Ciências Contábeis
UNIVERSIDADE FEDERAL DA BAHIA

Wecsley Otero Prates

MATF09

Estatísticas para Ciências Sociais Aplicadas I

ESTATÍSTICAS PARA AS CIÊNCIAS SOCIAIS APLICADAS I

UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE CIÊNCIAS CONTÁBEIS
BACHARELADO EM CIÊNCIAS CONTÁBEIS

ESTATÍSTICAS PARA AS CIÊNCIAS SOCIAIS APLICADAS I

Professor Doutor Wecsley Otero Prates

Salvador, 2017

UNIVERSIDADE FEDERAL DA BAHIA
Reitor: João Carlos Salles Pires da Silva
Vice-Reitor: Paulo César Miguez de Oliveira

Pró-Reitoria de Ensino de Graduação
Pró-Reitor: Penildon Silva Filho

Faculdade de Ciências Contábeis
Diretor: Prof. Joséliton Silveira da Rocha

Superintendência de Educação a Distância - SEAD
Superintendente: Márcia Tereza Rebouças Rangel

Coordenação de Tecnologias Educacionais
CTE-SEAD
Haenz Gutierrez Quintana
Coordenação Administrativa
CAD-SEAD
Sofia Souza
Coordenação de Design Educacional
CDE-SEAD
Lanara Souza

UAB-UFBA

Bacharelado em Ciências Contábeis EAD
Coordenadora:
Profª Inês Teresa Lyra Gaspar da Costa

Produção do Material Didático
Coordenação de Tecnologias Educacionais
CTE-SEAD

Núcleo de Estudos de Linguagens & Tecnologias - NELT/UFBA

Direção de Criação
Prof. Haenz Gutierrez Quintana

Projeto gráfico
Haenz Gutierrez Quintana
Capa: Alessandro Faria

Equipe Design
Supervisão
Alessandro Faria
Editoração/ Ilustração
Sofia Guimarães / Orlando Dantas
Camila Leite / Mariana Netto
Antonio Felix / Aline Corujas

Equipe Audiovisual
Direção
Prof. Haenz Gutierrez Quintana

Coordenação de estúdio:
Maria Christina Souza

Câmera/ Iluminação
Maria Christina Souza

Edição:
Franklin Matos Junior
Imagens de cobertura:
Maria Christina Souza;
Franklin Matos Junior.

Animação e videografismos:
Bianca Fernandes

Edição de som e Trilha Sonora:
José Balbino
Pedro Henrique Queiroz Barreto

Sistema de Bibliotecas - UFBA

P912 Prates, Wecsley Otero.
Estatísticas para as ciências sociais aplicadas I / Wecsley Otero Prates. - Salvador: UFBA, Faculdade de Ciências Contábeis, Superintendência de Educação a Distância, 2017.
156 p.: il.

ISBN 978-85-8292-140-1

1. Estatística matemática. 2. Probabilidades. 3. Estatística – Estudo e ensino. I. Universidade Federal da Bahia. Faculdade de Ciências Contábeis. Superintendência de Educação a Distância. II. Título.

CDU – 519.2

Carta de apresentação da disciplina

Caro(a) estudante,

Este material foi elaborado com muito carinho para você. O primeiro questionamento que deve estar se perguntando é: "Por que devo estudar estatística em Ciências Sociais? Não darei a você a resposta, vou relatar o que abordarei nesse material e ao final eu espero que você tenha construído a própria resposta, ok?

O módulo trata de pontos importantes no ensino da estatística dentro do curso de Ciências Sociais. Trataremos da grande importância da Estatística Descritiva e mostraremos o quão importante é saber sumarizar, e como sumarizar, uma grande quantidade de informações. Mostrarei conteúdos relacionados à estatística, para que eles servem e aonde vamos usá-los. O material foi preparado para que um profissional das Ciências Sociais pudesse conhecer um pouco da estatística e assim utilizá-la dentro do ensino e também dentro da pesquisa. Nesta apostila, procurei mostrar somente o que é básico, trabalhando com o que é mais simples e mais usado. Em um primeiro momento podem achar que tudo é muito difícil, cheio de fórmulas e de suas deduções. Entretanto, garanto que não é! Basta que se dediquem um pouco. O primeiro passo para você entender todo o trabalho, é se familiarizar com alguns termos e perder o medo da matemática e consequentemente da Estatística. O conteúdo programático da disciplina está organizado em três unidades, que por sua vez estão subdivididas de 3 a 7 temas, conforme apresentado no sumário.

De um modo geral, a dinâmica dos estudos contemplará discussões nos fóruns temáticos no Ambiente Virtual de Aprendizagem - AVA em cada uma das semanas previstas no cronograma do curso, tendo como referência as propostas de discussão contempladas ao final de cada tema.

Bons estudos a todos!

Professor Wecsley O. Prates

Mini Currículo do Professor

O Prof.º Wecsley Otero Prates é Doutor em Estatística, título obtido em 2016 pela Universidade Federal de Minas Gerais, tem Mestrado em Estatística pela Universidade Federal de Minas Gerais e Bacharel em Estatística pela Universidade Federal da Bahia. É professor adjunto do departamento de Estatística, Coordenador das disciplinas de Estatística para os cursos de Administração e Ciências Contábeis e representante do colegiado de Estatística no curso de Ciências da Computação da Universidade Federal da Bahia. Atua principalmente nas disciplinas de Probabilidade, Processos Estocásticos para o curso de Bacharel em Estatística e nas disciplinas de Estatística para cursos de serviços da Universidade Federal da Bahia. Participante do grupo de pesquisa do Métodos Estatísticos Aplicados e Computacionais do departamento de Estatística da Universidade Federal da Bahia. Tem experiência como professor EaD e como tutor em instituições públicas e privadas.

SUMÁRIO

1. INTRODUÇÃO À ESTATÍSTICA

1.1 Um pouco da História

A Estatística pode ser entendida de duas maneiras. Uma delas corresponde aos dados oficiais do Estado, da qual deriva o nome "estatística". A outra maneira refere-se à matemática empregada para análise dos dados oficiais, ou seja, a linguagem, que também pode ser empregada na análise de outros dados, não necessariamente do Estado, como os surveys ou pesquisas de opinião, por exemplo. Veremos a seguir o processo histórico do desenvolvimento deste recurso.

A origem da palavra Estatística está associada à palavra latina *Status* (Estado). Há indícios de que 3000 anos A.C. já se faziam censos na Babilônia, China e Egito e até mesmo o 4º livro do Velho Testamento faz referência a uma instrução dada a Moisés, para que fizesse um levantamento dos homens de Israel que estivessem aptos para guerrear. Usualmente, estas informações eram utilizadas para a taxação de impostos ou para o alistamento militar.

O Imperador César Augusto, por exemplo, ordenou que se fizesse o Censo de todo o Império Romano. A palavra *Censo* é derivada da palavra *Censere*, que em Latim significa Taxar. Em 1085, Guilherme, O Conquistador, solicitou um levantamento estatístico da Inglaterra, que deveria conter informações sobre terras, proprietários, uso da terra, empregados e animais. Os resultados deste censo foram publicados em 1086 no livro intitulado "Domesday Book" e serviram de base para o cálculo de impostos.

Contudo, mesmo que a prática de coletar dados sobre colheitas, composição da população humana ou de animais, impostos, etc., fosse conhecida pelos egípcios, hebreus, caldeus e gregos, e se atribuam a Aristóteles cento e oitenta descrições de Estados, apenas no século XVII a Estatística passou a ser considerada disciplina autônoma, tendo como objetivo básico a descrição dos bens do Estado.

A palavra Estatística foi cunhada pelo acadêmico alemão Gottfried Achenwall (1719-1772), que foi um notável continuador dos estudos de Hermann Conrig (1606-1681). A escola alemã atingiu sua maturidade com A. L. Von Schlozer (1735-1809), mas sempre com ideias diferentes daquelas que fundamentaram a Estatística Moderna. Com algum exagero, pode-se dizer que o seu principal legado foi o termo *Staatenkunde*, que deu origem à designação atual.

Ilustração: Orlando Dantas

Na Enciclopédia Britânica, o verbete Statistics apareceu em 1797. Em contraposição à natureza eminentemente qualitativa da escola alemã, na Inglaterra do século XVII surgiram os aritméticos políticos, dentre os quais se destacaram John Graunt (1620-1674) e William Petty (1623-1687). Eles preocuparam-se com o estudo numérico dos fenômenos sociais e políticos, na busca de leis quantitativas que pudessem explicá-los.

O estudo consistia essencialmente de exaustivas análises de nascimentos e mortes, realizadas através das Tábuas de Mortalidade, que deram origem às atuais tábuas usadas pelas companhias de seguros. Dessa forma, a escola dos aritméticos políticos pode ser considerada o berço da Demografia.

Um de seus mais notáveis adeptos foi o pastor alemão Sussmilch (1707-1767), com o qual se pode dizer que a Estatística aparece pela primeira vez como meio indutivo de investigação. Na última metade do século XIX, os alemães Helmert (1843-1917) e Wilhelm Lexis (1837-1914), o dinamarquês Thorvald Nicolai Thiele (1838-1910) e o inglês Francis Ysidro Edgeworth (1845- 1926), obtiveram resultados extremamente valiosos para o desenvolvimento da Inferência Estatística, muitos dos quais só foram completamente compreendidos mais tarde.

Contudo, o impulso decisivo deve-se a Karl Pearson (1857-1936), William S. Gosset (1876-1937) e, em especial, a Ronald A. Fisher (1890-1962).

1.2 Conceitos Fundamentais

A Estatística pode ser encarada como uma ciência ou como um método de estudo. Duas concepções para a palavra **Estatística**:

1. no plural ***(estatísticas)***, indica qualquer coleção consistente de dados numéricos, reunidos com a finalidade de fornecer informações acerca de uma atividade qualquer. Por exemplo, as estatísticas demográficas referem-se aos dados numéricos sobre nascimentos, falecimentos, matrimônios, desquites, etc.

2. no singular (***estatística)***, indica um corpo de técnicas, ou ainda uma metodologia técnica desenvolvida para a coleta, a classificação, a apresentação, a análise e a interpretação de dados quantitativos e a utilização desses dados para a tomada de decisões.

Qualquer ciência experimental não pode prescindir das técnicas proporcionadas pela Estatística, como por exemplo, a Física, a Biologia, a Administração, a Economia, etc. Todos esses ramos de atividade profissional tem necessidade de um instrumental que se preocupa com o tratamento quantitativo dos fenômenos de massa ou coletivos, cuja mensuração e análise requerem um conjunto de observações de fenômeno ou particulares.

1.3 Aplicações de Estatística em Ciências Sociais

Recentemente, ao buscar a palavra **estatística**, percebe-se que aproximadamente nove em cada 10 citações são uma relação de aversão para com a ciência, Raniere Ramos (2015). A grande maioria vem de alunos ou profissionais de outras áreas do conhecimento, como por exemplo, as Ciências Sociais. Muitos deles se perguntam: ***eu escolhi uma área das Ciências Sociais, preciso mesmo trabalhar com números? Eu vou trabalhar com pessoas, e não com computadores, preciso de estatística? Qual a importância da estatística na minha profissão? Ela é mesmo importante?***

A estatística aplicada às ciências sociais é importante sim! Poderíamos citar uma infinidade de exemplos, como:

- Qual é o efeito de uma terapia familiar entre jovens de 10 a 18 anos que presenciaram o divórcio de seus pais? Ela é eficiente? Para quais casos ela funciona mais?
- O quanto a ansiedade do aluno interfere nas provas do vestibular? Existe diferença entre homens e mulheres? Existe diferença entre quem está prestando o vestibular pela 1ª ou 3ª vez?
- Será que a taxa de natalidade de Porto Alegre é parecida com a de Recife? Existe diferença entre o Índice de Mortalidade Infantil entre as regiões Nordeste e Sudeste?

Onde entra a **estatística** com tudo isso acontecendo? Como ela poderia ajudar? É possível entender o cenário atual e prever o futuro? De maneira simples, a **estatística** ajudaria no planejamento e coleta, organização dos dados, processamento, análise, interpretação e visualização das informações. Apenas com técnicas descritivas já seria possível entender a evolução e o comportamento da dinâmica populacional de uma região ou país determinado.

Analisar o crescimento demográfico, as taxas de imigração e emigração, taxas de natalidade e mortalidade, expectativa de vida, distribuição por sexo e faixa etária, população economicamente ativa, crescimento vegetativo, epidemias e endemias, entre outras. Considerando o cenário atual, o governo do país terá subsídios e condições para prever os impactos de uma possível mudança demográfica na estrutura etária de uma nação. Seria possível dizer, quais os problemas poderiam surgir, e como solucioná-los (Sampaio e Danelon, 2015).

A estatística moderna é uma tecnologia quantitativa para a ciência experimental e observacional que permite avaliar e estudar as incertezas e os seus efeitos no planejamento e interpretação de experiências e de observações de fenômenos da natureza e da sociedade (Matsushita, 2004).

A estatística não é um ramo da matemática onde se investigam os processos de obtenção, organização e análise de dados sobre uma determinada população. A estatística também não se limita a um conjunto de elementos numéricos relativos a um fato social, nem a números, tabelas e gráficos usados para o resumo, à organização e apresentação dos dados de uma pesquisa, embora este seja um aspecto da estatística que pode ser facilmente percebido no cotidiano.

Ilustração: Orlando Dantas

1.4 Definição de Estatística

Estatística é a ciência que se preocupa com a coleta, a organização, descrição (apresentação), análise e interpretação de dados experimentais e tem como objetivo fundamental o estudo de uma população.

Este estudo pode ser feito de duas maneiras:

- Investigando todos os elementos da população ou
- Por amostragem, ou seja, selecionando alguns elementos da população.

Ilustração: Orlando Dantas

1.5 Divisão da Estatística

A estatística é a ciência que consiste na recolha, manipulação e classificação de dados tendo em vista o conhecimento de determinado fenômeno e a possibilidade de, a partir desse conhecimento, inferir possíveis novos resultados. É objetivo da Estatística extrair informação dos dados para obter uma melhor compreensão das situações que representam.

A Estatística é dividida em dois Grupos:

Estatística Descritiva: Por conta da quantidade de dados geralmente ser tão grande, é extremamente difícil captar intuitivamente as informações que os dados contêm. É necessário, portanto, que as informações sejam reduzidas até o ponto em que se possa interpretá-las mais claramente. A estatística descritiva vai resumi-las através do uso de certas medidas-síntese, que tornem possível a interpretação de resultados. No sentido mais amplo, suas funções são:

• coleta de dados;

• organização e classificação destes dados;

• apresentação através de gráficos e tabelas;

• cálculo de coeficientes (estatísticos), que permitem descrever resumidamente os fenômenos.

Estatística Indutiva (Amostral ou Inferencial)*:* é a aquela que partindo de uma amostra, estabelece hipóteses, tirar conclusões sobre a população de origem e que formula previsões fundamentando-se na teoria das probabilidades. A estatística indutiva cuida da análise e interpretação dos dados.

O processo de generalização do método indutivo está associado a uma margem de incerteza. Isto se deve ao fato de que a conclusão que se pretende obter para o conjunto de todos os indivíduos analisados quanto a determinadas características comuns baseia-se em uma parcela do total de observações.

Ilustração: Orlando Dantas

1.6 Fases do Trabalho Estatístico

O trabalho estatístico é um método científico, que consiste das cinco etapas básicas seguintes:

- Coleta e crítica de dados
- Tratamento dos dados
- Apresentação dos dados
- Análise e interpretação dos resultados
- Conclusão

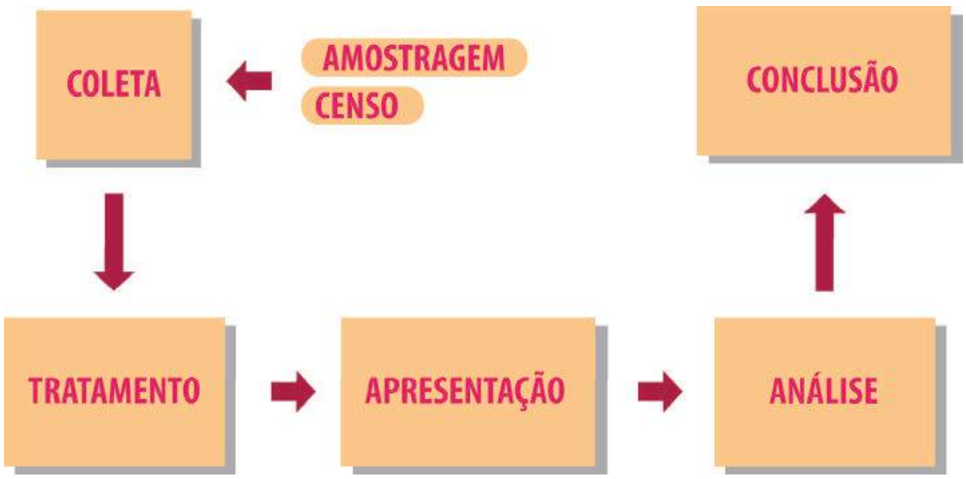


Ilustração: Orlando Dantas

Coleta e crítica dos dados

Após definirmos cuidadosamente o problema que se quer pesquisar, damos início à coleta dos dados numéricos necessários à sua descrição. A coleta pode ser direta ou indireta. A coleta é direta quando feita sobre elementos informativos de registro obrigatório (nascimentos, casamentos e óbitos, importação e exportação de mercadorias), elementos pertinentes aos prontuários dos alunos de uma escola ou, ainda, quando os dados são coletados pelo próprio pesquisador através de inquéritos e questionários.

A coleta direta de dados pode ser classificada relativamente ao fator tempo em:

a) Contínua – quando feita continuamente, tal como a de nascimentos e óbitos e a de frequência dos alunos às aulas.

b) Periódica – quando feita em intervalos constantes de tempo, como os censos e as avaliações mensais dos alunos.

c) Ocasional – Quando feita extemporaneamente, a fim de atender a uma conjuntura ou a uma emergência, como no caso de epidemias que assolam ou dizimam rebanhos inteiros.

A coleta se diz indireta quando é inferida de elementos conhecidos (coleta direta) e/ou do conhecimento de outros fenômenos relacionados com o fenômeno estudado. Como por exemplo, podemos citar a pesquisa sobre a mortalidade infantil, que é feita através de dados colhidos por uma coleta direta.

Obtidos os dados, eles devem ser cuidadosamente criticados, à procura de possíveis falhas e imperfeições, a fim de não incorrermos em erros grosseiros ou certo vulto, que possam influir sensivelmente nos resultados.

A crítica é externa quando visa às causas dos erros por parte do informante, por distração ou má interpretação das perguntas que lhe foram feitas; é interna quando visa observar os elementos originais dos dados da coleta.

Ilustração: Orlando Dantas

Tratamento dos dados

Nada mais é do que a soma e o processamento dos dados obtidos e a disposição mediante critérios de classificação Pode ser manual ou eletrônica.

Apresentação dos dados

Por mais diversa que seja a finalidade em vista, os dados devem ser apresentados sob forma adequada – tabelas e gráficos – tornando mais fácil o exame daquilo que está sendo objeto de tratamento estatístico.

Disponível em <http://www.freepik.com/free-photo/business-meeting_1202441.htm#term=statistics&page=1&position=27>. Acesso em: 01/09/2017

Análise dos resultados

Após a apresentação dos dados devemos calcular as medidas típicas convenientes para fazermos uma análise dos resultados obtidos, através dos métodos da Estatística Indutiva ou Inferencial, e tirarmos desses resultados conclusões e previsões.

Conclusão

É de responsabilidade de um especialista no assunto que está sendo pesquisado, que não é necessariamente um estatístico, relatar as conclusões de maneira que sejam facilmente entendidas por quem as for usar na tomada de decisões.

Ilustração: Orlando Dantas

2. ESTATÍSTICA DESCRITIVA

Por onde quer que se olhe ou escute uma coleção de números são normalmente enunciados como estatísticas. Estes números referem-se aos mais diversos campos de atividades: esportes, economia, finanças, etc. Assim tem-se por exemplo:

- O número de carros vendidos no país aumentou em 30%.
- A taxa de desemprego atinge, hoje, 13,1%.
- As ações da Telebrás caíram R$ 1,5, hoje.

2.1 População e Amostra

Definição 2.1.1: *O conjunto, finito ou infinito, de indivíduos ou objetos que apresentam em comum determinadas características definidas, cujo comportamento interessa ao pesquisador é denominado população ou universo.*

A população é estudada em termos de observações de características nos indivíduos (animados ou inanimados) que sejam relevantes para o estudo, e não em termos de pessoas ou objetos em si. O objetivo é tirar conclusões sobre o fenômeno em estudo, a partir dos dados observados. Pode-se definir uma população como sendo, uma coleção de todos os possíveis elementos, objetos ou medidas de interesse.

Assim, são exemplos de populações:

- o conjunto das rendas de todos os habitantes de Salvador;
- o conjunto de todas as notas dos alunos de Ciências Sociais;
- o conjunto das alturas de todos os alunos da Universidade Federal da Bahia; etc.

Ilustração: Orlando Dantas

Como em qualquer estudo estatístico temos em mente estudar uma ou mais características dos elementos de uma população, é importante definir bem essas características de interesse para que sejam delimitados os elementos que pertencem à população e quais os que não pertencem.

p.ex Exemplos

- Estudar os filhos tidos, tipo de moradia, condições de trabalho, tipo de sanitário. Números de quartos para dormir, estado civil, uso da terra, tempo de trabalho, local de nascimento, tipo de cultivo, etc., dos agricultores do Estado do Amazonas.

População: Todos os agricultores (proprietários de terra ou não) plantadores das culturas existentes no Estado do Amazonas.

- Estudar a precipitação pluviométrica anual (em mm) na cidade de Manaus.

População: Conjunto das informações coletadas pela Estação Pluviométrica, durante o ano.

Um levantamento efetuado sobre toda uma população é dito de levantamento censitário ou simplesmente censo.

Fazer levantamentos, estudos, pesquisas, sobre toda uma população (censo) é, em geral, muito difícil. Isto se deve à vários fatores. O principal é o custo. Um censo custa muito caro e demanda um tempo considerável para ser realizado. Assim, normalmente, se trabalha com partes da população denominadas de **amostras**.

Definição 2.1.2: É uma parte (um subconjunto finito) representativa de uma população selecionada segundo métodos adequados.

O objetivo é fazer inferências, tirar conclusões sobre populações com base nos resultados da amostra, para isso é necessário garantir que amostra seja representativa, ou seja, a amostra deve conter as mesmas características básicas da população, no que diz respeito ao fenômeno que desejamos pesquisar. As razões de se recorrer a amostras são: menor custo e tempo para levantar dados; melhor investigação dos elementos observados.

Ilustração: Orlando Danta

p.ex Exemplo

Uma pesquisa de opinião para saber o resultado das eleições para o governo do estado da Bahia em 2016, a população considerada foram todos os eleitores do estado e para constituir a amostra o IBOPE coletou a opinião de cerca de 1600 eleitores.

Sabendo um pouco mais

Link para auxiliar nas definições sobre População e Amostra:

https://www.youtube.com/watch?v=NvqHnM3TucE

O termo indução é um processo de raciocínio em que, partindo-se do conhecimento de uma parte, procura-se tirar conclusões sobre a realidade no todo.

Ao induzir estamos sujeitos a erros. Entretanto, a Estatística Indutiva, que obtém resultados sobre populações a partir das amostras, diz qual a precisão dos resultados e com que probabilidade se pode confiar nas conclusões obtidas.

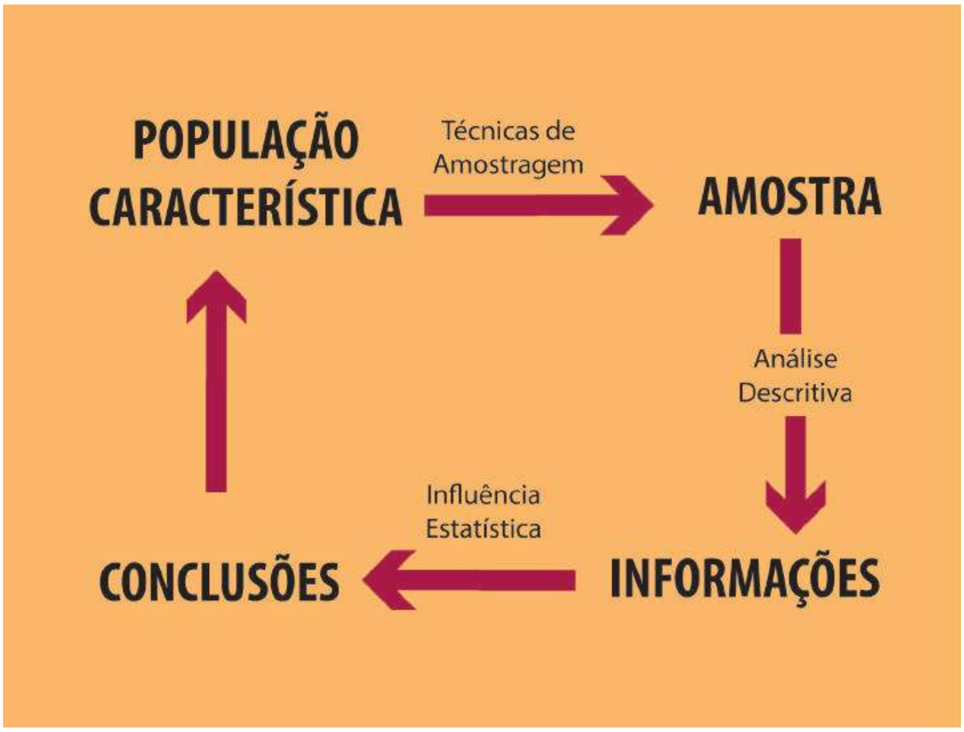


Ilustração: Sofia Guimarães

ATIVIDADE RESOLVIDA

1 - População ou universo é:

- Um conjunto de pessoas;
- Um conjunto de elementos quaisquer;
- Um conjunto de pessoas com uma característica comum;
- Um conjunto de elementos com pelo menos uma característica em comum;
- Um conjunto de indivíduo de um mesmo município, estado ou país.

2 - Uma parte da população retirada para analisá-la denomina-se:

- Universo;
- Parte;
- Pedaço;
- Dados Brutos;
- Amostra.

3 - A parte da estatística que se preocupa somente com a descrição de determinadas características de um grupo, sem tirar conclusões sobre um grupo maior denomina-se:

- Estatística de População;
- Estatística de Amostra;
- Estatística Inferencial
- Estatística Descritiva;
- Estatística Grupal.

ATIVIDADE

1 - Nos exercícios abaixo, identificar a população e a amostra.

I - Pretende-se estudar o efeito de um novo medicamento para curar determinada doença. É selecionado um grupo de 20 doentes, administrando-se o novo medicamento a 10 destes doentes escolhidos ao acaso e o medicamento habitual aos restantes.

População:

Amostra:

II - Uma pesquisa realizada com 18 mil cidadãos brasileiros com idade entre 15 e 20 anos mostrou que 2,5 mil têm acesso a computador.

População:

Amostra:

III - Pesquisa realizada com 500 famílias das classes A, B, C no Rio de Janeiro e em São Paulo mostra que 78% das crianças usam celulares.

População:

Amostra:

IV - Uma pesquisa sobre quais marcas o brasileiro considera as melhores em várias categorias de produtos. O resultado é fruto de 2 mil entrevistas feitas com a população acima de 16 anos de todas as classes sociais do país.

População:

Amostra:

2.2 Tipos de Variáveis

Os dados estatísticos se obtêm mediante um processo que envolve a observação ou outra mensuração de características de uma população ou amostra tais como renda anual numa comunidade, sexo dos indivíduos de uma tribo indígena, percentagem de açúcar em cereais, etc. Cada uma dessas características é chamada de variável, porque originam valores que tendem a exibir certo grau de variabilidade quando se fazem mensurações sucessivas.

Exemplo: Suponha que um questionário foi aplicado aos alunos do 3º período do curso de Serviços Sociais da UFBA fornecendo as seguintes informações

- Identificação do aluno;
- Turma em que o aluno foi alocado (A ou B);
- Sexo: F se feminino, M se masculino;
- Idade: Idade em anos;
- Altura: Altura em metros;
- Peso: Peso em quilogramas;
- Filhos: Número de filhos na família.

Claramente tais variáveis têm naturezas diferentes no que tange aos possíveis valores que podem assumir. Tal fato deve ser levado em conta nas análises dos dados, pois para cada tipo de variável existe um tratamento diferente. Vamos considerar dois grandes tipos de variáveis:

Definição 2.2.1: Quantitativas (numéricas): São as variáveis cujos v*alores são expressos em números.*

Elas podem ser subdivididas em quantitativas discretas e quantitativas contínuas. As variáveis discretas podem ser vistas como resultantes de contagens, assumindo assim, valores inteiros. Já as variáveis contínuas geralmente provêm de uma mensuração e podem assumir qualquer valor em intervalos dos números reais.

São variáveis quantitativas discretas

Número de irmãos, de alunos numa sala de aula, de defeitos num carro novo, etc.

São variáveis quantitativas contínuas:

Altura, peso, comprimento, espessura, velocidade, etc.

Definição 2.2.2: Qualitativas (não numéricas): São as variáveis cujos possíveis valores que assumem representam atributos e/ou quantidades.

Se tais variáveis têm uma ordenação natural, indicando intensidades crescentes de realização, então elas serão classificadas como qualitativas ordinais ou por postos. Caso contrário, quando não é possível estabelecer uma ordem natural entre seus valores definindo apenas uma categoria, elas são classificadas como qualitativas nominais.

São variáveis qualitativas nominais:

Turma (A ou B), sexo (F ou M), cor dos olhos, campo de estudo, etc.

São variáveis qualitativas ordinais:

Tamanho (pequeno, médio ou grande), Classe social (baixa, média ou alta), etc.

Podemos resumir a classificação das variáveis no seguinte esquema:

Ilustração: Sofia Guimarães

ATIVIDADE

Leia e responda as perguntas:

I - Uma empresa de prestação de serviços que tem 5.000 clientes cadastrados realizou uma pesquisa com 500 desses. A investigação procurou determinar: o sexo dos clientes; a idade dos clientes; o grau de satisfação dos clientes (insatisfeitos, satisfeitos e muito satisfeitos); a renda anual dos clientes; o número de automóveis dos clientes.

Agora responda:

a) o universo da pesquisa é de .... clientes.

b) a amostra da pesquisa é de .... clientes.

c) o sexo dos clientes representa uma variável

d) a idade dos clientes representa uma variável

e) o grau de satisfação dos clientes representa uma variável

f) a renda anual dos clientes representa uma variável

g) o número de automóveis dos clientes representa uma variável

II - Luciano é dono de uma loja de automóveis. Para ampliar a qualidade da loja Luciano resolveu pesquisar o perfil dos clientes em relação à renda mensal, ao modelo de automóvel preferido, ao número de automóveis que cada cliente possui e à qualidade dos serviços prestados. Dos 3000 clientes cadastrados nessa loja, 1600 foram entrevistados. Pergunta-se:

a) Quantas pessoas formam o universo da pesquisa?

b) A amostra pesquisada foi de quantas pessoas?

c) Determine as variáveis pesquisadas e classifique-as como qualitativa ou quantitativa (discreta ou contínua):

III - Para pesquisar o esporte preferido dos alunos de uma escola com 1500 alunos, foram selecionados, de modo imparcial, 900 alunos. Com base nessas informações, responda:

a) Quantas pessoas tem a população da pesquisa?

b) A amostra desta pesquisa é formada de quantas pessoas?

c) Qual variável foi estudada nesta pesquisa?

IV – Uma concessionária de automóveis tem cadastrados 3.500 clientes e fez uma pesquisa sobre a preferência de compra em relação à "cor" (branca, vermelha ou azul); "preço"; "número de portas" (duas ou quatro) e "estado de conservação" (novo ou usado). Foram consultados 210 clientes. Diante dessas informações, responda:

a) Qual é o universo estatístico e qual é a amostra desta pesquisa?

b) Quais são as variáveis e qual é o tipo de cada uma?

c) Quais os possíveis valores da variável "cor" nessa pesquisa?

V – Uma agência de turismo tem 2500 clientes cadastrados. Para melhor atendê-los, foi pesquisada a preferência em relação ao tempo de duração, ao preço, ao número de acompanhantes, ao número de passeios e à qualidade dos serviços prestados em uma viagem. Foram consultados, de modo imparcial, 700 pessoas.

a) Quantas pessoas tem a população envolvida na pesquisa?

b) A amostra pesquisada foi de quantas pessoas?

c) Quais foram as variáveis qualitativas pesquisadas?

d) Quais foram as variáveis quantitativas (discretas e contínuas)?

2.3 Coleta dos Dados

Refere-se a obtenção, reunião e registro sistemático de dados, com o objetivo determinado. A escolha da fonte de obtenção dos dados está diretamente relacionada ao tipo do problema, objetivos do trabalho, escala de atuação e disponibilidade de tempo e recursos.

Disponível em < https://static.pexels.com/photos/164686/pexels-photo-164686.jpeg >. Acesso em: 01/09/2017

Definição 2.3.1: Fontes Primárias: *é o levantamento direto no campo através de mensurações diretas ou de entrevistas ou questionários aplicados a sujeitos de interesse para a pesquisa.*

Vantagens: grau de detalhamento com respeito ao interesse dos quesitos levantados; maior precisão das informações obtidas.

Definição 2.3.2: Fontes Secundárias: *Quando são publicados ou registrados por outra organização.*

A coleta de dados secundários se realiza através de documentos cartográficos (mapas, cartas, imagens e fotografias obtidas por sensor remoto ou por fotogrametria e imagens de radar). Estas fontes de informação são de extrema importância.

Das fotografias aéreas em escalas reduzidas ou mais detalhadas, das imagens de radares ou satélite e de cartas obtêm-se informações quanto ao uso do solo, drenagem, estruturas viárias e urbanas, povoamento rural, recursos florestais, minerais e pedológicos, estrutura fundiária e de serviços, dados altimétricos, etc.

Vantagens: inclui um processo de redução e agregação de informações.

2.4 Série Estatística

Definição 2.4.1: *Denomina-se Série Estatística toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da Época, do Local, ou da Espécie (fenômeno).*

Numa série estatística observa-se a existência de três elementos ou fatores: o Tempo, o Espaço e a Espécie. Conforme varie um desses elementos, a série estatística classifica-se em Temporal, Geográfica e Específica.

Definição 2.4.2: Série Temporal (histórica ou evolutiva)*: Identifica-se pelo caráter variável do fator cronológico. O local e a espécie (fenômeno) são elementos fixos. Esta série também é chamada de cronológica.*

Tabela 2.4.2: Produção Brasileira de Trigo, 1988 - 1993

ANOS	QUANTIDADE (1000 ton)
1988[1]	2345
1989	2451
1990	2501
1991	2204
1992	2306
1993	2560

Fonte: IBGE
Nota: Produção voltada para o consumo interno.
(1) Parte da produção exportada.

Ilustração: Sofia Guimarães

- Elemento variável: tempo (fator cronológico)
- Elemento fixo: local (fator geográfico) e o fenômeno (espécie)

Definição 2.4.3: Série Geográfica*: Apresenta como elemento variável o fator geográfico. A época e o fato (espécie) são elementos fixos. Também é chamada de espacial, territorial ou de localização.*

Definição 2.4.4: Série Específica*: O caráter variável é apenas o fato ou espécie. Também é chamada de série categórica.*

Exemplos típicos de séries categóricas são: produção industrial por gênero de indústria; produção agrícola por tipo de lavoura; público dos estádios por séries do campeonato brasileiro. Nas séries específicas, os registros das ocorrências não levam em conta variações de espaço e tempo.

Tabela 2.4.4: Áreas das Regiões Fisiográficas - Brasil - 1966

REGIÕES FISIOGRÁFICAS	ÁREA (KM²)
NORTE	3.581.180
NORDESTE	965.652
SUDESTE	1.260.057
SUL	825.621
CENTRO-OESTE	1.879.965
BRASIL	8.511.965

Fonte: IBGE

Ilustração: Sofia Guimarães

Definição 2.4.5: Série Conjugada: *É à apresentação de duas ou mais séries de maneira conjugada, havendo duas ordens de classificação: uma horizontal e outra vertical.*

O exemplo abaixo é de uma série geográfica-temporal.

Tabela 2.4.5: Quantidade de carros vendidos da ABC Veículos LTDA, no 1º bimestre de 2011 e por cidade da filial.

