Estatística Básica com **Excel**

Flavio Alves Pozzi

Flavio Alves Pozzi
Estatística Básica com Excel

EDITORA FAEL

Estatística Básica com **Excel**

Flavio Alves Pozzi

2ª Edição

Curitiba
2018

EDITORA FAEL

Ficha Catalográfica elaborada pela Fael. Bibliotecária – Cassiana Souza CRB9/1501

P893e Pozzi, Flavio Alves
Estatística básica com excel / Flavio Alves Pozzi. – 2. ed. – Curitiba: Fael, 2018.

214 p.: il.

ISBN 978-85-5337-025-2

1. Estatística – Conceitos introdutórios 2. Excel – Conceitos introdutórios I. Título

CDD 519.5

FAEL

Direção Acadêmica	Francisco Carlos Sardo
Coordenação Editorial	Raquel Andrade Lorenz
Revisão	Veridiana Almeida
Projeto Gráfico	Sandro Niemicz
Imagem da Capa	Shutterstock.com/pedrosek
Arte-Final	Evelyn Caroline dos Santos Betim

Apresentação

A UTILIZAÇÃO DA Estatística é cada vez mais acentuada em qualquer atividade profissional da vida moderna. Nos seus mais diversificados ramos de atuação, as pessoas estão frequentemente expostas à Estatística, utilizando-a com maior ou menor intensidade. Isto se deve às múltiplas aplicações que o método estatístico proporciona àqueles que dele necessitam.

HÁ, EVIDENTEMENTE, A necessidade de especiais cuidados no manejo e na interpretação da Estatística; a interpretação não é monopólio dos estatísticos, sendo natural que, possuindo um maior conhecimento das técnicas estatísticas, levem vantagens no tocante à apreciação, análise e interpretação dos dados estatísticos. O raciocínio claro é indispensável para interpretar estatísticas, requerendo uma disposição mental receptiva e crítica.

Raramente, ou nunca, os dados estatísticos falam por si mesmos. A coisa mais importante acerca da interpretação dos dados estatísticos é saber que, se forem habilmente coletados e criticamente analisados, podem ser extremamente úteis.

Há muito tempo sinto a necessidade de um livro-texto claro e compreensível sobre estatística. O material pretende ser de fácil leitura, compreensão e assimilação. Além disso, deve ser interessante, e não maçante, contendo numerosos exemplos e aplicações.

Este livro foi escrito com base nessas ideias, procurei apresentar os diversos tópicos de maneira clara e suave, evitando, quando possível, incluir teoremas e demonstrações. Quando da necessidade de cálculos mais avançados foi apresentado o Excel para facilitar o cálculo, para que você pudesse prestar atenção nos resultados e nas interpretações, que são muito importantes para o analista.

O objetivo deste livro é o de apresentar uma introdução moderna à estatística e à probabilidade com um importante diferencial, que é a utilização do Excel como ferramenta de cálculo e, desenvolver no leitor, uma compreensão intuitiva dos assuntos tratados. Por conveniência, o livro está dividido em duas partes. A primeira refere-se à estatística e, a segunda, diz respeito à probabilidade.

O livro destina-se a ser usado como livro-texto para um curso formal em estatística e probabilidade ou como ferramenta de consulta para alunos e/ou profissionais que desejam aprender a utilizar não só as ferramentas da estatística, mas também utilizá-las no Excel.

Bem-vindo à estatística! Você está iniciando uma interessante e significativa aventura, pois começa a explorar um dos mais básicos instrumentos da tomada de decisão.

Flavio Alves Pozzi

Sumário

Glossário | 7

1 Introdução à Estatísitca | 9

2 Distribuição de Frequências | 37

3 Apresentação de Dados | 57

4 Medidas de Posição | 75

5 Medidas de Dispersão e Assimetria | 103

6 Correlação e regressão | 123

7 Introdução Geral de Probabilidade | 135

8 Probabilidade Condicionada | 169

9 Distribuição Binomial, Normal e Poisson | 181

Referências | 213

Glossário

Glossário de Símbolos

- $\sum$ – Somatório. Muitos procedimentos estatísticos necessitam do cálculo da soma de um conjunto de números, isto é feito da seguinte maneira: $\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$
- x_j – valores da variável estudada.
- f_j – Frequência em que os valores da variável estudada ocorreu, frequência simples absoluta, número de observações ocorridas de um valor.
- fr_j – Frequência simples relativa.
- F_j – Frequência absoluta acumulada.
- Fr_j – Frequência relativa acumulada.
- n – é o número total de valores observados.
- k – número de classes
- At – Amplitude total do universo ou da amostra.
- LI – Limite inferior de classe (para cada classe).
- LS – Limite superior de classe (para cada classe).
- Ac – Amplitude do intervalo de classe (para cada classe).
- PM_j – Ponto médio de classe (para cada classe).

1

Introdução à Estatística

A UTILIZAÇÃO DA Estatística é cada vez mais acentuada em qualquer atividade profissional da vida moderna. Nos seus mais diversificados ramos de atuação, as pessoas estão frequentemente expostas à Estatística, utilizando-a com maior ou menor intensidade. Isto se deve às múltiplas aplicações que o método estatístico proporciona àqueles que dele necessitam.

GERALMENTE, QUANDO APRESENTAMOS uma citação estatística, somos levados, de pronto, a desacreditar em qualquer argumentação em contrário destituída de base numérica.

DEVEMOS SER FRONTALMENTE contra os que afirmam que a "Estatística é capaz de provar qualquer coisa", ou que a "Estatística não prova coisa alguma".

Aquele que aceita dados estatísticos indiscriminadamente, muitas vezes, se deixará enganar sem necessidade; também aquele que rejeita qualquer informe estatístico de pronto estará dando prova de ignorância.

Atualmente, o público leigo (leitor de jornais e revistas) posiciona-se em dois extremos divergentes e igualmente errôneos quanto à validade das conclusões estatísticas: ou crê em sua infalibilidade ou afirma que elas nada provam. Os que assim pensam ignoram os objetivos, o campo e o rigor do método estatístico; ignoram a Estatística quer teórica quer prática, ou a conhecem muito superficialmente.

Há, evidentemente, a necessidade de especiais cuidados no manejo e na interpretação da Estatística; a interpretação não é monopólio dos estatísticos, sendo natural que, possuindo um maior conhecimento das técnicas estatísticas, levem vantagens no tocante à apreciação, análise e interpretação dos dados estatísticos. O raciocínio claro é indispensável para interpretar estatísticas, requerendo uma disposição mental receptiva e crítica.

Raramente, ou nunca, os dados estatísticos falam por si mesmos. A coisa mais importante acerca da interpretação dos dados estatísticos é saber que, se forem habilmente coletados e criticamente analisados, podem ser extremamente úteis.

Existem muitas concepções errôneas acerca da natureza desta disciplina. A ideia de que um leigo possa fazer da Estatística talvez seja diferente da ideia de um profissional. É comum, por exemplo, as pessoas formarem conceitos distorcidos a respeito de um estatístico profissional. Para alguns, trata-se de um indivíduo que tem a capacidade de manipular números para demonstrar seus pontos de vista. Alguns estudantes, por outro lado, tendem a admiti-lo como alguém que, auxiliado por sua calculadora, tem a faculdade de converter qualquer assunto em um estudo "científico".

Toda essa aura criada em torno da disciplina tem provocado, em estudantes e profissionais, uma dupla atitude: de apreensão, quanto à dificuldade de absorção de seu conteúdo, e de expectativa, quanto sua potencialidade como instrumento auxiliar de resolução de problemas.

Por essa razão, é extremamente difícil apresentar uma definição de Estatística, além do que muitos de seus conceitos fundamentais não apresentam uma definição explícita, ou, se a apresentam, esta não se revela suficientemente clara para dar uma ideia definitiva de seu significado.

Segundo Toledo (1985), é possível distinguir duas concepções para a palavra estatística:

a. 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.
   - as estatísticas econômicas consistem em dados numéricos relacionados com emprego, produção, preços, vendas e com outras atividades ligadas aos vários setores da vida econômica.

b. No singular (ciência), indica a atividade humana especializada ou um corpo de técnicas, ou ainda uma metodologia 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.

Segundo Costa Neto (2002), Estatística é a ciência que se preocupa com a organização, descrição, análise e interpretação dos dados experimentais.

A Estatística é uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumi-los, analisá-los, interpretá-los e deles extrair conclusões.

Para esclarecer e fixar melhor o conceito de Estatística é interessante salientar alguma coisa sobre aquilo que ela não é, pois a Estatística não é, de forma alguma, um método mediante o qual se pode provar tudo aquilo que se deseja. Na realidade, não há nada nos métodos estatísticos capaz de evitar que um indivíduo superficial ou inescrupuloso extraia de um estudo estatístico suas próprias conclusões, apesar da existência dos dados numéricos.

A Estatística também não é simplesmente uma coleção de dados (estatísticos) nem constitui um substituto do pensamento abstrato ou teórico dos casos excepcionais. Dessa forma, os métodos estatísticos não se opõem, de modo algum, à análise qualitativa dos casos particulares.

> O historiador Andrew Lang disse que algumas pessoas usam a Estatística "como um bêbado utiliza um poste de iluminação, para servir de apoio e não para iluminar". (DOWNING, 2002).

Alguns dos que abusam da Estatística o fazem simplesmente por descuido ou ignorância, outros por objetivos pessoais, pretendendo suprimir dados desfavoráveis enquanto dão ênfase aos dados que lhes são favoráveis.

Infelizmente, os maus empregos são tão numerosos quanto os usos válidos da Estatística. Ninguém - administrador, executivo, cientista ou pesquisador social deve deixar-se enganar pelas más Estatísticas, embora os casos de emprego indevido da Estatística sejam tantos que possam gerar a falsa impressão de que a Estatística é, raras vezes ou nunca, digna de confiança.

Para estabelecer o âmbito dos estudos da disciplina e adotando-se um esquema prático de raciocínio, pode-se dizer que a Estatística compreende duas funções (ou campos) bem amplas. A primeira função é descritiva e a segunda é indutiva.

1.2 Natureza dos Dados

Alguns conjuntos de dados consistem em números (como por exemplo: altura, largura, comprimento, peso, idade etc.), chamados de **dados quantitativos**; enquanto outros são não numéricos (como por exemplo: sexo, nome, cor, turma, preferências de uma população etc.), chamados de **dados qualitativos**.

- Os **dados quantitativos** consistem em números que representam contagens ou medidas, pois seus valores são expressos em números.
- Os **dados qualitativos** podem ser separados em diferentes categorias que se distinguem por alguma característica não numérica, já que é resultado de uma classificação por tipos ou atributos.

Os dados estatísticos também podem ser identificados pela sua espécie ou tipo, sendo eles contínuos ou discretos.

- Os **dados contínuos** resultam de um número infinito de valores possíveis que podem ser associados a pontos em uma escala contínua de tal maneira que não haja lacunas ou interrupções. São aqueles onde as variáveis podem assumir qualquer valor dentro de um intervalo de valores, digamos todos os valores x no intervalo $0 \leq x \leq 1$, assim, estas são, então, variáveis contínuas. Como exemplo de variável contínua, temos a altura das pessoas, tomando sua medida real sem arredondamento.
- Os **dados discretos** resultam de um conjunto finito de valores possíveis ou de um conjunto enumerável desses valores, onde as variáveis só podem assumir determinados valores num intervalo de valores, ou seja, a menor diferença não nula entre dois valores dessa variável é finita. Um exemplo de variável discreta, é o número de pontos obtidos em uma partida de voleibol.

Exemplo: uma empresa fabrica peças para automóveis, após a fabricação, uma amostra é retirada para análise de qualidade, ao contar o número de peças inspecionadas, teremos dados discretos, ao analisar as medidas das peças, teremos dados contínuos. Todos os dados deste exemplo são quantitativos.

Exemplo: em um concurso de beleza, ao analisarmos o(a) candidato(a) mais belo(a), estaremos trabalhando com um dado qualitativo.

Podemos ainda, em alguns casos, para simplificar o estudo, transformar dados contínuos em dados discretos, por exemplo, se a altura de uma pessoa é de 174,3784 centímetros, discretizamos a medida, dizendo que esta pessoa possui 174 centímetros ou um metro e setenta e quatro centímetros (1,74 metros).

1.3 Estatística Descritiva e a Estatística Indutiva

- **Estatística Descritiva**: Em um sentido mais amplo, a estatística descritiva pode ser interpretada como uma função, cujo objetivo é a observação de fenômenos de mesma natureza, a coleta de dados numéricos referentes a esses fenômenos, a organização e a classificação desses dados observados e a sua apresentação através de gráficos e tabelas, além do cálculo de coeficientes (estatísticas) que permitem descrever resumidamente os fenômenos.

Dado um conjunto de elementos, podemos, em relação a um certo fenômeno, estudar todos os seus elementos, classificando-os, fornecendo números indicativos que sumariem, certas características dos dados; são números sumariantes, que fornecem descrições de todo o conjunto sem a apresentação total dos elementos, ou mesmo medidas e relações do conjunto, não perceptíveis com a pura apresentação do rol de dados. A este setor da estatística, denomina-se **estatística descritiva**.

- **Estatística Indutiva ou Inferência Estatística**: o processo de generalização, que é característico do método indutivo, está associado a uma margem de incerteza. A existência da incerteza deve-se 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. A medida da incerteza é tratada mediante técnicas e métodos que se fundamentam na Teoria da Probabilidade.

É a parte da estatística que, baseando-se em resultados obtidos da análise de uma amostra da população, procura inferir, induzir ou estimar as leis de comportamento da população da qual a amostra foi retirada.

Portanto, a estatística indutiva refere-se a um processo de generalização, a partir de resultados particulares. Consiste em obter e generalizar conclusões, ou seja, inferir propriedades para o todo com base na parte, no particular. A inferência estatística implica, pois, um raciocínio muito mais complexo do que preside a Estatística Descritiva. Entretanto, bem compreendida e utilizada, pode converter-se em um instrumento muito importante para o desenvolvimento de uma disciplina científica.

Figura 1 – Interação Entre os Ramos da Estatística

1.4 População e Amostra

Ao coletar os dados referentes às características de um grupo de objetos ou indivíduos, tais como as alturas e pesos dos estudantes de uma universidade ou os números de parafusos defeituosos ou não produzidos por uma fábrica em um certo dia, é muitas vezes impossível ou impraticável observar todo o grupo por algum motivo, seja ele custo, tempo, dados inacessíveis, populações infinitas etc.

1.4.1 População ou Universo Estatístico

O conjunto da totalidade dos indivíduos sobre o qual se faz uma análise (ou inferência) recebe o nome de população ou universo. A população congrega todas as observações que sejam relevantes para o estudo de uma ou mais características dos indivíduos, os quais podem ser concebidos tanto como seres animados ou inanimados. Em linguagem mais formal, a população é o conjunto constituído por todos os indivíduos que apresentem pelo menos uma característica comum, cujo comportamento interessa analisar (inferir). ***Censo*** é uma coleção de dados relativos a todos os elementos de uma população.

Pode-se classificar pelo tamanho, sendo finita, quando a população possui um número determinado de elementos e, infinita, quando possui um número infinito de indivíduos, mas tal definição só existe na teoria, pois na prática, nunca encontraremos populações com infinitos elementos, mas sim, populações com um grande número de componentes, por isso, tais populações são tratadas como se fossem infinitas. Quanto maior a população, mais difícil a observação dos aspectos que se deseja estudar, devido ao alto custo, ao intenso trabalho e ao tempo gasto para tal.

1.4.2 Amostra

A amostra pode ser definida como um subconjunto, uma parte selecionada da totalidade de observações abrangidas pela população, através da qual se faz um juízo ou inferência sobre as características da população. As características da amostra são chamadas de estatísticas (descritivas).

A amostragem é um artifício ou uma técnica estatística que possibilita realizar a pesquisa em universos infinitos, quanto aos aspectos de custo e de tempo. Desta forma, a Estatística pode ser estendida ao estudo das populações chamadas “infinitas” nas quais não temos a possibilidade de observar todos os elementos do universo.

Mesmo no caso das populações finitas, passou-se a empregar o estudo por amostragem pela economia e rapidez dos resultados. Assim, o estudo da qualidade dos produtos de um lote de peças produzidas passou a ser feito a partir dos resultados obtidos pela inspeção dos elementos de uma amostra.

A teoria da amostragem é útil para determinar se as diferenças observadas entre duas amostras são realmente devidas a uma variação casual ou se são verdadeiras. No geral, ao estudo da inferência[1] de uma pesquisa a respeito de uma população mediante a utilização de amostras delas extraídas, junto com a precisão das inferências usando a teoria da probabilidade, denominamos inferência estatística.

1.4.3 Tipos de Amostragem

- **Amostra aleatória**: os elementos da população são escolhidos de tal forma que cada um deles tenha igual chance de figurar na amostra;
- **Amostra estratificada**: subdivide-se a população em, no mínimo, duas subpopulações (ou estratos) que compartilham das mesmas característicass (como sexo) e, em seguida, extrai-se uma amostra de cada estrato, proporcionalmente a população. Por exemplo, suponhamos que a população de mulheres em uma determinada escolar é de 54%, neste caso a amostra escolhida para análise terá 54% de mulheres;
- **Amostragem sistemática**: escolhe-se um ponto de partida e selecionamos cada K-ésimo elemento (como por exemplo cada 50º elementos) da população. Por exemplo, em uma pesquisa em uma sala de aula, a professora faz a pergunta a cada 5 alunos da lista da chamada, isto é os números 5, 10, 15, 20 etc.; o mesmo princípio, se utilizado em uma linha de produção, por exemplo, inspecionar uma peça a cada 10 produzidas.

1 Inferência estatística é um ramo da Estatística cujo objetivo é fazer afirmações a partir de um conjunto de valores representativo (amostra) sobre um universo. Tal tipo de afirmação deve sempre vir acompanhada de uma medida de precisão sobre sua veracidade.

- **Amostragem de conveniência**: utiliza-se de resultados de fácil acesso.

Exemplo de Amostra Estratificada: Em uma população de 180 estudantes na qual 108 são mulheres e 72 são homens. A amostra consiste em compor dois estratos (masculino e feminino) nos quais usaremos como amostra 10% da população, como demonstrado na tabela a seguir:

Sexo	População	Estratificação de 10% da População	Amostra
M	108	108 x 10% = 10,8	11
F	72	72 x 10% = 7,2	7
Total	180	180 x 10% = 18	18

Observação: perceba a necessidade de arredondamento da amostra, pois é impossível uma amostra de 10,8 pessoa.

1.5 Fenômenos Estatísticos

O fenômeno em Estatística relaciona-se com qualquer evento que se pretenda analisar, cujo estudo seja passível da aplicação da técnica estatística. A Estatística dedica-se ao estudo dos fenômenos de massa, que são resultantes do concurso de um grande número de causas, total ou parcialmente desconhecidas, que podem ser chamados de "fenômenos estatísticos".

É possível não se conhecerem exatamente as causas subjacentes aos fenômenos, pois pode-se estudá-los através de suas manifestações, descobrindo-se neles alguns aspectos globais, sem remontar a essas causas. O que caracteriza tais fenômenos (sociais, bio1ógicos etc.) é o fato de serem eles provenientes de um concurso de causas nem sempre totalmente conhecidas pelo analista.

Segundo Toledo (1985), os fenômenos estatísticos classificam-se em três tipos:

a. **Fenômenos Coletivos ou Fenômenos de Massa**

   Os fenômenos coletivos são aqueles que não podem ser definidos por uma simples observação. Exemplo: a natalidade; a mortalidade; o preço médio de veículos usados, vendidos diariamente em uma cidade, são fenômenos coletivos.

b. Fenômenos Individuais ou Particulares

Os fenômenos individuais são aqueles que irão compor os fenômenos coletivos. Exemplo: cada nascimento, cada indivíduo que morre, cada casamento que ocorre, cada veículo usado que se vende diariamente, são fenômenos individuais.

c. Fenômenos de Multidão

Os fenômenos de multidão distinguem-se dos fenômenos coletivos pelo fato de as características observadas para a massa não se verificarem para o particular, para o indivíduo isoladamente. Podem ser classificados sob dois aspectos:

- **Fenômenos Típicos**: são aqueles que se manifestam de forma regular, revelando um comportamento definido.
- **Fenômenos Atípicos**: são aqueles cuja manifestação se dá através de um comportamento irregular, não revelando uma tendência definida.

1.6 Fases do Método Estatístico (Estatística Descritiva)

Quando se pretende empreender um estudo estatístico completo, existem diversas fases do trabalho que devem ser desenvolvidas para se chegar aos resultados finais do estudo. Essas etapas ou operações são chamadas fases do trabalho estatístico e são de âmbito da Estatística Descritiva.

As fases principais são as seguintes:

a. Definição do Problema

A primeira fase do trabalho estatístico consiste em uma definição ou formulação correta do problema a ser estudado. Além de considerar detidamente o problema objeto do estudo, o analista deverá examinar outros levantamentos realizados no mesmo campo e análogos, uma vez que parte da informação de que necessita pode, muitas vezes, ser encontrada nesses últimos.

Por exemplo, um fabricante de sabonete, que deseja lançar um produto novo no mercado, poderia estar interessado em um estudo sobre as características dos consumidores atuais. Não havendo estudos semelhantes, ele deverá formular o problema com base em sua própria experiência. Uma lista de fatores relevantes deverá resultar dessa investigação preliminar: número de unidades consumidas por família em cada ano, número médio de pessoas que compõe cada família, número de membros adultos da família, as marcas preferidas e assim por diante. Saber exatamente aquilo que se pretende pesquisar é o mesmo que definir corretamente o problema.

Esta é uma fase muito importante, pois se o problema não for definido corretamente, o planejamento e a execução dos trabalhos jamais atingirão as metas desejadas pelo analista, causando assim, resultados distorcidos que levarão a uma análise incorreta.

b. Planejamento

Consiste em se determinar o procedimento necessário para resolver o problema e, em especial, como levantar informações sobre o assunto objeto do estudo. Abaixo, apresentamos algumas perguntas que devem ser respondidas nesta fase.

- Que dados deverão ser obtidos?
- Como se deve obtê-los?
- O que será pesquisado?
- Quem participará da pesquisa?
- Em que setores geográficos será feita a pesquisa?
- Qual o grau de precisão exigido na pesquisa?
- O estudo vai ser feito na população ou na amostra?
- Qual o tipo de amostragem?
- Qual o tamanho da amostra?
- Como será feita a coleta de dados?
- Quais as informações que serão coletadas?
- Quais são os resultados esperados?
- Quais materiais serão necessários para realizar a pesquisa?

- Qual o tempo disponível para fazer a pesquisa?
- Qual o custo previsto?
- Qual a verba destinada ao projeto? etc.

Mais especificamente, na fase do planejamento, a preocupação maior reside na escolha das perguntas, bem como sua correta formulação, qualquer que seja a modalidade de coleta dos dados.

É nessa fase que será escolhido o tipo de levantamento a ser utilizado, para tanto é conveniente utilizar as ferramentas de qualidade como o 5W1H.

c. Coleta dos Dados

O terceiro passo é essencialmente operacional, compreendendo a coleta das informações propriamente ditas. Formalmente, a coleta de dados se refere à obtenção, reunião e registro sistemático de dados, com um objetivo determinado.

É possível distinguir dois tipos de fontes externas, as quais darão origem a duas espécies de dados: dados primários e dados secundários.

- **Dados Primários**: Os dados são primários quando são coletados, publicados ou comunicados pela própria pessoa ou organização que os tenha recolhido.
- **Dados Secundários**: Os dados são secundários quando são coletados, publicados ou comunicados por outra organização.

A coleta de dados pode ser realizada de duas maneiras, direta ou indiretamente.

1. Coleta Direta

A coleta é direta quando é obtida diretamente da fonte, como no caso da empresa que realiza uma pesquisa para saber a preferência dos consumidores por sua marca. Há três tipos de coleta direta:

a. Coleta Contínua

A coleta de dados é contínua quando estes são obtidos ininterruptamente, automaticamente e na vigência de um determinado período: um ano, por exemplo. É aquela em que é feito o registro tão logo se verifique o fato. Ex.: registros de nascimento, de casamento, de óbito, baixa automática de estoque etc.

b. **Coleta Periódica**

A coleta de dados é periódica quando é realizada em períodos determinados, de tempos em tempos, com repetições cíclicas. Ex.: recenseamento a cada dez anos, o censo industrial, anualmente, balanço etc.

c. **Coleta Ocasional**

A coleta de dados é ocasional quando os dados forem colhidos esporadicamente, ocasionalmente, atendendo a uma conjuntura qualquer ou a uma emergência. Ex.: coleta de casos fatais em um surto epidêmico, registro de pedidos de um determinado artigo que uma grande empresa recebe em um dia de greve.

2. Coleta Indireta

A coleta dos dados é indireta quando é inferida a partir dos elementos conseguidos pela coleta direta, ou através do conhecimento de outros fenômenos que, de algum modo, estejam relacionados com o fenômeno em questão. É feita, portanto, para deduções e conjeturas, podendo ser realizada:

a. **Por analogia**

Quando o conhecimento de um fenômeno é induzido a partir de outro que com ele guarda relações de casualidade. Ex: No carnaval do ano passado 10.000 pessoas visitaram a cidade, espera-se a mesma quantidade este ano.

b. **Por proporcionalização**

Quando o conhecimento de um fato se induz das condições quantitativas de uma parte dele. É feito através de uma regra de três, em que se mede um elemento básico. Nada mais é que uma porcentagem. Ex: uma peça ocupa um espaço de $3m^2$, tenho um espaço de $30m^2$, portanto coloco 10 peças neste espaço.

c. **Por indícios**

A coleta por indícios se dá quando são escolhidos fenômenos sintomáticos para discutir um aspecto geral da vida social. Ex: reunião de elementos como prova de um crime para a descoberta dos culpados.

d. **Por avaliação**

A coleta é feita por avaliação quando, através de informações fidedignas ou estimativas cadastrais, presume-se o estado quantitativo de um fenômeno. Ex.: supor que existam 150 pessoas numa sala.

d. **Apuração dos Dados**

Antes de começar a analisar os dados, é conveniente que lhes seja posto algum tratamento prévio, a fim de torná-los mais expressivos. A quarta etapa do processo é, então, a da apuração ou sumarização, que consiste em resumir os dados, através de sua contagem e agrupamento. É um trabalho de condensação e de tabulação dos dados, que chegam ao analista de forma desorganizada.

Por conseguinte, através da apuração, tem-se 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.

Entretanto, a contrapartida da melhor apreciação dos dados em seu conjunto é a perda correspondente de detalhes, uma vez que se trata de um processo de sintetização.

e. **Apresentação dos Dados**

Há duas formas de apresentação, que não se excluem mutuamente:

1. **Apresentação Tabular (tabelas)**

A apresentação tabular é uma apresentação numérica dos dados. Consiste em dispor os dados em linhas e colunas distribuídas de modo ordenado, segundo algumas regras práticas adotadas pelos diversos sistemas estatísticos.

As tabelas têm a vantagem de conseguir expor, sinteticamente e em um só local, os resultados sobre determinado assunto, de modo a se obter uma visão global mais rápida daquilo que se pretende analisar.

De maneira mais formal, define-se com a tabela a disposição escrita que se obtém, fazendo-se referir uma coleção de dados numéricos a uma determinada ordem de classificação.

2. **Apresentação Gráfica**

A apresentação gráfica dos dados numéricos constitui uma apresentação geométrica. Embora a apresentação tabular seja de extrema importância, no sentido de facilitar a análise numérica dos dados, não permite ao analista obter uma visão tão rápida, fácil e clara do fenômeno e sua variação como a conseguida através de um gráfico.

Os gráficos tornam os dados mais visuais. Com um gráfico você pode transformar os dados de uma tabela para mostrar comparações, padrões e tendências etc.

Este tema será tratado em um capítulo específico.

f. **Análise e Interpretação dos Dados**

Nesta etapa, o interesse maior reside em tirar conclusões que auxiliem o pesquisador a resolver seu problema. A análise dos dados estatísticos está ligada essencialmente ao cálculo de medidas, cuja finalidade principal é descrever o fenômeno. Assim, o conjunto de dados a ser analisado pode ser expresso por números-resumos, as estatísticas, que evidenciam caraterísticas particulares desse conjunto. O significado exato de cada um dos valores obtidos através do cálculo das várias medidas estatísticas disponíveis deve ser bem interpretado.

1.7 Séries Estatísticas

Uma série estatística define-se como toda e qualquer coleção de dados estatísticos referidos a uma mesma ordem de classificação. No sentido mais amplo, série é uma sucessão de números referidos a qualquer variável. Se os números expressarem dados estatísticos, a série será chamada de série estatística.

Em outros termos, a palavra série é usada normalmente para designar um conjunto de dados dispostos de acordo com um caráter variável, residindo a qualidade serial na disposição desses valores, e não em uma disposição temporal ou espacial de indivíduos. Para diferenciar uma série estatística de outra, há que se levar em conta, então, os três caracteres presentes na tabela que as apresenta:

- A época (fator temporal ou cronológico) a que se refere o fenômeno analisado;
- O local (fator espacial ou geográfico) onde o fenômeno acontece;
- O fenômeno (espécie do fato ou fator especificativo) que é descrito.

1.7.1 As Quatro Tipos de Séries Estatísticas

As séries estatísticas, conforme visto, diferenciam-se de acordo com a variação de um desses três elementos: época, local e fato.

1. Série Temporal ou Cronológica

A série temporal, igualmente chamada série cronológica, série histórica, série evolutiva ou marcha, identifica-se pelo caráter variável do fator cronológico. Assim, deve-se ter:

a) Elemento Variável: época (fator cronológico);

b) Elementos Fixos: local (fator geográfico) ou fenômeno (fator especificativo).

Exemplo: Vendas – Mercado Interno – Comercial RF Ltda

Meses	Vendas (produtos)
Janeiro	10.000
Fevereiro	6.000
Março	4.000
Total do Trimestre	20.000

2. Série Geográfica

Também denominada série territorial, série espacial ou série localização, a série geográfica apresenta como elemento ou caráter variável somente o fator geográfico. Assim:

a) Elemento Variável: local (fator geográfico);

b) Elementos Fixos: época (fator cronológico) ou fenômeno (fator especificativo).

Exemplo: Exportação por Região do Brasil

Meses	Exportação (em milhares de Reais)
Norte	5.000
Nordeste	6.000
Sudeste	15.000
Centroeste	5.000
Sul	10.000
Total de Exportação	41.000

3. Série Específica

A série especifica recebe também outras denominações: série categórica ou série por categoria. Agora, o caráter variável é o fenômeno.

a) Elemento Variável: fenômeno (fator especificativo);

b) Elementos Fixos: época (fator cronológico) ou local (fator geográfico).

Exemplo: Vendas por produto – Comercial RF Ltda.

Meses	Vendas (produtos)
Produto A	10.000
Produto B	6.000
Produto C	4.000
Total de Produtos	20.000

4. Distribuição de Frequência

Neste caso, todos os elementos - época, local e fenômeno - são fixos. Embora fixo, o fenômeno ou fator especificativo apresenta-se agora através de gradações, isto é, os dados referentes ao fenômeno que se está representando são reunidos de acordo com sua magnitude. Normalmente, os problemas de tabulação são enquadrados nesse tipo de série.

Nas distribuições de frequências, os dados estatísticos são dispostos ordenadamente em linhas e colunas, de modo a permitir sua leitura nos sentidos horizontal e vertical. Na tabela resultante desse procedimento, são fixos a época, o local e o fenômeno, estando os dados agrupados de acordo com a intensidade ou variação quantitativa do fenômeno.

Exemplo: Número de empregados das várias classes de salários –Comercial RF Ltda.

Classes de Salários (R$)	Número de Empregados
até 500	30
de 500 a 1.000	20
de 1.000 a 1.500	10
de 1.500 a 2.000	5
Total de Empregados	65

Observação: se associarmos o termo *nominal* a "nome somente", o significado é fácil de memorizar. Por exemplo, respostas do tipo "sim" ou "não", sexo etc., como as categorias carecem de qualquer significado numérico, os dados precedentes não podem ser utilizados em cálculos. Assim é que não podemos tirar a média de 20 mulheres e 15 homens. CUIDADO, por vezes atribuem-se números a categorias que não têm qualquer significado para efeito de cálculo, e a média calculada com base neles em geral não tem sentido. Por exemplo: atribuir 0 para sexo masculino e 1 para sexo feminino.

1.8 Planejamento de Experimentos

Os estudos que utilizam métodos estatísticos vão desde os que são bem concebidos e executados, dando resultados confiáveis, aos que são concebidos deficientemente e mal executados, levando a conclusões enganosas e sem qualquer valor real. A seguir, apresentamos alguns pontos importantes para o planejamento de um estudo capaz de produzir resultados válidos:

- Identificar com precisão a questão a ser respondida e definir com clareza a população de interesse.
- Estabelecer um plano para coleta de dados. Esse plano deve descrever detalhadamente a realização de um estudo observacional ou de um experimento, e deve ser elaborado cuidadosamente, de modo que os dados coletados representem efetivamente a população em questão.
- Coletar os dados. Devemos ser extremamente cautelosos, para minimizar os erros que podem resultar de uma coleta tendenciosa.
- Analisar os dados e tirar conclusões. Identificar também possíveis fontes de erros.

Os pontos relacionados acima não invalidam as "Fases do Método Estatístico." Estas foram apresentadas no item 1.6 e reforçam pontos importantes para que o pesquisador reduza erros de planejamento do estudo.

Então, os estudos que requerem métodos estatísticos decorrem tipicamente de duas fontes comuns: **estudos observacionais** e **experimentos.**

Estudos Observacionais: verificamos e medimos características específicas, mas não tentamos manipular ou modificar os elementos a serem estudados.
Por exemplo: fazer uma pesquisa de cidadãos para determinar que porcentagem da população utiliza transporte público.

Experimento: aplicamos determinado tratamento e passamos então a observar seus efeitos sobre os elementos a serem pesquisados.
Por exemplo, realizar um tratamento com determinado remédio com um grupo de pacientes a fim de determinar sua eficiência na cura.

1.9 Arredondamento de Números

O arredondamento de um dado estatístico deve obedecer a algumas regras:

a. **Arredondamento por Falta**

Quando o primeiro dígito após aquele que será arredondado (identificados no quadro a seguir em negrito) for **menor ou igual a quatro**, não deverá ser alterado o dígito remanescente.

Número a arredondar	Arredondamento para	Números que serão eliminados	Número arredondado
12,**4**89	Inteiros	**4**89	12
20,7**3**3	Décimos	**3**3	20,7
35,99**2**	Centésimos	**2**	35,99

b. **Arredondamento por Excesso**

Quando o primeiro dígito após aquele que será arredondado (identificados no quadro a seguir em negrito) for **maior ou igual a cinco** seguido por dígitos maiores que zero, o dígito remanescente será acrescido de uma unidade.

Número a arredondar	Arredondamento para	Números que serão eliminados	Número arredondado
15,**5**04	Inteiros	**5**04	16
16,5**6**1	Décimos	**6**1	16,6
17,57**8**	Centésimos	**8**	17,58

c. **Arredondamento de Dígitos Seguidos do Cinco**

Quando o dígito situado mais à esquerda dos que serão eliminados for um cinco, o último dígito remanescente (identificados no quadro a seguir em negrito), se for par, não se altera, e se for ímpar, será aumentado de uma unidade.

Número a arredondar	Arredondamento para	Dígito remanescente	Número arredondado
1**5**,5	Inteiros	**5** - ímpar	16
1**6**,5	Inteiros	**6** - par	16
16,**5**5	Decimal	**5** - ímpar	16,6
16,**6**5	Decimal	**6** - par	16,6

d. Arredondamento de Soma

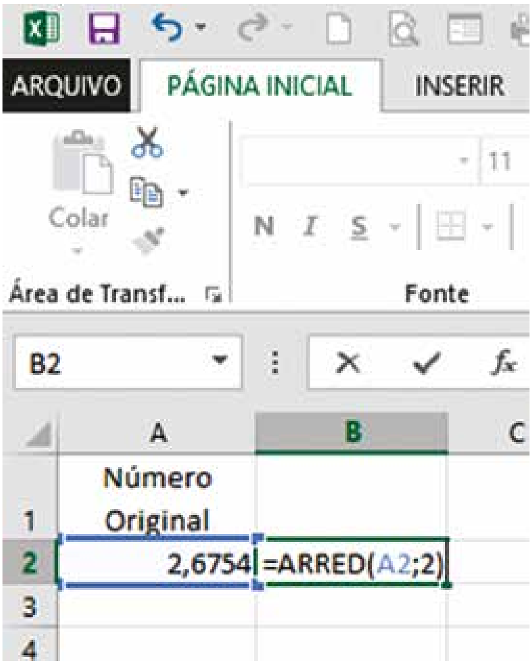
Quando se trata de soma, deve-se arredondar primeiro o total e, posteriormente, as parcelas. Por exemplo:

Nº	Série Original	Série Arredondada	Erro de Arredondamento	Série Corrigida
1	6,62 %	7	7 – 6,63 = 0,38	7
2	7,50 %	8	8 – 7,50 = 0,50	7
3	14,62 %	15	15 – 14,62 = 0,38	15
4	20,10 %	20	20 – 20,10 = – 0,10	20
5	24,64 %	25	25 – 24,73 = 0,36	25
6	26,52 %	27	27 – 26,52 = 0,48	26
	100,00 %	102 > 100		100

Podemos observar no exemplo acima, que na coluna "Série Arredondada" ocorreram cinco arredondamentos para cima e apenas um para baixo, acarretando um erro de soma de dois pontos, pois ao arredondar a soma da "Série Original" (100,00 %) deveríamos obter um valor arredondado de 100 % e não de 102 %.

Nesses casos devemos rever os arredondamentos para corrigir a soma final, calcular o erro causado pelo arredondamento (coluna de "Erro de Arredondamento"). Após calcular os erros causados pelo arredondamento, deve-se identificar as linhas com maior erro nominal, ajustando o arredondamento original de forma a atingir a soma final original. No caso apresentado acima, foram realizados ajustes para baixo nas linhas 2 e 6, com erros respectivos de 0,50 e 0,48 e os números arredondados de 8 para 7 e 27 para 26 respectivamente.

Exemplo no Excel: No Excel você pode utilizar a função "ARRED". Ao utilizá-la você irá arredondar para o número mais próximo especificado. A função ARRED requer dois argumentos, conforme apresentado na fórmula a seguir =**ARRED(a*rgumento1, argumento2*)**



- × O primeiro argumento é o número a ser arredondado. Pode ser um número que você especifique diretamente na fórmula ou uma referência de célula.
- × O segundo argumento é o número de casas decimais para as quais você deseja arredondar o resultado.

No exemplo ao lado, podemos perceber que para fazer o arredondamento do número original 2,6574, utilizamos a fórmula "=ARRED(A2;2)", onde "A2" refere-se a célula que contém o número a ser arredondado e, o número "2", representa o número de casas decimais que você deseja arredondar. O resultado desta fórmula será 2,68.

Se a fórmula fosse "**=ARRED(A2;1)**" o resultado seria **2,7**.

1.10 Algarismos Significativos

Se uma altura foi determinada com precisão como 1,66 metros, isto significa que seu valor verdadeiro está compreendido entre 1,655 e 1,665 metros. Os algarismos corretos, separados dos zeros necessários para a localização da vírgula, chamam-se **algarismos significativos** do número.

Exemplo:

a. 1,66 têm 3 algarismos significativos.

b. 4.5300 têm 5 algarismos significativos.

c. 0,0018 = 1,8 x 10-3 têm 2 algarismos significativos.

d. 0,001800 = 1,800 x 10-3 têm 4 algarismos significativos.

Os números resultantes de enumerações ou de contagens, ao contrário dos das medições, são naturalmente exatos e, assim, tem uma quantidade ilimitada de algarismos significativos. Em alguns destes casos, contudo, pode ser difícil decidir quais são os algarismos significativos sem informações adicionais. Por exemplo, o número 186.000.000 pode ter 3, 4,..., 9 algarismos significativos. Se souber que ele possui 5 algarismos significativos, será melhor escrever o número como 186,00 milhões ou $1{,}8600 \times 10^8$.

1.11 Notação Sigma (Σ) - Somatório

Muitos procedimentos estatísticos necessitam do cálculo da soma de um conjunto de números. O operador somatório facilita sobremaneira a indicação e a formulação de medidas, bem como algumas operações algébricas desenvolvidas pela estatística. Então, vamos definir matematicamente como isso é feito:

Definição: $$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + x_4 \ldots\ldots x_i$$

A equação acima, pode ser explicada da seguinte forma: somatório de X_i, com *i* variando de 1 até *n*. O 1 e o *n* indicam, respectivamente, o limite inferior e o superior do somatório, representando o número de ordem da primeira e da última parcela a serem somadas.

Exemplo: um professor deseja somar as idades de seus alunos. Sua turma possui exatamente 35 alunos. Vamos especificar as variáveis *x*, *i* e *n* e montar a equação acima representando a soma das idades. Então, *x* representa a idade de cada aluno, i são os alunos da turma e n é o número total de alunos desta turma, que neste caso n é igual a 35 alunos. A equação para este caso pode ser representada da seguinte forma:

$$\sum_{i=1}^{35} x_i = x_1 + x_2 + x_3 + x_4 \ldots\ldots x_{35} = 21 + 22 + 20 + 18 + \ldots + 24$$

Exemplo: se uma variável qualquer x tiver a sequência de valores (2, 5, 7 e 8). Então, podemos dizer que $\sum x = 22$.

Nesse exemplo, temos que $x_1 = 2$, $x_2 = 5$, $x_3 = 7$ e $x_4 = 8$. A leitura que devemos fazer de $\sum_{i=1}^{n} x_i$ é a soma dos valores da variável x começando de ($i = 1$) e terminando com ($i = 4$).

