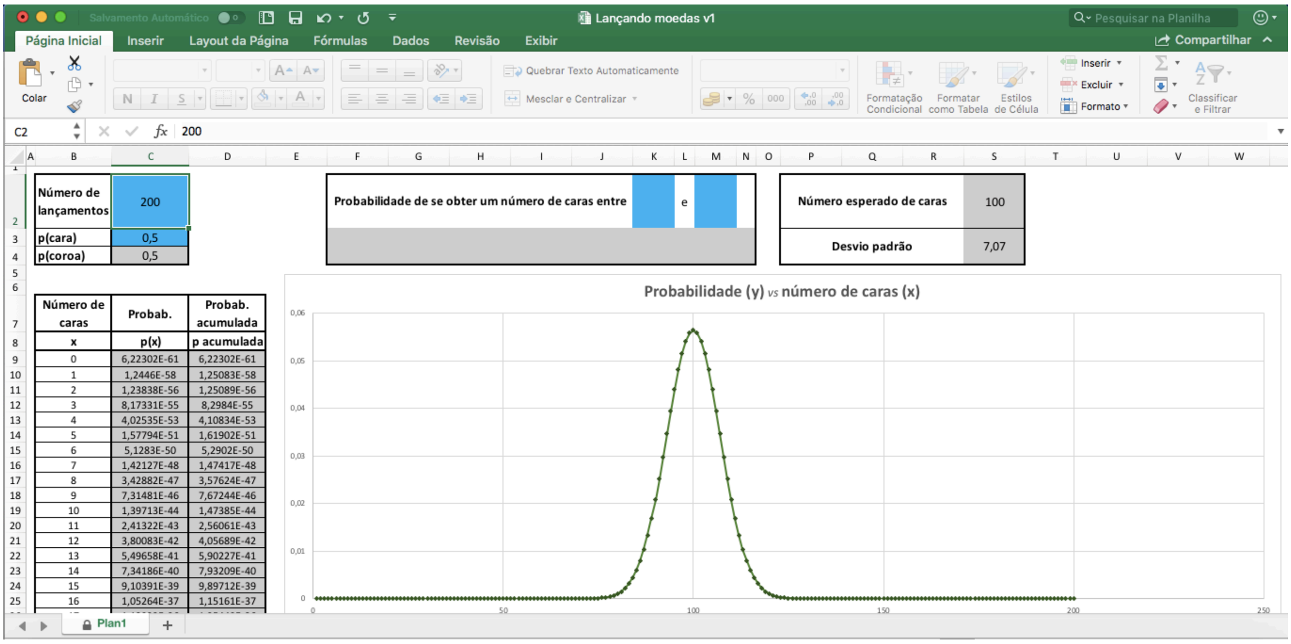
A tarefa do usuário é determinar o valor de n e p_{cara}. A planilha mostrará um gráfico probabilidade vs número de caras, o número esperado de caras e o desvio padrão.

Algumas simulações:

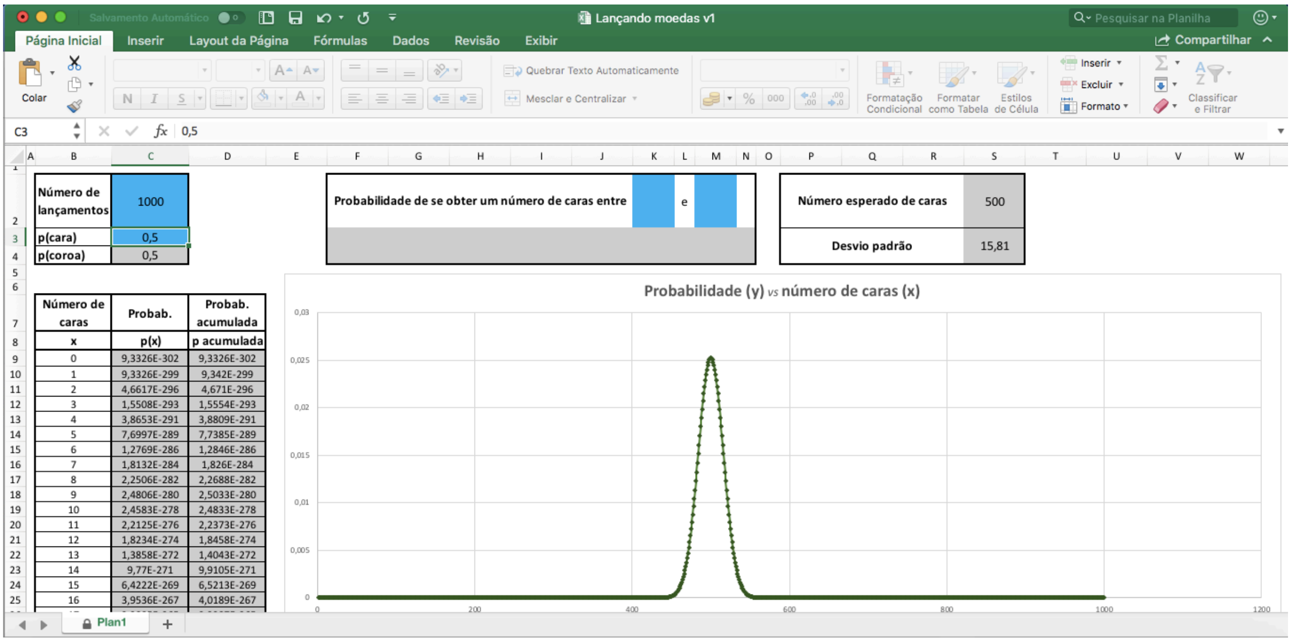
- $n = 50$ e $p_{cara} = 0{,}5$:



- $n = 200$ e $p_{cara} = 0{,}5$:



- $n = 1000$ e $p_{cara} = 0{,}5$:



Comentários pertinentes a serem feitos com os alunos:

- Esse é um problema de distribuição binomial. Entretanto, quanto maior o valor de n, mais próximo de uma distribuição normal essa distribuição fica;
- Para o caso $n = 1000$, justamente por ser comportamento próximo a uma distribuição normal, serão encontrados valores próximos aos teóricos quando estivermos calculando a probabilidade do número de caras estar ao redor de 1, 2 ou 3 desvios-padrão do valor esperado:

Probabilidade de se obter um número de caras entre	485	e	515
67,3063%			

Número esperado de caras	500
Desvio padrão	15,81

Probabilidade de se obter um número de caras entre	470	e	530
94,6322%			

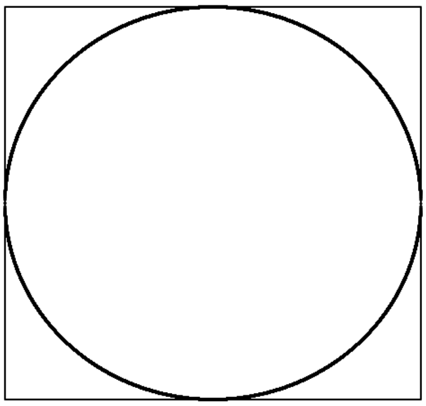
Número esperado de caras	500
Desvio padrão	15,81

Probabilidade de se obter um número de caras entre	455	e	545
99,6016%			

Número esperado de caras	500
Desvio padrão	15,81

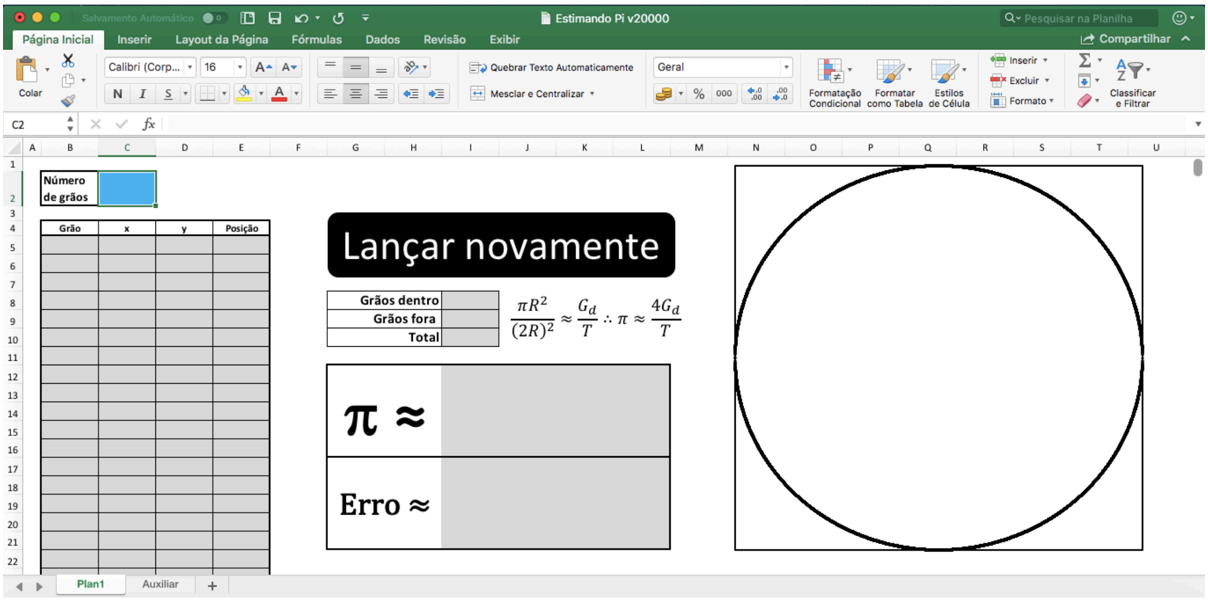
Atividade prática 3: Estimando o valor de π (Excel)

