Estatística

Marcos Antonio Barbosa

INSTITUTO FEDERAL
PARANÁ
Educação a Distância

Curitiba-PR
2012

Presidência da República Federativa do Brasil

Ministério da Educação

Secretaria de Educação a Distância

Este Caderno foi elaborado pelo Instituto Federal do Paraná para o Sistema Escola Técnica Aberta do Brasil – e-Tec Brasil.

Prof. Irineu Mario Colombo
Reitor

Profª. Mara Christina Vilas Boas
Chefe de Gabinete

Prof. Ezequiel Westphal
Pró-Reitoria de Ensino - PROENS

Prof. Gilmar José Ferreira dos Santos
Pró-Reitoria de Administração - PROAD

Prof. Paulo Tetuo Yamamoto
Pró-Reitoria de Extensão, Pesquisa e Inovação - PROEPI

Profª. Neide Alves
Pró-Reitoria de Gestão de Pessoas e Assuntos Estudantis - PROGEPE

Prof. Carlos Alberto de Ávila
Pró-reitoria de Planejamento e Desenvolvimento Institucional - PROPLADI

Prof. José Carlos Ciccarino
Diretor Geral de Educação a Distância

Prof. Ricardo Herrera
Diretor de Planejamento e Administração EaD - IFPR

Profª Mércia Freire Rocha Cordeiro Machado
Diretora de Ensino, Pesquisa e Extensão EaD - IFPR

Profª Cristina Maria Ayroza
Coordenadora Pedagógica de Educação a Distância

Prof. Adriano Stadler
Coordenador do Curso

Adriana Valore de Sousa Bello
Fábio Decker
Karmel Louise Pombo Schultz
Kátia Ferreira
Suelem Sousa Santana de Freitas
Assistência Pedagógica

Profª Ester dos Santos Oliveira
Profª Linda Abou Rejeili
Lídia Emi Ogura Fujikawa
Revisão Editorial

Flávia Terezinha Vianna da Silva
Diagramação

e-Tec/MEC
Projeto Gráfico

Catalogação na fonte pela Biblioteca do Instituto Federal do Paraná

Apresentação e-Tec Brasil

Prezado estudante,

Bem-vindo ao e-Tec Brasil!

Você faz parte de uma rede nacional pública de ensino, a Escola Técnica Aberta do Brasil, instituída pelo Decreto nº 6.301, de 12 de dezembro 2007, com o objetivo de democratizar o acesso ao ensino técnico público, na modalidade a distância. O programa é resultado de uma parceria entre o Ministério da Educação, por meio das Secretarias de Educação a Distância (SEED) e de Educação Profissional e Tecnológica (SETEC), as universidades e escolas técnicas estaduais e federais.

A educação a distância no nosso país, de dimensões continentais e grande diversidade regional e cultural, longe de distanciar, aproxima as pessoas ao garantir acesso à educação de qualidade, e promover o fortalecimento da formação de jovens moradores de regiões distantes, geograficamente ou economicamente, dos grandes centros.

O e-Tec Brasil leva os cursos técnicos a locais distantes das instituições de ensino e para a periferia das grandes cidades, incentivando os jovens a concluir o ensino médio. Os cursos são ofertados pelas instituições públicas de ensino e o atendimento ao estudante é realizado em escolas-polo integrantes das redes públicas municipais e estaduais.

O Ministério da Educação, as instituições públicas de ensino técnico, seus servidores técnicos e professores acreditam que uma educação profissional qualificada – integradora do ensino médio e educação técnica, – é capaz de promover o cidadão com capacidades para produzir, mas também com autonomia diante das diferentes dimensões da realidade: cultural, social, familiar, esportiva, política e ética.

Nós acreditamos em você!

Desejamos sucesso na sua formação profissional!

Ministério da Educação
Janeiro de 2010

Nosso contato
etecbrasil@mec.gov.br

Indicação de ícones

Os ícones são elementos gráficos utilizados para ampliar as formas de linguagem e facilitar a organização e a leitura hipertextual.

Atenção: indica pontos de maior relevância no texto.

Saiba mais: oferece novas informações que enriquecem o assunto ou "curiosidades" e notícias recentes relacionadas ao tema estudado.

Glossário: indica a definição de um termo, palavra ou expressão utilizada no texto.

Mídias integradas: sempre que se desejar que os estudantes desenvolvam atividades empregando diferentes mídias: vídeos, filmes, jornais, ambiente AVEA e outras.

Atividades de aprendizagem: apresenta atividades em diferentes níveis de aprendizagem para que o estudante possa realizá-las e conferir o seu domínio do tema estudado.

Sumário

Palavra do professor-autor

Este livro tem o objetivo de propor uma visão geral da estatística. Perceberá que esse material foi elaborado para que você se aproprie de mais informação, cujos conhecimentos serão fundamentais a sua prática profissional na Logística.

A linguagem adotada neste material foi criteriosamente determinada para facilitar sua compreensão; os assuntos abordados, as indicações de fontes de pesquisa, de leituras, de filmes, de *sites*, etc. foram cuidadosamente selecionados com o intuito de estabelecer um elo com a sua experiência de vida e os seus conhecimentos já adquiridos. Disponibilizamos, ainda, ao final de cada aula, exercícios para que você possa fixar o conteúdo estudado.

Por fim, lembre-se que sua prática está no exercício pleno de aplicar os conhecimentos adquiridos aqui, fazendo a diferença na sua postura e na sua prática profissional.

Bons estudos!

Prof. Marcos Barbosa

Aula 1 – Estatística aplicada à logística

Nesta aula apresentaremos a você um breve histórico da evolução da Estatística. Identificaremos os conceitos e fundamentos que são a base para o estudo da estatística.

1.1 Uma visão geral

A evolução das ciências tem suas raízes na história do homem. Com a Estatística, que entendemos como um ramo da Matemática aplicada, não é diferente.

Figura 1.1: Análise estatística
Fonte: www.decifrando.com

Desde as civilizações mais antigas, era comum que vários povos registrassem informações, através de rabiscos, de ossos e de pinturas, em cavernas, em formatos de desenhos. Com o passar do tempo, esses registros foram sendo cada vez mais utilizados, melhorados, para que pudessem responder muitas questões sociais, de natureza científica de controle e epistemológica.

Assim, esses registros passaram a ter uma importância circunstancial na evolução da humanidade, auxiliando na tomada de decisões individuais ou coletivas, como: controlar o número de habitantes de uma região, de natalidade, de óbitos, de acidentes, de produtos produzidos, estimativas de riqueza individual e social, de distribuição de uma classe ou comunidade, no modo de como distribuíam equitativamente terras aos povos, como cobravam impostos e como distribuíam suas rendas, que resumindo, hoje chamaríamos de processos de "estatísticas". (CRESPO, 2002)

Já na Idade Média, essas informações eram colhidas, como na maioria das vezes, para finalidades de ordem tributárias ou de ordem bélicas.

Somente a partir do século XVI, segundo Crespo (2002), começa a surgir às primeiras **análises sistemáticas** de fatos sociais: tais como batizados, casamentos e funerais originando as primeiras tábuas e tabelas.

A-Z

Análises sistemáticas
Entende-se por análises sistemáticas – como a condição do que pode ser analisado em um conjunto de elementos classificados e organizados entre si segundo um ou mais critérios.

Também para Junior, (2006), é no século XVI, que a estatística passa a ter um caráter mais científico, como pode observar em sua fala:

Inicialmente, no século XVI, pensada pelos ingleses como uma ciência política, destinava-se a descrever características de um país, tais como população, área, riquezas e recursos naturais. Deste papel histórico, origina-se a sua função de caracterização numérica de uma série de informações populacionais. Com esta abordagem, o termo é utilizado no plural, como as "estatísticas de saúde", as "estatísticas de mortalidade", as "estatísticas do registro civil", entre outras.

No século XVIII, auge do período iluminista, onde se entende a razão como a fonte de todo o conhecimento, e que teve como grande marco a Revolução Industrial. Época dos pensadores, como Descartes, Newton, Locke, Voltaire, Montesquieu, Rousseau, D'Alembert, entre outros, que buscavam uma explicação racional para o mundo e acreditavam por melhorias das condições existenciais do homem. Época da consolidação de descobertas como o cálculo integral, de Newton e Leibniz, tais fatos passaram a adquirir uma feição verdadeiramente científica.

Quem sabe, esse não foi o nascimento da Estatística?

Segundo contextos históricos, foi o economista alemão Godofredo Achenwall em seus estudos e registros que inicia a nova ciência, que seria conhecida como **Estatística.** A Estatística deixa de ser simples catalogação de dados numéricos coletivos para se tornar o **estudo de como chegar a conclusões sobre o todo** (população) **partindo de observações de parte deste todo** (amostras).

Figura 1.2: Godofredo Achenwall
Fonte: www.dipity.comc

Você sabia?

Godofredo Achenwall é considerado, para os alemães, o pai da Estatística Moderna.

1.2 Definição

Para muitos autores, podemos definir um conhecimento ou objeto através da observação que fazemos. Assim é comum termos diferentes olhares quando se trata de uma definição, porém essas diferenças são apenas de ordem contextual. Vejamos algumas definições.

Para Crespo (2002), a **Estatística** é uma parte da Matemática aplicada que fornece métodos para coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões.

Já para o dicionário Houaiss (2005), estatística é um ramo da matemática que trata da coleta, da análise, da interpretação e da apresentação de massas de dados numéricos.

Como professor eu prefiro entender como uma metodologia (método/modo/caminho) para realizar a coleta, classificação, apresentação, análise e interpretação de dados para a tomada de decisões, já que a estatística se tornou uma ciência devido a necessidade de registrar dados. Ou de modo bem simplista, entender a **Estatística como um ramo da Matemática aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados coletados, visando à utilização dos mesmos na tomada de decisões.**

Parece meio confuso termos várias definições, não é mesmo?

Porém, a definição é algo particular, é como descrevemos o modo de entendimento. Sendo assim cada indivíduo pode ter um olhar diferente sobre um objeto em análise.

Bom, fiquem tranquilos que iremos estudar o assunto com mais propriedade nas próximas aulas.

Outra questão importante é distinguir dados de informações. Entendemos dados como algo real. Pode ser imagem, som, texto, letras, números, ou seja, fatos em sua forma primária ou bruta. Já a informação é um conjunto de dados (bruta) que tem algum significado para quem observa ou para o detentor dentro de um determinado contexto.

Podemos dividir a estatística em duas partes, e de modo geral, entendê-las separadamente: uma parte se preocupa com a coleta, organização e descrição dos dados, e que chamamos de **Estatística Descritiva**; já a outra parte, trabalha com a análise e interpretação desses dados, que chamamos de **Estatística Indutiva** ou também conhecida como **Inferencial**.