Então:

$$\sum_{i=1}^{4} x_i = \sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4 = 2 + 5 + 7 + 8 = 22$$

Se quisermos calcular: teremos o seguinte cálculo:

$$\sum_{i=1}^{4} \left(x_i * 2\right) = (x_1 * 2) + (x_2 * 2) + (x_3 * 2) + (x_4 * 2) =$$

$$\sum_{i=1}^{4} \left(x_i * 2\right) = \left(2 * 2\right) + \left(5 * 2\right) + \left(7 * 2\right) + \left(8 * 2\right) = 44$$

Atividades

1. Faça a correlação correta:

A) Estatística Descritiva;	N) Variável quantitativa;
B) Estatística Indutiva;	O) Variável quantitativa contínua;
C) População;	P) Variável quantitativa discreta;

D) Amostra;	Q) Amostragem Aleatória;
E) Fonte primária;	R) Amostragem sistemática;
G) Fonte secundária;	S) Amostragem estratificada proporcional;
H) Séries estatísticas;	T) Dados brutos;
I) Séries regionais;	U) Variável;
J) Séries Categóricas;	V) Frequência;
K) Variável qualitativa;	W) Frequência Relativa;
L) Rol;	X) Limite Superior;
M) Arredondamento;	Y) Limite Inferior.

( ) É o conjunto de todos os elementos envolvidos no fenômeno a ser estudado;

( ) Toda e qualquer coleção de dados estatísticos referidos a uma mesma ordem de classificação;

( ) Cuida da análise e interpretação dos dados;

( ) É o número de vezes que cada observação ocorre no rol;

( ) O pesquisador utiliza relatórios, revistas, arquivos, livros ou coleta de dados realizados por instituições especializadas.

2. Estabelecer quais dos dados seguintes são discretos e quais são contínuos.

   a. Número de ações vendidas diariamente na Bolsa de Valores.
   b. Temperaturas registradas a cada meia hora em um posto de meteorologia.
   c. Vida média de uma televisão.
   d. Salários anuais de trabalhadores da construção civil.
   e. Comprimentos de 1.000 parafusos produzidos numa fábrica.
   f. Número de litros de água numa máquina de lavar roupas.
   g. Número de livros em uma estante de biblioteca.

h. Diâmetro de uma esfera.

3. A população envolvida em uma pesquisa sobre a incidência de cárie dentária em escolas da cidade de Morro Grande é apresentada no quadro a seguir:

Escola	População
A	500
B	250
C	440
D	360
Total	1550

Baseado nesses dados, estratifique uma amostra com 200 elementos.

4. Arredonde cada um dos números seguintes conforme a precisão indicada:

a. 48,6 para a unidade mais próxima.

b. 136,5 para a unidade mais próxima.

c. 2,484 para o centésimo mais próximo.

d. 0,0435 para o milésimo mais próximo.

e. 4,50001 para a unidade mais próxima.

f. 143,95 para o décimo mais próximo.

g. 368 para a centena mais próxima.

h. 24.448 pra o milhar mais próximo.

i. 5,56500 para o centésimo mais próximo.

j. 5,56501 para o centésimo mais próximo.

5. Some os números: 4,35; 8,65; 2,95; 12,45; 6,65; 7,55 e 9,75; nas seguintes condições:

a. Diretamente.

b. Arredondando para décimos de acordo com a convenção de número par.

c. Arredondando para décimos de maneira que o algarismo anterior a 5 cresça de uma unidade.

d. Arredondando para a unidade mais próxima.

6. Quantos algarismos significativos há em cada um dos seguintes números, supondo-se que eles foram registrados com precisão?

a. 1,498 centímetros.

b. 1,4980 centímetros.

c. 0,0028 metros.

d. 0,00280 metros.

e. 1,00280 metros.

f. 9 gramas.

7. Observando a tabela de distribuição por frequência abaixo, que representa a distribuição de renda de certo país, responda:

Faixa de Renda	**Habitações**
até 1 salário mínimo	224.740
de 1 a 3 salários mínimos	363.860
de 4 a 8 salários mínimos	155.700
mais de 8 salários mínimos	47.500
Total	791.800

a. Qual é a porcentagem de domicílios onde a renda é superior a 8 salários mínimos?

b. Quantos são os domiciliados onde a renda está entre 1 e 3 salários?

c. Quantos são os domicílios onde a renda está abaixo de 3 salários?

d. Faça uma amostra estratificada de 15% da população.

Respostas

1. C; L; A; V; G
2. a) Discreto; b) Contínuas; c) Contínua; d) Discretos; e) Contínuos; f) contínua; g) Discreta; h) Contínua.
3. A: 64; B: 32; C:57; D: 47
4. a) 49; b) 136; c) 2,48; d) 0,044; e) 5; f) 144,0; g) 400; h) 24.000; i) 5,56; j) 5,57
5. a) 52,35; b) 52,4; c) 52,7; d) 53
6. a) 4; b) 5; c) 2; d)3; e) 6; f) 1
7. a) 6,0%; b) 363.860; c) 558.600; d) 33.711; 54.579; 23.355; 7.125; 118.770

2

Distribuição de Frequências

Uma das vantagens das tabelas estatísticas é a de condensar, de forma consistente, as informações necessárias ao estudo desejado. Isto porque, frequentemente, o estudo de um determinado fenômeno requer a coleta de uma grande massa de dados numéricos, difícil de ser tratada se esses dados não forem organizados e condensados em uma tabela.

No caso específico das seriações, acontece normalmente que, ao coletar os dados referentes ao fenômeno objeto de estudo, o analista se defronta com valores que se repetem algumas vezes, sugerindo sua apresentação através de tabelas onde somente apareçam valores distintos uns dos outros.

Essa providência favorece evidentemente uma análise e interpretação mais rápida da natureza e comportamento do fenômeno observado. Neste capítulo, será desenvolvido um tipo de tabela que condensa uma coleção de dados conforme as frequências ou repetições de seus valores.

2.1 Dados Brutos

Dados brutos são aqueles que ainda não foram numericamente organizados. Feita a coleta, os dados originais ainda não se encontram prontos para a análise, por não estarem numericamente organizados.

Por essa razão, costuma-se chamá-los de dados brutos. Tomando-se, por exemplo, as alturas dos alunos em uma sala de aula e anotando-se os resultados em uma lista da qual constem os nomes dos alunos em ordem alfabética, ninguém garantirá que os valores correspondentes às alturas observarão uma determinada ordem numérica, crescente ou decrescente. Mais provável é que estejam desorganizados, uma vez que a ordem das alturas não corresponde necessariamente à ordem alfabética.

Na tabela abaixo, estão relacionados os valores correspondentes às notas individuais de 20 alunos de Estatística.

Tabela 2.1 - Notas de Estatística – turma A

9,8	7,6	8,2	8,1	3,4
6,8	4,5	2,3	5,5	7,6
9,6	4,7	3,2	1,7	5,4
5,2	6,7	7,8	8,2	8,5

Como pode ser observado, as notas estão dispostas de forma desordenada. Em razão disso, pouca informação se consegue obter inspecionando os dados anotados. Mesmo uma informação tão simples como a de saber as notas máxima e mínima requer um certo exame dos dados da tabela.

2.2 Rol (Dados Ordenados)

Um **rol** é um arranjo de dados numericamente brutos em ordem crescente ou decrescente de grandezas, isto é, é uma lista em que os valores estão dispostos em uma determinada ordem.

Dispondo os dados da Tabela 2.1 em ordem crescente de acordo com as notas, obtém-se uma ordenação conforme apresentado a seguir.

Tabela 2.2 - Notas de Estatística – turma A

1,7	2,3	3,2	3,4	4,5
4,7	5,2	5,4	5,5	6,7
6,8	7,6	7,6	7,8	8,1
8,2	8,2	8,5	9,6	9,8

2.3 Tabela de Frequências

As **tabelas de frequências** são representações nas quais os valores se apresentam em correspondência com suas repetições, evitando-se assim que eles apareçam mais de uma vez na tabela, como ocorre com o rol.

Exemplo: Uma empresa fabricante de instrumentos de precisão está interessada em saber o número de aparelhos defeituosos rejeitados pela seção encarregada do controle de qualidade. As estatísticas, fornecidas para essa seção, referem-se ao período de 1991 a 1994.

Tabela 2.3 - Empresa 2M – Número mensal de aparelhos defeituosos

Mês / Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2010	6	2	5	6	0	8	7	6	3	4	5	8
2011	10	9	7	6	3	4	6	4	5	4	0	1
2012	3	6	7	9	3	1	4	6	5	3	5	4
2013	7	2	5	8	6	4	2	5	1	6	5	2

Os dados brutos, apresentados na tabela acima, não informam muita coisa sobre o fenômeno "número de aparelhos defeituosos", sendo difícil extrair deles muitas conclusões, sem esforço de concentração. Observa-se, entretanto, que os valores que constam da tabela aparecem repetidos, como o zero. Esse fato irá sugerir, naturalmente, que se condensem todos as resultados em uma tabela, estabelecendo a correspondência entre o valor individual e o respectivo número de vezes que ele foi observado.

As tabelas de frequências podem representar tanto valores individuais como valores agrupados em classes, conforme apresentado a seguir.

a. Distribuição de Frequências de Dados Tabulados - Não-Agrupados em Classes

Utilizando os dados da tabela anterior, é possível construir uma tabela de frequência de valores não-agrupados em classe, ou seja, uma tabela onde os valores da variável aparecem individualmente. Este tipo de apresentação é utilizado para representar uma variável discreta ou descontínua.

Tabela 2.4 - Empresa 2M – Número mensal de aparelhos defeituosos

j	Nº de Aparelhos com Defeito (x_j)	Número de Meses (f_j)	Observações para construção da tabela
1	0	2	No rol (Tabela 3), zero defeitos aparecem duas vezes, nos meses de maio e novembro. Um defeito, três vezes, nos meses de junho, setembro e dezembro. Dois defeitos, quatro vezes, nos meses de fevereiro (dois), julho e dezembro, e assim por diante. Deve-se contar o número de defeitos apresentados no Rol e inserir na tabela de frequências.
2	1	3	
3	2	4	
4	3	5	
5	4	7	
6	5	8	
7	6	9	
8	7	4	
9	8	3	
10	9	2	
11	10	1	
		$\sum_{j=1}^{11} f_j = 48$	Na Tabela-3 aparecem 48 “números de aparelhos defeituosos”.

Na primeira coluna, encabeçada pelo índice *j*, aparecem os números correspondentes à ordem dos valores da variável. O índice *j* será utilizado sempre que se estiver trabalhando com tabelas de frequências (dados tabulados).

Na segunda coluna, encabeçada por x_i, são anotados os valores da variável, neste caso estamos estudando a variável "quantidade de defeitos". A terceira coluna, encabeçada por f_j, apresenta as frequências (número de vezes que os valores da segunda coluna aparecem no rol apresentado na Tabela 3), que são os resultados numéricos provenientes da contagem. A soma das frequências é sempre igual ao número total de valores observados, conforme equação a seguir:

$$\sum_{j=1}^{k} f_j = f_1 + f_2 + \ldots + f_j = n$$

onde:

k – é o extremo superior do intervalo de valores do índice *j;*

f_j – é o número de observações de um valor (frequência absoluta);

n – é o número total de valores observados.

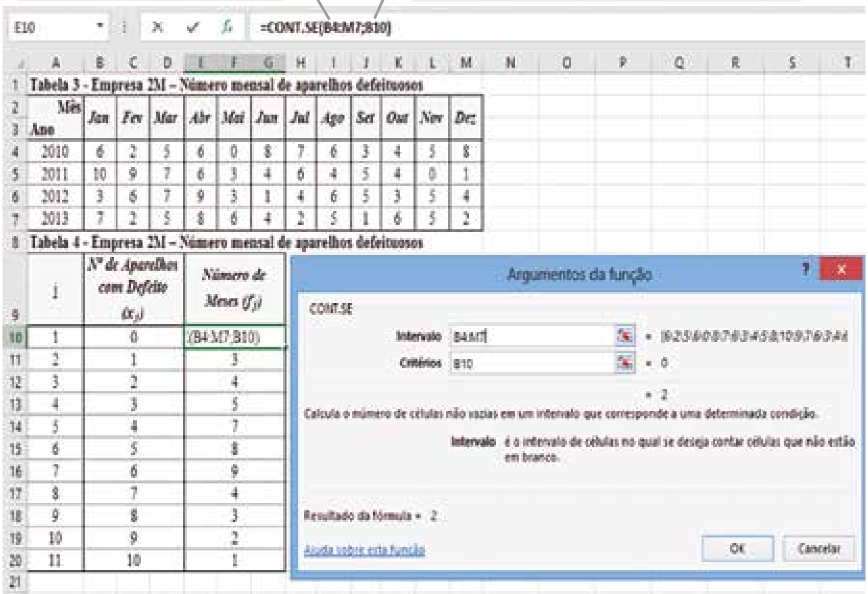
Exemplo: na Tabela 2.4 temos as seguintes notações:

- *j* varia de 1 a 11; então *k* = 11
- $f_1 = 2; f_2 = 3; f_3 = 4 \ldots . f_{11} = 1$
- $n = f_1 + f_2 + f_3 + \ldots f_{11} = 2 + 3 + 4 + 5 + 7 + 8 + 9 + 4 + 3 + 2 + 1 = 48$

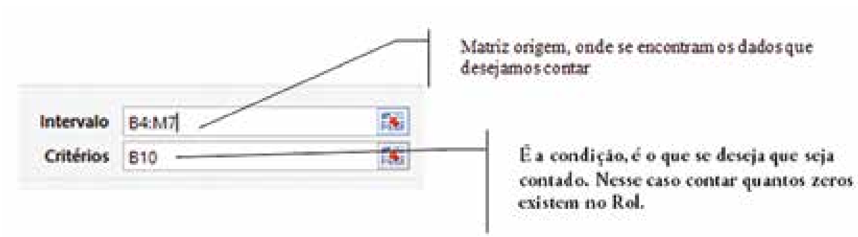
Exemplo no Excel: No Excel, podemos contar o número de meses utilizando a função "CONT.SE". A função CONT.SE conta o número de células dentro de um intervalo que atendem a um único critério que você especifica. Por exemplo, é possível contar todas as células que começam com uma certa letra ou todas as células que contêm um número maior do que ou menor do que um número que você especificar. No exemplo apresentado na Tabela 2.4, podemos contar quantos meses possuem zero erros no Rol da Tabela 2.3.

Matriz origem, onde se encontram os dados que desejamos contar

É a condição, é o que se deseja que seja contado. Nesse caso, contar quantos zeros existem no Rol.



Ampliando o quadro de configuração do Excel, temos:



b. Distribuição de frequências de Dados Agrupados em Classes

Muitas vezes, mesmo com o risco de se sacrificar algum detalhe manifestado na ordenação de valores individuais, há vantagem em resumir os dados em uma distribuição de frequências, onde os valores observados não aparecerão individualmente, mas agrupados em classes.

Quando a variável objeto do estudo for contínua, será sempre conveniente agrupar os valores observados em classes. Se, por outro lado, a variável for discreta e o número de valores representativos dessa variável for muito grande, recomenda-se o agrupamento dos dados em classes. Neste último caso, o procedimento visa a evitar certos inconvenientes, como:

- Grande extensão da tabela, dificultando, tanto quanto os dados brutos, a leitura e a interpretação dos resultados apurados.
- Aparecimento de diversos valores da variável com frequência nula.
- Impossibilidade ou dificuldade de visualização do comportamento do fenômeno como um todo, bem como de sua variação.

Exemplo: Um teste de estatística, contendo 100 perguntas do tipo certo/errado, foi aplicado em uma turma de 500 estudantes. A tabela abaixo apresenta os resultados do teste.

Tabela 2.5 – Exemplo de Tabela de Frequência por Classes

Classes Notas	Frequência f_j
0 |– 10	5
10 |–20	15
20 |– 30	20
30 |– 40	45
40 |– 50	100
50 |– 60	130
60 |– 70	100
70 |– 80	60
80 |– 90	15
90 |– 100	10
	$\sum_{j=1}^{10} f_j = 500$

Classes de Frequências (CLASSES)

Frequência das Classes (FREQUÊNCIAS) número de ocorrências do evento dentro do limite da classe.

Observação: Simbologias usadas nas distribuições por frequências, para determinação da amplitude de classes:

- 1 |— 5: o limite inferior (1) pertence à classe e o superior (5) pertence à classe seguinte.
- 1 —| 5: o limite inferior (1) não pertence à classe e o superior (5) pertence.
- 1 |—| 5: os dois limites (1 e 5) pertencem à classe.
- 1 — 5: os dois limites (1 e 5) não pertencem à classe.

2.4 Elementos de uma Distribuição de Frequências

Para construir uma tabela de frequências, é necessário conhecer alguns termos próprios e adequados.

a) Frequência Simples Absoluta - f_j

A frequência simples absoluta de uma classe ou de um valor individual é o número de observações correspondentes a essa classe ou a esse valor.

Exemplo: ao analisarmos a tabela anterior, podemos verificar que:

$f_1 = 5; f_2 = 15; \ldots\ldots; f_9 = 15; f_{10} = 10.$

b) Amplitude Total - *At*

A amplitude total ou intervalo total é a diferença entre o maior e o menor valor observado da variável em estudo.

Exemplo: na Tabela-2.5, o maior valor é 100 e o menor valor é zero, então, a amplitude total é igual a 100.

c) Classe

Classe de frequência, ou, simplesmente, classe, é cada um dos grupos valores em que se subdivide a amplitude total do conjunto de valores observados da variável.

Uma determinada classe pode ser identificada por seus extremos ou pela ordem em que ela se encontra na tabela.

Classe 0 |– 10 ou primeira classe (j = 1)

Classe 80 |– 90 ou nona classe (j = 9)

O número de classes, em uma distribuição de frequências, é representado por k. É importante que a distribuição conte com um número adequado de classes. Se esse número for escasso, os dados originais ficarão tão comprimidos que pouca informação se poderá extrair da tabela. Se, por outro lado, forem utilizadas muitas classes, haverá algumas com frequência nula ou muito pequena, e o resultado será um distribuído irregular e prejudicial à interpretação do fenômeno como um todo.

A regra de Sturges é um dos métodos que estabelece o número de classes, sendo calculada pela seguinte expressão:

$$k = 1 + 3{,}3.\log_{10} n$$

onde:

k = número de classes

n = número total de observações

Observação: k é sempre um **número inteiro**, pois se estivermos medindo número de pessoas, ou peças, não conseguimos medir uma pessoa e meia ou uma peça e meia por exemplo.

Exemplo: Se o número de observações for de 400:

n = 400

k = 1 + 3,3 x log (400)

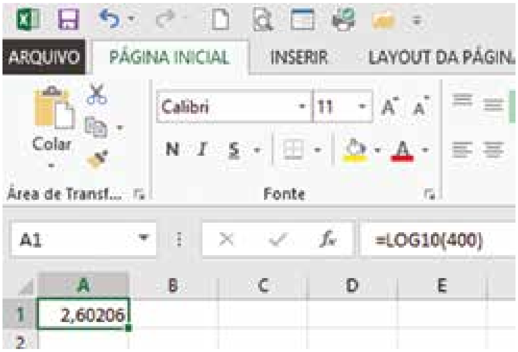
como log (400) = 2,60206, então:

k = 1 + (3,3 x 2,60206) = 1 + 8,586798

k = 9,586798, como k é sempre um número inteiro, temos que:

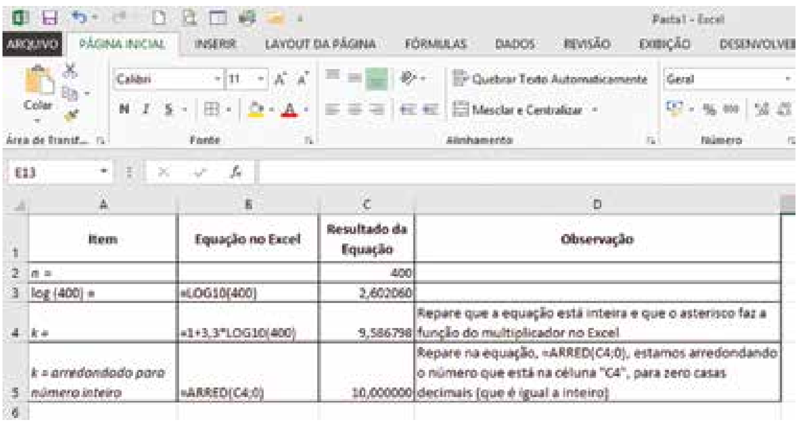
k = 10

Exemplo: como calcular logaritmo na base 10 (log) no Excel?



Para calcular o logaritmo na base 10 no Excel, você deverá utilizar a seguinte fórmula: = LOG10 (número desejado), conforme mostrado no quadro ao lado, na célula A1.

Exemplo: como calcular o número de classes no Excel, se o número de observações for de 400.



Item	Equação no Excel	Resultado da Equação	Observação
n =		400	
log (400) =	=LOG10(400)	2,602060	
k =	=1+3,3*LOG10(400)	9,586798	Repare que a equação está inteira e que o asterisco faz a função do multiplicador no Excel
k = arredondado para número inteiro	=ARRED(C4;0)	10,000000	Repare na equação, =ARRED(C4;0), estamos arredondando o número que está na céluna "C4", para zero casas decimais (que é igual a inteiro)

d) **Limites de Classes**

Os limites de classe são seus valores extremos. Por exemplo, na tabela anterior (Tabela 5), a segunda classe tem como limites os valores 10 e 20. O valor 10 é denominado limite inferior ou limite mínimo de classe **(*LI*),** enquanto o valor 20 é denominado limite superior da classe **(*LS*)**.

Os valores 0 e 100, por representarem, respectivamente, o limite inferior da primeira classe e o superior da última, são também denominados limite inferior e limite superior da distribuição ou do fenômeno estudado.

1. Limites Reais de Classe

Considere a tabela abaixo, na qual se exemplifica uma distribuição de frequências.

Classes	Frequência (f_j)
2,50 a 2,59	1
2,60 a 2,69	2
2,70 a 2,79	7
2,80 a 2,89	4
2,90 a 2,99	2

A primeira classe, cujos limites são 2,50 e 2,59, congregaria, na realidade, valores compreendidos no intervalo de 2,495 a 2,595, caso haja necessidade de arredondamento. Esses limites são denominados limites reais de classe.

2. Limites não Definidos

Uma classe com limite indefinido ou aberto é aquela que inclui todos os valores da variável menores que um certo limite superior especificado, ou maiores que um limite inferior especificado.

Classes	Frequência (f_j)
1 a 99	60
100 a 499	80
500 a 999	100
1.000 a 9.999	40
10.000 ou mais	20
Total	300

A utilização desse expediente prejudica a representação e análise dos dados.

e) Amplitude do Intervalo de Classe - *Ac*

A amplitude do intervalo de classe ou simplesmente intervalo de classe é o comprimento da classe, sendo geralmente definida como a diferença entre seus limites superior e inferior.

A amplitude de classe também pode ser calculada através da seguinte equação:

$$Ac = \frac{At}{k}$$

f) Ponto Médio de Classe - PM_j

O ponto médio ou valor médio de classe é o valor que a representa, para efeito de cálculo de certas medidas. Na distribuição de frequências com valores agrupados em classes, considera-se que os resultados incluídos em cada classe distribuem-se uniformemente por seu intervalo. Por essa razão, a escolha do ponto médio para representar todos os valores de uma classe é o procedimento mais coerente, uma vez que esse ponto, por suas caraterísticas, deve ser equidistante dos limites de classe.

Para obter o ponto médio de uma classe, basta acrescentar ao seu limite inferior a metade da amplitude do intervalo de classe. Esse procedimento pode ser adotado, qualquer que seja a representação tabular escolhida.

2.5 Tipos de Frequências

a) Frequência Simples Absoluta – f_j

A frequência simples absoluta é o número de repetições de um valor individual ou de uma classe de valores da variável. Trata-se do caso visto até o presente. A soma das frequências simples absolutas em uma tabela é chamada frequência total e corresponde ao número total de observações.

$$\sum_{j=1}^{k} f_j = n$$

b) Frequência Simples Relativa - fr_j

A frequência simples relativa representa a proporção de observações de um valor individual ou de uma classe, em relação ao número total de observações. Trata-se, portanto, de um número relativo. Para calcular a frequência relativa, basta dividir a frequência absoluta da classe ou do valor individual pelo número total de observações. Simbolicamente,

$$fr_j = \frac{f_j}{\sum f_j} = \frac{f_j}{n}$$

Desejando expressar o resultado em termos percentuais, multiplica-se o quociente obtido por 100.

$$fr_j = \frac{f_j}{\sum f_j}.100 = \frac{f_j}{n}.100$$

c) Frequência Absoluta Acumulada - F_j

A frequência absoluta acumulada de uma classe ou de um valor individual é a soma da frequência simples absoluta dessa classe ou desse valor com as frequências simples absolutas das classes ou dos valores anteriores.

d) Frequência Relativa Acumulada - Fr_j

A frequência relativa acumulada da classe ou do valor individual *j* é igual à soma da frequência simples relativa dessa classe ou desse valor com as frequências simples relativas das classes ou dos valores anteriores.

Exemplo: Um teste de estatística, contendo 100 perguntas do tipo certo/errado, foi aplicado em uma turma de 500 estudantes. A tabela abaixo apresenta os resultados do teste.

Classes Notas	Frequência f_j	F. Relativa fr_j (%)	F. Acumul. F_j	F. R. Acumul. Fr_j (%)	Ponto Médio – x_j
0 – 10	5	5÷500 =**1%**	**5**	5÷500 =**1%**	(0+10) ÷2=**5**
10 –20	15	15÷500 =**3%**	5+ 15 = **20**	20÷500 =**4%**	(10+20) ÷2=**15**
20 – 30	20	20÷500 =**4%**	20 + 20 = **40**	40÷500 =**8%**	(20+30) ÷2=**25**
30 – 40	45	45÷500 =**9%**	40 + 45 = **85**	85÷500 =**17%**	(30+40) ÷2=**35**
40 – 50	100	100÷500 =**20%**	85+ 100 = **185**	185÷500 =**37%**	(40+50) ÷2=**45**
50 – 60	130	130÷500 =**26%**	185+ 130 = **315**	315÷500 =**63%**	(50+60) ÷2=**55**
60 – 70	100	100÷500 =**20%**	315 + 100 = **415**	415÷500 =**83%**	(60+70) ÷2=**65**
70 – 80	60	60÷500 =**12%**	415 + 60 = **475**	475÷500 =**95%**	(70+80) ÷2=**75**
80 – 90	15	15÷500 =**3%**	475+ 15 = **490**	490÷500 =**98%**	(80+90) ÷2=**85**
90 – 100	10	10÷500 =**2%**	490 + 10 = **500**	500÷500 =**100%**	(90+100) ÷2=**95**
	$\sum_{j=1}^{10} f_j = 500$	$\sum_{j=1}^{10} fr_j = 100\%$			

2.6 Roteiro para a elaboração de uma Tabela de Frequências com Dados Agrupados em Classes

- Lista de dados brutos que pode ou não ser transformada em rol;
- Encontrar a Amplitude Total do conjunto de valores observados;

  At = maior valor do conjunto – menor valor do conjunto
- Calcular o número de classes através da regra de Sturges;

$$k = 1 + 3{,}3.\log_{10} n$$

- Determinar a amplitude do intervalo de classe. A amplitude do intervalo de classe será igual ao quociente entre a amplitude total da série e o número de classes escolhidas;

  Amplitude do Intervalo de Classe = $Ac = \dfrac{At}{k}$

Muitas vezes, ao efetuar a divisão acima, pode-se chegar a um resultado não muito conveniente, sob o aspecto de montagem das classes. Neste caso, convém arredondar o número correspondente à amplitude do intervalo de classe para um número mais adequado, que facilite os cálculos – arredondamento arbitrário.

- Determinar os limites das classes;
- Construir a tabela de frequências, conforme sugerido anteriormente.

Exemplo: A tabela abaixo representa a altura de 100 alunos (em metros) de uma determinada sala de aula, construa uma tabela com a distribuição de frequências, calculando o número de classes e sua amplitude, utilizando a regra de Sturges.

1,48	1,48	1,52	1,52	1,53	1,53	1,55	1,55	1,57	1,57
1,57	1,57	1,58	1,58	1,58	1,58	1,59	1,59	1,60	1,60
1,61	1,61	1,61	1,61	1,62	1,62	1,63	1,63	1,64	1,64
1,64	1,64	1,65	1,65	1,65	1,65	1,65	1,65	1,67	1,67
1,68	1,68	1,68	1,68	1,69	1,69	1,69	1,69	1,70	1,70
1,71	1,71	1,72	1,72	1,72	1,72	1,72	1,72	1,73	1,73
1,73	1,73	1,73	1,73	1,73	1,76	1,76	1,76	1,76	1,77
1,77	1,78	1,78	1,78	1,78	1,81	1,81	1,82	1,82	1,83
1,83	1,83	1,84	1,84	1,84	1,84	1,85	1,85	1,87	1,87
1,88	1,88	1,89	1,89	1,92	1,92	1,93	1,93	1,98	1,98

- Amplitude Total.

  At = 1,98 – 1,48 = 0,50

- Número de Classes.

  k = 1 + 3,3 x log (100) = 1 + 3,3 x 2,000 = 7,6

  Como k é sempre um número inteiro, temos que usar k = 8

- Amplitude do Intervalo de Classe.

$$Ac = \frac{At}{k} = \frac{0,50}{8} = 0,0625$$

Como a Amplitude do Intervalo de Classe deve ter o mesmo número de casas decimais que as medidas do Rol (alturas em metros) o número acima deve ser arredondado, então Ac = 0,06.

- Agora podemos determinar os limites das classes, conforme apresentado abaixo:

Limite Inferior de Cada Classe	Limite Superior de Cada Classe	Observação
1,48	1,48 + 0,06 = **1,54**	Devemos iniciar esta coluna com o menor valor encontrado no Rol, que é 1,48
1,54	1,54 + 0,06 = **1,60**	
1,60	1,60 + 0,06 = **1,66**	
1,66	1,66 + 0,06 = **1,72**	
1,72	1,72 + 0,06 = **1,78**	
1,78	1,78 + 0,06 = **1,84**	
1,84	1,84 + 0,06 = **1,90**	
1,90	1,90 + 0,06 = **1,96** Temos que usar **1,98**	Aqui temos que verificar se o número encontrado é pelo menos igual ao Limite Superior do Rol. No caso, como o Limite Superior do Rol é igual a 1,98, temos que fazer um ajuste na última classe para contemplar todos os números do Rol.

Vamos montar a Tabela de Frequências.

<table>
<tr><th>Limite Inferior de Cada Classe</th><th>Limite Superior de Cada Classe</th><th>Frequência</th><th>Observação</th></tr>
<tr><td>1,48</td><td>1,54</td><td>6</td><td rowspan="4">Como fazer para montar a coluna de Frequências?
→</td></tr>
<tr><td>1,54</td><td>1,60</td><td>14</td></tr>
<tr><td>1,60</td><td>1,66</td><td>18</td></tr>
<tr><td>1,66</td><td>1,72</td><td>20</td></tr>
</table>

Limite Inferior de Cada Classe	Limite Superior de Cada Classe	Frequência	Observação
1,72	1,78	17	Base para você verificar na tabela de dados quantos elementos existem dentro das classes especificadas ao lado. Por exemplo, existem 6 alunos com altura entre 1,48 e 1,54; 14 alunos com altura entre 1,54 e 1,60 e assim por diante.
1,78	1,84	11	
1,84	1,90	8	
1,90	1,98	6	
		100	

Atividades

1. As notas obtidas em matemática por 80 estudantes de uma escola X estão relacionadas abaixo:

53	57	59	60	60	60	61	61	62	62	62	62	63	63	65	65
65	66	67	67	68	68	68	69	71	71	71	72	72	73	73	73
73	74	74	74	75	75	75	75	75	75	75	76	76	76	76	77
77	78	78	78	78	78	79	79	79	80	81	82	82	83	84	85
85	85	86	87	88	88	88	89	90	93	93	93	94	95	95	97

   a. Organize os dados em classes considerando a regra de Stuges;

   b. Faça a distribuição por frequências.

2. Numa amostra de 20 recém-nascidos normais, foram observados os seguintes pesos (em Kg):

2,450	2,455	2,458	2,510	2,513	2,516	2,600	2,608	2,620	2,621
2,625	2,629	2,650	2,653	2,680	2,686	2,670	2,672	2,720	2,726
2,750	2,758	2,830	2,833	2,845	2,849	2,850	2,852	2,862	2,866
2,870	2,874	2,876	2,883	2,885	2,890	2,900	2,905	2,950	2,959
3,010	3,053	3,058	3,200	3,259	3,304	3,359	3,420	3,436	3,570
3,599	3,602	3,610	3,620	3,650	3,690	3,750	3,800	3,850	3,890

a. Organize os dados em classes considerando a regra de Stuges;

b. Faça a distribuição por frequências;

c. Faça a frequência relativa;

d. Faça a frequência acumulada.

3. Em uma fábrica foram testadas 400 lâmpadas e a duração delas aparece na distribuição por frequência abaixo:

Duração (em horas)		**Número de lâmpadas** f_i	**PM = X_i**	**F_i**	**fr_i**	**Fr_i**
300	400	14				
400	500	46				
500	600	58				
600	700	76				
700	800	68				
800	900	62				
900	1.000	48				
1.000	1.100	22				
1.100	1.200	6				
Total		400				

Observando a tabela, responda:

a. Qual a amplitude de cada classe?

b. Qual a amplitude total da distribuição?

c. Monte o restante da tabela.

4. Complete o quadro de frequências abaixo:

Classes	f	fr (%)	Fr (%)
0|– 2	8	__	__
2|– __	__	__	40%
4|–6	20	25%	__
6|– __	__	__	__
8|–|10	__	15%	__
Total	__	__	__

Onde:

f – frequência absoluta

fr – frequência relativa

Fr – frequência relativa acumulada

5. Certo pesquisador aplicou um teste aos alunos de um colégio e obteve os seguintes resultados:

Acerto em determinado teste de conhecimento.

1,1		1,1	1,1	1,1	1,2	1,4	1,4	1,4	1,6	1,6	2,0	2,0	2,0	2,1	2,2
2,4		2,5	2,7	2,9	3,0	3,0	3,0	3,1	3,2	3,3	3,3	4,0	4,1	4,1	
4,2															
4,2		4,3	4,4	4,6	4,9	5,0	5,0	5,0	6,0	6,0	6,5	6,5	7,0	7,0	
7,2															
7,2		7,6	7,8	8,0	8,2	8,2	8,4	8,4	8,6	9,0	9,2	9,4	9,6	9,8	
9,8															

Pede-se para:

a. indicar a amplitude total;

b. determinar a amplitude das classes (usar regra de sturges);

c. fazer a distribuição por frequência.

6. Complete a distribuição com as frequências absoluta e relativa e, também, de frequências acumuladas, relativa e absoluta, para a tabela a seguir que representa uma amostra dos salários de 25 funcionários selecionados em uma empresa.

Classe	Salários (R$)	Nº de Funcionários	Frequência Relativa	Frequência Acumulada	Freq. Acum. Relativa
1	1000 – 1200	2			
2	1200 – 1400	6			
3	1400 – 1600	10			
4	1600 – 1800	5			
5	1800 - 2000	2			

De acordo com a tabela anterior, é correto afirmar que:

a. mais da metade dos funcionários ganha mais que R$ 1.600,00.

b. a maior parte dos funcionários ganha entre R$ 1.000,00 e R$ 1.400,00.

c. 40% dos funcionários ganham entre R$ 1.400,00 e R$ 1.600,00.

d. apenas 8% dos funcionários ganham mais de R$ 1.600,00.

3

Apresentação de Dados

Os gráficos encontram-se presentes em quase todos os meios de divulgação de informação, designadamente nos jornais e revistas, nos manuais escolares, nas apresentações públicas e até os nossos relatórios individuais já não passam sem eles. Contudo, fazer um gráfico ou um mapa que de fato informe e seja, simultaneamente, apelativo, legível e coerente com os dados não é tarefa fácil.

No capítulo anterior fizemos referência à utilização das tabelas como instrumento de análise e apresentação de dados estatísticos. Neste capítulo, formalizaremos a construção de tabelas e gráficos, apresentando as diferentes formas de apresentação de uma informação para que a análise possa ser feita de forma simples e fácil.

A apresentação gráfica é um complemento importante da apresentação tabular. A principal vantagem de um gráfico sobre a tabela prende-se ao fato de que ele permite conseguir uma visualização imediata da distribuição dos valores observados. Em contrapartida, pode-se perder alguns detalhes importantes, necessitando, muitas vezes, apresentar uma tabela junto do gráfico para que o observador possa entender melhor o fenômeno estudado.

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

Por outro lado, podemos dizer que colunas de números ou dados estatísticos dispersos são conhecidos por despertarem temor, medo, ansiedade, mal entendidos etc. Algumas pessoas parecem não prestar atenção em informações estatísticas apresentadas em forma de tabela, mas podem prestar bastante atenção aos mesmos dados quando eles são apresentados em forma de gráfico ou imagem.

Por isso, muitas vezes, é preferível a apresentação das informações em forma de gráfico. Mas, antes de construir um gráfico, é muito importante responder algumas perguntas:

- Para quem iremos apresentar as informações?
- Qual o nível de detalhe exigido?
- Qual a melhor forma de apresentação? Tabelas simples, tabelas por classes, gráficos?
- Se gráfico, qual gráfico?

Portanto, a representação gráfica das séries estatísticas tem por finalidade representar os resultados obtidos, permitindo que se chegue a conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da série. A escolha da tabela e/ou gráfico mais apropriado fica a critério do analista. Contudo, os elementos simplicidade, clareza e veracidade devem ser considerados, quando da elaboração de um gráfico.

- Simplicidade: deve ser destituído de detalhes de importância secundária, assim como de traços desnecessários que possam levar o observador a uma análise morosa ou sujeita a erros.

- Clareza: deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.
- Veracidade: deve expressar a verdade sobre o fenômeno em estudo.

Diretrizes para a construção de um gráfico:

- O título do gráfico deve ser o mais claro e completo possível. Quando necessário, deve-se acrescentar subtítulos;
- A orientação geral dos gráficos deve ser da esquerda para a direita;
- Sempre que possível, a escala vertical há de ser escolhida de modo a aparecer a linha 0 (zero);
- Só devem ser incluídas no desenho as coordenadas indispensáveis para guiar o olhar do leitor ao longo da leitura. Muitas linhas de referência, sejam elas horizontais ou verticais, podem atrapalhar a leitura;
- A escala horizontal deve ser lida da esquerda para a direita, e a vertical de baixo para cima;
- Os títulos e marcações do gráfico devem ser dispostos de maneira que sejam facilmente lidos, partindo da margem horizontal inferior ou da margem esquerda.

Leitura e interpretação de um gráfico:

- Declarar qual o fenômeno ou fenômenos representados, a região considerada, o período de tempo, a fonte dos dados etc;
- Examinar o tipo de gráfico escolhido, verificar se é o mais adequado, criticar a sua execução, no conjunto e nos detalhes;
- Analisar cada fenômeno separadamente, fazendo notar os pontos mais em evidência, o máximo e o mínimo, assim como as mudanças mais bruscas.

Muitas vezes, o uso indevido dos gráficos pode trazer uma ideia falsa dos dados que estão sendo analisados, chegando mesmo a confundir o leitor. Vejamos, através de um exemplo, como esse fato pode ocorrer. Os três gráficos apresentados a seguir representam o mesmo fenômeno com mesmo grupo de dados. A primeira impressão é a de que os três gráficos representam dados nitidamente diferentes.

Gráfico 1 – Vendas Mensais

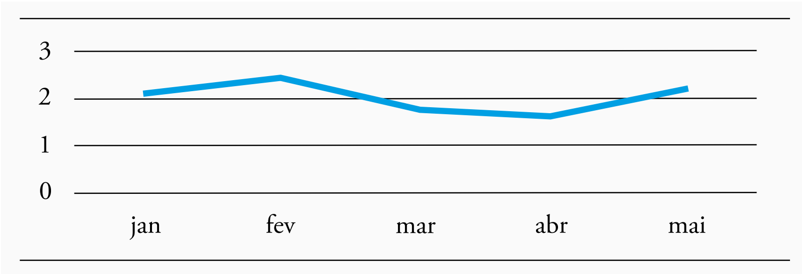


Gráfico 2 – Vendas Mensais

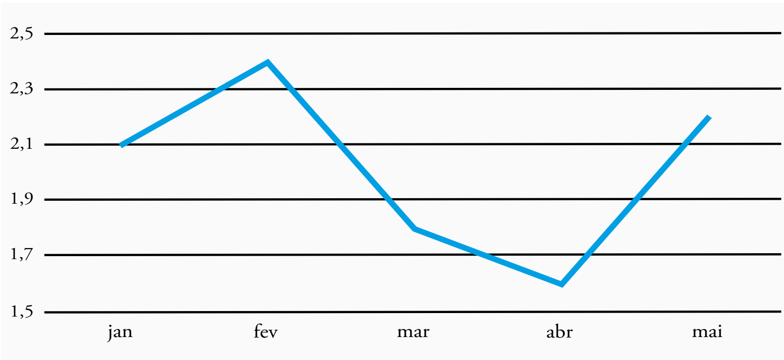
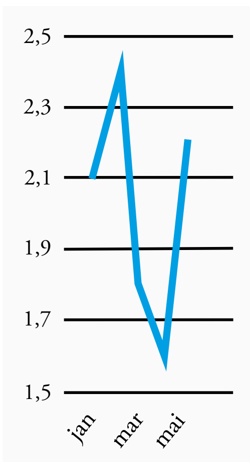


Gráfico 3 – Vendas Mensais



Podemos perceber que no gráfico 2, as flutuações das vendas parecem ser moderadas; já no gráfico 1, tem-se a impressão de que a flutuação das vendas não manifesta praticamente tendência nenhuma; no gráfico 3, a impressão é de as vendas flutuaram muito durante o período estudado.

Temos que tomar muito cuidado com a construção das escalas e no tamanho do gráfico para que o leitor possa ter a análise correta.

Examinando superficialmente os três gráficos e não prestando muita atenção aos detalhes, o leitor receberá impressões diferentes sobre a flutuação das vendas.

Mas não se esqueça de que, todo gráfico, assim como as tabelas, tem na sua parte superior um título. Na parte inferior, opcionalmente, é fornecida a fonte que apresentou o gráfico ou os dados que permitiram a sua construção.

Há uma infinidade de gráficos, nós vamos apresentar os principais tipos a serem utilizados na estatística.