Nessa atividade, será simulado o lançamento de grãos de feijão em uma região quadrada com um círculo inscrito, visando a estimação do número π. Caso os grãos caiam aleatoriamente sobre a região, quanto maior o número de grãos lançados, mais próxima de $\frac{\pi}{4}$ tenderá a razão entre o número de grãos que caem na região circular (G_d) e o total de grãos lançados (T):

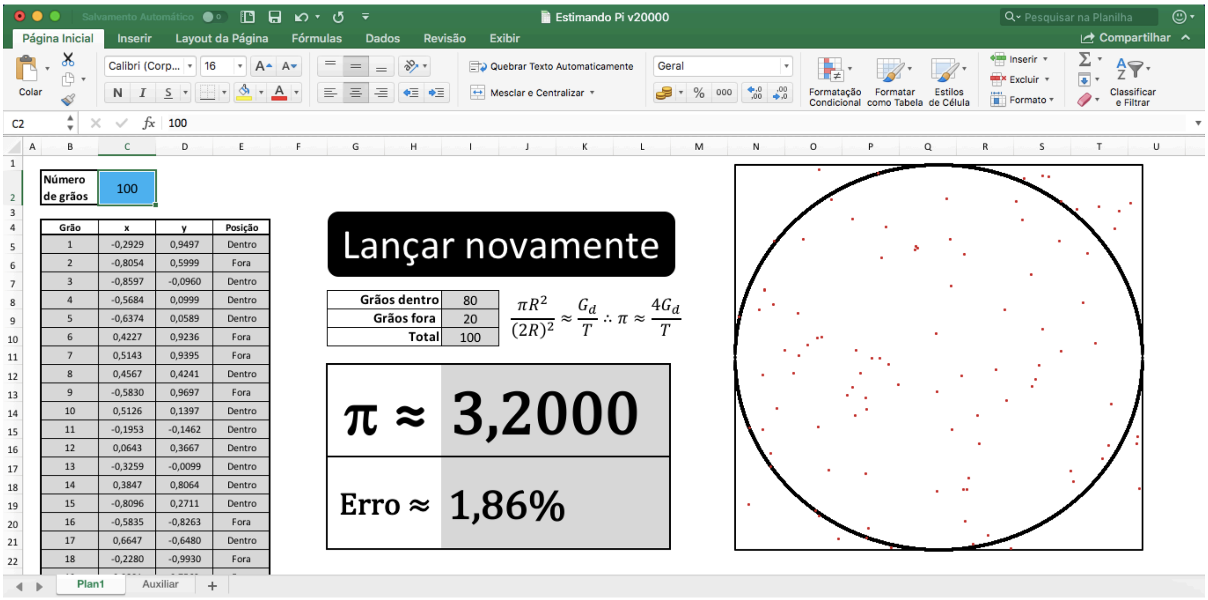


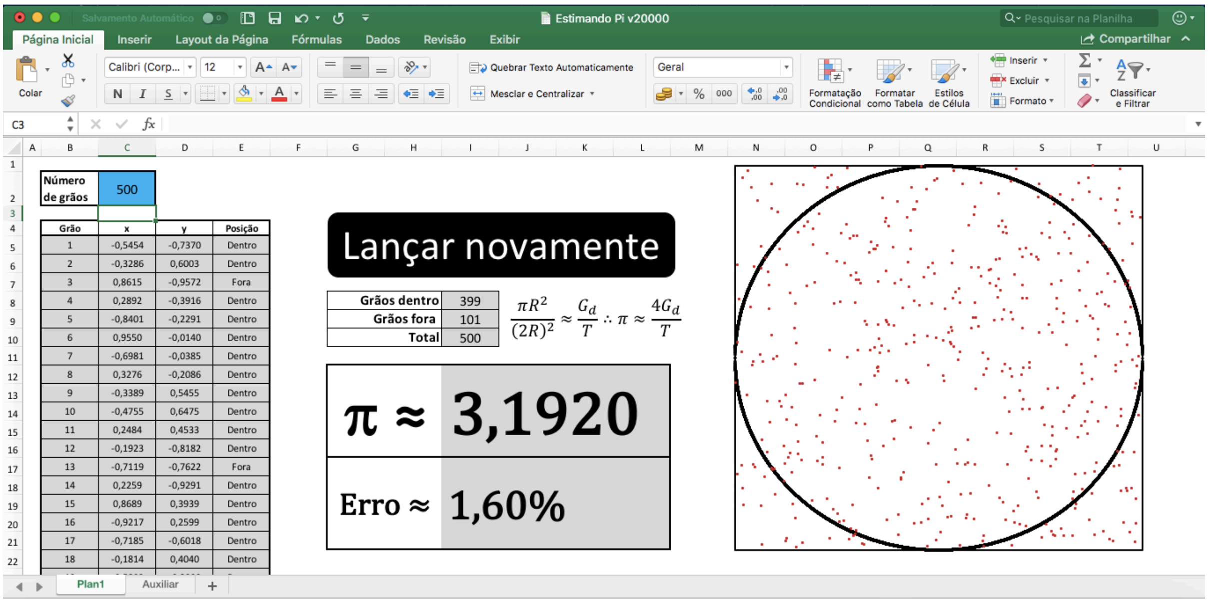
$$\frac{G_d}{T} \approx \frac{A_{círculo}}{A_{quadrado}} = \frac{\pi R^2}{(2R)^2} = \frac{\pi R^2}{4R^2} = \frac{\pi}{4}$$

A interface da planilha é a seguinte:

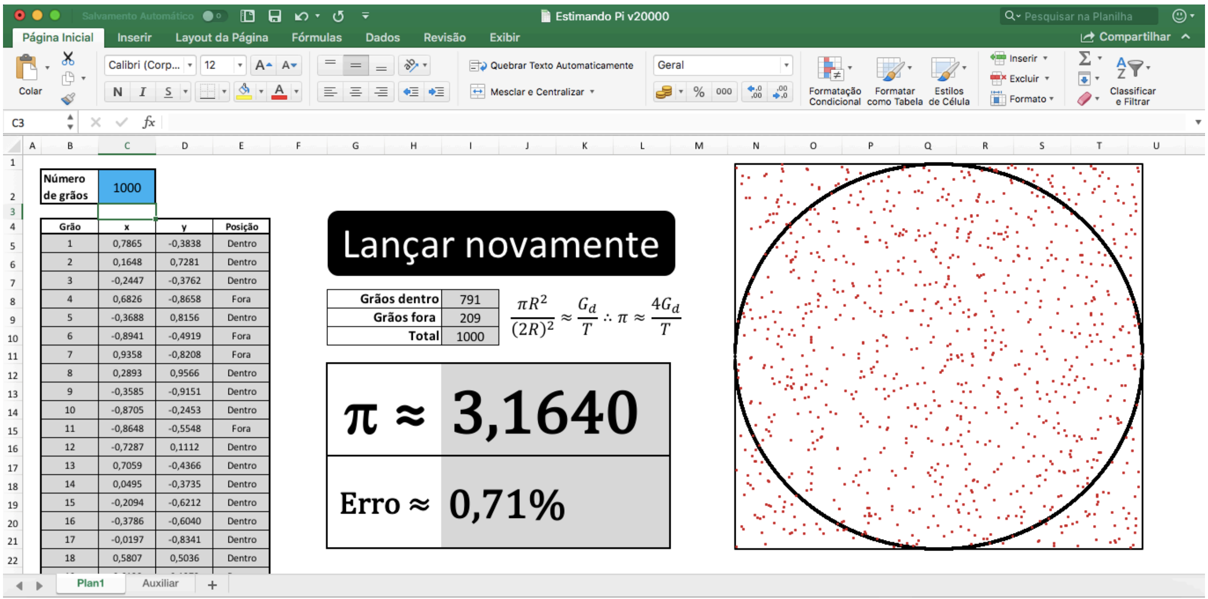


O usuário da planilha deve colocar o número de grãos desejados, e o programa simulará o lançamento dos mesmos. As imagens a seguir ilustram a simulação para 100, 500, 1000, 5000, 10000 e 20000 grãos:

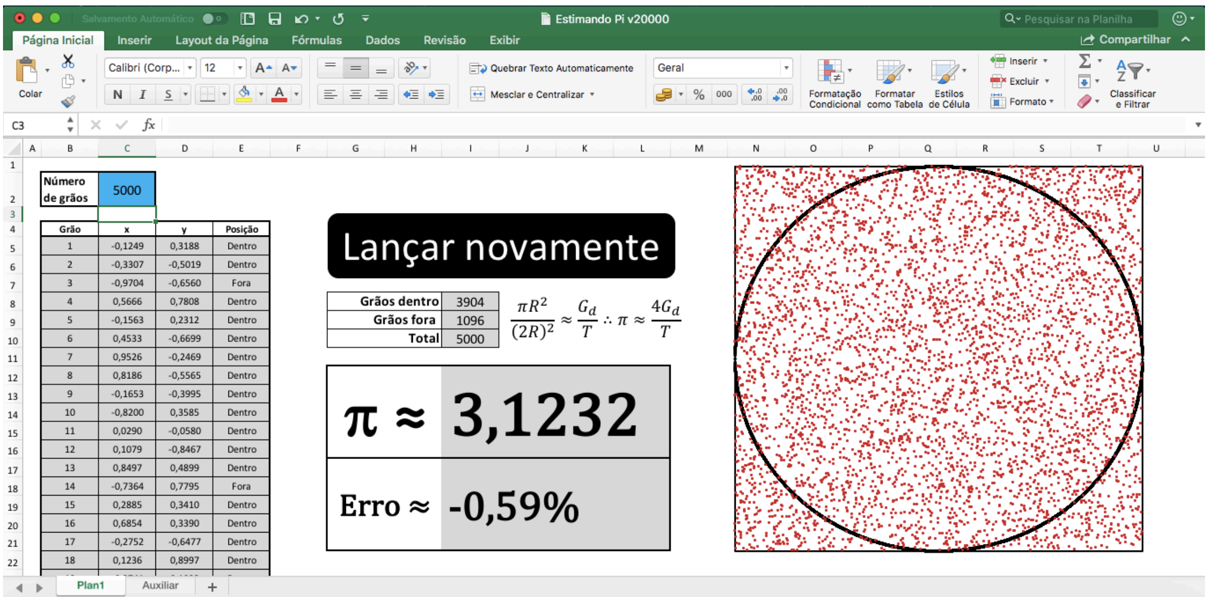




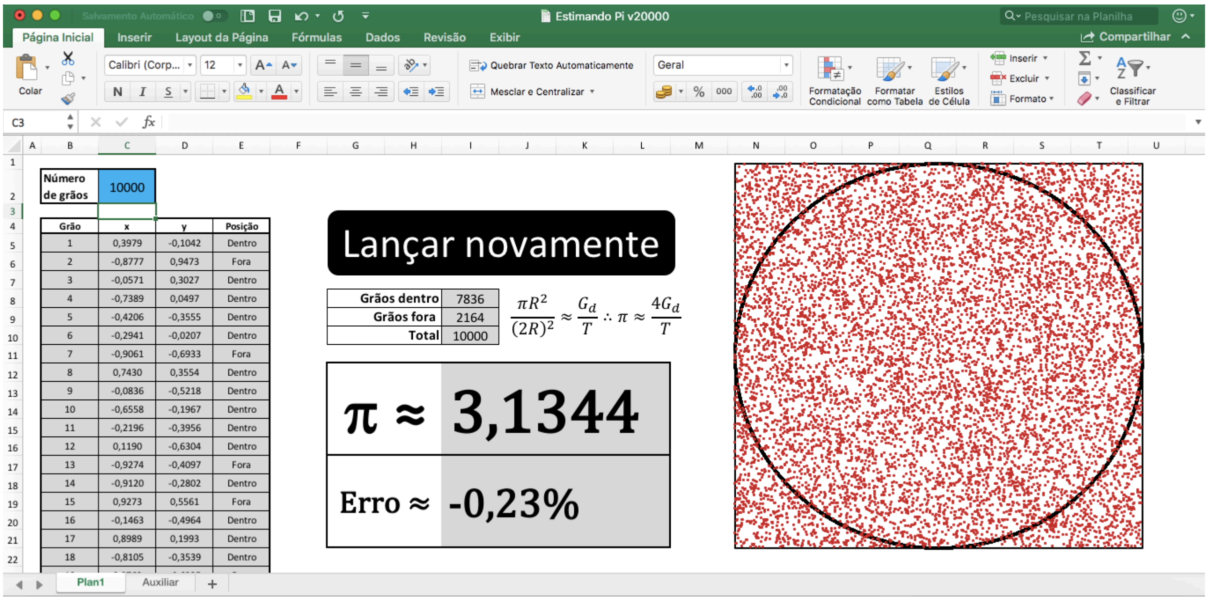
Grão	x	y	Posição
1	-0,5454	-0,7370	Dentro
2	-0,3286	0,6003	Dentro
3	0,8615	-0,9572	Fora
4	0,2892	-0,3916	Dentro
5	-0,8401	-0,2291	Dentro
6	0,9550	-0,0140	Dentro
7	-0,6981	-0,0385	Dentro
8	0,3276	-0,2086	Dentro
9	-0,3389	0,5455	Dentro
10	-0,4755	0,6475	Dentro
11	0,2484	0,4533	Dentro
12	-0,1923	-0,8182	Dentro
13	-0,7119	-0,7622	Fora
14	0,2259	-0,9291	Dentro
15	0,8689	0,3939	Dentro
16	-0,9217	0,2599	Dentro
17	-0,7185	-0,6018	Dentro
18	-0,1814	0,4040	Dentro

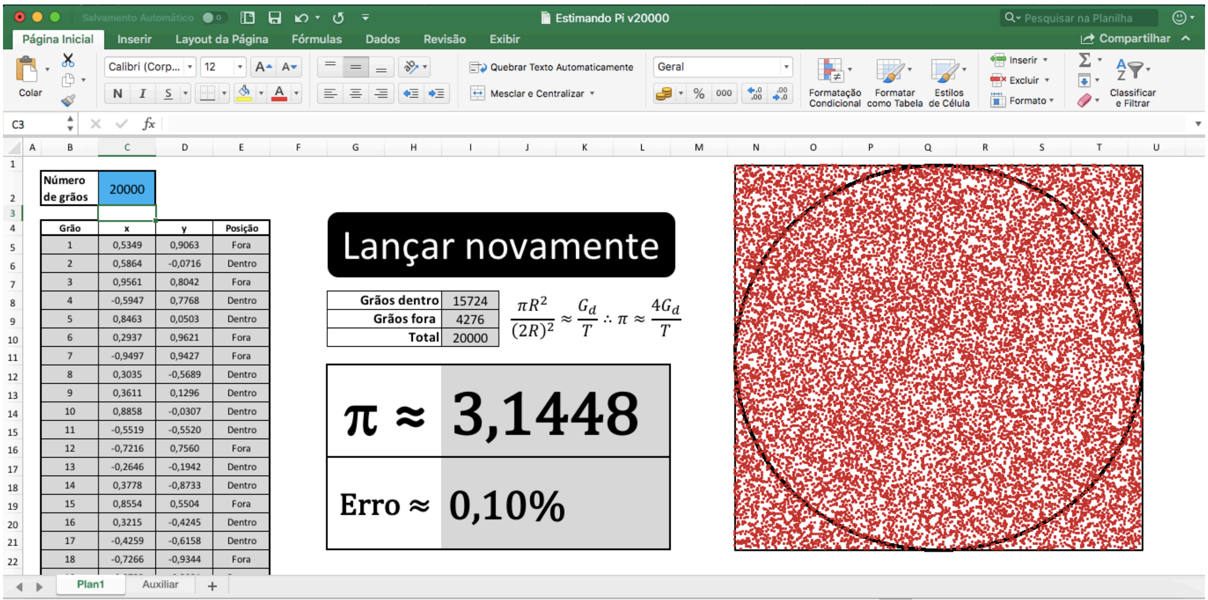


Grão	x	y	Posição
1	0,7865	-0,3838	Dentro
2	0,1648	0,7281	Dentro
3	-0,2447	-0,3762	Dentro
4	0,6826	-0,8658	Fora
5	-0,3688	0,8156	Dentro
6	-0,8941	-0,4919	Fora
7	0,9358	-0,8208	Fora
8	0,2893	0,9566	Dentro
9	-0,3585	-0,9151	Dentro
10	-0,8705	-0,2453	Dentro
11	-0,8648	-0,5548	Fora
12	-0,7287	0,1112	Dentro
13	0,7059	-0,4366	Dentro
14	0,0495	-0,3735	Dentro
15	-0,2094	-0,6212	Dentro
16	-0,3786	-0,6040	Dentro
17	-0,0197	-0,8341	Dentro
18	0,5807	0,5036	Dentro