A-Z

Estatística Indutiva ou Inferencial
É a parte da estatística que se baseia nos resultados obtidos pela análise das amostras, procurando comportamentos, leis, que induzam ou infiram em toda a população. São as chamadas generalizações da amostra para a população em estudo.

Em geral, as pessoas quando se referem ao termo **estatística** pensam no cálculo matemático; mas na verdade esquecem que ela tem como finalidade encontrar leis de comportamentos para todo um conjunto de dados. Isso se afirma em Crespo (2002) que descreve: (...) **o aspecto essencial da Estatística é o de proporcionar métodos inferenciais, que permitam conclusões que transcendam os dados obtidos inicialmente**.

Assim, a análise e interpretação dos dados estatísticos tornam possível diagnosticar comportamentos, problemas de processos, tanto em uma empresa como a um conjunto de elementos com características comuns, uma tribo indígena, por exemplo.

1.3 A Estatística na logística

Figura 1.3: O mundo logístico
Fonte: ©Franck Boston/Shutterstock

Atualmente, a estatística é usada na logística em seus vários setores, pois não se pode tomar decisões com base no "achismo" (qualitativa) ou decisões com mera suposição instintiva. Num mundo tão competitivo, dinâmico e rápido, as decisões devem ser tomadas, quando possível, sempre em cima de fatos, de dados ou de elementos que justifiquem essas decisões.

Você já percebeu que na maioria das vezes esses fatos, são valores numéricos que podem ser organizados e operados através do conhecimento específico? Por exemplo, em estoques utilizamos a média de saída, entrada, desvios e cálculos de lotes econômicos de compra, assim como de segurança. Já em cadeias de suprimentos, podemos levantar hipóteses através de cálculos probabilísticos para um melhor gerenciamento. Se pensarmos em controle e/ou qualidade temos várias aplicações de controle estatístico de qualidade.

Toda essa aplicação prática se dá por meio de **sondagem**, através de amostras, onde as informações vão sendo **coletadas.**

1.4. Algumas aplicações da estatística em outras áreas

Além da aplicação em logística, temos outras áreas que utilizam bastante à estatística. A princípio, citaremos apenas algumas, pois o uso se dá em quase todas as esferas das ciências. Vejamos:

- **Contabilidade**: análise e avaliação de auditorias.
- **Finanças**: simulações de investimento.
- **Marketing**: pesquisa de mercado, de preço, de praça, de produtos, de aceitação.

Pesquise no *site* do IBGE – http://www.ibge.gov.br –, dados nacionais para obter as normas técnicas para apresentação tabular de estatística brasileira, junto ao IBGE.

- **Produção**: controle de qualidade.
- **Economia**: "previsões" da economia.
- **Matemática**: provisões, comportamentos e leituras dos padrões com os números.

Resumo

Breve relato histórico da evolução da estatística. Vimos algumas vertentes que definem a estatística ora como ciência, ora como um instrumento matemático para auxiliar na tomada de decisão, e por fim direcionamos o uso da estatística no campo logístico e sua atuação nas demais áreas do conhecimento.

Atividades de aprendizagem

1. Conforme o que foi estudado nesta aula, defina o que você entendeu por estatística.

2. Como era feito na antiguidade o controle das informações? Pesquise e discuta com seus colegas.

3. Faça uma pesquisa em arquivos eletrônicos (internet) ou em bibliografias impressas (livro) para determinar diferentes definições da "Estatística". Depois discuta no fórum com os colegas se as definições encontradas apresentam coerência.

Aula 2 – Método experimental x método estatístico

Nesta aula definiremos o que é método, suas diferenças e as fases que compõem um método estatístico.

2.1 Introdução

Sistematicamente o conhecimento sempre resultou da observação e do estudo. A **epistemologia** explica isso, embora, sabemos que muito desse conhecimento, já foi descoberto por acaso. Hoje em dia, a verdade é estabelecida por **processos científicos,** testados sistematicamente, ou seja, analisados com muito cuidado.

A-Z

Epistemologia ou teoria do conhecimento
É um ramo da filosofia que trata dos problemas filosóficos relacionados com a crença e o conhecimento. É o estudo científico da ciência (conhecimento), sua natureza e suas limitações. A epistemologia estuda a origem, a estrutura, os métodos e a validade do conhecimento, motivo pelo qual também é conhecida como teoria do conhecimento.
Fonte: Dicionário eletrônico Houaiss

Para Macedo e Castanheira (2008), atualmente**,** qualquer acréscimo de conhecimento é fruto da observação e do estudo do fenômeno que o caracteriza. A partir desse estudo são desenvolvidos processos científicos que resultarão em conclusões sobre o fato que está sendo estudado.

Sendo assim, o modo sistemático é explicado por uma metodologia, que para você, é entendido como caminho e/ou processo, modo de fazer alguma coisa (pesquisa, descoberta, trabalho) através de regras bem definidas.

Na visão de Crespo (2002), o **método** é um conjunto de meios dispostos convenientemente para se chegar a um fim. Sendo assim, vários são os métodos existentes para validar um conhecimento. Já para Filho (2006), método pode ser definido como um conjunto de meios, processos e instrumentos usados pelos cientistas e estudiosos para formularem seus princípios, teorias e normas.

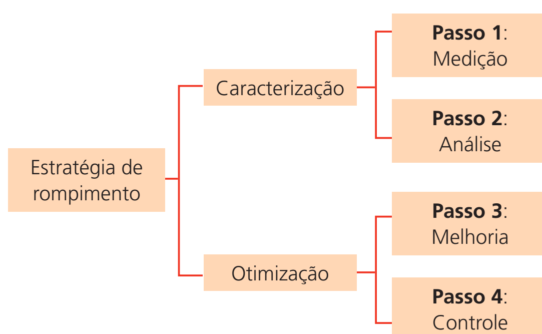
Com base nessas definições, podemos resumir que para cada processo, existe um método. Sendo assim, temos:

- **Método científico**: centraliza suas atenções nas observações;
- **Método experimental**: centraliza na realização e nas respostas das experiências;
- **Método estatístico**: se prende no relacionamento dos fatos de ordem quantificáveis, isto é, que podem ser contados ou enumerados.

Como estamos iniciando nossos estudos conheceremos, primeiramente, os métodos experimentais e estatísticos. O **método experimental** consiste em manter constantes todas as causas (fatores), menos uma, e variar esta causa de modo que o pesquisador possa descobrir seus efeitos, caso existam. É um método muito aplicado no estudo da Física, da Química, na psicologia, etc., já foi também utilizado muito na medicina.

Já o **método estatístico** não mantém nenhuma causa especifica, ou seja, deixa todas essas causas variando, porém registra todas as variações resultantes delas, e procura determinar no resultado final, qual foi à influência de cada uma delas sobre o resultado final. É muito utilizada no campo das ciências sociais. Por exemplo, na determinação das causas que definem o preço de uma mercadoria. Vários são os fatores, tais como demanda e oferta, sazonalidade, gosto do produto em função do consumidor, etc.

O método passo-a-passo:



Estratégia de rompimento:
1. Selecione um problema.
2. Defina o padrão de desempenho.
3. Valide um sistema de medição.
4. Estabeleça a capacidade.
5. Defina os objetivos de desempenho.
6. Identifica as fontes de variação.
7. Relacione as causas potenciais.
8. Descubra a relação entre as variáveis.
9. Estabeleça as tolerâncias operacionais.
10. Valide um sistema de medição.
11. Estabeleça a capacidade.
12. Implemente os processos de controle.

Figura 2.1: Métodos – DMAIC/6 SIGMA
Fonte: www.estatcamp.com.br

2.2 Fases do método estatístico

Vejamos cada uma das fases do método estatístico:

- **Definição do problema**

Determinação do objetivo da pesquisa, ou seja, definir com clareza o que se pretende pesquisar. Qual é a população ou amostra a ser pesquisada? Quais são as características mensuráveis (medidas) que se quer pesquisar?

- **Delimitação do problema**

É onde respondemos as perguntas: onde será realizada a pesquisa? Qual o local? Em que época? Com que objetos?

- **Planejamento**

Consiste na determinação do procedimento necessário para se fazer a pesquisa. A pesquisa se apropria um pouco da delimitação, embora seja muito mais abrangente, pois, como sabemos é preciso planejar o trabalho a ser realizado, tendo em vista o objetivo que se pretende atingir. Nesta fase, deve se levantar inúmeras informações sobre o objeto do estudo. Elencamos algumas perguntas mais comuns. Veja: Que dados deverão ser obtidos? Como obtê-los? O que será pesquisado? Quem participará da pesquisa? Em que setores geográficos serão feita a pesquisa? Qual o grau de precisão exigido na pesquisa? Qual o tipo de amostragem? Qual o tamanho da amostra? 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?

- **Coleta de dados**

Após a delimitação do problema, com a determinação das características (mensuráveis) do fenômeno que se quer pesquisar, é iniciada a **coleta de dados** numéricos. Essa coleta consiste na obtenção dos dados, seja por meio de questionário, observações, de entrevistas.

Existem várias ferramentas de coletas de dados, a saber: direta e indireta.

a) **Coleta direta**: informativos de registro obrigatório, tais como registros de nascimentos, de casamentos, prontuários médicos, de tabelas de controles ou dados obtidos pelo próprio pesquisador através de questionários.

b) **Coleta indireta**: quando é inferida de elementos conhecidos (coleta direta).

- **Crítica dos dados**

Obtidos os dados, eles devem ser cuidadosamente analisados, procurar possíveis falhas e imperfeições, a fim de não cometer erros grosseiros, que possam alterar sensivelmente os resultados.

Por exemplo, não podemos em uma pesquisa considerar um questionário que seja respondido pela metade. Temos que separar os dados por meios de contagem, por tipo, etc.

- **Apuração dos dados**

É o processamento dos dados obtidos, mediante critérios de classificação. Pode ser manual, eletromecânica ou eletrônica. Também conhecida como tabulação dos dados.

- **Exposição dos resultados**

Os dados são apresentados em forma de tabelas ou gráficos, com suas informações pertinentes de modo claro, objetivo e verídico da pesquisa, tornando mais fácil a observação daquilo que está sendo objeto de estudo, ou seja, do tratamento estatístico.

- **Análise dos resultados**

É o momento em que se tiram conclusões do todo (população) a partir de informações fornecidas por parte da amostra. Tais análises estão ligadas as medidas que permitem descrever o que está acontecendo na pesquisa com o grupo ou objeto de estudo. Fazemos aqui generalizações do comportamento desse grupo.

Resumo

Nesta aula vimos a definição da palavra método, a diferença entre métodos e suas respectivas fases.

Atividades de aprendizagem

1. Relacione as colunas abaixo:

a) Método	( ) Coleta direta
b) Fases do método estatístico	( ) Conjunto de meios para se chegar a um fim
c) Método experimental	( ) Gráficos
d) Exposição dos dados	( ) Realizações de experiências

2. Em que duas grandes áreas a ciência Estatística pode ser dividida? Descreva sucintamente do que trata cada uma destas áreas.