3.1 Gráfico de Colunas

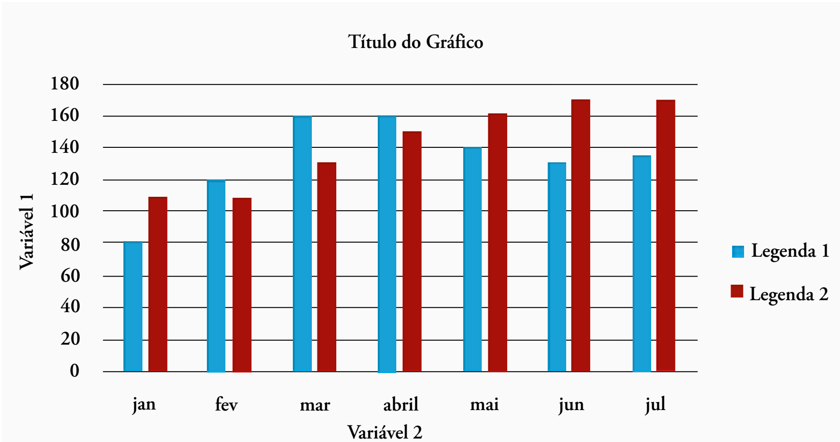
O Gráfico de Colunas é uma das formas mais populares de representar informação, em parte pela facilidade quer de execução, quer de leitura.

Um gráfico de colunas serve para representar a frequência de ocorrências de uma série expressa em números absolutos ou em porcentagens, em um eixo cartesiano.

Servem para apresentar um conjunto de dados e também para comparar vários conjuntos de dados.

Devem ser utilizadas para apresentar variáveis discretas ou qualitativas, em termos absolutos ou relativos, ou para comparar categorias de variáveis quantitativas. Podem, igualmente, representar a evolução de uma variável ao longo do tempo.

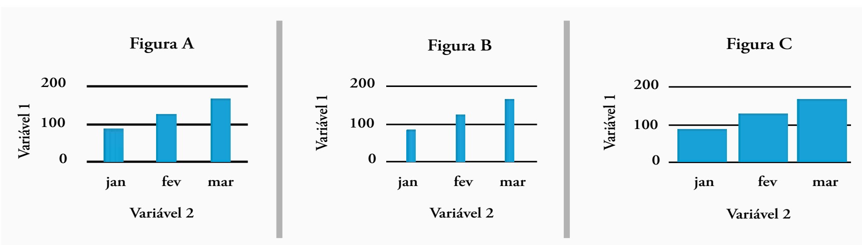
Aproveitamos o exemplo a seguir para demonstrar a localização dos principais componentes de um gráfico.



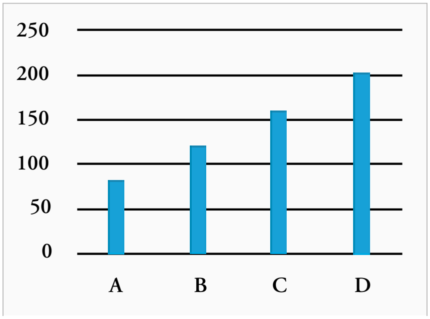
- **Título do Gráfico:** O título deve estar presente em qualquer tipo de representação gráfica e ser escrito com vista a orientar o leitor na sua interpretação. Para tal, deve ser redigido por forma a responder às perguntas: O Quê, Onde e Quando. Simultaneamente, deve ser conciso, relevante e claro, ou seja, conter apenas informação essencial para uma interpretação correta do gráfico.
- **Identificações (Variáveis 1 e 2):** Neste conceito genérico, enquadra-se toda informação escrita posicionada na área exterior: as designações dos eixos de valores e categorias, a referência às respectivas unidades e eventuais notas. Na maior parte dos gráficos ou tabelas não se justifica uma grande precisão nos dados apresentados. Um número excessivo de casas decimais (separadas das unidades por uma vírgula), ou mesmo uma casa decimal em valores elevados, envolve um rigor desnecessário e prejudicial à leitura. Para ser mais legível, a formatação de valores acima dos milhares pode ser feita com um espaço em vez de um ponto ou uma vírgula. Os valores da escala devem ser expressos com valores arredondados múltiplos de 1, 2 e 5 (ex. 5, 10, 25, 50, 100, etc.). Aconselha-se a que não se apresentem números com mais de 5 dígitos, adaptando, caso seja preciso, a unidade para milhares ou milhões.
- **Legenda:** Uma boa legenda deve fazer mais do que simplesmente etiquetar os componentes do gráfico. Deve dizer-nos o que é impor-

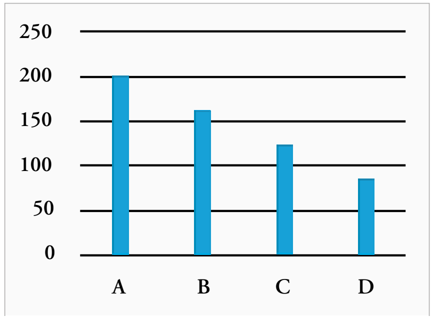
tante e qual é o objetivo do gráfico: informar o leitor e obrigar quem faz o gráfico a estruturar a informação.

- **Eixo de categorias ou variáveis:** Neste eixo estão posicionadas as variáveis ou categorias que se pretendem retratar. No caso de gráficos que representam séries que evoluem ao longo do tempo, a este eixo estão associados os períodos temporais, em que a cada mês, trimestre, ano ou outro, corresponderá apenas um ponto ou uma barra no gráfico.
- **Linhas auxiliares:** As linhas auxiliares existem para ajudar o sistema visual a fazer comparações e ler valores aproximados. Um gráfico com demasiadas linhas auxiliares dá mais peso visual do que deve a estes elementos secundários, sem que daí advenham vantagens significativas ao nível da leitura de valores aproximados. Por outro lado, um gráfico com poucas linhas auxiliares não traz grande valor acrescentado à leitura.
- **Espaço entre as barras:** Os espaços entre as barras devem estar construídos de forma a que não se dificulte a comparação (Figura B) nem se assemelhe a um histograma (C), sugerindo uma continuidade quando, afinal, a variável representada é discreta. É aconselhado um espaço entre as barras aproximadamente igual ao tamanho das mesmas (A)

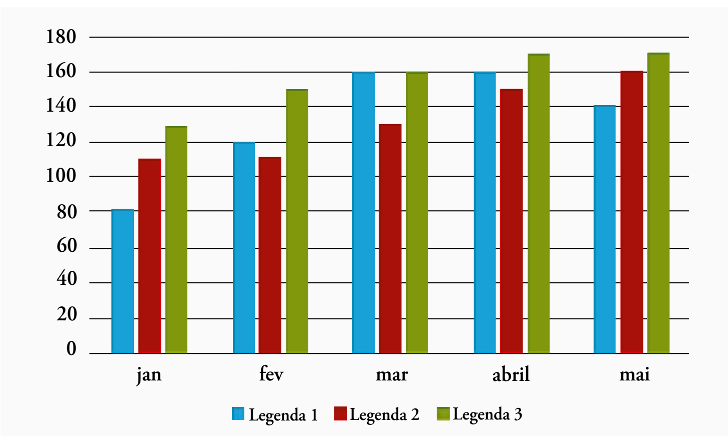


- **Ordenação:** Na representação da informação, por vezes, é importante organizar as categorias por ordem crescente ou decrescente para melhor compreender certos fenômenos implícitos. É igualmente comum ordenar alfabeticamente (ou geograficamente) as designações das categorias, nomeadamente nos casos em que se representam países ou outro tipo de unidades administrativas, mas isso nem sempre é a melhor opção. Se o mesmo conjunto de categorias é apresentado em mais do que um gráfico, então a posição relativa de cada categoria deve manter-se, ou seja, as categorias devem aparecer na mesma ordem em todos os gráficos. Da mesma forma, o tamanho e a escala dos gráficos devem ser os mesmos, se o objetivo for a comparação entre eles.



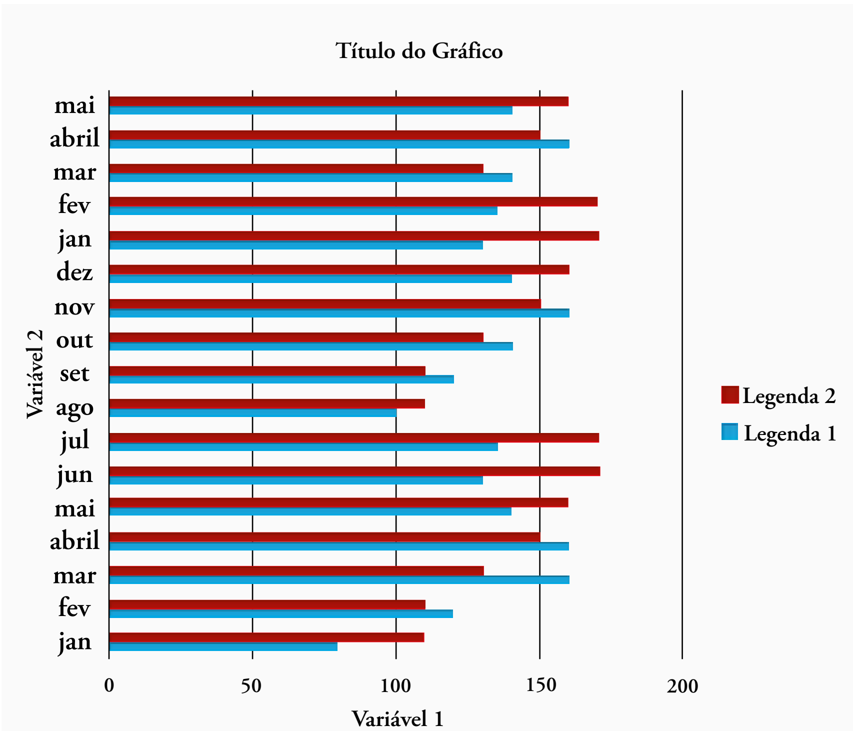


- **Gráfico de barras agrupadas:** Os gráficos de barras agrupadas são utilizados para descrever, simultaneamente, duas ou mais categorias, para uma dada variável discreta, ou quando se pretende realçar o valor das categorias em detrimento do valor total das variáveis. As diferentes categorias são representadas por barras sendo a distinção entre elas feita recorrendo às variáveis visuais (cor ou valor). Os grupos de entidades devem estar separados por um espaço em branco, mas não deve existir qualquer espaço entre as categorias de cada grupo.



3.2 Gráfico de Barras

O Gráfico de Barras é bem parecido com o Gráfico de Colunas, a diferença é que utilizam-se barras horizontais em vez de colunas verticais. Este gráfico deve ser utilizado quando o número da "Variável 2" for muito grande, conforme apresentado abaixo, ou quando deseja-se visualizar a diferença entre o valor mínimo e o valor máximo da variável.



3.3 Gráfico de Colunas ou Barras Agrupadas

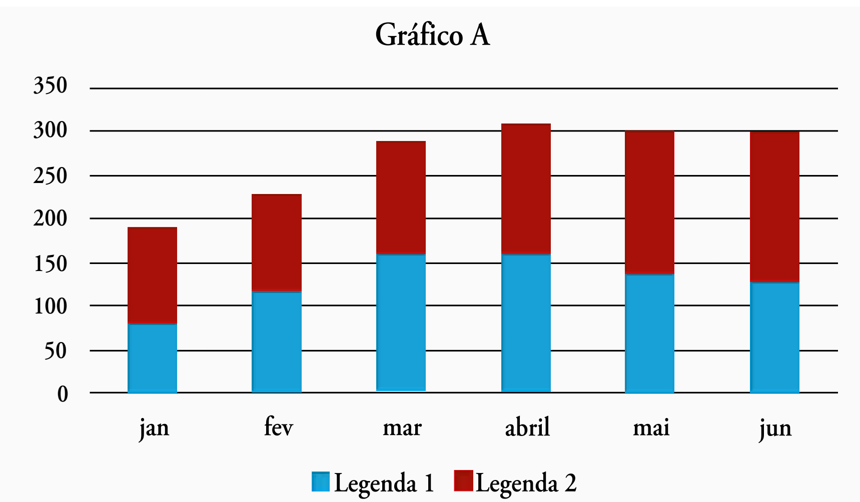
Recorre-se aos gráficos de colunas ou barras empilhadas quando o conjunto de dados contém duas ou mais categorias. Neste tipo de gráfico, cada barra subdivide-se em pelo menos duas categorias, com distintas cores ou padrões, permitindo mostrar a relação entre cada categoria (Homens/Mulheres) e o respectivo subtotal (ex: Comércio e Administração). As categorias surgem assim posicionadas umas sobre as outras, se for um gráfico de barras vertical (ou lado a lado, se o gráfico for horizontal), sendo que a altura (ou a largura) de cada componente corresponde ao valor absoluto ou relativo da categoria.

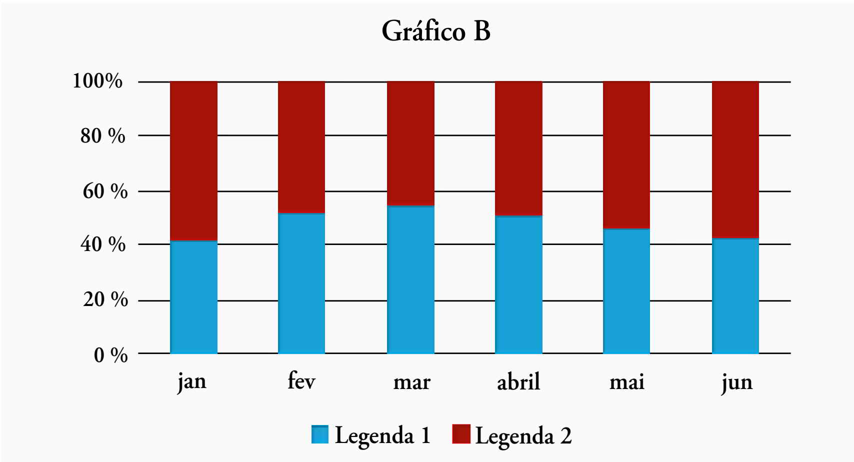
O gráfico em valor absoluto (A) adequa-se aos casos em que se pretende evidenciar mais o valor total das variáveis do que das respectivas categorias, dado que o todo é apreendido com maior precisão do que as partes. Tal precisão advém de, para o total, ser comparada a posição relativa numa mesma escala, enquanto que na estimação dos valores das categorias são confrontados e ordenados os tamanhos respectivos.

Se o maior objetivo destes gráficos é indicar graficamente a soma total, mais do que estimar visualmente as respectivas categorias, valerá então a pena questionar por que não se opta por representar apenas o total ou então substituir por outra forma de representação.

No gráfico em valor relativo (B) apenas se pode estimar o valor das categorias observando o tamanho das barras que lhes correspondem.

Com duas categorias torna-se mais fácil estimar os valores, dado que a base e o topo da escala servem de ponto de referência, mas com mais de duas categorias a leitura é consideravelmente mais difícil.





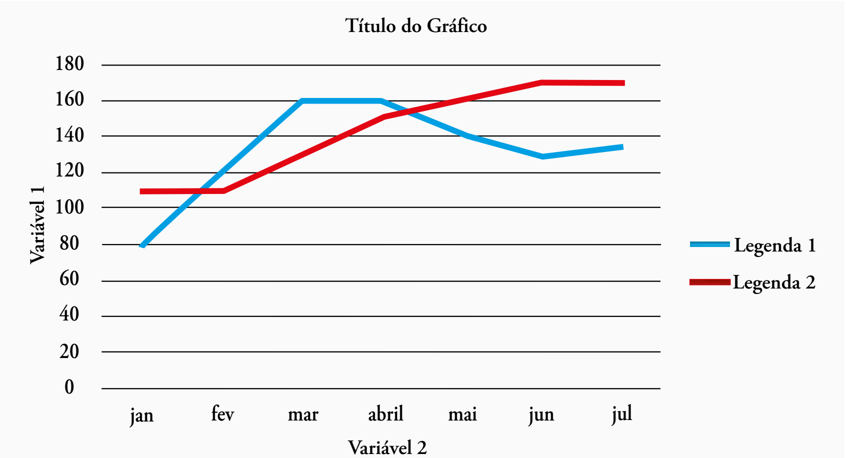
3.4 Gráfico de Linhas

O Gráfico de Linhas utiliza-se de linhas para unir pontos situados em um plano cartesiano, sendo excelente para apresentação de séries históricas ou temporais, pois permite avaliar com clareza a evolução de um fenômeno ao longo do tempo. Presta-se ainda para a representação da frequência acumulada de dados. Os procedimentos de elaboração são semelhantes aos do gráfico em colunas.

No caso das séries temporais, costuma-se dispor a variação quantitativa dos dados, no eixo y (vertical); no eixo x (horizontal), dispõe-se a variação temporal da série.

O mais comum é aquele que representa séries temporais (ou cronológicas), em que uma determinada variável contínua é analisada ao longo do tempo. O eixo do y mede as variáveis em estudo, enquanto o eixo do x apresenta as unidades temporais dispostas cronologicamente em intervalos iguais de tempo, começando à esquerda com a data mais antiga.

Num gráfico de linhas, ao contrário dos gráficos de barras, as séries podem ser longas. O objetivo nestes gráficos é comparar os declives das curvas por forma a responder a perguntas do tipo: Em que períodos a variação foi significativa? Quando foram os pontos de inflexão?



Utilizando-se de legenda composta por linhas de cores, formas ou de texturas diferentes, pode-se representar algumas variações de um mesmo tema, permitindo inclusive boas comparações entre estas, no entanto, é preciso cautela para não se exagerar no uso desse recurso. É importante lembrar que o título e a fonte são elementos que devem estar presentes em todos os gráficos.

3.5 Gráfico de Circular, Setorial ou de Pizza

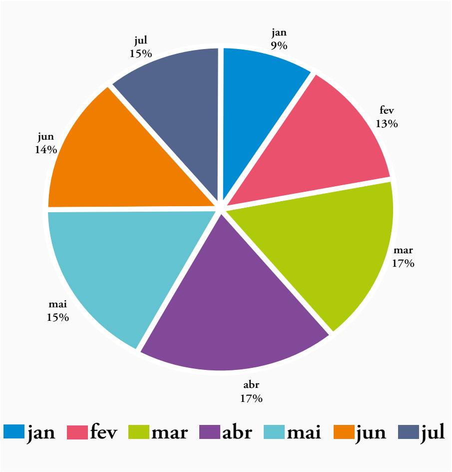


Gráfico circular é um excelente recurso para comparar os valores de cada parcela de um conjunto de dados com o total. É feito tomando por base a figura de um círculo dividido em setores de tamanhos proporcionais aos valores que representam. O somatório dos setores deverá ser sempre de 360º ou 100% dos dados.

Algumas recomendações devem ser feitas para a elaboração dos gráficos setoriais. Os valores devem ser apresentados em ordem decrescente a partir da parte superior do gráfico e no sentido horário. Ao lado de cada setor, podem-se colocar os percentuais e os nomes de cada parcela. Este gráfico não deve ser usado quando são muitas parcelas, nem quando existem muitas parcelas com valores muito semelhantes, sob pena de se perder uma das suas principais funções: a da comparação.

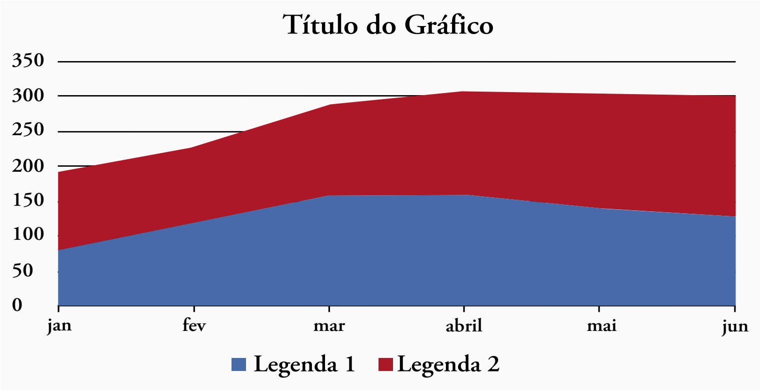
A sua utilização é desaconselhada quando se pretende comparar mais do que um período temporal, para variáveis que contenham mais de cinco componentes ou quando os componentes têm aproximadamente o mesmo peso, sendo neste caso, preferível substituir o gráfico circular por um gráfico de barras.

Muitas fatias ou fatias demasiadamente estreitas são dificilmente interpretáveis, sendo por isso necessário complementar o gráfico com os valores respectivos ou associar um subconjunto de valores a outro gráfico circular de tamanho proporcional à quantidade que representa.

3.6 Gráfico de Área

Recorre-se aos gráficos de área quando se pretende visualizar simultaneamente a evolução do total e dos respectivos componentes. Tal como nos gráficos de barras empilhados, existem poucas vantagens nesta forma de apresentação dado não ser possível responder de forma imediata a perguntas sobre o crescimento ou decréscimo ao longo do tempo, sobretudo quando o primeiro dos componentes apresenta oscilações significativas.

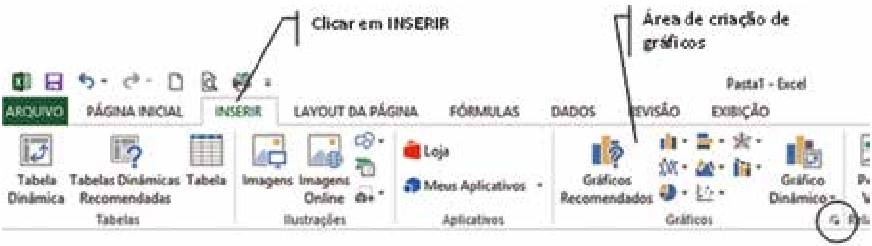
Os gráficos de área são utilizados como alternativa aos gráficos de linhas. No entanto, trazem dificuldades acrescidas quando as áreas se intersectam porque deixa de ser possível seguir a evolução dos componentes.



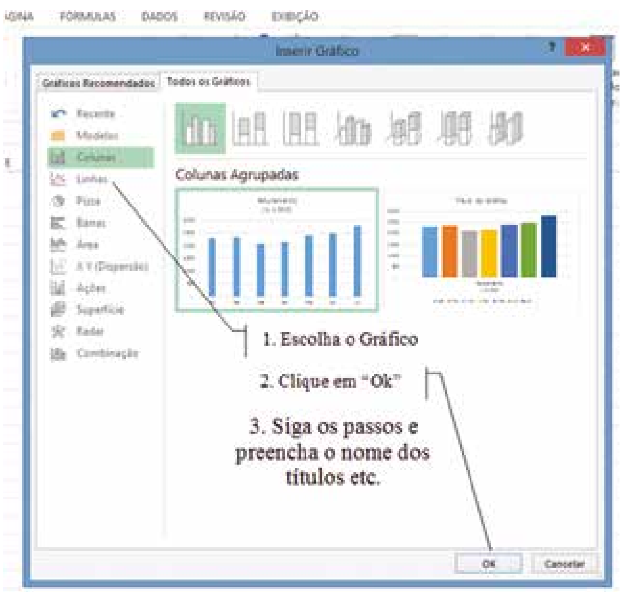
3.7 Gráficos no Excel

O Excel possui diversos tipos de gráficos e a escolha adequada depende basicamente do tipo de dado e da finalidade da apresentação. Para construir um gráfico, primeiramente você deve ter uma tabela preenchida no Excel.

Barra de Ferramentas do Excel



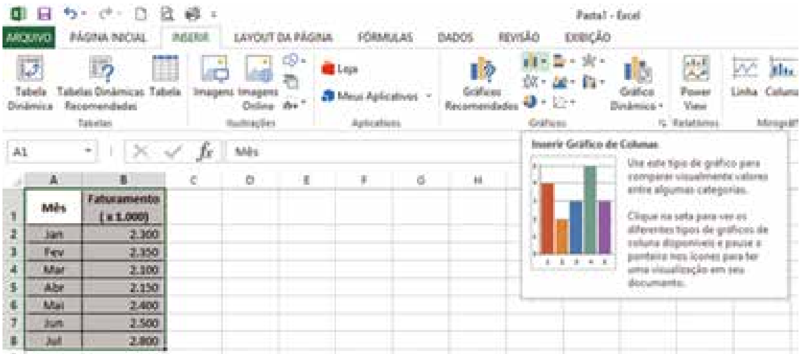
Após verificar a área de criação de gráficos, conforme figura acima, você pode simplesmente marcar a tabela desejada e clicar no gráfico desejado, na barra de ferramentas acima, ou clicar no botão a direita da parte inferior, marcada com um círculo, para verificar outras possibilidades de gráfico, conforme figura abaixo.



Exemplo: Gráfico no Excel

1. Primeiro marque uma tabela no Excel.

2. Selecione o Gráfico desejado, por exemplo, Gráfico de Colunas



3. Clique no Gráfico desejado, seu gráfico está pronto.

4. Verifique que você pode alterar o nome do gráfico simplesmente clicando no nome e digitando o novo nome.

5. Na parte superior da "Barra de Ferramentas, você pode formatar, alterar e modificar seu gráfico.
6. Se você clicar com o botão da direita do mouse nas barras ou em qualquer parte do gráfico, você poderá alterar cada parte para ficar do "seu jeito".

Atividades

1. Gráfico de Colunas

Arrecadação Bruta 20XX

Mês	R$ (Milhões)
Mar	34.888
Abr	52.334
Mau	85.023
Jun	95.254
Jul	136.126
Ago	162.643
Total	566.268

2. Gráfico Circular (Pizza)

Consumo Industrial de Energia 20XX

Empresas	KWH (Milhões)
SP – Light	13.617
Cemig	6.763
RJ – Light	3.226
Copel	1.183
Escelsa	1.258
Total	26.047

3. Gráfico Linear

Produção de Pneus (São Paulo) 20XX

Empresas	Pneus (× 1.000)
Jan	176
Fev	152
Mar	183
Abr	171
Mai	195
Jun	294
Total	1.171

4. Gráfico de Colunas

Estatura de Alunos de um Colégio

Altura	Alunos
150 ⊢ 155	05
156 ⊢ 161	09
162 ⊢ 167	19
168 ⊢ 173	18
174 ⊢ 179	14
180 ⊢ 185	12
186 ⊢ 191	04
Total	81

4

Medidas de Posição

As MEDIDAS DE posição, que serão objeto desse capítulo, podem-se apresentar de várias formas, dependendo daquilo que se pretende conhecer a respeito dos dados estatísticos. As mais importantes são as medidas de tendência central, as quais são assim denominadas, em virtude da tendência de os dados observados se agruparem em torno desses valores centrais.

A MODA, A média aritmética e a mediana são as três medidas mais utilizadas para resumir o conjunto de calores representativos do fenômeno que se deseja estudar.

4.1 Média

A medida de tendência central mais comumente usada para descrever resumidamente uma distribuição de frequências é a média aritmética, sendo que em certos casos podem ser utilizados outros tipos de médias: média geométrica, média harmônica, média quadrática, média cúbica ou média biquadrática.

4.1.1 Média Aritmética

a) **Média Aritmética Simples - ($\overline{x}$)**

A média aritmética simples ou simplesmente média aritmética de um conjunto de números é igual ao quociente entre a soma dos valores do conjunto e o número total de valores. Genericamente, podemos escrever:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}\text{, onde}$$

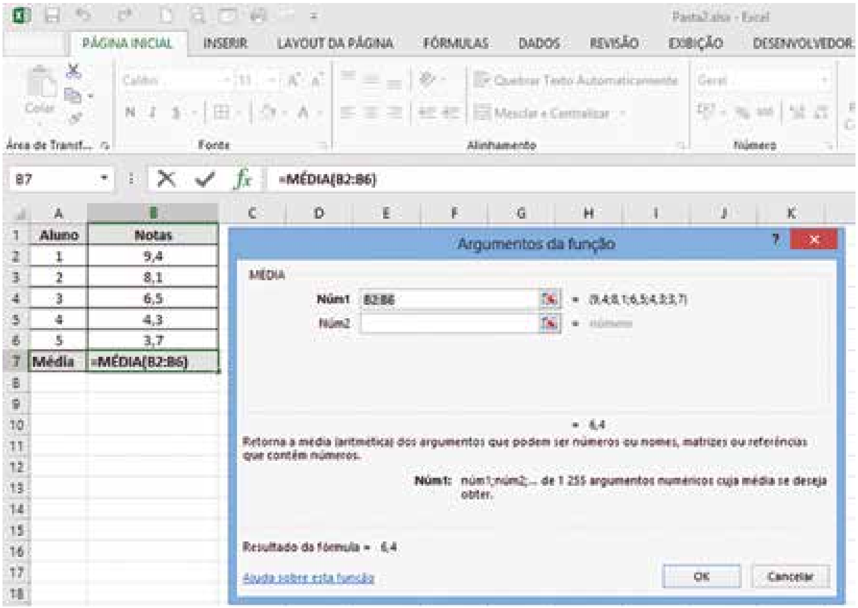
- x_i – representa cada valor do conjunto estudado;
- n – representa o total de elementos do conjunto.

Exemplo 1: Suponha que em uma sala de aula de 5 alunos, as notas em Estatística foram as seguintes: 9,4; 8,1; 6,5; 4,3 e 3,7. A média aritmética da turma na disciplina será calculada da seguinte forma:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{9,4+8,1+6,5+4,3+3,7}{5} = \frac{32}{5} = 6,4$$

No exemplo acima, n é igual a 5, cinco elementos do conjunto, que são respectivamente: x_1 = 9,4; x_2 = 8,1; x_3 = 6,5; x_4 =4,3 e x_5 =3,7.

Exemplo no Excel: No Excel, podemos calcular a média aritmética simples utilizando a função "MÉDIA". Conforme demonstrado abaixo, podemos verificar como calcular a média do exemplo anterior utilizando o Excel.



No Excel, a média é calculada com os valores compreendidos entre as células B2 e B6, com a notação "MÉDIA(B2:B6), o Excel entende que você deseja calcular a média dos valores contidos das células B2, B3, B4, B5 e B6. Então você pode entender que B2:B6 quer dizer "B2 até B6".

b) **Média Aritmética Ponderada -** $\left(\overline{x}_p\right)$

A média aritmética é considerada ponderada quando os valores do conjunto tiverem pesos diferentes. Tratando-se da média simples, todos os valores apresentam igual peso. Obtém-se uma média aritmética ponderada através do quociente entre o produto dos valores da variável pelos respectivos pesos e a soma dos pesos.

Exemplo 2: Um professor pode realizar 4 provas por ano em sua matéria, atribuindo a cada uma delas os seguintes pesos: 1, 2, 3 e 4; respectivamente. Se um aluno tiver recebido as notas: 8, 7, 9 e 9, nessa ordem, sua nota final será a média aritmética ponderada "8,5"; obtida da seguinte forma:

$$\text{Média Final} = \overline{x} = \frac{(8\times1)+(7\times2)+(9\times3)+(9\times4)}{1+2+3+4} = \frac{85}{10} = 8{,}5$$

O mesmo resultado seria obtido se fossem adotados pesos relativos, como indicado na tabela abaixo:

Provas	Pesos Relativos	Notas	Produtos
1ª	(1/10) = 0,1	8	0,1 x 8 = 0,8
2ª	(2/10) = 0,2	7	0,2 x 7 = 1,4
3ª	(3/10) = 0,3	9	0,3 x 9 = 2,7
4ª	(4/10) = 0,4	9	0,4 x 9 = 3,6
Soma dos pesos Relativos	1,0	Média Ponderada	0,8+1,4+2,7+3,6 = **8,5**

Quando se usam pesos relativos, o denominador será sempre igual à unidade, e a média aritmética ponderada será igual à soma dos produtos dos valores da variável pelos respectivos pesos relativos. A soma dos pesos relativos é sempre igual a 1.

Exemplo 3: No exemplo anterior, onde o professor realizou 4 provas no ano, atribuindo a cada uma delas os seguintes pesos: 1, 2, 3 e 4; respectivamente. O aluno obteve as seguintes notas: 8, 7, 9 e 9, nessa ordem, e ficou com a média igual a 8,5. Se outro aluno tirou as mesmas notas com a ordem inversa (9, 9, 7 e 8), qual seria a média desse aluno?

$$\text{Média Final} = \overline{x} = \frac{(9\times1)+(9\times2)+(7\times3)+(8\times4)}{1+2+3+4} = \frac{80}{10} = 8{,}0$$

Obs.: Por que o segundo aluno tirou média menor do que o primeiro, se as notas foram iguais? Como as notas foram iguais, mas, o peso de cada prova é diferente, a melhor nota na última prova aumenta o valor da média, mais do que uma melhor nota na primeira prova.

1. Dados Agrupados em Tabelas

Exemplo 4 : Admitamos que as notas atribuídas a vinte alunos em um teste de estatística sejam as seguintes:

Notas (x_j)	**Nº de Alunos f_j**	**$x_j.f_j$**
4	1	4x1 = 4
5	5	5x5 = 25
6	6	6x6 = 36
7	5	7x5 = 35
8	3	8x3 = 24
	$\sum_{j=1}^{5} f_j = 20 = n$	$\sum_{j=1}^{5} x_j.f_j = 124$

- K = número de classes = 5
- n = número total de observações = 20

então,
Média Aritmética = $\dfrac{\sum_{j=1}^{k} x_j.f_j}{\sum_{j=1}^{k} f_j} = \dfrac{\sum x.f}{n} = \dfrac{\sum_{j=1}^{5} x_j.f_j}{\sum_{j=1}^{5} f_j} = \dfrac{124}{20} = 6,2$

onde: K = número de classes = 5

n = número total de observações = 20

Exemplo 5: Quando os valores estão agrupados em classes, a tabela requer mais uma coluna, necessária para dispor os pontos médios de classes, como indicado na seguinte tabela:

Classes	**Frequência (f_j)**	**Ponto Médio (x_j)**	**$x_j.f_j$**
10 – 20	5	(10+20)÷2=**15**	15×5=**75**
20 – 30	10	(20+30)÷2=**25**	25×10=**250**
30 – 40	15	(30+40)÷2=**35**	35×15=**525**
40 – 50	10	(40+50)÷2=**45**	45×10=**450**
50 – 60	5	(50+60)÷2=**55**	55×5=275
	$\sum f_j = n = 45$		$\sum x_j.f_j = 1.575$

$$\text{Média Aritmética} = \overline{x} = \frac{\sum x_j . f_j}{n} = \frac{1.575}{45} = 35$$

c) **Propriedades da Média Aritmética**

1. **Primeira Propriedade:** a soma algébrica dos desvios de um conjunto de números tomados em relação à média aritmética é zero, simbolicamente:

para dados brutos: $\sum_{i=1}^{k} d_i = \sum (x_i - \overline{x}) = 0$

para dados tabulados: $\sum_{i=1}^{k} d_i . f_i = \sum (x_i - \overline{x}) . f_i = 0$, onde d_i = desvio;

Exemplo 6: Considerando o "exemplo 3", onde a média aritmética das notas é igual a 6,2, segue que:

Notas (x_j)	$d_j = x_j - \overline{x}$
4	4 – 6,2 = -2,2
5	5 – 6,2 = -1,2
6	6 – 6,2 = -0,2
7	7 – 6,2 = 0,8
8	8 – 6,2 = 2,8
$\sum_{j=1}^{5} d_j = \sum_{j=1}^{5} (x_j - 6,2) = 0$	

2. **Segunda Propriedade:** média ponderada de todas as médias. Se n_1 números têm média $\overline{x}_1$, n_2 números têm média $\overline{x}_2$, ... , n_k números têm média $\overline{x}_k$, a média do conjunto formado por todos os números é dada pela expressão.

$$\overline{x} = \frac{\overline{x}_1 n_1 + \overline{x}_2 n_2 + \ldots\ldots + \overline{x}_k n_k}{n_1 + n_2 + \ldots\ldots + n_k} = \frac{\sum_{j=1}^{k} \overline{x} j n_j}{\sum_{j=1}^{k} n_j}$$

3. **Terceira Propriedade:** somando-se (ou subtraindo-se) um valor constante e arbitrário a cada um dos elementos de um conjunto de números, a média aritmética fica somada (ou subtraída) por essa constante.

Exemplo 7: Considerando o "exemplo 3", onde a média aritmética das notas é igual a 6,2; segue que, se somarmos a todas as notas uma constante igual a 1,0; a média aritmética será acrescida de 1,0; ficando então igual a 7,2, confira no quadro abaixo:

Notas (x_j)	Nova Nota ($x_j + 1,0$)	Nº de Alunos f_j	$x_j.f_j$
4	4+1=5	1	5x1=5
5	5+1=6	5	6x5=30
6	6+1=7	6	7x6=42
7	7+1=8	5	8x5=40
8	8+1=9	3	9x3=27
		$\sum_{j=1}^{5} f_j = 20 = n$	$\sum_{j=1}^{5} x_j.f_j = 144$

- K = número de classes = 5
- n = número total de observações = 20

então, Média Aritmética = $\dfrac{\sum_{j=1}^{k} x_j.f_j}{\sum_{j=1}^{k} f_j} = \dfrac{\sum x.f}{n} = \dfrac{\sum_{j=1}^{5} x_j.f_j}{\sum_{j=1}^{5} f_j} = \dfrac{144}{20} = 7,2$

4. **Quarta Propriedade:** Multiplicando-se (ou dividindo-se) cada elemento de um conjunto de número por um valor constante e arbitrário, a média fica multiplicada (ou dividida) por essa constante.

Exemplo 8: Considerando o "exemplo 3", onde a média aritmética das notas é igual a 6,2; segue que, se dividirmos todas as notas uma constante igual a 2,0; a média aritmética será dividida de 2,0; ficando então igual a 3,1; confira no quadro a seguir:

Notas (x_j)	Nova Nota (x_j ÷ *2,0*)	Nº de Alunos f_j	$x_j.f_j$
4	4÷2=2	1	2x1=2
5	5÷2=2,5	5	2,5x5=12,5
6	6÷2=3	6	3x6=18
7	7÷2=3,5	5	3,5x5=17,5
8	8÷2=4	3	4x3=12
		$\sum_{j=1}^{5} f_j = 20 = n$	$\sum_{j=1}^{5} x_j.f_j = 62$

- K = número de classes = 5
- n = número total de observações = 20

então, Média Aritmética = $\dfrac{\sum_{j=1}^{k} x_j.f_j}{\sum_{j=1}^{k} f_j} = \dfrac{\sum x.f}{n} = \dfrac{\sum_{j=1}^{5} x_j.f_j}{\sum_{j=1}^{5} f_j} = \dfrac{62}{20} = 3,1$

4.1.2 Média Geométrica

A média geométrica de *n* valores é definida, genericamente, como a raiz *n-ésima* do produto de todos eles. A média geométrica pode ser simples ou ponderada, conforme se utilize ou não em seu cálculo uma tabela de frequências.

a) **Média Geométrica Simples**

Dados *n* valores x_1, x_2, ... x_n, a média geométrica desses valores será:

$$\overline{x}_g = \sqrt[n]{x_1 \times x_2 \times \ldots.. \times x_n}$$

b) **Média Geométrica Ponderada**

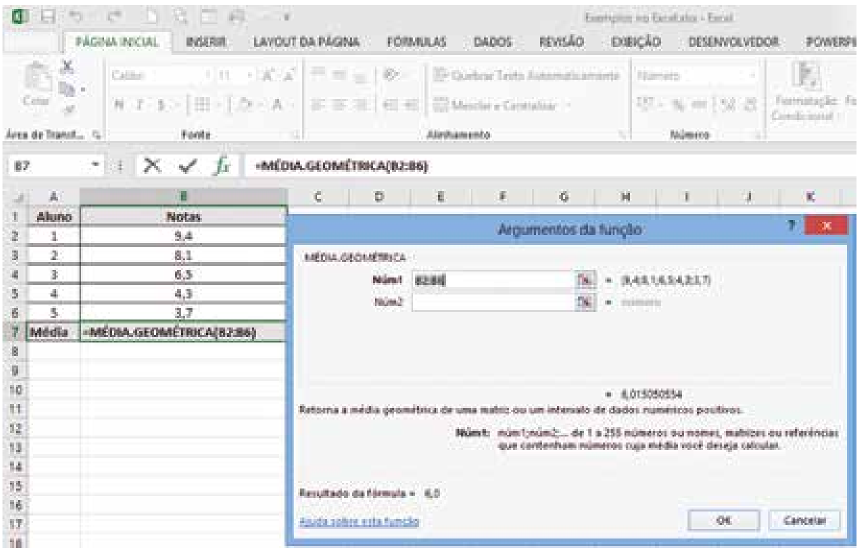
A média geométrica ponderada de um conjunto de números dispostos em uma tabela de frequência é calculada por intermédio da seguinte expressão:

$$\overline{x}_g = \sqrt[\sum_{j=1}^{k} f_j]{x_1^{f_1} \times x_2^{f_2} \times \ldots.. \times x_k^{f_k}} = \sqrt[n]{x_1^{f_1} \times x_2^{f_2} \times \ldots.. \times x_k^{f_k}}$$

Exemplo 9: Suponha que em uma sala de aula de 5 alunos, as notas em Estatística foram as seguintes: 9,4; 8,1; 6,5; 4,3 e 3,7. A média geométrica da turma na disciplina será calculada da seguinte forma:

$$\overline{x}_g = \sqrt[5]{9,4 \times 8,1 \times 6,5 \times 4,3 \times 3,7} = 6,0$$

Exemplo no Excel: No Excel, podemos calcular a média geométrica simples utilizando a função "MÉDIA.GEOMÉTRICA". Conforme demonstrado abaixo, podemos verificar como calcular a média do exemplo anterior utilizando o Excel:



No Excel, a média é calculada com os valores compreendidos entre as células B2 e B6, com a notação "MÉDIA.GEOMÉTRICA(B2:B6), o Excel entende que você deseja calcular a média dos valores contidos das células B2, B3, B4, B5 e B6. Então você pode entender que B2:B6 quer dizer "B2 até B6".