FILIAL	MÊS		TOTAL
	JANEIRO	FEVEREIRO	
SÃO PAULO	10.000	3.000	13.000
RIO DE JANEIRO	12.000	5.000	17.000
TOTAL	22.000	8.000	30.000

Fonte: Dados fictícios, Março/2011

Ilustração: Sofia Guimarães

2.5 Apresentação dos Dados

É um processo de apuração ou sumarização que consiste em resumir os dados através de sua contagem ou agrupamento. É um trabalho de condensação e de tabulação dos dados, que chegam ao analista de forma desorganizada. Por meio da apuração, se tem a oportunidade de condensar os dados de modo a obter um conjunto compacto de números, o qual possibilita distinguir melhor o comportamento do fenômeno na sua totalidade. Os dados de fenômenos geográficos podem ser organizados em mapas, tabelas, matrizes, disquetes ou fitas. Há duas formas de apresentação que não se excluem mutuamente:

2.5.1 Apresentação Tabular

É uma apresentação numérica dos dados. Consiste em dispor os dados em linhas e colunas distribuídos de modo ordenado, segundo algumas regras práticas adotadas pelo Conselho Nacional de Estatística. As tabelas têm a vantagem de conseguir expor, sistematicamente em um só local, os resultados sobre determinado assunto, de modo a se obter uma visão global mais rápida daquilo que se pretende analisar.

Uma tabela possui elementos essenciais e complementares. Os elementos essenciais são:

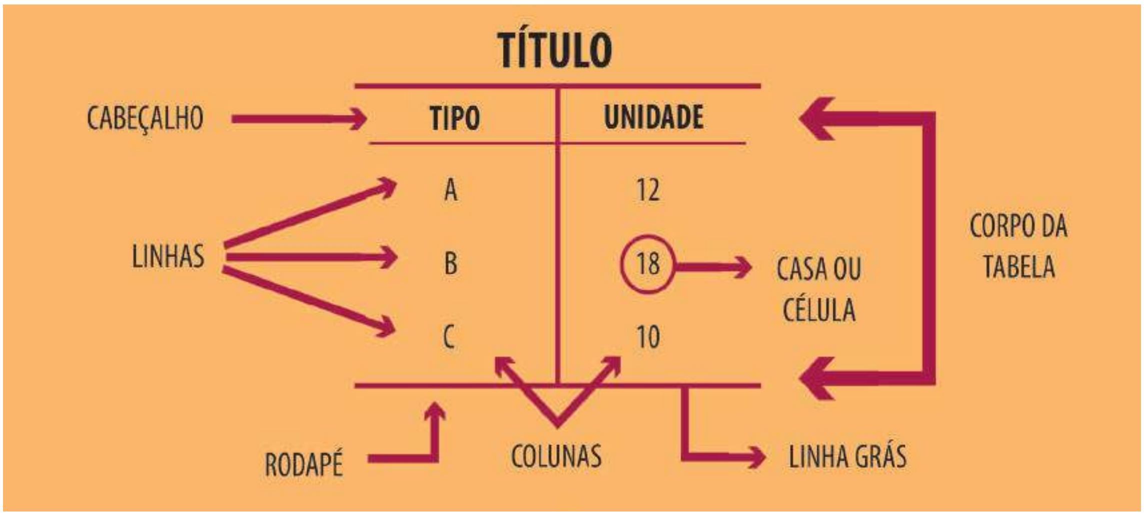


Ilustração: Sofia Guimarães

<u>**1. Título da Tabela:**</u> localizado no topo da tabela, deve conter informações, as mais completas possíveis, respondendo às perguntas: O que?, Quando? e Onde?, além de conter a palavra "Tabela" e sua respectiva numeração.

<u>**2. Corpo da Tabela:**</u> é o conjunto de linhas e colunas que contém informações sobre a variável em estudo, onde:

- na parte superior da tabela tem-se o cabeçalho da coluna, que especifica o conteúdo das colunas;
- verticalmente têm-se as colunas (indicadora e numérica), onde a coluna indicadora é aquela que especifica o conteúdo das linhas e na coluna numérica os valores numéricos destas linhas.

<u>3. Rodapé</u>: localizado na parte inferior da Tabela (fora) e contém informações sobre o responsável pela informação (Fonte), algum texto esclarecedor acerca do conteúdo da tabela (Nota) e por fim algum símbolo remissível atribuído a algum elemento da tabela que necessite de uma nota (Chamada).

Observação: Nenhuma casa deve ficar sem preenchimento. Todas devem ter o registro de algum número ou sinal:

- **- (hífen):** quando o valor numérico é nulo;
- **... (reticências):** quando não se dispõe de dado;
- **? (ponto de interrogação):** quando há dúvida sobre a exatidão do valor;
- **0; 0,0; ou 0,00 (zero)**: quando o valor numérico é pequeno para ser expresso pela unidade utilizada. Este deve conter o mesmo número de casas decimais padronizado pela tabela;
- **x (letra x):** quando o dado for omitido a fim de evitar individualização da informação.

Regras para Tabelas

a) em artigos ou publicações que contenham muitas tabelas, estas serão numeradas em ordem crescente, conforme o aparecimento;

b) tabelas são fechadas no alto e embaixo por linhas horizontais, mas não à esquerda e à direita por linhas verticais. Traços verticais para separar colunas no corpo da tabela podem ser empregados;

c) uma vez definido um determinado número de casas decimais, esse número será mantido para todas as casas de modo a assegurar uniformidade na apresentação dos dados;

d) totais e subtotais serão destacados.

ATIVIDADE

Elabore uma tabela (conforme as normas) para a seguinte situação:

1. O Instituto Brasileiro de Siderurgia forneceu dados relativos à produção brasileira de Ferro-Gusa (1.000 t) em 1993, Nas seguintes unidades da Federação: Minas Gerais (12.888), Espírito Santo (3.174), Rio de Janeiro (5.008) e São Paulo (2.912).

Classifique-a em cronológica, geográfica ou específica:

Faça uma tabela que represente os valores das exportações de açúcar do Brasil fornecidas pelo Instituto do Açúcar e do Álcool, nos anos de 1965 a 1971, em milhares de dólares: 60193; 80114; 82826; 106879; 112064; 126740; 149548.

Faça um comentário sobre o que você consegue concluir analisando visualmente os dados da tabela.

Monte uma tabela para representar o seguinte fato: população da região Norte do Brasil em 1970, sabendo-se que em Rondônia, Acre, Amazonas, Roraima, Pará e Amapá, temos, respectivamente:116620 – 218006 – 960934 – 41638 – 219707 e 116480 habitantes, segundo dados do IBGE.

Escreva um parágrafo comentando o que se pode observar e concluir com base nos dados tabelados.

Faça uma tabela estatística para representar o movimento religioso de certo município no período de 1975 – 1977, que apresentou os seguintes dados: em 1975, houve 56738 habitantes batizados (dos quais 26914 eram do sexo feminino), 15884 casamentos e 13678 extremas-unções. Em 1976, houve 33915 batizados do sexo masculino e 29568 do sexo feminino; os casamentos foram em número de 17032 e as extremas-unções 14328. Em 1977, em um total de 71232, 34127 eram batizados do sexo masculino; as extremas-unções foram 16107 e os casamentos 16774.

Escreva um parágrafo comentando o que se pode observar e concluir com base nos dados tabelados.

2.5.2 Apresentação Gráfica

Constitui uma apresentação geométrica dos dados. Permite ao analista obter uma visão rápida e clara do fenômeno e sua variação. Nessa etapa, o interesse maior consiste em tirar conclusões que auxiliem o pesquisador a resolver seu problema.

Disponível em < https://static.pexels.com/photos/163032/office-pen-calculator-computation-163032.jpeg >. Acesso em: 01/09/2017

A apresentação gráfica é um complemento importante da apresentação tabular. A vantagem de um gráfico sobre a tabela está em possibilitar uma rápida impressão visual da distribuição dos valores ou das frequências observadas.

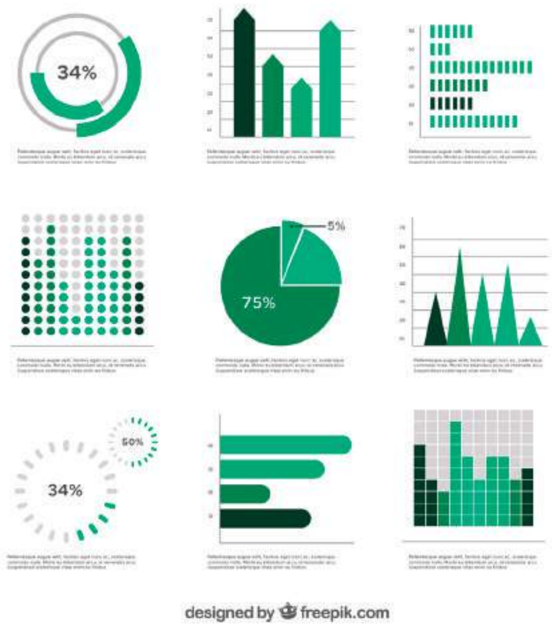
Os gráficos propiciam uma ideia inicial mais satisfatória da concentração e dispersão dos valores, uma vez que através deles os dados estatísticos se apresentam em termos de grandezas visualmente interpretáveis.

Requisitos Fundamentais em um Gráfico

- **Simplicidade:** possibilitar a análise rápida do fenômeno observado. Deve conter apenas o essencial.
- **Clareza:** possibilitar a leitura e interpretação correta dos valores do fenômeno.
- **Veracidade**: deve expressar a verdade sobre o fenômeno observado.

Tipos de Gráficos Quanto a Forma

Diagramas: gráficos geométricos dispostos em duas dimensões. São mais usados na representação de séries estatísticas.



Disponível em <https://static.pexels.com/photos/163032/office-pen-calculator-computation-163032.jpeg http://www.freepik.com/free-vector/set-of-green-graphics-in-flat-design_1256949.htm#term=statistics&page=1&position=6 >. Acesso em: 01/09/2017

Cartogramas: é a representação sobre uma carta geográfica, sendo muito usado na Geografia, História e Demografia.

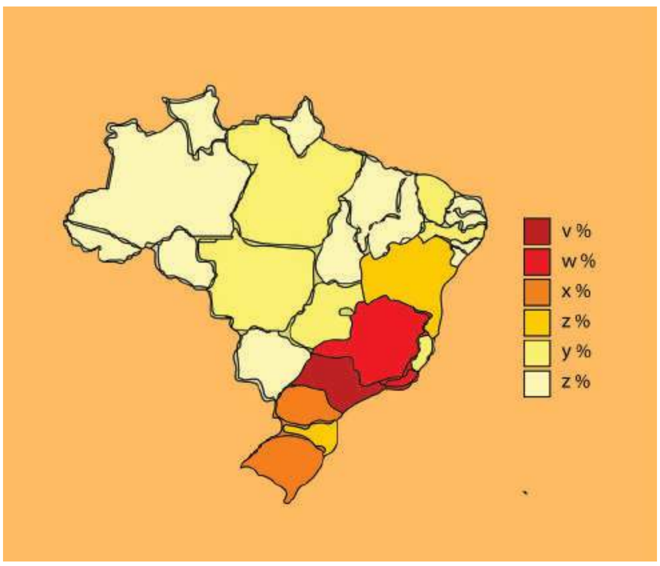
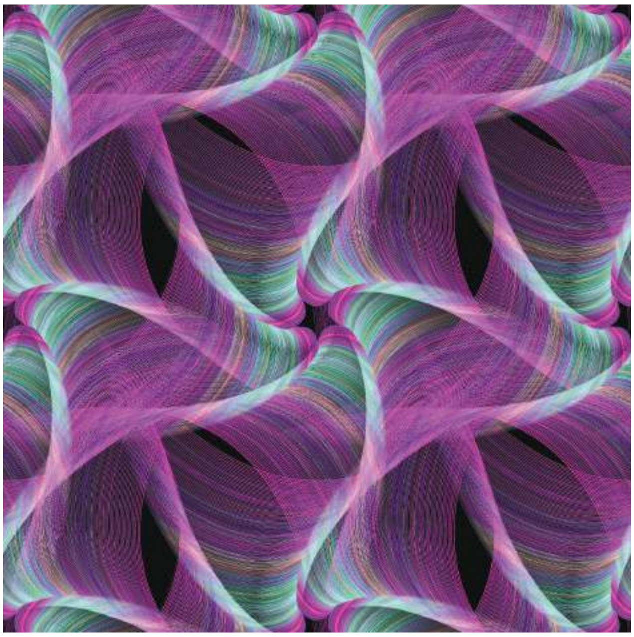


Ilustração: Orlando Dantas

Estereogramas: representam volumes e são apresentados em três dimensões.



Disponível em <http://www.freepik.com/free-vector/coloured-fractal-background-design_913511.htm>. Acesso em: 04/09/2017

Pictogramas: a representação gráfica consta de figuras representativas do fenômeno. Desperta logo a atenção do público.

Disponível em <http://www.freepik.com/free-vector/education-icons-set_1148964.htm#term=pictogram&page=1&position=11>. Acesso em: 01/09/2017

Os Diagramas

Gráfico de Barras: O gráfico de barras é composto por duas linhas ou eixos, um vertical e outro horizontal. No eixo vertical são construídas as barras que representam a variação de um fenômeno ou de um processo de acordo com sua intensidade. Essa intensidade é indicada pela altura da barra. No eixo horizontal especificam-se as categorias da variável. As barras devem sempre possuir a mesma largura e a distância entre elas deve ser constante.

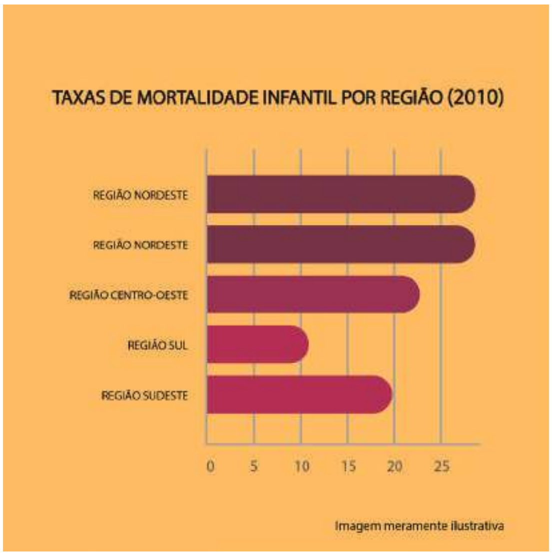


Ilustração: Orlando Dantas

Gráficos de barras empilhadas mostram o relacionamento de itens individuais com o todo

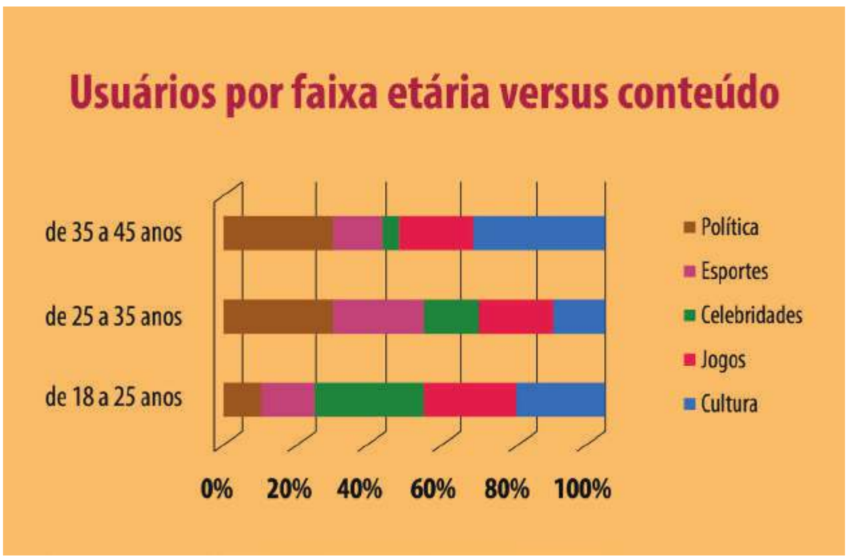


Ilustração: Orlando Dantas

Gráfico de Coluna: Um gráfico de colunas mostra as alterações de dados em um período de tempo ou ilustra comparações entre itens. As categorias são organizadas na horizontal e os valores são distribuídos na vertical, para enfatizar as variações ao longo do tempo.

Ilustração: Orlando Dantas

Gráficos de colunas empilhadas mostram o relacionamento de itens individuais com o todo

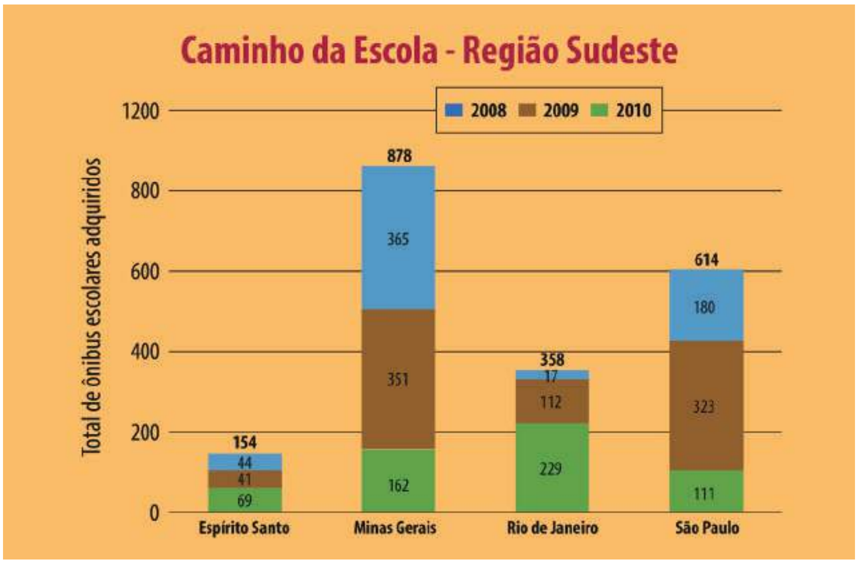


Ilustração: Orlando Dantas

Gráfico de Pizza: Um gráfico de pizza mostra o tamanho proporcional de itens que constituem uma série de dados para a soma dos itens. Ele sempre mostra somente uma única série de dados, sendo útil quando você deseja dar ênfase a um elemento importante.

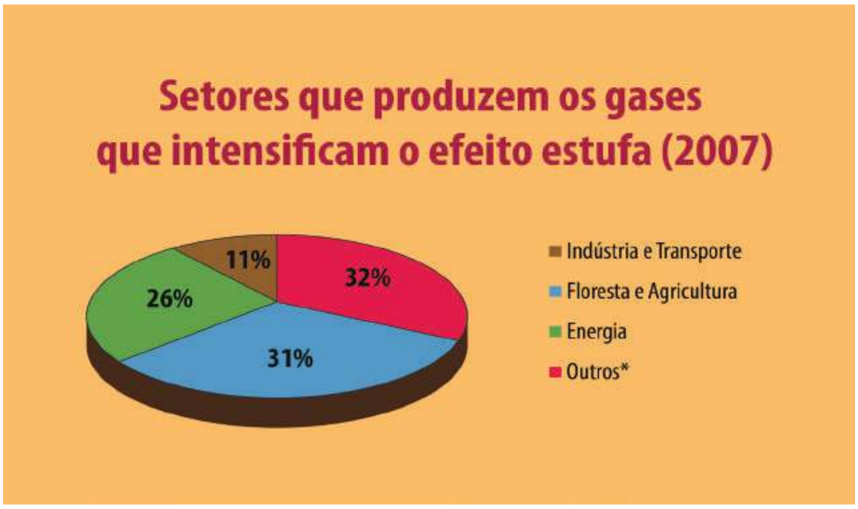


Ilustração: Orlando Dantas

Para facilitar a visualização de fatias pequenas, você pode agrupá-las em um único item do gráfico de pizza e subdividir esse item em um gráfico de pizza ou de barras menor, ao lado do gráfico principal.

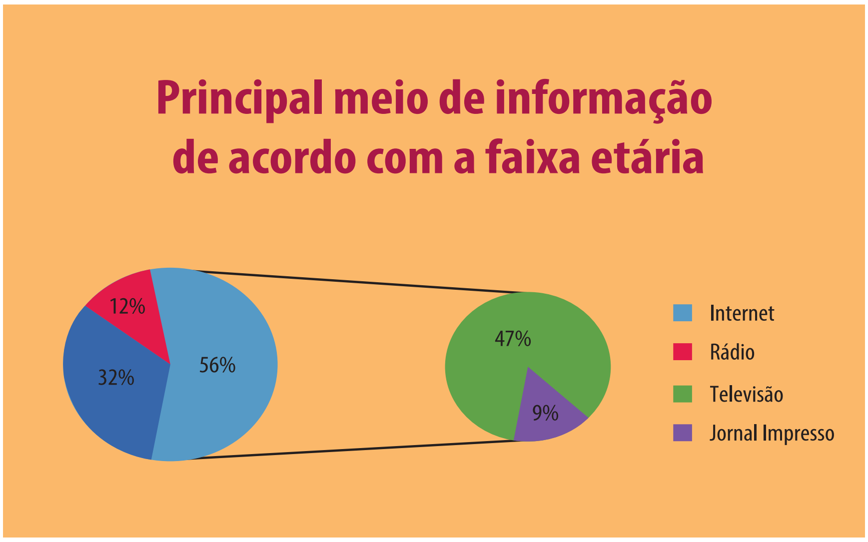


Ilustração: Antonio Felix

Gráfico de Rosca: Como um gráfico de pizza, o gráfico de rosca mostra o relacionamento das partes com o todo, mas pode conter mais de uma série de dados. Cada anel do gráfico de rosca representa uma série de dados.

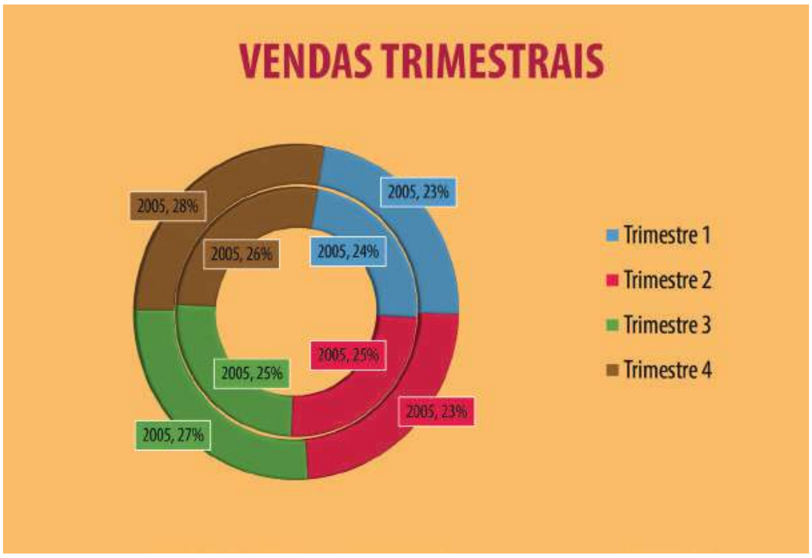


Ilustração: Orlando Dantas

Gráfico de Linha: Um gráfico de linhas é utilizado para mostrar evolução, ou tendências, nos dados em intervalos iguais.

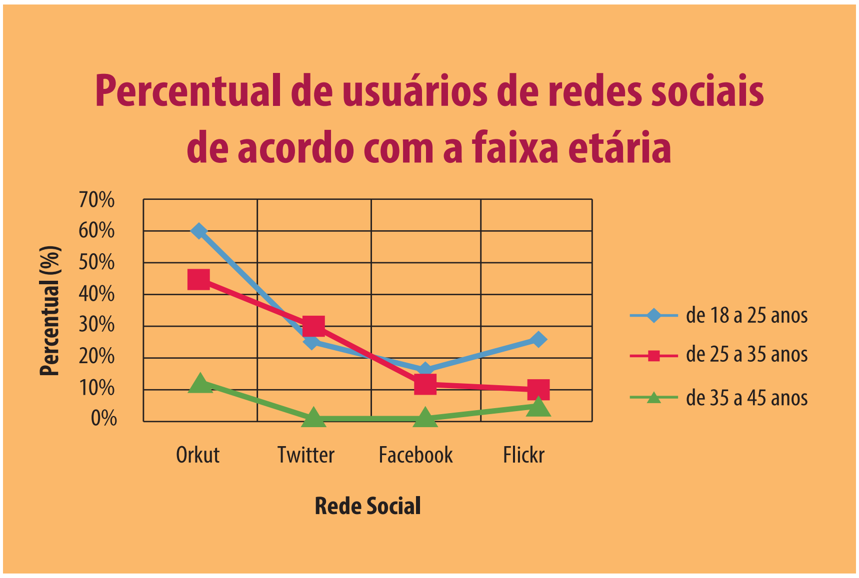


Ilustração: Antonio Felix

Gráfico de Dispersão XY: Um gráfico xy (dispersão) mostra a relação existente entre os valores numéricos em várias séries de dados ou plota dois grupos de números como uma série de coordenadas xy. Este gráfico mostra intervalos irregulares ou clusters de dados e é usado geralmente para dados científicos. Quando ordenar seus dados, coloque valores x em uma linha ou coluna e insira valores y correspondentes nas linhas ou colunas adjacentes.

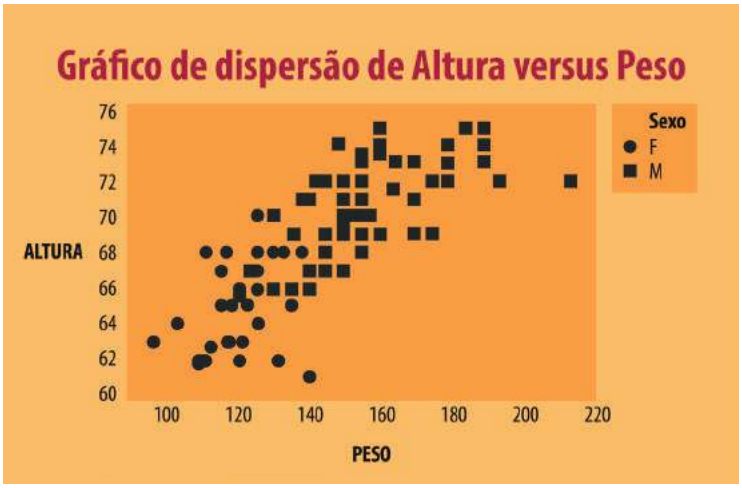


Ilustração: Orlando Dantas

Gráfico de Bolhas: Um gráfico de bolhas é um tipo de gráfico xy (dispersão). O tamanho do marcador de dados indica o valor de uma terceira variável. Para organizar seus dados, coloque os valores de x em uma linha ou coluna e insira os valores de y e os tamanhos das bolhas correspondentes nas linhas ou colunas adjacentes.

Quantidade de produtos	Vendas	% da Participação no Mercado
14	R$ 12 200	15%
20	R$ 60 000	33%
18	R$ 24 400	10%
22	R$ 32 000	42%

Ilustração: Orlando Dantas

Gráfico de Superfície: Um gráfico de superfície é útil quando você deseja localizar combinações vantajosas entre dois conjuntos de dados. Como em um mapa topográfico, as cores e os padrões indicam áreas que estão no mesmo intervalo de valores.

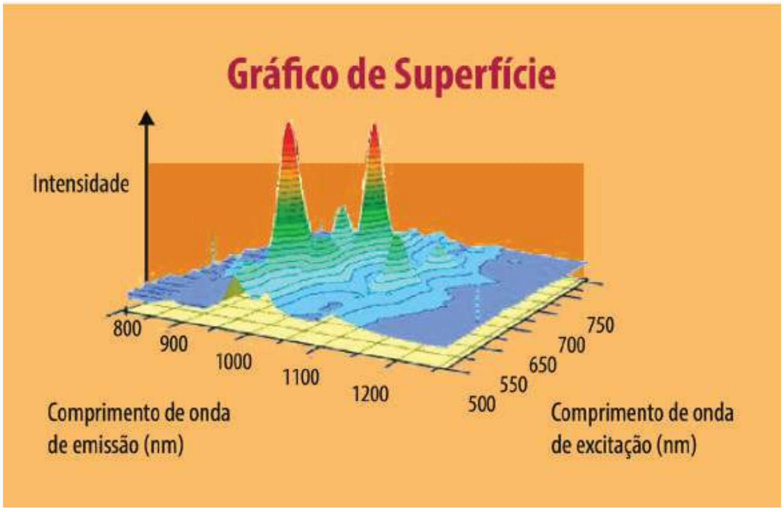


Ilustração: Orlando Dantas

2.6 Distribuição de frequências

As tabelas estatísticas, geralmente, condensam informações de fenômenos que necessitam da coleta de grande quantidade de dados numéricos. No caso das distribuições de frequências que é um tipo de série estatística, os dados referentes ao fenômeno objeto de estudo se repetem na maioria das vezes sugerindo a apresentação em tabela onde aparecem valores distintos uns dos outros.

Representação dos Dados Amostrais ou Populacionais

Definição 2.6.1: Dados Brutos: *são aqueles que não foram numericamente organizados, ou seja, estão na forma com que foram coletados.*

Exemplo: Número de filhos de um grupo de 50 casais.

.

2	3	0	2	1	1	1	3	2	5
6	1	1	4	0	1	5	6	0	2
1	4	1	3	1	7	6	2	0	1
3	1	3	5	7	1	3	1	1	0
3	0	4	1	2	2	1	2	3	2

Ilustração: Sofia Guimarães

Definição 2.6.2: Rol: é a organização dos dados brutos em ordem de grandeza crescente ou decrescente.

Exemplo: Número de filhos de um grupo de 50 casais.

0	0	0	0	0	0	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	2	2	2	2	2	2	2	2
2	3	3	3	3	3	3	3	3	4
4	4	5	5	5	6	6	6	7	7

Ilustração: Sofia Guimarães

2.6.1 Distribuição de frequência por Intervalos

Definição 2.6.1.1: *É a série estatística que condensa um conjunto de dados conforme as frequências ou repetições de seus valores.*

Os dados encontram-se dispostos em classes ou categorias junto com as frequências correspondentes. Neste caso **todos** os elementos são fixos (época, local e fenômeno). A distribuição de frequência pode ser por intervalo ou por pontos, dependendo da quantidade de informações que se tenha e/ou do tipo de variável. É um tipo de tabela que condensa uma coleção de dados conforme as frequências (repetições de seus valores).

É utilizada quando se tem dados quantitativos discretos ou contínuos. Exemplo de Distribuição de frequência por Intervalo:

Tabela 2.6.1: Tempo de Estudo (em dias) de 160 alunos da Escola W, no ano de 2012.

Classe (i)	Tempo (em dias)	Nº de Alunos (F_i)	$Fac_{(i)}$	Fr_i	$Fr_{(ac)i}$	X_i
1ª	10 |-20	38				
2ª	20 |-30	45				
3ª	30 |-40	30				
4ª	40 |-50	22				
5ª	50 |-60	10				
6ª	60 |-|70	15				
	Total	160				

Ilustração: Sofia Guimarães

Construção de Distribuições de frequências por Intervalos:

1º Passo: montar o Rol (organizar os dados em ordem crescente ou decrescente).

2º Passo: calcular a Amplitude Total da distribuição de frequência (ΔT), que é a diferença existente entre o maior ($X_{máx}$) e o menor valor (X_{min}) observado.

3º Passo: Determinar o Número de Classes da Distribuição de frequência (), que são os subintervalos nos quais são contadas as observações da variável. Existem várias maneiras de se calcular o número de classes, neste curso será utilizado o método prático.

- Se $n < 25$, utiliza-se k = 5classes;

- Se $n \geq 25$, utiliza-se $k = \sqrt{n}$

- Fórmula de Sturges $k = 1 + 3{,}22\log(n)$.

Observação: Sempre arredondar o valor de (é um número inteiro) seguindo as regras de arredondamento.

Arredondamento de Dados

Conforme IBGE (1993) deve-se utilizar as seguintes regras de arredondamento:

Regra I - Se o 1º algarismo a ser abandonado for menor que "5" o último a permanecer fica inalterado.

Exemplo: Arredondar para centésimos os números abaixo.

a) 47,3227 47,32 b) 0,29364 0,29 c) 53,77474 53,77 d) 30,00132 30,00

Regra II - Se o último algarismo a permanecer for maior ou igual a "5" será acrescido de uma unidade o algarismo a permanecer.

Exemplo: Arredondar para décimos os números abaixo.

a) 1,4632 1,5 b) 23,09425 23,1 c) 38,97777 39,0 d) 74,28583 74,3

4º Passo: calcular o intervalo de Classe ou Amplitude do Intervalo de Classe (h), que é o comprimento da classe.

$$h = \frac{\Delta T}{k}$$

5º Passo: Construção das Classes

1ª Classe: Limite Inferior (l_i) = menor valor do Rol e Limite Superior (l_s) = l_i da 1ª Classe + h;

2ª Classe: l_i = l_s da 1ª Classe e l_s = l_i da 2ª Classe + h ;

kª Classe: l_i = l_s da (k-1)ª Classe e l_s = l_i da kª Classe + h.

Convenção:

|- inclui o número da esquerda e exclui o da direita.

-| exclui o número da esquerda e inclui o da direita.

- exclui ambos.

|-| inclui ambos.

6º Passo: obtenção da frequência Simples ou frequência Absoluta da Classe (f_i) que é o número de observações contadas dentro da classe.

7º Passo: Ponto Médio de Classe (X_i)**:** é a média aritmética calculada entre o limite inferior (l_i) e o superior (l_s) da classe. É o valor em estatística que representa os valores da variável dentro da classe.

$$X_i = \frac{(l_i + l_s)}{2}$$

Tipos de frequências

Definição 2.6.1.2: frequência Absoluta Acumulada de Classe *(Fac_i): é a acumulação sucessiva, a partir da primeira classe até uma classe qualquer, das frequências simples ou absoluta das classes.*

$$\begin{gathered} Fac_1 = f_1 \\ Fac_2 = f_1 + f_2 \\ \dots \\ Fac_k = f_1 + f_2 + \dots + f_k \end{gathered}$$

***Definição 2.6.1.3: frequência Relativa de Classe** (fr_i): é a relação existente entre a frequência absoluta ou simples de classe i e o número de observações da variável.*

$$fr_i = \frac{f_i}{\sum_{i=1}^{k} f_i}$$

Observação: Temos que $\sum_{i=1}^{k} fr_i = 1$e $n = \sum_{i=1}^{k} f_i$

***Definição 2.6.1.4: frequência Relativa Acumulada** ($Fr_{(ac)i}$): é a acumulação sucessiva, a partir da primeira classe até uma classe qualquer das frequências relativas das classes.*

$$Fr_{(ac)1} = fr_1$$
$$Fr_{(ac)2} = fr_1 + fr_2$$
$$\dots$$
$$Fr_{(ac)k} = fr_1 + fr_2 + \dots + fr_k$$

Percentual

***Definição 2.6.1.5: Percentual Simples:** é o percentual de cada intervalo e é calculada em função da frequência simples;*

***Definição 2.6.1.6: Percentual Acumulado:** é o percentual de cada intervalo e é calculada em função da frequência acumulada.*

2.6.2 Distribuição por Pontos

Definição 2.6.2.1: Distribuição de frequência por Pontos: É uma série est*atística na quais as frequências observadas estão associadas a um ponto real observado.*

Na construção da distribuição por ponto, o número linhas (classes) na tabela é igual ao número de pontos existentes, e utilizam-se os mesmos elementos da distribuição por intervalo, com a diferença que o próprio ponto já é o valor de X_i (ponto médio na distribuição por intervalo). Exemplo de Distribuição de frequência por Pontos:

Tabela 2.6.2.1: Número de Dependentes dos Professores da E.E.E.F. ABC em Dezembro/2011.

Ponto	Nº de Dependentes (X_i)	Professores (F_i)
1º	0	40
2º	1	50
3º	2	30
4º	3	20
5º	4	10
	Total	150

Fonte: Dados Hipotéticos, JAN/2012

Ilustração: Sofia Guimarães

Exemplo: Considere os dados brutos que representam a safra, em alqueires, por árvore, para um conjunto de 40 pessegueiros.

Safra atual em (alqueire/ árvore) para 40 Pessegueiros

11,1 12,5 32,4 7,8 21,0 16,4 11,2 22,3

4,4 6,1 27,5 32,8 18,5 16,4 15,1 6,0

10,7 15,8 25,0 18,2 12,2 12,6 4,7

23,5 14,8 22,6 16,0 19,1 7,4 9,2 10,0

26,2 3,5 16,2 14,5 3,2 8,1 12,9 19,1 13,7

Vamos construir uma tabela para representar esses dados:

1- Construção do Rol em ordem crescente

Safra atual em (alqueire/ árvore) para 40 Pessegueiros

3,2 3,5 4,4 4,7 6 6,1 7,4 7,8

8,1 9,2 10,0 10,7 11,1 11,2 12,2 12,5

12,6 12,9 13,7 14,5 14,8 15,1 15,8 16,0

16,2 16,4 16,4 18,2 18,5 19,1 19,1 21,0

22,3 22,6 23,5 25,0 26,2 27,5 32,4 32,8

2- Amplitude Total:

$$\Delta T = 32{,}8 - 3{,}2 = 29{,}6$$

3- Número de classes:

$$k = \sqrt{n}$$

4- Amplitude da classe:

$$h = \frac{29{,}6}{6} = 4{,}93 \approx 5$$

5- Intervalos de variação das classes:

- **1ª classe:** Limite inferior: 3 Limite superior: 3 + = 3 + 5 = 8
- **2ª classe:** Limite inferior = Limite superior da classe anterior: 8 Limite superior: 8 + 5 = 13
- **3ª classe:** Vai de 13 a 18
- **4ª classe:** Vai de 18 a 23
- **5ª classe:** Vai de 23 a 28
- **6ª classe:** Vai de 28 a 33

6 – Construção da tabela

Safra Anual (em alqueire/árvore) para 40 Pessegeiros

Número de alqueires	Número de árvores
$3 \vdash 8$	8
$8 \vdash 13$	10
$13 \vdash 18$	9
$18 \vdash 23$	7
$23 \vdash 28$	4
$28 \vdash 33$	2
Total	40

Ilustração: Sofia Guimarães

6 – Graficamente podemos utilizar o Histograma, que é um gráfico de colunas formado pelas classes da tabela e pela frequência simples de cada classe, como pode ser visto abaixo.