3. Selecione um trabalho (relatório de pesquisa, artigo científico na sua área, etc.) e procure identificar cada uma das fases do método estatístico.

Aula 3 – População e amostra

Nesta aula definiremos o que vem a ser amostra e população, e apresentaremos as diferenças entre variáveis qualitativas das quantitativas.

3.1 Universo estatístico

Sabemos que a Estatística tem por finalidade estudar os fenômenos coletivos e as relações existentes entre eles. Entendemos por coletivo, a reunião de um grupo de pessoas ou objetos, com características semelhantes, e com um grande número de elementos, os quais na estatística passa a se chamar população ou universo.

Crespo (2002) define como sendo o conjunto de entes (pessoas, coisas) portadores de, pelo menos, uma característica em comum denominada de **população estatística** (ou universo estatístico). Já para Castanheira (2010), população é o conjunto de elementos que desejamos observar para obtermos determinados dados. A seguir, alguns exemplos:

- Estudantes, pois, constituem uma população; apresentam pelo menos uma característica comum: são os que estudam.
- Mulheres que fazem musculação, constituem uma população, pois apresentam pelo menos uma característica comum: são os que se exercitam em academias.
- Produtos entregues na região sul, constituem uma população, pois apresentam pelo menos uma característica comum: vão para a região sul.

Figura 3.1: População
Fonte: ©Andrea Danti/Shutterstock

Como em qualquer estudo ou análise estatística, temos que pesquisar uma ou mais características dos elementos de alguma população. Esta característica deve estar perfeitamente definida. Segundo Crespo (2002, p.19) isto se dá quando, considerado um elemento qualquer, podemos afirmar, sem ambiguidade, se esse elemento pertence ou não à população. É necessário, pois, existir um critério de constituição da população, válido para qualquer pessoa, no tempo ou no espaço.

Por isso, se fizéssemos uma pesquisa sobre produção de peças defeituosas, teríamos primeiramente que definir quais seriam as peças que desse universo, isto é, peças que atualmente estão sendo produzidas, ou incluir também as que já foram e apresentaram defeitos? É claro que a solução do problema vai depender de cada caso em particular.

A população pode ser classificada também em finita ou infinita. Entendemos **população finita,** aquela que conhecemos o seu número total de elementos. Por exemplo: estamos analisando o aproveitamento nas entregas de um determinado produto em 1000 municípios. Veja que sabemos exatamente quantos municípios estão sendo observados. Logo, a população é de 1000, ou seja, tem um fim, um limite, portanto é finita. No entanto, se a população possui um número infinito de elementos, ela é uma **população infinita**. Exemplo: satisfação dos clientes que utilizam compras via internet. É impossível mensurar um único número para essa variável.

Na maioria das vezes, por dificuldade de atender todo o universo, impossibilitando seu controle ou inviabilizando-o economicamente, ou ainda em função do tempo, limitamos as observações referentes a uma determinada pesquisa a apenas uma parte da população. A essa parte chamamos de amostra. Logo, podemos definir amostra como sendo parte de uma população que se quer observar e levantar características.

Para Crespo (2002), a amostra é um subconjunto finito de elementos extraídos de uma população. Já Milone (1993, p.16) define amostra como um "subconjunto, representativo ou não, da população em estudo. Essa representatividade da amostra ocorre quando ela apresenta as mesmas características gerais da população da qual foi extraída".

Você viu que a definição depende muito do olhar da pessoa que define. Experimente você definir sob sua ótica, o que é uma amostra?

Para facilitar, vejamos o exemplo. Em época de eleições, sejam elas municipais, ou não, é comum vermos pesquisa de televisão ou jornais, que se utiliza de uma **amostra** de 1000 eleitores. Com base nos resultados, os responsáveis pela pesquisa formulam conclusões acerca da **população** de todos os milhares de eleitores em questão, para saber quem tem a preferência eleitoral.

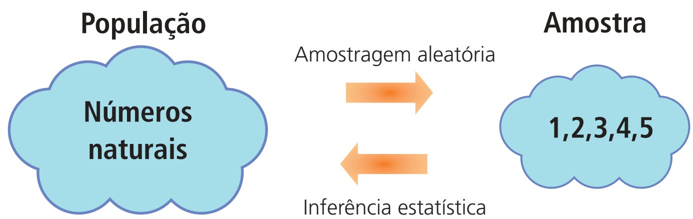


Figura 3.2: População e amostra
Fonte: Elaborado pelo autor

3.2 Tipos de variáveis

Em matemática entendemos variável como sendo o valor desconhecido de um número que varia; já na estatística, a variável está associada à característica de interesse dos elementos que está sendo estudado, pesquisado, sejam eles, pessoas ou objetos.

Os dados ou elementos são os fatos e os números coletados, analisados e interpretados; são os resultados possíveis para a variável que se quer achar. Podemos definir alguns exemplos de variáveis, a saber: nome, sexo, peso, comprimento, quantidade, entre outros.

Figura 3.3: Variáveis quantitativas e qualitativas
Fonte: http://img20.imageshack.us

As variáveis podem ser:

a) Qualitativas – são aquelas cujo valores são expressos por atributos, como sexo (masculino ou feminino), estado civil (casado, viúvo, separado, solteiro, divorciado), itens inspecionados (conformes ou não conformes), itens inspecionados (defeituosos ou não-defeituosos).

As variáveis qualitativas são subdivididas em:

- **Nominais**: quando não têm uma ordem, ou seja, uma hierarquia. A variável sexo, por exemplo, tanto faz escrever - em um questionário de pesquisa - quem vem primeiro, se é masculino ou feminino, ou vice versa.

- **Ordinais**: quando possuem ordenamento, ou seja, tem uma ordem a ser seguida. Por exemplo, a variável escolaridade sempre vem primeiro: ensino fundamental incompleto, ensino fundamental completo; depois vem ensino médio incompleto, ensino médio completo; e por último, superior incompleto, superior completo, mestrado, doutorado.

b) Quantitativas – são aquelas cujo valores são expressos em números , por exemplo, a idade das pessoas, a altura de determinados produtos, o volume, etc.

Elas são subdivididas em:

- **Contínuas**: quando o valor numérico é um valor decimal, com vírgula, sempre associada à medição. Ex. salário, altura, diâmetro das peças analisadas pelo controle de qualidade, etc.

- **Discretas**: são valores numéricos inteiros, utilizados nas contagem. Ex. número de alunos do polo, quantidade de peças defeituosas, quantidade em estoque, etc.

Resumo

População – um conjunto que admite característica de um grupo de pessoas ou objetos na sua totalidade.

Amostra – um subconjunto da população.

Variáveis – quantitativas (discretas e contínuas) – qualitativas (nominais e ordinais).

Atividades de aprendizagem

- Conhecendo a população pesquisada, classifique as variáveis em qualitativas (nominais ou ordinais) ou quantitativas (contínuas ou discretas).

Observe o exemplo dado:

População: peças produzidas por certa máquina

Variável: diâmetro externo.

Classificação: ____________________________________?

Trata-se de uma medida que é numérica e pode variar decimalmente, logo, é quantitativa e contínua.

1. Agora é a sua vez de responder.

a) População: bibliotecas da sua cidade

Variável: número de volumes (livros/obras)

Classificação: ____________________________________.

b) População: autopeças

Variável: preço das peças

Classificação: ____________________________________.

c) População: modelos

Variável: cor dos olhos

Classificação: ____________________________________.

d) População: uma caixa com 20.000 parafusos

Variável: comprimento dos parafusos

Classificação: ____________________________________.

I) a) quantitativa contínua; b) quantitativa discreta; c) quantitativa discreta; d) qualitativo nominal; e) quantitativa discreta.

2. Identifique cada número como discreto ou contínuo, completando a lacuna:

a) Cada barra de chocolate tem 16,13mg de cacau ________________ (discreto/contínuo).

b) O velocímetro de um automóvel indica uma velocidade de 80km/h ________________ (discreto/contínuo).

c) Uma pesquisa efetuada com 1.000 pessoas indica que 50 delas praticam atividades esportivas regularmente. ________________ (discreto/contínuo).

d) Das 1.000 peças de televisores digitais enviadas a um comerciante, 12 retornam com defeitos elétricos. ________________ (discreto/contínuo).

II) a) contínuo; b) discreto; c) discreto; d) discreto

Anotações

Aula 4 – Técnicas de amostragem

Nesta aula iremos conhecer as técnicas de amostragem mais utilizadas na estatística descritiva.

4.1 Separando uma amostra

Fundamentalmente, o processo pelo qual são estudadas as características de alguns fenômenos da população, através da amostra, é denominado de **amostragem**.

A-Z

Amostragem
É o estudo das relações existentes entre a população e as amostras dela extraídas. É o conjunto de técnicas utilizadas para a seleção de uma amostra.

Essa amostragem pode ser a probabilística ou não probabilística. É denominada **amostragem probabilística**, quando todos os elementos da população têm a mesma probabilidade de serem sorteados ou selecionados. Dessa forma, cada elemento da população passa a ter a mesma chance de ser escolhido, garantindo à amostra o caráter de representatividade, e isto é muito importante, pois nossas conclusões referentes à população vão estar baseadas nos resultados obtidos nas amostras da população.

Devemos garantir que a amostra seja selecionada por processos adequados e criteriosos. Caso contrário, o estudo poderá levar a resultados errados, tornando a pesquisa não representativa.

A amostragem probabilística também conhecida como amostragem aleatória, garante o acaso na escolha, ou seja, cada elemento tem a mesma chance de pertencer à amostra, de ser selecionado.

Os três tipos de amostragem probabilísticos ou aleatórios mais utilizados são: amostragem aleatória simples, amostragem proporcional estratificada e amostragem sistemática.

Embora os tipos de amostragem acima sejam os mais utilizados, outros tipos podem ser evidenciados conforme sua necessidade e facilidade em decorrência da pesquisa.

Vamos conhecer os tipos de amostras probabilísticas mencionados acima.

4.1.2 Amostragem aleatória simples

Este tipo de amostragem é equivalente a um sorteio. Para Crespo (2002), a **amostragem aleatória** pode ser realizada numerando-se a população de **1** a **N** e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, **n** números dessa sequência, os quais corresponderão aos elementos pertencentes à amostra.

Exemplo: Vamos obter uma amostra representativa de 20% para a pesquisa de estatura de 50 alunos de uma escola. Primeiramente, numeramos os alunos de 01 a 50. Em segundo lugar, escrevemos os números, de 01 a 50, em pedaços iguais de um mesmo papel, colocando-os dentro de um saco ou urna. Agitamos, sempre para misturar bem os pedaços de papel e retiramos, um a um, dez (10) números que formarão a amostra.