Exemplo 10: Calcular a média geométrica com os dados do "Exemplo 4".

$$\overline{x}_g = \sqrt[20]{4^1 \times 5^5 \times 6^6 \times 7^5 \times 8^3} = 6,09577 = 6,1$$

c) Propriedades da Média Geométrica

1. **Primeira Propriedade:** o produto dos quocientes de cada valor de um conjunto de números pela média geométrica do conjunto é igual a um.

$$\frac{x_1}{\overline{x}_g} \times \frac{x_2}{\overline{x}_g} \times \ldots\ldots \times \frac{x_n}{\overline{x}_g} = 1$$

2. **Segunda Propriedade:** a média geométrica é menor ou igual à média aritmética, $\overline{x}_g \leq \overline{x}$.

d) Aplicações da Média Geométrica

- A Média Geométrica deve ser utilizada quando os dados se desenvolvem segundo uma progressão geométrica, como é o caso dos preços num período de inflação grande.
- Média em Distribuição Assimétrica: uma distribuição de frequência pode encontrar-se deformada à direita ou à esquerda (assimétrica). Contudo se usarmos os logaritmos dos valores da variável, com um intervalo de classe constante para os logaritmos, a curva se transformará em simétrica. Neste caso, a média geométrica revela-se mais apropriada que a aritmética.
- Médias de Taxas de Variação: A média geométrica é usada em certas ocasiões para determinar taxas médias. Assim, por exemplo, suponhamos que um indivíduo tenha investido um capital de R$ 500,00 em 1990. Após um ano de aplicação, essa importância ascendeu a R$ 650,00. Reaplicando, ao final de mais um ano seu montante ficou igual a R$910,00. A taxa média de aumento de capital será obtida mediante o cálculo de uma média geométrica:

Período	Taxa
1990 – 1991	650 ÷ 500 = **1,3**
1991 - 1992	910 ÷ 650 = **1,4**

Taxa Média = $\sqrt{1,3 \times 1,4} = 1,3491$

4.1.3 Média Harmônica

A média harmônica de um conjunto de valores é o inverso da média aritmética dos inversos.

a. Média Harmônica Simples

Dado o conjunto de n valores $x_1, x_2, \ldots\ldots x_n$, a média harmônica do conjunto será:

$$\bar{x}_h = \frac{1}{\frac{1/x_1 + 1/x_2 + \ldots + 1/x_n}{n}} = \frac{n}{1/x_1 + 1/x_2 + \ldots + 1/x_n} = \frac{n}{\sum_{i=1}^{n} 1/x_i}$$

b. Média Harmônica Ponderada

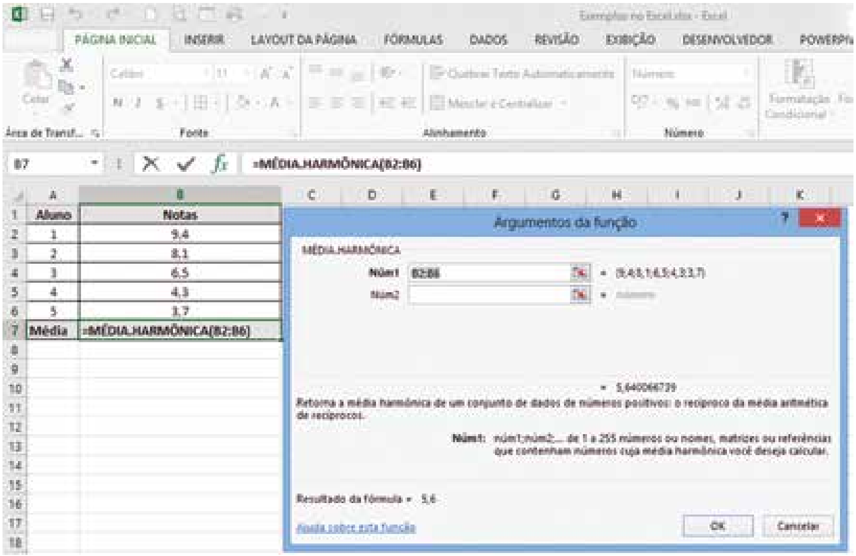
A média harmônica ponderada de um conjunto de números, dispostos em uma tabela de frequências, é dada pela seguinte expressão:

$$\bar{x}_h = \frac{n}{\sum_{j=1}^{k} f_j/x_j} = \frac{n}{f_1/x_1 + f_2/x_2 + \ldots + f_n/x_n}$$

Exemplo 11: Suponha que em uma sala de aula de 5 alunos, as notas em Estatística foram as seguintes: 9,4; 8,1; 6,5; 4,3 e 3,7. A média geométrica da turma na disciplina será calculada da seguinte forma:

$$\bar{x}_h = \frac{5}{1/9{,}4 + 1/8{,}1 + 1/6{,}5 + 1/4{,}3 + 1/3{,}7} = 5{,}6$$

Exemplo no Excel: No Excel, podemos calcular a média harmônica simples utilizando a função "MÉDIA.HARMÔNICA". Conforme demonstrado abaixo, podemos verificar como calcular a média do exemplo anterior utilizando o Excel:



No Excel, a média é calculada com os valores compreendidos entre as células B2 e B6, com a notação "MÉDIA.HARMÔNICA(B2:B6), o Excel entende que você deseja calcular a média dos valores contidos das células B2, B3, B4, B5 e B6. Então você pode entender que B2:B6 quer dizer "B2 até B6".

Exemplo 12: Calcular a média geométrica com os dados do "Exemplo 4".

$$\bar{x}_h = \frac{20}{1/4 + 5/5 + 6/6 + 5/7 + 3/8} = 5{,}9$$

c) Propriedades da Média Harmônica

Genericamente, podemos enunciar: A média harmônica de um conjunto de números positivos é menor ou igual à média geométrica, e esta, por sua vez, é menor ou igual à média aritmética.

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x}$$

d) Aplicação da Média Harmônica

A média harmônica é particularmente recomendada para série de valores que são inversamente proporcionais, como para o cálculo de velocidade média, tempo médio de escoamento de estoques, custo médio de bens comprados com uma quantia fixa etc.

4.2 Moda (Mo)

Pode-se definir moda como o valor mais frequente, quando comparada sua frequência com a dos valores de um conjunto ordenados.

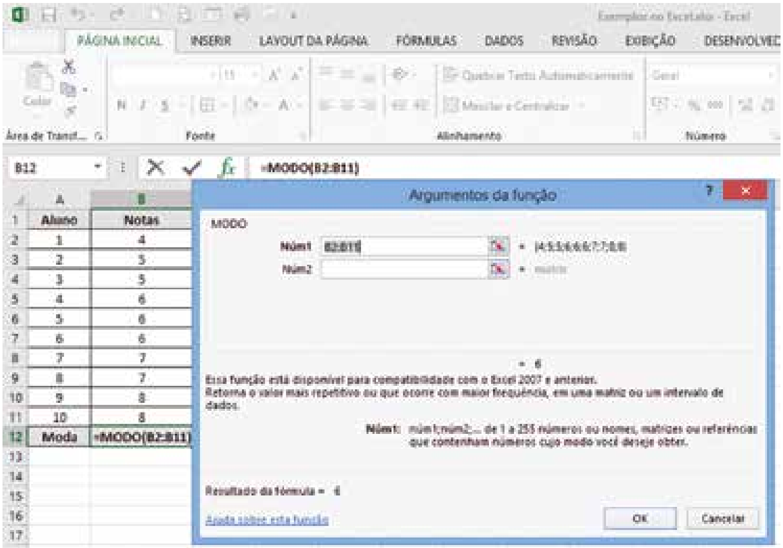
4.2.1 Determinação da Moda de valores Não-Tabulados

Considerando um conjunto ordenado de valores, a moda será o valor predominante, o valor mais frequente desse conjunto. Evidentemente, um conjunto de valores pode não apresentar moda, sendo denominado amodal.

Exemplo 13: Calcular a moda dos seguintes conjunto de valores:

- $X = \{4, 5, 5, 6, 6, 6, 7, 7, 8, 8\}$
- $Y = \{4, 4, 5, 5, 6, 6\}$
- $Z = \{1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6, 6, 7\}$
- $W = \{1, 2, 3, 4, 5, 6\}$
- Moda de X =: Mo = 6; o valor 6 é o mais frequente;
- Moda de Y = : conjunto amodal, pois seus três valores apresentam duas vezes;
- Moda de Z = : $Mo_1 = 2$ e $Mo_2 = 5$; é um conjunto bimodal;
- Moda de W = : Conjunto amodal.

Exemplo no Excel: No Excel, podemos calcular a moda utilizando a função “MODO”. Conforme demonstrado abaixo, podemos verificar como calcular a MODA do conjunto “X” acima:



No Excel, a moda é calculada com os valores compreendidos entre as células B2 e B11, com a notação "MODA(B2:B11), o Excel entende que você deseja calcular a MODA dos valores contidos das células B2, B3, B4, B5, B6, B7, B8, B9, B10 e B11. Então você pode entender que B2:B11 quer dizer "B2 até B11".

4.2.2 Determinação da Moda de valores Tabulados

O método de King, para o cálculo da moda, baseia-se na influência das classes adjacentes sobre a classe modal, como mostrado na equação abaixo:

$$Mo = LI + Ac \frac{f_{post}}{f_{ant} + f_{post}}$$

onde:

- LI – limite inferior da classe modal;
- Ac – amplitude da classe modal;

f_{ant} – frequência simples da classe anterior à classe modal;

f_{post} – frequência simples da classe posterior à classe modal.

Exemplo 14: Calcular, pelo método de King, a moda dos valores constantes da tabela abaixo:

Classes	f_i	
10⊢ 20	2	
20⊢30	4	f_{ant}
30⊢40	8	Classe Modal
40⊢50	5	f_{post}
50⊢ 60	1	
	N = 20	

Classe modal é a terceira: 30⊢40. A moda, segundo a fórmula de King, será:

$$Mo = LI + Ac\frac{f_{post}}{f_{ant} + f_{post}} = 30 + 10\frac{5}{4+5} = 35,556$$

4.3 Mediana (Md)

A mediana é a terceira medida de tendência central e pode ser definida como o valor que divide uma série ordenada de tal forma que pelo menos a metade (ou 50%) dos itens sejam iguais ou maiores do que ela, e que haja pelo menos outra metade (ou 50%) de itens menores do que ela. A medida é considerada uma separatriz, por dividir uma distribuição ou um conjunto de dados em partes iguais.

4.3.1 Determinação da Mediana de Valores Não-Tabulados

a. O Número de Observações é Ímpar

Neste caso, deve-se encontrar o valor do elemento mediano (E_{Md}), o que é feito da seguinte forma:

$$E_{Md} = \frac{n+1}{2}$$

O passo seguinte será localizar a mediana na lista de valores, de acordo com o resultado obtido na cálculo do elemento mediano.

Exemplo 15: Calcular a mediana do seguinte conjunto de números:

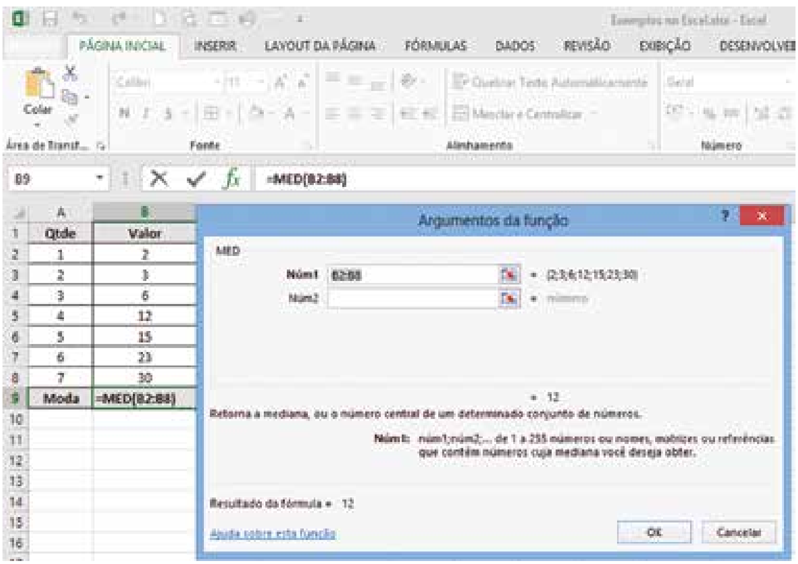
$X = \{2, 3, 6, 12, 15, 23, 30\}$

$$E_{Md} = \frac{n+1}{2} = \frac{7+1}{2} = 4$$

$E_{Md} = 4$ indica que a mediana é o valor que se encontra na quarta posição da lista ordenada de valores, é o quarto número da série, então:

Md = 12

Exemplo no Excel: No Excel, podemos calcular a mediana utilizando a função "MED". Conforme demonstrado abaixo, podemos verificar como calcular a mediana do exemplo anterior utilizando o Excel:



No Excel, a mediana é calculada com os valores compreendidos entre as células B2 e B8, com a notação "MED(B2:B8), o Excel entende que você deseja calcular a média dos valores contidos das células B2, B3, B4, B5, B6, B7 e B8. Então você pode entender que B2:B8 quer dizer "B2 até B8".

b. O Número de Observações é Par

Agora, o elemento mediano (E_{Md}) será determinado através da seguinte expressão:

$$E_{Md} = \frac{n}{2}$$

Exemplo 16: Calcular a mediana do seguinte conjunto de números:

$X = \{3, 6, 9, 12, 14, 15, 17, 20\}$

Como vemos, $n = 8$; então o elemento mediano será:

$$E_{Md} = \frac{n}{2} = \frac{8}{2} = 4$$

Seguindo a mesma linha de raciocínio do exemplo anterior, identificaríamos a mediana como o quarto elemento da lista, ou seja: $Md = 12$. O problema, é que se verificarmos na série acima, notamos que 12 não representa exatamente o meio da série, contendo três valores menores do que doze (3, 6 e 9) e quatro maiores (14, 15, 17 e 20).

Toda vez que houver um número par de observações, a lista apresentará dois valores centrais e a mediana será determinada calculando a média aritmética deles. Então, teríamos:

$$Md = \frac{12 + 14}{2} = 13$$

4.3.2 Determinação da Mediana de Valores Tabulados – não agrupados em classes

Quando os valores da variável estiverem já tabulados, o procedimento a ser adotado será praticamente idêntico ao anterior.

Em primeiro lugar, deve-se verificar se o número de observações é ímpar ou par e, conforme o caso, aplicar a fórmula adequada para o cálculo do elemento mediano (E_{Md}).

Em seguida, acrescentamos uma coluna à tabela de frequência original, a coluna referente à frequência acumulada. Comparando o resultado obtido no cálculo do elemento mediano com os valores constantes dessa coluna, determinaremos a mediana.

Exemplo 17: Calcular a mediana dos valores apresentados nas tabelas abaixo:

Caso 1		
Valores (x_j)	Frequência (f_j)	Frequência Acumulada (F_i)
2	5	5
3	10	15
4	15	30
5	12	42
6	5	47
7	3	50
	n=50	

Caso 2		
Valores (xj)	Frequência (fj)	Frequência Acumulada (Fj)
3	3	3
4	6	9
5	9	18
6	8	26
7	6	32
8	3	35
	n=35	

Solução:

- **Caso 1:**

  N = 50, então $E_{Md} = \frac{n}{2} = \frac{50}{2} = 25$. A mediana deverá ser o 25º elemento.

  O passo seguinte será verificar a mediana na coluna da frequência acumulada. Neste caso, o 25º elemento se encontra na terceira linha, pois a terceira linha compreende do 16º ao 30º elemento.

  Então, *Md* = 4.

- **Caso 2:**

  N = 35, então $E_{Md} = \frac{n+1}{2} = \frac{35+1}{2} = 18$

A mediana deverá ser o décimo oitavo elemento. O passo seguinte será verificar a mediana na coluna da frequência acumulada.

Então, $Md = 5$

4.3.3 Determinação da Mediana de Valores Tabulados – agrupados em classes

Para determinar a mediana de valores tabulados em classes, utiliza-se, geralmente, a seguinte expressão: $Md = LI + Ac\dfrac{E_{Md} - F_{ant}}{f_{Md}}$

onde:

- LI – Limite inferior da classe;
- Ac – Amplitude do intervalo de classe;
- E_{Md} – elemento mediano;
- F_{ant} – frequência acumulada até a classe anterior à classe mediana;
- f_{Md} – frequência absoluta da classe mediana.

Exemplo 18: Calcular a mediana dos valores apresentados na tabela abaixo:

Classes	fj	Fj
10 ⊢ 20	10	
20⊢ 30	20	30
30⊢ 40	35	65
40⊢ 50	40	105
50⊢ 60	25	130
60⊢70	15	145
70⊢ 80	5	150
	N = 150	

$$E_{Md} = \frac{150}{2} = 75$$, então:

A mediana deverá ser o 75º elemento. O passo seguinte será verificar a mediana na coluna da frequência acumulada. Neste caso, o 75º elemento se encontra na Quarta classe.

$$Md = LI + Ac\frac{E_{Md} - F_{ant}}{f_{Md}} = 40 + 10\frac{75-65}{40} = 42,5 \;\Rightarrow\; Md = 42,5$$

4.4 Considerações Finais Sobre a Média Aritmética, a Moda e a Mediana

Qual dessas medidas de tendência central é a melhor? Infelizmente, não há uma resposta única, porque não há critérios objetivos para determinar a medida mais representativa para todos os conjuntos de dados, podemos dizer que elas possuem diferentes vantagens e desvantagens. Uma vantagem importante da média é que ela leva em conta todos os valores da amostra ou população, mas uma desvantagem é que às vezes pode ser seriamente afetada por alguns valores extremos. Dentre as várias medidas de tendência central, seguramente a média aritmética é a mais utilizada.

Resumidamente, podemos dizer que a média aritmética é a soma dos valores observados dividida pelo número total deles; a mediana é o valor que divide a série em duas partes iguais quanto ao número de valores de cada parte e a moda é o valor em cuja vizinhança tendem a se concentrar os valores da série.

a) **Conceitos Físicos**

A média aritmética corresponde ao centro de gravidade estudado em física. Esse fato nos revela que essa medida é afetada de maneira acentuada pelos valores extremos da série, o que não acontece com a moda e a mediana. Assim, por exemplo:

Exemplo 19: Duas turmas fizeram prova de Estatística, verifique as notas e compare-as.

Turma 1		
Notas (x_j)	Nº de Alunos Frequência (f_j)	Frequência Acumulada (F_j)
0	0	0
1	0	0
2	0	0
3	2	2
4	3	5
5	4	9
6	8	17
7	10	27
8	3	30
9	0	30
10	0	30
	$n = 30$	

Turma 2		
Notas (x_j)	Nº de Alunos Frequência (f_j)	Frequência Acumulada (F_j)
0	4	4
1	2	6
2	4	10
3	0	10
4	3	13
5	2	15
6	1	16
7	2	18
8	3	21
9	6	27
10	8	35
	$n = 35$	

Se verificarmos, iremos achar que a média aritmética da Turma 1 é igual ao da Turma 2, isto é, Média = 6,0. Então, aparentemente, pela análise somente da média aritmética, podemos concluir que as duas turmas tiveram um rendimento igual, o que não é verdade.

Turma 1	Turma 2
Média = 6,0	Média = 6,0
Moda = 7,0	Moda = 10,0
Mediana = 6,0	Mediana = 7,0

Podemos concluir que a Turma 1 é muito mais homogênea do que a Turma 2, pois na Turma 1 as medidas estão próximas, o que não ocorre na turma 2.

A média aritmética é preferível às demais medidas, para estimar a tendência central, quando se trata de muitas classes de populações, por haver menos variabilidade entre as médias aritméticas calculadas a partir de várias amostras aleatórias do que entre as medianas e as modas.

A mediana é preferível à medida que se está interessado em conhecer exatamente o ponto médio da distribuição, aquele valor que a divide em duas partes exatamente iguais. É preferível, ainda, quando os resultados extremos são tais que podem afetar sensivelmente o valor da média.

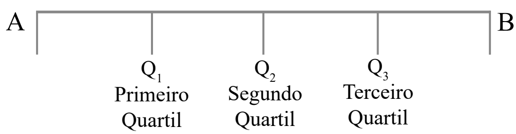
A moda é utilizada essencialmente quando pretendemos apenas uma medida rápida e aproximada da tendência central.

4.5 Quartis, Decis e Percentis (ou Centis)

Há uma série de medidas de posição semelhantes na sua concepção (a mediana), embora não sejam medidas de tendência central. Como se sabe, a mediana divide a distribuição em duas partes iguais quanto ao número de elementos de cada parte. Já os quartis permitem dividir a distribuição em quatro partes iguais, os decis em dez partes e os centis em cem partes iguais.

4.6 Quartis (Qi)

Para dividir uma série ordenada de valores em quatro partes iguais, precisamos de três separatrizes – quartis.



Genericamente, para determinar a ordem ou posição do quartil a ser calculado, usaremos a seguinte expressão:

$$E_{Qi} = \frac{i.n}{4}$$ onde:

- E_{Qi} – Elemento que indica a ordem do quartil;
- i – Número do quartil a ser calculado;
- n – Número de observações.

Então, para dados tabulados em classes, calcula-se o valor do quartil através da expressão: $Q_i = LI + Ac\dfrac{E_{Qi} - F_{ant}}{f_{Qi}}$

onde:

- LI – Limite inferior da classe;
- Ac – Amplitude do intervalo de classe;
- E_{Qi} – elemento que indica a ordem do quartil;
- F_{ant} – frequência acumulada até a classe anterior à classe do quartil;
- f_{Qi} – frequência absoluta da classe do quartil.

Exemplo 20: Determinar o primeiro e terceiro quartil da distribuição abaixo.

i	Estaturas	f_i	F_i
1	150 ⊢160	4	4
2	160 ⊢ 170	10	14
3	170 ⊢ 180	20	34
4	180 ⊢190	6	40
5	190 ⊢200	6	46
6	200 ⊢ 210	2	48
Total	48		

O 1º quartil está no 12º elemento, conforme calculado abaixo. Deve-se contar a frequência acumulada até encontrar o valor superior, que neste caso é 14. Isto quer dizer que o 1º quartil está na segunda classe.

O 3º quartil está no 36º elemento, conforme calculado abaixo. Deve-se contar a frequência acumulada até encontrar o valor superior, que neste caso é 40. Isto quer dizer que o 3º quartil está na quarta classe.

1º Quartil

$$E_{Q1} = \frac{i.n}{4} = \frac{1 \times 48}{4} = 12$$

$$Q1 = LI + Ac\frac{E_{Qi} - F_{ant}}{f_{Qi}} = 160 + 10 \times \frac{12 - 4}{10} = 168$$

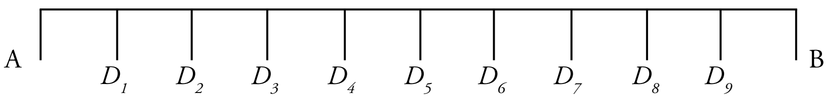
3º Quartil

$$E_{Q3} = \frac{i.n}{4} = \frac{3 \times 48}{4} = 36$$

$$Q3 = LI + Ac\frac{E_{Qi} - F_{ant}}{f_{Qi}} = 180 + 10 \times \frac{36-34}{6} = 183$$

4.7 Decis (Di)

Para dividir uma série ordenada de valores em dez partes iguais, precisamos de nove separatrizes – decis.



O primeiro decil de um conjunto ordenado de valores é o valor que divide um conjunto em duas partes de modo que dez por cento dos valores sejam menores e noventa por cento sejam maiores do que ele.

De uma forma geral, para calcular os decis, recorremos à seguinte expressão que define a ordem em que o decil se encontra:

$$E_{Di} = \frac{i.n}{10}$$

onde:

- E_{Di} – Elemento que indica a ordem do decil;
- i – Número do quartil a ser calculado;
- n – Número de observações.

Então, para dados tabulados em classes, calcula-se o valor do decil através da expressão:

$$D_i = LI + Ac\frac{E_{Di} - F_{ant}}{f_{Di}}$$

onde:

- LI – Limite inferior da classe;
- Ac – Amplitude do intervalo de classe;

- E_{Di} – elemento que indica a ordem do decil;
- F_{ant} – frequência acumulada até a classe anterior à classe do decil;
- F_{Di} – frequência absoluta da classe do decil.

4.8 Percentis o Centis (Ci)

Para dividir uma série ordenada de valores em cem partes iguais, precisamos de noventa e nove separatrizes – centis.

Neste caso, cada parte em que foram subdivididos os valores do conjunto, através dos noventa e nove centis, contará com um centésimo ou um por cento dos valores do conjunto.

O elemento que definirá a ordem do centil, em uma distribuição de frequências de valores tabulados agrupados em classes, será encontrado pelo emprego da expressão:

$$E_{Ci} = \frac{i.n}{100}$$

onde:

- E_{Ci} – Elemento que indica a ordem do centil;
- i – Número do quartil a ser calculado;
- n – Número de observações.

Então, para dados tabulados em classes, calcula-se o valor do decil através da expressão:

$$C_i = LI + Ac\frac{E_{Ci} - F_{ant}}{f_{Ci}}$$

onde:

- LI – Limite inferior da classe;
- Ac – Amplitude do intervalo de classe;
- E_{Ci} – elemento que indica a ordem do centil;
- F_{ant} – frequência acumulada até a classe anterior à classe do centil;

- F_{Ci} – frequência absoluta da classe do centil.

Atividades

1. Os graus de um estudante em seis exames foram: 84, 91, 72, 68, 87 e 78. Determine a média aritmética, média geométrica e a média harmônica dos graus.

2. Dez medidas do diâmetro de um cilindro foram anotadas por um cientista como 3,88; 4,09; 3,92; 3,97; 4,02; 3,95; 4,03; 3,92; 3,98 e 4,06 centímetros. Determinar a média aritmética das medidas.

3. Entre 100 números, vinte são 4, quarenta são 5, trinta são 6 e o restante são 7. Determine a média aritmética dos números.

4. Os graus finais de um estudante em matemática, física, química e biologia são, respectivamente, 82; 86; 90 e 70. Se os pesos atribuídos a essas matérias são, respectivamente, 3; 5; 3 e 1, determine a média aritmética.

5. Quatro grupos de estudantes, constituídos de 15, 20, 10 e 18 indivíduos tem pesos médio de 81, 74, 77 e 70 kg, respectivamente. Determinar o peso médio de todos os estudantes.

6. A partir da distribuição de frequências das alturas apresentada na tabela abaixo para determinar:
   a. Média;
   b. Mediana;
   c. Moda.

Salário	Número de estudantes
151 – 158	5
159 – 166	18
167 – 174	42
175 – 182	27

Salário	Número de estudantes
183 – 190	8
Total	100

7. A tabela abaixo, apresenta uma distribuição de frequência dos graus de um exame final de álgebra.
   a. Calcular a média, mediana e a moda;
   b. Determinar os quartis da distribuição;
   c. Interpretar claramente o significado de cada um.

Grau	Número de estudantes
30 —— 40	9
40 —— 50	32
50 —— 60	43
60 —— 70	21
70 —— 80	11
80 —— 90	3
90 —— 100	1
Total	**120**

8. Organize o Rol e calcule a Média, Mediana e Moda dos seguintes dados:
   a. 3, 5, 2, 6, 5, 9, 5, 2, 8, 6;
   b. 51, 6, 48, 7, 3, 50, 49, 5, 22.

9. Certo pesquisador aplicou um teste aos alunos de um colégio e obteve os seguintes resultados:

   Acerto num teste de conhecimento.

1,1	1,1	1,1	1,1	1,2	1,4	1,4	1,4	1,6	1,6	2,0	2,0	2,0	2,1	2,2
2,4	2,5	2,7	2,9	3,0	3,0	3,0	3,1	3,2	3,3	3,3	4,0	4,1	4,1	4,2
4,2	4,3	4,4	4,6	4,9	5,0	5,0	5,0	6,0	6,0	6,5	6,5	7,0	7,0	7,2
7,2	7,6	7,8	8,0	8,2	8,2	8,4	8,4	8,6	9,0	9,2	9,4	9,6	9,8	9,8

Pede-se para calcular:

a. média aritmética;

b. moda;

c. mediana;

d. 1º Quartil;

e. 2º Quartil;

f. 3º Quartil.

Respostas

1. MA = 80,0; MG = 79,6 ; MH = 79,2
2. MA = 3,98 centímetros; MG = 3,98 centímetros; MH =3,98 centímetros
3. MA = 5,3; MG = 5,2; MH = 5,1
4. MA = 84,7; MG = 84,5; MH = 84,3
5. MA = 75,0 kg; MG = 74,9; MH = 74,8
6. 171,70 cm
7. A) média = 54,4; mediana = 53,95 = 54,0; moda = 53,96 = 54,0; b) 1º quartil = 46,25 = 46,2; 2º quartil = 53,95 = 54,0; 3º quartil = 61,43 = 61,4.
8. a) média = 5,1; mediana = 5; moda = 5; b) média = 27,8; mediana = 22; moda = amodal;
9. a) 4,75 = 4,8; b) 1,1; c) 4,2; d) 2,2; e) 4,2; f) 7,2

5

Medidas de Dispersão e Assimetria

Como se sabe, os fenômenos em cuja análise intervém o método estatístico, bem como os dados estatísticos a ele referentes, caracterizam-se tanto pela sua semelhança quanto pela sua variabilidade. Não há razão alguma para se calcular a média de um conjunto de dados onde não haja variação desses elementos. Ocorre, que se a variabilidade dos dados for muito grande, sua média terá um grau de confiabilidade tão pequeno que será inútil calculá-la.

É importante ressaltar que a análise completa dos dados requer não apenas sua apresentação, através de gráficos e tabelas, ou o cálculo de medidas de posição. Caracterizar um conjunto de valores apenas através de uma média, por exemplo, é decrevê-lo inadequadamente, uma vez que os dados diferem entre si, em maior ou menor grau.

Portanto, para avaliar o grau de variabilidade ou dispersão dos valores de um conjunto de números, lançaremos mão das medidas de dispersão. Essas proporcionarão um conhecimento mais completo do fenômeno a ser analisado, permitindo estabelecer comparações entre fenômenos de mesma natureza e mostrando até que ponto os valores se distribuem acima ou abaixo da tendência central.

Medidas de dispersão mais usadas:

- Amplitude total;
- Amplitude quartílica;
- Desvio médio;
- Desvio padrão e variância.

5.1 Amplitude Total ou Intervalo Total (At)

A amplitude total de um conjunto de números é a diferença entre os valores extremos do conjunto.

At = limite superior do fenômeno – limite inferior do fenômeno.

Exemplo 1: Calcular a amplitude total do conjunto de números abaixo:

$$A = \{10, 12, 15, 17, 21, 25, 28\} \quad \Rightarrow \quad At = 28 - 10 = 18$$
$$B = \{1,71; 1,75; 1,78; 1,80; 1,85\} \quad \Rightarrow \quad At = 1,85 - 1,71 = 0,14$$

Exemplo 2: Calcular a amplitude total dos valores dispostos na tabela abaixo:

Classes	f_i
10 – 20	2
20 – 30	4
30 – 40	8
40 – 50	14
50 – 60	12
60 – 70	6

$At = 70 - 10 = 60$

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 ele leva em conta apenas os valores extremos da série.

A amplitude total é sensível ao tamanho da amostra, variando com possíveis valores extremos anormais, além da insensibilidade entre valores extremos.

A vantagem da amplitude é seu cálculo rápido e fácil, é também sua mais importante desvantagem. Isto é, a amplitude depende totalmente de apenas dois valores, o maior e o menor valor de um determinado conjunto de dados. Como resultado, a amplitude normalmente proporciona meramente um índice aproximado da variabilidade de uma distribuição.

5.2 Desvio Quartil ou Amplitude Semi-interquartílica (Dq)

O desvio quartil é uma medida de dispersão baseada no quartil e calculada como a média aritmética das diferenças entre a mediana e os dois quartis.

$$Dq = \frac{Q_3 - Q_1}{2}, \ \ onde: \ Q_3 = terceiro \ \ quartil \ e \ Q_1 = primeiro \ \ quartil.$$

O desvio quartil deverá ser usado preferencialmente quando a medida de tendência central for a mediana.

5.3 Desvio Médio (Dm)

O desvio médio ou média dos desvios é igual à média aritmética dos valores absolutos dos desvios tomados em relação a média aritmética do fenômeno estudado.

a) Desvio Médio para Dados Brutos

Quando os dados não vierem dispostos em uma tabela de frequências, o desvio médio será calculado através da seguinte fórmula:

$$Dm = \frac{\sum_{i=1}^{n} \left| d_i \right|}{n} = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right|}{n}, \quad onde:$$

$$d_i = \left(x_i - \overline{x} \right) = desvio\ em\ relação\ à\ média\ aritmética$$
$$\overline{x} = média\ aritmética$$

Exemplo 3: Como exemplo de aplicação, calculemos o desvio médio do seguinte conjunto de números: 4, 6, 8, 9, 10 e 11. Inicialmente devemos calcular a média aritmética:

$$\overline{x} = \frac{4+6+8+9+10+11}{6} = 8$$

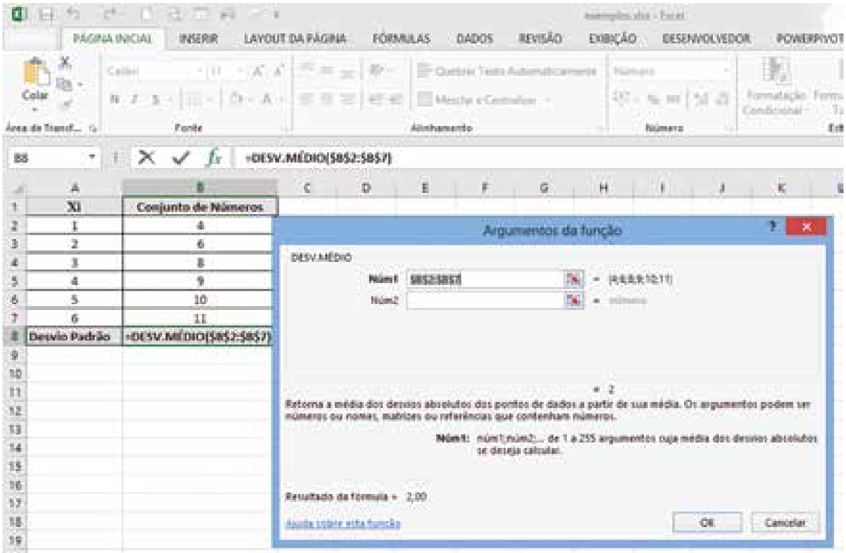
Temos então,

i	$\mathbf{x_i}$		
1	$x_1 = 4$	4 – 8 = -4	4
2	$x_2 = 6$	6 – 8 = -2	2
3	$x_3 = 8$	8 – 8 = 0	0
4	$x_4 = 9$	9 – 8 = 1	1
5	$x_5 = 10$	10 – 8 = 2	2
6	$x_6 = 11$	11 – 8 = 3	3
	$\sum$	0	12

Então:

$$Dm = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right|}{n} = \frac{12}{6} = 2$$

Exemplo no Excel: No Excel, podemos calcular desvio médio de conjunto de números utilizando a função "DESV.MÉDIO". Conforme demonstrado abaixo, podemos verificar como calcular o desvio médio do exemplo anterior utilizando o Excel.



No Excel, a desvio médio é calculado com os valores compreendidos entre as células B2 e B7, com a notação "DESV.MÉDIO(B2:B7), o Excel entende que você deseja calcular o desvio médio dos valores contidos das células B2, B3, B4, B5, B6 e B7. Então você pode entender que B2:B7 quer dizer "B2 até B7".

b) Desvio Médio para Dados Tabulados

Se os valores vierem dispostos em uma tabela de frequência, agrupados ou não em classes, será usada a seguinte fórmula:

$$Dm=\frac{\sum_{j=1}^{k}\left| d_j \right| \times f_j}{n} = \frac{\sum_{j=1}^{k}\left| x_j - \overline{x} \right| \times f_j}{n}, \text{ onde:}$$

x_j = valor individual ou um ponto médio da classe

Exemplo 4: calcule o desvio médio dos dados apresentados na tabela a seguir.

j	Idades	Frequência (f_j)
1	18 ⊢ 21	9
2	21 ⊢24	12
3	24 ⊢ 27	12
4	27 ⊢ 30	17
5	30 ⊢ 33	16
6	33 ⊢ 36	14
7	36 ⊢39	11
8	39 ⊢ 42	9
	$\sum$	**100**

Primeiramente temos que calcular a média;

j	Idades	Frequência (f_j)	Ponto Médio (x_i)	$x_j.f_j$
1	18 ⊢ 21	9	19,5	19,5 × 9 = 176
2	21 ⊢24	12	22,5	22,5 × 12 = 270
3	24 ⊢ 27	12	25,5	25,5 × 12 = 306
4	27 ⊢ 30	17	28,5	28,5 × 17 = 485
5	30 ⊢ 33	16	31,5	31,5 × 16 = 504
6	33 ⊢ 36	14	34,5	34,5 × 14 = 483
7	36 ⊢39	11	37,5	37,5 × 11 = 413
8	39 ⊢ 42	9	40,5	40,5 × 9 = 365
	$\sum$	**100**		**3.000**

Então a média é: $\overline{x} = \dfrac{3.000}{100} = 30$

Agora podemos calcular o desvio médio conforme apresentado a seguir.

j	Idades	Frequência (f_j)	Ponto Médio (x_j)	$x_j - \overline{x}$	$\left\|x_j - \overline{x}\right\|.f_j$
1	18 ⊢ 21	9	19,5	19,5 – 30 = **-10,5**	10,5 × 9 = 94,5
2	21 ⊢24	12	22,5	22,5 – 30 = **-7,5**	7,5 × 12 = 90
3	24 ⊢ 27	12	25,5	25,5 – 30 = **-4,5**	4,5 × 12 = 54
4	27 ⊢ 30	17	28,5	28,5 – 30 = **-1,5**	1,5 × 17 = 25,5
5	30 ⊢ 33	16	31,5	31,5 – 30 = **1,5**	1,5 × 16 = 24
6	33 ⊢ 36	14	34,5	34,5 – 30 = **4,5**	4,5 × 14 = 63
7	36 ⊢39	11	37,5	37,5 – 30 = **7,5**	7,5 × 11 = 82,5
8	39 ⊢ 42	9	40,5	40,5 – 30 = **10,5**	10,5 × 9 = 94,5
	$\sum$	**100**			**528**

O desvio médio é:

$$Dm = \frac{\sum_{j=1}^{k}\left|d_j\right| \times f_j}{n} = \frac{\sum_{j=1}^{k}\left|x_j - \overline{x}\right| \times f_j}{n} = \frac{528}{100} = 5,28$$

5.4 Desvio Padrão (S ou σ)

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 em relação a 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 ou a média quadrática dos desvios, ou afastamento em relação à média aritmética desse conjunto, serão definidas fórmulas que veremos a seguir.

5.4.1 Desvio Padrão para Dados Brutos

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n}}$$

Quando o desvio padrão representar uma descrição da amostra e não da população, caso mais comum, o denominador 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. Então, quando estivermos trabalhando com uma amostra, devemos utilizar a equação abaixo:

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

Exemplo 5: Como exemplo de aplicação, calculemos o desvio padrão do seguinte conjunto de números: 4, 6, 8, 9, 10 e 11.

Inicialmente devemos calcular a média aritmética:

$$\bar{x} = \frac{4+6+8+9+10+11}{6} = 8$$

Temos então,

i	$\mathbf{x_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{\left(x_i - \bar{x}\right)^2}$
1	$x_1 = 4$	4 – 8 = -4	16
2	$x_2 = 6$	6 – 8 = -2	4
3	$x_3 = 8$	8 – 8 = 0	0
4	$x_4 = 9$	9 – 8 = 1	1
5	$x_5 = 10$	10 – 8 = 2	4
6	$x_6 = 11$	11 – 8 = 3	9
	$\sum$	**0**	**34**

- **População:** se o conjunto de números apresentados fosse população, o cálculo do desvio padrão deve ser feito da seguinte forma:

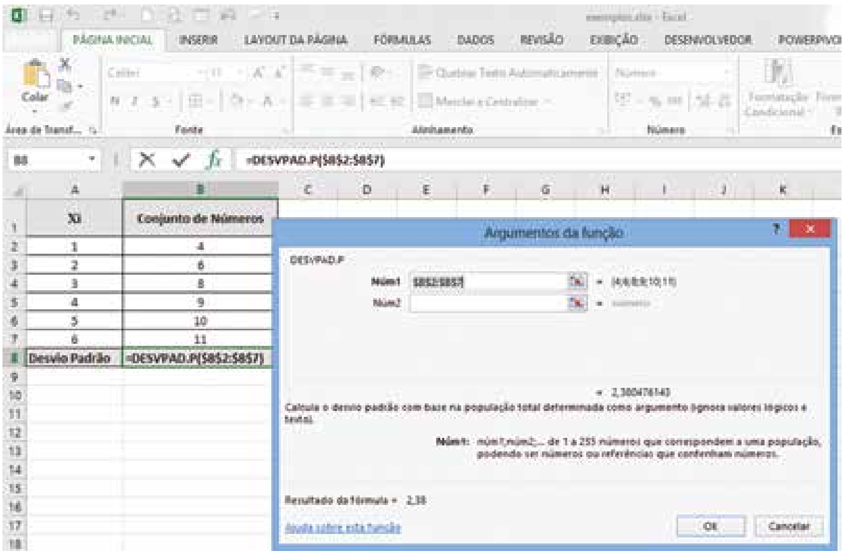
$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}} = \sqrt{\frac{34}{6}} = 2{,}38$$

- **Amostra:** se o conjunto de números apresentados fosse população, o cálculo do desvio padrão deve ser feito da seguinte forma:

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}} = \sqrt{\frac{34}{6-1}} = 2{,}61$$

Exemplo no Excel: resolver o exemplo 5 utilizando o Excel.

- **Para População:** No Excel, podemos calcular desvio padrão de uma população utilizando a função "DESVPAD.P". Conforme demonstrado abaixo, podemos verificar como calcular o desvio padrão de uma população do exemplo anterior utilizando o Excel:



No Excel, a desvio padrão é calculado com os valores compreendidos entre as células B2 e B7, com a notação "DESVPAD.P (B2:B7), o Excel entende que você deseja calcular o desvio padrão dos valores contidos das células B2, B3, B4, B5, B6 e B7. Então você pode entender que B2:B7 quer dizer "B2 até B7".

- **Para Amostra:** Se você estiver calculando o desvio padrão de uma amostra é só utilizar o procedimento apresentado acima utilizando a função "DESVPAD.A".

5.4.2 Desvio Padrão para Dados Tabulados

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

Para População:

$$S = \sqrt{\frac{\sum_{j=1}^{k} \left(x_i - \overline{x}\right)^2 \times f_j}{n}}$$

Para Amostra:

$$S = \sqrt{\frac{\sum_{j=1}^{k} \left(x_i - \overline{x}\right)^2 \times f_j}{n-1}}$$

Onde *k* representa o número de classes ou de linhas de uma tabela.