Grão	x	y	Posição
1	-0,1249	0,3188	Dentro
2	-0,3307	-0,5019	Dentro
3	-0,9704	-0,6560	Fora
4	0,5666	0,7808	Dentro
5	-0,1563	0,2312	Dentro
6	0,4533	-0,6699	Dentro
7	0,9526	-0,2469	Dentro
8	0,8186	-0,5565	Dentro
9	-0,1653	-0,3995	Dentro
10	-0,8200	0,3585	Dentro
11	0,0290	-0,0580	Dentro
12	0,1079	-0,8467	Dentro
13	0,8497	0,4899	Dentro
14	-0,7364	0,7795	Fora
15	0,2885	0,3410	Dentro
16	0,6854	0,3390	Dentro
17	-0,2752	-0,6477	Dentro
18	0,1236	0,8997	Dentro





Observação: como a planilha gera os pontos aleatoriamente, duas simulações com o mesmo número de grãos certamente mostrarão resultados diferentes. Inclusive, é perfeitamente possível que uma simulação com um determinado número de grãos dê uma aproximação melhor do que outra com um número maior de grãos. Entretanto, ao repetir os dois experimentos várias vezes, a probabilidade de isso ocorrer novamente será baixa.

Sugestão: na sala de aula, seria interessante realizar essa experiência com material concreto, isto é, o professor ou professora pode levar uma folha grande de papel pardo, por exemplo, com um quadrado e um círculo inscrito, e pedir para os alunos lançarem os grãos de feijão aleatoriamente na região.

Atividade prática 4: Quantos peixes há no lago? (Excel)

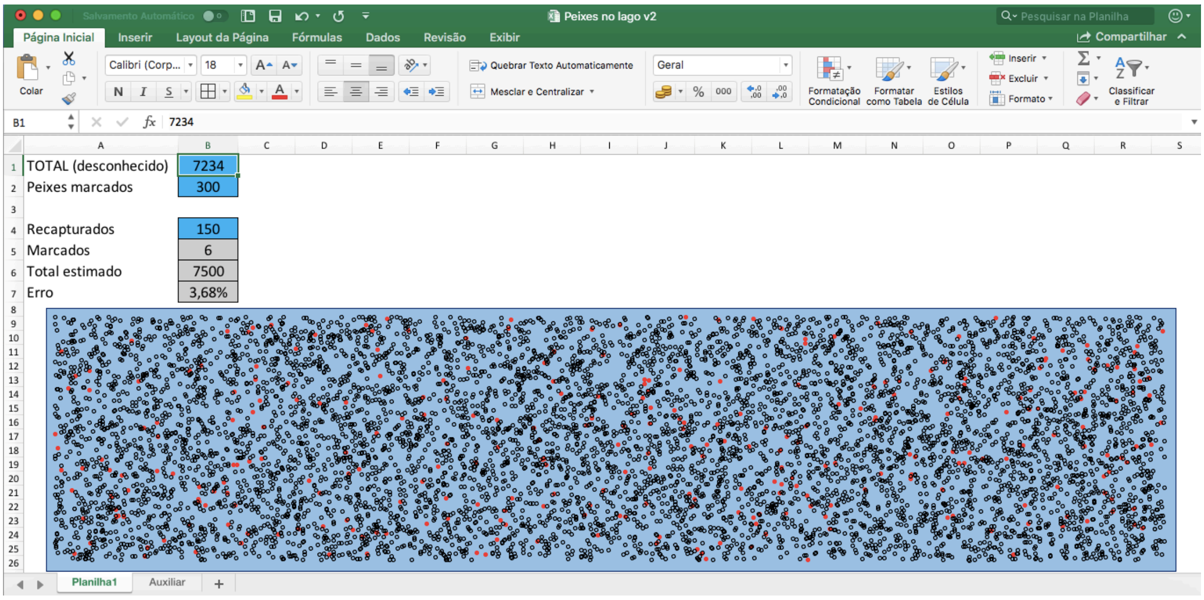
Essa atividade, originalmente proposta por CORDANI (2006), foi desenvolvida para que os alunos entendam como se pode estimar o tamanho de uma população desconhecida. Para isso, o professor ou professora colocará um número hipotético de peixes, que deverá ser estimado pelos alunos.

O programa simula o processo de "captura e recaptura": captura-se uma parcela de peixes, os quais receberão uma marca. Após isso, eles são soltos novamente no lago e, após misturarem-se, há a recaptura. Conta-se, então, quantos peixes marcados foram recapturados, e através de uma proporção, estima-se o total de peixes.

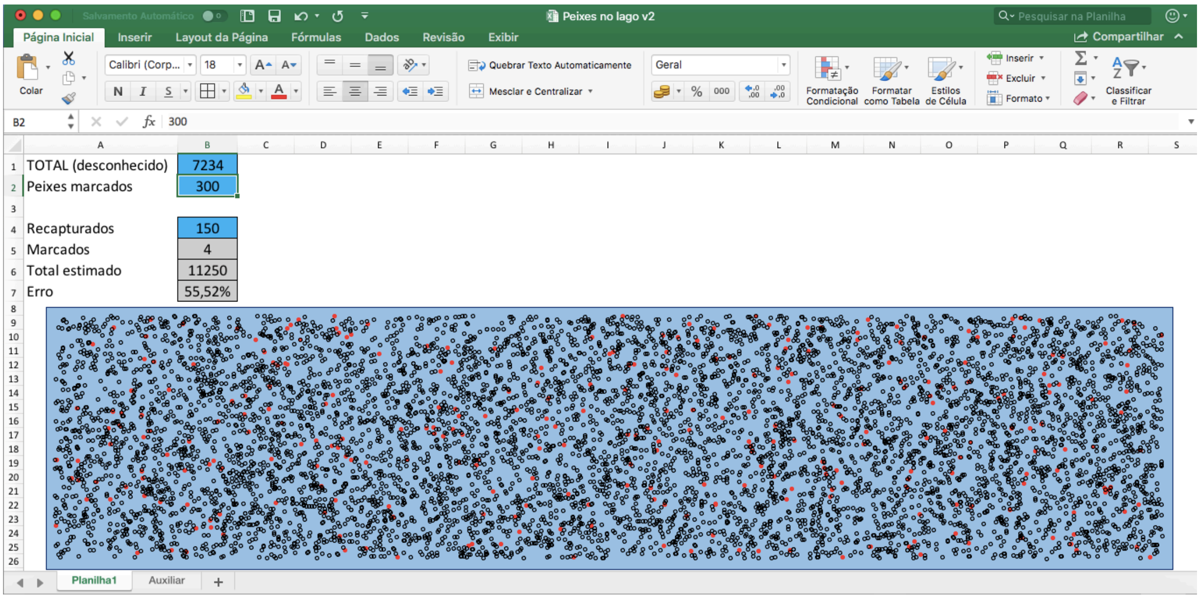
Por exemplo: o professor coloca na planilha que o lago possui 7234 peixes (deixando claro para os alunos que esse valor seria desconhecido, e que vamos estima-lo). Esse total de peixes recebe o nome N. Na etapa de captura são marcados 300 deles. Esses peixes são soltos e, na etapa de recaptura, selecionam-se 150 peixes aleatoriamente, e observa-se que 6 estavam marcados. Então:

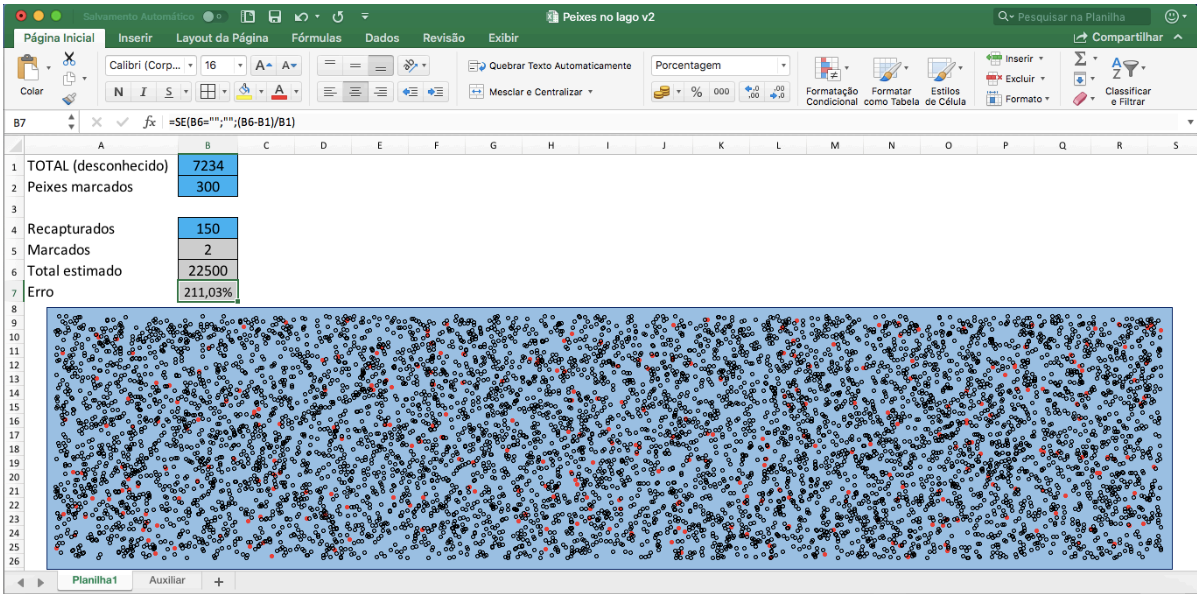
$$\frac{6}{150} = \frac{300}{N} \therefore N = 7500$$

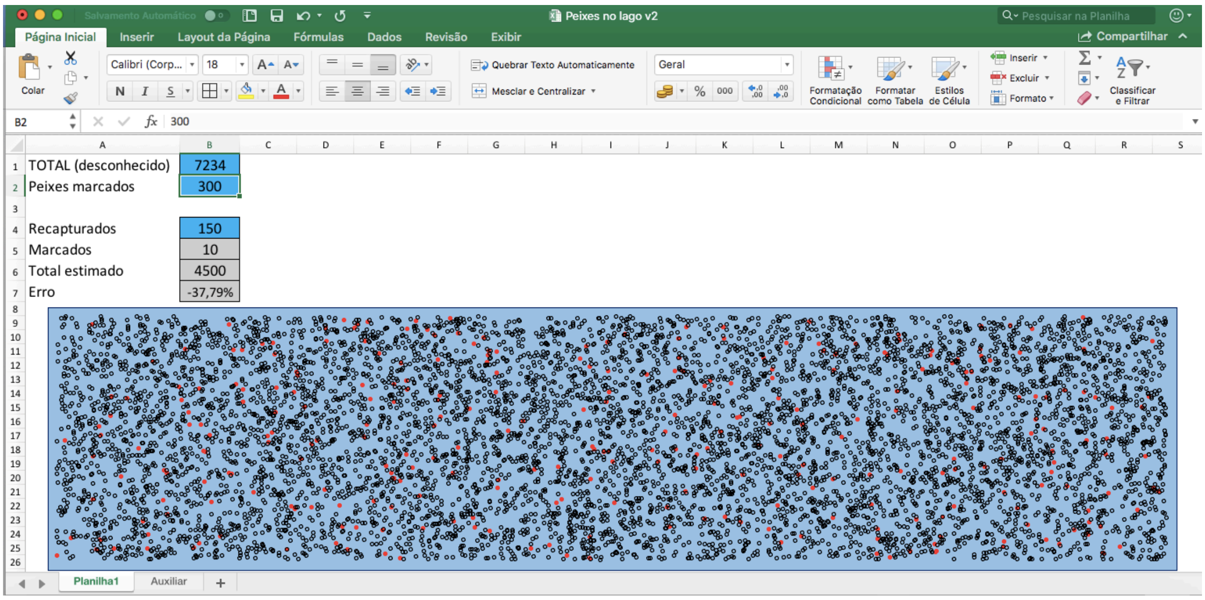
$$\text{Erro: } \frac{7500-7234}{7234} \approx 3{,}68\%$$



Refazendo a mesma simulação mais 3 vezes, obtemos:


Peixes no lago v2
B2
300
TOTAL (desconhecido) 7234
Peixes marcados 300
Recapturados 150
Marcados 4
Total estimado 11250
Erro 55,52%
Planilha1
Auxiliar


Peixes no lago v2
B7
=SE(B6="";"";(B6-B1)/B1)
TOTAL (desconhecido) 7234
Peixes marcados 300
Recapturados 150
Marcados 2
Total estimado 22500
Erro 211,03%
Planilha1
Auxiliar


Peixes no lago v2
B2
300
TOTAL (desconhecido) 7234
Peixes marcados 300
Recapturados 150
Marcados 10
Total estimado 4500
Erro -37,79%
Planilha1
Auxiliar

O melhor resultado foi o primeiro. Porém, em uma situação real não há como saber se um determinado resultado é satisfatório ou não. Então, uma possibilidade seria repetir o experimento várias vezes e tomar a média aritmética, por exemplo. Com o uso da planilha, podemos fazer isso várias vezes, e ainda variar os parâmetros. A tabela a seguir indica os valores estimados para a mesma população de peixes, quando são variadas as quantidades de peixes marcados e/ou recapturados (foram feitos 4 experimentos de cada, e então obtida a média aritmética para cada caso):

		Peixes recapturados						
		150	**300**	**450**	**600**	**750**	**900**	**1050**
Peixes marcados	**150**	13125	6514	9844	6599	8047	6530	7948
	300	9777	8555	6615	6700	8180	6258	8716
	450	7628	7646	7153	7120	7452	7324	7320
	600	10010	6923	6999	7021	7582	7403	7137
	750	7857	7622	6898	7137	7116	7687	7063
	900	5395	7063	7338	7575	7821	7719	7741
	1050	7418	7081	8005	6935	7243	7600	7739

Algumas ponderações:

- Nessa atividade, supõe-se que todos os peixes têm mesma probabilidade de serem sorteados, o que não necessariamente é verdade;
- Considera-se que a população de peixes está bem distribuída no lago, o que também não é necessariamente verdade;
- A estimação do tamanho de uma população a partir de uma amostra sempre está sujeita a erro. Porém, à medida que aumentamos o tamanho da amostra, o erro diminui. Isso pode ser constatado na tabela anterior.

Atividade prática 5: Pesquisa eleitoral (Excel)

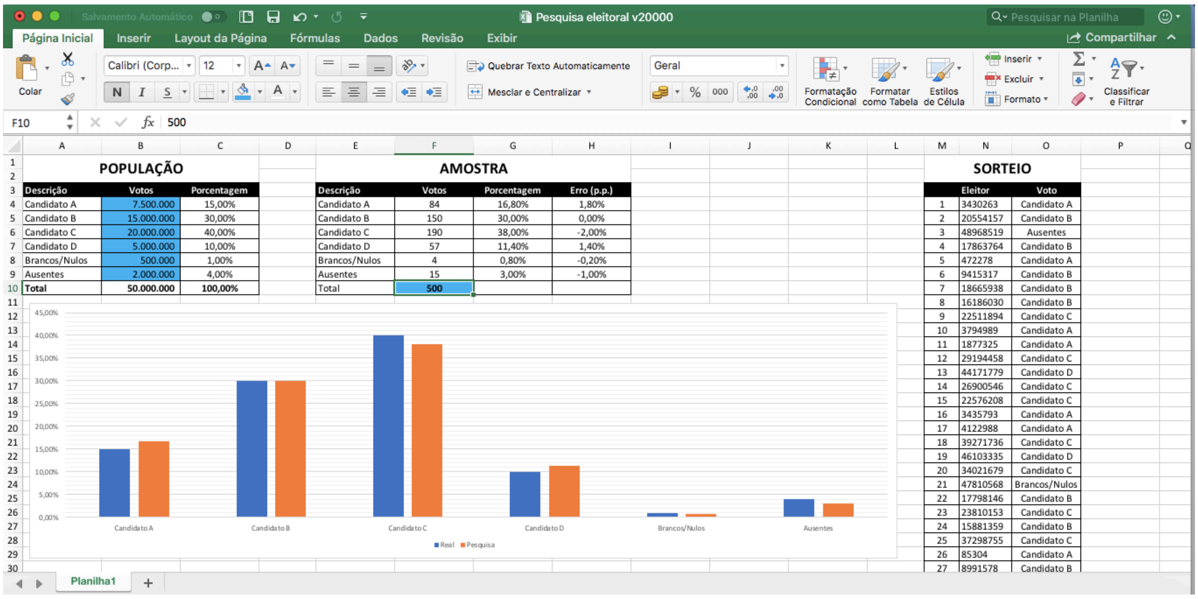
Essa atividade pode ser encarada como uma variante da atividade 4: como estimar a percentagem de eleitores que votarão em um determinado candidato?