Quando o número de elementos da amostra é grande, esse tipo de sorteio torna-se muito trabalhoso. A fim de facilitá-lo, utiliza-se de uma tabela de (**figura 4.1**) números aleatórios.

Assim, para o nosso exemplo, considerando a 18ª linha, tomamos os números de dois algarismos, obtendo:

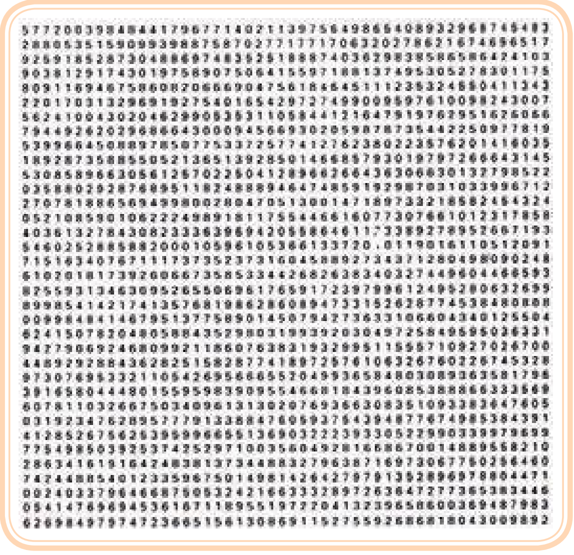


Figura 4.1: Tabela de números aleatórios
Fonte: http://tic4-39188.blogspot.com.

4.1.3 Amostragem proporcional estratificada

Muitas vezes a população se divide, em subpopulações, subconjuntos ou estratos. Dessa forma, a amostragem estratificada seleciona quantos elementos da amostra serão retirados de cada estrato.

Quando a população se divide em estratos, é interessante que o sorteio dos elementos que irão compor a amostra leve em consideração os estratos, ou seja, obtêm-se os elementos da amostra de modo proporcional ao número de elementos de cada estrato, consistindo assim em quantos elementos da amostra serão retirados de cada estrato.

A determinação dos elementos dessa amostra pode ser feita de vários modos, entre elas destacamos a uniforme e a proporcional. A uniforme se caracteriza por um sorteio que estabelece um número igual de elementos em cada estrato; já a proporcional, se resume em obter o número de elementos sorteados em cada estrato pela proporção, ou seja, o número de elementos sorteados da amostra deve ser proporcional ao número de elementos dos estratos. Vejamos a **figura 4.2** que resume bem isso!

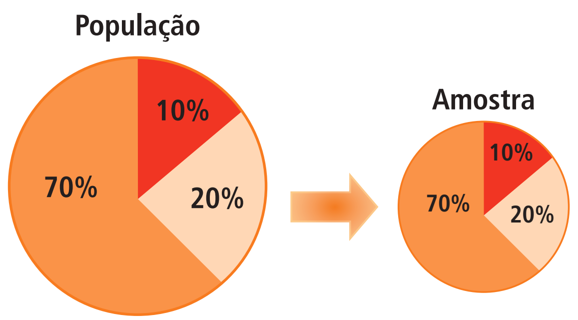


Figura 4.2: Amostra estratificada
Fonte: Adaptado de www.mbi.com.br

Exemplo: Se considerarmos, por exemplo, que, de 70 alunos, 40 sejam meninas e 30 meninos, vamos obter amostra proporcional estratificada.

Temos dois estratos (sexo masculino e feminino) se queremos uma amostra de 10%, teremos: três homens (10% de 30) e quatro mulheres (10% de 40).

4.1.4 Amostragem sistemática

Esse sistema é igual ao da amostragem aleatória simples, porém os elementos da população já se encontram ordenados. São exemplos: os prontuários médicos de um hospital, os prédios de uma rua, as linhas de produção de um fábrica, etc.

Nestes casos, a seleção dos elementos que constituirão a amostra pode ser feito por um sistema imposto pelo pesquisador ou utilizando de regras de escolha, pode-se indicar a regra da determinação por divisão. Essa regra consiste em achar um coeficiente de ordenação ou de escolha. Suponha que K, seja o coeficiente de escolha, podemos determiná-lo pela divisão do número de elementos (N) pelo número tamanho da amostra (n). Observe:

$$K = \frac{N}{n}$$

Onde:

N → tamanho da população

n → tamanho da amostra

K → coeficiente de escolha

Vamos entender esse conceito através de um exemplo:

Uma indústria produz 900 brinquedos de artefatos de madeira por dia, dos quais desejamos obter uma amostra formada por 50 brinquedos. Podemos usar o seguinte procedimento: como, $K = \frac{N}{n} = \frac{900}{50} = 18$ escolhemos por sorteio casual um número de 1 a 18 (inclusive), o qual indicaria o primeiro elemento (brinquedo) sorteado para a amostra; os demais elementos seriam periodicamente considerados de 18 em 18. Assim, se o número sorteado fosse o sétimo brinquedo produzido (7º), tomaríamos como segundo brinquedo para a amostra o 25º produzido, e assim por diante, até completar os 50 brinquedos da amostra.

As amostras não probabilísticas são técnicas de amostragem em que a escolha é deliberada dos elementos da amostra. Elas são também utilizadas devido a sua simplicidade, ou devido a impossibilidade de se obterem amostras probabilísticas. Para Crespo (2007), em muitos casos os efeitos da utilização de uma amostra não probabilística podem ser equivalentes ao de uma amostragem probabilística.

São amostragens não probabilísticas, as técnicas de:

a) Amostragem por conveniência ou acidental.

b) Amostragem intencional ou por julgamento.

c) Amostragem por cotas.

4.2 Modelo de cálculo aproximado para determinação do tamanho de uma amostra qualquer

Embora a definição do tamanho de uma amostra seja bastante subjetiva, pois depende muito do pesquisador, cabe a ele responder a pergunta: **Qual é o tamanho da amostra que eu preciso?**

O cálculo do tamanho da amostra deve fazer parte de qualquer projeto de pesquisa. O objetivo principal é estabelecer qual o número de indivíduos que necessitam ser estudados, para que se tenha uma representatividade da população.

Uma forma aproximada para determinar o **tamanho de uma amostra,** obtida a partir de uma amostragem aleatória, é determinada pelo cálculo matemático abaixo. Desde que se conheça o erro amostral e o tamanho da população. Fórmula para a primeira aproximação do tamanho amostral:

$$n_o = \frac{1}{e_o^{\,2}}$$

Onde:

$n_o \rightarrow$ tamanho da amostra (1ª aproximação)

$e_o^{\,2} \rightarrow$ erro amostral

Você sabia?

Sobre o tamanho de uma amostra:

Silveira Júnior *et. al.* (1980), ao dimensionar uma amostra, diz que se necessita do conhecimento prévio da variância da população e do grau de precisão desejado, mas quando não se dispõem de informações sobre esta variabilidade, deve-se realizar uma pré-amostragem, em pequena escala para que se possa obter as estimativas dos parâmetros populacionais (média e variância) que serão usados na obtenção do melhor tamanho da amostra, pois quanto maior o tamanho amostral, maiores serão o tempo e os gastos despendidos com a amostragem. Por outro lado, amostras pequenas podem resultar em menor precisão, o que é indesejável.

Após achar o, n_o , aplica-se um fator de correção, o qual leva em conta o tamanho da população:

$$n = \frac{N \cdot n_o}{N + n_o}$$

Onde:

$n_o \rightarrow$ tamanho da amostra (1ª aproximação)

$n \rightarrow$ tamanho da amostra final

$N \rightarrow$ tamanho da população

Exemplo:

Em uma empresa que contém 2.000 colaboradores, deseja-se fazer uma pesquisa de satisfação. Quantos colaboradores devem ser entrevistados para tal estudo?

Resolução:

N = 2.000

Definindo o erro amostral tolerável em 2%

$E_0 = 0{,}04$

$$n_0 = \frac{1}{(E_o)^2}$$

$$n_0 = \frac{1}{(0{,}04)^2}$$

$n_0 = 625$

$$n = \frac{N \cdot n_o}{N + n_o}$$

$$n = \frac{2.000 \cdot 625}{2.000 + 625}$$

$$n = \frac{1.250.000}{2625}$$

n = 476,19 que arredondado vai para 476 colaboradores, esse seria o tamanho da amostra, tendo uma representatividade de 23,8% da população.

Para uma melhor fixação, observe o exercício resolvido abaixo, e depois resolva a alternativa 5 utilizando o tipo de amostragem sistemática.

Uma linha de produção de uma fábrica de parafusos produz 1.000 peças por turno. Como você obteria uma amostra correspondente a 2% das peças para realização de teste de controle de qualidade?

Calculamos 2% de 1.000 = 20, agora utilizamos a expressão $K = \dfrac{N}{n}$

Substituindo temos, $K = \dfrac{1.000}{20} = 50$.

Agora, separamos os parafusos de 50 em 50, partindo do primeiro: 50, 100, 150, 200, 250, ...1.000, totalizando assim 20 parafusos para nossa amostra.

(Fonte: http://www.datalyzer.com.br/site/suporte/administrador/info/arquivos/info60)

Resumo

- Tipos de técnicas de amostragem probabilística: amostragem aleatória simples; amostragem proporcional estratificada; amostragem sistemática.

- Determinação do tamanho de uma amostra: $n_o = \dfrac{1}{e_o^{\,2}}$ e $n = \dfrac{N \,.\, n_o}{N + n_o}$.

Atividades de aprendizagem

1. Identifique nas alternativas abaixo, o tipo de amostragem utilizado.

a) Ao escalar um júri, o tribunal de justiça decidiu selecionar aleatoriamente 4 pessoas brancas, 3 morenas e 4 negras.

Resposta:____.

b) Um cabo eleitoral escreve o nome de cada senador do Brasil, em cartões separados, mistura e extrai 10 nomes.

Resposta:____.

c) Um administrador hospitalar faz uma pesquisa com pessoas que estão na fila de espera para serem atendidas pelo sistema SUS, entrevistando uma a cada 10 pessoas da fila.

Resposta:____.

a) Amostragem estratificada; b) Amostragem simples ao acaso; c) Amostragem sistemática. Se a escolha dessa fila não partiu de nenhum tipo de amostragem probabilística, poderíamos até considerar que a amostragem é inicialmente por conveniência (escolha da fila), a qual foi então seguida por uma amostragem sistemática.

2. Se uma população se encontra dividida em quatro estratos (N), com tamanhos $N_1 = 90$, $N_2 = 120$, $N_3 = 60$ e $N_4 = 480$ e temos possibilidade de retirar no total 100 elementos para compor a amostras, quantos elementos devem ser retirados de cada estrato?