Exemplo 6: Calcule o desvio médio dos dados apresentados na tabela abaixo, sabendo-se que é uma amostra.

j	Idades	Frequência (f_j)
1	18 ⊢ 21	9
2	21 ⊢24	12
3	24 ⊢ 27	12
4	27 ⊢ 30	17
5	30 ⊢ 33	16
6	33 ⊢ 36	14
7	36 ⊢39	11
8	39 ⊢ 42	9
	$\sum$	**100**

Primeiramente temos que calcular a média.

j	Idades	Frequência (f_j)	Ponto Médio (x_j)	$x_j . f_j$
1	18 ⊢ 21	9	19,5	19,5 × 9 = 176
2	21 ⊢24	12	22,5	22,5 × 12 = 270
3	24 ⊢ 27	12	25,5	25,5 × 12 = 306
4	27 ⊢ 30	17	28,5	28,5 × 17 = 485
5	30 ⊢ 33	16	31,5	31,5 × 16 = 504
6	33 ⊢ 36	14	34,5	34,5 × 14 = 483
7	36 ⊢39	11	37,5	37,5 × 11 = 413
8	39 ⊢ 42	9	40,5	40,5 × 9 = 365
	$\sum$	**100**		**3.000**

Então a média é: $\overline{x} = \dfrac{3.000}{100} = 30$

Agora podemos calcular o desvio médio conforme apresentado abaixo:

j	Idades	Frequência (f_j)	Ponto Médio (x_j)	$x_j - \overline{x}$	$\left(x_j - \overline{x}\right)^2 . f_j$
1	18 ⊢ 21	9	19,5	19,5 – 30 = **-10,5**	$-10,5^2 \times 9 = 992,25$
2	21 ⊢24	12	22,5	22,5 – 30 = **-7,5**	$-7,5^2 \times 12 = 675$
3	24 ⊢ 27	12	25,5	25,5 – 30 = **-4,5**	$-4,5^2 \times 12 = 243$
4	27 ⊢ 30	17	28,5	28,5 – 30 = **-1,5**	$-1,5^2 \times 17 = 38,25$
5	30 ⊢ 33	16	31,5	31,5 – 30 = **1,5**	$1,5^2 \times 16 = 36$
6	33 ⊢ 36	14	34,5	34,5 – 30 = **4,5**	$4,5^2 \times 14 = 283,5$
7	36 ⊢39	11	37,5	37,5 – 30 = **7,5**	$7,5^2 \times 11 = 618,75$
8	39 ⊢ 42	9	40,5	40,5 – 30 = **10,5**	$10,5^2 \times 9 = 992,25$
$\sum$		**100**			**3.879**

O desvio padrão é:

$$S = \sqrt{\frac{\sum_{j=1}^{k}\left(x_i - \overline{x}\right)^2 \times f_j}{n-1}} = \sqrt{\frac{3.879}{100}} = 6,23$$

5.5 Variância (S2 ou σ2)

Conforme se pode perceber pelo símbolo, a variância é o quadrado do desvio padrão, ou se preferir, o desvio padrão é a raiz quadrada da variância. Dessa forma, pode-se dizer que a fórmula da variância é igual à expressão do desvio padrão, sem o sinal do radical.

5.6 Significado do desvio padrão e da variância

O desvio padrão é mais fácil de interpretar do que a variância, porque ele está na unidade correta de medida. Mesmo assim, a série de passos necessários para calcular o desvio padrão pode causar uma sensação de desconforto em relação ao significado de seu resultado. Por exemplo, suponha que em uma certa distribuição de escores tenhamos $S = 4$. O que esse número indica? O que exatamente podemos dizer agora a respeito daquela distribuição que não poderíamos ter dito antes?

O capítulo que fala da "Distribuição Normal" buscará esclarecer o significado completo de desvio padrão. Por ora, observamos brevemente que o desvio padrão representa a variabilidade do "termo médio" em uma distribuição, pois ele mede a média dos desvios a contar da média. Os procedimentos de elevar ao quadrado e cálculo da raiz quadrada também entram no contexto, mas, sobretudo, para eliminar sinais de menos e voltar à unidade mais conveniente de medida, a unidade de escore bruto.

Observamos também que, quanto maior a variabilidade em torno da média de uma distribuição, maior o desvio padrão. Desse modo, $S = 4{,}5$ indica maior variabilidade do que $S = 2{,}5$. Por exemplo, a distribuição das temperaturas diárias em Curitiba, tem um desvio padrão maior do que a distribuição de temperaturas para o mesmo período em Brasília.

Analisemos a tabela abaixo, que apresenta as sentenças dadas por duas juízas a dois conjuntos de seis réus em julgamentos por roubo:

Alunos	Juíza A	Juíza B
Alunos 1	34 meses	26 meses
Alunos 2	30 meses	43 meses
Alunos 3	31 meses	22 meses

Alunos	Juíza A	Juíza B
Alunos 4	33 meses	35 meses
Alunos 5	36 meses	20 meses
Alunos 6	34 meses	34 meses
Média	**33,0**	**30,0**
Variância	**4,0**	**65,0**
Desvio Padrão	**2,0**	**8,1**

Observe primeiro a vantagem do desvio padrão sobre a variância. Mesmo que eles sejam iguais em suas aptidões para medir a dispersão, o desvio padrão tem uma interpretação mais tangível. Nesse caso, ele é expresso em termos de meses, algo que faz sentido para nós. A variância, entretanto, é declarada em termos de meses ao quadrado, o que a torna mais difícil de ser compreendida.

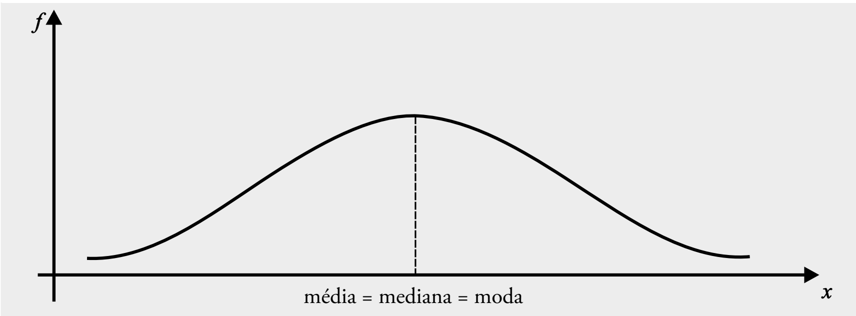
Voltando a comparação, podemos dizer que a Juíza-A tem uma média maior, mas um desvio padrão menor do que a Juíza-B. Alguém poderia dizer, pelo menos baseado somente nesses dados, que a Juíza-A é mais rigorosa, apesar de mais justa, e a Juíza-B é mais clemente, apesar de mais inconsistente. Para um advogado, a sua melhor aposta poderia ser a Juíza-A. embora possa esperar uma sentença mais longa (devido a média mais alta), talvez você não queira arriscar receber as sentenças severas pelas quais a Juíza-B é conhecida, para alguns casos específicos.

Desse modo, o desvio padrão é uma ferramenta útil para medir o grau de variabilidade em uma distribuição ou para comparar a variabilidade em diferentes distribuições. Ele também é empregado, e com bastante frequência, para ajustar a posição relativa de escores individuais dentro de uma distribuição. Nesse sentido, ele é um padrão contra o qual avaliamos a colocação de um escore (somo sua nota em uma prova) dentro de toda a distribuição (como as notas das provas da turma inteira).

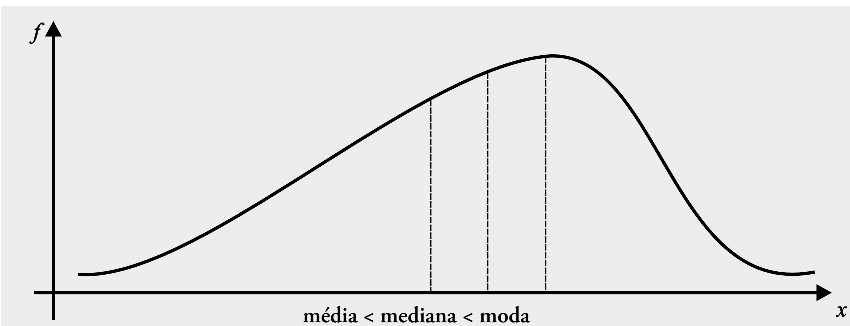
5.7 Assimetria

Assimetria é o grau de desvio, ou afastamento da simetria, de uma distribuição. Quando a distribuição for **simétrica** (a), a média e a moda coincidem; quando ela for **assimétrica à esquerda ou negativa** (b) a média é menor que a moda e quando for **assimétrica à direita ou positiva** (c) a média é maior que a moda como mostram as figuras a seguir:

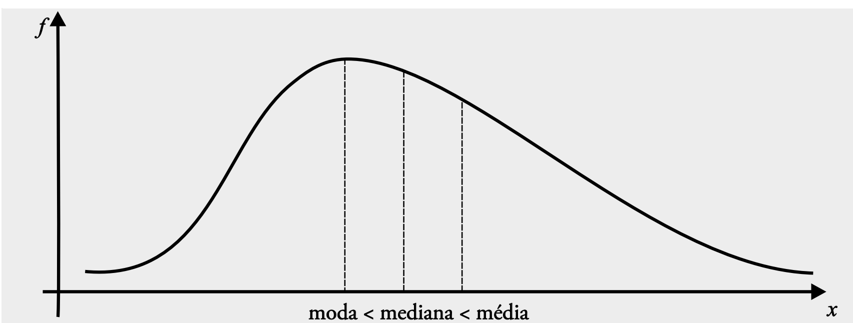
(a)



(b)



(c)



- No caso (a) a distribuição é classificada de simétrica.
- No caso (b) a distribuição é classificada de assimétrica positiva.
- No caso (c) a distribuição é classificada de assimétrica negativa.

Para termos um parâmetro quantitativo a respeito da assimetria, basta fazermos a diferença entre a média e a moda da distribuição.

Média – Moda

Com o resultado dessa diferença, temos a seguinte classificação:

- Média – Moda = 0 ⇒ Distribuição simétrica;
- Média – Moda < 0 ⇒ Distribuição assimétrica negativa;
- Média – Moda > 0 ⇒ Distribuição assimétrica positiva.

Exemplo 7: analise as distribuições abaixo sob a ótica da assimetria.

Distribuição 1		
i	Estaturas (cm)	*fi*
1	150 ⊢ 160	4
2	160 ⊢ 170	10
3	170 ⊢ 180	20
4	180 ⊢ 190	10
5	190 ⊢ 200	4
	Total	48
	Média	175
	Mediana	175
	Moda	175
	Desv. Padrão	10,4

Distribuição 2		
i	Estaturas (cm)	*fi*
1	150 ⊢ 160	4
2	160 ⊢ 170	10
3	170 ⊢ 180	20
4	180 ⊢ 190	40
5	190 ⊢ 200	4
	Total	78
	Média	178,8
	Mediana	181,2
	Moda	185
	Desv. Padrão	9,5

Distribuição 3		
i	Estaturas (cm)	*fi*
1	150 ⊢ 160	4
2	160 ⊢ 170	40
3	170 ⊢ 180	20
4	180 ⊢ 190	10
5	190 ⊢ 200	4
	Total	78
	Média	171,2
	Mediana	168,8
	Moda	165
	Desv. Padrão	9,5

Logo:

- Média – Moda = 175 – 175 = 0 (distribuição simétrica);
- Média – Moda = 178,8 – 185 = -6,15 < 0 (assimetria negativa);
- Média – Moda = 171,2 – 165 = +6,15 > 0 (assimetria positiva).

5.7.1 Coeficiente de assimetria

Para melhor compararmos duas ou mais medidas de dispersão, usaremos o Coeficiente de Assimetria (ou Coeficiente de Pearson) definido por:

$$A_V = \frac{3(\overline{x} - Md)}{S}$$

Para efeito de comparação, usaremos o exemplo anterior das distribuições 1, 2 e 3.

- **Distribuição 1**

$$A_{V1} = \frac{3 \times (175 - 175)}{10,4} = 0 \text{ (distribuição simétrica)}$$

- **Distribuição 2**

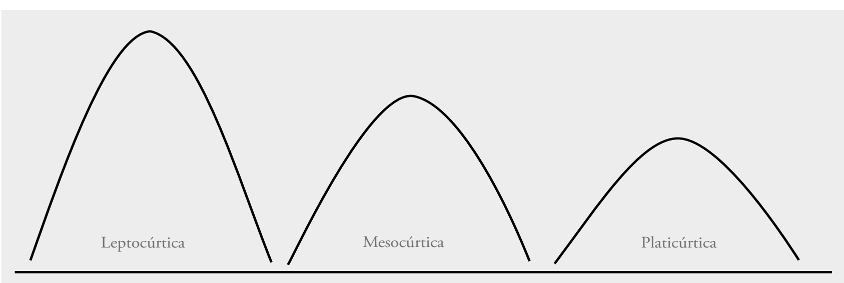
$$A_{V2} = \frac{3 \times (178,8 - 181,2)}{9,5} = -0,75 \text{ (distribuição assimétrica negativa)}$$

- **Distribuição 3**

$$A_{V3} = \frac{3 \times (171,2 - 168,8)}{9,5} = +0,75 \text{ (distribuição assimétrica positiva)}$$

5.8 Curtose

É o grau de achatamento de uma distribuição em relação a uma distribuição padrão.



5.8.1 Coeficiente de Curtose

Também conhecido como coeficiente percentílico de curtose, que é encontrado por meio da equação:

$$C = \frac{Q_3 - Q_1}{2 \times \left(P_{90} - P_{10}\right)}$$

Onde:

- Q_1 – 1º quartil;
- Q_3 – 3º quartil;
- P_{10} – 10º Percentil;
- P_{90} – 90º Percentil.

O coeficiente percentílico para curva normal é dado pelo valor 0,263.

Logo:

- C = 0,263 ⇒ Curva mesocúrtica;
- C < 0,263 ⇒ Curva leptocúrtica;
- C > 0,263 ⇒ Curva platicúrtica.

Exemplo 8: De acordo com as distribuições na tabela abaixo, calcule os graus de curtose e classifique as distribuições em relação à curva normal.

Distribuições	Q_1	Q_3	P_{10}	P_{90}
1	28,8	45,6	20,5	49,8
2	814	935	772	1012
3	63,7	80,3	55	86,6

- Distribuição 1:

$$C_1 = \frac{Q_3 - Q_1}{2 \times \left(P_{90} - P_{10}\right)} = \frac{45,6 - 28,8}{2 \times \left(49,8 - 20,5\right)} = 0,287$$

A distribuição 1 é platicúrtica.

× Distribuição 2:

$$C_2 = \frac{Q_3 - Q_1}{2 \times \left(P_{90} - P_{10}\right)} = \frac{935 - 814}{2 \times \left(1012 - 772\right)} = 0,252$$

A distribuição 2 é leptocúrtica.

× Distribuição 3:

$$C_3 = \frac{Q_3 - Q_1}{2 \times \left(P_{90} - P_{10}\right)} = \frac{80,3 - 63,7}{2 \times \left(86,6 - 55\right)} = 0,263$$

A distribuição 2 é mesocúrtica.

Atividades

1. Determine a amplitude total, o desvio médio e o desvio padrão (população) de cada um dos conjuntos de números abaixo:

   a. 12, 6, 7, 3, 15, 10, 18, 5.

   b. 9, 3, 8, 8, 9, 8, 9, 18.

2. A tabela abaixo, apresenta uma distribuição de frequência das alturas de 100 estudantes de estatística (população).

Grau	Número de estudantes
150 — 158	18
158 — 166	42
166 — 174	27
174 — 182	8
182 — 190	5
Total	100

Calcular:

a. Desvio Padrão

b. Variância

c. Desvio Médio

d. Amplitude Total

3. Faça uma análise dos dados abaixo:

Turma	Média Finais	Desvio Médio
A	7,2	4
B	8,3	6
C	4,3	2

4. Certo pesquisador aplicou um teste aos alunos de um colégio e obteve os seguintes resultados:

Acerto num teste de conhecimento.

37	38	38	39	40	41	41	42	42	43	43	44	44	44	45
45	46	46	46	46	46	49	50	50	50	53	53	55	56	57
58	58	60	60	63	63	64	64	64	65	65	66	67	67	68
68	70	70	71	72	72	72	74	75	75	76	77	78	78	79
80	82	82	84	84	86	90	92	94	96	98	98	99	99	99

Pede-se para determinar os seguintes itens:

a. Amplitude Total;

b. Desvio Médio;

c. Desvio Padrão;

d. Variância;

e. Desvio Quartil.

5. Na tabela a seguir, existem dados relativos aos quartis e percentis de duas distribuições de frequência. Calcular o coeficiente de curtose das distribuições abaixo.

Distribuições	Q_1	Q_3	P_{10}	P_{90}
1	32,8	54,4	30,5	69,8
2	730	820	680	912

6. Determine o coeficiente de assimetria de uma distribuição cujos dados estão apresentados abaixo:

   Média = 172, Mediana = 176, Moda = 175, S = 12,4.

7. Dado o conjunto de números 8, 4, 6, 9, 10 e 5 (considerando uma amostra). Determine:
   a. Desvio médio;
   b. Variância;
   c. Desvio padrão;
   d. Amplitude total;
   e. Desvio quartil.

Respostas

1. a) Ampliture = 15; Desvio Médio = 4,25; Desvio Padrão = 4,87. b) Ampliture = 15; Desvio Médio = 2,25; Desvio Padrão = 3,87.
2. Desvio Padrão = 8,24; Variância = 67,84; Desvio médio = 6,72; Amplitude Total = 40.
3. Podemos perceber que a turma "B" possui a melhor média, mas um desvio médio alto; isto quer dizer que existe uma variação grande de notas nesta turma. A turma "C", que possui a menor média e o menor desvio médio, é a turma mais homogênea, pois podemos perceber pelo desvio médio que os alunos tiraram nota em torno da média.
4. Amplitude total = 62; Desvio médio = 15,16; Desvio Padrão = 17,99; Variância = 323,82; Desvio quartil = 15,5.
5. C_1 = 0,27 - Curva platicúrtica; C_2 = 0,19 - Curva leptocúrtica.
6. Av = -0,97 - distribuição assimétrica negativa.
7. a) 2; b) 5,6; c) 2,366; d) 6; e) 2.

6

Correlação e regressão

A REGRESSÃO E a correlação são duas técnicas estreitamente relacionadas que envolvem uma forma de estimação. Mais especificamente, a análise de dados amostrais para saber se e como duas ou mais variáveis estão relacionadas uma com a outra numa população.

A ANÁLISE DE correlação dá um número que resume o grau de relacionamento entre duas variáveis. A análise de regressão tem como resultado uma equação matemática que descreve o relacionamento. A equação pode ser usada para estimar, ou predizer, valores futuros de uma variável quando se conhecem ou se supõem conhecidos valores da outra variável.

A ANÁLISE DE correlação é útil em trabalho exploratório, quando se pretende determinar quais variáveis são potencialmente importantes e o interesse está basicamente no grau ou força do relacionamento.

EM EDUCAÇÃO E psicologia, frequentemente, se dá maior ênfase ao grau ou força do relacionamento. Em outras áreas, como administração, economia, pesquisa média, agricultura, focaliza-se mais a natureza do relacionamento (isto é, a equação de predição), e a análise de regressão é o instrumento principal.

Relações Estatísticas e Correlações: São relações estabelecidas após uma pesquisa. Com base nos resultados da pesquisa, são feitas comparações que eventualmente podem conduzir (ou não) à ligação entre as variáveis.

Exemplo: relação entre a idade e a estatura de uma criança, ou a relação entre a classe social de uma pessoa e o número de viagens por ela realizado.

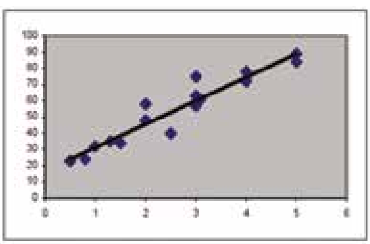
No estudo estatístico, a relação entre duas ou mais variáveis denomina-se correlação. A utilidade e importância das correlações entre duas variáveis podem conduzir à descoberta de novos métodos, cujas estimativas são vitais em tomadas de decisões.

6.1 Diagrama de dispersão

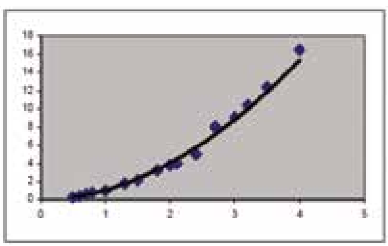
O diagrama de dispersão é um gráfico cartesiano em que cada um dos eixos corresponde às variáveis correlacionadas. A variável dependente (Y) situa-se no eixo vertical e o eixo das abscissas é reservado para a variável independente (X). Os pares ordenados formam uma nuvem de pontos.

A configuração geométrica do diagrama de dispersão pode estar associada a uma linha reta (correlação linear), uma linha curva (correlação curvilínea) ou, ainda, ter os pontos dispersos de maneira que não definam nenhuma configuração linear; nesta última situação, não há correlação.

Exemplos de diagramas de dispersão



Correção Linear

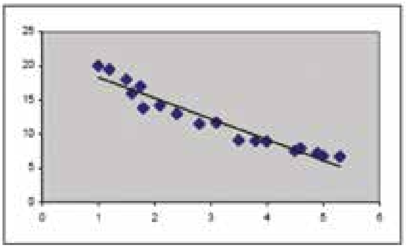


Correção Curvilínea

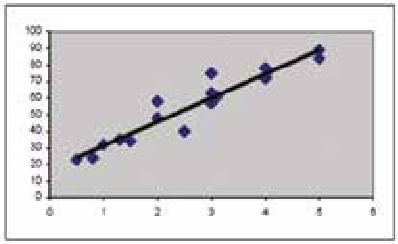
6.2 Correlação Linear

Correlação linear é uma correlação entre duas variáveis, cujo gráfico aproxima-se de uma linha. É uma linha de tendência, porque procura acompanhar a tendência da distribuição de pontos, que pode corresponder a uma reta ou a uma curva. Por outro lado, é, também, uma linha média, porque procura deixar a mesma quantidade de pontos abaixo e acima da linha.

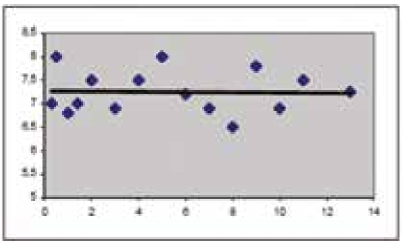
Exemplos:



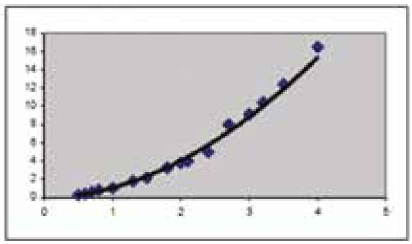
Correção Linear Positiva



Correção Linear Negativa



Não há Correlação



Relação Curvilínea direta

Para definir se a correlação entre as variáveis corresponde a uma linha reta ou a uma curva, pode-se utilizar modos qualitativos ou quantitativos. No modo qualitativo, vai imperar o "bom senso" do pesquisador para verificar qual o grau de intensidade na correlação entre as variáveis; isso significa o estabelecimento de uma relação numérica que medirá o nível da correlação.

6.3 Coeficiente de Correlação Linear (r)

O coeficiente de correlação linear pode ser apresentado como uma medida de correlação, pois tem como objetivo indicar o nível de intensidade que ocorre na correlação entre as variáveis. O coeficiente de correlação linear pode ser positivo ou negativo. O sinal positivo do coeficiente de correlação linear indica que o sentido da correlação corresponde a uma reta de inclinação descendente, e o sinal negativo corresponde a uma reta de inclinação ascendente. Uma das formas de medir o coeficiente de correlação linear foi desenvolvido por Pearson e recebe o nome de coeficiente de correlação de Pearson. O coeficiente de correlação de Pearson mede o grau de ajustamento dos valores em torno de uma reta.

Coeficiente de Correlação de Pearson (r):

$$r = \frac{n\sum x_i y_i - (\sum x_i)(y_i)}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right] \times \left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}}$$

Onde:

- r = o coeficiente de Pearson;
- n = o número de observações;
- x_i = variável independente;
- y_i =variável dependente.

O valor do coeficiente de correlação r tem a variação entre +1 e –1, ou seja, está limitado entre os valores do Intervalo[-1,+1].

- r = +1 (correlação positiva entre as variáveis);
- r = - 1 (correlação perfeita negativa entre as variáveis);
- r = 0 (não há correlação entre as variáveis ou, ainda, a correlação não é linear, caso exista).

Quanto mais próximo o valor de r estiver do valor “1”, mais forte a correlação linear. Quanto mais próximo o valor de r estiver do valor “0”, mais fraca a correlação linear.

Em geral, multiplica-se o valor de r por 100; dessa forma, o resultado passa a ser expresso em porcentagem. Na prática, estabelecem-se critérios para verificar os diversos níveis do fraco ao forte, chegando até o perfeito:

- $0<|r|<0{,}3$: a correlação é fraca e fica difícil estabelecer relação entre as variáveis. Em porcentagem: $0<|r|< 30\%$;
- $0{,}3\ |r|< 0{,}6$: a correlação é fraca, porém, podemos considerar a existência de relativa correlação entre as variáveis. Em porcentagem: $30\% \leq |r| <60\%$;
- $0{,}6\ |r| <1$: a correlação é de média para forte; a relação entre as variáveis é significativa, o que permite coerência com poucos conflitos na obtenção das conclusões. Em porcentagem: $60\% \leq |r| \leq 100\%$.

6.4 Regressão

A correlação linear é uma correlação entre duas variáveis, cujo gráfico aproxima-se de uma linha. O gráfico cartesiano que representa essa linha é denominado diagrama de dispersão. A regressão linear simples constitui uma tentativa de estabelecer uma equação matemática linear (linha reta) que descreva o relacionamento entre duas variáveis.

Para poder avaliar melhor a correlação entre as variáveis, é interessante obter a equação da reta; essa reta é chamada de reta de regressão e a equação que a representa é a equação de regressão. O diagrama de dispersão é construído de acordo com os dados amostrais de n observações e a equação de regressão é dada pela expressão:

$$Y= aX + b \rightarrow Y` =aX + b$$

Onde:

- X é a variável independente;
- $Y \rightarrow Y`$ é a variável dependente; na verdade, é a variável correlacionada com a variável X e sobre a qual se obtém um valor estimado.

Esse tipo de notação, de Y para Y`, caracteriza que não se trata de uma relação funcional para a determinação da reta, e sim, de uma relação estatística em que a distribuição está baseada em estimativas de dados colhidos por amostragem. Sendo “a” e “b” os parâmetros de equação da reta.

Para cálculo da equação da reta ou de uma curva, a partir de um gráfico, existem muitas maneiras diferentes, desde o mais simples método ao muito complexo. Este tipo de cálculo não faz parte da apresentação deste livro, pois, neste capítulo, limitaremo-nos a apresentar a maneira de calcular a equação de um gráfico e o coeficiente de regressão utilizando o Excel.

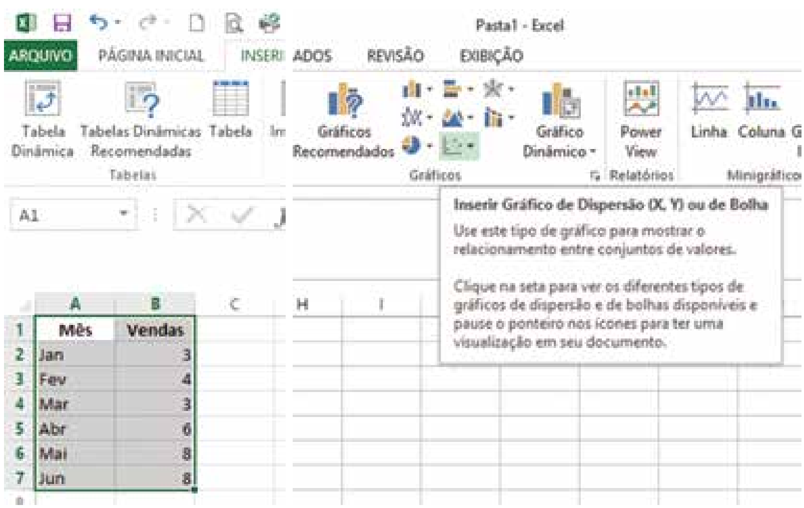
6.5 Utilizando o Excel

Os cálculos das equações e correlação e regressão são muito complicados e requerem, às vezes, conhecimento de matemática avançada, já, no Excel, basta clicar em um botão que a equação da reta ou da curva é calculado, bem como o coeficiente de correlação, como veremos a seguir.

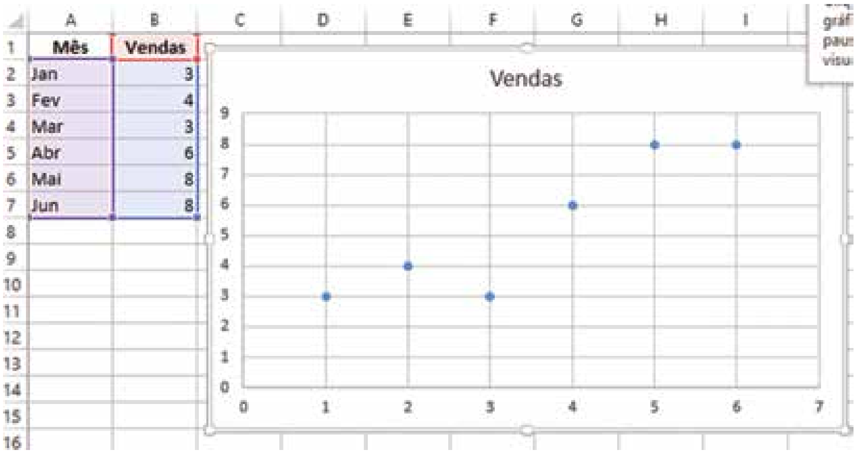
Exemplo: Suponha que você queira analisar as vendas de um determinado produto, conforme apresentado na tabela abaixo.

Mês	Vendas
Jan	3
Fev	4
Mar	3
Abr	6
Mai	8
Jun	8

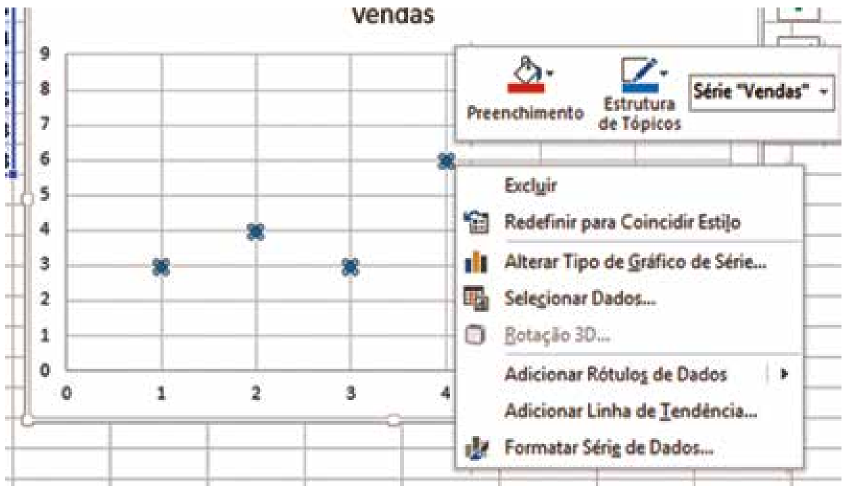
Primeiramente, você deve fazer um gráfico de dispersão. Conforme figura abaixo, marque a tabela, clique em "INSERIR", em seguida clique em "Inserir Gráfico de Dispersão".



Após clicar no botão “Inserir Gráfico de Dispersão”, você terá o gráfico conforme exposto a seguir, que representa exatamente a tabela apresentada.

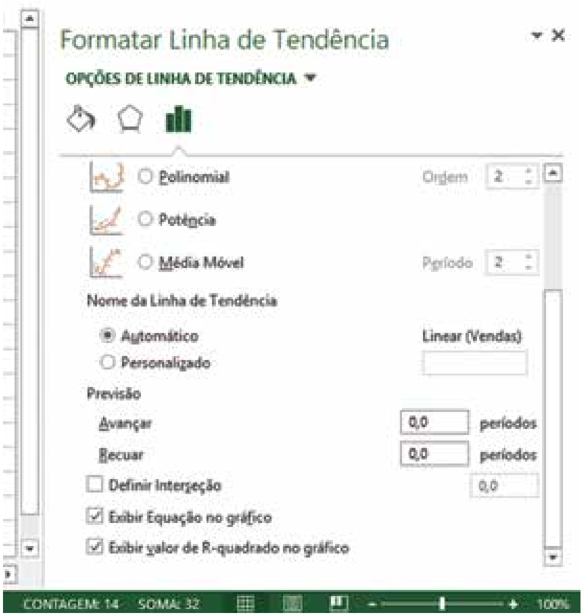


Clique com o botão da direita do mouse em cima de um dos pontos e escolha a opção “Adicionar Linha de Tendência”, conforme apresentado abaixo.

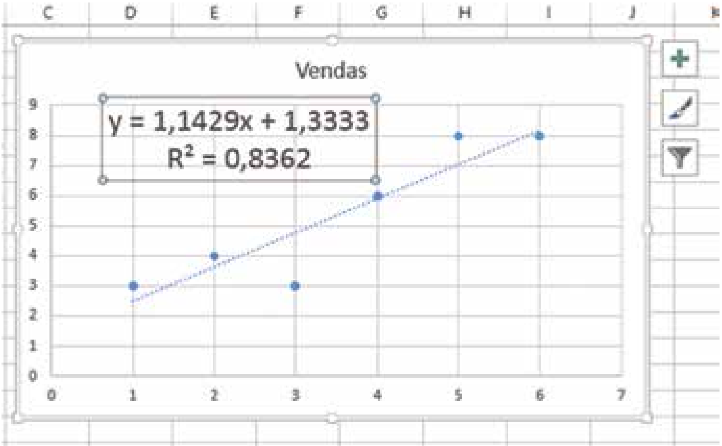


Vai aparecer no lado direito da tela um “menu” de opções, conforme apresentado abaixo.

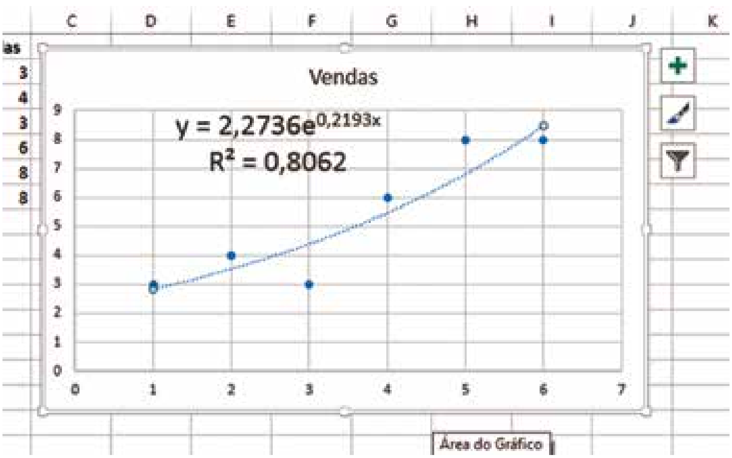
Neste "menu", você deve escolher o tipo de gráfico que melhor se adapta aos seus dados, por exemplo, se os dados tiverem o perfil "Exponencial", você deve escolher a tendência Exponencial e assim por diante. No caso deste exemplo, o gráfico de dispersão parece tender para uma reta linear. Então escolha o botão linear, conforme apresentado acima, desça a barra de rolagem a direita e escolha as opções "Exibir Equação no Gráfico" e "Exibir valor de R-quadrado no gráfico", conforme apresentado na figura abaixo.



Pronto, agora você já tem um gráfico com a linha de tendência, a equação da reta (regressão) e o coeficiente de regressão, que neste caso é apresentado como R^2.



Como saber se você escolheu a melhor opção de linha de tendência? Por exemplo, se neste caso você tivesse escolhido uma linha de tendência do tipo "Exponencial", conforme abaixo, o novo R^2 seria de 0,8062, que menor do que o R^2 da reta acima, que é de 0,8362. Lembre-se de que quanto mais perto de 1 for o R^2 melhor é sua linha de tendência e, consequentemente, melhor a equação que demonstra seus dados.



Exemplo: Suponhamos que uma cadeia de supermercados tenha financiado um estudo dos gastos com mercadoria para famílias de quatro pessoas. A investigação que limitou a famílias com renda líquida entre R$ 1.500 e R$ 8.000. Obteve-se a seguinte equação:

$$y=-200+2{,}3x, \text{ onde:}$$

- *y* - despesa anual estimada com mercadorias;
- *x* - renda líquida anual.

Suponha que a equação proporcione um ajustamento razoavelmente bom e que os dados tenham sido obtidos por métodos de amostragem aleatória. Estime a despesa de uma família de quatro pessoas com renda anual de R$ 2.000.

Solução: y= -200 + 2,3 . 2000= R$ 4.400

Atividades

1. A tabela abaixo apresenta os dados hipotéticos de quilometragem e preço de vendas de carros. Calcule o R^2 e a equação conforme solicitado:
   a. Regressão linear;
   b. Regressão exponencial;
   c. Qual a melhor opção.

Observação	Quilometragem (x_i)	Preço de Venda (y_i)
1	40	1000
2	30	1500
3	30	1200
4	25	1800
5	50	800
6	60	1000
7	65	500
8	10	3000
9	15	2500

Observação	Quilometragem (x_i)	Preço de Venda (y_i)
10	20	2000
11	55	800
12	40	1500
13	35	2000
14	30	2000

2. Uma equipe de engenheiros estabeleceu a seguinte relação para quilometragem urbana de carros de quatro cilindros com peso entre 1.500 e 3.000 kg.

   $y= 12 - 0{,}002x$, onde:

   y – estimativa do consumo km por litro;

   x – peso do carro.

   Calcule a estimativa do consumo de um carro que pese:

   a. 2.000 kg;

   b. 2.500 kg;

   c. 3.000 kg.

3. Determine pelo método de regressão linear, a equação que descreva a relação entre a frequência de acidentes e o nível de esforço preventivo (educacional) com base nos dados abaixo.

Homens/horas por mês para educação	Acidentes por milhão de homens/hora
200	7
500	6,4
450	5,2
800	4
900	3,1
150	8
300	6,5
600	4,4

4. A tabela abaixo apresenta os pesos de uma amostra de 12 pais e de seus filhos mais velhos.
   a. construir um diagrama de dispersão.
   b. determinar a equação de regressão e o coeficiente de correlação R^2.

Peso dos Pais	Peso dos Filhos
65	68
63	66
67	68
64	65
68	69
62	66
70	68
66	65
68	71
67	67
69	68
71	70

Respostas

1. a) $y= -38{,}555x + 2933{,}6$; $R^2 = 0{,}8086$ b)$y= 3753{,}1e^{-0{,}028x}$; $R^2 = 0{,}8366$ c) a regressão exponencial apresenta o R^2 mais próximo de 1, portanto é a curva que melhor se aproxima dos dados apresentados.

2. a) 8 kg/l; b) 7 kg/l; c) 6 kg/l.

3. $y= -0{,}0059x + 8{,}4431$; $R^2 = 0{,}9084$.

4. $y= 0{,}4764x + 35{,}825$/ $R^2 = 0{,}4937$.

7

Introdução Geral de Probabilidade

A FIM DE expor os conceitos básicos do modelo probabilístico que desejamos desenvolver, será conveniente o estudo de algumas ideias e conceitos matemáticos, com o objetivo de ajudar o aluno a entender melhor os conceitos dos modelos probabilísticos.

7.1 Revisão de Conjuntos

Um conjunto é uma coleção de objetos. Usualmente, conjuntos são representados por letras maiúsculas *A e B* etc. Existem três maneiras de descrever que objetos estão contidos no conjunto *A*:

a. Podemos fazer uma lista dos elementos de *A*. Por exemplo: $A = \{1, 2, 3, 4\}$; descreve o conjunto formado pelos inteiros positivos *1, 2, 3, 4*.

b. Podemos descrever o conjunto *A* por meio de palavras. Por exemplo: podemos dizer que *A* é formado de todos os números reais entre *0* e *1*, inclusive.

c. Para descrever o conjunto acima poderemos simplesmente escrever $A = \{x \mid 0 \leq x \leq 1\}$; isto é *A* é o conjunto de todos os "*x*", onde "*x*" é um número real entre 0 e 1, inclusive.

Os objetivos que individualmente formam a coleção ou conjunto *A* são denominados *membros ou elementos de A*. Quando "*a*" forma um elemento de *A*, escreveremos $a \in A$, e quando "*a*" não for um elemento de A, escreveremos $a \notin A$.

Definiremos o *conjunto universal* como o conjunto de todos os objetos que estejam sendo estudados. Este conjunto é geralmente representado pela letra "*U*". O *conjunto vazio* é aquele que não contém qualquer elemento, e geralmente é representado por "$\varnothing$".

Pode acontecer que, quando dois conjuntos *A* e *B* sejam considerados, ser elemento de *A* implique ser elemento de *B*. Nesse caso, diremos que *A* é um *subconjunto* de *B*, e escreveremos $A \subset B$. Interpretação semelhante será dada para $B \subset A$. Diremos que dois conjuntos constituem o mesmo conjunto, $A = B$, se e somente se, $A \subset B$ e $B \subset A$. Desse modo, dois conjuntos serão iguais se, e somente se, eles contiverem os mesmos elementos.

As duas seguintes propriedades do conjunto vazio e do conjunto fundamental são imediatas:

a. Para todo conjunto *A*, temos que $\varnothing \subset A$.

b. Desde que se tenha definido o conjunto fundamental, então, para todo conjunto *A*, considerado na composição de *U*, teremos $A \subset U$.

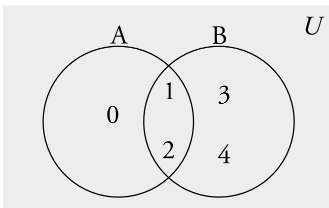
Exemplo: Suponha-se que U = {todos os números reais}, A = $\{x|x^2 + 2x - 3 = 0\} = \{-3, 1\}$, B = $\{x|(x-2)(x^2 + 2x - 3) = 0\} = \{-3, 1, 2\}$ e C = $\{x|\ x = -3, 1, 2\}$. Então, $A \subset B$ e B = C.