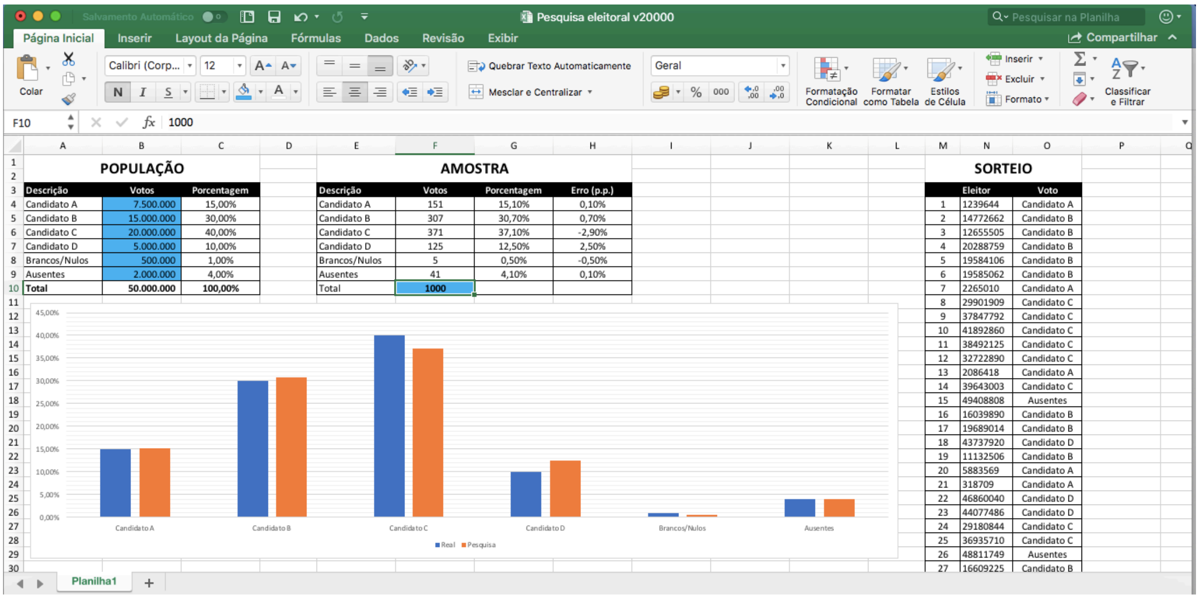
Nessa última atividade, o objetivo é mostrar aos alunos que uma pesquisa eleitoral bem feita deve retornar resultados bastante satisfatórios. Por exemplo, na última pesquisa eleitoral realizada pelo Datafolha nas eleições presidenciais de 2014, foram entrevistadas 19318 pessoas, e os resultados indicaram que a candidata Dilma Rousseff teria 52% das intenções de voto, e o candidato Aécio Neves 48%. No dia da eleição, mais de 100 milhões de eleitores compareceram, e os resultados foram 51,64% para Dilma e 48,36% para Aécio.

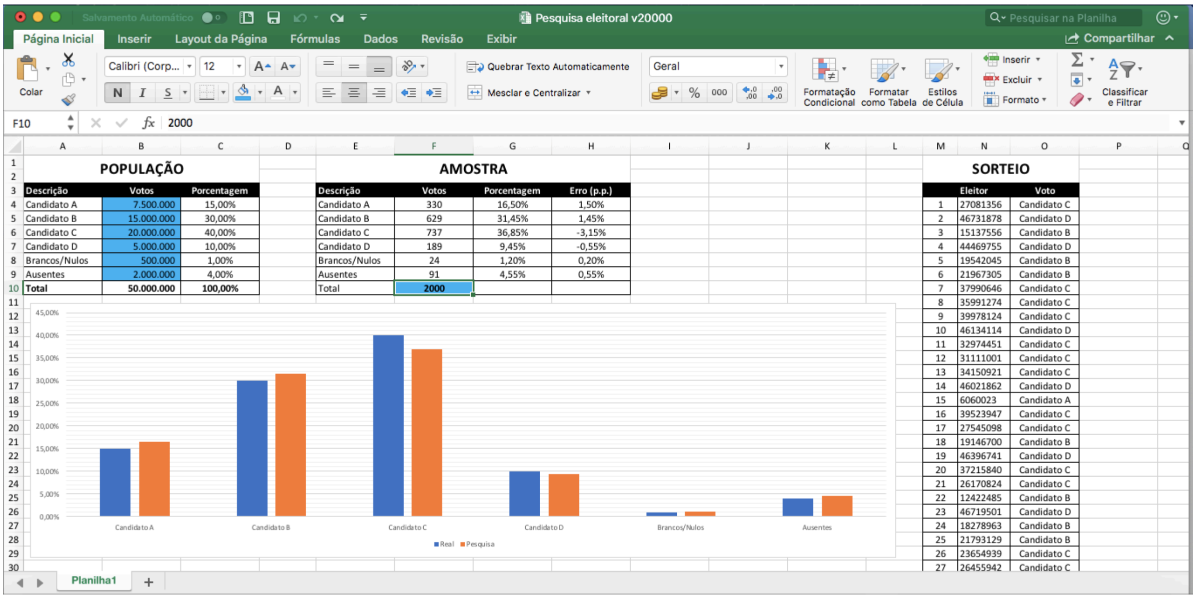
Por exemplo: suponha que as reais intenções de voto (ou não-voto) dos 50.000.000 de eleitores de um país sejam:

Candidato A	7.500.000
Candidato B	15.000.000
Candidato C	20.000.000
Candidato D	5.000.000
Brancos/Nulos	500.000
Não votarão	2.000.000

Se forem selecionadas aleatoriamente 500, 1000, 2000, 3000 ou 5000 pessoas nessa cidade, e supondo-se também que elas respondam corretamente a pesquisa, os resultados encontrados pela planilha seriam:







POPULAÇÃO

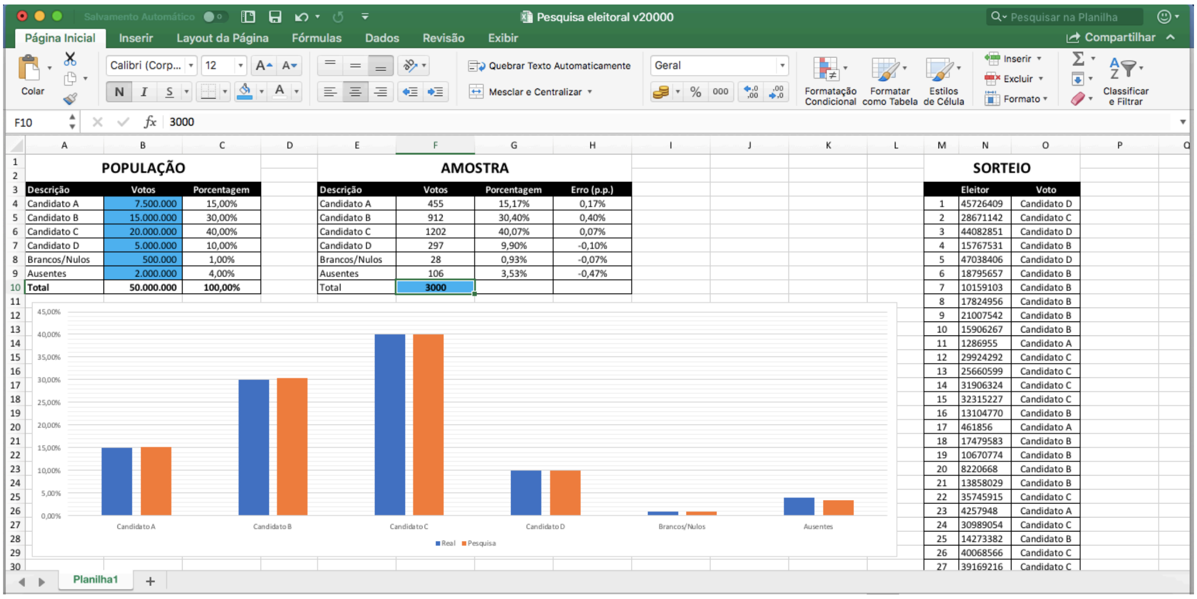
Descrição	Votos	Porcentagem
Candidato A	7.500.000	15,00%
Candidato B	15.000.000	30,00%
Candidato C	20.000.000	40,00%
Candidato D	5.000.000	10,00%
Brancos/Nulos	500.000	1,00%
Ausentes	2.000.000	4,00%
Total	50.000.000	100,00%