Resposta: a) n1=12 ; b) n2 = 16 ; c) n3 = 8 ; d) n4 = 64

3. Numa população que apresenta grande variação, em relação a certa característica em estudo, qual a técnica de amostragem mais adequada?

Resposta: Amostragem estratificada, sendo que em cada estrato tem que ter a características homogêneas, no que se deseja estudar.

4. Uma determinada empresa possui 240 funcionários em seu quadro. Obtenha uma amostra representativa dessa população, que corresponda a 20%.

Resposta: 48 funcionários sorteados pela amostragem simples.

5. Em uma fábrica de luminárias existe um compartimento com 350 lâmpadas, numeradas em ordem crescente. Obtenha uma amostra de 18 lâmpadas.

Resposta: Utilizamos a amostragem sistemática, determinando K = N/n, temos K = 19,44 que arredondado pode ser 20. Listamos as 20 primeiras lâmpadas e sorteamos uma, depois escolhemos as outras 17 amostras faltantes pulando de 20 em 20.

6. Obter o tamanho mínimo de uma amostra aleatória simples, admitindo alto grau de confiança, erro amostral de 5,5%, supondo que a população tenha:

- 550 elementos
- 27000 elementos

Resposta: Para 550° tamanho de N= 206,48 e para 27000 o tamanho de N = 326,57

Aula 5 – Tabelas

Nesta aula iremos conhecer os elementos considerados importantes para a elaboração de uma tabela, bem como diferenciar as tabelas pelas suas variáveis.

5.1 Organizando os dados

As tabelas são formas retangulares de apresentar dados de modo agrupado por linhas e colunas. Seu objetivo é resumir valores, que podem ser apresentados por uma ou mais variáveis. Para começarmos a construir as tabelas é essencial levar em consideração algumas informações imprescindíveis que devem conter nas tabelas. Observe abaixo:

A tabela é um quadro que resume um conjunto de observações. Ela é composta por:

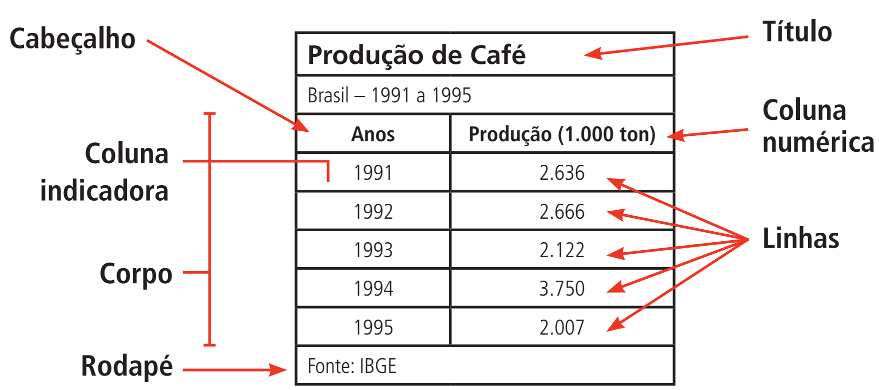


Figura 5.1: Componentes de uma tabela
Fonte: Adaptado de Estatística fácil, 2002.

5.2 Séries estatísticas

É a denominação que se dá as tabelas que têm critérios distintos que as específicas e as diferencia. Segundo Castanheira (2010), as séries estatísticas se diferenciam principalmente, pelas variações dos seguintes elementos: tempo, local e fato (fenômeno), fornecendo assim informações rápidas e seguras.

Vejamos algumas delas:

- **Séries históricas**

Também chamadas de temporal ou cronológicas, descrevem os valores da variável em função do tempo, permanecendo fixo as variáveis locais e o fenômeno (ou fato). Os dados são produzidos ao longo do tempo.

Tabela 5.1: Série cronológica

Custo de produção da Empresa Barbosa S/A Ltda.	
Ano	**Valores (em R$)**
1993	120.000
1994	124.000
1995	118.000
1996	132.000

Fonte: Elaborado pelo autor

- **Séries geográficas**

Chamadas também de espaciais ou territoriais. Têm como característica a variação do local, ou região. As variáveis tempo e fato permanecem sempre fixas. Vejamos o exemplo.

Tabela 5.2: Série geográfica

Vendas previstas por região – RJ / 1995	
Mercado	**Vendas (t/ano)**
Petrópolis	20.000
Teresópolis	35.000
Três Rios	18.000
Volta Redonda	42.000

Fonte: FGV Management

- **Séries específicas**

Leva em consideração a característica do fato ou fenômeno, considerando o tempo e o local de modo fixo. Isso que dizer que, leva em consideração os dados segundo as suas especificidades de ocorrência ou categorias. São conhecidas também como séries categóricas ou de qualidade.

Exemplo: Número de candidatos ao vestibular em uma Universidade do Estado do Paraná – 2010 na área de Ciências Sociais.

Tabela 5.3: Séries específicas

ÁREAS OFERTADAS	N° DE CANDIDATOS
Contabilidade	5.400
Economia	3334
Administração	7985
Ciências Atuárias	3569

Fonte: Elaborado pelo autor

- **Séries conjugadas ou de dupla entrada ou mistas**

Quando temos a conjunção de duas séries estatísticas, essa conjunção se dá entre as séries temporais, geográficas e especificas, em uma única tabela. Nesse tipo série, ficam criadas duas ordens de classificação: uma horizontal (linha) e uma vertical (coluna).

Tabela 5.4: Série conjugada

Terminais telefônicos em serviço 1991 - 1993			
Regiões	**1991**	**1992**	**1993**
Norte	342.938	375.658	403.494
Nordeste	1.287.813	1.379.101	1.486.649
Sudeste	6.234.501	1.379.101	1.486.649
Sul	1.497.315	1.608.989	1.746.232
Centro-oeste	713.357	778.925	884.822

Fonte: Ministério das comunicações

- **Distribuição de frequência**

É uma série particular que apresenta os resultados de uma pesquisa através de uma tabela que mostra a frequência (o número de vezes) de ocorrência de cada resultado. Ela pode ser elaborada com ou sem intervalos. Veremos mais adiante essa série em especial, por enquanto podemos visualizar a tabela 5.5 como exemplo prático.

Tabela 5.5: Distribuição de frequência com intervalos

Diâmetros (em cm.), de certa peça circular produzida por uma indústria			
Diâmetro (cm)	**f**	**fr**	**fa**
1,810 ⊢— 1,818	5	0,10	5
1,818 ⊢— 1,826	5	0,10	10
1,826 ⊢— 1,834	11	0,22	21
1,834 ⊢— 1,842	13	0,26	34
1,842 ⊢— 1,850	7	0,14	41
1,850 ⊢— 1,858	5	0,10	46
1,858 ⊢— 1,866	4	0,08	50
	$\Sigma f_i = 50$	$\Sigma fr_i = 1{,}00$	

Fonte: Elaborado pelo autor

Resumo

Nesta aula vimos:

- Tabelas – forma tabular de apresentar dados.
- Séries estatísticas – são tabelas com especificidades estatísticas.
- São séries estatísticas – a série temporal, geográfica, específica, mista e a distribuição de frequência.

Atividades de aprendizagem

1. Confeccione uma tabela, que contenha todos os elementos essenciais para sua identificação.

2. Analise em jornais ou em revistas, os tipos de tabelas que podem ser identificados como séries estatísticas.

Aula 6 – Gráficos

Nesta aula veremos alguns tipos de gráficos, bem como sua importância na apresentação dos dados.

6.1 Apresentando os gráficos

O gráfico é uma forma de apresentar dados estatísticos. Um de seus objetivos é uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão do que as séries.

Para Castanheira (2010), a representação gráfica deve obedecer três critérios:

- **Simplicidade** – ser despoluído de detalhes de importância secundária, tais como traços desnecessários que possam levar o observador a uma análise errônea.
- **Clareza** – possibilitar correta interpretação dos valores representativos do estudo.
- **Veracidade** – expressar a verdade da pesquisa ou de suas variáveis.

Segue uma lista dos gráficos de linhas, barras e setores mais utilizados, porém não queremos esgotar o assunto, que requer muitos modelos e formatos. Sua criação se dá pelos softwares estatísticos ou pelo próprio Excel. Vejamos:

- **Gráfico em linha ou em curva:** são os que normalmente representam as séries temporais.

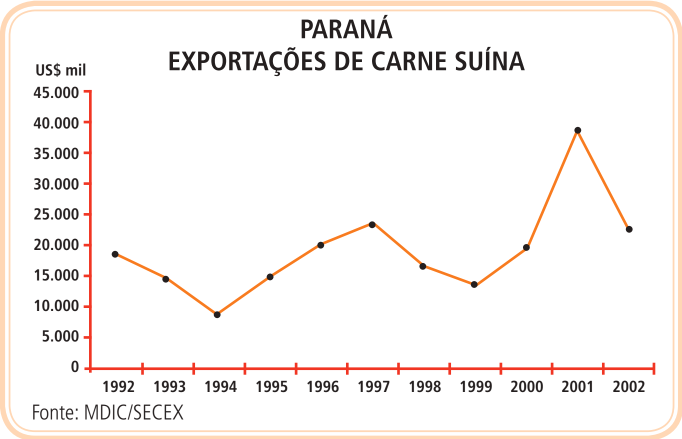


Figura 6.1: Gráfico em linha ou em curva
Fonte: MDIC/SECEX

- **Gráfico em colunas ou em barras:** são utilizados em séries geográficas e específicas.

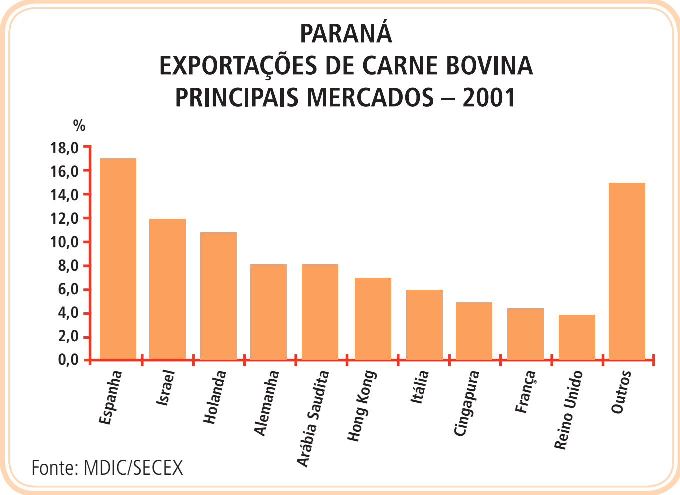


Figura 6.2: Gráfico em colunas ou em barras
Fonte: MDIC/SECEX

- **Gráfico em setores:** (pizzas) é utilizado para dados em percentuais; usado também em séries geográficas, específicas e mistas.