Definiremos *C* com a *união* de *A* e *B* (algumas vezes denominada a soma de *A* e *B*), da seguinte maneira:

$$C = x \mid x \in A \ ou \ x \in B \text{ (ou ambos)}\}$$

Escreveremos a união de *A* e *B*, assim : C= A ∪ B. Desse modo, *C* será formado de todos os elementos que estejam em *A*, ou em *B*, ou em ambos.

Exemplo: Suponha que A = {0, 1, 2} e B = {1, 2, 3, 4}; então, C = {0, 1, 2, 3, 4}.

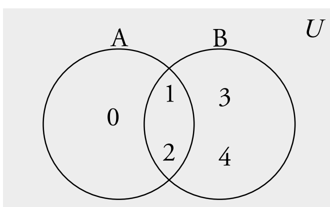


Definiremos *D* como a *interseção* de *A* e *B* (algumas vezes denominado o produto de *A* e *B*), da seguinte maneira:

$$D = x \mid x \in A \ e \ x \in B\}$$

Escreveremos a união de *A* e *B*, assim : D= $A \cap B$. Desse modo, *D* será formado de todos os elementos que estejam em *A*, ou em *B*, ou em ambos.

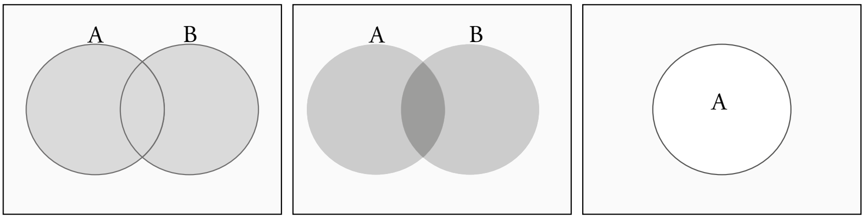
Exemplo: Suponha que *A* = {0, 1, 2} e B = {1, 2, 3, 4}; então, *D* = {1, 2}



Finalmente, introduziremos a noção de *complemento* de um conjunto *A*, na forma seguinte: o conjunto denotado por $\bar{A}$, constituído por todos os elementos que não estejam em A (mas que estejam no conjunto fundamental U) é denominado complemento de A. isto é, $\bar{A} = \{x \mid x \notin A\}$.

Exemplo: Suponha que $A = \{0, 1, 2\}$ e $U = \{1, 2, 3, 4\}$; então, $\bar{A} = \{3, 4\}$.

Um recurso gráfico, conhecido como D*iagrama de Venn*, poderá ser vantajosamente empregado quando estivermos combinando conjuntos, na maneira indicada acima. Em cada diagrama na figura abaixo, a região sombreada representa o conjunto sob exame.



$A \cup B$ $\qquad$ $A \cap B$ $\qquad$ $\bar{A}$

Exemplo: Suponha-se que $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$. Então encontraremos que:

$\bar{A} = \{5, 6, 7, 8, 9, 10\}$
$A \cup B = \{1, 2, 3, 4, 5, 6\}$
$A \cap B = \{3, 4\}$

Observe que ao descrever um conjunto (tal como $A \cup B$), cada elemento é relacionado apenas uma vez.

Identidades de conjuntos:

a. $A \cup B = B \cup A$

b. $A \cap B = A \cap B$

c. $A \cup (B \cup C) = (A \cup B) \cup C$

d. $A \cap (B \cap C) = (A \cap B) \cap C$

e. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

f. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

g. $A \cap \varnothing = \varnothing$

h. $A \cup \varnothing = A$

i. $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$

j. $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$

k. $\overline{\overline{A}} = A$

Definição*:* Sejam dois conjuntos *A* e *B*. Denominaremos *produto cartesiano* de *A* e *B*, denotando-o por *A* x *B*, o conjunto$\{(a, b), a \in A, b \in B\}$, isto é, o conjunto de todos os pares ordenados nos quais o primeiro elemento é tirado de *A* e segundo de *B*.

Exemplo: Suponha-se que *A* = *{1, 2, 3}* e *B* = *{1, 2, 3, 4}*. Então, *A* x *B* = *{(1,1); (1,2); (1, 3); (1,4); (2,1); ....;(2,4); (3,1); (3,2); ....(3,4)}.*

Observação: Em geral, $A \times B \neq B \times A$.

Número de elementos de um conjunto terá grande importância para nós. Se existir um número finito de elementos no conjunto A, digamos a_1, a_2, ..., an, diremos que A é finito. Se existir um número infinito de elementos em A, os quais possam ser postos em correspondência biunívoca com os inteiros positivos, diremos que A é numerável ou infinito numerável. Finalmente, deveremos considerar o caso de um conjunto infinito não-inumerável; este tipo de conjunto possui um número infinito de elementos que não podem ser enumerados. Pode-se mostrar, por exemplo, que para quaisquer dois números reais b > a, o conjunto $A = \{x | a \leq x \leq b\}$ contém um número não-inumerável de elementos.

7.2 Experimentos Determinísticos

Por essa expressão podemos nos referir a um modelo que estipule as condições sob as quais um experimento seja executado, *determinando* o resultado desse experimento. Então, podemos admitir que em um modelo determinístico é o resultado efetivo (numérico ou de outra espécie) seja determinado pelas condições sob as quais o experimento ou o procedimento seja executado.

Na natureza, existem muitos exemplos de "experimentos", para os quais modelos determinístico são apropriados. Por exemplo, as leis da gravitação explicam bastante precisamente o que acontece a um corpo que cai sob determinadas condições. As leis de Kepler nos dão o comportamento dos planetas. Em cada situação, o modelo especifica que as condições, sob as quais determinado fenômeno acontece, determinam o calor de algumas variáveis observáveis: a grandeza da velocidade, a área varrida durante determinado período de tempo, a aceleração de um corpo etc. Esses números aparecem em muitas das fórmulas com as quais estamos familiarizados.

7.3 Experimentos Não-Determinísticos

Em modelo não-determinístico ou aleatório, as condições da experimentação determinam somente o comportamento probabilístico do resultado observável.

Exemplos de experimentos não-determinísticos:

E_1: Jogue um dado e observe o número mostrado na face de cima;

E_2: Jogue uma moeda quatro vezes e observe o número de caras obtidas;

E_3: Uma lâmpada é fabricada. Em seguida é ensaiada à duração da vida, pela colocação em um soquete e anotação do tempo decorrido (em horas) até queimar.

E_4: Um lote de 0 peças contém 3 defeituosas. As peças são retiradas uma a uma (sem reposição da peça retirada) até que a última peça defeituosa seja encontrada. O número total de peças retiradas do lote é contado;

E_5: Peças são fabricadas até que 10 peças perfeitas sejam produzidas. O número total de peças fabricadas é contado;

E_6: De uma urna, que só contém bolas pretas, tira-se uma bola e verifica sua cor.

O que os experimentos acima têm em comum?

a. Cada experimento poderá ser repetido indefinidamente sob condições essencialmente inalteradas;

b. Muito embora não sejam capazes de afirmar que resultado particular ocorrerá, seremos capazes de descrever o conjunto de todos os possíveis resultados do experimento;

c. Quando o experimento for executado repetidamente, os resultados individuais parecerão ocorrer de uma forma acidental. Contudo, quando o experimento for repetido um grande número de vezes, uma configuração definida ou regularidade surgirá.

7.4 Espaço Amostral

Para cada experimento "E" do tipo que estamos considerando, definiremos o espaço amostral como o conjunto de todos os resultados possíveis de "*E*". Geralmente representaremos esse conjunto por "*S*".

Vamos considerar cada um dos experimentos acima e descrever um espaço amostral para cada um deles. O espaço amostral S_i se referirá ao experimento E_i.

S_1: {1, 2, 3, 4, 5, 6}

S_2: {0, 1, 2, 3, 4}

S_3: $\{t \mid t > 0\}$

S_4: {3, 4, 5, 6, 7, 8, 9, 10}

S_5: {10, 11, 12, 13, ...}

S_6: {bola preta}.

7.5 Eventos

Outra noção fundamental é o conceito de evento. Um evento "*A*" relativo a um particular espaço amostral *S*, associado a um experimento "*E*" é simplesmente um conjunto de resultados possíveis. Na terminologia dos conjuntos, um evento é um subconjunto de um espaço amostral *S*. Considerando nossa exposição anterior, isto significa que o próprio *S* constitui um evento, bem como 0 é o ∅. Qualquer resultado individual pode também ser tomado como um evento.

Alguns exemplos de eventos são dados a seguir. Novamente, nos referimos aos experimentos relacionados acima: A_i se referirá ao evento associado ao experimento E_i:

E_1: Jogue um dado e observe o número mostrado na face de cima;

S_1: {1,2,3,4,5,6}

A_1: Um número par ocorre, isto é, $A_1 = \{2, 4, 6\}$.

E_2: Jogue uma moeda quatro vezes e observe o número de caras obtidas;

S_2: {0, 1, 2, 3, 4}

A_2: {2}; isto é, duas caras.

E_3: Uma lâmpada é fabricada. Em seguida é ensaiada à duração da vida, pela colocação em um soquete e anotação do tempo decorrido (em horas) até queimar.

S_3: $\{t \mid t > 0\}$

A_3: $\{t \mid t < 3\}$; isto é, a lâmpada queima em menos de 3 horas.

E_4: Um lote de 0 peças contém 3 defeituosas. As peças são retiradas uma a uma (sem reposição da peça retirada) até que a última peça defeituosa seja encontrada. O número total de peças retiradas do lote é contado;

S_4: {3, 4, 5, 6, 7, 8, 9, 10}

A_4: {4}; o número de peças retiradas é 4.

E_5: Peças são fabricadas até que 10 peças perfeitas sejam produzidas. O número total de peças fabricadas é contado;

S_5: {10, 11, 12, 13, ...}

A_5: {20}; 20 peças foram produzidas.

E_6: De uma urna, que só contém bolas pretas, tira-se uma bola e verifica sua cor.

S_6: {bola preta}.

A_6: {bola preta}; a bola retirada é preta.

Agora, poderemos empregar as várias técnicas de combinar conjuntos, (eventos) e obter novos conjuntos (eventos), os quais já apresentamos anteriormente.

a. Se A e B forem eventos, $A \cup B$ será o evento que ocorrerá se, e somente se, A ou B (ou ambos) ocorrerem;

b. Se A e B forem eventos, $A \cap B$ será o evento que ocorrerá se, e somente se, A e B ocorrerem;

c. Se A for um evento, $\bar{A}$ será o evento que ocorrerá se, e somente se, não ocorrer A.

d. Se A_1, ..., A_n for qualquer coleção finita de eventos, então, $\bigcup_{i=1}^{n} A_i$ será o evento que ocorrerá se, e comente se, ao menos um dos eventos A_i ocorrer.

e. Se A_1, ..., A_n for qualquer coleção finita de eventos, então, $\bigcap_{i=1}^{n} A_i$ será o evento que ocorrerá se, e comente se, todos os eventos A_i ocorrerem.

f. Se A_1, ..., A_n for qualquer coleção infinita (numerável) de eventos, então, $\bigcup_{i=1}^{\infty} A_i$ será o evento que ocorrerá se, e comente se, ao menos um dos eventos A_i ocorrer.

g. Se A_1, ..., A_n.... for qualquer coleção infinita (numerável) de eventos, então, $\bigcap_{i=1}^{\infty} A_i$ será o evento que ocorrerá se, e comente se , todos os eventos A_i ocorrerem.

Definição: Dois eventos, A e B, são denominados mutuamente excludentes, se eles não puderem ocorrer juntos. Exprimiremos isso escrevendo $A \cap B = \varnothing$, isto é, a interação de A e B é o conjunto vazio.

Exemplo*:* Um dispositivo eletrônico é ensaiado e o tempo total de serviço t é registrado. Admitiremos que o espaço amostral seja $\{t \mid t \geq 0\}$. Sejam A, B e C três eventos definidos da seguinte maneira:

$A = \{t \mid t < 100\}$; $B = \{t \mid 50 \leq t \leq 200\}$; $C = \{t \mid t > 150\}$.

Consequentemente,

- $A \cup B = \{t|\ t \leq 200\}$
- $A \cap B = \{t|\ 50 \leq t \leq 100\}$
- $B \cup C = \{t|\ t \geq 50\}$
- $B \cap C = \{t|\ 150 \leq t \leq 200\}$
- $A \cap C = \varnothing$
- $A \cup C = \{t|\ t < 100 \text{ ou } t > 150\}$

7.6 Frequência Relativa

A fim de motivar a maneira de tratar o assunto, considera-se o seguinte procedimento: suponha-se que repetimos n vezes o experimento E. Admitamos que sejam, respectivamente, n_A e n_B o número de vezes que o evento A e o evento B ocorram nas n repetições.

Definição: $f_A = n_A/n$ é denominada frequência relativa do evento A nas n repetições de ε. A frequência relativa f_A apresenta as seguintes propriedades, de fácil verificação:

a. $0 \leq f_A \leq 1$;

b. $f_A = 1$ se, e somente se, A ocorrer em todas as n repetições;

c. $f_A = 0$ se, e somente se, A nunca ocorrer nas n repetições;

d. Se A e B forem eventos **mutuamente excludentes**, e se $f_{A \cup B}$ for a frequência relativa associada ao evento $A \cup B$, então, $f_{A \cup B} = f_A + f_B$.

7.7 Noções Fundamentais de Probabilidade

Definição: Seja ε um experimento. Seja S um espaço amostral associado a ε. A cada evento A associaremos um número real representado por P(A) e denominado **probabilidade de A**, que satisfaça às seguintes propriedades:

a. $0 \leq P(A) \leq 1$

b. $P(S) = 1$

c. Se A e B forem eventos **mutuamente excludentes**, $P(A \cup B) = P(A) + P(B)$

- **Teorema 1**: Se $\varnothing$ for o conjunto vazio, então $P(\varnothing) = 0$.
- **Teorema 2**: Se $\bar{A}$ for o evento complementar de A, então $P(A) = 1 - P(\bar{A})$.
- **Teorema 3**: Se A e B forem dois eventos quaisquer, então

$$P(A \cup B) = P(A) + P(B) - P(A \cup B)$$

- **Teorema 4**: Se A e B e C forem três eventos quaisquer, então

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- **Teorema 5**: Se $A \subset B$, então $P(A) \leq P(B)$.

7.7.1 Espaço Amostral Finito

Quando trabalhamos com experimentos em que o espaço amostral S seja formado de um número finito de elementos. Isto é, admitiremos que S possa ser escrito sob a forma $S = \{a_1, a_2, ..., a_k\}$.

A fim de caracterizar $P(A)$ para este modelo, deveremos inicialmente considerar o evento por um *resultado simples* (evento simples ou complementar), $A = \{a_i\}$.

A cada evento simples $\{a_i\}$ associaremos um número p_i, denominado probabilidade de $\{a_i\}$, que satisfaça às seguintes condições:

a. $p_i \geq 0, \quad i = 1, 2, ..., k$;

b. $p_1 + p_2 + ... + p_k = 1$.

Em seguida, suponha-se que um evento A seja constituído por r resultados, $1 \leq r \leq k$, a saber: $A = \{a_{j1}, a_{j2}, ..., a_{jr}\}$, onde $j_1, j_2, ..., j_r$ representam um qualquer dos r índices, de 1 até k, consequentemente, conclui-se que:

$$P(A) = p_{j1} + p_{j2} + ... + p_{jr}$$

7.8 Resultados Igualmente Verossímeis (igualmente provável, equiprovável)

Frequentemente, a característica física de um experimento sugere que aos vários resultados do espaço amostral sejam associadas probabilidades iguais. Esse espaço de probabilidade finito *S*, onde cada ponto tem a mesma probabilidade, será chamado de espaço igualmente verossímil ou equiprovável. Em particular, se *S* contém *n* pontos, então sua probabilidade é $\frac{1}{n}$. Além disso, se um evento *A* contém *r* pontos, então sua probabilidade é $r \times \frac{1}{n} = \frac{r}{n}$.

A hipótese comumente feita para espaços amostrais finitos é a de que todos os resultados sejam igualmente verossímeis. Esta hipótese não pode ser, contudo, tomada como segura; ela deve ser cuidadosamente justificada.

Se todos os *k* resultados forem igualmente verossímeis, segue-se que cada probabilidade será $p_1 = 1/k$. Consequentemente, a condição $p_1 + p_2 + ... + p_k = 1$ torna-se $k \times p_1 = 1$ para todos os *i*. Disto decorre que, para qualquer evento *A* formado de *r* resultados, teremos:

$$P(A) = \frac{r}{k}$$

Este método de avaliar *P(A)* é frequentemente enunciado da seguinte maneira:

$$P(A) = \frac{número\ de\ casos\ favoráveis\ a\ A\ pelos\ quais\ E\ pode\ ocorrer}{número\ total\ de\ casos\ pelos\ quais\ E\ pode\ ocorrer}$$

É muito importante compreender que a expressão de *P(A)* acima é apenas uma consequência da suposição de que todos os resultados sejam igualmente verossímeis, e ela é aplicável somente quando essa suposição for atendida. Ela certamente não serve como uma definição geral de probabilidade.

Exemplos:

1. Um dado é lançado e todos os resultados se supõem igualmente verossímeis. O evento A_1, ocorrerá se, e somente se, um número maior que 4 aparecer. Calcule a probabilidade de *A*.

$S = \{1, 2, 3, 4, 5, 6)$ A = {5, 6}; então:

$P(A) = 1/6 + 1/6 = 2/6$

2. Uma moeda equilibrada é retirada duas vezes. Seja *A* o evento: {aparece uma cara}. Calcule a probabilidade de *A*.

Resolução 1: $S = \{0, 1, 2\}$, aparece zero caras, uma cara ou duas caras

então $P(A) = 1/3$ **SOLUÇÃO INCORRETA !!!!!!!**

Esta análise é obviamente incorreta, porque no espaço amostral considerado acima, todos os resultados não são igualmente verossímeis. Podemos verificar no espaço amostral abaixo, que há dois eventos possíveis com uma cara.

Resolução 2: S = {CC, CK, KC, KK}; onde: C = cara e K = coroa, então;

$P(A) = 2/4 = ½$ **SOLUÇÃO CORRETA**

Esta análise é correta, porque neste espaço amostral todos os resultados são igualmente verossímeis.

3. Selecione aleatoriamente uma carta de um baralho comum de 52 cartas. Sejam *A* = {a carta é uma espada} e *B* = {a carta é uma figura, isto é, valete, dama ou rei}. Calcule $P(A)$, $P(B)$ e $P(A \cap B)$.

$$P(A) = \frac{\textit{número de espadas}}{\textit{número de cartas}} = \frac{13}{52} = \frac{1}{4}$$

$$P(A) = \frac{\textit{número de figuras}}{\textit{número de cartas}} = \frac{12}{52} = \frac{3}{13}$$

$$P(A \cap B) = \frac{\textit{número de figuras de espadas}}{\textit{número de cartas}} = \frac{3}{52}$$

4. Sejam os conjuntos $A = \{x \mid x \text{ é ímpar}\}$, $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, mostre:

   a. $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

   b. $A \cap B = \{1, 3, 5, 7, 9\}$

   c. $\bar{A} \cap B = \{2, 4, 6, 8, 10\}$

5. De um baralho comum de 52 cartas extrai-se uma carta ao acaso. Descreva o espaço amostral: a) não se levando em conta os naipes. B) levando-se em conta os naipes.

   a. O espaço amostral consiste de ás, dois, três, ... dez, valete, dama e rei. Podendo ser representado por: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13}, onde 11 é o valete, 12 é a dama e o 13 representa o rei.

   b. O espaço amostral consistirá de ás de copas, ás de ouros, ás de espada, ás de paus, ..., rei de copas, rei de ouros, rei de espada, rei de paus. Denotando copas, espada, ouro e paus respectivamente por C, E, O, P, por exemplo, podemos identificar o valete de espadas por (11, E). O espaço amostral consistirá da tabela abaixo:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
C	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	13C
E	1E	2E	3E	4E	5E	6E	7E	8E	9E	10E	11E	12E	13E
O	1O	2O	3O	4O	5O	6O	7O	8O	9O	10O	11O	12O	13O
P	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P	11P	12P	13P

Seja A o evento {extração de um rei} ou simplesmente {rei} e B {extração de uma carta de paus} ou simplesmente {carta de paus}. Descreva os eventos e calcule a probabilidade: d1) $A \cup B$, d2) $A \cap B$, d3) $\bar{A} \cap B$, d4) $A \cup \bar{B}$, d5) $A - B$, d6) $\bar{A} - \bar{B}$, d7) $(A \cap B) \cup (A \cup \bar{B})$

1) $A \cup B$ = {rei ou paus}

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
C	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	13C
E	1E	2E	3E	4E	5E	6E	7E	8E	9E	10E	11E	12E	13E
O	1O	2O	3O	4O	5O	6O	7O	8O	9O	10O	11O	12O	13O
P	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P	11P	12P	13P

$$P(A \cup B) = \frac{16}{52}$$

2) $A \cap B$ = {rei e paus)

	1	2	3	4	5	6	7	8	9	10	11	12	13
C	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	13C
E	1E	2E	3E	4E	5E	6E	7E	8E	9E	10E	11E	12E	13E
O	1O	2O	3O	4O	5O	6O	7O	8O	9O	10O	11O	12O	13O
P	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P	11P	12P	13P

$$P(A \cap B) = \frac{1}{52}$$

3) $\bar{A} \cap B$ = {não é rei e paus}. Como A = {rei}, então $\bar{A}$ = {não é rei}.

	1	2	3	4	5	6	7	8	9	10	11	12	13
C	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	13C
E	1E	2E	3E	4E	5E	6E	7E	8E	9E	10E	11E	12E	13E
O	1O	2O	3O	4O	5O	6O	7O	8O	9O	10O	11O	12O	13O
P	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P	11P	12P	13P

$$P(\bar{A} \cap B) = \frac{12}{52}$$

4) $A \cup \bar{B}$ = {rei ou copas ou ouros ou espadas}. Como B {extração de uma carta de paus}, então $\bar{B}$ {extração de uma carta que não seja paus} = {copas, ouros ou espadas}.

	1	2	3	4	5	6	7	8	9	10	11	12	13
C	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	13C
E	1E	2E	3E	4E	5E	6E	7E	8E	9E	10E	11E	12E	13E
O	1O	2O	3O	4O	5O	6O	7O	8O	9O	10O	11O	12O	13O
P	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P	11P	12P	13P

$$P(A \cup \bar{B}) = \frac{40}{52}$$

5) $A - B$ = {rei mas não paus}. É o mesmo resultado que $A \cap \bar{B}$.

	1	2	3	4	5	6	7	8	9	10	11	12	13
C	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	13C
E	1E	2E	3E	4E	5E	6E	7E	8E	9E	10E	11E	12E	13E
O	1O	2O	3O	4O	5O	6O	7O	8O	9O	10O	11O	12O	13O
P	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P	11P	12P	13P

$$P(A-B)=\frac{3}{52}$$

6) $\bar{A}-\bar{B}$ = {não rei "menos" não paus} = {não rei "não" não paus} = {não rei paus} =$\bar{A}\cup B$ que está apresentado no exemplo d3.

7) $(A \cap B) \cup (A \cup \bar{B})$ = {(rei e paus) ou (rei e não paus)} = {rei}. Chega-se a este resultado notando que d7)$(A \cap B) \cup (A \cup \bar{B})$= A

$$P\left[(A\cap B)\cup\left(A\cup\bar{B}\right)\right]=\frac{4}{52}$$

7.9 Técnicas de contagem

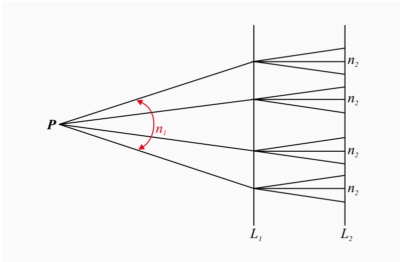
7.9.1 Princípios Fundamentais

Se algum procedimento pode ser realizado de n_1 maneiras diferentes; se, seguindo este, um segundo procedimento pode ser realizado de n_2 maneiras diferentes; se ainda, seguindo este segundo, um terceiro pode ser realizado de n_3 maneiras diferentes, e assim por diante; então, o número de maneiras nas quais podem realizar os procedimentos na ordem dada é o produto n_1 x n_2 x n_3 ....

7.9.2 Regra da Multiplicação

Suponha-se que um procedimento designado por 1 possa ser executado de n_1 maneiras. Admita-se que um segundo procedimento, designado por 2, possa ser executado de n_2 maneiras. Suponha-se, também, que cada maneira de executar 1 possa seguida por qualquer daquelas para executar 2.

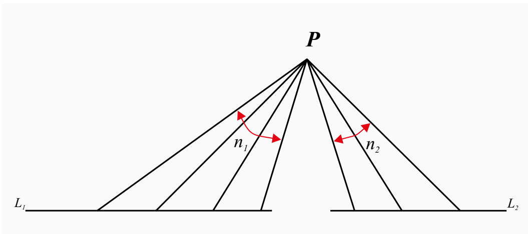
Então, o procedimento formado por 1 seguido de 2 poderá ser executado de n_1 x n_2 maneiras.



7.9.3 Regra da Adição

Suponha-se que um procedimento, designado por 1 possa ser executado de n_1 maneiras. Admita-se que um segundo procedimento, designado por 2, possa ser executado de n_2 maneiras. Além disso, suponha-se que não seja possível que ambos os procedimentos 1 e 2 sejam realizados em conjunto.

Então, o número de maneiras pelas quais poderemos realizar ou 1 ou 2 será $n_1 + n_2$.



Exemplos:

1. Suponhamos que uma placa de carro contenha duas letra distintas, seguidas por três dígitos, com o primeiro diferente de zero. Quantas placas podem ser impressas?

   Resolução:

   - A primeira letra pode ser apresentada de 26 maneiras diferentes;
   - A segunda letra pode ser apresentada de 25 maneiras diferentes, já que a letra impressa primeiro não pode ser escolhida para a segunda;

- O primeiro dígito pode ser apresentado de 9 maneiras, pois o primeiro deve ser diferente de zero;
- O segundo e o terceiro dígitos podem ser apresentados de 10 maneiras diferentes.

  Resposta*: Portanto, podem ser impressas 26 x 25 x 9 x 10 x 10 = 585.000 placas diferentes.*

2. Uma peça manufaturada deve passar por três estações de controle. Em cada estação, a peça é inspecionada para determinada característica e marcada adequadamente. Na primeira estação, três classificações são possíveis, enquanto nas duas últimas, quatro classificações são possíveis. De quantas maneiras uma peça pode ser marcada?

   Resposta: *Consequentemente, existem 3 x 4 x 4 = 48 maneiras pelas quais uma peça pode ser marcada.*

3. Suponha-se que estejamos planejando uma viagem e devamos escolher entre o transporte por ônibus ou por trem. Se existem três rodovias e duas ferrovias, quantos caminhos disponíveis existem para a viagem?

   Resposta: *existem 3 + 2 = 5 caminhos disponíveis para a viagem.*

7.9.4 Notação Fatorial

O produto dos inteiros positivos de 1 a n, inclusive, aparece frequentemente em matemática e, por isso, é representado pelo símbolo especial $n!$ (lê-se "n fatorial"):

$n! = 1 \times 2 \times 3 \times \ldots\ldots \times (n - 2) \times (n - 1) \times n$

É conveniente definir, também, $0! = 1$.

Exemplos:

a. $3! = 1 \times 2 \times 3 = 6$

b. $5! = 5 \times 4 \times 3!$

c. $\frac{8!}{6!} = \frac{8 \times 7 \times 6!}{6!} = 8 \times 7 = 56$

d. $12 \times 11 \times 10 = \dfrac{12 \times 11 \times 10 \times 9!}{9!} = \dfrac{12!}{9!}$

7.9.5 Permutações e Arranjos

Suponha-se que nós temos n objetos diferentes. De quantas maneiras ${}_nP_n$ poderemos dispor (permutar) esses objetos? Dessa maneira, o número de permutações de n objetos diferentes é dado por:

$${}_nP_n = n\,!$$

Considerem-se novamente n objetos diferentes. Agora desejamos escolher r desses objetos, $0 \leq r \leq n$ e permutar os r escolhidos. Denotaremos o número de maneiras de fazer isso (arranjos) por ${}_nA_r$.

$${}_nA_r = \frac{n!}{(n-r)!}$$

7.9.6 Combinações

Considere-se, novamente, n objetos diferentes. Agora, trataremos da contagem do número de maneiras a escolher r entre esses n objetos **sem considerarmos a ordem**. Definido para n inteiro positivo e r um inteiro tal que $0 \leq r \leq n$.

Portanto, o número de maneiras de escolher r dentre n objetos diferentes, não se considerando a ordem, é dado por:

$${}_nC_r = \frac{n!}{r!\,(n-r)!} = \binom{n}{r}$$

Exemplo:

1. Temos os objetos *a, b,* e *c*; calcule:

   a. de quantas maneiras ${}_nP_n$ podemos dispor esses objetos?

   ${}_3P_3 = n! = 3! = 6 \Rightarrow$ {abc, acb, bac, bca, cab, cba}.

No Excel

b. de quantas maneiras ${}_nA_r$ podemos dispor esses objetos, se r = 2?

${}_3A_2 = 6 \Rightarrow$ {ab, ba, ac, ca, bc, cb).

${}_3A_3 = {}_3P_3 = 6 \Rightarrow$ {abc, acb, bac, bca, cab, cba}.

c. de quantas maneiras ${}_nA_r$ podemos dispor esses objetos, se não considerarmos a ordem, dado que r = 2?

${}_3C_2 = 3 \Rightarrow$ {ab, ac, bc,).

${}_3C_3 = 1 \Rightarrow$ {abc}.

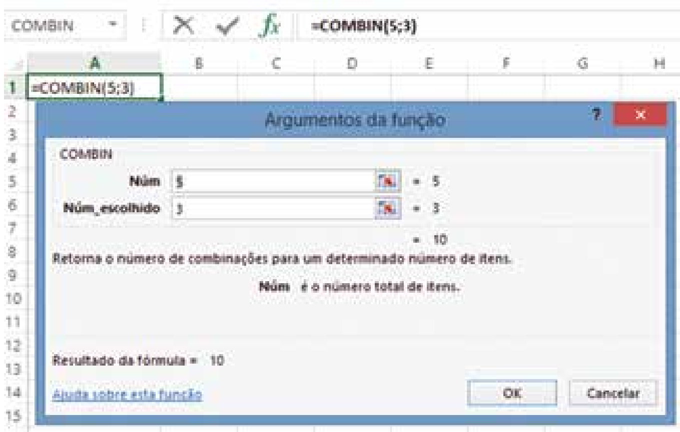
2. Dentre cinco pessoas, quantas comissões de três membros podem ser escolhidas?

Resposta: $${}_5C_3 = \frac{5!}{3!\,(5-3)!} = \frac{5!}{3! \times 2!} = \frac{5 \times 4 \times 3}{3!} = 10$$

isto é, 10 comissões disponíveis

Suponha que as 5 pessoas são chamadas de a, b, c, d, e; as comissões possíveis são: abc, abd, abe, acd, ace, ade, bcd, bce, bde, cde.

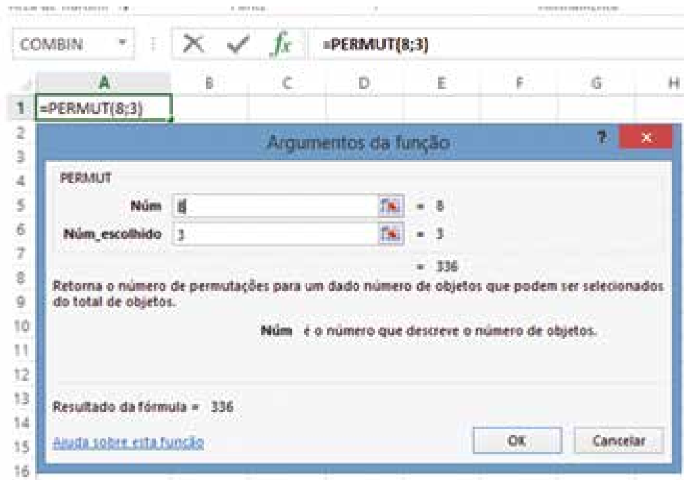
No Excel, podemos escrever a seguinte função em uma célula "= COMBIN(5;3)"



3. Com oito bandeiras diferentes, quantos sinais feitos com três bandeiras se podem fazer?

$$\textbf{Resposta:}\ {}_8A_3 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{8\times7\times6}{1} = 336\ \textit{sinais diferentes}.$$

No Excel, podemos escrever a seguinte função em uma célula "= PERMUT(8;3)"



4. Um grupo de oito pessoas é formado de cinco homens e três mulheres. Quantas comissões de três pessoas podem ser constituídas, incluindo exatamente dois homens?

$$\textbf{Resposta:}\ \binom{5}{2}\times\binom{3}{1} = \frac{5!}{2!(5-2)!}\times\frac{3!}{1!(3-1)!} = \frac{5\times4}{2!}\times\frac{3}{1!}$$

$$\binom{5}{2}\times\binom{3}{1} = 10\times3 = 30\ \textit{comissões}$$

5. Em um lote de 100 peças, formado por 20 peças defeituosas e 80 peças perfeitas, escolhemos ao acaso 10 peças sem reposição. Qual a probabilidade de acharmos exatamente 5 peças defeituosas e 5 peças perfeitas entre as 10 escolhidas?

Resposta: O número de maneiras de fazer isso é $S = \binom{100}{10}$

A = {achar exatamente 5 peças defeituosas e 5 peças perfeitas entre as 10 escolhidas}

$$P(A) = \frac{\binom{20}{5}\binom{80}{5}}{\binom{100}{10}} = 0{,}0215$$

6. Admita-se que se escolham ao acaso dois objetos, dentre os quatro denominados *a, b, c* e *d*.

   a. Se escolhermos sem reposição, o espaço amostral *S* poderá ser representado da forma abaixo:

   S = {(a,b);(a,c); (a,d); (b,c);(b,d);(c,d)}.

Então, existem ${}_4C_6 = 6$ resultados possíveis.
Cada um desses resultados indica somente quais
os dois objetos que foram escolhidos e não a
ordem em que eles foram escolhidos.

   b. Se escolhermos com reposição, o espaço amostral *S'* poderá ser representado por:

   S' = {(a,a); (a,b); (a,c); (a,d); (b,a); (b,b); (b,c); (b,d); (c,a); (c,b); (c,c); (c,d); (d,a); (d,b); (d,c); (d,d)}

   Existem $4^2 = 16$ resultados possíveis.

Definição: O número de maneiras de escolher
coisas de n, com reposição, é dado por nr. Neste caso, estaremos interessados na ordem em que as peças sejam escolhidas.

7. Uma caixa contém seis bolas numeradas de 1 a 6, uma pessoa vai pegar duas bolas com reposição (ela vai pegar uma bola, anotar o resultado, colocar a bola de novo na caixa, pegar uma bola e anotar o resultado). Responda as perguntas abaixo:

   a) Defina o espaço amostral. O espaço amostral é o conjunto de todos os resultados possíveis.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

= 36 resultados possíveis

   b) Calcule o número de todos os resultados possíveis.

   Vamos lembrar a definição da regra da multiplicação "suponha-se que um procedimento designado por 1 possa ser executado de n_1 maneiras. Admita-se que um segundo procedimento, designado por 2, possa ser executado de n_2 maneiras. Suponha-se, também, que cada maneira de executar 1 possa seguida por qualquer daquelas para executar 2.

   Como um dado possui 6 possibilidades de resultado possíveis e ele foi lançado duas vezes, o número de resultados possíveis será igual a 6^2=36.

   Então, podemos concluir que se este dado for lançado quatro vezes o número de resultados possíveis será igual a 6^4.

8. Uma caixa contém seis bolas numeradas de 1 a 6, uma pessoa vai pegar duas bolas sem reposição (ela vai pegar uma bola, anotar o resultado, pegar outra bola e anotar o resultado), sendo que a **ordem da retirada importa**, isto é 12 é diferente de 21. Responda as perguntas abaixo:

   a. Defina o espaço amostral. O espaço amostral é o conjunto de todos os resultados possíveis.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

= 30 resultados possíveis

Neste caso os resultados 11, 22, 33, 44, 55, 66 não fazem parte dos resultado possíveis, pois se a bola 1 for retirada de primeira, ela não estará na caixa para ser retirada na segunda vez e, assim por diante, para os outros resultados.

Verifique que a diferença entre o exercício 7 e 8 é somente a reposição, no exercício 7 as bolas voltam a caixa para a segunda retirada, o que não acontece no exercício 8.

b. Calcule o número de todos os resultados possíveis.

$$_nA_{r=}\frac{n!}{(n-r)!}=\frac{6!}{(6-2)!}=\frac{6\times5\times4!}{4!}=6\times5=30$$

Onde n = 6 (número total de bolas) e r = 2 (número de bolas retiradas).

Perceba que no espaço amostral foram pintados de cinza-escuro os 30 resultados possíveis.

9. Uma caixa contém seis bolas numeradas de 1 a 6, uma pessoa vai pegar duas bolas sem reposição (ela vai pegar uma bola, anotar o resultado, pegar outra bola e anotar o resultado), sendo que a **ordem da retirada não importa**, isto é, se cada bola representa uma pessoa, não importa se você escolheu a pessoa 1 depois a 2 ou o inverso, neste caso 12 é igual a 21. Responda as perguntas abaixo:

a. Defina o espaço amostral. O espaço amostral é o conjunto de todos os resultados possíveis.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

= 15 resultados possíveis

Neste caso os resultados 11, 22, 33, 44, 55, 66 não fazem parte dos resultado possíveis, pois se a bola 1 for retirada de primeira, ela não estará na caixa para ser retirada na segunda vez e, assim por diante, para os outros resultados. Os resultados 21, 31, 32, 41, 42, 43, 51, 52, 53, 54, 61, 62, 63, 64 e 65 também não fazem parte dos resultados possíveis porque estão representados por seus inversos na parte cinza mais escura do espaço amostral apresentado na tabela acima.

b. Calcule o número de todos os resultados possíveis.

$$_nC_{r=}\frac{n!}{r!(n-r)!}=\frac{6!}{(6-2)!}=\frac{6\times5\times4!}{2!4!}=\frac{6\times5}{2!}=\frac{6\times5}{2\times1}=\frac{30}{2}=15$$

Onde n = 6 (número total de bolas) e r = 2 (número de bolas retiradas).

Perceba que no espaço amostral foram pintados de cinza-escuro os 15 resultados possíveis.

Verifique que a diferença entre o exercício 8 e 9 é somente a importância ou não da ordem da retirada das bolas, no exercício 8 é importante e no 9 não.

10. Com relação ao apresentado no Exercício 7, calcule as seguintes probabilidades.

a. Sair números pares

A = {números pares}

11	12	13	14	15	16
21	22	23	24	25	26

31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

$$P(A) = \frac{15(\textit{números pares})}{30(\textit{total de possibilidades})} = \frac{1}{2}$$

b. a soma das bolas ser igual a 10.

A = {soma igual a 10}.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

$$P(A) = \frac{3(\textit{soma igual a}\,10)}{36(\textit{total de possibilidades})} = \frac{1}{12}$$

c. a soma das bolas ser menor que 7.

A = {soma menor do que 7}.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

$$P(A) = \frac{15(\textit{soma menor do que}\,7)}{36(\textit{total de possibilidades})} = \frac{15}{36} = \frac{5}{12}$$

11. Com relação ao apresentado no Exercício 8, calcule as seguintes probabilidades.

a. Sair números pares.

A = {números pares}.

	12	13	14	15	16
21		23	24	25	26
31	32		34	35	36
41	42	43		45	46
51	52	53	54		56
61	62	63	64	65	

$$P(A)=\frac{18(\textit{números pares})}{36(\textit{total de possibilidades})}=\frac{1}{2}$$

b. a soma das bolas ser igual a 10.

A = {soma igual a 10}.

	12	13	14	15	16
21		23	24	25	26
31	32		34	35	36
41	42	43		45	46
51	52	53	54		56
61	62	63	64	65	

$$P(A)=\frac{2(\textit{soma igual a}\,10)}{30(\textit{total de possibilidades})}=\frac{1}{15}$$

c. a soma das bolas ser menor que 7.

A = {soma menor do que 7}.

	12	13	14	15	16
21		23	24	25	26
31	32		34	35	36
41	42	43		45	46
51	52	53	54		56
61	62	63	64	65	

$$P(A)=\frac{12(\textit{soma menor do que}\,7)}{30(\textit{total de possibilidades})}=\frac{12}{30}=\frac{2}{5}$$

12. Com relação ao apresentado no Exercício 9, calcule as seguintes probabilidades.

a. Sair números pares.

A = {números pares}.

	12	13	14	15	16
		23	24	25	26
			34	35	36
				45	46
					56

$$P(A)=\frac{9(\textit{números pares})}{15(\textit{total de possibilidades})}=\frac{3}{5}$$

b. a soma das bolas ser igual a 10.

A = {soma igual a 10}.

12	13	14	15	16
	23	24	25	26
		34	35	36
			45	46
				56

$$P(A)=\frac{1(\textit{soma igual a}\,10)}{15(\textit{total de possibilidades})}=\frac{1}{15}$$

c. a soma das bolas ser menor que 7.

A = {soma menor do que 7}.