AMOSTRA

Descrição	Votos	Porcentagem	Erro (p.p.)
Candidato A	330	16,50%	1,50%
Candidato B	629	31,45%	1,45%
Candidato C	737	36,85%	-3,15%
Candidato D	189	9,45%	-0,55%
Brancos/Nulos	24	1,20%	0,20%
Ausentes	91	4,55%	0,55%
Total	2000		

SORTEIO

	Eleitor	Voto
1	27081356	Candidato C
2	46731878	Candidato D
3	15137556	Candidato B
4	44469755	Candidato D
5	19542045	Candidato B
6	21967305	Candidato B
7	37990646	Candidato C
8	35991274	Candidato C
9	39978124	Candidato C
10	46134114	Candidato D
11	32974451	Candidato C
12	31111001	Candidato C
13	34150921	Candidato C
14	46021862	Candidato D
15	6060023	Candidato A
16	39523947	Candidato C
17	27545098	Candidato C
18	19146700	Candidato B
19	46396741	Candidato D
20	37215840	Candidato C
21	26170824	Candidato C
22	12422485	Candidato B
23	46719501	Candidato D
24	18278963	Candidato B
25	21793129	Candidato B
26	23654939	Candidato C
27	26455942	Candidato C



POPULAÇÃO

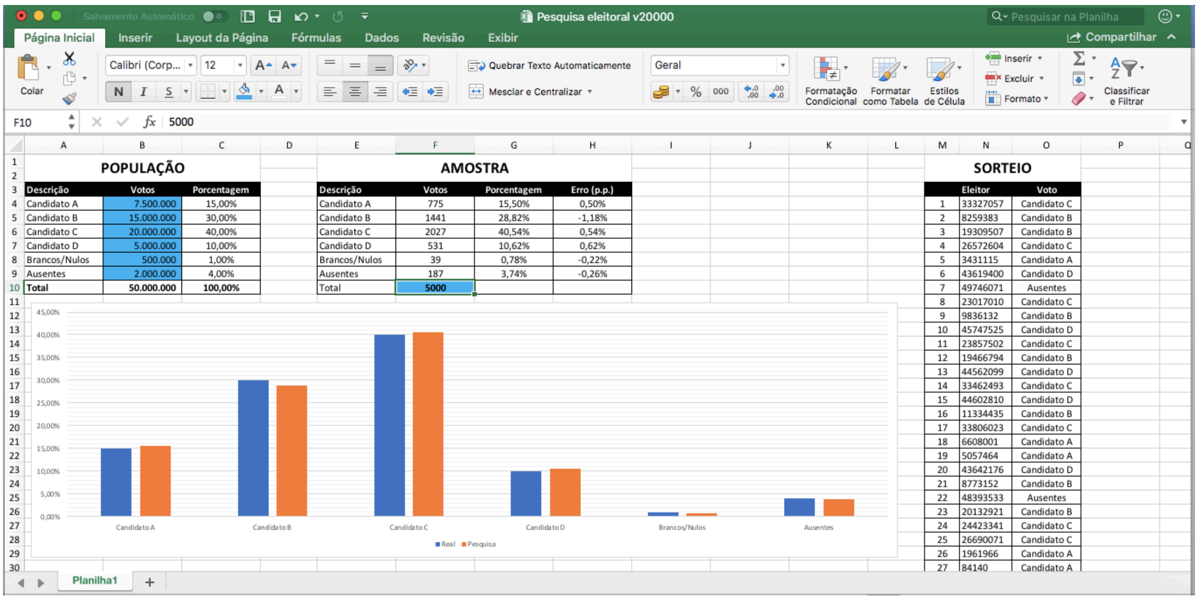
Descrição	Votos	Porcentagem
Candidato A	7.500.000	15,00%
Candidato B	15.000.000	30,00%
Candidato C	20.000.000	40,00%
Candidato D	5.000.000	10,00%
Brancos/Nulos	500.000	1,00%
Ausentes	2.000.000	4,00%
Total	50.000.000	100,00%

AMOSTRA

Descrição	Votos	Porcentagem	Erro (p.p.)
Candidato A	455	15,17%	0,17%
Candidato B	912	30,40%	0,40%
Candidato C	1202	40,07%	0,07%
Candidato D	297	9,90%	-0,10%
Brancos/Nulos	28	0,93%	-0,07%
Ausentes	106	3,53%	-0,47%
Total	3000		

SORTEIO

	Eleitor	Voto
1	45726409	Candidato D
2	28671142	Candidato C
3	44082851	Candidato D
4	15767531	Candidato B
5	47038406	Candidato D
6	18795657	Candidato B
7	10159103	Candidato B
8	17824956	Candidato B
9	21007542	Candidato B
10	15906267	Candidato B
11	1286955	Candidato A
12	29924292	Candidato C
13	25660599	Candidato C
14	31906324	Candidato C
15	32315227	Candidato C
16	13104770	Candidato B
17	461856	Candidato A
18	17479583	Candidato B
19	10670774	Candidato B
20	8220668	Candidato B
21	13858029	Candidato B
22	35745915	Candidato C
23	4257948	Candidato A
24	30989054	Candidato C
25	14273382	Candidato B
26	40068566	Candidato C
27	39169216	Candidato C



POPULAÇÃO

Descrição	Votos	Porcentagem
Candidato A	7.500.000	15,00%
Candidato B	15.000.000	30,00%
Candidato C	20.000.000	40,00%
Candidato D	5.000.000	10,00%
Brancos/Nulos	500.000	1,00%
Ausentes	2.000.000	4,00%
Total	50.000.000	100,00%

AMOSTRA

Descrição	Votos	Porcentagem	Erro (p.p.)
Candidato A	775	15,50%	0,50%
Candidato B	1441	28,82%	-1,18%
Candidato C	2027	40,54%	0,54%
Candidato D	531	10,62%	0,62%
Brancos/Nulos	39	0,78%	-0,22%
Ausentes	187	3,74%	-0,26%
Total	5000		

SORTEIO

	Eleitor	Voto
1	33327057	Candidato C
2	8259383	Candidato B
3	19309507	Candidato B
4	26572604	Candidato C
5	3431115	Candidato A
6	43619400	Candidato D
7	49746071	Ausentes
8	23017010	Candidato C
9	9836132	Candidato B
10	45747525	Candidato D
11	23857502	Candidato C
12	19466794	Candidato B
13	44562099	Candidato D
14	33462493	Candidato C
15	44602810	Candidato D
16	11334435	Candidato B
17	33806023	Candidato C
18	6608001	Candidato A
19	5057464	Candidato A
20	43642176	Candidato D
21	8773152	Candidato B
22	48393533	Ausentes
23	20132921	Candidato B
24	24423341	Candidato C
25	26690071	Candidato C
26	1961966	Candidato A
27	84140	Candidato A

Conclusão

O que se percebe é que não é dada a devida importância ao tema Estatística, embora ele esteja presente em inúmeras situações do cotidiano. Com o que foi abordado nesse trabalho, espera-se que possa ser uma ferramenta para reverter esse cenário.

Cada vez mais o uso de recursos computacionais no ensino se faz necessário. Aqui, foram abordados apenas os temas Probabilidade e Estatística. Entretanto, várias outras áreas também podem se beneficiar com o uso de planilhas eletrônicas, *softwares*, etc.

Espera-se que esse trabalho possa contribuir para o bom entendimento dos temas Probabilidade e Estatística (principalmente) por parte dos alunos.