Figura 6.3: Gráfico em setores
Fonte: MDIC/SECEX

6.2 Histograma

É a representação gráfica das distribuições de frequências. Uma ferramenta para a análise e representação de dados quantitativos, agrupados em classes de frequência (com intervalos) que permite diferenciar a forma, o ponto central e a variação da distribuição, além de outros dados como amplitude e simetria na distribuição dos dados. O histograma faz parte das ferramentas básicas da qualidade que podem ser aplicadas em situações reais e menos complexas.

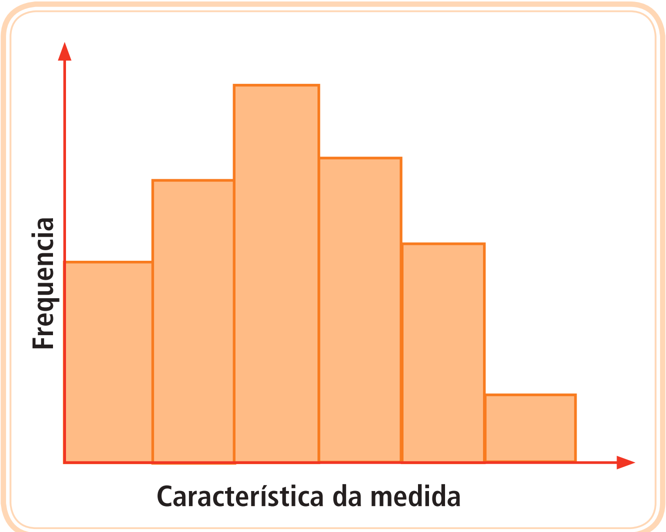


Figura 6.4: Histograma
Fonte: http://mathforum.forumais.com/

Resumo

Vimos nesta aula os critérios de simplicidade, clareza e veracidade para a construção de um gráfico. Você conheceu também os tipos de gráficos: de linha, de barras, de setores e o histograma.

Atividades de aprendizagem

- Utilize as séries da aula 5 para fazer gráficos com a ferramenta **Assistente Gráfico** do Excel.

Anotações

Aula 7 – Dados, rol, frequência e tabulação

Nesta aula você aprenderá a determinar os conceitos de dados, rol, frequência e dados tabulados.

7.1 Organizando os dados

Entendemos por dados, os elementos (objetos) da pesquisa estatística, os quais podem ser qualitativos (atributos) ou quantitativos (números), como já vimos.

Esses dados são chamados de brutos, pois ainda não sofreram nenhum tipo de organização para serem usados. Vejamos um exemplo:

Suponha que você fez uma pesquisa de opinião querendo saber sobre a satisfação no atendimento de um funcionário do comércio **X**. Ao recolher esse questionário, você não se preocupou em organizá-lo. Esse tipo de procedimento é entendido como sendo dados brutos. No momento em que os dados brutos são organizados por ordem alfabética ou numérica, passo a dar um tratamento mais adequado.

Em se tratando de pesquisa quantitativa, a partir do momento que ordenamos numericamente, em ordem crescente, decrescente, ou outra condição qualquer, criamos uma relação de ordenamento, que chamamos de **Rol**.

A-Z

Rol
Enumeração um tanto minuciosa; catálogo, lista, relação, designação de coisas, quantias ou circunstâncias, segundo determinada ordem, para registro, fixação ou recordação.
Fonte: dicionário Houaiss, 2010.

Observe o exemplo:

Um professor da disciplina de controle e armazenamento de materiais, aplica um teste para 30 alunos. Esse professor lista em uma planilha (poderia ser uma folha de caderno),as notas dos alunos, como podemos ver:

70 – 60 – 80 – 90 – 60 – 50 – 70 – 40 – 60 – 70 – 90 – 80 – 70 – 100 – 40 – 30 –40 – 40 – 50 – 60 – 50 – 60 –100 – 70 – 80 – 60 – 60 – 50 – 80 – 90.

A essa relação chamamos de dados brutos. Porém percebemos que essa relação não nos permite tirar maiores conclusões.

Vamos melhorar essa relação colocando em ordem: 30 – 40 – 40 – 40 – 40 – 50 – 50 – 50 – 50 – 60 – 60 – 60 – 60 – 60 – 60 – 60 –70 – 70 – 70 – 70 – 70 – 80 –80 – 80 – 80 – 90 – 90 – 90 – 100 – 100. Você percebeu as diferenças?

7.1 Frequências

Entendemos por frequência a quantidade de vezes que um dado ou objeto se repete. As frequências podem ser absolutas, relativas ou acumuladas.

Segundo Castanheira (2010, p. 25), a **frequência absoluta** é o número de vezes que um mesmo resultado acontece durante uma pesquisa. Já a frequência relativa é a frequência absoluta representada de modo proporcional em relação ao conjunto de dados observados. A frequência acumulada, o próprio nome já diz, é o cálculo da frequência acumulando termo a termo.

Mas voltando ao exemplo das notas dos alunos. Fica muito mais fácil se montarmos uma tabela (que chamamos de distribuição de frequência), com as devidas frequências de suas notas. Observe:

Tabela 7.1: Distribuição de frequência das notas dos alunos

Notas	Frequência das notas dos alunos
30	1
40	4
50	4
60	7
70	5
80	4
90	3
100	2
Total	**30**

Fonte: Elaborado pelo autor

Essa tabulação, além de organizar os dados, facilita responder rapidamente algumas perguntas, como:

- Quantos alunos tiraram nota 30, 60 e 70?
- Quantos alunos estão acima da média escolar?

Viram como é fácil entender que a tabela acima é uma distribuição de frequência? Porém, quando temos muitos dados, é difícil fazer tabelas como a mencionada anteriormente, por isso fazemos uso de intervalos. Este assunto será visto mais a diante.

Sabemos que a frequência absoluta, representada por f, é simplesmente o número de vezes que o resultado correspondente à variável em estudo aparece na pesquisa. Já frequência relativa (fr) é a frequência na forma percentual. Ela é obtida dividindo-se a frequência da classe pelo somatório de todas as frequências. Elas podem ser calculadas pela fórmula :

$$fr = \frac{f_i}{\sum f_i}$$

A frequência acumulada normalmente representada por **fa** é o somatório das frequências das classes inferiores, ou iguais à classe que está sendo examinada. Veja a **tabela 7.2**:

Tabela 7.2: distribuição de frequência com intervalos e suas respectivas frequências. Observação: O símbolo ∑ indica somatório.

Diâmetros (em cm.), de certa peça circular produzida por uma indústria			
Diâmetro (cm)	**f**	**fr**	**fa**
1,810 ⊢ 1,818	5	0,10	5
1,818 ⊢ 1,826	5	0,10	10
1,826 ⊢ 1,834	11	0,22	21
1,834 ⊢ 1,842	13	0,26	34
1,842 ⊢ 1,850	7	0,14	41
1,850 ⊢ 1,858	5	0,10	46
1,858 ⊢ 1,866	4	0,08	50
	$\sum f_{i = 50}$	$\sum fr_{i = 1,00}$	

Fonte: Elaborado pelo autor

Resumo

- Dados brutos – é o elementos observado na sua ordem primária ou primitiva.
- Rol – ordenação de dados numéricos.
- Frequência – número de vezes que um dado aparece.
- Tabulação – modo de organizar os dados de uma pesquisa.

Atividades de aprendizagem

1. Determine o rol dos seguintes números: 1, 3, 5, 5, 7, 9, 11, 1, 2, 4.

2. Dados vinte valores de variáveis (x) quaisquer abaixo:

1, 2, 2, 2, 2, 4, 10, 11, 10, 4, 10, 10, 9, 9, 8, 8,10, 8, 4, 9, 8

Determine o rol e construa uma distribuição de frequência, contendo uma coluna para a frequência relativa e outra para acumulada.

Anotações

Aula 8 – Distribuição de frequência

Nessa aula, além de nos aprofundarmos um pouco mais sobre o assunto, veremos como confeccionar uma distribuição de frequência com ou sem intervalos.

8.1 Diferenciando as tabelas

A tabela que apresenta os dados com sua respectiva frequência é denominada de Distribuição de Frequência. Pois bem, essa distribuição pode ser com ou sem intervalos. O que isso quer dizer? Quer dizer que vários dados deixam as tabelas com muitas linhas, tornando inviável sua construção e análise, assim utilizamos as tabelas com intervalos. Você tem na sequência alguns exemplos de distribuição de frequência:

Tabela com poucos dados e sem intervalos, normalmente utilizados para dados qualitativos ou quantitativos.

Tabela 8.1: Tabela com poucos dados e sem intervalos

Número de filhos	
Quantidade	**Frequência (f_i)**
0	2
1	3
2	5
3	11
4	2
5	1
6	1
	$\sum f_{i = 25}$

Fonte: Dados fictícios do autor

Tabela com muitos dados e com intervalos, normalmente utilizados para dados quantitativos.

Tabela 8.2: Tabela com muitos dados e intervalos

Diâmetro (em cm.) de uma peça produzida por certa indústria	
Quantidade	Frequência
1,810 ├── 1,818	5
1,818 ├── 1,826	5
1,826 ├── 1,834	11
1,834 ├── 1,842	13
1,842 ├── 1,850	7
1,850 ├── 1,858	5
1,858 ├── 1,866	4
	$\sum f_{i = 50}$

Fonte: Dados fictícios do autor

8.1 Elementos de uma distribuição de frequência com intervalos

Para elaborar uma distribuição de frequência com intervalos, é preciso que você conheça alguns elementos. Para simplificar, listamos esses elementos na distribuição abaixo.

- **Classes (k)** – classes de frequência ou simplesmente classes. São intervalos da variável, popularmente chamada de linhas da distribuição. As classes têm por objetivo determinar a quantidade de linhas que terá a distribuição. Este número (quantidade de linhas) pode ser determinado por três modos:

  – pelo método empírico, onde o pesquisador determina a quantidade de linhas;
  – pela regra de Sturges, que determina pelo cálculo da equação **k = 1 + 3,3 log n** onde **n** é o numero de dados da amostra ou população.
  – pela regra prática, que consiste em calcular a raiz quadrada de **n**, como na expressão $k = \sqrt{n}$.

Exemplo: Pela regra prática, se **n** tivesse 40 dados, a tabela seria determinada por um número de linhas k igual a 6, pois: $k = \sqrt{n} = \sqrt{40} = 6$. Inicia-se então a montagem da distribuição de frequência.

- **Limites de classe** – são os extremos de cada classe. Cada classe (linha) possui um limite inferior (L_i) e um limite superior (L_s). Observe a **figura 8.1**:

Limite inferior L_i Limite superior L_s

1ª classe
2ª classe

Estaturas (cm)	Frequência
150 $\vdash$ 154	4
154 $\vdash$ 158	9
158 $\vdash$ 162	11
162 $\vdash$ 166	8
166 $\vdash$ 170	5
170 $\vdash$ 174	3
Total	40

Figura 8.1: Distribuição de frequência com intervalos
Fonte: Elaborado pelo autor.