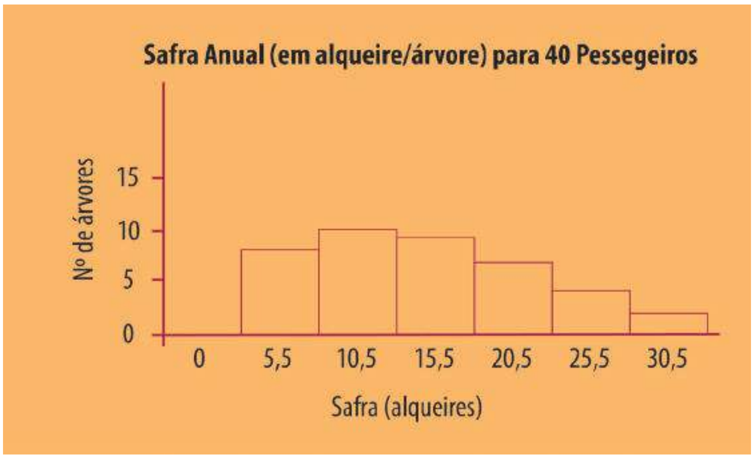


Ilustração: Sofia Guimarães

Sabendo um pouco mais

Link para auxiliar no tópico Distribuição de frequências

https://www.youtube.com/watch?v=hdR3V86dOg0

Usando o Excel para construir a Tabela de Distribuição de frequência

Vamos supor que uma empresa fabricante de lâmpadas deseja testar uma parte de sua produção. Selecionou 60 lâmpadas de 100W e deixou-as ligadas te que queimassem. O tempo de vida útil de cada uma delas está registrado na tabela abaixo. O objetivo é construir a tabela de Distribuição de frequências completa usando o Excel.

Para construir a tabela de Distribuição de frequência usando o aplicativo Excel você deverá seguir os seguintes passos:

Tempo de Vida Útil de Lâmpadas de 100 W

---

684 796 859 939 773 697 693 721 832 902 1004 857 819 907 1038 912

1005 952 836 888 962 1096 994 918 868 905 926 786 852 870 893

922 1016 760 860 899 911 938 1093 1016 859 971 924 1041 742 920

848 977 1005 1080 821 852 876 1014 1052 773 909 762 984 954

Passo 1: Digite todos os valores que a variável **"duração ou vida útil"** assume em uma única coluna, a partir da célula A1 até a célula A60.

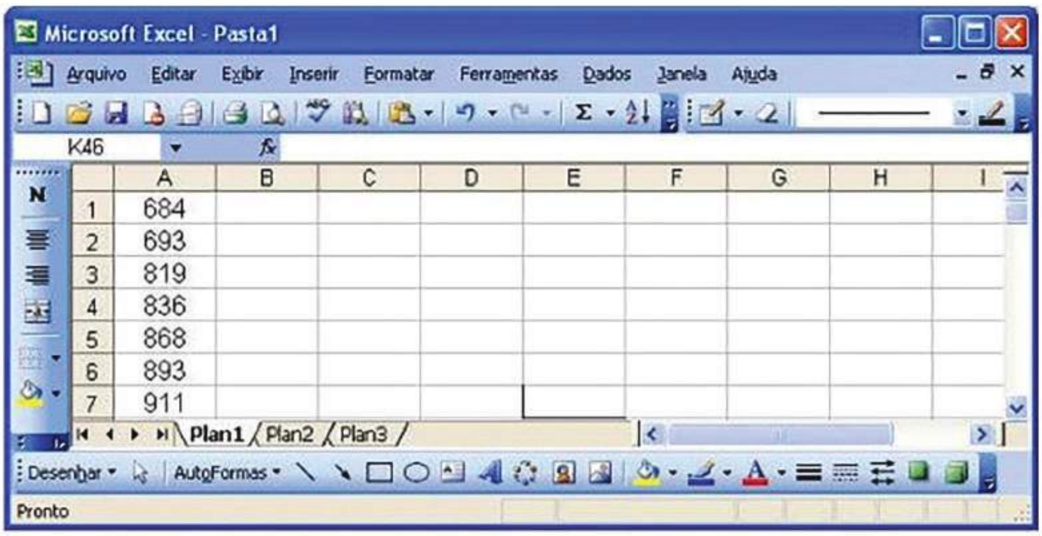


Ilustração: Mariana Netto

Passo 2: Selecione a coluna de A1 até A60 e na barra de **ferramentas** clique no ícone **Classificação Crescente.** Esta função colocará os valores em ordem crescente, ou seja, do menor para o maior.

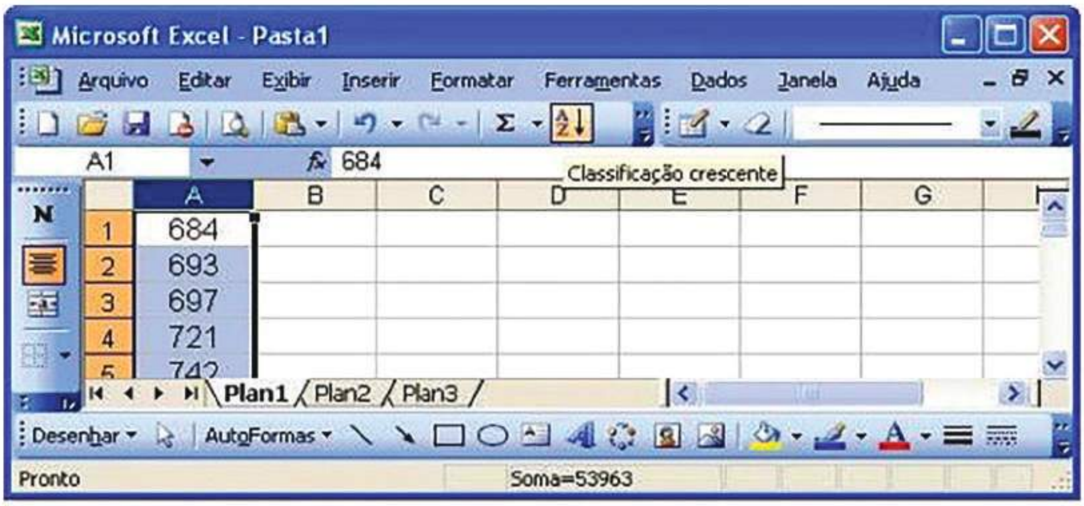


Ilustração: Mariana Netto

Passo 3: Na célula C5, digite n. Na célula à direita, ou seja, D5 digite = e clique no ícone **Função**. Aparecerá uma tela chamada **Inserir Função**. Em categoria selecione **todas** e em função escolha **CONT.NÚM.** Para finalizar clique OK.

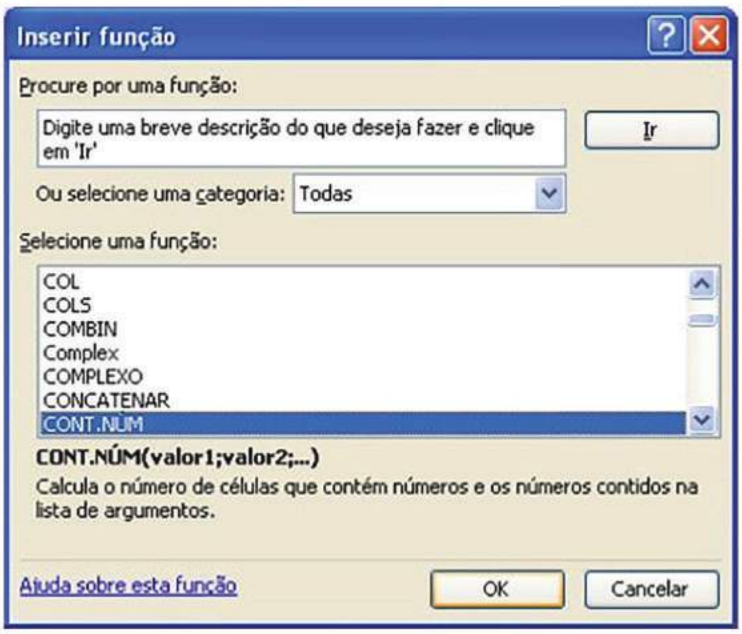


Ilustração: Mariana Netto

Passo 4: Em Valor 1 selecione a tabela do A1 até A60 ou digite A1:A60. Clique OK novamente. A função **CONT.NÚM** conta quantos elementos fazem parte da amostra.

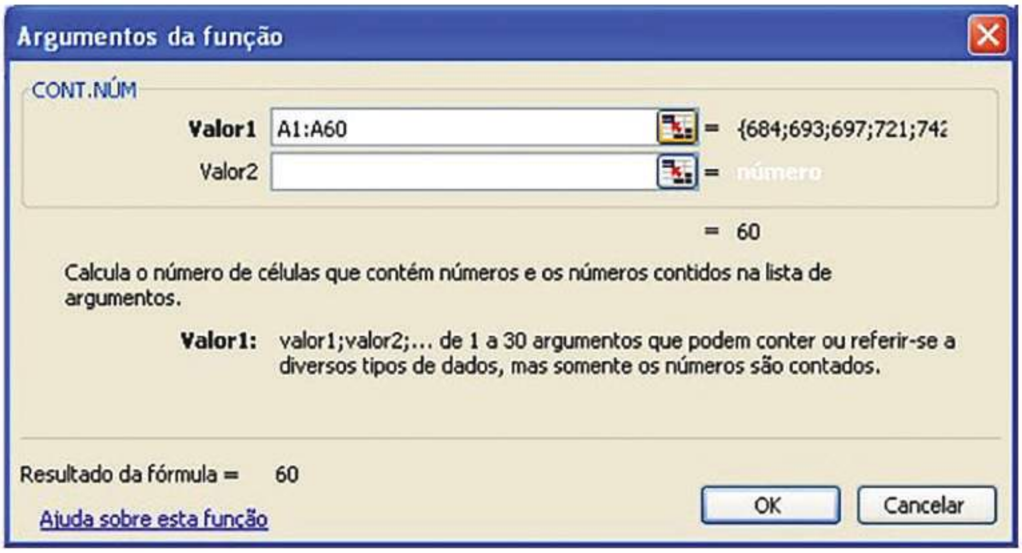


Ilustração: Mariana Netto

Passo 5: Aparecerá o resultado 60 na célula D5.

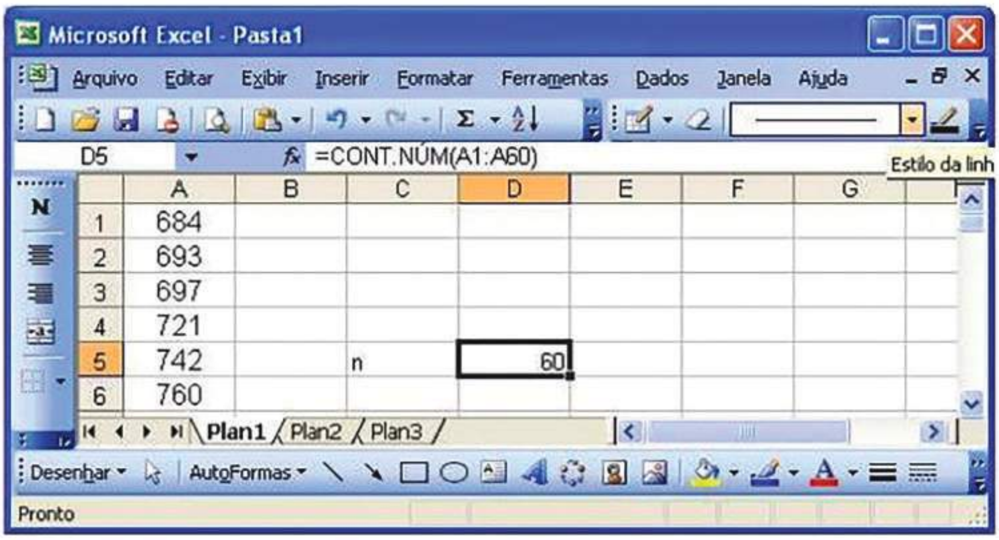


Ilustração: Mariana Netto

Passo 6: Na célula C6, digite xmin. Na célula à direita, ou seja, D6 digite = e clique no ícone **Função**. Aparecerá uma tela chamada **Inserir Função**. Em categoria clique escolha todas e em função escolha **MÍNIMO**. Para finalizar clique OK.

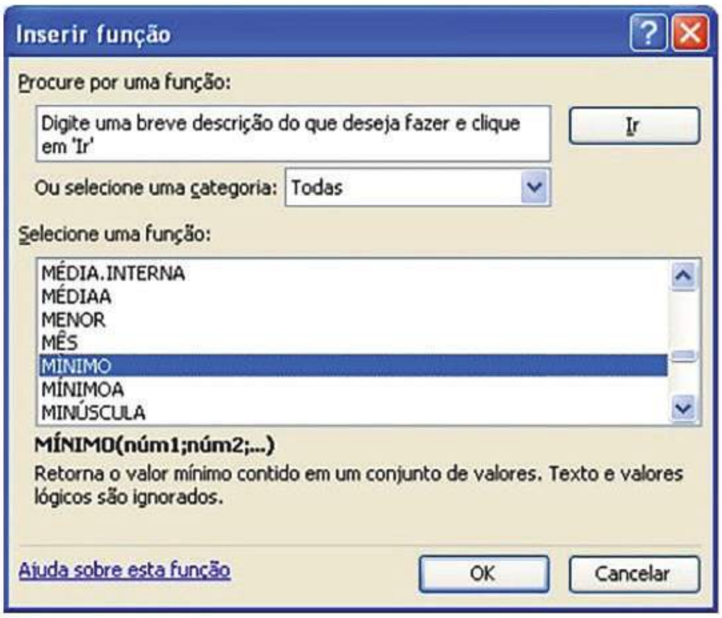


Ilustração: Mariana Netto

Passo 7: Em **Num 1** selecione a tabela do A1 até A60 ou digite A1:A60. Clique OK novamente. Esta função procura o valor **mínimo** de uma amostra.

Ilustração: Mariana Netto

Passo 8: Aparecerá o valor 684 na célula D6.

Ilustração: Mariana Netto

Passo 9: Na célula C7 digite **xmax.** Na célula à direita, ou seja, D7 digite = e clique no ícone **Função.** Aparecerá uma tela chamada **Inserir Função**. Em categoria selecione **todas** e em função escolha **MÁXIMO**. Para finalizar clique OK.

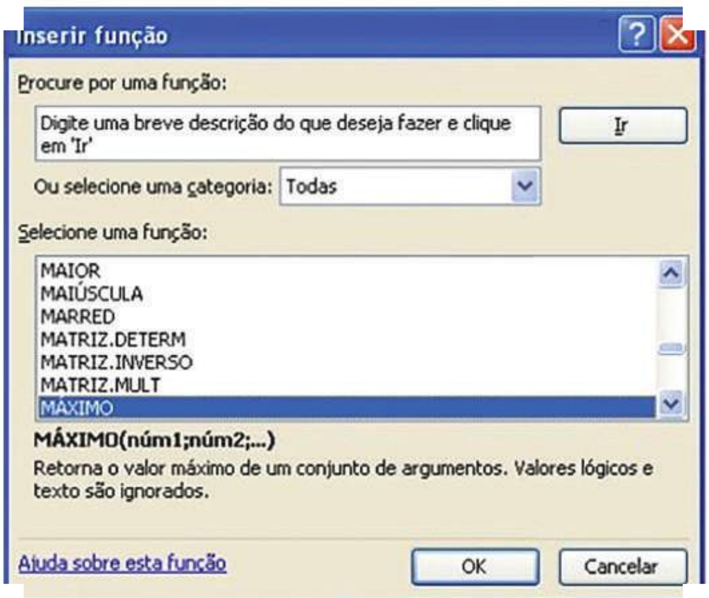


Ilustração: Mariana Netto

Passo 10: Em **Núm 1** selecione a tabela do A1 até A60 ou digite A1:A60. Clique OK novamente. Esta função procura o valor **máximo** de uma série de dados.

Ilustração: Mariana Netto

Passo 11: Na célula D7 aparecerá o valor 1096.

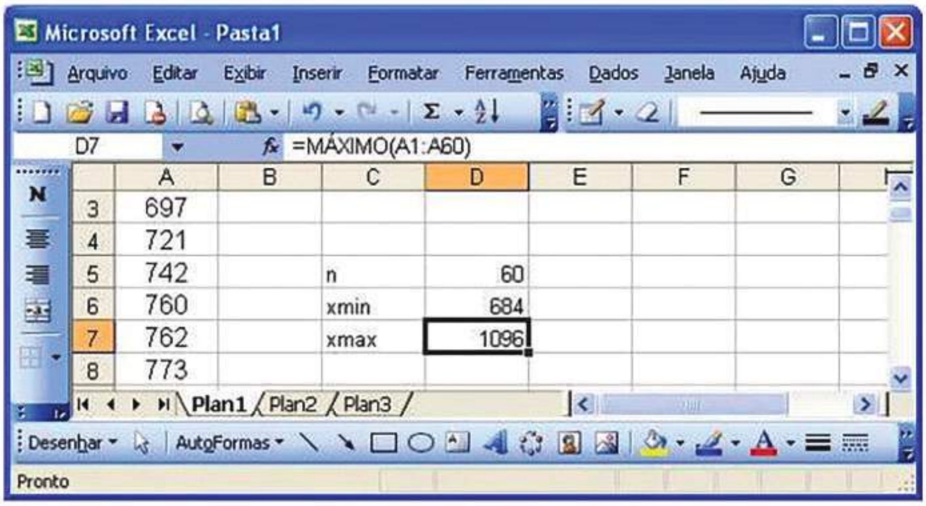


Ilustração: Mariana Netto

Passo 12: Na célula C8, digite AA. Na célula à direita, ou seja, D8 digite o sinal de igual (=) , então selecione a célula D7, digite o sinal de menos e selecione a célula D6. Finalize com **Enter**.

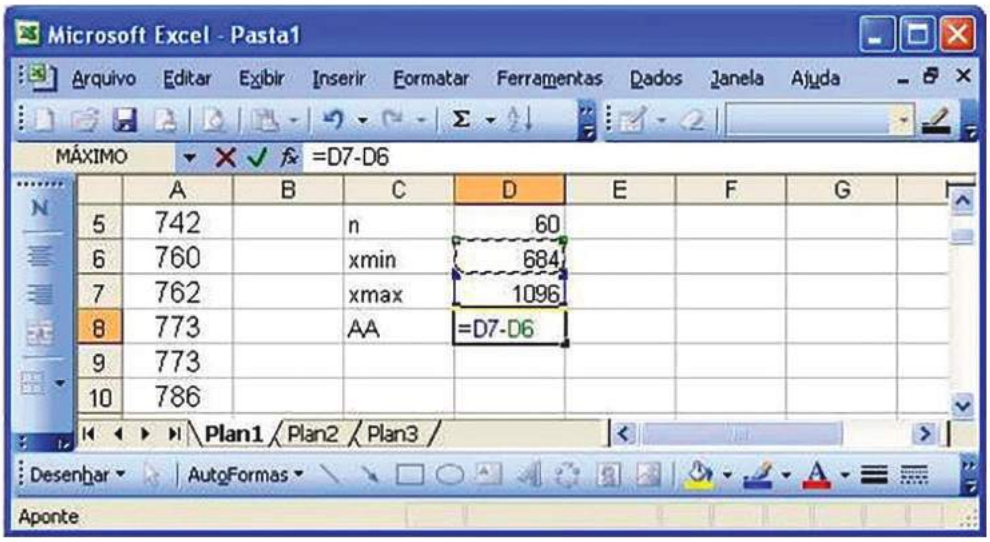


Ilustração: Mariana Netto

Passo 13: O valor dessa célula corresponde à amplitude amostral, ou seja, a variação da duração das 60 lâmpadas analisadas (em Watts).

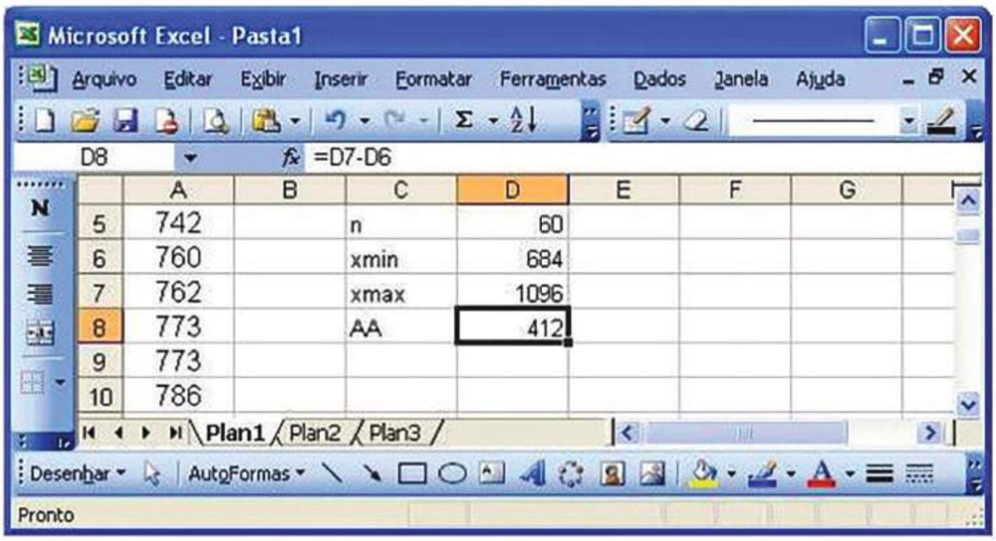


Ilustração: Mariana Netto

Passo 14: Na célula C9 digite **log n**. Na célula D9 digite = e clique no ícone **Função**. Selecione a função **log10** e clique OK.

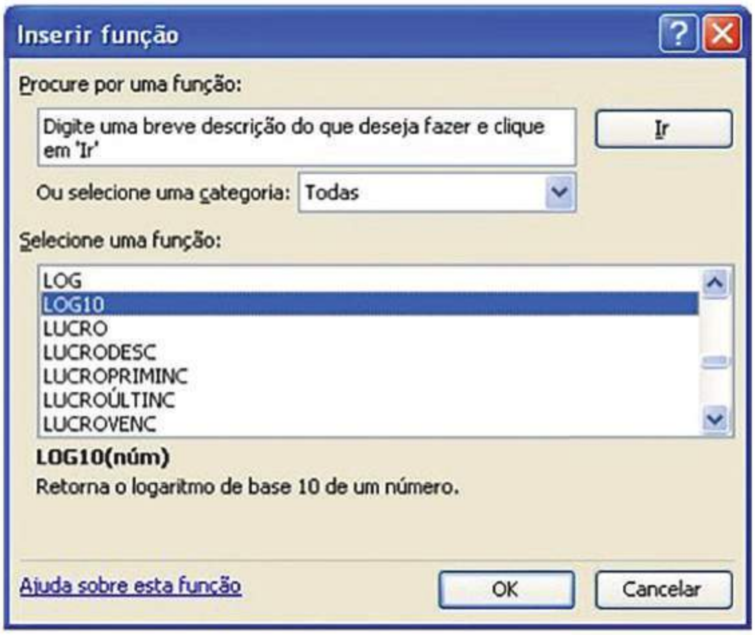


Ilustração: Mariana Netto

Passo 15: Em **Num**. selecione a célula D5 e clique OK.

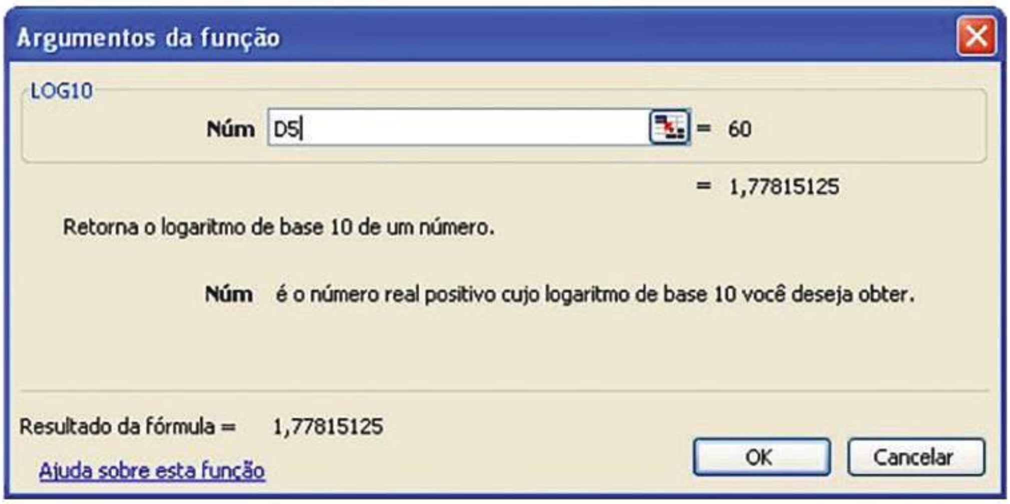


Ilustração: Mariana Netto

Passo 16: Esta função calcula o logaritmo de 60 que será usado na fórmula que calcula o número de classes (i).

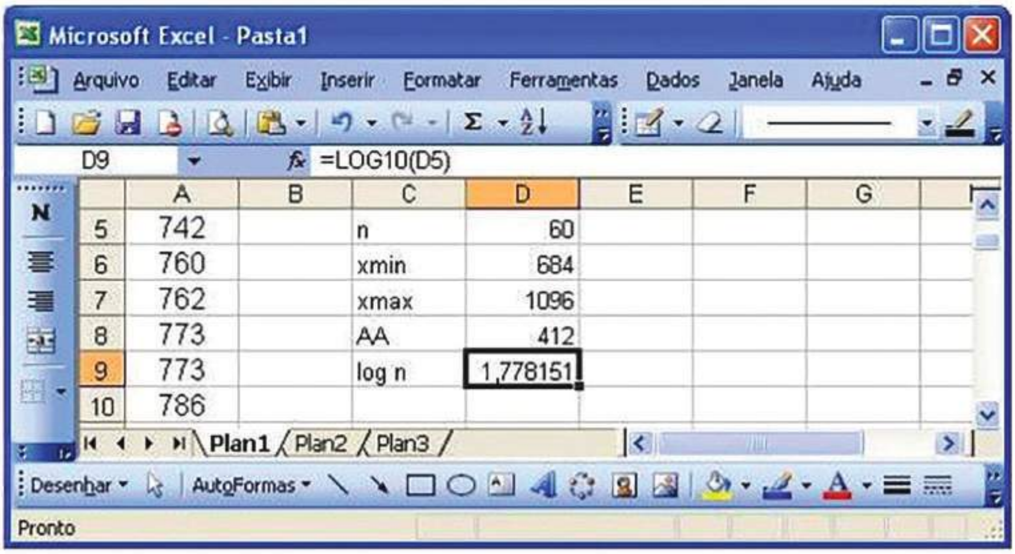


Ilustração: Mariana Netto

Passo 17: Na célula C10, digite i. Na célula à direita, ou seja, D10 digite = e clique no ícone **Função**. Escolha a função **ARRED.** Para finalizar clique OK.

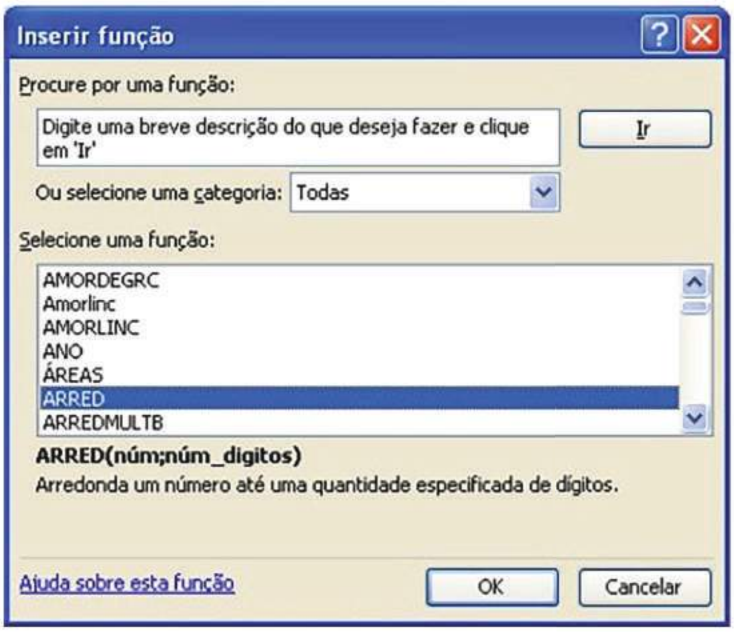


Ilustração: Mariana Netto

Passo 18: Em Núm digite (**1+(3,3*D9)**). Essa é a fórmula que calcula o número de classes, já conhecida por todos. Em **Num_dígitos** digite **0 (zero)**. Finalize com OK

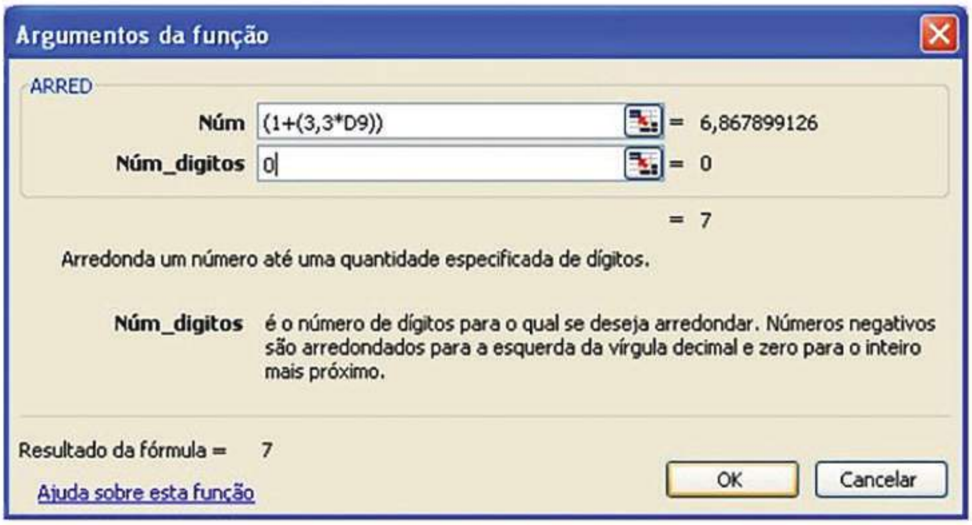


Ilustração: Mariana Netto

Observação: O zero representa um número inteiro, assim como o número 1 representa um número com uma casa decimal, o 2 representa duas casas decimais e assim por diante. Lembre-se que o número de classes deve ser sempre um número inteiro, por isso usamos o 0 (zero).

Passo 19: Na célula D10 aparecerá o número de classes que a tabela de distribuição de frequências terá.

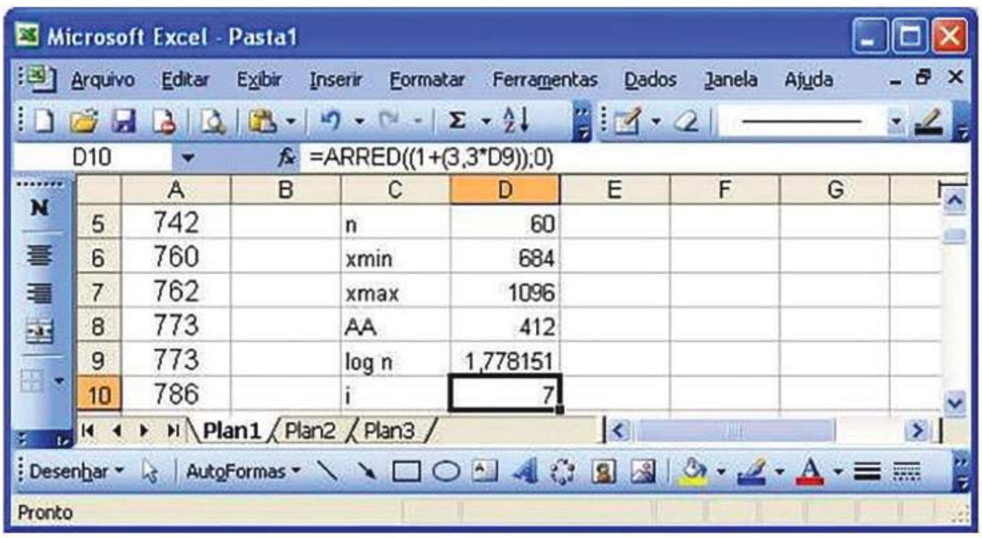


Ilustração: Mariana Netto

Passo 20: Na célula C11, digite **h**. Na célula à direita, ou seja, D11 digite = e clique no ícone **Função**. Selecione a função **ARREDONDAR.PARA.CIMA** . Para finalizar clique OK. Em **Num** digite D8/D10. Em **Num_dígitos** digite **0 (zero)**.

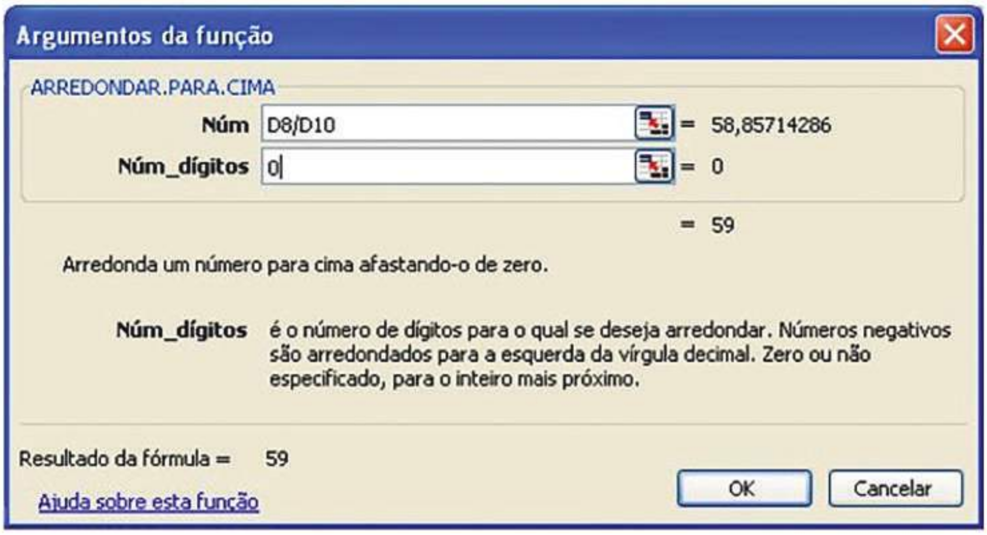


Ilustração: Mariana Netto

Passo 21: Na célula D11 aparecerá o valor da amplitude do intervalo de classe (**h**).

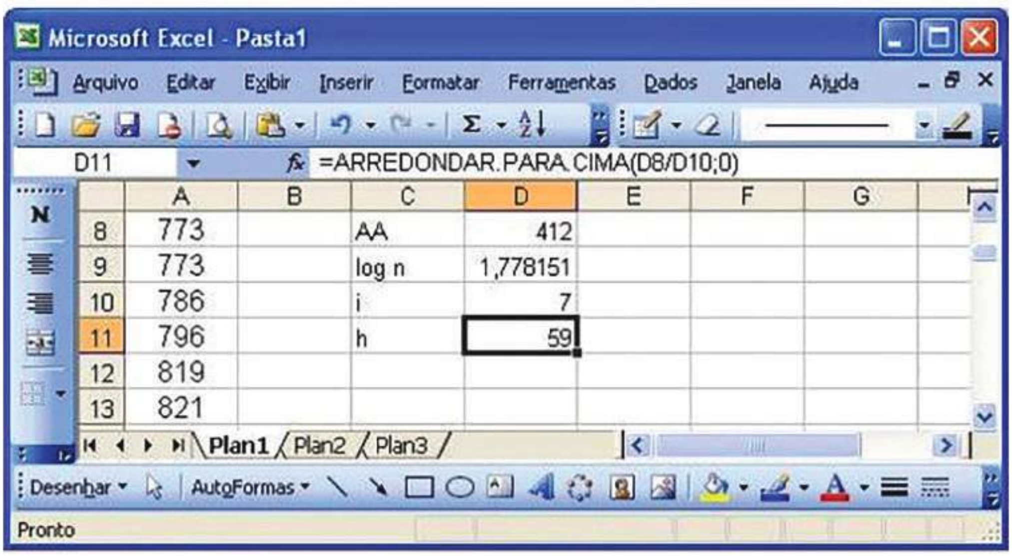


Ilustração: Mariana Netto

Passo 22: Na célula C13, digite li, em D13, **li**, em E13, **fi**, em F13, **fri**, em G13, **Fi**, em H13, **Fri** e em I13, **Xi**.