	12	13	14	15	16
		23	24	25	26
			34	35	36
				45	46
					56

$$P(A)=\frac{6(\textit{soma menor do que }7)}{15(\textit{total de possibilidades})}=\frac{6}{15}=\frac{2}{5}$$

13. Extrai-se ao acaso uma carta de um baralho de 52 cartas. Determine a probabilidade de a carta ser:

Para este exercício vamos representar os naipes como O – ouros; P – paus; C – copas; E – espadas e números de 1 a 13 para representar as cartas de "ás" ao "rei".

a. Um ás

A = {1}

$$P(A)=P(A\cap C\ ou\ A\cap E\ ou\ A\cap O\ ou\ A\cap P)$$
$$P(A)=P(A\cap C)+P(A\cap E)+P(A\cap O)+P(A\cap P)$$
$$P(A)=\frac{1}{52}+\frac{1}{52}+\frac{1}{52}+\frac{1}{52}=\frac{4}{52}=\frac{1}{13}$$

Poderíamos chegar a esta conclusão simplesmente contando o número de "ás" que existe no baralho (4) dividido pelo total de cartas do baralho (espaço amostral = 52).

b. Valete de copas

B = {11 de copas}

$$P(11\cap C)=\frac{1}{52}$$

c. Três de paus ou seis de ouros

C = {3 de paus ou 6 de ouros}

$$P(C)=P(3\cap P)+P(6\cap O)=\frac{1}{52}+\frac{1}{52}=\frac{2}{52}=\frac{1}{26}$$

d. Um dez ou uma carta de espada

D = {10 ou carta de espada}

$P(D)$=P(10∪E), como 10 e E não são mutuamente excludentes, temos que utilizar o Teorema 3 do item 7.7 deste capítulo, conforme reapresentado abaixo:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

então:

$$P(D) = P(10 \cup E) = P(10) + P(E) - P(10 \cup E) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{4}{13}$$

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
C	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C	11C	12C	13C
E	1E	2E	3E	4E	5E	6E	7E	8E	9E	**10E**	11E	12E	13E
O	1O	2O	3O	4O	5O	6O	7O	8O	9O	10O	11O	12O	13O
P	1P	2P	3P	4P	5P	6P	7P	8P	9P	10P	11P	12P	13P

Verificando a tabela acima, podemos comprovar o Teorema 3 apresentado, pois se apenas somássemos estaríamos somando duas vezes o 10 de espadas, por isto devemos retirar a interseção para retirar 1 dez de espada da soma e fazer a conta corretamente.

14. Extrai-se ao acaso uma bola de uma caixa que contém 6 bolas vermelhas, 4 brancas e 5 azuis. Determine a probabilidade de a bola extraída ser:

a. Vermelha

V = {vermelha}

$$P(V) = \frac{N^{o}\,de\,bolas\,vermelhas}{Total\,de\,bolas\,na\,caixa} = \frac{6}{15} = \frac{2}{5}$$

b. Não vermelha

B = {branca}

A = {azu}

$$P(\bar{A}) = \frac{N^{o}\,de\,bolas\,não\,vermelhas}{Total\,de\,bolas\,na\,caixa} = P(B) + P(A) = \frac{4}{15} + \frac{5}{15} = \frac{9}{15} = \frac{3}{5}$$

Atividades

1. Seja $U = \{1, 2, 3, 4, 5\}$ um universo e suponha que $A = \{1, 5\}$, $B = \{2, 3, 5\}$, $C = \{4, 2\}$ sejam subconjuntos de U. Determine:

   a) $A \cup (B \cup C)$

   b) $(A \cup B) \cup C$

   c) $A \cap (B \cup C)$

   d) $(A \cap B) \cup (A \cap C)$

   e) $\bar{A} \cup (\bar{B} \cap \bar{C})$

   f) $(A \cup B) - (A \cup C)$

2. Determine a probabilidade de sair um rei, um ás, um valete de paus ou uma dama de ouros na extração de uma carta de um baralho normal.

3. Determine a probabilidade do aparecimento do total de 8 na jogada de dois dados.

4. Determine a probabilidade de extração de uma peça perfeita, se em 600 peças previamente examinadas, foram encontradas 12 defeituosas.

5. Determine a probabilidade de aparecimento de 7 ou 11 em uma jogada de um par de dados.

6. Determine a probabilidade de ao menos uma cara sair em três jogadas de uma moeda.

7. Extrai-se, ao acaso, uma bola de uma caixa que contém 10 bolas vermelhas, 30 brancas, 20 azuis e 15 laranjas. Determine a probabilidade de extração de uma bola:

   a. laranja ou vermelha;

   b. não vermelha ou azul;

   c. não azul;

   d. branca.

8. Extrai-se, ao acaso, duas bolas **com** reposição de uma caixa que contém 10 bolas vermelhas, 30 brancas, 20 azuis e 15 laranjas. Determine a probabilidade de que:
   a. ambas sejam brancas;
   b. a primeira seja vermelha e a segunda branca;
   c. nenhuma seja laranja;
   d. ambas vermelhas, ou brancas, ou vermelha e branca (em qualquer ordem).

9. Extrai-se, ao acaso, duas bolas **sem** reposição de uma caixa que contém 10 bolas vermelhas, 30 brancas, 20 azuis e 15 laranjas. Determine a probabilidade de que:
   a. ambas sejam brancas;
   b. a primeira seja vermelha e a segunda branca;
   c. nenhuma seja laranja;
   d. ambas vermelhas, ou brancas, ou vermelha e branca (em qualquer ordem).

10. Calcule utilizando as regras de contagem:
    a. De quantas maneiras podemos escolher 6 questões de 10?
    b. De quantas maneiras podemos colocar 3 moedas distintas em 2 bolsas diferentes?
    c. De quantas maneiras podem 5 pessoas sentar-se em um sofá que dispõe apenas de 3 lugares?
    d. De quantas maneiras podemos escolher, dentre 8 homens e 6 mulheres, uma comissão composta de 3 homens e 4 mulheres?

11. Uma caixa contém 9 fichas numeradas de 1 a 9. Extraem-se três sucessivamente. Determinar a probabilidade de serem, alternadamente, ímpar-par-ímpar ou par-ímpar-par.

Respostas

1. a. {1,2,3,4,5}; b. {1,2,3,4,5}; c. {5}; d. {5}; e. ø; f. {3}
2. $\frac{5}{26}$
3. $\frac{5}{36}$
4. 0,98
5. $\frac{2}{9}$
6. $\frac{7}{8}$
7. a. $\frac{1}{3}$; b. $\frac{3}{5}$; c. $\frac{11}{15}$; d. $\frac{2}{5}$
8. a. $\frac{4}{25}$; b. $\frac{4}{75}$; c. $\frac{16}{25}$; d. $\frac{64}{225}$
9. a. $\frac{29}{185}$; b. $\frac{2}{37}$; c. $\frac{118}{185}$; d. $\frac{52}{185}$
10. a. 210; b. 8; c. 60; d. 840
11. $\frac{5}{18}$

8

Probabilidade Condicionada

Probabilidade condicional é um segundo evento de um espaço amostral que ocorre em um evento depois que já tenha ocorrido o primeiro.

Para melhor compreensão do que seja probabilidade condicional, considere um espaço amostral S finito, não vazio e um evento A de S, se quisermos outro evento B desse espaço amostral S, essa nova probabilidade é indicada por $P(B \mid A)$ e dizemos que é a probabilidade condicional de B em relação a A.

Em outras palavras, podemos simplificar da seguinte forma: a probabilidade condicionada refere-se à probabilidade de um evento A sabendo que ocorreu um outro evento B e representa-se por $P(A|B)$, lida "probabilidade condicional de A dado B" ou ainda "probabilidade de A dependente da condição B".

Essa probabilidade condicional irá formar um novo espaço amostral, pois agora o espaço amostral será A e os elementos do evento B irão pertencer a $B \cap A$.

8.1 Probabilidade Condicionada

Sejam A e B dois eventos (figura abaixo), com P(A) >0 . Denotemos por P($B|A$) a probabilidade de ocorrência de B, na hipótese de A ter ocorrido. Ora, como A ocorreu, A passa a ser o novo espaço amostral que vem substituir o espaço original S.

Estabeleceremos a seguinte *Definição*:

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$, desde que $P(A) > 0$.

Deste modo, temos duas maneiras de calcular a probabilidade condicionada $P(B \text{ I } A)$:

a. Diretamente, pela consideração da probabilidade de B em relação ao espaço amostral induzido A.

b. Empregando a definição acima, onde $P(A \cap B)$ e $P(A)$ são calculados em relação ao espaço amostral original S.

Vamos reexaminar a diferença entre extrair uma peça de um lote, ao acaso, com ou sem reposição, a seguir.

Exemplo*:*

1. Uma partida de 100 peças é composta de 20 peças defeituosas e 80 peças perfeitas. Suponha-se que escolhemos duas peças desse lote: (a) com reposição; (b) sem reposição.

   Definamos os dois eventos seguintes:

   A = {a primeira peça é defeituosa}; B = {a segunda peça é defeituosa}.

   Se estivermos extraindo com reposição, P(A) = P(B) = 20/100 =1/5, porque cada vez que extrairmos do lote, existirão 20 peças defeituosas no total de 100. No entanto, se estivermos extraindo sem reposição, os resultados não serão tão imediatos. É ainda verdade, naturalmente, que P(A) =1/5. Mas e sobre P(B)?- É evidente que, a fim de calcularmos P(B), deveremos conhecer a composição do lote no momento de se extrair a segunda peça. Isto é, deveremos saber se A ocorreu ou não.

Este exemplo mostra a necessidade de se introduzir o seguinte importante conceito.

Sejam A e B dois eventos associados ao experimento. Denotaremos por $P(B|A)$ a probabilidade condicionada do evento B, quando A tiver ocorrido.

No exemplo acima, $P(B|A) = 19/99$, porque se A tiver ocorrido, então para a segunda extração restarão somente 99 peças, das quais 19 delas serão defeituosas.

Sempre que calcularmos $P(B|A)$, estaremos essencialmente calculando P(B) em relação ao espaço amostral reduzido A, em lugar de fazê-lo em relação ao espaço amostral original S.

Exemplo:

2. Dois dados equilibrados são lançados, registrando-se o resultado como (x_1, x_2), onde x_i; é o resultado do *i-ésimo* dado, i =1, 2. Por isso, o espaço amostral S pode ser representado pela seguinte lista de 36 resultados igualmente prováveis.

$$s = \begin{Bmatrix} (1,1) & (1,2) & \cdots & (1,6) \\ (2,1) & (2,2) & \cdots & (2,6) \\ \vdots & \vdots & \vdots & \vdots \\ (6,1) & (6,2) & \cdots & (6,6) \end{Bmatrix}$$

Consideremos os dois eventos seguintes:

$A = \{(x_1 + x_2) \mid x_1 + x_2 = 10\}$ e $A = \{(x_1 + x_2) \mid x_1 > x_2\}$

Assim, $A = \{(5,5); (4,6); (6, 4)\}$ e $B = ((2,1); (3, 2); (3, 1); ....; (6, 5)\}$.

Portanto, $P(A) = 3/36$ e $P(B) = 15/36$. E $P(B|A) = 1/3$, uma vez que o espaço amostral é, agora, formado por A (isto é, três resultados), e somente um desses três resultados é coerente com o evento B. De modo semelhante, poderemos calcular $P(A \mid B) = 1/15$.

Exemplo:

3. Suponha-se que um escritório possua 100 máquinas de calcular. Algumas dessas máquinas são elétricas *(E)*, enquanto outras são manuais *(M)*;

e algumas são novas *(N)*, enquanto outras são muito usadas *(U)*. A tabela abaixo, dá o número de máquinas de cada categoria. Uma pessoa entra no escritório, pega uma máquina ao acaso e descobre que é nova. Qual será a probabilidade de que seja elétrica? Em termos da notação introduzida, desejamos calcular $P(E \mid N)$.

	Elétrica	Manual	Total
Nova	40	30	70
Usada	20	10	30
Total	60	40	100

Considerando-se somente o espaço amostral reduzido N (isto é, as 70 máquinas novas), temos $P(E \mid N)= 40 / 70 = 4 / 7$. Empregando a definição de probabilidade condicionada, temos que:

$$P(E|N)=\frac{P(E\cap N)}{P(N)}=\frac{40/100}{70/100}=\frac{4}{7}$$

A mais importante consequência da definição de probabilidade condicionada acima, é obtida ao se escrever:

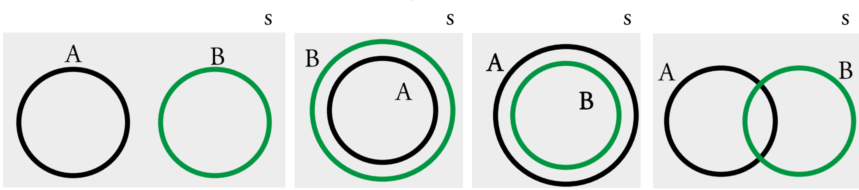
$\boldsymbol{P(A \cap B) = P(B \mid A) \times P(A)}$ ou, equivalentemente:

$\boldsymbol{P(A \cap B) = P(A \mid B) \times P(B)}$

Isto é, algumas vezes, mencionado como o teorema da multiplicação de probabilidades.

Podemos aplicar esse teorema para calcular a probabilidade da ocorrência conjunta dos eventos A e B.

Examinaremos agora, rapidamente, se poderemos fazer uma afirmação geral sobre a grandeza relativa de $P(A|B)$ e $P(A)$. Consideraremos quatro casos, que estão ilustrados pelos Diagramas de Venn abaixo:



a. $A \cap B = \varnothing$ b. $A \subset B$ c. $B \subset A$ d. $A \cap B \neq \varnothing$

Onde,

a. $P(A \mid B) = 0$, $0 \leq P(A)$; porque A não poderá ocorrer se B tiver ocorrido.

b. $P(A \mid B) = P(A \cap B) \div P(B) = [P(A) \div P(B)] \geq P(A)$; já que $0 \leq P(B) \leq 1$.

c. $P(A \mid B) = P(A \cap B) \div P(B) = [P(B) \div P(B)] = 1 \geq P(A)$.

d. Neste caso nada poderemos afirmar sobre a grandeza relativa de $P(A \mid B)$ e $P(A)$.

Observe-se que em dois dos casos acima, $P(A) \leq P(A \mid B)$; em um caso $P(A) \geq P(A \mid B)$; e no quarto caso, não podemos fazer qualquer comparação.

Até aqui, empregamos o conceito de probabilidade condicionada a fim de avaliar a probabilidade de ocorrência conjunta de dois eventos. Poderemos aplicar esse conceito em outra maneira de calcular a probabilidade de um evento simples A.

Necessitaremos da seguinte definição:

Definição: Dizemos que os eventos: $B_1, B_2, \ldots, B_k$, representam uma partição do espaço amostral S, quando:

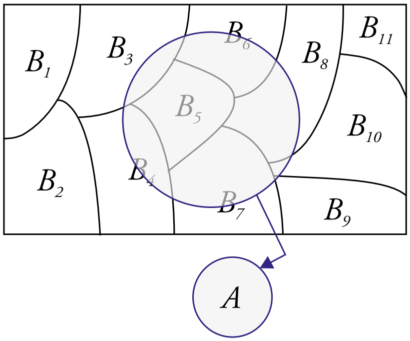
a. $B_i \bigcap B_j = \varnothing$, para todo $i \neq j$.

b. $\bigcup_{i=1}^{k} B_i = S$.

c. $P(B_i) > 0$ para todo i.

Explicando: Quando o experimento ***E*** é realizado um, e somente um, dos eventos B_i ocorre.

Por exemplo: na jogada de um dado, $B1 = \{1, 2\}$; $B2 = \{3, 4, 5\}$ e $B3 = \{6\}$ representariam uma partição do espaço amostral, enquanto $C1 = \{1, 2, 3, 4\}$ e $C2 = \{4, 5, 6\}$ não representariam.



Consideremos A um evento qualquer referente a S, e B_1, B_2, ..., B_k, uma partição de S.n O Diagrama de Venn acima ilustra isso para k = 8. Portanto, poderemos escrever:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup .... \cup (A \cap B_k).$$

Naturalmente, alguns dos conjuntos $(A \cap B_i)$ poderão ser vazios, mas isso não invalida essa decomposição de A. O ponto importante é que todos os eventos $(A \cap B_1) \cup$ $(A \cap B_2) \cup .... \cup (A \cap B_k)$ são dois a dois mutuamente excludentes. Por isso, poderemos aplicar a propriedade da adição de eventos mutuamente excludentes, e escrever:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_k).$$

Contudo, cada termo $P(A \cap B_i)$ pode ser expresso na forma $P(A \cap B_i)$ x $P(B_i)$ e, daí, obteremos o que se denomina o teorema da *PROBABILIDADE TOTAL*:

$$P(A) = P(A \mid B_1)\, P(B_1) + P(A \mid B_2)\, P(B_2) + ...+ P(A \mid B_K)\, P(B_K)$$

Este resultado representa uma relação extremamente útil, porque frequentemente, quando $P(A)$ é pedida, pode ser difícil calculá-la diretamente. No entanto, com a informação adicional de que B_j tenha ocorrido, seremos capazes de calcular $P(A \mid B_j)$ e, em seguida, empregar a fórmula acima.

Exemplo:

4. Consideremos o lote de 20 peças defeituosas e 80 não-defeituosas, do qual extrairemos duas peças, sem reposição. Definindo-se os eventos A e B como iguais a:

A = {a primeira peça extraída é defeituosa};

B = {a segunda peça extraída é defeituosa}.

poderemos, agora, calcular P(B), assim:

$$P(B)=P(B|A)P(A)+P(B|\overline{A})P(\overline{A})=\frac{19}{99}\cdot\frac{1}{5}+\frac{20}{99}\cdot\frac{4}{5}=\frac{1}{5}$$

Exemplo:

5. Uma determinada peça é manufaturada por três fábricas, digamos 1, 2 e 3. Sabe-se que 1 produz o dobro de peças que 2; e 2 e 3 produziram o mesmo número de peças (durante um período de produção especificado). Sabe-se também que 2% das peças produzidas por 1 e por 2 são defeituosas, enquanto 4% daquelas produzidas por 3 são defeituosas. Todas as peças produzidas são colocadas em um depósito, e depois uma peça é extraída ao acaso. Qual é a probabilidade de que essa peça seja defeituosa?

   Vamos introduzir os seguintes eventos: A = {a peça é defeituosa), B_1 = {a peça provem da fábrica 1}; B_2 = {a peça provém da fábrica 2}; B_3 = {a peça provém da fábrica 3}.

   Pede-se $P(A)$; e empregando-se o resultado acima, poderemos escrever:

   $P(A) = P(A \mid B_1)\ P(B_1) + P(A \mid B_2)\ P(B_2) + P(A \mid B_3)\ P(B_3)$

   $P(B_1)$ = 1/2, enquanto $P(B_2)=P(B_3)$= 1/4. Também, $P(A \mid B_1) = P(A \mid B_2)$ = 0,02; enquanto $P(A \mid B_3)$ = 0,04. Levando-se esses valores à expressão acima, encontraremos:

   $P(A) = 0{,}025$.

8.2 Teorema de Bayes

Poderemos empregar o *exemplo anterior* para sugerir outro importante resultado. Suponha-se que uma peça seja retirada do depósito e se verifique ser ela defeituosa. Qual é a probabilidade de que tenha sido produzida na fábrica 1?

Empregando a notação já introduzida, pede-se $P(B_1 \mid A)$. Poderemos calcular esta probabilidade como uma consequência da seguinte exposição: Seja B_1, B_2, ..., B_k, uma partição do espaço amostral S e seja A um evento associado a S. Aplicando-se a definição de probabilidade condicionada, poderemos escrever:

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}; i = 1, 2, 3, \ldots\ldots, k.$$

Este resultado é conhecido como Teorema de Bayes. É também denominado fórmula da probabilidade das "causas" (ou dos "antecedentes"). Desde que os B_i constituam uma partição do espaço amostral, um e somente um, dos eventos B_i ocorrerá. (Isto é, um dos eventos B_i deverá ocorrer e somente um poderá ocorrer.) Portanto, a expressão acima nos dá a probabilidade de um particular B_i (isto é, uma "causa"), dado que o evento A tenha ocorrido.

A fim de aplicar esse teorema, deveremos conhecer os valores das $P(B_i)$. Frequentemente, esses valores são desconhecidos, e isso limita a aplicabilidade do teorema. Tem havido considerável controvérsia sobre o Teorema de Bayes; ele é perfeitamente correto matematicamente; somente a escolha imprópria dos $P(B_i)$ pode tornar o resultado discutível.

Exemplo

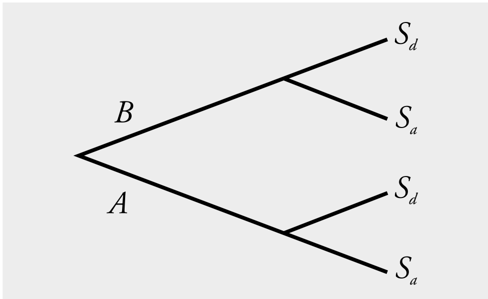
6. Voltando ao problema proposto acima, e agora aplicando o teorema de Bayes, obtemos:

$$P(B_1 \mid A) = \frac{(0{,}02)(1/2)}{(0{,}02)(1/2) + (0{,}02)(1/4) + (0{,}04)(1/4)} = 0{,}40$$

Exemplo:

7. Você agora se defronta com o seguinte problema de decisão: uma caixa do tipo desconhecido lhe é oferecida. Você terá permissão para tirar uma amostra de bombom (uma situação reconhecidamente irrealística, mas que nos permitirá introduzir ideias importantes, sem ficar muito complicado), e com esta informação você deve decidir se adivinha que a caixa que lhe foi oferecida é do tipo A ou se do tipo B. O seguinte "diagrama

de árvore” (assim denominado por causa dos vários passos ou ramos que aparecem) nos ajudará a analisar o problema. *(S_d e S_a correspondem,* respectivamente, a escolher um bombom de sabor doce ou um bombom de sabor amargo.)



Façamos alguns cálculos:

$P(A) = 0{,}6;$	$P(B) = 0{,}4;$	$P(S_d \text{ I } A) = 0{,}7;$
$P(S_a \text{ I } A) = 0{,}3;$	$P(Sd \mid B) = 0{,}3;$	$P(Sa \mid B) = 0{,}7.$

Desejamos realmente saber:

$P(A \mid S_d)$, $P(A \mid S_a)$, $P(B \mid S_d)$ e $P(B \mid S_a)$.

Suponha-se que realmente retiremos um bombom de sabor doce. Qual decisão seríamos mais tentados a tomar? Vamos comparar:

$$P(A \mid S_d) \quad \text{e} \quad P(B \mid S_d).$$

Empregando a fórmula de Bayes, teremos:

$$P(A \mid S_d) = \frac{P(S_d \mid A)P(A)}{P(S_d \mid A)P(A) + P(S_d \mid B)P(B)} = \frac{(0{,}7)(0{,}6)}{(0{,}7)(0{,}6) + (0{,}3)(0{,}4)} = \frac{7}{9}$$

Cálculo semelhante dará:

$$P(B \mid S_d) = \frac{2}{9}.$$

Dessa maneira, baseados na evidência que tivemos (isto é, a tirada de um bombom de sabor doce) é 2½ vezes mais provável que nós estejamos diante de uma caixa do tipo *A*, em vez de uma do tipo *B*. Consequentemente, poderíamos presumivelmente decidir que uma caixa do tipo A foi apresentada.

Exemplo:

8. A urna I contém três fichas vermelhas e 2 fichas azuis, e a urna II contém 2 fichas vermelhas e 8 fichas azuis. Joga-se uma moeda "honesta". Se a moeda der "cara", extrai-se uma ficha da urna I; se der "coroa", extrai-se uma ficha da urna II. Determine a probabilidade de escolha de uma ficha vermelha.

   Se R o evento "escolha de uma ficha vermelha", enquanto que I e II denotam, respectivamente, escolha da urna I e da urna II. Como se pode extrair uma ficha vermelha tanto da urna I como da urna II, podemos utilizar os resultados abaixo:

$$P(R) = P(I) \times P(R \mid I) + P(II) \times P(R \mid II) = \frac{1}{2}\left(\frac{3}{3+2}\right) + \frac{1}{2}\left(\frac{2}{2+8}\right) = \frac{2}{5}$$

9. Uma urna contém 5 fichas vermelhas e 4 brancas. Extraem-se sucessivamente duas fichas, sem reposição, constatando-se que a segunda é branca. Qual a probabilidade de a primeira também ser branca?

   Sejam B_1, B_2 os eventos "branca na 1ª extração", "branca na 2ª extração", respectivamente, desejamos $P(B_1|B_2)$. Essa probabilidade nos é dada por:

$$P(B_1 \mid B_2) = \frac{P(B_1 \cap B_2)}{P(B_2)} = \frac{\left(4/9\right)\left(3/8\right)}{4/9} = \frac{3}{8}$$

Atividades

1. Uma caixa contém 2 bolas vermelhas e 3 azuis. Extraem-se ao acaso duas bolas, sem reposição. Determine a probabilidade de serem:

   a. ambas azuis;

   b. ambas vermelhas;

   c. uma vermelha e uma azul.

2. Se, em uma família com dois filhos, ao menos um deles é menino, qual a probabilidade de o outro ser também menino? (considerar que o nascimento de meninos e meninas é igual a ½).

3. Determine a probabilidade de extração de 3 ases de um baralho usual de 52 cartas.

   a. com reposição.

   b. sem reposição.

4. A caixa I contém 3 bolas vermelhas e 5 brancas, e a caixa II, 4 vermelhas e 2 brancas. Extrai-se ao acaso uma bola da primeira caixa e coloca-se na segunda, sem observar a cor. Extrai-se então uma bola da segunda caixa. Qual a probabilidade de ser branca?

5. Uma caixa contém 3 bolas azuis e 2 vermelhas, e outra caixa contém 2 bolas azuis e 3 vermelhas. Extrai-se ao acaso uma bola de uma das caixas: é azul. Qual a probabilidade de ter sido extraída da primeira caixa?

6. A urna I contém 2 bolas brancas e 3 pratas; a urna II, 4 brancas e 1 preta; a urna III, 3 brancas e 4 pretas. Escolhe-se uma urna ao acaso e extrai-se uma bola: é branca. Qual a probabilidade de ter sido escolhida a primeira urna?

7. A probabilidade de um indivíduo atingir um alvo é 2/3. Se ele deve atirar até atingir o alvo pela primeira vez, determine a probabilidade de serem necessários cinco tiros.

8. Mulheres solteiras, com curso superior e renda até R$2.000,00. Esse é o perfil do brasileiro que busca aperfeiçoamento profissional gratuito na Internet, como mostra levantamento feito pelo FGV on-line, de março a setembro de 2009. (Jornal O Globo, 03 mar. 2010). O resultado desse levantamento é apresentado abaixo.

   × São mulheres: 58,3%.

   × Ganham até R$2 mil por mês: 77,7%.

- Têm graduação: 68,1%.
- Concentram-se em SP, RJ e MG: 62,8%.
- Ocupam o cargo de analista: 34,1%.

Considere que 2.000 pessoas participaram dessa entrevista e que, do total de pessoas que se concentram em São Paulo, Rio de Janeiro e Minas Gerais, 50% são homens. Escolhendo-se, ao acaso, um dos homens entrevistados,qual é, aproximadamente, a probabilidade de que ele seja de São Paulo, Rio de Janeiro ou Minas Gerais?

9. Um jogo consiste em lançar uma moeda honesta até obter duas caras consecutivas ou duas coroas consecutivas. Na primeira situação, ao obter duas caras consecutivas, ganha-se o jogo. Na segunda, ao obter duas coroas consecutivas, perde-se o jogo. A probabilidade de que o jogo termine, com vitória, até o sexto lance, é?

10. Uma urna contém 10 bolas brancas, 5 bolas amarelas e 10 bolas pretas. Uma bola é escolhida ao acaso da urna e verifica-se que não é preta, qual a probabilidade de ser amarela?

Respostas

1. a) 3/10; b) 1/10; c) 3/5
2. 1/3
3. a) 1/2197; b) 1/17.576
4. 21/56
5. 21/31
6. 14/57
7. 2/243
8. 75,3%
9. 31/64
10. 1/3

9

Distribuição Binomial, Normal e Poisson

Na maioria dos problemas estatísticos, a amostra não é suficientemente grande para determinar a distribuição da população de maneira muito precisa. Contudo, há normalmente bastante informação na amostra, juntamente com a informação obtida de outras fontes, de modo a sugerir o tipo geral da distribuição da população correspondente.

Combinando a experiência e a informação fornecidas pela amostra, pode-se comumente convencionar a natureza geral da distribuição da população. Esta convenção leva ao que é conhecido como distribuição de probabilidade ou distribuições teóricas.

Uma distribuição de probabilidade é um modelo matemático para a distribuição real de frequências. Neste capítulo, trataremos das três distribuições mais utilizadas: distribuição binomial, distribuição de Poisson e distribuição normal.

9.1 Distribuição Binomial (ou Distribuição de Bernoulli)

Suponhamos um experimento tal como a jogada repetida de uma moeda, ou a escolha de repetidas fichas de uma urna etc. Cada jogada ou extração é chamada de **prova**. Em cada prova em particular há uma probabilidade associada a um determinado evento, tal como aparecimento de "cara" na moeda, a escolha de uma ficha vermelha etc.

Em certos experimentos, a probabilidade não varia de prova para prova, tal como no caso da jogada de uma moeda, ou de um dado. Tais provas dizem-se então **independentes** e costumam designar-se como *provas de Bernoulli*.

Um processo de Bernoulli é um processo de amostragem no qual:

a. Em cada tentativa existem dois resultados possíveis e mutuamente exclusivos; eles são denominados, por conveniência de sucesso e fracasso.

b. As séries de tentativas, ou observações, são constituídas de eventos independentes.

c. A probabilidade de **sucesso**, é indicado por $\boldsymbol{p}$, permanece constante de tentativa para tentativa.

Em outras palavras, seja $\boldsymbol{p}$ a probabilidade de ocorrência de um evento em uma prova de Bernoulli (chamada de probabilidade de sucesso). Então, $\boldsymbol{q - 1 - p}$ é a probabilidade de não ocorrência do evento (chamado de probabilidade de fracasso).

A probabilidade de o evento ocorrer exatamente $\boldsymbol{x}$ vezes em $\boldsymbol{n}$ provas é dada pela função de probabilidade:

$$P(X=x) = {}_nC_x \times p^x \times q^{(n-x)} = \frac{n!}{x!(n-x)!} \times p^x \times q^{(n-x)}$$

Onde a variável aleatória $\boldsymbol{X}$ denota o número de sucessos em $\boldsymbol{n}$ provas, e $x = 0, 1, 2, ..., n$.

A seguir relacionamos algumas das importantes propriedades da distribuição binomial:

- Média: $\mu = np$
- Variância: $\sigma^2 = npq$
- Desvio Padrão: $\sigma = \sqrt{npq}$
- Coeficiente de Assimetria: $\alpha_3 = \dfrac{q-p}{\sqrt{npq}}$
- Coeficiente de Achatamento: $\alpha_4 = 3 + \dfrac{1-6pq}{\sqrt{npq}}$

Exemplos:

1. A probabilidade de obter exatamente 2 "caras" em 6 jogadas de uma moeda é:
   - X – sair cara (sucesso)
   - x – obter exatamente 2 caras
   - n – 6 jogadas
   - p – probabilidade de sucesso "sair cara", $p = ½$
   - q – probabilidade de fracasso "sair coroa", $q = 1 - p = ½$

   então, utilizando a equação:

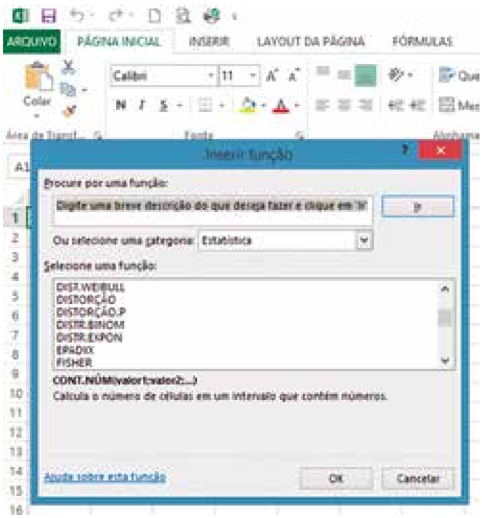
$$P(X = x) = {}_nC_x \times p^x \times q^{(n-x)} = \frac{n!}{x!(n-x)!} \times p^x \times q^{(n-x)}$$

obtemos:

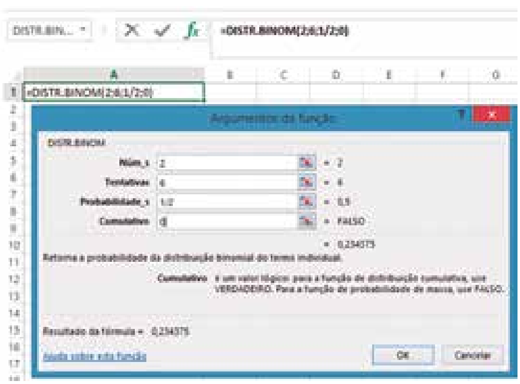
$$P(X = 2) = {}_6C_2 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^{(6-2)} = \frac{6!}{2!(6-2)!} \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^{(4)}$$

$$P(X = 2) = \frac{6 \times 5 \times 4!}{2!4!} \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^{(4)} = \frac{6 \times 5}{2!} \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^{(4)} = \frac{15}{64}$$

Exemplo no Excel: No Excel, podemos calcular desvio médio de conjunto de números utilizando a função "DISTR.BINOM" (clicar em função, depois escolher a categoria "Estatística" e, em seguida, escolher "DISTR.BINOM", conforme apresentado abaixo).



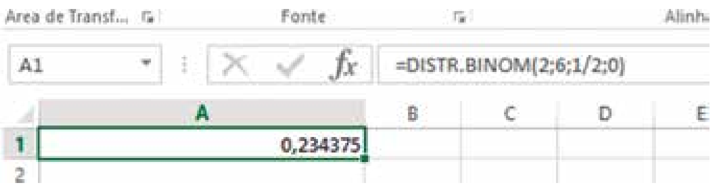
Conforme demonstrado abaixo, podemos verificar como calcular o desvio médio do exemplo anterior utilizando o Excel:

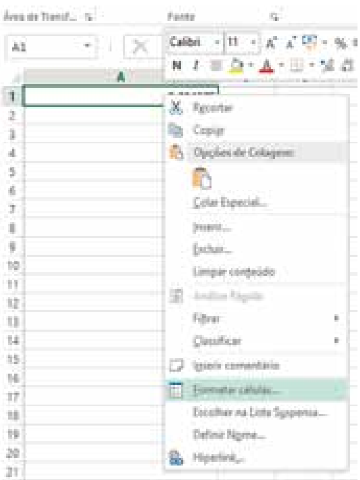


Onde:

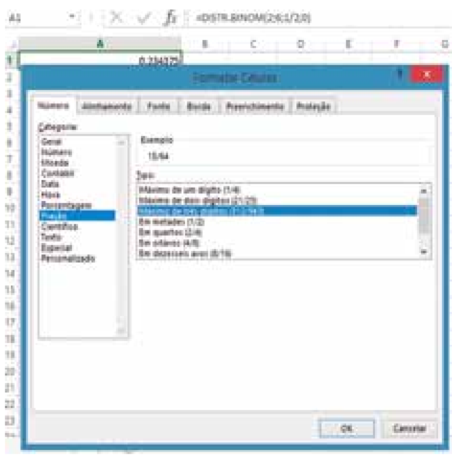
- Núm_s (x) – número de repetições de sucesso
- Tentativas (n) – número total de tentativas
- Probabilidade_s (p) – probabilidade de sucesso
- Cumulativo – sempre colocar 0 (zero)

Após você inserir as informações solicitadas acima, e clicar em “OK”, o Excel retornará a resposta apresentada abaixo:



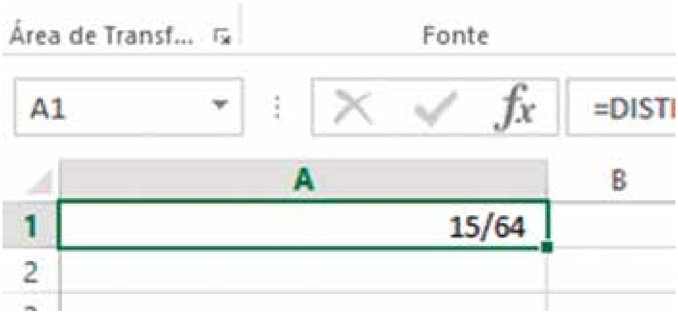


Caso você deseje verificar o resultado acima em forma de fração, clique com o botão direito do mouse em cima da célula, aparecerá a tela da figura ao lado, clique em “Formatar Células”, conforme indicado na figura. Aparecerá a tela da Figura abaixo.



Escolha a aba “Número”, clique em “Fração” e escolha a opção “Máximo de três dígitos (312/943). Clique em “OK”. Pronto, seu resultado sairá em forma de fração, conforme figura abaixo.

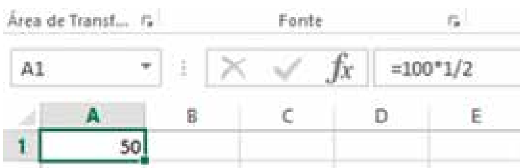
Verifique e tente fazer estes procedimentos para obter o resultado em forma de fração, conforme apresentado na figura abaixo.



2. Em 100 jogadas de uma moeda, calcule o número esperado de caras (número médio) e o desvio padrão.

   Sabe-se que o sucesso deste exercício é sair caras e que a probabilidade de sair uma cara é de ½ = 50%, então podemos calcular a média e o desvio padrão, pois p = ½ e q = ½.

× Média: $\mu = np = 100 \times \frac{1}{2} = 50\, caras$



No Excel você deve escrever dentro da célula a seguinte sintaxe: “=100*1/2”.

× Desvio Padrão: $\sigma = \sqrt{npq} = \sqrt{100 \times \frac{1}{2} \times \frac{1}{2}} = 5$

No Excel você deve escrever dentro da célula a seguinte sintaxe: “= RAIZ (100*1/2*1/2)”.

3. Determinar a probabilidade de ocorrência de três 6 em cinco lances de um dado honesto.

   × X – sair seis

   × x – obter exatamente três seis

   × n – 5 jogadas

   × p – probabilidade de sucesso "sair seis", p= 1/6

   × q – probabilidade de fracasso "não sair seis", $q = 1 - p = 5/6$

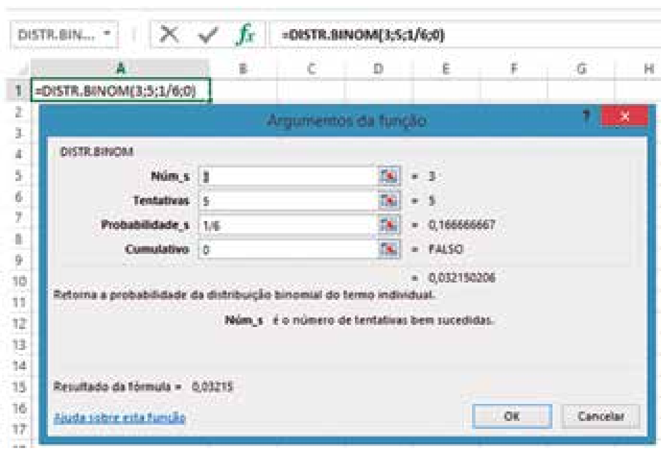
   Obs.: lembre-se que ao jogar um dado uma única vez, a probabilidade de ocorrer qualquer resultado é igual a 1/6 (uma chance em seis resultados possíveis).

$$P(X=3) = {}_5C_3 \times \frac{1}{6}^3 \times \frac{5}{6}^{(5-3)} = \frac{5!}{3!(5-3)!} \times \frac{1}{6}^3 \times \frac{5}{6}^{(5-3)}$$

$$P(X=3) = \frac{5\times4\times3\times2!}{3!\times2!} \times \frac{1}{6}^3 \times \frac{5}{6}^2 = \frac{5\times4\times3}{3\times2\times1} \times \frac{1}{6}^3 \times \frac{5}{6}^2 = 10 \times \frac{1}{216} \times \frac{25}{36}$$

$$P(X=3) = \frac{250}{7776} = 0{,}03215 = 3{,}215\%$$

No Excel, você pode escrever a seguinte sintaxe dentro da célula:



4. Verifica-se, em uma fábrica, que, em média, 10% dos parafusos produzidos por uma determinada máquina não satisfazem a certas especificações. Se forem selecionados, ao acaso, 10 parafusos da produção diária dessa máquina, determinar a probabilidade de exatamente três serem defeituosas.

   - X – parafuso ser defeituoso (sucesso)
   - x – obter exatamente 3 parafusos defeituosos
   - n – 10 parafusos serão analisadas
   - p – probabilidade de sucesso "sair seis", p= 10%=0,1
   - q – probabilidade de fracasso "não sair seis", $q = 1 - p = 0{,}9$

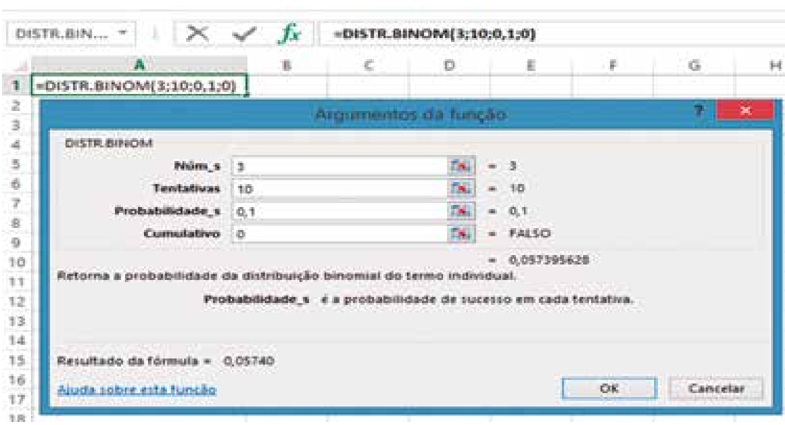
Note que para o exercício, o sucesso é "peça ser defeituosa", isto faz sentido pois estamos procurando a probabilidade de peças defeituosas.

$$P(X=3) = {}_{10}C_3 \times 0{,}1^3 \times 0{,}9^{(10-3)} = \frac{10!}{3!(10-3)!} \times 0{,}1^3 \times 0{,}9^{(10-3)}$$

$$P(X=3) = \frac{10 \times 9 \times 8 \times 7!}{3! \times 7!} \times 0{,}1^3 \times 0{,}9^7 = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} \times 0{,}1^3 \times 0{,}9^7$$

$$P(X=3) = 120 \times 0{,}1^3 \times 0{,}9^7 = 0{,}0574 = 5{,}74\%$$

No Excel, você pode escrever a seguinte sintaxe dentro da célula:



5. Devido às altas taxas de juros, uma empresa informa que 30% de suas contas a receber de outras empresas comerciais se encontram vencidas. Se um contador escolhe aleatoriamente uma amostra de cinco contas, determine a probabilidade de exatamente 20% das contas estarem vencidas.