- **Amplitude total da distribuição (A_T)**

É o valor da diferença entre o limite superior da última classe pelo limite inferior da primeira classe. Segue abaixo a generalização:

$$A_t = L_{max} - L_{min} = 174 - 150 = 24$$

- **Amplitude de um intervalo de classe(h)**

É o que indica o tamanho do intervalo, ele é determinado pela diferença do limite superior da classe (L_s) pelo limite inferior da classe (L_i).

$$h = L_s - L_i = 154 - 150 = 4$$

Ou $h = \dfrac{A_T}{k} = \dfrac{24}{6} = 4.$

Observação: Todas as classes apresentam a mesma amplitude, ou seja, o mesmo tamanho.

- **Ponto médio de uma classe**

É utilizado para o cálculo da média aritmética para dados agrupados (tabelados) com intervalos. Para a primeira classe temos:

$$x_i = \frac{L_i + L_s}{2} = \frac{150 + 154}{2} = 152$$

8.2 Reforçando os cálculos que determinam outros tipos de frequências

Utilizando a **tabela 8.2**, indicamos como se calcula as frequências:

Limite inferior L_i Limite superior L_s

1ª classe
2ª classe

Estaturas (cm)	Frequência
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
Total	40

Figura 8.2: Cálculo das frequências
Fonte: Elaborado pelo autor.

- **Frequência simples ou absoluta (f_i)**

É o valor que representa o número de dados de cada classe. Sempre teremos: $\sum f_i = n$, onde **n** é o número total de observações.

- **Frequência relativa (f_r)**

É o valor da razão entre a frequência simples e a frequência total.

$$\mathbf{fr = \frac{f_i}{\Sigma f_i} = \frac{f_i}{n}}$$

Da distribuição acima temos que a frequência relativa da 3ª classe é

$$fr = \frac{11}{40} = 0{,}275$$

O propósito das frequências relativas é permitir a análise ou facilitar comparações de dados.

- **Frequência acumulada** (F_i)

É o total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe:

$$\mathbf{F_k = f_1 + f_2 + \ldots + f_k}$$

A frequência acumulada da 3ª classe da distribuição acima ilustrada é $F_i = 24$ (4 + 9 + 11), o que significa que existem 24 alunos com altura inferior a 162cm (limite superior do intervalo da 3ª classe).

- **Frequência acumulada relativa (F_r):**

É o valor da razão entre a frequência acumulada simples e a frequência total.

$$F_r = \frac{f_i}{\Sigma f_i}$$

Assim, para a terceira classe poderíamos ter: $F_r = \frac{24}{40} = 0{,}60.$

Observe um exercício resolvido:

A tabela apresenta uma distribuição de frequência das áreas de 370 lotes de uma determinada cidade.

Áreas (m^2)	Número de lotes
300 ⊢ 400	14
400 ⊢ 500	16
500 ⊢ 600	58
600 ⊢ 700	76
700 ⊢ 800	68
800 ⊢ 900	62
900 ⊢ 1000	48
1000 ⊢ 1100	22
1100 ⊢ 1200	6
Total	**370**

Com referência a essa tabela, temos que determinar:

1. a amplitude total;

2. o limite superior da 5ª classe;

3. o limite inferior da 8ª classe;

4. o ponto médio da 7ª classe;

5. a amplitude do intervalo da 2ª classe;

6. a frequência da 4ª classe;

7. as frequências relativas e acumuladas, bem como a acumulada relativa.

8. a frequência relativa da 6ª classe;

9. a frequência acumulada da 5ª classe;

10. o número de lotes cuja área não atinge 700m²;

11. o número de lotes cuja área atinge e ultrapassa 800m²;

12. a percentagem dos lotes que não atingem 600m²;

13. a percentagem dos lotes cuja área é de 500m², no mínimo, mas inferior a 100m²;

14. a classe do 72º lote.

Resolução:

A primeira coisa a se fazer é calcular as frequências relativa, acumulada e acumulada relativa. Para isso monta-se a tabela estendida, com o auxílio dos tópicos abordados acima:

Salário dos funcionários da empresa ABC				
Áreas (m²)	f_i	f_r	F_i	F_r
300 ⊢ 400	14	0,0378	14	0,0378
400 ⊢ 500	16	0,0432	30	0,0810
500 ⊢ 600	58	0,1567	88	0,2378
600 ⊢ 700	76	0,2054	164	0,4432
700 ⊢ 800	68	0,1837	232	0,6270
800 ⊢ 900	62	0,1675	294	0,7945
900 ⊢ 1000	48	0,1’297	342	0,9243
1000 ⊢ 1100	22	0,0594	364	0,9837
1100 ⊢ 1200	6	0,0162	370	1
	∑370			

Resolvendo o exercício temos:

1. a amplitude total – R: $A_t = L_{max} - L_{min} = 1200 - 300 = 900$

2. o limite superior da 5ªclasse – R: $L_i = L_5 = 800$

3. o limite inferior da 8ªclasse – R: $l_i = l_8 = 1000$

4. o ponto médio da 7ªclasse – R: $x_i = \dfrac{L_i + L_s}{2} = \dfrac{900 + 1000}{2} = 950$

5. a amplitude do intervalo da 2ªclasse – R: $h = L_s - L_i = 500 - 400 = 100$

6. a frequência da 4ªclasse – R: $f_4 = 76$

7. as frequências relativas e acumuladas, bem como a acumulada relativa – R: Já está na tabela.

8. a frequência relativa da 6ª classe – R: $fr_6 = 0{,}1675$ ou 16,75%

9. a frequência acumulada da 5ª classe – R: $Fa_5 = 232$

10. o número de lotes cuja área não atinge 700m² – R: $Fa_4 = 164$

11. o número de lotes cuja área atinge e ultrapassa 800m² – R: 138

12. a percentagem dos lotes que não atingem 600m² – R: $Fa_3 = 0{,}2378$ ou 23,78%

13. a percentagem dos lotes cuja área é de 500m², no mínimo, mas inferior a 1000m² – R: 0,9243 – 0,2378 = 0,6865 ou 68,65%

14. a classe do 72º lote – R: 3º classe

Resumo

Você aprendeu a elaboração de uma distribuição de frequência para dados quantitativos e os passos que devem ser seguidos:

1. A partir da tabela primitiva, elaborar o rol dos dados;

2. Determinar o número de classes (k). Este número pode ser determinado pelo método empírico, pela regra de Sturges: $k = 1 + 3{,}3 \log n$ ou pela regra prática $k = \sqrt{n}$;

3. Determinar a amplitude do intervalo de classe (A_T);

4. Determinar os limites inferiores e superiores de cada classe;

5. Realizar a contagem dos elementos de cada classe;

6. Montar a distribuição de frequência nas formas tabular e gráfica.

Atividades de aprendizagem

1. Os valores da distribuição de frequência abaixo se referem aos preços de locação de kitnets de um ambiente com banheiro no bairro Jamacuri, pela imobiliária Barbosa, entre os períodos de 01/03 a 30/02/2009.

Preço de locação (em R$)	Quantidade de imóveis
130 ⊢ 185	3
185 ⊢ 240	7
240 ⊢ 295	4
295 ⊢ 350	6
350 ⊢ 405	9
405 ⊢ 460	1
Total	**30**

Responda:

a) Qual a amplitude total da distribuição?

b) Qual a amplitude da 3ª classe?

c) Quantos sobrados possuem seu preço de locação que estão entre R$240,00 e R$405,00?

d) Qual a frequência relativa da 6ª classe?

e) Qual a frequência acumulada da 4ª classe?

f) Qual o percentual de sobrados cujo preço de locação está entre R$240,00 (inclusive) e R$460,00?

2. Uma distribuição de frequência está representada abaixo:

Salário dos funcionários da empresa Começar Empreendimentos Logísticos Ltda.				
Faixa salarial, em R$	f_i	f_r	F_i	F_r
0,00 ⊢ 240,00	4			
240,00 ⊢ 480,00	8			
480,00 ⊢ 720,00	14			
720,00 ⊢ 960,00	28			
960,00 ⊢ 1200,00	21			
1200,00 ⊢ 1440,00	12			
1440,00 ⊢ 1680,00	7			
Acima de 1680,00	6			
	100			

a) Complete as lacunas restantes.

b) Com base na tabela, responda qual é o número de pessoas que recebe menos que R$960,00?

c) Qual a porcentagem de pessoas que recebe entre R$1.200,00 (inclusive) e R$1.440,00?

d) Qual o número de pessoas que recebe de R$720,00 para cima?

3. Uma confecção contratou uma empresa para fazer uma pesquisa para saber a estatura das pessoas adultas de determinado bairro da sua cidade, a fim de saber como encaminhar a produção das roupas que produz. A pesquisa, feita com uma amostra de 40 adultos escolhidos ao acaso, revelou os seguintes dados:

Estaturas (em m) de 40 pessoas adultas									
1,66	1,50	1,62	1,79	1,64	1,85	1,68	1,65	1,61	1,68
1,63	1,61	1,68	1,76	1,70	1,70	1,75	1,62	1,63	1,70
1,58	1,57	1,73	1,82	1,72	1,68	1,70	1,70	1,67	1,67
1,56	1,58	1,69	1,65	1,58	1,65	1,64	1,64	1,64	1,69

Complete:

a) A variável em questão é ______________ (qualitativa/quantitativa).

b) Essa variável é do tipo ______________ (contínua/discreta).

c) Elabore uma distribuição de frequência relativa.

d) Elabore o histograma de frequência relativa.

4. Os dados para os números de unidades produzidas por um determinado funcionário da área de produção durante vinte dias são apresentados a seguir:

160	170	181	156	176
148	198	179	162	150
162	156	179	178	151
157	154	179	148	156

Resuma os dados construindo em seu caderno:

a) Uma distribuição de frequência relativa

b) Uma distribuição de frequência relativa acumulada

c) Os respectivos histogramas para os itens (a) e (b).

5. O diretor de produção de uma grande fábrica de chocolates resolveu fazer uma inspeção surpresa na linha para verificar o peso dos chocolates, os quais deveriam ter 200 ±1g. Para tanto, coletou uma amostra com 40 unidades e verificou, um a um, o peso dos chocolates. Os resultados obtidos foram os seguintes:

199,4	198,5	199,7	200,3	197,8	199,2	201,2	200,5
198,2	197,9	199,8	201,4	199,4	199,4	200,1	198,9
197,9	199,2	200,2	200,3	199,7	199,5	197,8	199,5
199,5	199,1	200,4	198,5	198,9	200,3	198,6	199,2
198,6	198,8	200,1	199,1	198,8	201,4	199,8	198,5

a) Organizar os resultados um uma tabela de frequência acumulada relativa e fazer o respectivo histograma.

b) Como diretor de produção avaliar os seguintes itens:

- Quantos % do total da amostra pesam menos do que 199g?
- Quantos % pesam mais do que 201g?
- Sabendo que o valor admitido é 200±1g como você avalia a sua produção?