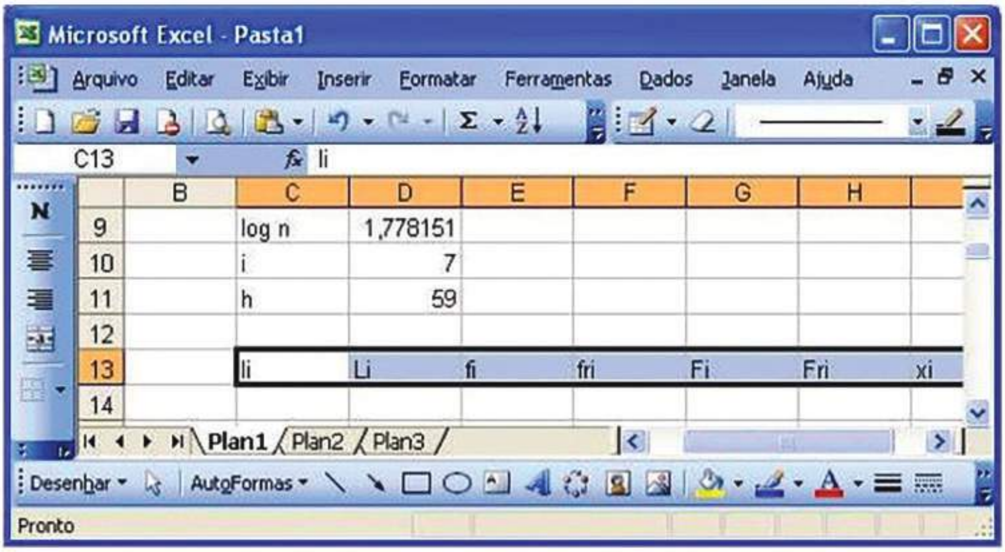


Ilustração: Mariana Netto

Passo 23: Na célula C14, digite o sinal de =, selecione a célula D6 e clique Enter. O valor que aparecerá na célula é o limite inferior do primeiro intervalo de classe.

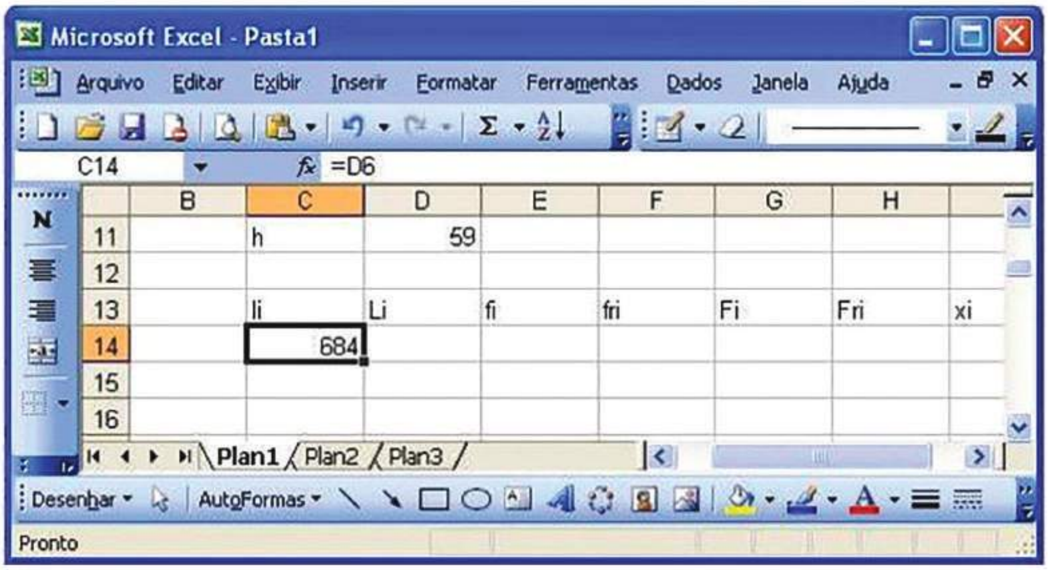


Ilustração: Mariana Netto

Passo 24: Na célula D14, digite o sinal de =, selecione a célula C14 e digite **+ D11** e clique Enter. O cifrão antes e depois do nome da coluna serve para congelar está célula. Quando calculamos os limites da classe, a partir do anterior sempre somamos o valor de **h**, aqui representado pela célula D11. Aparecerá na célula D14 o limite superior da 1ª classe.

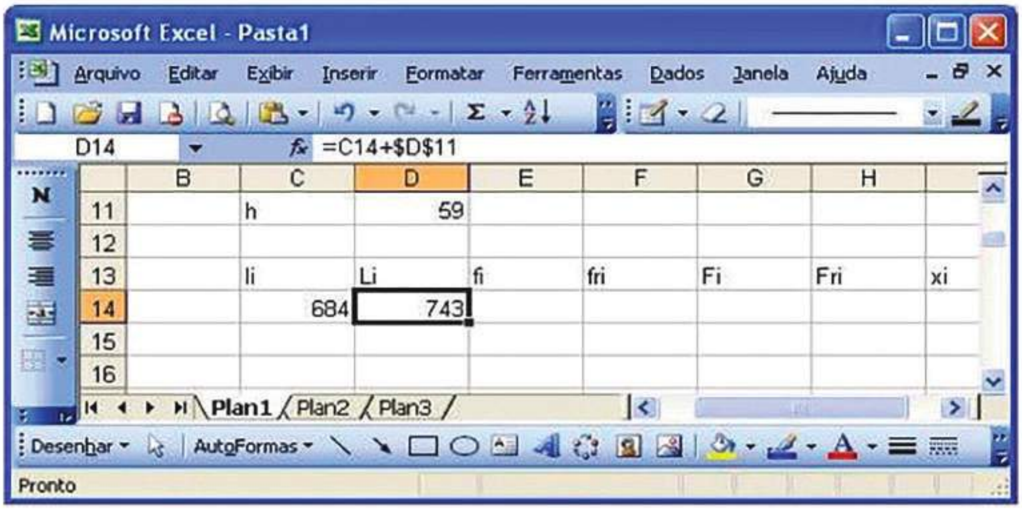


Ilustração: Mariana Netto

Passo 25: Na célula C15, digite o sinal de =, selecione a célula D14 e clique Enter. Aparecerá na célula C15 o limite inferior da 2ª classe.

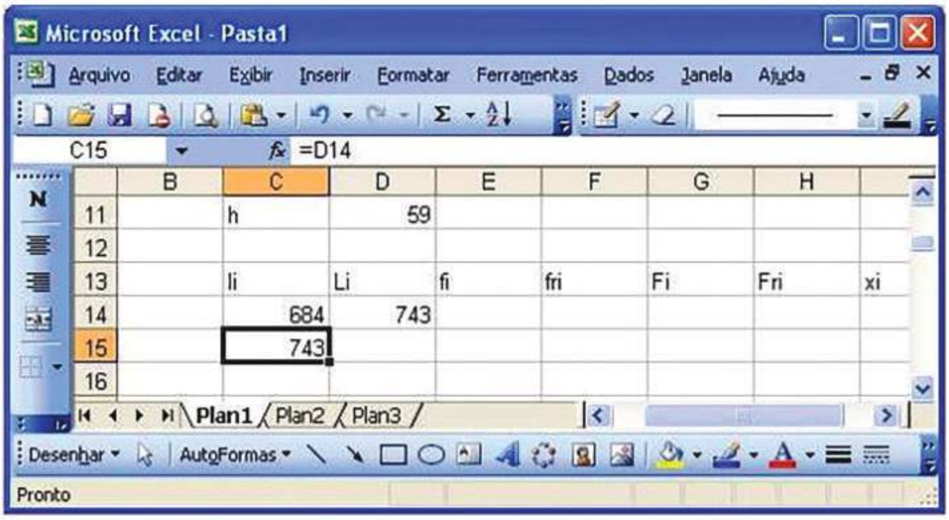


Ilustração: Mariana Netto

Passo 26: Selecione a célula C15, vá com o mouse no canto inferior direito até que apareça uma cruz e puxe até a célula C20.

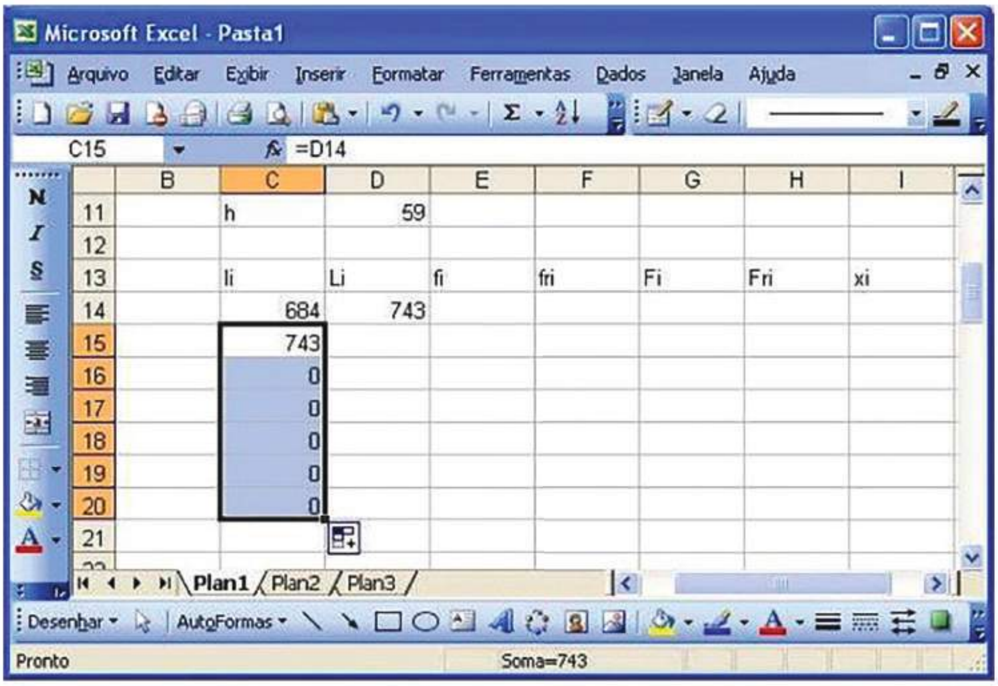


Ilustração: Mariana Netto

Passo 27: Selecione a célula D14, vá com o mouse no canto inferior direito até que apareça uma cruz e puxe até a célula D20. Os intervalos de classe estão prontos.

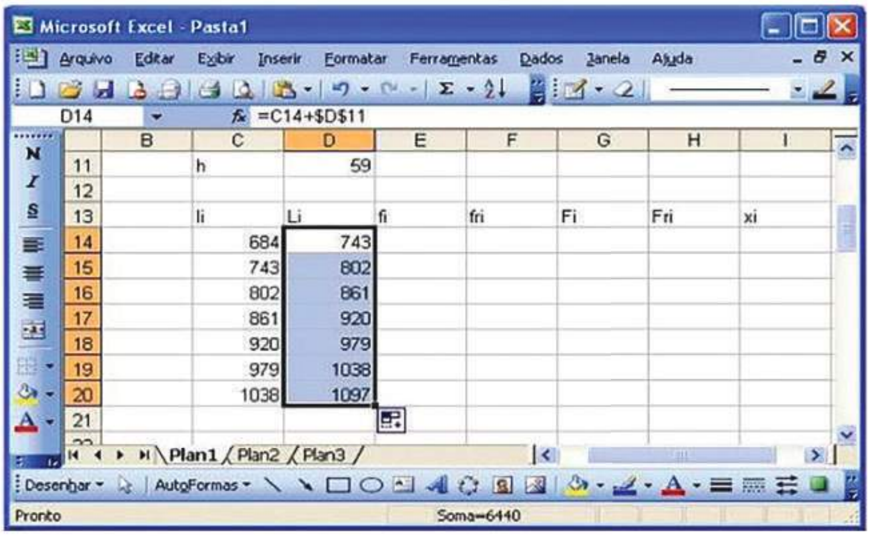


Ilustração: Mariana Netto

Observação: Note que você puxará até a célula C20 porque a tabela tem apenas 7 classes. Dependendo do número de elementos na amostra o número de classes pode aumentar ou diminuir.

Passo 28: Selecione a coluna que vai da célula E14 até E20 (coluna correspondente às frequências simples).

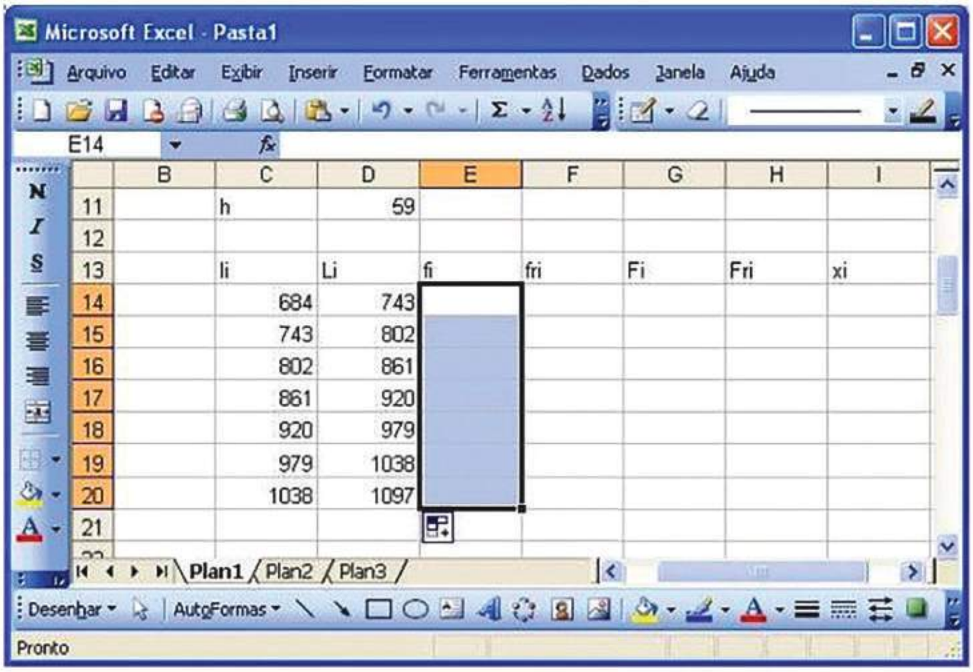


Ilustração: Mariana Netto

Passo 29: Com o mouse, selecione o ícone **Função**. A função escolhida deve ser **frequência**. Clique OK.

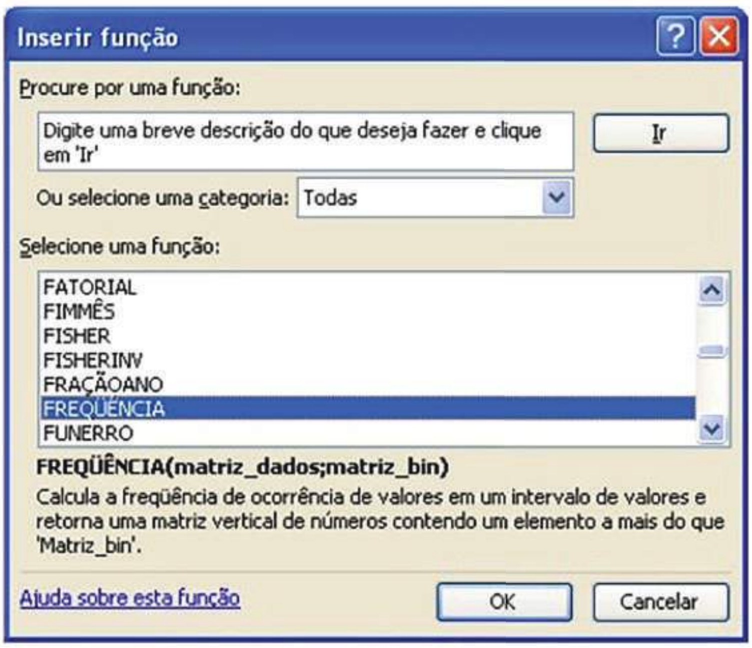


Ilustração: Mariana Netto

Passo 30: Em Matriz_dados selecione a coluna de A1 até A60 ou digite A1:A60. Em **Matriz_bin** selecione a coluna de D14 até D20 (coluna referente aos limites superiores de classe) ou digite D14:D20.

Não clique enter e sim CTRL SHIFT ENTER.

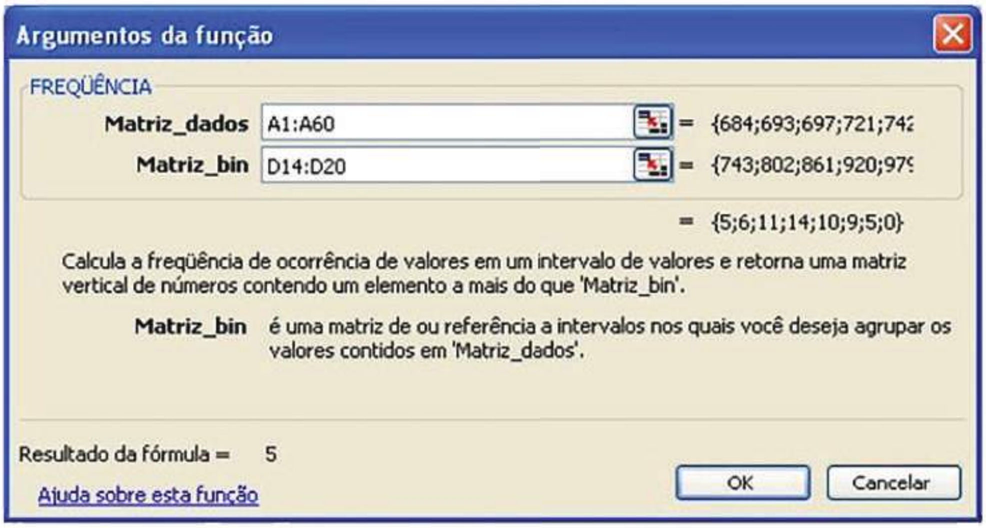


Ilustração: Mariana Netto

Passo 31: A coluna da frequência deve estar completa a partir daí.

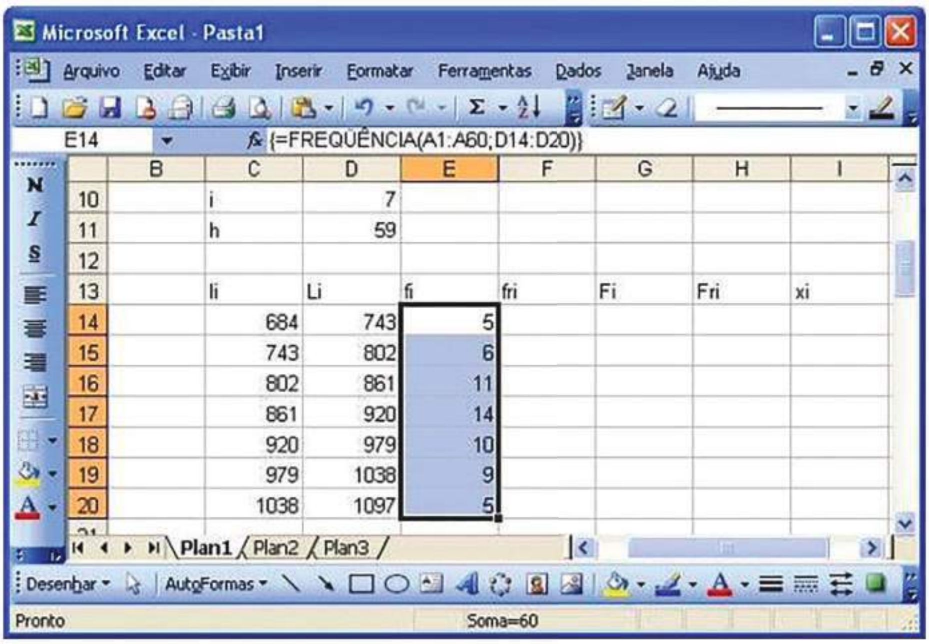


Ilustração: Mariana Netto

Passo 32: Selecione a célula E21 e clique no ícone **Auto Soma**.

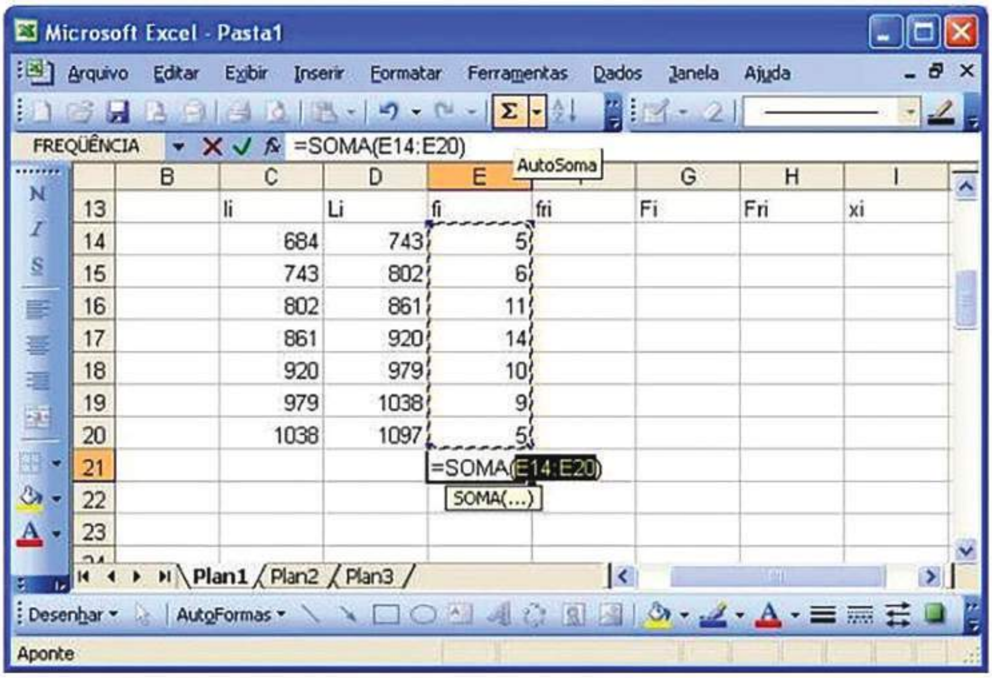


Ilustração: Mariana Netto

Passo 33: Clique *Enter* a aparecerá o valor 60 na célula E21, que corresponde ao somatório das frequências simples, ou seja, ao tamanho da amostra.

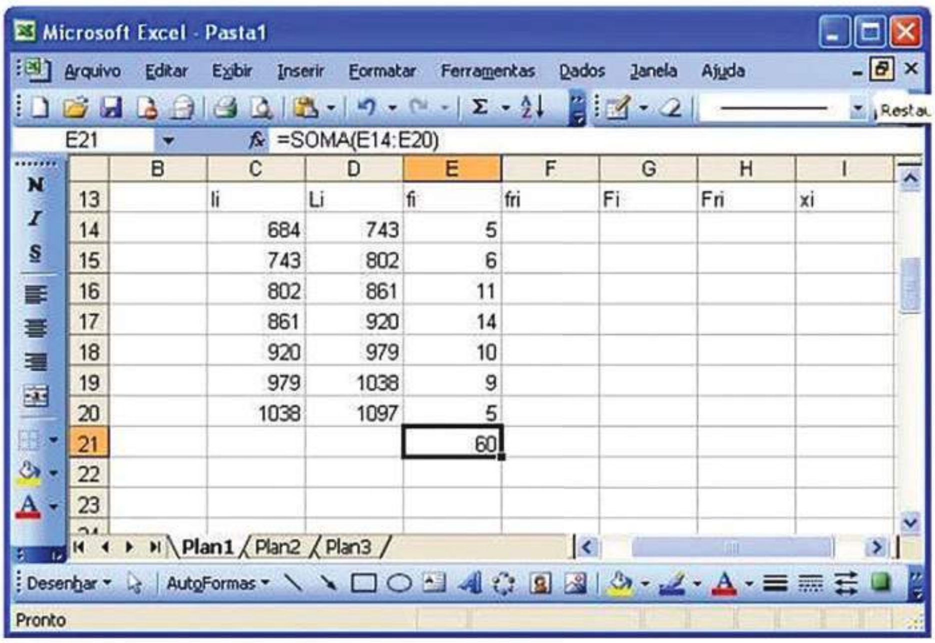


Ilustração: Mariana Netto

Passo 34: Na célula F14 digite o sinal de =, selecione a célula E14, digite /, selecione a célula E21 e clique em F4 (F4 congela a célula e funciona exatamente como o cifrão antes e depois da letra que representa a célula) no teclado, digite ***100**. Aperte a tecla *Enter*.

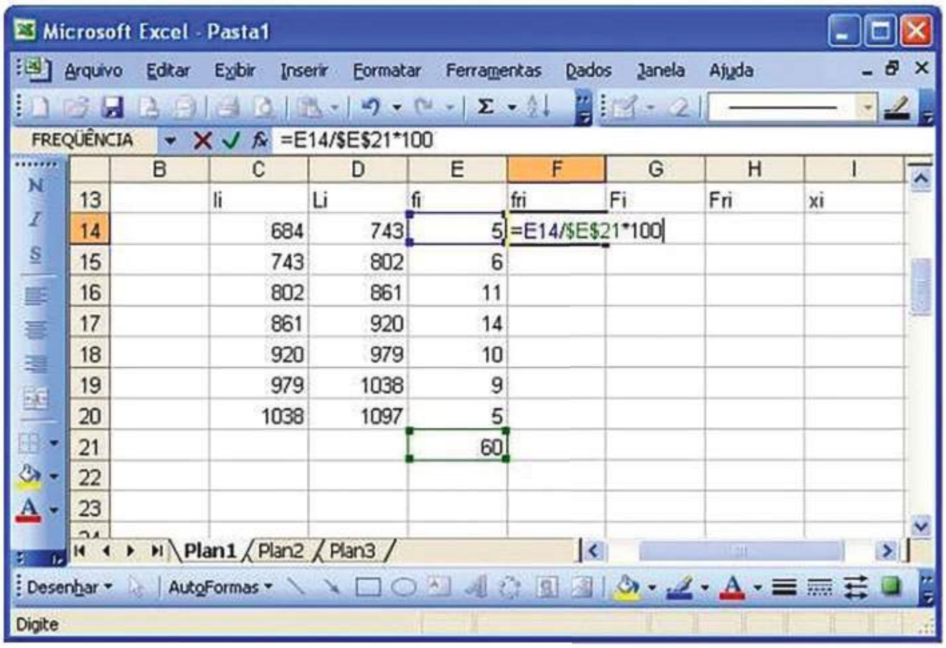


Ilustração: Mariana Netto

Passo 35: Aparecerá na célula F14 o resultado da frequência relativa da 1ª. classe.

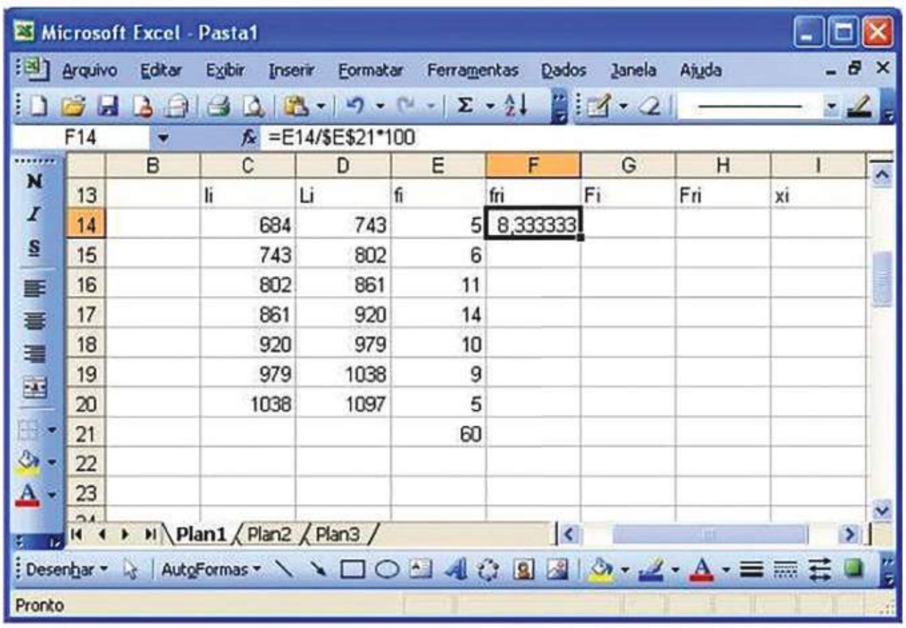


Ilustração: Mariana Netto

Passo 36: Selecione a célula F14, vá com o mouse no canto inferior direito até que apareça uma cruz e puxe até a célula F20. Está pronta a coluna correspondente às frequências relativas da tabela.

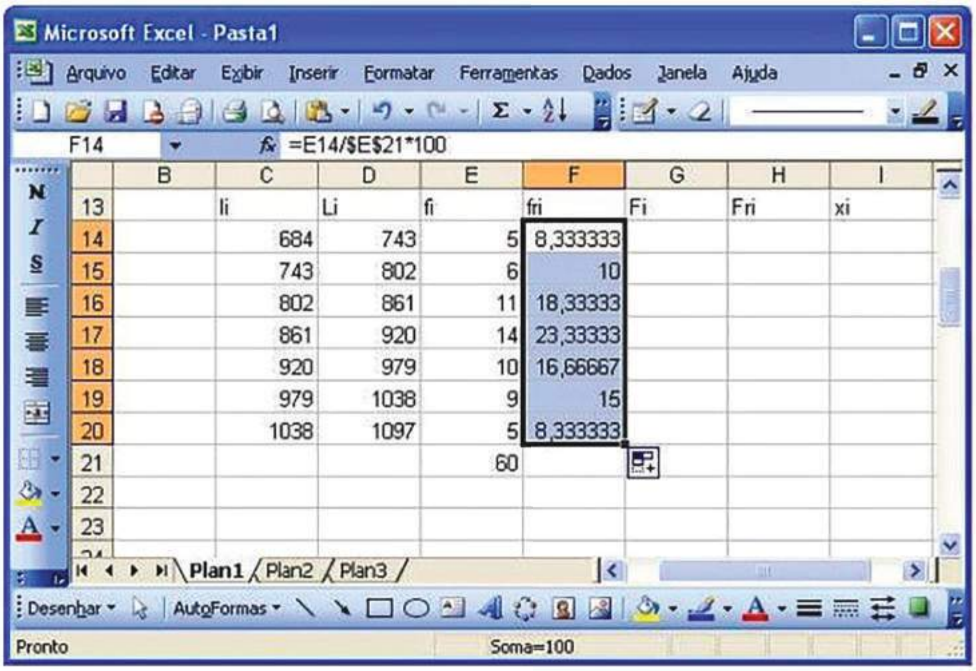


Ilustração: Mariana Netto

Passo 37: Selecione a célula F21 e clique no ícone **Auto Soma.**

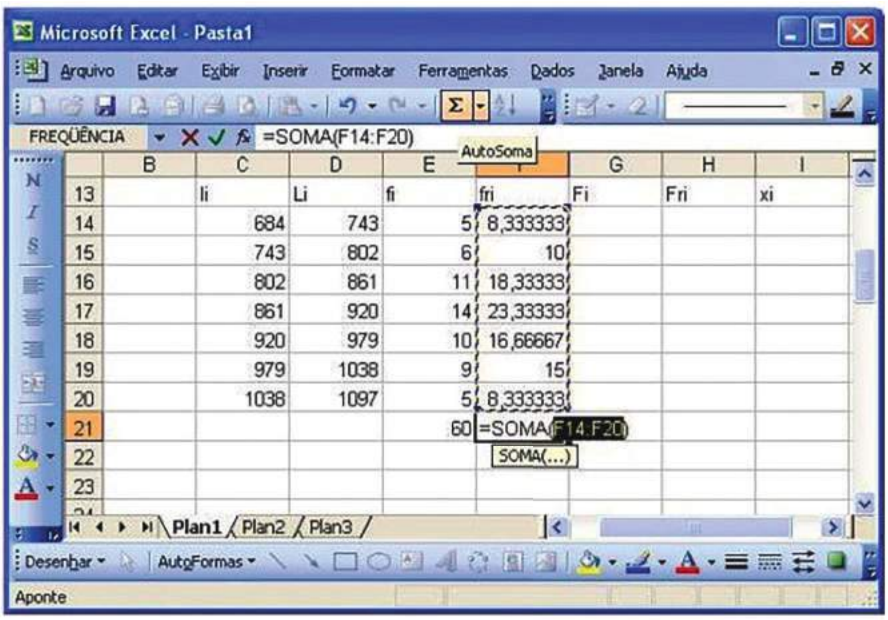


Ilustração: Mariana Netto

Passo 38: Aperte a tecla Enter a aparecerá o somatório das frequências relativas na célula F21. Esta frequência deve ser sempre igual a 100%.

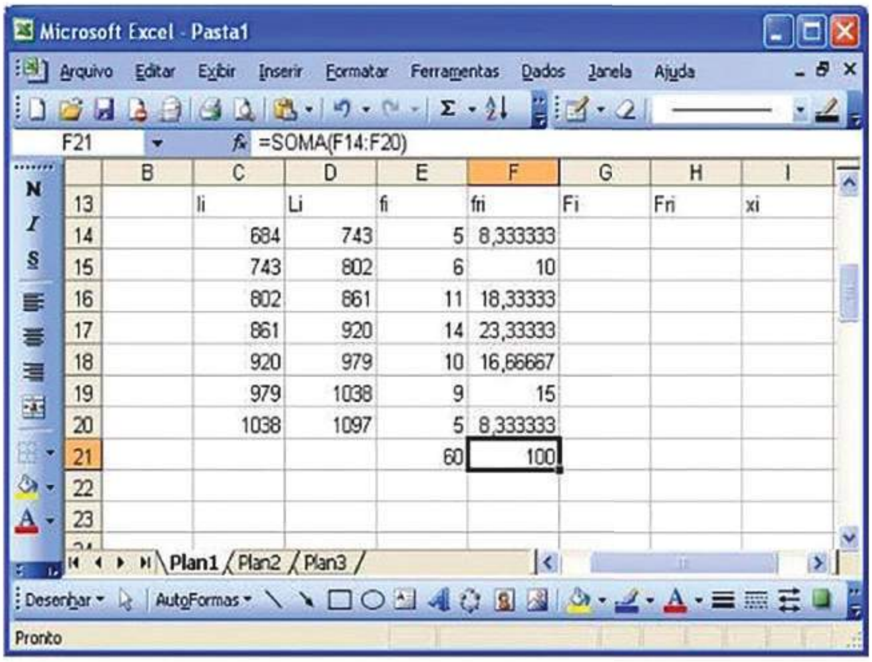


Ilustração: Mariana Netto

Passo 39: Selecione a coluna que vai de F4 até F21. Na barra de ferramentas clique em formatar, em seguida célula, número, 2 casas, OK.

Ilustração: Mariana Netto

Passo 40: Esse último passo padroniza a frequência relativa de todas as classes com 2 casas decimais.

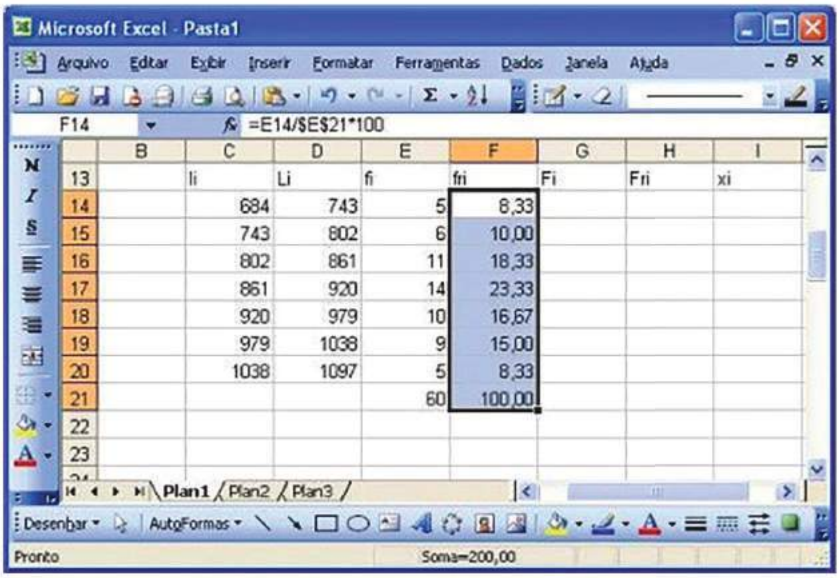


Ilustração: Mariana Netto

Passo 41: Na célula G14 digite o sinal de =, selecione a célula E14. Clique *Enter*.