Anexo 1: cursos da UNICAMP com disciplinas obrigatórias na área de Estatística

Fonte: Catálogo dos Cursos de Graduação 2018

Curso	Área	Nome da(s) disciplina(s)	**Temas abordados:** **Descritiva (D)** **Probabilidade (P)** **Inferência (I)**
Arquitetura	Exatas	Fundamentos de Estatística Aplicada ao Planejamento Urbano	D / P
Ciência da computação	Exatas	Introdução aos modelos probabilísticos	D / P / I
Tecnologia em Construção de Edifícios	Exatas	Estatística	D / P / I
Saneamento Ambiental	Exatas	Estatística	D / P / I
Engenharia Agrícola	Exatas	Estatística para Experimentalistas	D / P / I
Engenharia Ambiental	Exatas	Estatística	D / P / I
Engenharia Civil	Exatas	Estatística para Experimentalistas	D / P / I
Engenharia de Alimentos	Exatas	Estatística para Experimentalistas	D / P / I
Engenharia de Computação	Exatas	Introdução aos modelos probabilísticos	D / P / I
Engenharia de Controle e Automação	Exatas	Estatística para Experimentalistas	D / P / I
Engenharia de Manufatura	Exatas	Estatística e Probabilidade para Engenharia	D / P / I
Engenharia de Produção	Exatas	Controle Estatístico da Qualidade Estatística e Probabilidade para Engenharia	D / P / I
Engenharia de Telecomunicações	Exatas	Probabilidades e Teoria da Informação	D / P
Engenharia Elétrica	Exatas	Introdução aos modelos probabilísticos	D / P / I
Engenharia Física	Exatas	Física Estatística Probabilidade	P
Engenharia Mecânica	Exatas	Estatística para Experimentalistas	D / P / I
Engenharia Química	Exatas	Estatística para Experimentalistas	D / P / I
Estatística	Exatas	Várias	D / P / I
Física	Exatas	Probabilidade I	P
Geologia	Exatas	Estatística para Experimentalistas	D / P / I
Licenciatura em Física	Exatas	Estatística para Experimentalistas	D / P / I
Licenciatura em Matemática	Exatas	Estatística e Probabilidade I	D / P
Matemática	Exatas	Probabilidade I	P

Matemática Aplicada e Computacional	Exatas	Probabilidade I	P
Química	Exatas	Estatística Aplicada à Química Analítica	D
Química Tecnológica	Exatas	Estatística Aplicada à Química Analítica	D
Sistemas de Informação	Exatas	Estatística	D / P / I
Ciências Biológicas	Biológicas	Matemática Aplicada para Biologia	P
Ciências do Esporte	Biológicas	Matemática Aplicada para Biologia	P
Educação Física	Biológicas	Estatística Aplicada à Educação Física	D / I
Enfermagem	Biológicas	Introdução à Estatística Vital	D
Farmácia	Biológicas	Estatística para Experimentalistas Matemática Aplicada para Biologia	D / P / I
Licenciatura em Ciências Biológicas	Biológicas	Matemática Aplicada para Biologia	P
Nutrição	Biológicas	Estatística e Bioestatística	D / P / I
Odontologia	Biológicas	Bioestatística e Metodologia de Pesquisa	D
Administração	Humanas	Estatística para Administração	D / P / I
Administração Pública	Humanas	Estatística para Administração	D / P / I
Ciências Econômicas	Humanas	Estatística Econômica e Introdução à Econometria Estatística para Experimentalistas	D / P / I
Ciências Sociais	Humanas	Estatística para Ciências Sociais	D / P
Geografia	Humanas	Estatística Descritiva	D / P
Licenciatura Integrada Química / Física	Humanas	Matemática Elementar	P

Anexo 2: cursos da UERJ com disciplinas obrigatórias na área de Estatística

Fonte: Ementário UERJ

Curso	Área	Nome da(s) disciplina(s)	Temas abordados: Descritiva (D) Probabilidade (P) Inferência (I)
Ciência da Computação	Exatas	Matemática Discreta	P
Ciências Atuariais	Exatas	Várias	D / P / I
Ciências Contábeis	Exatas	Estatística Aplicada a Negócios	D / P / I
Engenharia Ambiental e Sanitária	Exatas	Probabilidade e Estatística III	D / P / I
Engenharia Cartográfica	Exatas	Probabilidade e Estatística III	D / P / I
Engenharia de Computação	Exatas	Probabilidade e Estatística	D / P / I
Engenharia Civil	Exatas	Probabilidade e Estatística III	D / P / I
Engenharia Elétrica	Exatas	Probabilidade e Estatística III	D / P / I
Engenharia Mecânica	Exatas	Probabilidade e Estatística III	D / P / I
Engenharia de Produção	Exatas	Probabilidade e Estatística III Estatística Aplicada na Produção	D / P / I
Engenharia Química – Rio	Exatas	Introdução à Quimiometria	D / P / I
Engenharia Química – Resende	Exatas	Probabilidade e Estatística	D / P / I
Estatística	Exatas	Várias	D / P / I
Física – Bacharelado	Exatas	Física Estatística A	D / P
Física – Licenciatura	Exatas	Mecânica Estatística Métodos Matemáticos	D / P
Geologia	Exatas	Estatística	D / P
Matemática – Bacharelado	Exatas	Estatística I	D / P / I
Matemática – Licenciatura – Rio	Exatas	Estatística I	D / P / I
Matemática – Licenciatura – Baixada Fluminense	Exatas	Probabilidade e Estatística	D / P / I
Matemática – Licenciatura – São Gonçalo	Exatas	Estatística	D / P / I
Oceanografia	Exatas	Estatística II Estatística III	D / P / I
Ciências Biológicas	Biológicas	Bioestatística	D / P
Ciências Biológicas – Licenciatura – Rio	Biológicas	Bioestatística	D / P
Ciências Biológicas – Licenciatura – CEDERJ	Biológicas	Elementos de Matemática e Estatística	D / P / I

Ciências Biológicas – Licenciatura – São Gonçalo	Biológicas	Introdução à Bioestatística	D / P / I
Educação Física – Bacharelado	Biológicas	Estatística Aplicada à Pesquisa em Educação Física	D / I
Educação Física – Licenciatura	Biológicas	Estatística Aplicada à Pesquisa em Educação Física	D / I
Enfermagem	Biológicas	Bioestatística	D / P / I
Nutrição	Biológicas	Bioestatística	D / P / I
Administração	Humanas	Estatística Aplicada a Negócios	D / P / I
Arqueologia	Humanas	Estatística Aplicada a Arqueologia	D / P
Arquitetura	Humanas	Estatística para Arquitetura	D / P
Ciências Econômicas	Humanas	Estatística Econômica e Introdução à Econometria	D / P
Ciências Sociais – Bacharelado	Humanas	Estatística Aplicada à Pesquisa em Ciências Sociais	D / P
Ciências Sociais – Licenciatura	Humanas	Estatística Aplicada à Pesquisa em Ciências Sociais	D / P
Geografia – Bacharelado	Humanas	Estatística Descritiva I	D
Geografia – Licenciatura – Rio	Humanas	Estatística Descritiva I	D
Geografia – Licenciatura – São Gonçalo	Humanas	Estatística	D
Pedagogia – Licenciatura – Baixada Fluminense	Humanas	Estatística Aplicada à Educação	D
Relações Internacionais	Humanas	Estatística Aplicada à Pesquisa em Relações Internacionais	D / P
Turismo	Humanas	Estatística Aplicada ao Turismo	D / I

Bibliografia

[1] MAGALHÃES, Marcos Nascimento; DE LIMA, Antonio Carlos Pedroso. *Noções de Probabilidade e Estatística*. 6 ed. São Paulo: Editora da Universidade de São Paulo, 2004.

[2] HOLANDA, Francisco Bruno; MUNIZ NETO, Antonio Caminha. *Introdução à Estatística.* Portal da Matemática – OBMEP. 2017.

[3] CORDANI, Lisbeth Kaiserlian. *Oficina Estatística Para Todos*. Disponível em <https://www.ime.usp.br/~abe/ce-arquivos/Oficina.pdf>. 2006. Acesso em dez. 2017.

[4] IEZZI, Gelson et al. *Matemática: volume único.* 5 ed. São Paulo: Atual, 2011.

[5] GIOVANNI, José Ruy; BONJORNO, José Roberto. *Matemática: uma nova abordagem, vol. 2*. São Paulo: FTD, 2000.

[6] DANTE, Luiz Roberto. *Matemática, volume único*. 1 ed. São Paulo: Ática, 2005.

[7] *Base Nacional Comum Curricular (versão definitiva de 20 dez. 2017).* Disponível em <http://basenacionalcomum.mec.gov.br/images/BNCC_20dez_site.pdf>. Acesso em jan. 2017.

[8] *Base Nacional Comum Curricular (proposta preliminar – segunda versão revista – abr. 2016).* Disponível em <http://historiadabncc.mec.gov.br/documentos/bncc-2versao.revista.pdf>. Acesso em jan. 2018