   - *X* – contas vencidas (sucesso)
   - *x* – obter exatamente 20% de contas vencidas. Se o contador vai retirar 5 contas, 20% de 5 é igual a **1** conta vencida.
   - *n* – 5 contas serão analisadas
   - *p* – probabilidade de sucesso "ter conta vencida", *p*= 30%=0,3
   - *q* – probabilidade de fracasso "não ter conta vencida", $q = 1 - p = 0,7$

   $$P(X=1) = {}_5C_1 \times 0,3^1 \times 0,7^{(5-1)} = \frac{5!}{1!(5-1)!} 0,3^1 \times 0,7^4 = \frac{5 \times 4!}{1!4!} 0,3^1 \times 0,7^4$$

   $P(X=1) = 5 \times 0,3^1 \times 0,7^4 = 0,36015 = 36,015\%$

   No Excel, você pode escrever a seguinte sintaxe dentro da célula: "= DISTR.BINOM(1;5;0,3;0)"

6. Uma fábrica de pedidos de Correios envia uma carta circular que terá uma taxa de resposta de 10%. Suponha que 20 cartas circulares são endereçadas a uma nova área geográfica como um teste de mercado. Supondo que na nova área é aplicável a taxa de resposta de 10%, determinar a probabilidade de apenas uma pessoa responder.

   - *X* – cartas respondidas (sucesso)
   - *x* – uma pessoa responder
   - *n* – 20 cartas são enviadas
   - *p* – probabilidade de sucesso "carta respondida", *p*= 10%=0,1
   - *q* – probabilidade de fracasso "carta não respondida", $q = 1 - p = 0,9$

   $$P(X=1) = {}_{20}C_1 \times 0,1^1 \times 0,9^{(20-1)} = \frac{20!}{1!(20-1)!} \times 0,1^1 \times 0,9^{19}$$

$$P(X=1)=\frac{20\times 19!}{1!19!}\times 0,1^{1}\times 0,9^{19}=20\times 0,1^{1}\times 0,9^{19}=27,017\%$$

No Excel, você pode escrever a seguinte sintaxe dentro da célula: "=DISTR.BINOM(1;20;0,1;0)"

7. Durante um certo ano, 70% das ações ordinárias negociadas na Bolsa de Valores de São Paulo tiveram aumentadas suas cotações, enquanto 30% tiveram suas cotações diminuídas ou estáveis. No começo do ano, um serviço de assessoria financeira escolhe 10 ações como sendo especialmente recomendadas. Se as 10 ações representam uma seleção aleatória, qual a probabilidade de que todas as 10 ações escolhidas tenham tido suas cotações aumentadas?

   - X – ações com cotação aumentada (sucesso)
   - x – 10 ações escolhidas tiveram suas cotações aumentadas
   - n – 10 ações escolhidas
   - p – probabilidade de sucesso "carta respondida", p= 70%=0,7
   - q – probabilidade de fracasso "carta não respondida", q = 1 – p = 0,3

$$P(X=10)={}_{10}C_{10}\times 0,7^{10}\times 0,3^{(10-10)}=\frac{10!}{10!(10-10)!}\times 0,7^{10}\times 0,3^{0}$$

P(X=10)= 1 × $0,7^{10}$×1=2,8248%

No Excel, você pode escrever a seguinte sintaxe dentro da célula: "=DISTR.BINOM (10;10;0,7;0)"

9.2 Distribuição de Poisson

A distribuição de Poisson pode ser usada para determinar a probabilidade de um dado número de sucessos quando os eventos ocorrem em um tempo contínuo. Tal processo, é similar ao processo Binomial exceto que os eventos ocorrem de forma contínua ao invés de ocorrerem em tentativas ou observações fixadas.

Distribuição de Poisson é a probabilidade estatística usada para registrar a ocorrência de eventos imprevisíveis em um grande número de tentativas que se repetem.

Se a probabilidade de sucesso for muito pequena e o número de experiências grande, teremos então a Distribuição de Poisson.

Uma aplicação comum é prever o número de eventos em um determinado período de tempo, como, por exemplo, o número de carros que chegam a um posto de gasolina em uma determinada hora.

Somente um valor é necessário para determinar a probabilidade de um dado número de sucessos em um processo de Poisson: o número médio de sucessos para a específica dimensão de interesse. Este número médio é geralmente representado por λ ou μ. A equação para se determinar a probabilidade de um dado número ***X*** de sucessos em uma distribuição de Poisson é:

$$P(X \mid \mu) = \frac{\mu^X \times e^{-\mu}}{X!}$$

Onde "e" é a constante 2,71828 usada em conexão com os logaritmos naturais, sendo que os valores de $e^{-\mu}$ podem ser obtidos na Tabela 1 no final deste capítulo. A representação lê-se a probabilidade de acontecer ***X*** tal que a média conhecida é μ.

9.2.1 Premissas da Distribuição de Poisson

- A probabilidade de ocorrência de um sucesso em cada unidade de medida é a mesma.
- A ocorrência de um sucesso numa unidade de medida é independente das que ocorrem em qualquer outra unidade.

Com Distribuição de Poisson analisamos a ocorrência de um evento aleatório com probabilidade por unidade de tempo constante.

Por exemplo, o número de clientes que entram na agência durante a primeira hora do horário de atendimento do banco, o número de solicitações por hora dos usuários de microcomputadores de uma grande empresa de telemarketing, o número de clientes atendidos das 8hs às 9hs da manhã em um posto de gasolina etc.

9.2.2 Características da Distribuição de Poisson

- O número de sucessos que ocorrem num intervalo ou uma região especificada é independente daqueles que ocorrem em qualquer outro intervalo de tempo ou região disjunto (que não tenham elemento em comum).
- A probabilidade de ocorrência de um único sucesso durante um intervalo pequeno ou numa região é proporcional ao comprimento do intervalo ou região e não depende do número de sucessos que ocorrem fora deste intervalor ou região.
- A probabilidade de mais de um sucesso ocorrendo em tais intervalos pequenos ou regiões pequenas é desprezível.

Em todas estas situações, temos um conjunto de ocorrências que satisfazem as seguintes condições:

- O número de ocorrências de um evento em um intervalo de tempo (espaço) é independente do número de ocorrências do evento em qualquer outro intervalo disjunto – ocorrências independentes umas das outras.
- A probabilidade de duas ou mais ocorrências simultâneas é praticamente zero.
- O número médio de ocorrências por unidade de tempo (espaço) é constante ao longo do tempo (espaço) – ocorrências distribuídas uniformemente sobre o intervalo considerado.
- O número de ocorrências durante qualquer intervalo depende somente da duração ou tamanho do intervalo; quanto maior o intervalo, maior o número de ocorrências.

A distribuição de Poisson é muito usada nas seguintes situações:

- Carros que passam por um cruzamento por minuto, durante uma certa hora do dia;
- Erros tipográficos por página, em um material impresso;
- Defeitos por unidade (m^3, m^2, m, etc.) por peça fabricada;
- Colônias de bactérias numa dada cultura por $0{,}01mm^2$, numa plaqueta de microscópio;

- Mortes por ataque de coração por ano, numa cidade;
- Problemas de filas de espera em geral, e outros;
- Clientes chegando ao caixa de um supermercado;
- Acidentes com automóveis em uma determinada estrada;
- Número de carros que chegam a um posto de gasolina;
- Número de falhas em componentes por unidade de tempo;
- Número de requisições para um servidor em um intervalo de tempo;
- Número de peças defeituosas substituídas num veículo durante o primeiro ano de vida;

9.2.3 Propriedades

Média = μ

Variância = μ

Desvio Padrão = $\sqrt{\mu}$

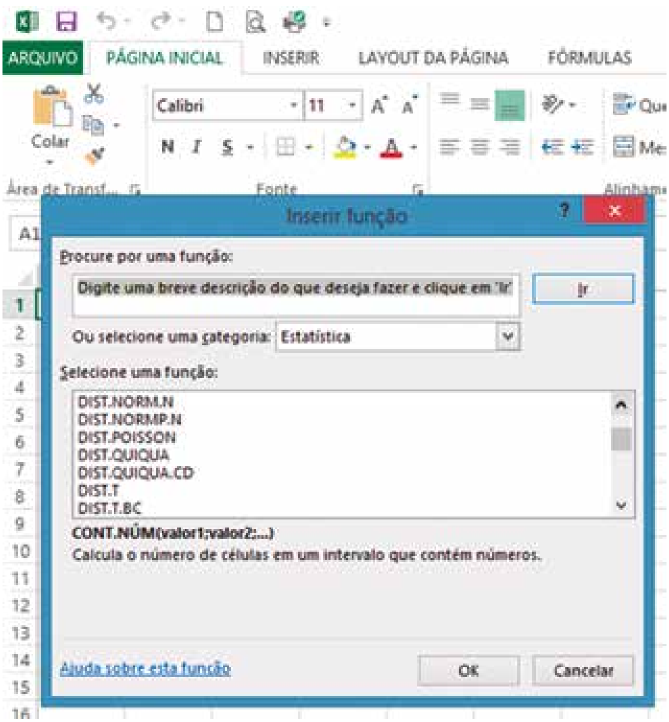
9.2.3.1 Qual é a diferença entre as distribuições de Poisson e Binomial?

Enquanto a distribuição binomial pode ser usada para encontrar a probabilidade de um número designado de sucessos em n tentativas, a distribuição de Poisson é usada para encontrar a probabilidade de um número designado de sucessos por unidade de intervalo. As outras condições exigidas para se aplicar a distribuição Binomial são também exigidas para se aplicar a distribuição de Poisson; isto é, devem existir somente dois resultados mutuamente exclusivos, os eventos devem ser independentes, e o número médio de sucessos por unidade de intervalo deve permanecer constante.

9.2.3.2 Alguns exemplos de quando podemos aplicar a distribuição de Poisson

A distribuição de Poisson é frequentemente usada em pesquisa operacional na solução de problemas administrativos. Alguns exemplos são o número de chamadas telefônicas para a polícia por hora, o número de clientes chegando a uma bomba de gasolina por hora, e o número de acidentes de tráfego num cruzamento por semana.

Exemplo no Excel: No Excel, podemos calcular desvio médio de conjunto de números utilizando a função "DIST.POISSON" (clicar em função, depois escolher a categoria "Estatística" e, em seguida, escolher "DIST.POISSON", conforme apresentado abaixo).



Exemplos:

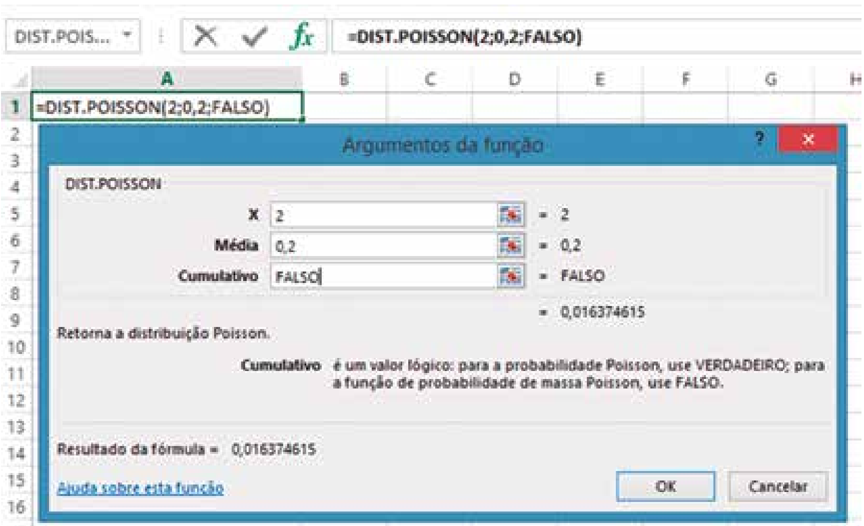
1. Considere um processo que tem uma taxa de 0,2 defeitos por unidade. Qual a probabilidade de uma unidade qualquer apresentar:

Neste caso temos que μ = 0,2

a. dois defeitos?

$$P(X=2)=\frac{\mu^{X}\times e^{-\mu}}{X!}=\frac{0,2^{2}\times e^{-0,2}}{2!}=0,0164$$

No Excel



b. um defeito?

$$P(X=1)=\frac{\mu^{X}\times e^{-\mu}}{X!}=\frac{0,2^{1}\times e^{-0,2}}{1!}=0,1637$$

No Excel, você pode escrever a seguinte sintaxe dentro da célula: "=DIST.POISSON(1;0,2;FALSO)". Para calcular a probabilidade corretamente no Excel, você sempre terá que escrever "FALSO" no campo "Cumulativo". A sintaxe para escrever diretamente na célula é "=DIST.POISSON(X;μ;FALSO)", no caso do exemplo, X =1 e μ = 0,2.

c. zero defeito?

$$P(X=0)=\frac{\mu^{X}\times e^{-\mu}}{X!}=\frac{0,2^{0}\times e^{-0,2}}{0!}=0,8187$$

No Excel, você pode escrever a seguinte sintaxe dentro da célula: "=DIST.POISSON(0;0,2;FALSO)".

2. Suponha que uma aplicação de tinta em um automóvel é feita de forma mecânica, e pode produzir defeitos de fabricação, como bolhas ou áreas mal pintadas, de acordo com uma variável aleatória X que segue uma distribuição de Poisson de parâmetro μ = 1. Suponha que sorteamos um carro ao acaso para que sua pintura seja inspecionada, qual a probabilidade de:

   a. Encontrarmos 1 defeito?

   $$P(X=1)=\frac{\mu^X \times e^{-\mu}}{X!}=\frac{1^1 \times e^{-1}}{1!}=0,3679$$

   No Excel você pode escrever "=DIST.POISSON(1;1;FALSO)"

   b. Encontrarmos, pelo menos, 1 defeito?

   Pelo menos 1 defeito é igual a 1 defeito ou mais, então:

   P(X≥1)=P(X=1)+P(X=2)+P(X=3)+ ⋯ =1-P(X<1)

   P(X≥1)=1-P(X<1)=1-P(X=0)

   $$P(X\geq 1)=1-P(X=0)=1-\frac{1^0 \times e^{-1}}{0!}=1-0,3679=0,6321$$

   No Excel você pode escrever "=1-DIST.POISSON(0;1;FALSO)"

   c. encontrarmos de 2 a 4 defeitos?

   Encontrar de 2 a 4 defeitos corresponde a encontrar 2 ou 3 ou 4 defeitos, então:

   P(2 ≤ X ≤ 4) = P(X = 2) + P(X = 3) + P(X = 4)

   $$P(2\leq X\leq 4)=\frac{1^2 \times e^{-1}}{2!}+\frac{1^3 \times e^{-1}}{3!}+\frac{1^4 \times e^{-4}}{4!}=0,1839+0,0613+0,0153$$

   $$P(2\leq X\leq 4)=0,2605$$

$P(2 \leq X \leq 4) = 0,2605$

No Excel você pode escrever "=DIST.POISSON(2;1;FALSO) + DIST.POISSON(3;1;FALSO) + DIST.POISSON(4;1;FALSO)"

3. Uma central telefônica tipo PABX recebe uma média de 5 chamadas por minuto. Qual a probabilidade deste PABX não receber nenhuma chamada durante um intervalo de 1 minuto?

$$P(X=0)=\frac{\mu^{X}\times e^{-\mu}}{X!}=\frac{5^{0}\times e^{-5}}{0!}=0,0067$$

No Excel você pode escrever "=DIST.POISSON(0;5;FALSO)"

4. Um departamento de conserto de máquinas recebe uma média de cinco chamadas por hora. Qual a probabilidade de que em uma hora selecionada aleatoriamente, sejam recebidas exatamente três chamadas?

$$P(X=3)=\frac{\mu^{X}\times e^{-\mu}}{X!}=\frac{5^{3}\times e^{-5}}{3!}=0,1404$$

No Excel você pode escrever "=DIST.POISSON(3;5;FALSO)"

5. Em média, quatro pessoas utilizam, por hora, um determinado telefone público. O mesmo, apresentando defeito, será reparado após uma hora de comunicação do defeito. Qual a probabilidade de que somente o reclamante não possa usar o telefone devido ao defeito?

$$P(X=1)=\frac{\mu^{X}\times e^{-\mu}}{X!}=\frac{4^{1}\times e^{-4}}{1!}=0,0733$$

No Excel você pode escrever "=DIST.POISSON(1;4;FALSO)"

6. Em média, um digitador comete 3 erros a cada 1.000 números teclados. Qual a probabilidade de que, na digitação de um importante relatório, composto por 2.000 números, não ocorram erros?

Como são cometidos, em média, 3 erros a cada 1.000 números teclados, para cada 2.000 números são esperados 6 erros, em média. Então, $\mu = 6$ para 2.000 erros.

$$P(X=0)=\frac{\mu^X \times e^{-\mu}}{X!}=\frac{6^0 \times e^{-6}}{0!}=0{,}0025$$

No Excel você pode escrever "=DIST.POISSON(0;6;FALSO)"

7. Em cada dez dias chegam, em média, trinta navios à determinada doca. Qual a probabilidade de que, em um dia aleatoriamente escolhido, cheguem à doca:

a. Exatamente 4 navios?

Como em 10 dias chegam, em média, 30 navios, em um dia espera-se que cheguem 3 navios ($\mu = 3$).

$$P(X=4)=\frac{\mu^X \times e^{-\mu}}{X!}=\frac{3^4 \times e^{-3}}{4!}=0{,}1680$$

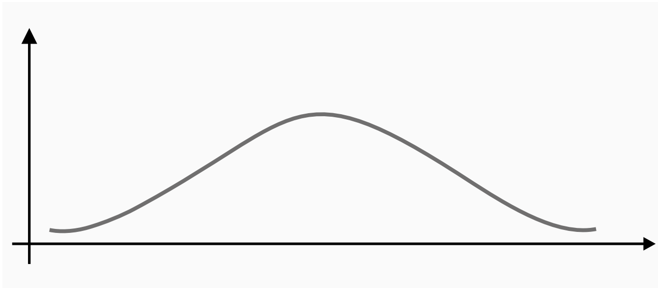
No Excel você pode escrever "=DIST.POISSON(4;3;FALSO)"

b. Mais 4 navios?

$$P(X>4)=P(X=5)+P(X=6)+P(X=7)+\cdots=1-P(X\leq 4)$$
$$P(X>4)=1-P(X=0)-P(X=1)-P(X=2)-P(X=3)-P(X=4)$$
$$P(X>4)=1-\frac{3^0 \times e^{-3}}{0!}-\frac{3^1 \times e^{-3}}{1!}-\frac{3^2 \times e^{-3}}{2!}-\frac{3^3 \times e^{-3}}{3!}-\frac{3^4 \times e^{-3}}{4!}$$
$$0{,}8152$$

9.3 Distribuição Normal

A distribuição normal de probabilidade é uma distribuição de probabilidade contínua que é simétrica e mesocúrtica. A curva que representa a distribuição normal de probabilidade é frequentemente descrita como tendo uma forma de "sino" é, também, conhecida como Curva de Gauss, ou Curva Gausssiana. Veja figura abaixo:



A distribuição normal é, sem dúvida, a mais importante distribuição contínua, pois diversos estudos práticos têm como resultado uma distribuição normal. Podemos citar como exemplo a altura de uma determinada população em geral segue uma distribuição normal. Entre outras características físicas e sociais tem um comportamento gaussiano, ou seja, segue uma distribuição normal.

A variação natural de muitos processos industriais é realmente aleatória. Embora as distribuições de muitos processos possam assumir uma variedade de formas, muitas variáveis observadas possuem uma distribuição de frequências que é, aproximadamente, uma distribuição de probabilidade Normal.

A distribuição de probabilidade normal é importante na inferência estatística por três razões distintas:

- As medidas produzidas em diversos processos aleatórios seguem esta distribuição.
- Probabilidades normais podem ser usadas frequentemente como aproximações de outras distribuições de probabilidade, tais como a binomial e a de poisson.
- As distribuições de estatísticas da amostra, tais como a média e a proporção, frequentemente seguem a distribuição normal independentemente da distribuição da população.

Uma variável aleatória contínua $ X $ tem distribuição Normal se sua função densidade de probabilidade for dada por:

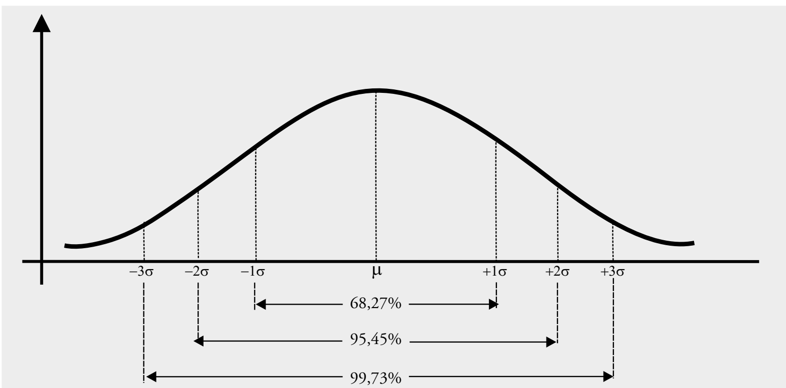
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$

Onde μ e σ são, respectivamente, a média e o desvio padrão.

Para cada valor de μ e/ou σ temos uma curva de distribuição de probabilidade. Porém, para se calcular áreas específicas, faz-se uso de uma distribuição particular: a "distribuição normal padronizada", também chamada de Standartizada ou reduzida, o qual é a distribuição normal com μ = 0 e σ = 1. Para obter tal distribuição, isto é, quando se tem uma variável ***X*** com distribuição normal com média μ diferente de 0 (zero) e/ou desvio padrão σ diferente de 1 (um), devemos reduzi-la a uma variável **Z**, efetuando o seguinte cálculo:

$$Z = \frac{X - \mu}{\sigma}$$

Para achar a área sob a curva normal, devemos conhecer dois valores numéricos, a média μ e o desvio padrão σ. A Figura a seguir mostra algumas áreas importantes:



A figura acima exibe o gráfico da função densidade da distribuição normal padronizada. Nesse gráfico indicamos as áreas 1, 2 e 3 desvios padrões a contar da média, isto é, entre *z* = -1 e 1, *z* = -2 e 2, *z* =-3 e 3, iguais respectivamente, a 69,27%, 95,45% e 99,73% da área total, que é 100%. Isto significa que:

- $P(-1 \leq Z \leq 1) = 0{,}6825 = 68{,}25\%$
- $P(-2 \leq Z \leq 2) = 0{,}9545 = 95{,}45\%$
- $P(-3 \leq Z \leq 3) = 0{,}9973 = 99{,}73\%$

Assim, a distribuição passa a ter média $\mu = 0$ e desvio padrão $\sigma = 1$. Pelo fato da distribuição ser simétrica em relação à média $\mu = 0$, a área à direita é igual a área à esquerda de μ. Por ser uma distribuição muito usada, existem tabelas a qual encontramos a resolução de suas integrais. Assim, a "Tabela 2 – Área sob a Curva Normal Padronizada de 0 a Z" apresenta as áreas sob esta curva, limitadas pelas ordenadas em $z = 0$ e um qualquer valor positivo z. Por essa tabela, podemos determinar a área entre duas ordenadas quaisquer, valendo-nos da simetria da curva em relação a $z = 0$.

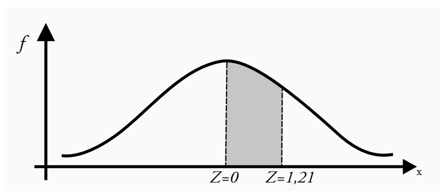
Exemplos:

Exemplo 1: Determine a área sob a curva normal padronizada conforme apresentado abaixo:

a. entre $Z = 0$ e $Z = 1{,}21$

Usando a Tabela-2, apresentada ao final deste capítulo, percorramos as linhas pintadas de cinza, até encontrar 1,21 e caminhamos nela até a coluna marcada por 0; o resultado será 0,3869, é a área procurada e representa a probabilidade de Z estar entre 0 e 1,2. A tabela sempre apresenta a área entre Z igual a zero até um determinado número. Então, na Tabela-2, reduzida abaixo, você deve procurar o número $Z = 1{,}21$, pois nas linhas estão apresentados os dois primeiros dígitos do número "1,2" e na coluna está o terceiro número 0,01.

Z	0,00	0,01	0,02
0,0	0,0000	0,0040	0,0080
...	...	...	...
1,0	0,3413	0,3438	0,3461
1,1	0,3643	0,3665	0,3686
1,2	0,3849	$\boxed{0{,}3869}$	0,3888
1,3	0,4032	0,4049	0,4066

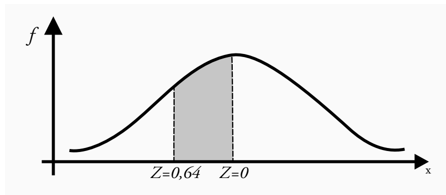


Então, $P(0 \leq Z \leq 1{,}21) = 0{,}3869 = 38{,}69\%$

b. Entre $Z = -0{,}64$ e $Z = 0$

Como área procurada entre "$Z = -0{,}64$ e $Z = 0$" é igual a área "$Z = 0$ e $Z = 0{,}64$" e a Tabela-2 só apresenta os valores positivos de Z, pois as áreas são simétricas. Percorrendo a coluna *Z* até a entrada 0,6 e, em seguida, caminhando à direita até a coluna 0,04, chega-se ao valor 0,2389

Z	0,02	0,03	0,04
0,0	0,0080	0,0120	0,0160
0,1	0,0478	0,0517	0,0557
0,2	0,0871	0,0910	0,0948
0,3	0,1255	0,1293	0,1331
0,4	0,1628	0,1664	0,1700
0,5	0,1985	0,2019	0,2054
0,6	0,2324	0,2357	$\boxed{0,2389}$
0,7	0,2642	0,2673	0,2704



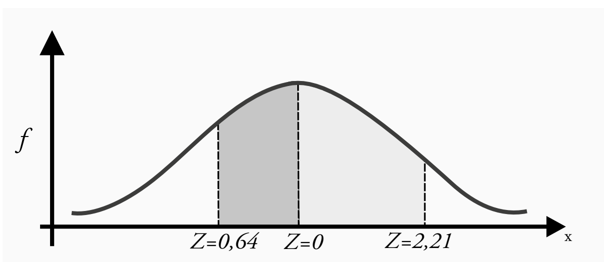
Então, $P(-0{,}64 \le Z \le 0) = 0{,}2389 = 23{,}89\%$

c. entre *Z* = -0,46 e *Z* = 2,21

Neste caso, temos que dividir o problema em duas etapas, pois a tabela apresenta a área entre zero e o *Z* desejado:

P(-0,46 ≤ Z ≤ 2,21)=P(-0,46 ≤ Z ≤ 0)+ P(0 ≤ Z ≤ 2,21), por simetria temos que:

P(-0,46 ≤ Z ≤ 2,21)= P(0 ≤ Z ≤ + 0,46) + P(0 ≤ Z ≤ 2,21)



1ª etapa: P(0 ≤ Z ≤ + 0,46)

Como área procurada entre *Z* = 0 e *Z* = 0,46". Percorrendo a coluna *Z* na Tabela-2 até a entrada 0,4 e, em seguida, caminhando à direita até a coluna 0,06, chega-se ao valor 0,1772.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879

Então,

2ª etapa: P(0 ≤ Z ≤ 2,21)

Como área procurada entre *Z* = 0 e *Z* = 2,21". Percorrendo a coluna *Z* na Tabela-2 até a entrada 2,2 e, em seguida, caminhando à direita até a coluna 0,01, chega-se ao valor 0,4864.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936

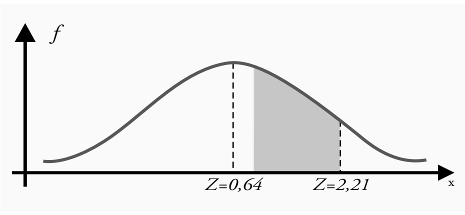
Então, P(0 ≤ Z ≤ 2,21)= 0,4864

Como: P(- 0,46 ≤ Z ≤ 2,21)= P(0 ≤ Z ≤ + 0,46)+P(0 ≤ Z ≤ 2,21)= 0,1772 + 0,4864

P(-0,46 ≤ Z ≤ 2 ,21)= 0,6363

d. entre *Z* = 0,46 e *Z* = 2,21

Neste caso, temos que dividir o problema em duas etapas, pois a tabela apresenta a área entre zero e o *Z* desejado:



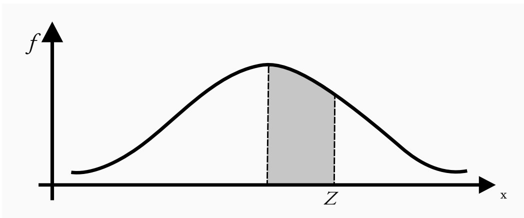
P(0,46 ≤ Z ≤ 2,21)= P(0 ≤ Z ≤ 2,21) - P(0 ≤ Z ≤ + 0,46)

Como os calores foram calculados no exemplo "c" acima, temos que:

$P(0{,}46 \leq Z \leq 2{,}21) = 0{,}4864 - 0{,}1772 = 0{,}3092 = 30{,}92\%$

Exemplo 2: Se a expressão "área" se refere à área sob a curva normal padronizada, determine o valor ou os valores de *Z* tais que:

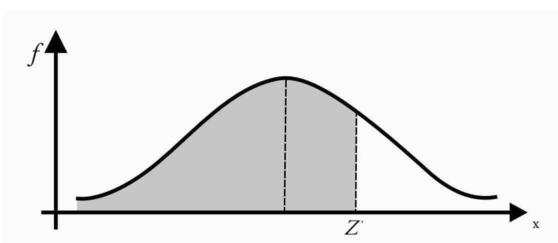
a. área entre 0 e 0,3770



Na Tabela-2, a entrada 0,3770 está localizada à direita da linha marcada 1,1 e sob a coluna 0,06. Então, *Z* = 1,16. Por simetria, *Z* também pode ser -1,16, portanto ***Z* = ±1,16**.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	$\boxed{0{,}3770}$	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319

b. área a esquerda de 0,8621

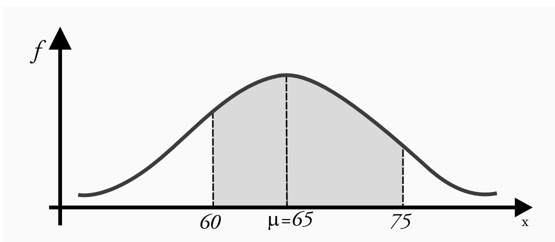


Como a área é maior do que 0,5, *Z* deve ser positivo, conforme apresentado acima (cada metade da curva acima tem área de 0,5). Então, nós temos duas áreas possíveis a analisar: a primeira até a *Z* = 0 com área de 0,5 e outra área entre *Z* = 0 e *Z'*, com área de 0,3621 (0,8621 – 0,5). Para uma área de 0,3621 temos um *Z'* = 1,09 (conforme pode ser confirmado a partir da Tabela-2, apresentada abaixo).

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015

Exemplo 3: O peso médio de 500 estudantes de sexo masculino de certo colégio é 65 kg e o desvio padrão de 6 kg. Supondo os pesos distribuídos normalmente, determine quantos estudantes pesarão:

a. entre 60 e 75 kg



Para resolvermos este problema, primeiramente devemos padronizá-lo, através da equação abaixo:

$$Z = \frac{X - \mu}{\sigma} \text{ ,então:}$$

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{60 - 65}{6} = -0,83333 = -0,83$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{75 - 65}{6} = 1,6666 = 1,67$$

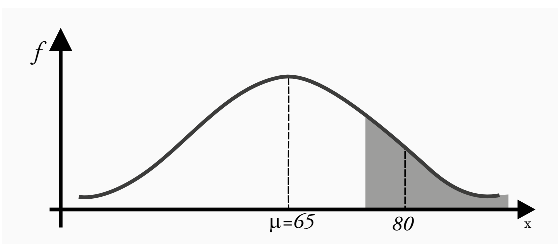
Temos agora que calcular a área da curva:

$P(-0{,}83 \leq Z \leq 1{,}67) = P(0 \leq Z \leq 0{,}83) + P(0 \leq Z \leq 1{,}67) = 0{,}2967 + 0{,}4524$

$P(-0{,}83 \leq Z \leq 1{,}67) = 0{,}7492$

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	$\boxed{0{,}2967}$	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	$\boxed{0{,}4525}$	0,4535	0,4545

a. mais que 80 kg



Para resolver este problema, primeiro precisamos transformar o gráfico acima em entendimento matemático.

$P(X \geq 80) = P(X \geq 65) - P(0 \leq X \leq 80)$

Por definição, sabemos que a probabilidade de qualquer evento ocorrer maior do que a média é igual a 50%; o mesmo podemos dizer para um evento ocorrer menor do que a média, então:

$P(X \geq 65) = 0{,}5 = 50\%$

Precisamos agora padronizar o 80 kg:

$$Z = \frac{X - \mu}{\sigma} = \frac{80 - 65}{6} = 2,50$$

Temos agora que calcular a área da curva:

P(Z ≥ 2,50)= P(Z ≥ 0)-P(0 ≤ Z ≤ 2,50)=0,5 - 0,4938=0,0062=0,62%

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964

Atividades

1. Em 10 lançamentos, qual a probabilidade de obter:
   a. exatamente 3 caras?
   b. pelo menos 3 coroas?
   c. Qual o número mais provável de caras? Com que desvio padrão?

2. Se em novembro chove 4 dias, em média, calcule a probabilidades de:
   a. chover pelo menos 5 dias de novembro;
   b. não chover em novembro.
   c. Qual a probabilidade de chover num dia qualquer de novembro?
   d. Qual o desvio padrão desta distribuição?

3. Determine a probabilidade de, em três jogadas de uma moeda, aparecer:
   a. 3 caras
   b. 2 coroas e 1 cara

c. ao menos 1 cara

d. no máximo 1 coroa

4. Se a média numa distribuição de Poisson é 16,7, calcule a probabilidade de:

a. aparecer o valor 8

b. aparecer valores menores que 5

c. aparecer valores maiores ou igual a 5

d. Qual o desvio padrão desta distribuição?

5. Em média, chegam 30 clientes por hora numa fila com distribuição de Poisson. Qual a probabilidade de, na próxima hora,

a. chegar 20 clientes?

b. chegar mais de 9 clientes?

6. Se a probabilidade de um indivíduo acusar reação negativa à injeção de determinado soro é 0,001, determine a probabilidade de que, em 2.000 indivíduos:

a. exatamente 3 indivíduos tenham reação negativa

b. mais de 2 indivíduos tenham reação negativa

7. A altura média de certa população masculina adulta é de 1,70m, com desvio padrão de 10cm. Calcule a porcentagem de homens com altura:

a. entre 1,60m e 1,75m;

b. maior que 1,70m;

c. menor que 1,50m.

8. Suponha que a renda média de uma grande comunidade possa ser razoavelmente aproximada por uma distribuição normal com média de R$1.500,00 e desvio padrão de R$300,00, calcule:

a. Qual a porcentagem da população que terá renda superior a R$ 1.860,00?

b. Qual a porcentagem da população que terá renda inferior a R$1.000,00?

c. Qual a porcentagem da população que terá renda entre R$1.050,00 e R$ 1.780?

9. Analisando os resultados das avaliações dos 2.000 alunos de certa escola, verificou-se que a distribuição das notas tem uma distribuição aproximadamente normal com média igual a 6 e desvio padrão igual a 1. Quantos alunos podemos esperar que tenham tirado nota:

a. inferior a 5?

b. superior a 7,5?

c. entre 6,5 e 8,5?

Respostas

1. Binomial; a) $P(X = 3) = 0{,}1172$; b) $P(X \geq 3) = 0{,}9453$; c) $\mu = 5$, $\sigma = \pm 1{,}58$

2. Binomial; a) $P(X \geq 5) = 0{,}3708$; b) $P(X = 0) = 0{,}01366$; c) $P(X = 1) = 0{,}06306$; d) $\sigma = \pm 1{,}86$

3. Binomial; a) $P(X = 3$ cara$) = 1/8$; b) $P(X = 2$ coroa$) = 3/8$; c) $P(X \geq 1$ cara$) = 7/8$; d) $P(X \leq 1$ coroa$) = 1/2$

4. Poisson; a) $P(X = 8) = 0{,}008385$; b) $P(X < 5) =0{,}0002333$; c) $P(X \geq 5) = 0{,}999766$; d) $\sigma = \pm 4{,}08$

5. Poisson; a) $P(X = 20) = 0{,}01341$; b) $P(X > 9) = 0{,}999992878$

6. Poisson; a) $P(X = 3) = 0{,}18$; b) $P(X > 2) = 0{,}323$

7. Normal; a) $P(1{,}60 < X < 1{,}75) = 53{,}28\%$; b) $P(X > 1{,}70) = 50\%$; c) $P(X < 1{,}50)= 2{,}275\%$

8. Normal; a) $P(X > 1.860) = 11{,}51\%$; b) $P(X < 1.000) = 4{,}75\%$; c) $P(1.050 < X < 1.780) = 75{,}7\%$

9. Normal; a) $P(X < 5) = 317{,}4$ alunos $(15{,}87\%)$; b) $P(X > 7{,}5) = 134$ alunos $(6{,}68\%)$; c) $P(6{,}5 < X < 8{,}5) =605$ alunos $(30{,}23\%)$

Tabela 1 – Valores de $e^{-\mu}$ para alguns valores de μ.

μ	$e^{-\mu}$	μ	$e^{-\mu}$	μ	$e^{-\mu}$	μ	$e^{-\mu}$
0,0	1,00000	2,5	0,08208	5,0	0,00674	7,5	0,00055
0,1	0,90484	2,6	0,07427	5,1	0,00610	7,6	0,00050
0,2	0,81873	2,7	0,06721	5,2	0,00552	7,7	0,00045
0,3	0,74082	2,8	0,06081	5,3	0,00499	7,8	0,00041
0,4	0,67032	2,9	0,05502	5,4	0,00452	7,9	0,00037
0,5	0,60653	3,0	0,04979	5,5	0,00409	8,0	0,00034
0,6	0,54881	3,1	0,04505	5,6	0,00370	8,1	0,00030
0,7	0,49659	3,2	0,04076	5,7	0,00335	8,2	0,00027
0,8	0,44933	3,3	0,03688	5,8	0,00303	8,3	0,00025
0,9	0,40657	3,4	0,03337	5,9	0,00274	8,4	0,00022
1,0	0,36788	3,5	0,03020	6,0	0,00248	8,5	0,00020
1,1	0,33287	3,6	0,02732	6,1	0,00224	8,6	0,00018
1,2	0,30119	3,7	0,02472	6,2	0,00203	8,7	0,00017
1,3	0,27253	3,8	0,02237	6,3	0,00184	8,8	0,00015
1,4	0,24660	3,9	0,02024	6,4	0,00166	8,9	0,00014
1,5	0,22313	4,0	0,01832	6,5	0,00150	9,0	0,00012
1,6	0,20190	4,1	0,01657	6,6	0,00136	9,1	0,00011
1,7	0,18268	4,2	0,01500	6,7	0,00123	9,2	0,00010
1,8	0,16530	4,3	0,01357	6,8	0,00111	9,3	0,00009
1,9	0,14957	4,4	0,01228	6,9	0,00101	9,4	0,00008
2,0	0,13534	4,5	0,01111	7,0	0,00091	9,5	0,00007
2,1	0,12246	4,6	0,01005	7,1	0,00083	9,6	0,00007

2,2	0,11080	4,7	0,00910	7,2	0,00075	9,7	0,00006
2,3	0,10026	4,8	0,00823	7,3	0,00068	9,8	0,00006
2,4	0,09072	4,9	0,00745	7,4	0,00061	9,9	0,00005

Tabela 2 – Área sob a Curva Normal Padronizada de 0 a *Z*

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Referências

BUSSAB, Wilton De Oliveira. Estatística básica. São Paulo: Editora Saraiva, 2013.

CASTANHEIRA, Nelson Pereira. Estatística aplicada a todos os níveis.

2ª Ed. Curitiba: IBEPEX, 2005.

COSTA NETO, Pedro Luíz de Oliveira. Estatística e Probabilidades.

2ª ed. São Paulo: Editora Blucher, 2002.

DOWNING, Douglas. Estatística aplicada. 2. Ed. São Paulo: Saraiva, 2002.

LEVIN, Jack. Estatística para ciências humanas. 11ª Ed. São Paulo: Pearson Education do Brasil, 2012

SPIEGEL, Murray Rakph. Probabilidade e Estatística. São Paulo: McGraw--Hill do Brasil, 1978.

STEVENSON, William J. Estatística aplicada à administração. São Paulo: Harper & Row do Brasil, 1981.

TOLEDO, Geraldo Luciano; Ovalle, Ivo Izidoro. Estatística Básica. 2ª ed. São Paulo: Atlas, 1985

A utilização da Estatística é cada vez mais acentuada em qualquer atividade profissional da vida moderna. Nos seus mais diversificados ramos de atuação, as pessoas estão frequentemente expostas à Estatística, utilizando-a com maior ou menor intensidade. Isto se deve às múltiplas aplicações que o método estatístico proporciona àqueles que dele necessitam.

Há, evidentemente, a necessidade de especiais cuidados no manejo e na interpretação da Estatística; a interpretação não é monopólio dos estatísticos, sendo natural que, possuindo um maior conhecimento das técnicas estatísticas, levem vantagens no tocante à apreciação, análise e interpretação dos dados estatísticos. O raciocínio claro é indispensável para interpretar estatísticas, requerendo uma disposição mental receptiva e crítica.

Raramente, ou nunca, os dados estatísticos falam por si mesmos. A coisa mais importante acerca da interpretação dos dados estatísticos é saber que, se forem habilmente coletados e criticamente analisados, podem ser extremamente úteis.

Há muito tempo sinto a necessidade de um livro-texto claro e compreensível sobre estatística. O material pretende ser de fácil leitura, compreensão e assimilação. Além disso, deve ser interessante, e não maçante, contendo numerosos exemplos e aplicações.

O livro destina-se a ser usado como livro-texto para um curso formal em estatística e probabilidade ou como ferramenta de consulta para alunos e/ou profissionais que desejam aprender a utilizar não só as ferramentas da estatística, mas também utilizá-las no Excel.

Bem-vindo à estatística! Você está iniciando uma interessante e significativa aventura, pois começa a explorar um dos mais básicos instrumentos da tomada de decisão.

ISBN 978-85-53370-25-2