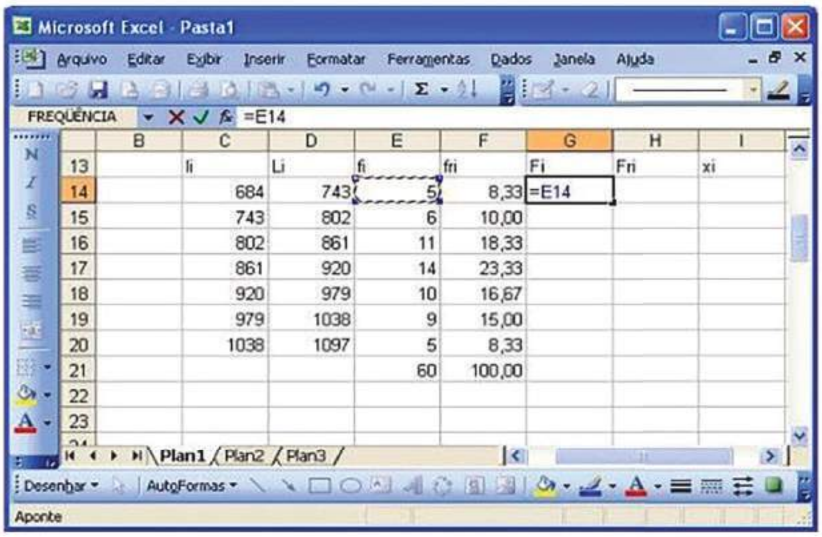


Ilustração: Mariana Netto

Passo 42: Na célula G14 aparecerá a frequência acumulada da 1ª classe.

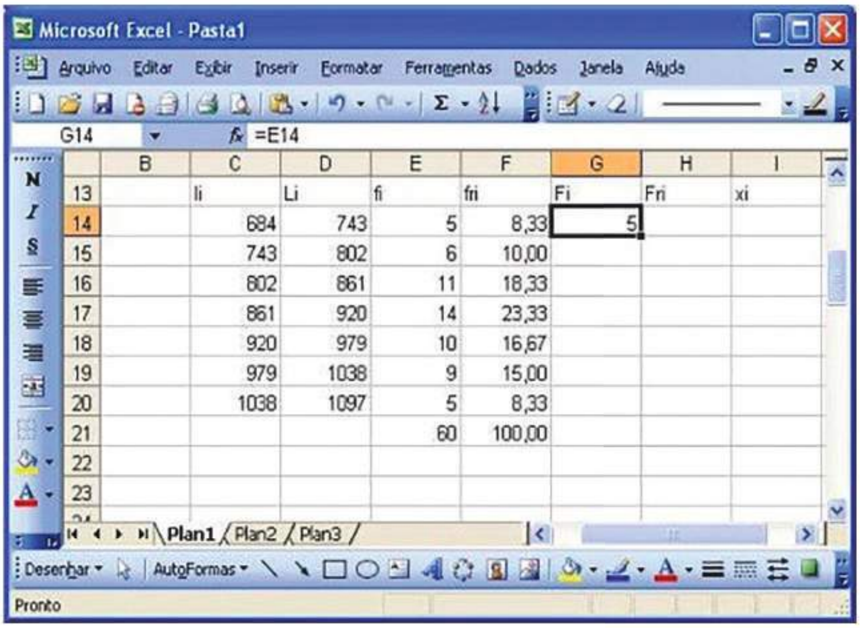


Ilustração: Mariana Netto

Passo 43: Na célula G15 digite igual (=), selecione a célula G14, digite + e selecione a célula E15. Clique *Enter*.

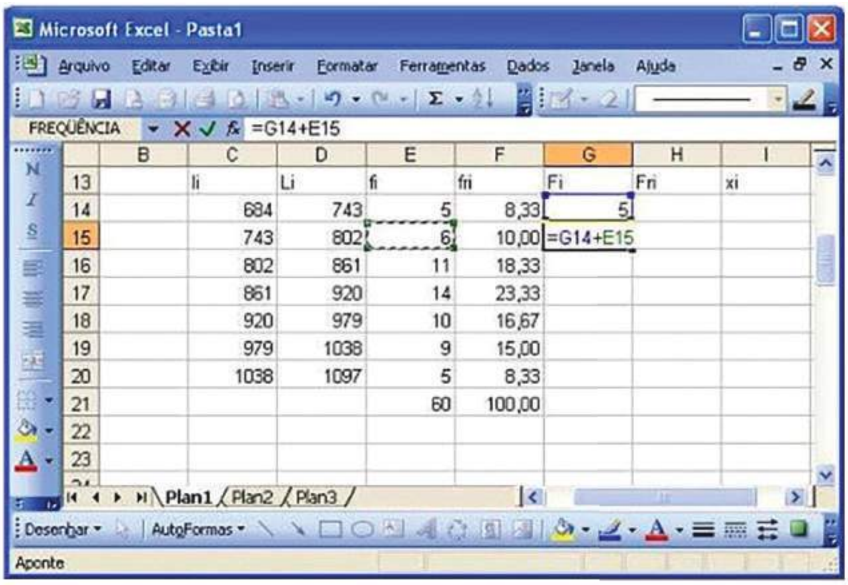


Ilustração: Mariana Netto

Passo 44: Aparecerá na célula G15 a frequência acumulada da 2ª classe.

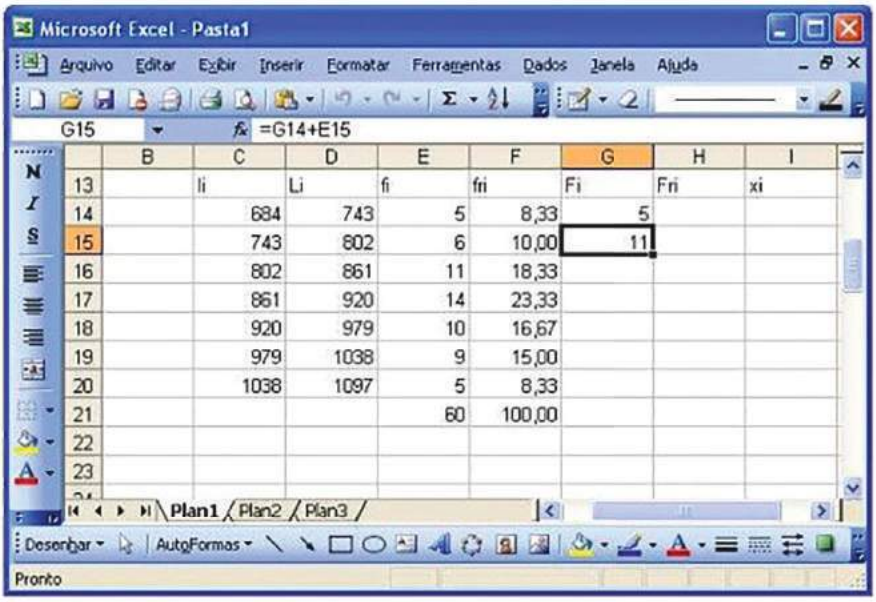


Ilustração: Mariana Netto

Passo 45: Selecione a célula G15, vá com o mouse até o canto inferior direito até que apareça uma cruz e puxe até a célula G20. Assim a coluna da frequência relativa fica completa.

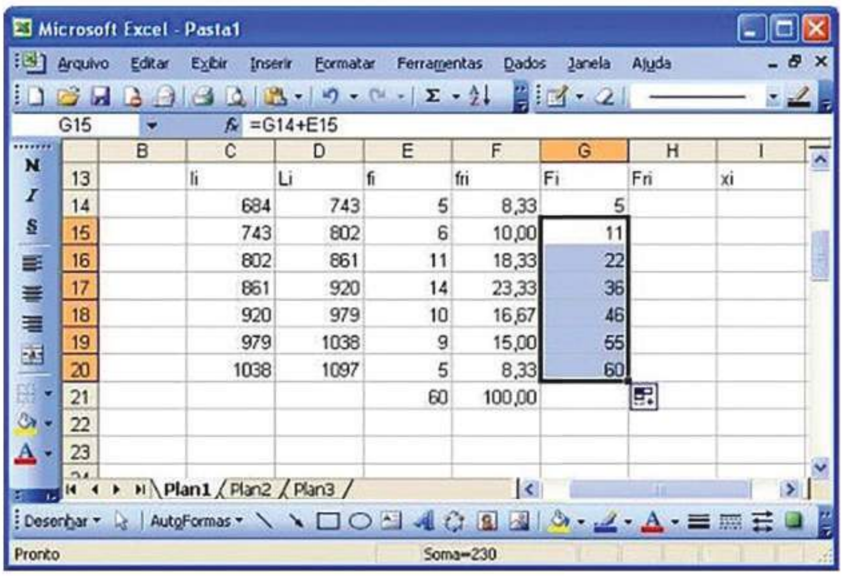
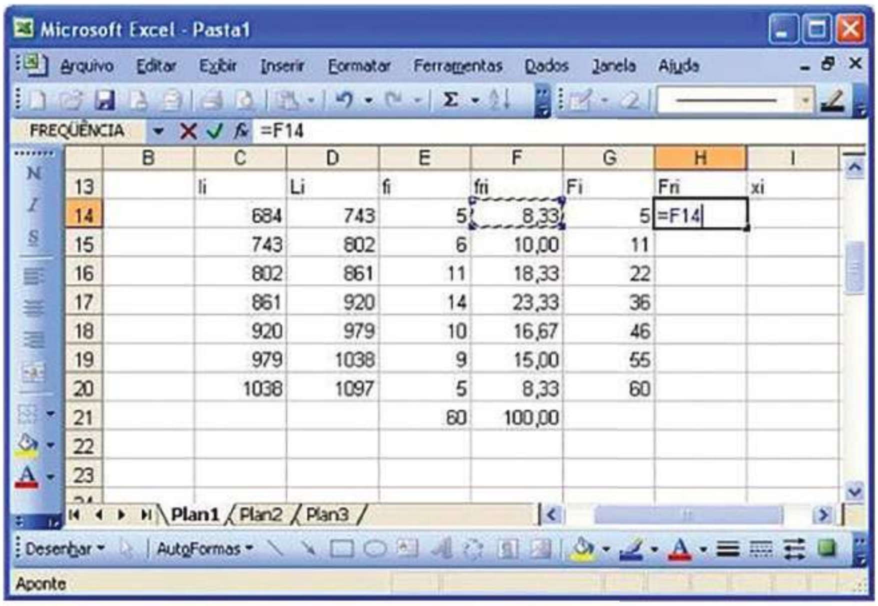


Ilustração: Mariana Netto

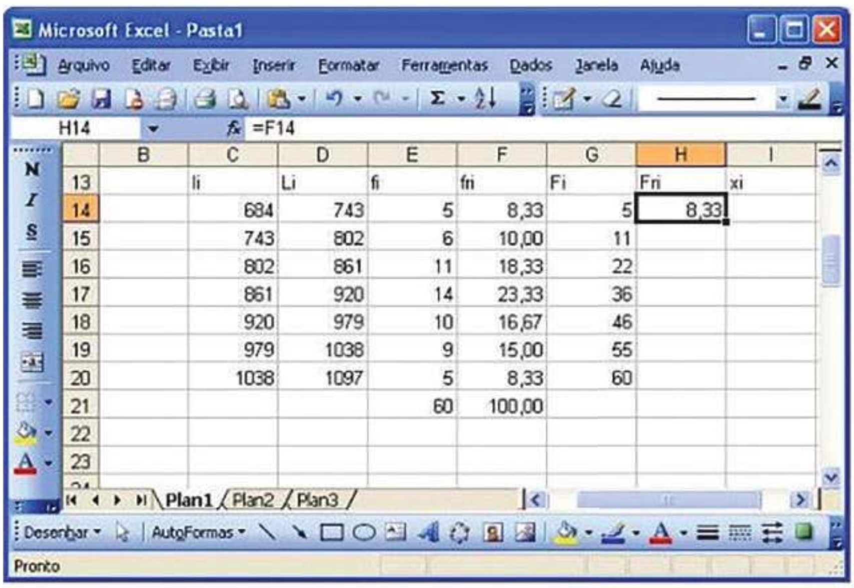
Passo 46: Na célula H14 digite o sinal de =, selecione a célula F14. Clique *Enter*.



	B	C	D	E	F	G	H	I
13		li	Li	fi	fri	Fi	Fri	xi
14		684	743	5	8,33	5	=F14	
15		743	802	6	10,00	11		
16		802	861	11	18,33	22		
17		861	920	14	23,33	36		
18		920	979	10	16,67	46		
19		979	1038	9	15,00	55		
20		1038	1097	5	8,33	60		
21				60	100,00			
22								
23								

Ilustração: Mariana Netto

Passo 47: Aparecerá a frequência relativa acumulada da 1ª. classe na célula H14.



	B	C	D	E	F	G	H	I
13		li	Li	fi	fri	Fi	Fri	xi
14		684	743	5	8,33	5	8,33	
15		743	802	6	10,00	11		
16		802	861	11	18,33	22		
17		861	920	14	23,33	36		
18		920	979	10	16,67	46		
19		979	1038	9	15,00	55		
20		1038	1097	5	8,33	60		
21				60	100,00			
22								
23								

Ilustração: Mariana Netto

Passo 48: Na célula H15 digite o sinal de =, selecione a célula H14, digite o sinal de + e selecione a célula F15. Clique *Enter*.

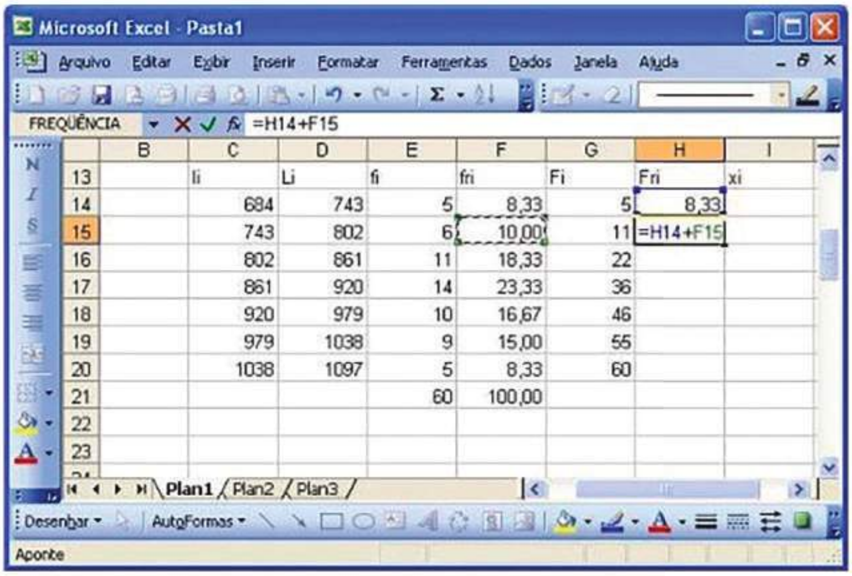


Ilustração: Mariana Netto

Passo 49: Aparecerá a frequência relativa acumulada da 2ª. classe na célula H15.

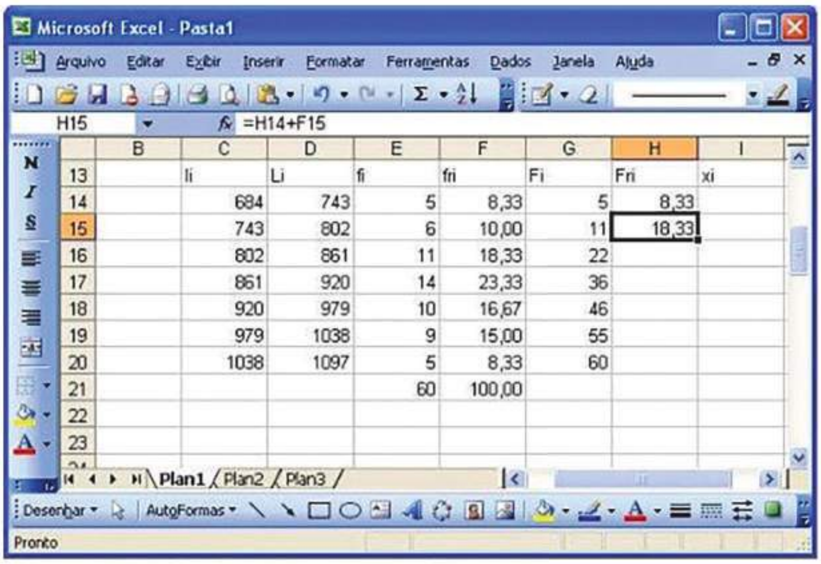


Ilustração: Mariana Netto

Passo 50: Selecione a célula H15, vá com o mouse até o canto inferior direito até que apareça uma cruz e puxe até a célula H20. Assim a coluna da frequência relativa acumulada está completa.

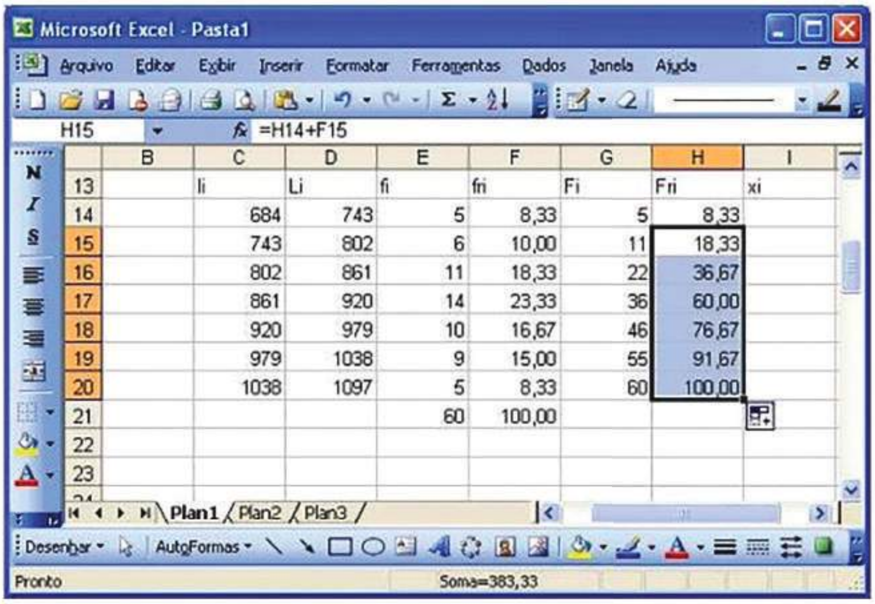


Ilustração: Mariana Netto

Passo 51: Na célula I14 digite o sinal de = e clique no ícone **função**. Selecione a função **média**. Clique em OK.

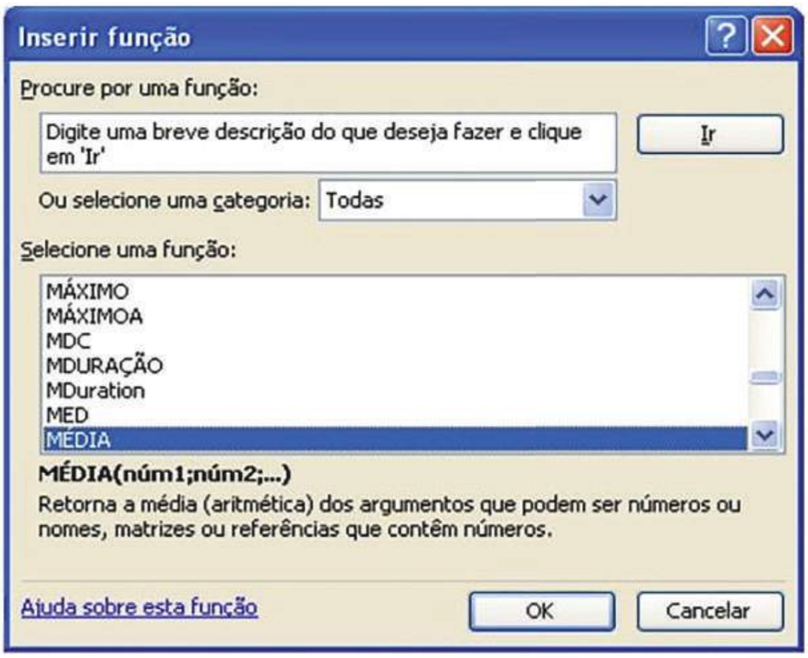


Ilustração: Mariana Netto

Passo 52: Em Núm 1 selecione as células C14 e D14 ou digite C14: D14.

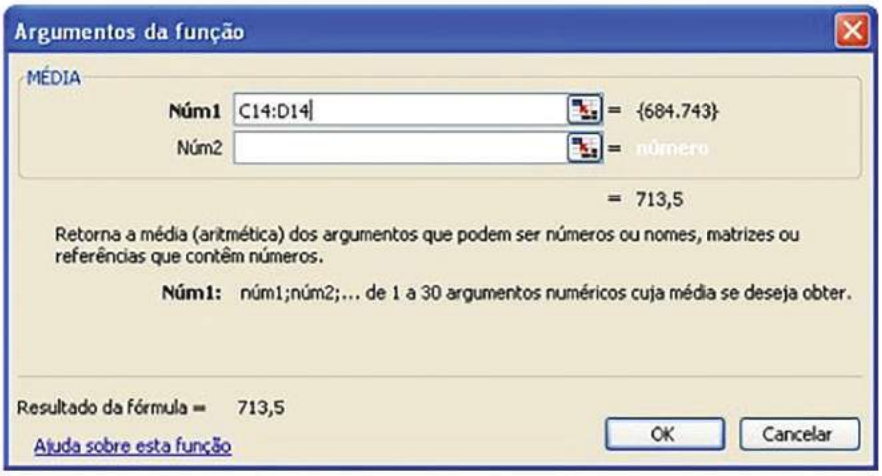


Ilustração: Mariana Netto

Passo 53: Clique *Enter* e na célula I14 aparecerá o ponto médio da 1ª classe.

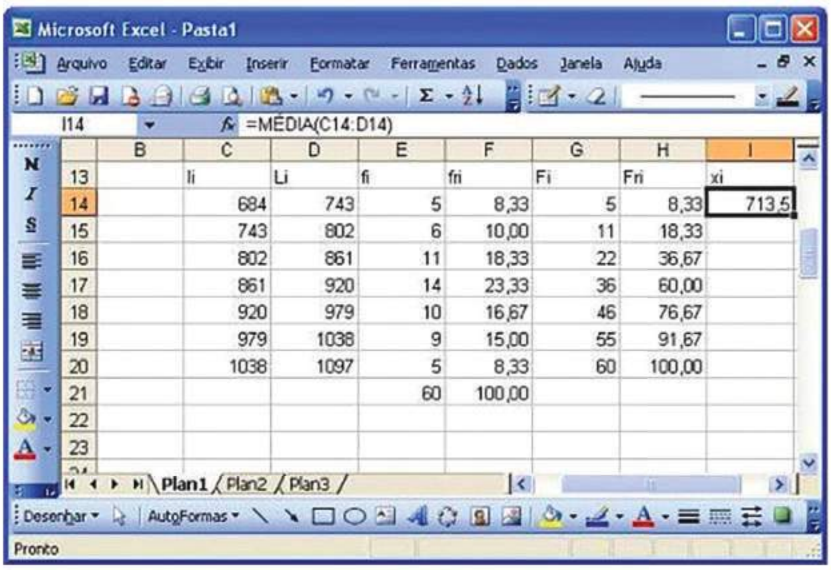


Ilustração: Mariana Netto

Passo 54: Selecione a célula I14, vá com o mouse ao canto inferior direito até aparecer uma cruz e puxe até a célula I20.

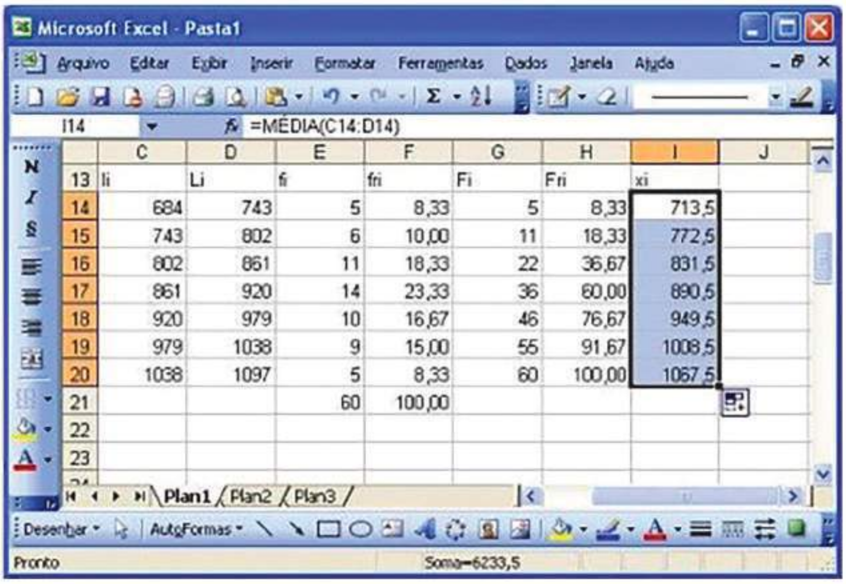


Ilustração: Mariana Netto

A Tabela de Distribuição de frequências está pronta. Agora que você já conhece os passos para construí-la no Excel, é só praticar.

2.7 Medidas de Tendência Central (ou de Posição)

Definição 2.7.1: *As medidas de posição mais importantes* são as medidas de ***tendência central****, que recebem tal denominação pelo fato de os dados observados tendem, em geral, a se agrupar em torno dos valores centrais.*

Dentre as medidas de tendência central, destacamos:

- a média aritmética;
- a mediana;
- a moda.

As outras medidas de posição são as **separatrizes**, que englobam:

a própria mediana;

os quartis;

os percentis.

2.7.1 Para dados não-agrupados (Quando os dados não estiverem na forma de distribuição de frequência)

Definição 2.7.1.1: Média aritmética ($\bar{X}$): é o *quociente da divisão da soma dos valores (dados, observações) da variável pelo número deles:*

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

sendo:

$\bar{X}$ = a média aritmética;

x_i = os valores da variável;

n = o número de valores.

Definição 2.7.1.2: Moda (Mo): *Denominamos **moda** de um conjunto de dados o valor (ou valores) que ocorre com maior frequência.*

Por ex..: o salário modal dos empregados de uma indústria é o salário mais comum, isto é, o salário recebido pelo maior número de empregados dessa indústria.

Definição 2.7.1.3: Mediana (Md): *A **mediana*** é outra med*ida de posição definida como o número que se encontra no centro de uma série de números, estando estes dispostos segundo uma ordem. Em outras palavras, a mediana de um conjunto de valores, <u>ordenados</u>, é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos.*

Observação: Se o nº de elementos for ímpar, então a mediana será exatamente o valor "do meio", ou seja, será o elemento que está na posição . $\frac{n+1}{2}$

- Se o nº de elementos for par, então a mediana será exatamente a média "dos dois valores do meio", ou seja, será a média entre os elementos $\frac{n}{2}$ e $\frac{n}{2}+1$.

Exemplo: Sabendo-se que a produção leiteira diária da vaca A, durante uma semana, foi de: 10, 14, 13, 15, 16, 18 e 12 litros, pergunta-se: Encontre a média, a moda e a mediana para a produção diária de leite desta vaca.

Temos que a média é dada por:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{10 + 14 + 13 + 15 + 16 + 18 + 12}{7} = 14$$

Portanto, a vaca A produz uma média de 14 litros de leite por dia.

Observação.: a média pode ser um número diferente de todos os valores da amostra que ela representa.

Moda: Como não existe um valor que aparece com maior frequência que os outros, não há valor de moda para este exemplo.

Observação: Um conjunto de dados que não tem moda, chama-se o conjunto de dados amodal.

Mediana: Ordenando os dados temos:

10 12 13 14 15 16 18

Como o número de elementos é ímpar, então o valor mediano é o valor central dos dados, ou seja, $\frac{n+1}{2} = \frac{7+1}{2} = 4^{\underline{o}}$ elemento, ou seja, 14 litros de leite por dia.

2.7.2 Para dados agrupados (Quando os dados estiverem na forma de distribuição de frequência)

Cálculo da Média: Quando os dados estiverem agrupados, ou seja, na forma de distribuição de frequências a forma de calcular a média aritmética muda um pouco. Nestes casos, como as frequências são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva a calcular a **média aritmética ponderada,** dada pela fórmula:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i f_i}{n}$$

Em que nesse caso, temos que

x_i= ponto médio da classe ;

f_i= frequência simples absoluta da classe ;

n= Quantidade de classes

Exemplo: Calcule a média da seguinte tabela de intervalos de classes.

Custos R$ Classes	X_i	F_i	X_iF_i
450 |- 550	500	8	
550 |- 650	600	10	
650 |- 750	700	11	
750 |- 850	800	16	
850 |- 950	900	13	
950 |- 1050	1000	5	
1050 |- 1150	1100	1	
Total		**64**	

Ilustração: Sofia Guimarães

Temos nesse caso que a média é dada por:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i f_i}{n} = \frac{(500.8) + (600.10) + (700.11) + \ldots + (1000.5) + (1100.1)}{8 + 10 + 11 + 16 + 13 + 5 + 1} = 754{,}687$$

Observação: O valor da média é sempre influenciado por valores extremos, ou seja, caso haja um valor muito acima dos demais, isso fará com a média seja aumentada e caso o valor seja muito inferior aos demais, o valor da média será consequentemente baixo.

No entanto, podemos verificar, às vezes, se um cálculo da média está incorreto ou não, apenas olhando para os banco de dados, ou para a tabela de intervalos de classes.

Exemplo: Os dados abaixo se referem às variações (%) de crescimento da Empreendedora XXX do município da Microrregião Oeste Baiana (Dados Fictícios) ao longo de 43 meses, sempre em relação ao mês anterior.

Variação	Nº de Meses
9,9 |- 15,9	8
15,9 |- 21,9	12
21,9 |- 27,9	13
27,9 |- 33,9	5
33,9 |- 39,9	3
39,9 |- 45,9	2
Total	**43**

Ilustração: Sofia Guimarães

O responsável pelos resultados forneceu a média da variação de crescimento que foi de 27,95%. Você concorda com o resultado?

Percebe-se que essa informação é falsa, sem precisar necessariamente fazer os cálculos. Pois o valor fornecido da média está em classe de valores que tem apenas 5 informações. Nota-se que a grande maioria dos dados está concentrada entre as classes 15,9 a 27,9. Temos nessas classes um total de 25 informações dos dados, que é praticamente 58% da quantidade total de informações.

Portanto, é muito importante observar os dados e não somente o resultado em si, para ter uma determinada conclusão a respeito de alguma informação.

Algumas Propriedades da Média

1ª propriedade: Somando-se (ou subtraindo-se) uma constante (c) a todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante.

Exemplo continuação: Voltando ao exemplo da produção leiteira diária da vaca A. Se no exemplo original somarmos a constante **2** a cada um dos valores da variável temos:

Os valores originais são: 10, 14, 13, 15, 16, 18 e 12. Então, a média dos novos valores somados com a constante 2 será:

$Y = (12+16+15+17+18+20+14) / 7 = 16$ litros ou

$Y = \bar{X} + 2 = 14 + 2 = 16$ litros

2ª propriedade: Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada (ou dividida) por essa constante.

Exemplo continuação: Se no exemplo original multiplicarmos a constante **3** a cada um dos valores da variável temos:

Y = (30+42+39+45+48+54+36) / 7 = 42 litros ou

Y = $\bar{X}$ x 3 = 14 x 3 = 42 litros

Cálculo da Moda: A classe que apresenta a maior frequência é denominada **classe modal.** Pela ***Definição 2.7.1.2***, podemos afirmar que a moda, neste caso, é o valor dominante que está compreendido entre os limites da classe modal. O método mais simples para o cálculo da moda consiste em tomar o ponto médio da classe modal. Damos a esse valor a denominação de **moda bruta**.

$$M_o = \frac{l_i + l_s}{2}$$

onde l_i = limite inferior da classe modal e l_s = limite superior da classe modal.

Exemplo continuação: Em relação ao exemplo das variações (%) de crescimento da Empreendedora XXX do município da Microrregião Oeste Baiana.

Variação	Nº de Meses
9,9 |- 15,9	8
15,9 |- 21,9	12
21,9 |- 27,9	13
27,9 |- 33,9	5
33,9 |- 39,9	3
39,9 |- 45,9	2
Total	**43**

Ilustração: Sofia Guimarães

Temos nesse caso que a classe modal é a classe 21,9 a 27,9, pois é a de maior frequência. Nesse caso, a moda bruta é dada por:

$$M_o = \frac{l_i + l_s}{2} = \frac{21{,}9 + 27{,}9}{2} = 24{,}9$$

Ou seja, a moda das variações é de 24,9%.

$$M_o = l_{mo} + \frac{f_{mo} - f_{ant}}{2f_{mo} - (f_{ant} + f_{post})} h$$

Para o cálculo da moda através do **Método de Czuber**, identifica-se o maior elemento em (frequência simples) na tabela. Identificamos esta classe como **classe modal** e assim aplicamos a fórmula chamada de método de Czuber para calcular a moda que é dada por:

onde

- l_{mo} = limite inferior da classe modal;
- f_{ant} = frequência simples anterior da classe modal;
- f_{mo} = frequência simples da classe modal;
- f_{post} = frequência posterior da classe modal;
- h = Amplitude de intervalo da classe modal;

Exemplo continuação: Calcular a moda na tabela abaixo.

Variação	Nº de Meses
9,9 |- 15,9	8
15,9 |- 21,9	12
21,9 |- 27,9	13
27,9 |- 33,9	5
33,9 |- 39,9	3
39,9 |- 45,9	2
Total	**43**

Ilustração: Sofia Guimarães

\Para calcular a moda na série identificamos o maior elemento na frequência (f_i) da série que neste caso é o número 13 que se encontra na terceira classe. Identificamos esta classe como **classe modal** e assim calculamos a moda aplicando o método de Czuber, utilizando como referência a classe 3.

$$M_o = l_{mo} + \frac{f_{mo} - f_{ant}}{2f_{mo} - (f_{ant} + f_{post})}h = 21{,}9 + \frac{13 - 12}{2.13 - (12 + 5)}.6 = 22{,}56$$

Assim, a variação que aparece com maior frequência é 22,56%.

Cálculo da Mediana: Para determinar a mediana dos dados agrupados em tabelas com intervalos de classes, devemos seguir os seguintes passos:

a) Determinar as frequências acumuladas absolutas.

b) Calcular a ordem do elemento mediano **n/2** ou equivalentemente **n x P**, onde **P=50%** (0,50)

c) Identificar a classe correspondente à frequência acumulada imediatamente superior a **n/2**, que é a ***classe que contém a mediana***, para em seguida aplicar a fórmula:

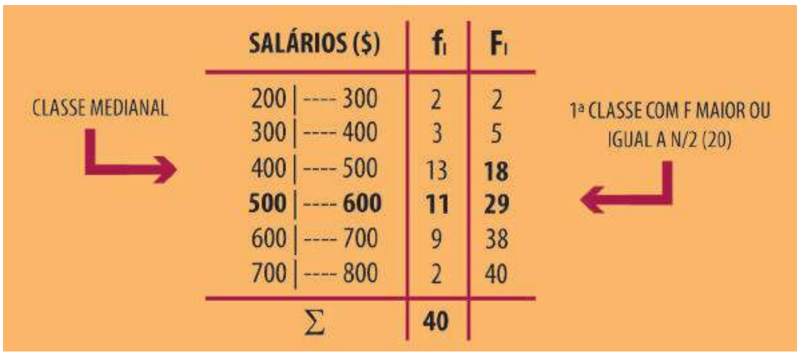
$$M_d = l_{med} + \frac{(n/2 - f_{ant})}{f_{med}} h$$

onde

- l_{med} = limite inferior da ***classe mediana***
- f_{med} = frequência absoluta da ***classe mediana***
- f_{ant} = frequência acumulada da classe anterior à ***classe mediana***
- h = amplitude do intervalo da ***classe mediana***

Observação: Não é necessário identificar se **n** é ***<u>par</u>*** ou <u>ímpar</u>.

Exemplo: Vamos calcular a mediana da seguinte tabela com dados salariais.



SALÁRIOS ($)	f_i	F_i
200 |---- 300	2	2
300 |---- 400	3	5
400 |---- 500	13	**18**
500 |---- 600	**11**	**29**
600 |---- 700	9	38
700 |---- 800	2	40
Σ	**40**	

Ilustração: Sofia Guimarães

Identificando componentes da fórmula:

(**n/2**) ou (50% de n)=**20**; Classe Mediana=**4ª**; l_{med}=**500**; f_{med}=**11**; h=**100**; f_{ant}=**18.**

$$M_d = l_{med} + \frac{(n/2 - f_{ant})}{f_{med}} h = 500 + \frac{(20 - 18)}{11}.100 = 500 + \frac{200}{11} = 500 + 18{,}18 = 518{,}18$$

Portanto, temos que 50% das pessoas, desta amostra, tem salário abaixo de R$ 518,18 ou 50% das pessoas, desta amostra, ganham acima de R$ 518,18.

Análise Comparativa das Medidas de Tendência Central

Analisando as definições e formas de cálculo das três medidas estatísticas **Média**, **Moda** e **Mediana**, podemos destacar as vantagens e desvantagens de cada uma destas medidas. A escolha sobre qual (ou quais) destas medidas utilizar para mostrar (ou enfatizar) a tendência dos valores de se agruparem em torno de valores centrais, ou mesmo para nos dar uma idéia sobre como o conjunto de valores está posicionado em seu eixo de valores, depende fundamentalmente do tipo de análise que se pretende desenvolver na pesquisa realizada.

O quadro a seguir reúne as principais características de cada uma das três medidas estatísticas estudadas.

Principais Características das Medidas de Tendência Central

MODA

VANTAGENS	DESVANTAGENS
Fácil de calcular. Não é afetada pelos dados extremos. Pode ser aplicada em qualquer escala: nominal, ordinal, intervar e razão (ou proporcional)	Pode estar afastada do centro dos dados. Difícil de incluir em funções matemáticas. A amostra pode ter mais de uma moda e algumas amostras podem não ter moda. Não utiliza todos os dados da amostra.

MEDIANA

VANTAGENS	DESVANTAGENS
Fácil de calcular. Não é afetada pelos dados extremos. É um valor único. Aplicável em escalas ordinal, intervalar e razão.	Difícil de incluir em funções matemáticas. Não utiliza todos os dados da amostra.

MÉDIA

VANTAGENS	DESVANTAGENS
Fácil de compreender e aplicar. Utiliza todos os dados da amostra. É um valor único. Aplicável em escalas intervalar e razão.	É afetada pelos dados externos da amostra. Requer o conhecimento de todos os dados da amostra.

Ilustração: Sofia Guimarães

Tipos de Simetria: Posição Relativa das Medidas de Tendência Central

Embora o conceito de simetria não tenha sido introduzido ainda, é muito fácil de ser entendido e será necessário para verificarmos como a forma de uma distribuição de frequências influencia os valores das três medidas de posição.

Numa distribuição de frequências **Simétrica** em relação à **Média**, os valores da **Média, Moda** e **Mediana** ***coincidem (ou são aproximadamente iguais)***, e não é difícil de visualizarmos este fato. Uma distribuição é **Assimétrica Positiva** quando os valores de seu ***extremo superior estão mais afastados*** da **Média** que seus valores do extremo inferior, resultando num contorno que apresenta uma ***cauda mais longa*** na direção do ***extremo superior dos valores***.

No item anterior foi enfatizado que a **Média** é a única medida de posição que é afetada pelos valores extremos. A **Assimetria Positiva** afeta a **Média** '***puxando-a***' em sua direção, ou seja, ***na direção da cauda mais longa*** que é ***a superior***. Observe o que acontece com as três medidas de posição neste caso.

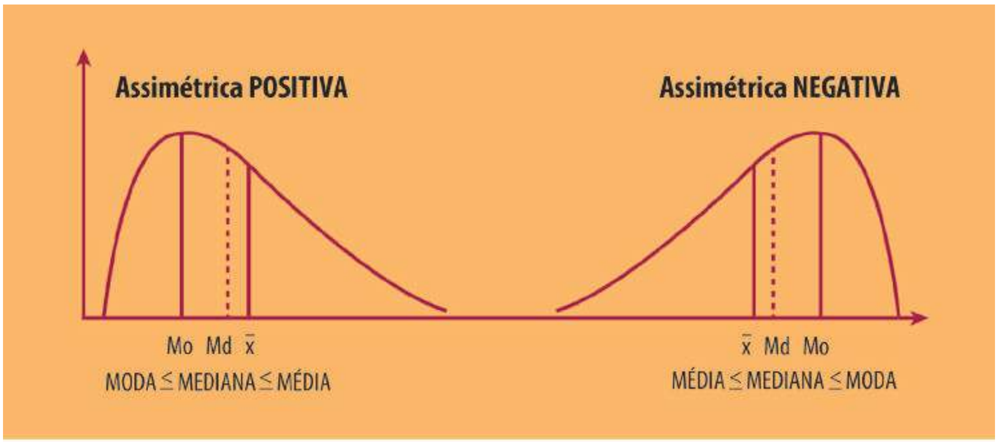


Ilustração: Sofia Guimarães

Reciprocamente, uma **Assimetria Negativa** apresenta uma cauda mais longa na direção dos valores extremos inferiores, o que afeta a **Média** '***puxando-a***' também na direção dos valores ***extremos inferiores***.

A **Moda** é facilmente identificável em ambos os casos: é a ***ordenada*** correspondente ao ***pico*** (**máximo**) ***de frequência***. Diante da constatação de que a cauda mais longa traz a **Média** para seu lado, é fácil percebermos que a cauda mais longa acaba por puxar para seu lado a **Mediana** também, quando confrontada com a **Moda**. Assim, mesmo sem calcular os valores das três medidas, dependendo da forma (gráfico) da DF, podemos estabelecer as relações acima entre as três medidas.

2.7.3 Medidas de Separatrizes

Medidas Separatrizes são na Estatística também são chamadas de Medidas de Posição.

Definição 2.8.1: *Separatrizes são aquelas medidas que "separam" ou que dividem o conjunto em um certo número de partes iguais.*

No caso da Mediana, vimos que ela divide o conjunto em duas metades. Já o Quartil, separa o conjunto em quatro partes iguais; o Decil, em dez partes e, finalmente, o Centil (ou Percentil), em cem partes iguais.

Recordando disso, lembraremos também que aprendemos uma relação importantíssima entre as quatro Medidas Separatrizes. Na verdade é uma relação até visual, que não precisamos fazer esforço para "decorar", bastando traçar uma reta (que representará o conjunto), e depois fazer as divisões, exatamente como mostramos no Ponto 15 e transcrevemos abaixo:

```
!------------------!------------------!
                   Md

!---------!---------!---------!---------!
          Q1        Q2         Q3

!----!----!-----!----!-----!-----!----!-----!----!----!
          D1 D2 D3 D4  D5  D6  D7  D8  D9

!----!----!-----!-----!-----!-----!-----!-----!----!----!
  C10 C20 C30 C40 C50 C60 C70 C80 C90
```

Daí, concluímos sem maiores dificuldades que:

$$Md = Q2 = D5 = C50$$

Quartis São Separatrizes que dividem um conjunto de dados em quatro partes iguais

0%	25%	50%	75%	100%
	Q1	Q2	Q3	

Primeiro Quartil - Q1

Separatriz que divide a distribuição em duas partes, tal que 25% dos valores sejam menores que ele e 75% maiores que ele.

Segundo Quartil - Q2

O segundo quartil coincide exatamente com a mediana. É o valor que divide a distribuição em exatamente metade dos elementos.

Terceiro Quartil - Q3

Valor que deixa 75% dos valores à sua esquerda e os 25% restante à sua direita

Determinação do Quartil

Já sabemos que para dividir um conjunto em quatro partes iguais, precisamos marcar três pontos apenas (como vimos no desenho acima!). Portanto, já sabemos que existem três quartis, os quais designaremos por Q1 (primeiro quartil), Q2 (segundo quartil) e Q3 (terceiro quartil).

Quando estudamos a Mediana, vimos que as questões que exigiam o cálculo desta medida costumavam dizer apenas algo como "determine o valor da Mediana deste conjunto" (e só!). Isso porque existe somente uma Mediana. Porém, em se tratando do Quartil, o enunciado jamais poderia dizer apenas "determine o valor do Quartil". Se assim o fizesse, ficaria no ar a pergunta: "Qual deles?". Se existem três quartis, uma questão de prova teria, logicamente, que explicitar qual deles está exigindo.

Para calcular o primeiro quartil, temos antes que determinar qual será a Classe do Primeiro Quartil. Agora, sabemos que o Quartil divide o conjunto em quatro partes. Portanto, a conta que faremos (para o primeiro quartil) é a seguinte:

$$(n/4)$$

Para fazer esta conta, também não nos preocuparemos se n é um valor par ou ímpar. Feita esta continha, passaremos a comparar seu resultado com os valores da frequência acumulada (f_{ac}), exatamente da mesma forma que fizemos para achar a Classe Mediana! A pergunta, agora adaptada ao Quartil, será a seguinte:

Esta f_{ac} é maior ou igual a (n/4)?

Enquanto a resposta for negativa, passaremos para a classe seguinte, e repetiremos a pergunta, até o momento em que a resposta for Sim. Ao chegarmos à resposta afirmativa, pararemos e procuraremos a classe correspondente. Esta será a Classe do Primeiro Quartil! Ou seja, será desta classe que iremos extrair os dados para usar na fórmula do Q1. Uma vez constatada qual é a Classe do Primeiro Quartil, só nos resta aplicar a fórmula! A facilidade em se memorizar a fórmula do Q1 é absoluta. Vamos recordar a fórmula da Mediana:

$$M_d = l_{med} + \frac{(n/2 - f_{ant})}{f_{med}} h$$

Agora é só pensar o seguinte: o que mudou até aqui para o Quartil foi que (n/2) passou a ser (n/4). Então também será apenas isso que irá mudar na fórmula. Daí, o primeiro quartil será determinado por:

$$Q_1 = l_{inf} + \frac{(n/4 - f_{ant})}{f_i} h$$

Ora, esta fórmula nos fala em limite inferior (l_{inf}), nos fala em amplitude da classe (h), além de duas frequências - e . A única coisa que teremos que lembrar é que todos esses dados serão retirados, tomando como referência a Classe do Primeiro Quartil.

Em suma, os passos para determinação do Q1 de um conjunto serão os seguintes:

- Determinamos o n (somando a coluna da);
- Calculamos a posição de (n/4) (independentemente de n ser par ou ímpar);
- Construímos a coluna da ;
- Comparamos o valor da posição de (n/4) com os valores da , iniciando da fac da primeira classe (a mais de cima.) e fazendo a seguinte pergunta: “esta é maior ou igual a (n/4)?". Se a resposta for Não, passamos à da classe seguinte. Quando a resposta for Sim, pararemos e procuraremos a classe correspondente. Esta será a nossa Classe do Primeiro Quartil.
- Finalmente, aplicamos a fórmula do Q1, extraindo os dados desta classe do Q1, que acabamos de encontrar.

Exemplo continuação: Para o conjunto de dados abaixo, determine o valor do primeiro quartil.

X_i	f_i
0! --- 10	2
10! --- 20	5
20! --- 30	8
30! --- 40	6
40! --- 50	3

1º Passo) Encontraremos n e calcularemos a posição (n/4):

X_i	f_i
0! --- 10	2
10! --- 20	5
20! --- 30	8
30! --- 40	6
40! --- 50	3
	n = 24

Daí, achamos que n=24 e, portanto, (n/4)=6ª posição.

2º Passo) Construímos a f_{ac} :

X_i	f_i	fac
0! --- 10	2	2
10! --- 20	5	7
20! --- 30	8	15
30! --- 40	6	21
40! --- 50	3	24
	n = 24	

3º Passo) Comparamos os valores da f_{ac} com o valor de (n/4), fazendo a pergunta de praxe, adaptada ao primeiro quartil:

X_i	f_i	fac	
0! --- 10	2	2	→ 2 é maior ou igual a 6? NÃO!
10! --- 20	5	7	→ 7 é maior ou igual a 6? SIM!
20! --- 30	8	15	
30! --- 40	6	21	
40! --- 50	3	24	
	n = 24		

4º Passo) Só nos resta agora aplicar a fórmula do Primeiro Quartil, tomando como referência a Classe do Q1, que acabamos de encontrar. Teremos:

$$Q_1 = l_{inf} + \frac{(n/4 - f_{ant})}{f_i} h = 10 + \frac{(6-2)}{5} . 10 = 18$$

.Ou seja, temos que 25% das observações da amostra têm valores abaixo de 18, e também que 75% das observações têm valores acima de 18.

E assim é feito para o cálculo dos demais Quartis. Para qualquer separatrizes, seja para calcular algum quartil, decil ou percentil, o que irá mudar nas fórmulas dos Quartis, Decis e Percentis será apenas a posição da separatriz. Seguindo o mesmo raciocínio, para o cálculo das posições de alguns Decis, temos que as seguintes posições:

- Fração do Quarto Decil: D4 à (4n/10)
- Fração do Quinto Decil: D5 à (5n/10)
- Fração do Decil: D6 à (6n/10)
- Fração do Sétimo Decil: D7 à (7n/10)
- Fração do Oitavo Decil: D8 à (8n/10)
- Fração do Nono Decil: D9 à (9n/10)

Finalmente, aplicamos a fórmula do Decil extraindo os dados desta classe do , que acabamos de encontrar. Ou seja

$$D_i = l_{inf} + \frac{(i.n/10 - f_{ant})}{f_i} h$$

E para os Percentis, temos que:

- Fração do Segundo Percentil: P2 (2n/100)
- Fração do Terceiro Percentil: P3 (3n/100)
- Fração do Quarto Percentil: P4 (4n/100)
- Fração do Nonagésimo Percentil: P90 (90n/100).
- Fração do Nonagésimo Oitavo Percentil: P98 (98n/100)
- Fração do Nonagésimo Nono Percentil: P99 (99n/100)

Finalmente, aplicamos a fórmula do Percentil, extraindo os dados desta classe do , que acabamos de encontrar. Ou seja

Medidas de Posição no Excel

Quando se dispõe de um aplicativo como o Excel, utilizamos a **Distribuição de frequências** ***apenas*** para a ***apresentação tabular*** e ***gráfica*** dos dados. Todas as Medidas Estatísticas podem (e devem) ser calculadas através de ***funções do Excel***, pois além do cálculo ser muito mais rápido que os cálculos manuais, as funções podem ser aplicadas diretamente aos Dados Brutos (Tabela Primitiva ou Banco de Dados), o que torna os resultados mais precisos, como vamos observar adiante. As funções do Excel para Medidas de Posição são:

MÉDIA

Função MÉDIA (núm 1; núm 2; ...)
núm 1; núm 2; ... são de 1 a 30 argumentos numéricos para os quais se deseja obter a média.
COMENTÁRIOS: a) Os argumentos devem ser números ou eles devem ser nomes, matrizes ou referências que contenham números.
b) Se uma matriz ou argumento de referência contiver texto, valores lógicos ou células vazias, estes valores serão ignorados; no entanto, as células com valor zero serão incluídas.

MODA

Função MODO (núm 1; núm 2; ...)
núm 1; núm 2; ... são argumentos de 1 a 30 para os quais se deseja calcular a Moda. Pode-se usar também uma única matriz ou referência auma matriz em vez de argumentos separados por pontos e vírgulas.
COMENTÁRIOS: a) Os argumentos devem ser números ou nomes, matrizes ou rereferências que contenham números.
b) Se uma matriz ou argumento de referência contiver texto, valores lógicos ou células vazias, estes valores serão ignorados; no entanto, as células com valor zero serão incluídas.
c) Se o conjunto de dados não contiver pontos de dados duplicados, MODO retornará o valor de erro #N/D.

MEDIANA

Função MED (núm 1; núm 2; ...)
núm 1; núm 2; ... são de 1 a 30 números dos quais se deseja obter a mediana.
COMENTÁRIOS: a) Os argumentos devem ser números ou nomes, matrizes ou rereferências que contenham números. O Excel examina todos os números em cada argumento de referência ou matriz.
b) Se uma matriz ou argumento de referência contiver texto, valores lógicos ou células vazias, estes valores serão ignorados; no entanto, as células com valor zero serão incluídas.
c) Se houer quantidade par de números no conjunto, MED calculará a média dos dois números do meio.

Ilustração: Sofia Guimarães

PERCENTIL

Função PERCENTIL (matriz;k)

matriz é a matriz ou intervalo de dados que se quer definir a posição relativa.
k é o valor do percentil no intevalo 0...1, inclusive.

COMENTÁRIOS:

a) Se *matriz estiver vazio ou contiver mais de 8.191 pontos de dados*. PERCENTIL retornará o valor de erro #NÚM!.
b) Se *k não for numérico*, PERCENTIL retornará o valor de erro #VALOR!.
c)Se *k for <0 ou se k>1*, PERCENTIL retornará o valor de erro #NÚM!.
d) Se *k não for múltiplo de 1/(n-1)*, PERCENTIL interpolará para determinar o valor no k-ésimo percentil.

QUARTIL

Função QUARTIL (matriz;quarto)

matriz é a matriz ou intervalo de célula de valores numéricos cujo valor quartil você deseja obter.
quarto indica o valor a ser retornado.
Se quarto for igual a (0 a 4). QUARTIL retornará
0 Valor mínimo
1 Primeiro quartil (25º percentil)
2 Valor Médio (50º percentil)
3 Terceiro quartil (75º percentil)

COMENTÁRIOS:

Se a matriz estiver vazia. QUARTIL retornará o valor de erro #NÚM!.
Se quarto não for um número inteiro, será truncado.
Se quarto <0 ou se quarto >4. QUARTIL retornará o valor de erro #NÚM!.
MÍNIMO, MED e MÁXIMO retornarão o mesmo valor que QUARTIL quando quarto for igual a 0, 2 e 4, respectivamente.

Ilustração: Orlando Dantas

Vamos verificar os resultados dos cálculos feitos à mão com os resultados obtidos com as funções do Excel.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
23		Tabela Primitiva da Pesquisa sobre Salários da Empresa X													
24															
25		625	569	338	593	462	420	579	392						
26		591	477	404	450	480	323	450	760						
27		619	573	484	600	246	632	654	499						
28		513	458	565	500	610	420	604	673						
29		580	683	470	450	220	541	782	552						
30															
31		**Elemento Calculado**										**Resultado**			
32		Acesso à Tabela										B25:$1$29			
33															
34															
35															
36															
37		**Medida Calculada**						**Função**				**Resultado**		**Célula**	**Fórmula**
38		Média						MÉDIA ()				521,03		L38	=MÉDIA (INDIRETO(L32))
39		Moda						MODA ()				450,00		L39	=MODO (INDIRETO(L32))
40		Mediana						MED ()				527,00		L40	=MED (INDIRETO(L32))
41		Mediana						QUARTIL ()				527,00		L41	=QUARTIL (INDIRETO(L32);2)
42		Mediana						PERCENTIL ()				527,00		L42	=PERCENTIL (INDIRETO(L32);50%)
43		Percentil 71 %						PERCENTIL ()				592.38		L43	=PERCENTIL (INDIRETO(L32);71%)
44															

Ilustração: Sofia Guimarães

Nos cálculos de medidas feito com os dados brutos, podem ocorrer, do que denominamos de ***erros de agrupamento***, introduzido pelo agrupamento de valores da variável em **Intervalos**. Por isso, o agrupamento não deve ser feito de forma aleatória, e sim através do método visto na **seção 2.6.1**

ATIVIDADE RESPONDIDA

1 - Considere a seguinte distribuição de frequência correspondente aos diferentes preços de um determinado produto em vinte lojas pesquisadas.

PREÇOS	Nº DE LOJAS
50	2
51	5
52	6
53	6
54	1
TOTAL	**20**

a) Quantas lojas apresentaram um preço de R$52,00? ***2***

b) Construa uma tabela de frequências simples relativas.

c) Construa uma tabela de frequências absolutas acumuladas.

d) Quantas lojas apresentaram um preço de até R$52,00 (inclusive)? ***13***

e) Qual o percentual de lojas com preço maior de que R$51,00 e menor de que R$54,00? ***6%***

PREÇOS	Nº DE LOJAS	fri	Fi	Fri
50	2	0,1	2	0,1
51	5	0,25	7	0,35
52	6	0,3	13	0,65
53	6	0,3	19	0,95
54	1	0,05	20	1
TOTAL	**20**	**1**		

2 - O quadro seguinte representa as alturas (em cm) de 40 alunos de uma classe.

162	163	148	166	169	154	170	166
164	165	159	175	155	163	171	172
170	157	176	157	157	165	158	158
160	158	163	165	164	178	150	168
166	169	152	170	172	165	162	164

Calcular a amplitude total.

Admitindo-se 6 classes, qual a amplitude do intervalo de classe?

Construir uma tabela de frequência das alturas dos alunos.

Determinar os pontos médios das classes.

At = 178 - 148 = 30
K = 6 · h = 30/6 = 5

CLASSES	fi	Xi
148 |- 153	3	150,5
153 |- 158	5	155,5
158 |- 163	7	160,5
163 |- 168	13	165,5
168 |- 173	9	170,5
173 |-| 178	3	175,5
TOTAL	40	

3 - Responda as questões abaixo: Média, Mediana e Moda são medidas de :

a) ( ) Dispersão b) ( ) posição c) ( ) assimetria d) ( ) curtose

Na série 10, 20, 40, 50, 70, 80 a mediana será:

a) ( ) 30 b) ( ) 35 c) ( ) 40 d) ( ) 45

50% dos dados da distribuição situa-se:

a) ( ) abaixo da média c) ( ) abaixo da moda

b) ( ) acima da mediana d) ( ) acima da média

4 - Calcule para cada caso abaixo a respectiva média.

a) 7, 8, 9, 12,14 Média = 10

b) Xi 3 4 7 8 12

Fi 2 5 8 4 3 Média = 6,8

c) Classes: 68 - 72 72 - 76 76 - 80 80 - 84

Fi: 8 20 35 40 Média = 78,15

5 - Calcule o valor da mediana.

a) 82, 86, 88, 84,91, 93 Mediana = 87

b) Xi 73 75 77 79 81

Fi 2 10 12 5 2 Mediana = 77

c) Classes: 1 - 3 3 - 5 5 - 7 7 - 9 9 - 11 11 - 13

Fi: 3 5 8 6 4 3 Mediana = 6,63

6 - Calcule a moda

a) 3, 4, 7, 7, 7, 8,9, 10 Moda = 7

b) Xi 2,5 3,5 4,5 6,5

Fi 7 17 10 5 Moda = 3,5

c) Classes: 10 - 20 20 - 30 30 - 40 40 - 50

Fi: 7 19 28 32 Moda = 41,11

7 - Para as distribuições abaixo foram calculados

Distrib. A Distrib. B Distrib. C

CLASSES	Fi
02 |- 06	6
06 |- 10	12
10 |- 14	24
14 |- 18	12
18 |- 22	6

$\bar{x}$ = 12 Kg
Med = 12 Kg
Mo = 12 Kg

CLASSES	Fi
02 |- 06	6
06 |- 10	12
10 |- 14	24
14 |- 18	30
18 |- 22	6

$\bar{x}$ = 12,9 Kg
Med = 13,5 Kg
Mo = 16 Kg

CLASSES	Fi
02 |- 06	6
06 |- 10	30
10 |- 14	24
14 |- 18	12
18 |- 22	6

$\bar{x}$ = 11,1 Kg
Med = 10,5 Kg
Mo = 8 Kg

Ilustração: Sofia Guimarães

Marque a alternativa correta:

a) a distribuição I é assimétrica negativa;

b) a distribuição II é assimétrica positiva;

c) a distribuição III é assimétrica negativa moderada.

d) a distribuição I é simétrica;

ATIVIDADE

1 - Construir a distribuição de frequência para as idades, em anos, de um grupo de amigos.

14 15 16 16 16 14 14 15 17 14

15 16 17 17 16 15 14 15 15 15

Com os dados organizados, responda as perguntas abaixo:

a) o grupo pesquisado é formado por _____ pessoas;

b) a pessoa mais velha tem _____ anos e a mais nova tem _____ anos;

c) a maioria tem ___ anos;

d) a minoria tem ___ anos.

e) variável estudada;

f) tipo de variável

2 - Os dados abaixo referem-se ao número de horas gastas por jovens assistindo a programas de TV durante um fim de semana.

6 8 2 7 10 5 6 7 2 10 6 8 7 7 6

5 2 7 8 10 8 7 7 7 6 10 5 5 5 5

Construir uma tabela de distribuição de frequência com frequência relativa em porcentagem correspondente aos dados fornecidos. Depois responder.

a) O menor tempo gasto pelos jovens assistindo a programas de TV é ________ horas e o maior tempo é _______ horas.

b) _______% dos jovens assistem a programas de TV durante 7 horas.

c) _______% dos jovens gastam 8 horas assistindo a programas de TV e _____% gastam 10 horas diante da televisão.

3 – Os dados abaixo se referem às variações (%) de crescimento da Empreendedora XXX do município da Microrregião Oeste Baiana (Dados Fictícios) ao longo de 43 meses.

VARIAÇÃO	Nº DE MESES
9,9 |-- 15,9	8
15,9 |-- 21,9	12
21,9 |-- 27,9	13
27,9 |-- 33,9	5
33,9 |-- 39,9	3
39,9 |-- 45,9	2
TOTAL	43

a) O responsável pela parte Estatística afirmou que o valor que me fornece 45% das maiores variações de crescimento é 27,97%. Ele está correto?

b) Complete a Tabela com as frequências relativas simples e frequências acumuladas absoluta e relativa.

c) Verifique qual o tipo de Assimetria dos dados.

2.8 Medidas de Dispersão

Já estudamos que um conjunto de valores pode ser sintetizado por meio de procedimento matemático, como o cálculo da média, moda, mediana, quartis e percentis. No entanto, a interpretação de dados estatísticos exige que se realize um número maior de estudos, além dos estudados nas unidades precedentes. Torna-se necessário ter uma ideia de como se apresentam os dados, qual a variação em torno da média, qual a concentração. Vejamos o seguinte exemplo:

Foram avaliados três grupos de executivos, cada um com cinco elementos, no que se refere à criatividade e os testes mostraram os seguintes resultados:

Grupo A: 5 5 5 5 5

Grupo B: 3 4 5 6 7

Grupo C: 1 2 5 7 10

Para representar cada grupo, podemos calcular a média e vamos verificar que os três grupos têm a mesma média = 5, entretanto observando a variação dos dados podemos perceber que os grupos se comportam de forma diferente, apesar de todos terem a mesma média. Nesse caso, a média ainda que considerada como um número que pode representar uma sequência de números, não pode destacar o grau de homogeneidade ou heterogeneidade que existe entre os valores que compõem o conjunto. Desse modo, precisamos efetuar outros procedimentos matemáticos para caracterizar melhor os dados de cada grupo com o objetivo de tirarmos conclusões qualitativas.

As medidas que mostram a variação dos dados de um conjunto são chamadas de Medidas de Dispersão ou Variabilidade:

I) Medida de Dispersão Absoluta:

- Amplitude total;
- Desvio médio;
- Variância e desvio-padrão.

II) Medidas de Dispersão relativa:

- Coeficiente de variação de Pearson;

2.8.1 Medidas de Dispersão Absoluta

Amplitude Total ou Intervalo Total

O **Símbolo** da Amplitude Total é**: AT**

Definição 2.8.1.1: *A amplitude total de um conjunto de números é a diferença entre os valores extremos do conjunto observado:*

$$AT = X_{máx} - X_{mín}$$

Exemplo: Calcular a amplitude total dos seguintes conjuntos de números:

A = {10, 12, 13, 20, 25, 34, 45}

B = {17, 18, 19, 20, 21, 22, 23}

C = {-4, -3, -2, 3, 5}

Para o conjunto A, temos: AT = 45 - 10 = 35

Para o conjunto B, temos: AT = 23 - 17 = 6

Para o conjunto C, temos: AT = 5 - (-4) = 9

Se os dados vierem dispostos em uma tabela de frequências, com os valores agrupados em classes, há duas formas de se definir a amplitude total:

Primeiro Método: AT = Ponto médio da última classe **-** ponto médio da primeira classe.

Segundo Método: AT = Limite superior da última classe - limite inferior da primeira classe.

Exemplo: Calcular a amplitude total dos valores dispostos na **Tabela 2.8.1.1**

Tabela 2.8.1.1 - Consumo de água do Bairro Santa Mônica /06

CLASSES	F_J	X_J
10 |---- 20	5	15
20 |---- 30	12	25
30 |---- 40	20	35
40 |---- 50	14	45
50 |---- 60	10	55
60 |---- 70	4	65
	n = 65	

Pelo **Primeiro Método:** AT = 65 - 15 = 50

Nesse método, os valores extremos são eliminados.

Pelo **Segundo Método:** AT = 70 - 10 = 60

Restrições ao uso da Amplitude Total

Embora a amplitude total seja a mais simples das medidas de dispersão, há uma forte restrição ao seu uso em virtude de sua grande instabilidade, uma vez que ela leva em conta apenas os valores extremos da série. Comparemos os conjuntos A e B do exemplo 1:

CONJUNTO	MÉDIA	AMPLITUDE TOTAL: A_t
A = {10, 12, 13, 15, 20, 25, 45}	$\bar{x} = 20$	$AT_A = 35$
B = {17, 18, 19, 20, 21, 22, 23}	$\bar{x} = 20$	$AT_B = 6$

A média aritmética de cada um desses conjuntos é igual a 20. Portanto, no que diz respeito a uma medida de posição, ambos os conjuntos podem ser considerados idênticos. Ao calcularmos a amplitude total, verificaremos que os valores do conjunto **A** apresentam maior dispersão. Todavia, no cálculo da amplitude total não são levados em consideração os valores da série que se encontram entre os extremos, o que poderia conduzir o analista a interpretações equivocadas. Muitas vezes, um valor particularmente anormal poderá afetar de maneira acentuada a medida. O conjunto **A**, por exemplo, apresenta o último valor (45) sensivelmente distante do penúltimo (25), fato que talvez tenha provocado uma amplitude total de tal magnitude (35).

Além da insensibilidade aos valores entre os extremos anormais, a amplitude total é sensível ao tamanho de amostra. Ao aumentar essa última, a amplitude total tende a aumentar, ainda que não proporcionalmente. Finalmente, a amplitude total apresenta muita variação de uma amostra para outra, mesmo que ambas sejam extraídas da mesma população.

Apesar dos inconvenientes dessa medida, os quais não justificam, na maioria das vezes, seu uso, há situações especiais em que ela resulta satisfatória. É o caso, por exemplo, da amplitude da temperatura em um dia ou no ano. Outra situação seria aquela em que os dados são raros ou demasiadamente esparsos para justificar o emprego de uma medida mais precisa.

É importante acrescentar que, ao descrever uma série por uma medida de posição (média, por exemplo) e de dispersão, se essa última for a amplitude total, é recomendável que se indiquem os valores extremos da série.

Exemplo: Determine a amplitude total em cada um dos casos.

1 - A= $\{12,15,25,32,45,18,36,19\} \rightarrow AT = 45 - 12 = 33$

2 - **Número de faltas dos acadêmicos da Turma A**

FALTAS	ACADÊMICOS
2	15
3	10
5	8
6	6
9	4
TOTAL	**43**

$$AT = 9 - 2 = 7$$

3 - Notas da atividade 1, dos acadêmicos da Turma A

NOTAS	ACADÊMICOS
3,8 |- 4,8	5
4,8 |- 5,8	6
5,8 |- 6,8	12
6,8 |- 7,8	15
7,8 |- 8,8	5
TOTAL	**43**

Pelo **Primeiro Método:** AT = 8,3 – 4,3 = 4

Nesse método, os valores extremos são eliminados.

Pelo **Segundo Método:** AT = 8,8 - 3,8 = 5

2.8.2 Desvio Médio (Símbolo: D_m)

Definição 2.8.2.1: *O desvio médio ou média dos desvios é igual à média aritmética dos valores absolutos dos desvios tomados em relação a uma das seguintes medidas de tendência central: média ou mediana.*

Desvio Médio para Dados Brutos

Quando os valores não vierem dispostos em uma tabela de frequências, o desvio médio será calculado, de acordo com a definição, através do emprego de uma das seguintes fórmulas:

$$D_m = \frac{\sum_{i=1}^{n} \mid x_i - \bar{x} \mid}{\sum_{i=1}^{n} f_i}$$ (1) Desvio em relação á média aritmética;

$$D_m = \frac{\sum_{i=1}^{n} \mid x_i - x_{med}}{\sum_{i=1}^{n} f_i}$$ (2) Desvio em relação á mediana.

As barras verticais indicam que são tomados os valores absolutos dos desvios.

Exemplo: Calcular o desvio médio do conjunto de números apresentado a seguir:

A= {10, 12, 13, 20, 25, 34, 45}

$$\bar{x_A} = \frac{10+12+13+20+25+34+45}{7} = 22{,}71, \text{ e } x_{medA} = 20$$

É conveniente colocarmos os valores dispostos em uma tabela, considerando:

Elementos do Conjunto A

X_i	$(X_i - \bar{X})$	$\lvert X_i - \bar{X} \rvert$	$(X_i - X_{ma})$	$\lvert X_i - X_{ma} \rvert$
10	10 - 22,71 = -12,71	12,71	10 - 20 = -10	10
12	12 - 22,71 = - 10,71	10,71	12 - 20 = -8	8
13	13 - 22,71 = -9,71	9,71	13 - 20 = -7	7
20	20 - 22,71 = 2,71	2,71	20 - 20 = 0	0
25	25 - 22,71 = 2,29	2,29	25 - 20 = 5	5
34	34 - 22,71 = 11,29	11,29	34 - 20 = 15	14
45	45 - 22,71 = 22,29	22,29	45 - 20 = 25	25
		$\sum = 61,71$		$\sum = 69$

Ilustração: Sofia Guimarães

Usando as fórmulas (1) e (2), chegaremos a:

Pela Média

$$D_m = \frac{\sum_{i=1}^{n} \mid x_i - 22,71 \mid}{7} = 10,24$$

Pela Mediana

$$D_m = \frac{\sum_{i=1}^{n} \mid x_i - 20 \mid}{7} = 9,86$$

Desvio Médio para Dados Tabulados sem Intervalo de Classe

Se os valores vierem dispostos em uma tabela de frequências, agrupados ou não em classes, serão usadas as seguintes fórmulas:

Cálculo pela Média:

$$D_m = \frac{\sum_{i=1}^{n} \mid x_i - \bar{x} \mid f_i}{\sum_{i=1}^{n} f_i}$$

Onde X_i representa um valor individual ou um ponto médio da classe.

Cálculo pela Mediana

$$D_m = \frac{\sum_{i=1}^{n} \mid x_i - x_{med} \mid f_i}{\sum_{i=1}^{n} f_i}$$

Exemplo: Calcular o desvio médio em relação à média e em relação à mediana do número de empregados por estabelecimentos comerciais conforme a tabela abaixo.

Tabela 2.8.2.1 - Número de empregados por estabelecimentos comerciais

Emp/estab	f	Fa	$x_i f_i$	$\lvert x_i - 4{,}71 \rvert . f_i$	$\lvert x_i - 5 \rvert . f_i$
1	2	2	2	3,71 . (2) = 7,42	4 . (2) = 8
2	3	5	6	2,71 . (3) = 8,13	3 . (3) = 9
3	3	8	9	1,71 . (3) = 5,13	2 . (3) = 6
4	5	13	20	0,71 . (5) = 3,55	1 . (5) = 5
5	6	19	30	0,29 . (6) = 1,74	0 . (6) = 0
6	4	23	24	1,29 . (4) = 5,16	1 . (4) = 4
7	3	26	21	2,29 . (3) = 6,87	2 . (3) = 6
10	2	28	20	5,29 . (2) = 10,58	5 . (2) = 6
	28		$\sum x_i f_i = 132$	$\sum = 48{,}58$	$\sum = 44$

Ilustração: Orlando Dantas

Inicialmente completamos a **Tabela 2.8.2.1** Após determinarmos os valores da mediana e da média devemos colocar mais duas colunas para o cálculo do desvio com a média e com a mediana.

Temos que a média é dada por 4,71. Para encontrarmos o valor do desvio em relação à média completamos a tabela com a coluna 5.

$$D_m = \frac{48{,}58}{28} = 1{,}74$$

Cálculo da Mediana: Posição da mediana é dada por $\frac{n}{2} = \frac{28}{2} = 14(posição)$.

Na coluna da Fa vemos que o elemento de ordem 14° está na classe 5, onde o valor de x é 5, ou seja, $x_{med} = 5$

.Para encontrarmos o valor do desvio em relação à mediana completamos a tabela com a coluna 6.

Desvio Médio para dados Tabulados com Intervalo de Classe.

Exemplo: Calcular o desvio médio dos valores representativos do consumo de energia elétrica (em Kwh) de 80 usuários.

Tabela 2.8.2.2 - Consumo de energia elétrica de consumidores de Campo Grande/MS/06

Consumo (Kwh)	f1	x1	x1f1	$\lvert x_i - \bar{x} \rvert f_i$	$\lvert x_i - x_{md} \rvert f_i$	F_i
5 ⊢ 25	4	15	60	64,5 x 4 = 258	62,31 x 4 = 249,24	4
25 ⊢ 45	6	35	210	44,5 x 6 = 267	42,31 x 6 = 253,86	10
45 ⊢ 65	14	55	770	24,5 x 14 = 343	22,31 x 14 = 312,34	24
65 ⊢ 85	26	75	1950	4,5 x 27 = 114	2,31 x 27 = 60,06	50
85 ⊢ 105	14	95	1330	15,5 x 14 = 217	17,69 x 14 = 247,66	64
105 ⊢ 125	8	115	920	35,5 x 8 = 284	37,69 x 8 = 301,52	72
125 ⊢ 145	6	135	810	55,5 x 6 = 333	57,69 x 6 = 346,14	78
145 ⊢ 165	2	155	310	75,5 x 2 = 150	77,69 x 2 = 155,38	80
	80		∑ = 6360	1970	∑ = 1926,2	

Ilustração: Orlando Dantas

Iniciamos com o cálculo da média e da mediana

. $Média = \bar{x} = \dfrac{6.360}{80} = 79,5$

$Mediana \rightarrow x_{med} = 65 + \frac{40-24}{26} . 20 \rightarrow x_{med} = 77,31.$

Completamos as tabelas com as diferenças e os produtos necessários para o cálculo do desvio médio.

Cálculo pela média: $D_m = \dfrac{\sum_{1=1}^{n} \lvert\, x_i - 79,5 \,\rvert f_i \,=\, 24,63kwh}{80}$

Cálculo pela média: $D_m = \dfrac{\sum_{1=1}^{n} \lvert\, x_i - 77,31 \,\rvert f_i \,=\, 24,08kwh}{80}$

Podemos observar novamente que o desvio médio, calculado com base na mediana, é menor que o calculado com base na média aritmética.

Observações:

1. O desvio médio apresenta resultado mais vantajoso que as medidas de dispersão precedentes, principalmente pelo fato de, em seu cálculo, levar em consideração todos os valores da distribuição.

2. O desvio médio, calculado levando-se em consideração os desvios em torno da mediana, é mínimo, ou seja, é menor do que qualquer desvio médio calculado com base em qualquer outra medida de tendência central.

3. Apesar de o desvio médio expressar aceitavelmente a dispersão de uma amostra não é tão frequentemente empregado como o desvio-padrão, o qual será estudado a seguir, pois este se adapta melhor a uma ampla gama de aplicações. Além disso, o desvio médio não considera o fato de alguns desvios serem negativos e outros positivos, pois essa medida os trata como se fossem todos positivos, como valores absolutos. Contudo, será preferível o uso do desvio médio em lugar do desvio-padrão, quando esse for indevidamente influenciado pelos desvios extremos.

2.8.3 Variância (Símbolo: S^2)

Vimos que a Amplitude total e o Desvio Médio são medidas que se deixam influenciar pelos valores extremos, que em grande maioria são devidos ao acaso.

Definição 2.8.3.1: *A variância é uma medida que leva em consideração valores extremos e os valores intermediários, isto é, expressa melhor os resultados obtidos. A variância relaciona os desvios em torno da média populacional, ou mais claramente, é a média aritmética dos quadrados dos desvios.*

Variância de uma população: $S^2 = \frac{\sum_{i=1}^{n} \quad (x_i - \mu)^2}{n}$

Observação: É mais comum na estatística o trabalho com amostra e não com a população. Neste caso o denominador passa a ser (n - 1) em vez de **n**, pois assim teremos uma melhora na estimativa do parâmetro da população:

Variância de uma amostra: $S^2 = \frac{\sum_{i=1}^{n} \quad (x_i - \bar{x})^2}{n-1}$

2.8.4 Desvio-padrão (Símbolo: S)

O desvio-padrão é a medida de dispersão mais usada, tendo em comum com o desvio médio o fato de em ambos serem considerados os desvios com relação à média. Só que, no cálculo do desvio-padrão, em lugar de serem usados os valores absolutos das discrepâncias ou desvios, calculam-se os quadrados desses. O desvio-padrão não é senão uma média quadrática dos desvios em relação à média aritmética de um conjunto de números, ou seja, é a raiz quadrada da média aritmética dos quadrados dos desvios, tomados a partir da média aritmética.

Desvio-padrão de Dados Brutos

Seja o seguinte conjunto de números: X = $\{x_{1,} x_2 ... x_n\}_.$ O desvio-padrão ou a média quadrática dos desvios ou afastamento em relação à média aritmética desse conjunto será definido por:

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

Exemplo: Calcular o desvio padrão do conjunto de números A={1,3,5,7}.

Temos nesse caso, que a média é dada por 4.

Cálculo do Desvio Padrão:

X_i	$(x_i - 4)^2$
1	$(1-4)^2 = 9$
3	$(3-4)^2 = 1$
5	$(5-4)^2 = 1$
7	$(7-4)^2 = 9$
	$\sum (x_i - 4)^2 = 20$

Aplicando a fórmula do desvio padrão temos: $s = \sqrt{\frac{20}{4}} = 2{,}23$

Observação: quando o desvio-padrão representar uma descrição da amostra e não da população, caso mais frequente em estatística, o denominador das expressões será igual a n - 1, em vez de n. A razão desse procedimento reside no fato de que, utilizando o divisor (n - 1), obtém-se uma estimativa melhor do parâmetro de população. Além do mais, apenas n - 1 das discrepâncias ($x_i - x$) são independentes, uma vez que essas (n - 1) discrepâncias determinam automaticamente a n-ésima. Para valores grandes de n ($n > 30$) não há grande diferença entre os resultados proporcionados pela utilização de qualquer dos dois divisores, n ou n - 1. Entretanto, daremos preferência para a fórmula que proporciona uma estimativa mais justa do desvio-padrão da população, ou seja:

Desvio-padrão de Dados Tabulados sem Intervalo de Classe

Quando os valores vierem dispostos em uma tabela de frequências, o cálculo do desvio-padrão se fará através de uma das seguintes fórmulas:

Desvio padrão para dados populacionais:

$$S = \sqrt{\frac{\sum_{i=1}^{n} \quad (x_i - \mu)^2 f_i}{n}}$$

Desvios padrão para dados amostrais:

$$S = \sqrt{\frac{\sum_{i=1}^{n} \quad (x_i - \bar{x})^2 f_i}{n - 1}}$$

Exemplo: Calcular o desvio padrão da tabela a seguir:

Tabela 2.8.4.1: Número de faltas/mês dos funcionários da empresa Agro Sul / 08

Nº DE FALTAS / MÊS	f	$X_i f_i$
0	4	0
1	3	3
2	2	4
3	1	3
4	1	4
5	2	10
6	1	6
7	1	7
*PESQUISA POPULACIONAL	$\sum f = 15$	$\sum x_i f_i = 37$

Ilustração: Sofia Guimarães

Inicialmente calculamos a média

$$\bar{x} = \frac{37}{15} = 2{,}47 \; faltas/mês$$

Em seguida, inserimos mais uma coluna na tabela 4.9 onde vamos fazer os cálculos necessários para o cálculo do desvio padrão:

Tabela 2.8.4.2: Número de faltas/mês dos funcionários da empresa Agro Sul / 08

Nº de faltas / mês	f	$x_i f_i$	$(x_i - 2{,}47)^2 f_i$
0	4	0	$(0 - 2{,}47)^2 . (4) = 24{,}40$
1	3	3	$(1 - 2{,}47)^2 . (3) = 6{,}48$
2	2	4	$(2 - 2{,}47)^2 . (2) = 0{,}44$
3	1	3	$(3 - 2{,}47)^2 . (1) = 0{,}28$
4	1	4	$(4 - 2{,}47)^2 . (1) = 2{,}34$
5	2	10	$(5 - 2{,}47)^2 . (2) = 12{,}80$
6	1	6	$(6 - 2{,}47)^2 . (1) = 12{,}46$
7	1	7	$(7 - 2{,}47)^2 . (1) = 20{,}52$
* pesquisa populacional	$\sum f = 15$	$\sum x_i f_i = 37$	79,72

Ilustração: Orlando Dantas

$$S = \sqrt{\frac{79}{15}} = 2{,}3$$

Ou seja, cada funcionário se distancia do número médio de faltas, em média, de 2,3 faltas/mês.

Desvio padrão de Dados Tabulados com Intervalo de Classes

Quando tivermos que calcular o desvio padrão para tabelas de dados tabulados com intervalos de classes usaremos as mesmas fórmulas para dados sem intervalos de classes, utilizando para os pontos médios de cada classe, seguindo com os mesmos procedimentos.

Exemplo: Com dados da tabela a seguir, calcule o desvio-padrão da distribuição de frequências do consumo de energia elétrica (Kwh) :

Tabela 2.8.4.3: Distribuição de frequências do consumo de energia elétrica

CONSUMO	NÚMEROS DE USUÁRIOS f_i	X_i	$X_i f_i$	$(X_i - 79,5)$	$(X_i - 79,5)^2$	$(X_i - 79,5)^2 f_i$
$5 \vdash 25$	4	15	60	-64,5	4160,25	16641,0
$25 \vdash 45$	6	35	210	-44,5	1980,25	11881,5
$45 \vdash 65$	14	55	770	-24,5	600,25	8403,5
$65 \vdash 85$	26	75	1950	-4,5	20,25	526,5
$85 \vdash 105$	14	95	1330	15,5	240,25	3363,5
$105 \vdash 125$	8	115	920	35,5	1260,25	10082,0
$125 \vdash 145$	6	135	810	55,5	3080,25	18481,5
$145 \vdash 165$	2	155	310	75,5	5700,25	11400,5
$\sum$			6360			80780

Ilustração: Sofia Guimarães

A média aritmética do consumo já foi calculada anteriormente, e é dada por 79,5.

Cálculo do Desvio-padrão pela Fórmula Original : $S = \sqrt{\dfrac{80.780}{79}} = 31{,}98$

.O desvio-padrão do consumo de energia elétrica é 31,98 Kwh. Lembre-se que o desvio médio já foi calculado, resultando em $D_m = 24{,}63$ Kwh.

2.8.5 Coeficiente de Variação de Pearson

O desvio-padrão por si só não revela muita coisa. Assim, um desvio padrão pode ser considerado pequeno para uma média e para outra é extremamente grande. Por exemplo, um desvio-padrão de 40 pode ser considerado pequeno para uma média de 35, entretanto, se a média for 4, este se torna muito grande.

Quando precisamos comparar duas ou mais séries de valores quanto à sua dispersão e variabilidade e esses conjuntos estão expressos em grandezas diferentes é preciso dispor de outra medida. Para contornar essas dificuldades e limitações, podemos caracterizar a dispersão ou variabilidade dos dados de maneira relativa ao seu valor médio.

Definição 2.8.5.1: *Uma medida que mede o grau de concentração dos valores em torno da média é denominada de Coeficiente de Variação.*

$$C.V = \frac{S}{\bar{x}}$$

Podemos realizar interpretações do coeficiente de variação através de algumas regras empíricas:

- **Se: C.V < 15% tem-se baixa dispersão**
- **Se: 15% ≤ C.V. < 30% tem-se média dispersão**
- **Se: C.V ≥ 30% tem-se elevada dispersão**

Podemos classificar as distribuições em homogêneas ou heterogêneas da seguinte forma:

Distribuição homogênea: tem coeficiente da variação com baixa ou média dispersão (até 30% de variação) .

Distribuição heterogênea: tem coeficiente da variação com elevada dispersão (acima de 30% de variação) .

Exemplo . Na Empresa Carrefour, o salário médio dos homens é de R$ 1500,00 com desvio-padrão de R$ 650,00 e o salário médio das mulheres é de R$ 1200,00 com desvio padrão de 580,00. A dispersão relativa dos salários é maior para os homens?

Homens: Média = 1.500 e Desvio Padrão = 650

Mulheres: Média = 1.200 e Desvio Padrão = 580.

Então, temos que

$$C.V_H = \frac{650}{1.500} = 43{,}3 \text{ e } C.V_M = \frac{580}{1.200} = 48{,}3$$

Portanto, os salários das mulheres têm dispersão relativa maior que os salários dos homens. As duas distribuições apresentam alta dispersão (C.V. ≥ 30%).

ATIVIDADE

1 - Um fabricante de embalagens recebeu uma encomenda de um cliente que fabrica margarina. Para isso, apresentou ao cliente três tipos diferentes de embalagens, A, B e C, segundo a pressão média de rompimento para cada uma delas. O cliente optou pelo tipo de embalagem que possuísse a menor variação absoluta na pressão de ruptura.

TIPOS DE EMBALAGENS	A	B	C
PRESSÃO MÉDIA DE RUPTURA (BÁRIA)	300	150	200
DESVIO-PADRÃO DAS PRESSÕES (BÁRIA)	60	40	50

a) Qual das três embalagens o cliente optou? Por quê?

b) Se a sua escolha fosse apoiada na maior variação relativa, qual das três embalagens ele teria escolhido? Por quê?

2 - Um estudo realizado na Estação da Lapa para verificar a quantidade de Unidade Formadora de Colônia (UFC) em dois níveis da Estação. O objetivo é caracterizar a qualidade do ar e verificar a quantidade de UFC encontrada nos dois níveis, ou seja, quanto maior o número de UFC pior é o ar. O estudo foi realizado ao longo de 10 semanas durante a Primavera. Registraram-se as seguintes quantidades de UFC por semana.

NÍVEL	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	MÉDIA	DESVIO
1	48	48	49	51	54	55	57	58	58	58	53,9	4,24
2	55	58	58	59	60	60	61	65	65	70	61,1	4,38

a) Qual a melhor medida para saber em qual nível o ar é mais homogêneo?

b) Com base nestes dados, utilizando a medida de tendência central e de dispersão, que conclusão faria quanto à homogeneização do ar?

3 - Uma dona de casa pesou 10 potes de manteiga e verificou que a média dos pesos dos potes era de 500 g, com variação entre cada pesagem, indicando um desvio padrão de 25 g. Ela repetiu a experiência com pacotes de arroz e verificou que a média dos pesos dos pacotes de arroz era 5000 g com variação de peso entre os pacotes representados pelo desvio padrão de 100 g.

Manteiga	**Arroz**
média = 500	média = 5000
desvio padrão = 25	desvio padrão = 100

Qual dos produtos apresentou maior variação em seus pesos? Justifique a sua resposta.

A planilha Excel para procedimentos de estatística descritiva

O programa Excel do Windows é uma planilha que pode ser utilizada pelo pesquisador, pelo estudante, pelo gerente, enfim, por todos aqueles que queiram executar procedimentos de cálculo estatístico. Abordar-se neste item, algumas funções matemático-estatísticas que possibilita a obtenção de aplicações de praticamente, todas as ferramentas da estatística descritiva, tratadas nesta seção.

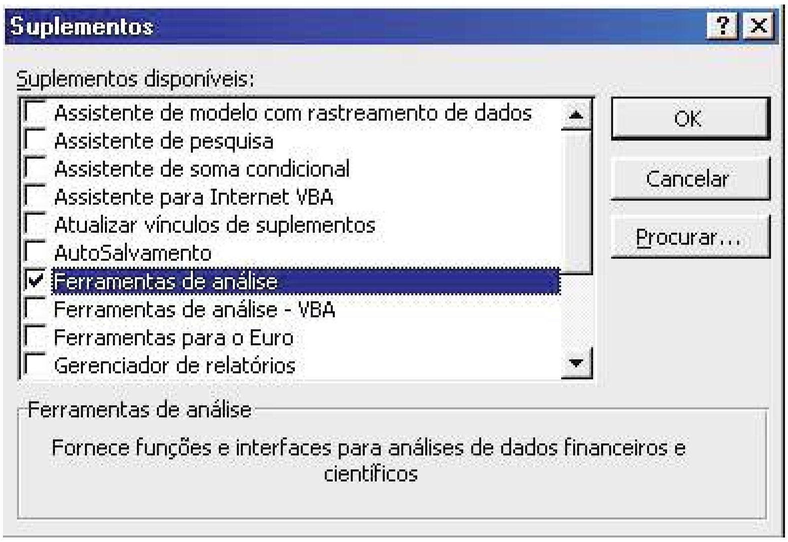
Uma vez que os dados foram inseridos na planilha de dados do Excel, uma das funções que poderá ser utilizada é a função , ou seja, a função ***Colar Função***. Primeiramente, o pesquisador deve escolher uma célula, que uma selecionada será o local onde o programa fará a inserção da operação escolhida. A partir de um clik no ícone colar função, abre-se a janela ***Colar Função*** do programa. No lado esquerdo da janela, o pesquisador seleciona a opção ***Estatística*** no quadro ***Categoria da função***, procedimento que exibe, automaticamente, no quadro à direita, diversas opções de operações estatísticas. Entre as opções, várias funções estatísticas poderão ser executas.

O Excel também possui uma opção de cálculo macro para um grupo de argumentos, que realiza de uma só vez, várias operações estatísticas.

Primeiramente, o Excel precisa ter acionada a opção ***análise de dados***, que está disponível no ícone ***Ferramentas*** da barra de menu. Caso esta opção não esteja disponível, click em ***Ferramentas***, seguido de ***Suplementos***, selecionando a opção ***Análise de dados*** nesta janela – Figura caixa de Diálogo Suplementos. Executados estes procedimentos, o Excel exib, na opção ***Ferramentas*** da barra de menu, o comando ***Análise de dados***.

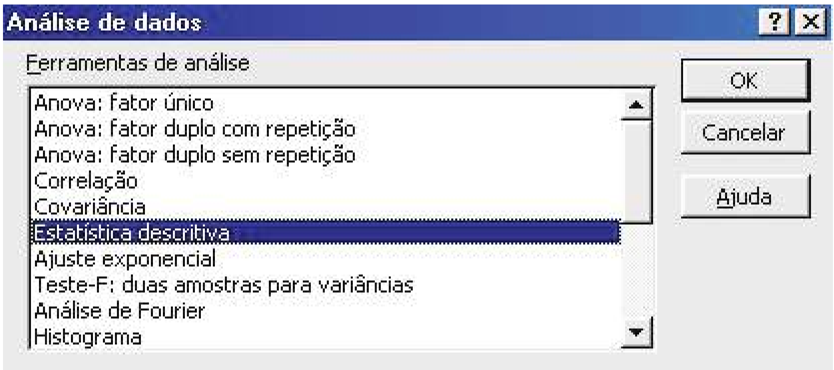
Para proceder aos cálculos da estatística descritiva através desta opção, insira o conjunto de dados em uma coluna de dados. Acione a opção ***Ferramentas*** seguido do comando ***Análise de dados***.

CAIXA DE DIÁLOGO SUPLEMENTOS



Estes procedimentos abrem a janela ***Análise de dados*** – Figura Caixa de Diálogo Análise de Dados, na qual o pesquisador deve escolher a opção estatística descritiva.

CAIXA DE DIÁLOGO ANÁLISE DE DADOS



Clicando a opção *Ok*, o Excel abre a janela ***Estatística descritiva***.

Nesta janela, o pesquisador deve indicar no quadro *Entrada* desta janela, na caixa de diálogo ***Intervalo de entrada***, em quais células se encontram os argumentos que deseja proceder aos cálculos da estatística descritiva.

Escolhe também se os cálculos serão efetuados e inseridos em colunas ou linhas. Nas opções de saída desta janela, o pesquisador pode escolher a mesma área de trabalho do Excel, ou inserir tais resultados em uma nova planilha. Ainda pode acionar opções como o intervalo de confiança para a média em níveis de probabilidade, o resumo estatístico e os valores máximo e mínimo do conjunto, que já são previstos no resumo estatístico.

A opção de estatística descritiva da planilha Excel permite ao pesquisador agilizar procedimentos de cálculo, com vistas à realização de análises. Quando se processa uma quantidade de argumentos reduzidos, como no exemplo em questão, não se pode mensurar a importância desta opção, enquanto ferramenta de análise e agilidade no processo. Entretanto, se os argumentos em estudo forem, por exemplo, da ordem de 300 ou 1000, o cálculo do desvio padrão já demandaria muito tempo para ser calculado manualmente ou com auxílio de uma calculadora comum. Portanto, o Excel permite a realização da estatística descritiva de forma ágil e rápida quando da realização de procedimentos estatísticos com vistas à realização de análises e inferência.

Ao acionar a sequência de comandos ao conjunto de dados 1, 3, 5, 6, e 8, o resultado gerado pelo programa é dado como na tabela a seguir.

FUNÇÃO ESTATÍSTICA	RESUMO ESTATÍSTICO
Média	4,6
Erro padrão	1,208304597
Mediana	5
Moda	#N/D
Desvio padrão	2,701851217
Variância da amostra	7,3
Curtose	-0,681178457
Assimetria	-0,182523257
Intervalo	7
Mínimo	1
Máximo	8
Soma	23
Contagem	5
Maior (1)	8
Menor (1)	1
Nível de confiança (95,0%)	3,354798334

#N/D: notação do Excel que, neste caso, significa não existir o valor modal.

Ilustração: Camila Leite

Note-se que a partir do resumo estatístico da tabela anterior, o pesquisador poderá realizar a análise dos resultados sobre: relações, variabilidades, médias quartis, entre outros, com vistas à interpretação e inferência.

Exercícios Propostos sobre Estatística Descritiva para Discussão no AVA

1 - Estabelecer quais das variáveis seguintes são discretas e quais são contínuas, justificando sua opção.

- Salários de empregados de uma indústria;
- Número de ações vendidas diariamente na Bolsa de Valores;
- Temperaturas registradas cada meia hora em um posto de Meteorologia;
- Comprimentos de 1000 parafusos produzidos numa fábrica;
- Número G de litros de água numa máquina de lavar roupa;
- Número B de livros em uma estante de biblioteca;
- Diâmetro D de uma esfera.

2 - Quais são as fases do método estatístico?

3 - Para os valores (escores) 205, 6, 5, 5, 5, 2 e 1 calcule a moda, a mediana e a média aritmética. Além disso, responda que medida de tendência central *não* deveria ser usada para descrever esse conjunto de escores? Por quê?

4 - Numa distribuição de valores iguais, o desvio-padrão é...

a) negativo;

b) positivo;

c) a unidade;

d) zero.

5 - Das afirmações abaixo:

A) Quando se ordenam valores não-agrupados segundo sua grandeza, a mediana é ponto médio desta série.

B) Quando os valores de uma série contínua estão agrupados em uma distribuição de frequência, a mediana é, por definição, o ponto que corresponde a 50% da distribuição.

C) Quando desejamos o ponto médio exato de uma distribuição de frequência, basta calcular a mediana.

D) Quando existem valores extremos que afetam muito o cálculo de uma média, para representá-la devemos dar preferência a mediana.

a. Todas incorretas

b. Todas corretas

c. Apenas **A** correta

d. Apenas **B** correta

e. Apenas **D** correta.

6 - Classifique as séries a seguir:

Produção de Café
Brasil 1991-1995

Anos	Produção (1000t)
1991	2.535
1995	2.666
1993	2.122
1994	3.750
1994	2.007

Fonte: IBGE
Classificação:

Preço do Acém no varejo
São Paulo - 1989/1993

Anos	Preço (US$)
1989	2,24
1990	2,73
1991	2,12
1992	1,89
1993	2,04

Fonte: APA
Classificação:

Duração Média dos Estudos
Superiores - 1994

Países	Anos
Itália	7,5
Alemanha	7,0
França	7,0
Holanda	5,9
Inglaterra	Menos de 4

Fonte: Revista Veja
Classificação:

Aquecimento de um motor
de avião de marca X

Minutos	Temperatura (°C)
0	20
1	27
2	34
3	41
4	49

Fonte: hipotética
Classificação:

Vacinação contra a
Poliomielite - 1993

Regiões	Quantidade
1989	2,24
1990	2,73
1991	2,12
1992	1,89
1993	2,04

Fonte: MS
Classificação:

Rebanhos Brasileiros - 1992

Espécie	Quantidade (1.000 Cabeças)
Bovinhos	154.440,8
Equinos	549,5
Suínos	34.532,2
Ovínos	19.955,9
Caprinos	12.159,6

Fonte: IBGE
Classificação:

Ilustração: Camila Leite

7 - Considere o faturamento, em milhões de reais, de algumas empresas brasileiras:

191 230 191 230 145 150 150 150 191 230 145 150 145 191 130

145 150 191 191 230 230 145 130 130 145 130 130 130 130 130

Elabore a tabela de distribuição de frequência com a frequência relativa em porcentagem e responda:

a) Qual é o faturamento que apresentou a maior frequência relativa?

b) Qual é a frequência relativa do maior faturamento registrado?

c) Qual faturamento apresentou a menor frequência absoluta?

8 - Os dados a seguir foram obtidos em indivíduos contaminados pelo veneno de um certo tipo de inseto e submetidos a tratamento. A variável de interesse Recup é definida como o tempo (em horas) entre a administração do tratamento e a recuperação do indivíduo. Os valores de Recup são os seguintes:

3, 90, 23, 46, 2, 42, 47, 37, 12, 51, 11, 1, 3, 3, 45, 3, 4, 11, 2, 8, 56, 39, 22, 16, 5 e 52.

a) Determine a média, mediana, desvio padrão e interprete os resultados.

b) Separe o conjunto de dados em três grupos denominados ***cura rápida***, com valor de Recup menor ou igual a 12, ***cura normal***, se o valor de Recup for maior do que 12 e menor ou igual a 45, e ***cura lenta***, se o valor de Recup estiver acima de 45. Compare a variabilidade desses três grupos através de seus coeficientes de variação.

9 - Durante o processo de certificação de qualidade de um laboratório de análise de água, o órgão certificador solicitou a dosagem de chumbo em várias alíquotas de uma amostra padrão de água. Os resultados das análises realizadas em triplicata por cinco analistas estão listados na tabela abaixo:

Analista	Concentração de Pb^{-2} (ppm)			Média	Desvio padrão
A	4,85	5,02	5,15	5,01	0,15
B	4,94	5,02	5,09	5,02	0,08
C	4,85	5,20	5,08	5,04	0,18
D	5,07	4,95	5,20	5,07	0,13
E	5,21	5,13	5,35	5,23	0,11

Ilustração: Camila Leite

a) Sabendo-se que o teor real de Pb^{-2} na amostra em questão é 5,00 ppm, qual analista apresentou resultado mais preciso?

b) E qual analista foi mais exato?

10 - Uma pesquisa de opinião foi realizada para avaliar os níveis de audiência de alguns canais de televisão, entre 20 h e 21 h, durante uma determinada noite. Os resultados obtidos estão representados no gráfico abaixo.

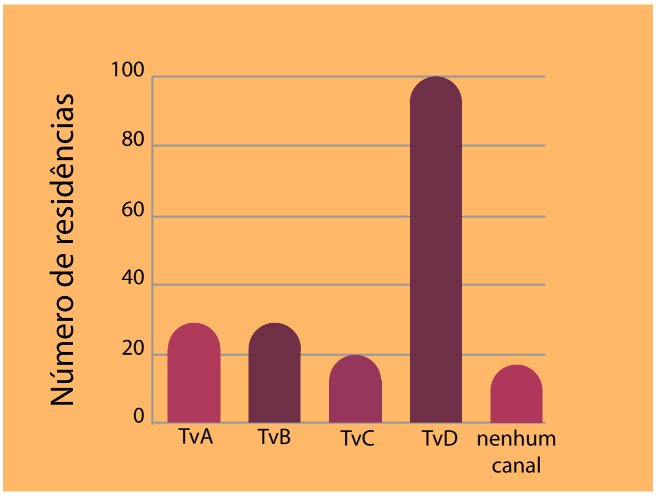


Ilustração: Camila Leite

A) (ENEM) O número de residências atingidas nessa pesquisa foi aproximadamente de:

a) 100 **c)** 150 **e)** 220

b) 135 **d)** 200

B) (ENEM) A percentagem de entrevistados que declararam estar assistindo a TV_B é aproximadamente igual a:

a) 15% **c)** 22% **e)** 30%

b) 20% **d)** 27%

C) Este gráfico é um gráfico do tipo:

a) barras **c)** setores **e)** linhas

b) colunas **d)** pizza

11 - Examine o gráfico abaixo. Depois responda **Verdadeiro ou Falso**, conforme o caso. No período considerado,

( ) houve um contínuo déficit na balança comercial brasileira

( ) a maior movimentação financeira ocorreu no ano de 1991

( ) houve um contínuo crescimento no valor das exportações

( ) os maiores saldos na balança comercial ocorreram em 1990 e 1993

( ) o menor valor de exportação brasileira verificou-se em 1990 e 1991

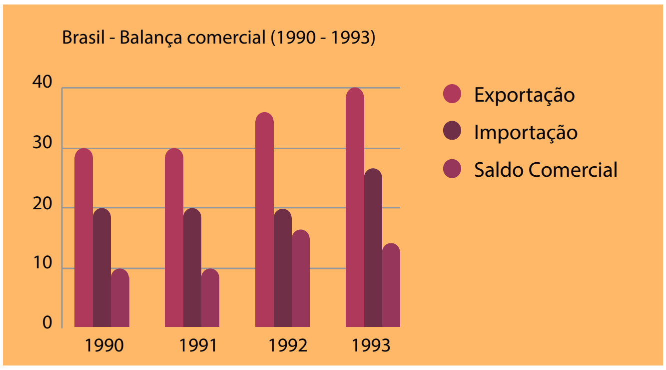


Ilustração: Camila Leite

12 - No gráfico de linhas abaixo, foi registrado o número de contatos em relação aos consórcios de carros e motos no período de 1995 à 2001.

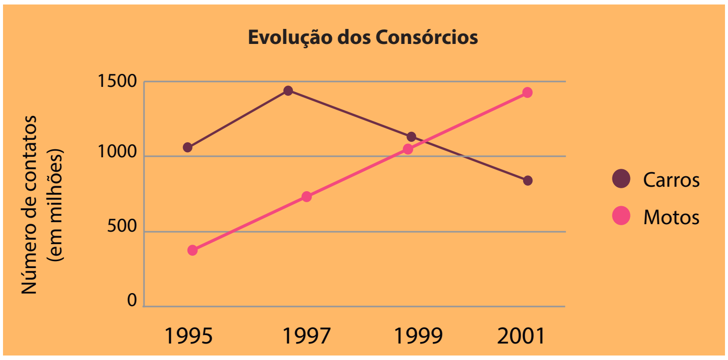


Ilustração: Camila Leite

Interpretando esse gráfico, podemos concluir que o número de contratos de motos ________________ (aumentou/diminuiu) desde 1995 e, em 2001, foi ________________ (maior/menor) que o número de contratos de carro. Também, em 2001, houve o _____________ (maior/menor) número de contratos.

13 - Em uma série de valores iguais onde o número de observações é ímpar, podemos afirmar que:

a) não existe mediana, pois os valores são todos iguais;

b) a média aritmética é menor que a mediana;

c) a mediana é menor que à média aritmética;

d) a mediana é igual a média aritmética;

e) não existe média aritmética

14 - Observe os conjuntos de dados abaixo, e depois marque a resposta correta:

I - 1, 3, 5, 9

II - 20, 14, 15, 19, 21, 22, 20

a) O variância de I é menor que a variância de **II**

b) O desvio padrão de I é maior que o desvio padrão de **II**

c) A média de **I** e **II** são iguais

d) O desvio padrão de **I** é menor que o desvio padrão de **II**

e) Os desvios padrões são iguais

15 - Calcule o desvio padrão de cada um dos grupos abaixo, depois responda:

Grupo A 7,5; 8,5; 6,0; 8,0; 5,0

Grupo B 5,5; 7,5; 9,0; 9,5; 4,5

a) O grupo A é mais homogêneo em relação ao grupo B

b) O grupo B é o mais homogêneo em relação ao grupo A

c) Nenhum grupo foi melhor que o outro

d) Os dois grupos foram igualmente homogêneos

16 - Considere a tabela 3.8 que representa os salários de funcionários de uma empresa de reciclagem.

Tabela: Salários da empresa de reciclagem Coisas & Tal

SALÁRIOS	FUNCIONÁRIOS
$500 \vdash 600$	3
$600 \vdash 700$	8
$700 \vdash 800$	12
$800 \vdash 900$	17
$900 \vdash 1000$	10
$1000 \vdash 1100$	8
$1100 \vdash 1200$	6
	$\sum f_i$ = **64**

Ilustração: Sofia Guimarães

Responda:

a) Qual o salário de 60% dos funcionários que ganham menos?

b) Qual o salário de 90% dos funcionários que ganham mais?

c) Qual o salário de 35% dos funcionários que ganham menos?

d) Qual o salário de 25% dos funcionários que ganham mais?

3. Teoria Elementar da Probabilidade

O termo probabilidade é usado de modo muito amplo na conversação diária para sugerir um certo grau de incerteza no qual ocorreu no passado, o que ocorrerá no futuro ou o que está ocorrendo no presente.

A ideia de probabilidade desempenha papel importante em muitas situações que envolvem uma tomada de decisão. Suponhamos que um empresário deseja lançar um novo produto no mercado. Ele precisará de informações sobre a probabilidade de sucesso para seu novo produto. Os modelos probabilísticos podem ser úteis em diversas áreas do conhecimento humano, tais como: Administração de Empresas, Economia, Psicologia, Biologia e outros ramos da ciência.

Disponível em < http://www.freepik.com/index.php?goto=41&idd=573692&url=aHR0cDovL3d3dy5zeGMuaH-UvcGhvdG8vMTIyMzE3Mw== >. Acesso em: 01/09/2017

3.1 Conceitos Básicos: Tipos De Experimentos

Definição 3.1.1: Determinísticos: *Ocorre quando, dadas as condições de experimentação, pode-se determinar ou predizer com certeza o resultado final do experimento, antes mesmo dele ocorrer. Exemplo: Estudar a queda e movimentos de corpos, saber o resultado da distância de um objeto quando aplicado certa força.*

Definição 3.1.2: Aleatórios: *São aqueles que não são possíveis determinar ou predizer com exatidão o resultado antes da realização do experimento. Exemplo: Investigação sobre o efeito de um novo tratamento em pacientes, efeito de um fertilizante químico no solo.*

Considere os seguintes experimentos:

E1. Jogar uma moeda e observar se dá cara ou coroa;

E2. Jogar um dado e observar a face voltada para cima;

E3. Inspecionar uma lâmpada, buscando determinar se está boa ou se tem defeito;

E4. Retirar uma carta de um baralho com 52 cartas e observar o seu naipe;

Características de um Experimento Aleatório

1. Cada experimento poderá ser repetido um grande número de vezes sob as mesmas condições;
2. Não podemos afirmar que um resultado particular ocorrerá, porém podemos descrever o conjunto de todos os resultados possíveis do experimento – as possibilidades de resultado;
3. Quando o experimento é repetido um grande número de vezes surgirá uma regularidade nos resultados. Esta regularidade, chamada de *Regularidade Estatística*, é que torna possível construir um modelo matemático preciso com o qual analisará o experimento.

Definição 3.1.3: Espaço Amostral: É um conjunto *de todos os possíveis resultados de um experimento aleatório.*

Exemplos:

a. E = jogar um dado e observar o nº da face de cima, então:

Ω = { 1, 2, 3, 4, 5, 6, }.

b. E = jogar duas moedas e observar o resultado, então:

Ω = { (c,c), (c,k), (K,c), (k,k) } em que c = cara e k = coroa.

Observe que sendo S um conjunto, poderá ser finito ou infinito, trataremos apenas dos conjuntos finitos.

Definição 3.1.4: Evento: *Evento é um conjunto de resultados do experimento; em termos de conjunto, é um subconjunto de .*

Em particular, e o conjunto vazio são eventos; é dito o evento certo e o evento impossível.

Exemplo: Seja o experimento E ; jogar um dado e observar o resultado. Então

Ω = {1, 2, 3, 4, 5, 6}. Cada subconjunto de constitui um evento.

A1 = {ponto 1} ou {1}

A2 = {ponto menor que 3} ou {1,2}

A3 = {ponto par} ou {2, 4, 6}

A4 = {ponto ímpar} ou {1 ,3, 5}

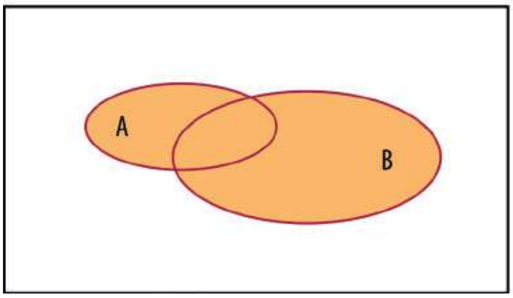
A5 = {divisores de 6} ou {1, 2, 3, 6}

A6 = {dos múltiplos de 1} ou {1, 2, 3, 4, 5, 6}

A7 = {dos pares múltiplos de 5} ou

Observação: Se o número de elementos do espaço amostral for **n**, então o número de eventos a ele associados é **2n.**

Definição 3.1.5: Evento União: *Chama-se evento união de dois eventos A e B, o evento AUB, formado pelos elementos comuns e não comuns aos eventos A e B. Ou seja, Dados os conjuntos A e B, define-se o conjunto união AUB = {x; xA ou xB}.*



$A \cup B$

Exemplos continuação:

$A_1 \cup A_2 = \{1,2\}$,

$A_2 \cup A_3 = \{1, 2, 4, 6\}$,

$A_3 \cup A_4 = \{1, 2, 3, 4, 5, 6\}$.

Propriedades imediatas:

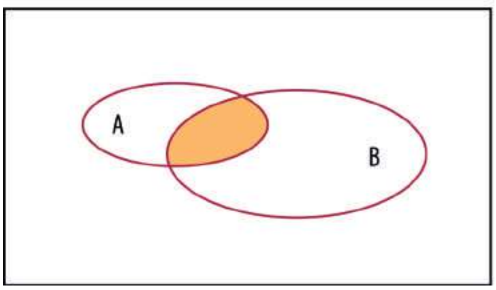
i) $A \cup A$ =A;

ii) $A \cup \phi$ = A;

iii) $A \cup B = B \cup A$;

iv) $A \cup \Omega = \Omega$.

Definição 3.1.6: Evento Intersecção: *Chama-se evento intersecção de dois eventos A e B, o evento A∩B formado pelos elementos comuns aos eventos A e B. Ou seja, Dados os conjuntos A e B , define-se o conjunto interseção A∩B ={x; xA e xB}.*



$A \cap B$

Exemplos continuação:

$A1 \cap A2 = \{1\}$;

$A3 \cap A5 = \{2, 6\}$;

$A1 \cap A3 = \emptyset$

Propriedades imediatas:

i) $A \cap A = A$;

ii) $A \cap \emptyset = \emptyset$;

iii) $A \cap B = B \cap A$;

iv) $A \cap \Omega = A$.

São importantes também as seguintes propriedades:

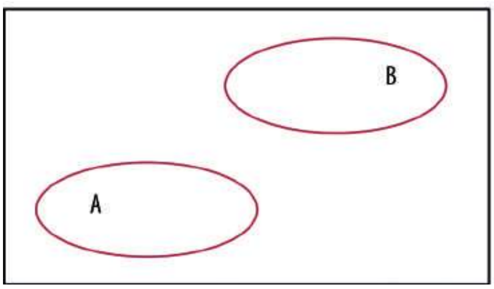
i) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (propriedade distributiva);

ii) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (propriedade distributiva);

iii) $A \cap (A \cup B) = A$ (lei da absorção);

iv) $A \cup (A \cap B) = A$ (lei da absorção).

Definição 3.1.7: Eventos Mutuamente Exclusivos: *Dois eventos são mutuamente exclusivos, se eles não puderem ocorrer simultaneamente, isto é, $A \cap B = \emptyset$.*



$$A \cap B = \emptyset$$

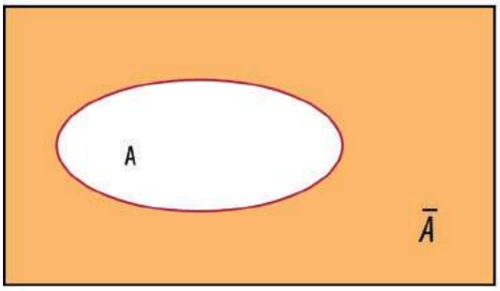
Exemplos:

A1 e A3 são mutuamente exclusivos,

A3 e A4 são mutuamente exclusivos,

A2 e A3 não são mutuamente exclusivos.

Definição 3.1.8: Evento Complementar: *Chama-se evento complementar de um evento A, denotado por* $\bar{A}$ *ou* A^c*, o conjunto , constituído pelos elementos* Ω *de que não pertencem ao conjunto A.*



Exemplos:

$A_1^c = \{2, 3, 4, 5, 6\}$

$A_3^c = \{1, 3, 5\}$

Definição 3.1.9: Evento Simples: *É o evento formado por um único ponto amostral.*

Exemplos:

A1={1}

A2={2}

A3={5}

A4={6}

Definição 3.1.10: Evento Certo: *É um evento formado por todos os pontos amostrais.*

Nesse caso, temos que o evento certo é o próprio espaço amostral Ω .

Definição 3.1.11: Evento Impossível: *É um evento que não possui elementos em* Ω .

Exemplo:

B={A1∩A3} = ∅

ATIVIDADE RESPONDIDA

1 - Determinar o espaço amostral relativo aos experimentos abaixo:

a. Três lançamentos consecutivos de uma moeda comum.

Solução: Sendo ca = cara e co = coroa, temos:

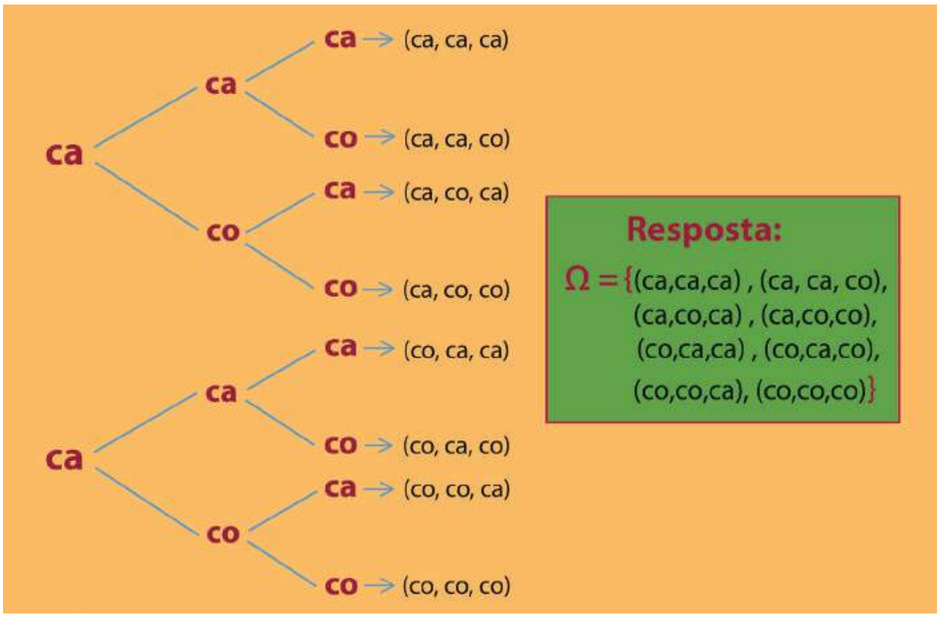


Ilustração: Orlando Dantas

b. Duas retiradas consecutivas e sem reposição de bolas de uma urna que contém 3 bolas brancas, 2 bolas azuis e 4 bolas vermelhas.

Solução:

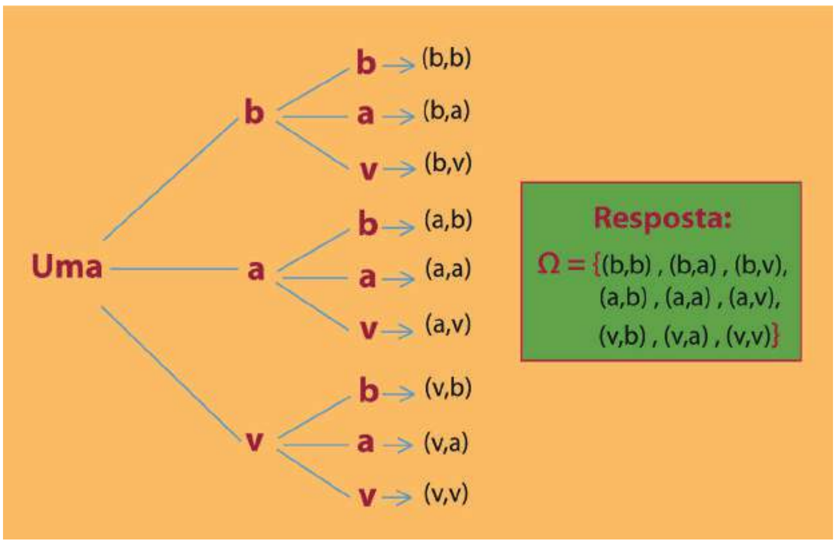


Ilustração: Orlando Dantas

ATIVIDADE

1 - Descrever o espaço amostral (Ω) e eventos (**A**) associados a cada um dos experimentos (**E**) a seguir:

E1: Lançam-se dois dados perfeitos e observam-se os números nas faces voltadas para cima;

A1: A soma das faces é sete

E2: Lançar uma moeda três vezes, sucessivamente, e anotar a sequência de caras (**K**) e coroas (**C**); **A2:** Sair pelo menos duas caras

E3: Lançar uma moeda e um dado, simultaneamente, e registrar os resultados;

A3: Obtenção de face ímpar no dado

E4: Numa linha de produção conta-se o número de peças defeituosas num período de 1 hora;

A4: Obter menos de 3 defeituosas

E5: Um fabricante produz um determinado artigo. Da linha de produção são retirados 3 artigos e cada um é classificado como bom(B) ou defeituoso(D).

A5: Pelo menos dois artigos são bons.

3.2 Definição Clássica de Probabilidade

Definição 3.2.1: *Seja um espaço amostral finito e equiprovável (aqueles espaços onde os eventos elementares possuem a mesma chance de ocorrerem.) e A um determinado evento ou seja, um subconjunto de Ω. A probabilidade de ocorrência do evento A será calculada pela fórmula:*

$$P(A) = \frac{n(A)}{n(\Omega)}$$

onde $n(A)$= número de elementos do evento A e $n(\Omega)$= número de elementos do espaço amostral Ω;

Observação: A probabilidade pode ser representada na forma de fração, número decimal ou em porcentagem.

Exemplos:

1 - No lançamento de um dado, determine a probabilidade de se obter:

a) o nº 2;

b) um nº par

c) um nº múltiplo de 3

Solução:

a) Temos que Ω={1,2,3,4,5,6} e seja A={*Onúmero*é2}. Portanto, A={2}

Percebe-se que é um espaço equiprovável, pois todos os elementos têm mesma probabilidade de ocorrência. Portanto, temos que

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{1}{6}$$

b) Seja B={*Onúmeroépar*. Portanto, B={2,4,6}. Logo, temos que

$$P(B) = \frac{n(B)}{n(\Omega)} = \frac{3}{6} = \frac{1}{2}$$

c) Seja C={*Onúmeroémúltiplode3*}. Portanto, C={3,6}.. Logo, temos que

$$P(C) = \frac{n(C)}{n(\Omega)} = \frac{2}{6} = \frac{1}{3}$$

2 - Um número inteiro é escolhido aleatoriamente entre os números 1, 2, 3, ... , 30. Qual a probabilidade de:

a) o número ser divisível por 5;

b) terminar em 3;

Solução:

a) Temos que Ω={1,2,...,30} e percebe-se que é um espaço equiprovável. Seja $A=\{Onúmeroserdivisívelpor5\}$. Logo $A=\{5,10,15,20,25,30\}$. Portanto, temos que:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{6}{30} = \frac{1}{5}$$

b) Seja $B=\{Onúmeroterminaem3\}$. Logo $B=\{3,13,23\}$. Portanto, temos que

$$P(B) = \frac{n(B)}{n(\Omega)} = \frac{3}{30} = \frac{1}{10}$$

Axiomas de Probabilidade

Sejam e eventos de um mesmo espaço amostral Ω. Dizemos que um número real p, que pode ser P(A) ou P(B), representa uma probabilidade, se e somente se, p satisfizer às seguintes condições:

1. $0<p<1$;
2. $P(\Omega)=1$;
3. Se Ae Bsão eventos mutuamente exclusivos, então $P(A \cup B)=P(A)+P(B)$

Teoremas da Probabilidade

1. Se $\emptyset$ é um evento impossível então $P(\emptyset) = 0$;
2. Se A^c é o complementar de A, então $P(A) = 1 - P(A)$;
3. Se A e B são eventos quaisquer em Ω, então $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Exemplo:

1 - Seja **E:** Lançar um Dado.

A= { sair um número 3}

B= { sair um nº par }

C= { sair um nº ímpar}

Temos que Ω={1,2,3,4,5,6}. Portanto, temos que

a) $P(A)=\frac{n(A)}{n(\Omega)}=\frac{1}{6}$;

b) $P(B)=\frac{3}{6}=\frac{1}{2}$;

c) $P(C)=\frac{3}{6}=\frac{1}{2}$;

d) $P(A\cup B)=P(A)+P(B)$,pois $P(A\cap B)=\emptyset$. Logo,

$$P(A\cup B)=\frac{1}{6}+\frac{1}{2}=\frac{2+6}{12}=\frac{8}{12}=\frac{2}{3} \text{ ;}$$

e) $P(A\cap C)$, temos que o conjunto $\{A\cap C\}=\{3\}$. Portanto temos que

$$P(A\cap C)=\frac{n(A\cap C)}{n(\Omega)}=\frac{1}{6} \text{ ;}$$

f) $P(A\cup C)=P(A)+P(C)-P(A\cap C)=\frac{1}{6}+\frac{1}{2}-\frac{1}{6}=\frac{1}{2}$;

g) $P(A^c)=1-P(A)=1-\frac{1}{6}=\frac{7}{6}$.

2 - Expresse os seguintes eventos em termos de conjunto:

I) A ocorre, mas B não ocorre: $\{A\cap B^c\}$;

II) A ou B ocorrem; $\{A\cup B\}$;

III) B e C ocorrem; $\{B \cap C\}$;

IV) A e C ocorrem, mas B não ocorre: $\{A \cap C \cap B^C\}$

3 - Quais dos eventos A, B e C são mutuamente exclusivos?

Nesse caso, temos que os eventos $\{A \cap B\}$ e $\{B \cap C\}$ são mutuamente exclusivos, pois $\{A \cap B\} = \{B \cap C\} = \emptyset$.

4 - Em uma sala de aula, a professora de Matemática decidiu fazer um levantamento dos lanches comprados pelos alunos. A professora verificou que, de um total de 35 alunos, dezenove compraram salgado; destes, quatro compraram pizza e salgado, e sete alunos não compraram lanche nesse dia. Qual a probabilidade de selecionar um aluno que comprou apenas pizza?

Solução: Primeiro precisamos saber qual a quantidade de alunos que compraram apenas pizza.

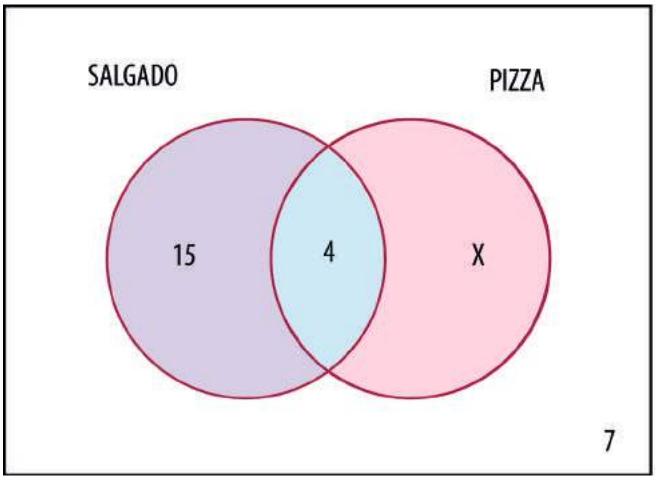


Ilustração: Sofia Guimarães

Observe a figura abaixo e a compare com as informações do enunciado. Fazer isto poderá lhe ajudar na resolução de outros problemas:

Somando os alunos que não compraram lanche, aqueles que compraram apenas pizza, os que compraram apenas salgado e os que compraram salgado e pizza, o resultado deve ser 35. Sendo assim, chegamos à seguinte equação:

$$x + 15 + 4 + 7 = 35$$
$$x = 35 - 26$$
$$x = 9$$

Portanto, nove alunos compraram apenas pizza. Assim, seja = Quantidade de alunos que compraram apenas pizza. Temos então que n(A)=9. Assim

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{9}{35}$$

5 - Em um curso de idiomas, foi feita uma pesquisa com adolescentes para verificar quais línguas estrangeiras eles gostariam de aprender. O resultado foi:

- 23 gostariam de aprender inglês;
- 24 gostariam de aprender espanhol;
- 25 gostariam de aprender italiano;
- 12 gostariam de aprender inglês e italiano;
- 10 gostariam de aprender italiano e espanhol;
- 9 gostariam de aprender inglês e espanhol;
- 7 gostariam de aprender inglês, espanhol e italiano.

Qual a probabilidade de selecionar um aluno que gostaria de aprender inglês?

Solução: Seja A={*OalunogostariadeaprenderInglês*}. Nesse caso temos que n(A)=23. No entanto, precisamos saber qual a quantidade de adolescentes que foram entrevistados.

Vamos montar um diagrama de Venn para organizar as informações fornecidas pelo problema. Dessa forma, há três círculos que se interceptam, e a intersecção dos três círculos é composta pelo número 7, pois essa é a quantidade de adolescentes que se interessaram pelas três línguas.

Em relação aos adolescentes que se interessaram por dois idiomas, doze interessaram-se por inglês e italiano e, dentre esses, sete interessaram-se por espanhol também. Logo, apenas cinco gostariam de fazer inglês e italiano. Assim, se 10 gostariam de aprender italiano e espanhol, retirando os sete que se interessaram pelas três línguas, restam apenas três que gostariam de aprender italiano e espanhol. De modo análogo, sabemos que nove pessoas interessaram-se por inglês e espanhol. Desconsiderando as sete que se interessaram por inglês também, restam apenas dois que gostariam de aprender inglês e espanhol.

Vamos analisar quantos gostariam de aprender apenas um idioma. Sabemos que 23 gostariam de aprender inglês e, desconsiderando os dois que se interessaram também por espanhol, cinco, por italiano, e sete, por ambos, restam apenas nove que gostariam de aprender apenas inglês. Se 24 gostariam de aprender espanhol e, desses, dois interessaram-se por inglês, três, por italiano, e sete, por ambos, 12 adolescentes gostariam de aprender apenas espanhol. Se dos 25 que se interessaram por italiano, retirarmos os cinco que também se interessaram por inglês, os três que gostariam de aprender espanhol e os sete que têm interesse em todas as línguas, restaram dez adolescentes que gostariam de aprender exclusivamente italiano. Com essas informações, podemos montar o seguinte diagrama de Venn:

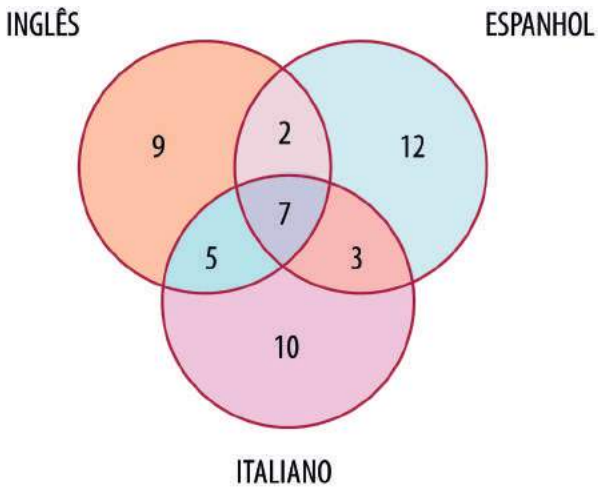


Ilustração: Sofia Guimarães

Somando todas as quantidades apresentadas no diagrama, temos:

9 + 2 + 7 + 5 + 12 + 3 + 10 = 48.

Portanto, 48 adolescentes foram entrevistados pelo curso de idiomas. Assim, temos que

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{23}{48}$$

ATIVIDADE

1 - Numa cidade, 30% dos homens são casados, 40% são solteiros, 20% são desquitados e 10% são viúvos. Um homem é escolhido ao acaso. Qual a probabilidade;

a) de ele ser solteiro? **R: 40%**

b) dele não ser casado? **R: 70%**

c) de ele ser solteiro ou desquitado? **R: 60%**

2 - Um colégio tem 1000 alunos. Destes: 200 estudam matemática, 180 estudam física, 200 estudam química, 20 estudam matemática, física e química 50 estudam matemática e física, 50 estudam física e química e 70 estudam somente química. Um aluno do colégio é escolhido ao acaso. Qual a probabilidade de;

a) ele estudar só matemática? **R: 7%**

b) ele estudar só física? **R: 10%**

c) ele estudar matemática e química? **R: 10%**

3 - Um empresário supersticioso, quando necessita viajar, escolhe o modelo de transporte através do lançamento de uma moeda: se sair cara viaja de ônibus, se sair coroa, viaja de avião. Numa semana em que tiver de fazer exatamente 4 viagens, ache a probabilidade de que ele faça:

a) Nenhuma de avião; **R: 6,25%**

b) Duas de ônibus; **R: 37,5%**

c) Pelo menos uma de avião. **R: 93,75%**

4 - Um trabalhador possui 4 calças (azul, preta, marrom, cinza) e 3 camisas (branca, azul, cinza) que podem ser utilizadas no trabalho. Este trabalhador se veste de forma aleatória. Ache a probabilidade de em um dia utilizar:

a) calça e camisa de mesma cor; **R: 16,67%**

b) não usar calça marrom; **R: 75%**

c) usar calça ou camisa azul. **R: 50%**

5 - Para fazer escolha do dia da semana (6 dias úteis) de folga de um funcionário o Diretor de uma empresa joga um dado e a face voltada para cima indicará o dia de folga. Sendo que nestas empresas trabalha um casal (Esposo/Esposa). Em uma semana ache a probabilidade de que, para este casal ocorra a seguinte folga:

a) esposo e esposa no mesmo dia; **R: 16,67%**

b) esposa antes do esposo; **R: 41,67%**

c) esposo e esposa em dias sequenciais. **R: 27,78%**

6 - Para conduzir o destino de um empresa, existem 4 pessoas disponíveis a saber: Alfredo, Beatriz, Carlos e Daniel. Os cargos existentes são: Presidente e Tesoureiro. Se a escolha for feita através de sorteio, ache a probabilidade de que:

a) presidente seja o Alfredo ou a Beatriz; **R: 50%**

b) tesoureiro não seja mulher; **R: 75%**

c) Carlos ou o Daniel fique de fora. **R: 16,67%**

3.3 Probabilidade Condicional

Definição 3.3.1: Probabilidade Condicional: *É a probabilidade de ocorrer determinado evento sob dada condição, ou seja, para dois eventos quaisquer AeB , sendo P(B)≠0, definimos a probabilidade condicional de A, dado que o evento B ocorreu, P(A/B) , como sendo:*

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Exemplos:

1 - A tabela a seguir mostra os resultados de um levantamento no qual foi indagado a 102 homens e 103 mulheres, trabalhadores, com idade entre 25 e 64 anos, se tinham poupado para emergência pelo menos um mês de salário.

POUPADO	SEXO		TOTAL
	HOMENS	MULHERES	
MENOS DE UM SALÁRIO MENSAL	47	59	106
UM SALÁRIO MENSAL OU MAIS	55	44	99
TOTAL	102	103	205

Ilustração: Sofia Guimarães

a) Dado que um trabalhador selecionado ao acaso é homem, obtenha a probabilidade dele ter poupado menos de um salário mensal:

Solução:

Seja A={*Umtrabalharterpoupadomenosdeumsaláriomensal*};

Seja B={*Otrabalhoéhomem*}.

Temos então que a Probabilidade pedida é dada por:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Nesse caso, temos que $P(B) = \frac{102}{205}$, e temos que $P(A \cap B) = \frac{47}{205}$.

Portanto temos que

$$P(A/B) = \frac{\frac{47}{205}}{\frac{102}{205}} = \frac{47}{102} = 0{,}4608$$

b) Dado que um trabalhador poupou um mês ou mais, obtenha a probabilidade de se tratar de uma mulher; (0,4444)

Solução:

Seja A={*Otrabalhadorémulher*};

Seja B={*Umtrabalhadorpoupouummêsoumais*}.

Temos então que a Probabilidade pedida é dada por:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Nesse caso, temos que $P(B) = \frac{99}{205}$, e temos que $P(A \cap B) = \frac{44}{205}$.

Portanto temos que

$$P(A/B) = \frac{\frac{44}{205}}{\frac{99}{205}} = \frac{44}{99} = 0{,}4444$$

ATIVIDADE PARA DISCUSSÃO NO AVA

1 - Numa certa faculdade, 25% dos estudantes foram reprovados em Matemática, 15% dos estudantes foram reprovados em Química e 10% dos estudantes foram reprovados tanto em matemática como em Química. Um estudante é selecionado ao acaso.

a) Se ele é reprovado em Matemática, qual a probabilidade de que ele tenha sido reprovado em Química? **R: 40%**

b) Se ele foi reprovado em Química, qual a probabilidade de que ele tenha sido reprovado em Matemática? **R: 66,67%**

c) Qual a probabilidade de que ele tenha sido reprovado em Matemática ou Química? **R: 30%**

2 - . Um grupo de pessoas está classificado da seguinte maneira:

	Professor	Advogado	Dentista
Homem	60	80	50
Mulher	90	40	30

Considerando: **H** = a pessoa é homem ; **M** = a pessoa é mulher ; **P** = a pessoa é professor; **A** = a pessoa é advogado; **D** = a pessoa é dentista. Calcule cada uma das probabilidades:

a) P(A/H) =

b) P(P/M) =

c) P(D/H) =

d) P(A/M) =

4 - Uma pesquisa realizada entre 1000 consumidores, registrou que 650 deles trabalham com cartões de crédito da bandeira *MasterCard*, que 550 trabalham com cartões de crédito da bandeira VISA e que 200 trabalham com cartões de crédito de ambas as bandeiras. Qual a probabilidade de ao escolhermos deste grupo uma pessoa que utiliza a bandeira *Visa*, ser também um dos consumidores que utilizam cartões de crédito da bandeira *MasterCard*? **(Dica: Use o Diagrama de Venn)**

3.4 Independência Estatística

Definição 3.4.1: *Dados dois eventos A e B de um espaço amostral Ω, dizemos que A independe de B quando a probabilidade de que eles ocorram simultaneamente for igual ao produto de suas probabilidades, ou seja, A e B são independentes quando*

$$P(A \cap B) = P(A).P(B)$$

Exemplo:

Seja o experimento lançar dois dados. Seja:

A: sair o número 6 no primeiro dado;

B: sair o número 3 no segundo dado.

Então Ω= { (1,1);(1,3); . . . ;(6,6)}

A = {(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

B= {(1,3),(2,3),(3,3),(4,3),(5,3),(6,3)}

A∩B= {(6,3)}

Então $P(A \cap B) = \frac{1}{36}$, e temos que $P(A) = \frac{6}{36}$ e $P(B) = \frac{6}{36}$, ou seja,

$$P(A \cap B) = P(A).P(B) = \frac{6}{36}.\frac{6}{36} = \frac{1}{36}$$

Observação: Da definição de probabilidade condicional (**Definição 3.3.1**), segue-se facilmente que se *A* e *B* são dois eventos independentes, com probabilidades positivas, teremos: P(A/B)=P(A)e P(B/A)=P(B)

Nota: Independência é hipótese (não é de natureza estatística).

3.5 Teorema do Produto

Uma conseqüência importante da definição formal de probabilidade condicional é a seguinte: Se *A* independe de *B*, logo *B* independe de *A*, então:

$$P(A \cap B) = P(B).P(A/B)$$

Isto é, a probabilidade de ocorrência simultânea de dois eventos $P(A \cap B)$ é o produto da probabilidade de um deles pela probabilidade do outro, dado o primeiro. pelo que uma generalização natural conduz a:

. $P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1).P(A_2/A_1).P(A_3/A_1 \cap A_2)\ldots P(A_n/A_1 \cap A_2 \cap \ldots \cap A_{n-1})$

Exemplos:

1 - Uma moeda é lançada 3 vezes. Sejam os eventos:

A: ocorrem pelo menos duas caras;

B: ocorrem resultados iguais nos 3 lançamentos.

Verificar se os eventos são independentes.

Solução:

Primeiro vamos calcular a probabilidade do evento *A*. Temos que o evento *A* é dado por

A={(C,C,K),(C,K,C),(K,C,C),(C,C,C)}. Portanto $P(A) = \frac{4}{8}$. Uma vez que

Ω={(C,C,C),(C,C,K),...,(K,K,K)} no qual se tem 8 elementos.

Vamos calcular agora a probabilidade do evento *B*. Temos que o evento *B* é dado por:

B={(C,C,C),(K,K,K)}. Então, temos que $P(B) = \frac{2}{8}$

Temos que o evento $A \cap B = \{(C,C,C)\}$, e temos que a probabilidade desse evento é dada por

$P(A \cap B) = \frac{1}{8}$,

Temos que para os eventos *A* e *B* serem independentes é preciso que

$P(A \cap B) = P(A).P(B)$. E temos que $P(A).P(B) = \frac{4}{8}.\frac{2}{8} = \frac{1}{8} = P(A \cap B)$.

Logo, os eventos *A* e *B* são eventos independentes.

2 - Um lote contém 50 peças boas (B) e 10 defeituosas (D). Uma peça é escolhida ao acaso e, sem reposição desta, outra peça é escolhida ao acaso. Qual a probabilidade de ambas serem defeituosa?

Solução:

Seja $D_i = \{a peça i é defeituosa\}, i = 1,2$.

Temos então que a probabilidade pedida é dada por

$$P(D_1 \cap D_2) = P(D_1).P(D_2/D_1) = \frac{10}{60}.\frac{9}{59} = \frac{3}{118}$$

ATIVIDADE PARA DISCUSSÃO NO AVA

1 - Considere uma urna contendo 3 bolas pretas e 5 bolas vermelhas. Retire duas bolas da urna, sem reposição.

a) Obtenha os resultados possíveis e as respectivas probabilidades;

b) Mesmo problema, para extrações com reposição.

2 - No problema anterior, calcule as probabilidades dos eventos:

a) bola preta na primeira e segunda extração;

b) bola preta na segunda extração;

c) bola vermelha na primeira extração.

3 - A probabilidade de que A resolva um problema é de ⅔ e a probabilidade de que B resolva é de ¾. Se ambos tentarem independentemente, qual a probabilidade do problema ser resolvido? (91,67%)

4 - Um restaurante popular apresenta apenas dois tipos de refeições: salada completa ou um prato à base de carne. 20% dos fregueses do sexo masculino preferem salada; 30% das mulheres escolhem carne; 75% dos fregueses são homens. Considere os seguintes eventos:

H: freguês é homem A: freguês prefere salada
M: freguês é mulher B: freguês prefere carne.

Calcular:

a) P(H), P(A/H), P(B/M); **R: 75%, 20%, 30%**

b) P(AH), P(AH); **R: 15% ,92,5%**

c) P(M/A); **R: 53,85%**

5 - Uma companhia de seguros analisou a frequência com que 2.000 segurados (1.000 homens e 1.000 mulheres) usaram o hospital. Os resultados são apresentados na tabela:

	Homens	Mulheres
Usaram o hospital	100	150
Não usaram o hospital	900	850

a) Qual a probabilidade de que uma pessoa segurada use o hospital? **R: 12,5%**

b) O uso do hospital independe do sexo do segurado?

6 - Numa cidade, 20% da população são mulheres que não podem votar (menores de 18 anos). Se 60% da população são mulheres, qual é a probabilidade de que uma mulher selecionada ao acaso não possa votar? **R: 33,33%**

7 - Um levantamento revela as seguintes informações sobre um grupo de pessoas:

	Gosta de música	Gosta de TV	Gosta de cinema
Homem	50	40	30
Mulher	30	60	40

Considerando: **H:** a pessoa é homem; **M:** a pessoa é mulher; **A:** a pessoa gosta de música; **B:** a pessoa gosta de TV e **C:** a pessoa gosta de cinema, determine:

a) P(M/C); **R: 57,14%**

b) P(B/M); **R: 46,15%**

c) P(H/A); **R: 62,5%**

d) P(A/H); **R: 41,67%**

e) P(C/H); **R: 25%**

8 - Uma família planeja ter 3 crianças. Qual é a probabilidade de que a família tenha exatamente 2 meninas, dado que a primeira criança que nasceu é menina?

9 - Considere uma experiência aleatória e dois acontecimentos A e B tais que o Diagrama de Venn correspondente é:

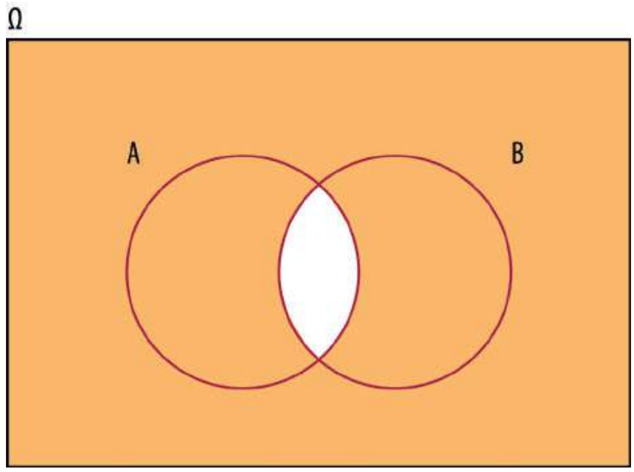


Ilustração: Sofia Guimarães

Indique o valor lógico das seguintes afirmações:

a) A região sombreada corresponde a não se realizar nenhum dos dois acontecimentos.
b) A e B são acontecimentos independentes.

c) A região sombreada corresponde a que se realize pelo menos um dos dois acontecimentos.

d) A região sombreada corresponde a que se realize quanto muito um único acontecimento.

10 - Considere uma experiência aleatória e dois acontecimentos A e B tais que o Diagrama de Venn correspondente é:

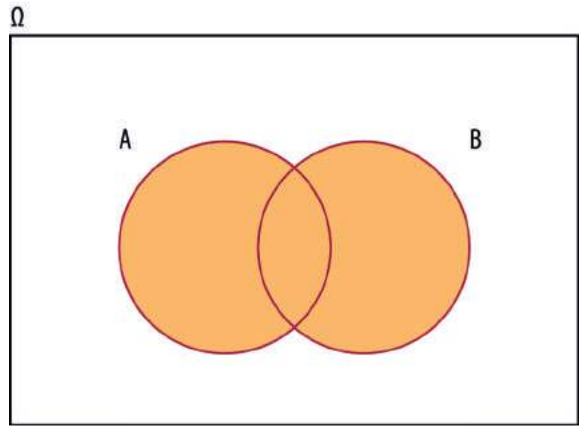


Ilustração: Sofia Guimarães

Indique o valor lógico das seguintes afirmações:

a) A e B são acontecimentos mutuamente exclusivos.

b) A e B podem ser acontecimentos independentes.

c) A probabilidade de se realizar apenas um dos acontecimentos é: P(A)+P(B)-2P(AB).

d) A região sombreada corresponde a realizar-se um e um só dos dois acontecimentos.

Sabendo um pouco mais

Link para auxílio com questões, simulados e vídeo-aulas de Probabilidade

http://www.soexercicios.com.br/plataforma/questoesSemelhantes/30678/ENEM/-diagrama-de-venn-probabilidade

ANOTAÇÕES

ANOTAÇÕES

Referências Bibliográficas

Barbetta, Pedro Alberto. *Estatística Aplicada às Ciências Sociais (5ª edição revisada).* Editora da UFSC. Florianópolis (SC), 2003.

Bunchaft, Guenia & Kellner, Sheilah Ribno de Oliveira. *Estatística Sem Mistérios (2ª edição corrigida)* vol. I, II, III e IV. Editora Vozes. Petrópolis (RJ), 1999.

Davis, J. A. *Levantamento de Dados em Sociologia: uma análise estatística elementar.* Rio de Janeiro – RJ: Zahar.

Hoel, Paul.G. *Estatística Elementar.* Rio de Janeiro, Editora Atlas, 1989.

Larson, Ron; Farber, Betsy. *Estatística aplicada.* 4. ed. São Paulo: Pearson Prentice Hall, 2010.

Meyer, P. L., *Probabilidades. Aplicações à Estatística.* Livros Técnicos e Científicos Editora S.A., Rio de Janeiro, 1981.

Montgomery, D.C. e Runger, G. C., *Applied Statistics and Probability for Engineers.* John Wiley & Sons, New York, 2ª Edição, 1999.

Morettin, Pedro Alberto; Bussab, Wilton de Oliveira. *Estatística básica* . 6. ed. São Paulo: Saraiva, 2010. 540 p.

Mundim, M. J. *Estatística com o BrOffice*. Editora Ciência Moderna, 2010.

Murteira, Bento J., *Probabilidades e Estatística.* Vol. I e Vol. II, 2ª Edição, McGraw-Hill de Portugal Ltda, Lisboa, 1990.

Oliveira L. A . *Estatística Aplicada À Educação* - Apostila, Des-UFSCar.

Rosental, Claude. & Frémontier-Murphy, Camille. *Introdução aos Métodos Quantitativos em Ciências Sociais.* Editora Instituto Piaget. Lisboa – Portugal, 2001.

Ross, Sheldon M., *Introduction to Probability and Statistics for Engineers and Scientists.*John Wiley & Sons, 1987.

Triola, Mario F. *Introdução a estatística*. 7. ed. Rio de Janeiro, RJ: Livros Técnicos e Científicos, 1999. 410 p.

Universidade Federal da Bahia

Estatísticas para as Ciências Sociais Aplicadas I

O módulo trata de pontos importantes no ensino da estatística dentro do curso de Ciências Sociais. Trataremos da grande importância da Estatística Descritiva e mostraremos o quão importante é saber sumarizar, e como sumarizar, uma grande quantidade de informações. Mostrarei conteúdos relacionados à estatística, para que eles servem e aonde vamos usá-los. O material foi preparado para que um profi ssional das Ciências Sociais pudesse conhecer um pouco da estatística e assim utilizá-la dentro do ensino e também dentro da pesquisa.

PROGRAD
PRÓ-REITORIA DE GRADUAÇÃO
Ciências Contábeis
UNIVERSIDADE FEDERAL DA BAHIA
SEAD
Superintendência de Educação a Distância | UFBA
UAB UNIVERSIDADE ABERTA DO BRASIL
NELT NÚCLEO DE ESTUDOS DE Linguagens & Tecnologias