Aula 9 – Medidas centrais

Nesta aula abordaremos o conceito de média aritmética e calcular a média para dados simples, agrupados sem intervalo e agrupado com intervalos.

9.1 Média aritmética

A média aritmética é uma das medidas de posição mais importante entre as medidas de tendência central, que recebem esta denominação pelo fato de os dados observados tenderem, em geral, a se agrupar em torno dos valores centrais.

Embora existam várias tipos de médias, tais como a ponderada, a geométrica, a harmônica, a quadrática, a cúbica, a biquadrática, iremos concentrar somente na média aritmética.

9.1.1 Média aritmética para dados não agrupados ou soltos

Representada por $\overline{x}$, é a medida central mais utilizada. Seu cálculo para dados soltos ou denominados de não agrupados é feito pela soma dos valores dividida pelo número deles. Genericamente é determinada pela expressão numérica:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Onde $\mathbf{x_i}$ indica o dado e **n** o número total de dados.

Exemplo 1: Determinar a média aritmética dos conjuntos de valores abaixo:

a) 70, 80, 120

$$\overline{x} = \frac{\sum x_i}{n} = \frac{70 + 80 + 120}{3} = \frac{270}{3} = 90$$

b) 5, 8, 10, 12, 15

$$\overline{x} = \frac{\sum x_i}{n} = \frac{5 + 8 + 10 + 12 + 15}{5} = \frac{40}{5} = 8$$

Logo, as médias aritméticas acima são 90 e 8.

Observação: O processo de cálculo da média aritmética é o mesmo, quer se trate de um conjunto de valores que traduzam representações amostrais, quer se trate de todos os valores de uma população. Temos então, Média amostral $\overline{x} = \frac{\sum x}{n}$ e média populacional $\mu = \frac{\sum x}{N}$.

Observe, no **quadro 9.1**, as vantagens e desvantagens da utilização da média aritmética:

Quadro 9.1: Vantagens e desvantagens da média aritmética

Vantagens	Desvantagens
Fácil de compreender e calcular.	É afetada por valores extremos.
Utiliza todos os valores da variável.	É necessário conhecer todos os valores da variável.
É um valor único.	
Fácil de incluir em equações matemáticas	

Fonte: CRESPO, 2007.

A média é utilizada quando **(1)** desejamos obter a medida de posição que possui maior estabilidade; e **(2)** houver necessidade de um tratamento algébrico ulterior.

Exemplo de aplicação: Em controle de qualidade, a média é utilizada para determinar se o processo está operando ao redor de um valor esperado, o alvo.

9.2 Média (aritmética) em distribuição de frequência

Para o cálculo da média aritmética, quando os dados estão agrupados em uma tabela, ou seja, em uma distribuição de frequência, procede-se um pouco diferente do cálculo para os dados soltos.

Também sabemos que a distribuição pode ou não conter intervalos. Frente a essas diferenças, veremos separadamente cada distribuição.

Dados agrupados sem intervalos de classe

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i}$$

Onde $\mathbf{x_i}$ $\mathbf{f_i}$ é a média aritmética ponderada pela respectiva frequência absoluta. Abaixo, veremos um exercício resolvido para entender melhor esse cálculo.

Exemplo: Considere a distribuição relativa de 34 famílias de 4 filhos, tomando como variável o número de filhos do sexo feminino.

Número de meninas (x_i)	Frequência (f_i)	$x_i f_i$
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
S	34	78

Calculando, $\overline{x} = \dfrac{\sum x_i f_i}{\sum f_i} = \dfrac{78}{34} = 2{,}29$ que arredondado podemos escrever 2,3. Sendo x uma variável quantitativa discreta, como interpretar o resultado obtido, 2 meninos e 3 décimos de menino? O valor médio 2,3 meninos sugere, neste caso, que o maior número de famílias tem 2 meninos e 2 meninas, sendo, porém, a tendência geral de uma leve superioridade numérica em relação ao número de meninos.

Dados agrupados com intervalos de classe

Nada mais é que uma tabela apresentada com intervalos. No cálculo da média aritmética, é necessário que você identifique o ponto médio. A fórmula que calcula a média é:

$$\overline{x} = \frac{\sum P_m f_i}{\sum f_i}$$

Onde P_m é o ponto médio da classe. Veremos como isso se aplica na prática através de um exemplo.

Considere a distribuição de estaturas de 40 alunos, que já foram distribuídos na tabela abaixo. Observe que essa distribuição é com intervalos de classe, devido ao número de elementos serem significativos (quantidade de dados 40 alunos).

Estatura (cm)	f_i
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
Σ	**40**

Para facilitar nosso cálculo, iremos construir novas colunas na tabela (tabela estendida) para determinar o ponto médio e o produto definido por $P_m f_i$.

Logo, a nova tabela fica assim:

Estatura (cm)	f_i	P_m	$P_m f_i$
150 ⊢ 154	4		
154 ⊢ 158	9		
158 ⊢ 162	11		
162 ⊢ 166	8		
166 ⊢ 170	5		
170 ⊢ 174	3		
Σ	**40**		

Para encontrarmos o ponto médio, vamos relembrar o que estudamos na aula 7. Perceba que o primeiro ponto médio da 1ª Classe (linha) é 152. Dessa forma encontramos todos os pontos médios, e passamos para a nova tabela, observe:

Estatura (cm)	f_i	P_m	$P_m f_i$
150 ⊢ 154	4	152	
154 ⊢ 158	9	156	
158 ⊢ 162	11	160	
162 ⊢ 166	8	164	
166 ⊢ 170	5	168	
170 ⊢ 174	3	172	
Σ	**40**		

Elaborando o produto de $P_m f_i$ de cada classe, chegamos a completar a tabela:

Estatura (cm)	f_i	P_m	$P_m f_i$
150 ⊢ 154	4	152	608
154 ⊢ 158	9	156	1404
158 ⊢ 162	11	160	1760
162 ⊢ 166	8	164	1312
166 ⊢ 170	5	168	840
170 ⊢ 174	3	172	516
Σ	**40**		**6440**

Calculando, $\overline{x} = \frac{\sum P_m f_i}{\sum f_i} = \frac{6.440}{40} = 161$, sendo assim a média das alturas dos alunos é de 161 cm.

Resumo

Nesta aula você conheceu a média aritmética, e aprendeu a determinar o seu cálculo nos três modos de distribuição.

Atividades de aprendizagem

1. Suponha que uma empresa está realizando uma pesquisa para saber qual é a média dos salários pagos aos seus funcionários. Para facilitar, o setor de RH, enviou a você uma planilha no formato de uma distribuição de frequência com intervalos dos respectivos níveis salariais e a quantidade de funcionários. Veja:

Salários (R$)	f_i
240 ⊢ 480	15
480 ⊢ 720	22
720 ⊢ 960	30
960 ⊢ 1200	18
1200 ⊢ 1440	15
Σ	**100**

Podemos observar que 15 funcionários recebem salários entre R$240,00 e R$480,00. Sabemos também que estão sendo pesquisados 100 funcionários. Agora é só calcular a média. Então, mãos à obra!

Agora preencha e calcule a tabela estendida abaixo:

Salários (R$)	f_i	P_m	$P_m f_i$
240 ⊢ 480	15		
480 ⊢ 720	22		
720 ⊢ 960	30		
960 ⊢ 1200	18		
1200 ⊢ 1440	15		
Σ	**100**		

Com base nestes resultados, determine o salário médio desses funcionários.

2. Um produto é vendido em três supermercados por R$13,00/kg, R$13,20/kg e R$13,50/kg. Determine o valor por quilo que se paga em média pelo produto.

3. O salário de 40 funcionários de um escritório está distribuído segundo o quadro abaixo. Calcule o salário médio destes funcionários.

Salário (R$)	Número de funcionários fi
400 ⊢ 500	12
500 ⊢ 600	15
600 ⊢ 700	8
700 ⊢ 800	3
800 ⊢ 900	1
900 ⊢ 1000	1
Total	**40**

4. Uma imobiliária gerencia o aluguel de residências particulares segundo o quadro abaixo. Calcule o valor médio do aluguel por residência.

Aluguel (R$)	Número de casas fi
0 ⊢ 200	30
200 ⊢ 400	52
400 ⊢ 600	28
600 ⊢ 800	7
800 ⊢ 1000	3
Total	120

5. Calcule o número médio de acidentes por dia em uma determinada esquina.

Número de acidentes por dia xi	Número de dias fi
0	30
1	5
2	3
3	1
4	1
Total	**50**

Anotações

Aula 10 – Mediana

Nesta aula abordaremos o conceito de mediana e calcularemos os tipos de mediana para dados simples, agrupados sem intervalo e agrupado com intervalos.

10.1 Mediana

Mediana é uma medida que divide um conjunto de dados em duas partes iguais. Em outras palavras, a mediana de um conjunto de valores, ordenados segundo uma ordem de grandeza, é o valor situado no meio, de tal forma, que separa em dois subconjuntos de mesmo número de elementos. Podemos representá-la por $M_{d.}$

10.1.1Determinações da Mediana (M_d) para dados não agrupados ou simples

Para determinar a mediana de dados soltos, ou seja, não agrupados, temos que primeiramente encontrar o rol desses valores. Você lembra o que era Rol? É a ordenação dos valores em ordem crescente (do menor para o maior). Em seguida precisamos saber se a quantidade de dados é par ou ímpar, pois para cada quantidade existe um procedimento. Vejamos!

Se a quantidade de valores n for um **número ímpar,** a mediana será o valor que se encontra no meio ou podemos determinar esse valor utilizando a expressão da variável situada na posição $\frac{(n+1)}{2}$; se a quantidade de valores n for um **número par,** a mediana será igual ao resultado de dividir por dois a soma dos valores das posições $\frac{n}{2}$ e $\frac{n}{2} + 1$.

Veja o exemplo:

Calcular a mediana dos conjuntos de dados abaixo:

a) 5, 8, 6

b) 7, 8, 9, 10

Resolução:

Na alternativa (a), a quantidade de elementos é ímpar, então colocamos em ordem crescente: 5, 6, 8. Agora é só pegar o termo do meio, que é o 6, logo a $M_d = 6$.

Poderíamos calcular também pela expressão $\frac{(n+1)}{2}$, sendo **n** o número de elementos dos dados, temos: n = 3, substituindo fica, $\frac{(3+1)}{2} = \frac{4}{2} = 2$, ou seja, a mediana é o elemento que ocupa a 2ª posição, que nesse caso é o 6.

Mas lembre-se que você também precisa achar o rol por esse processo.

Já na alternativa (b), a quantidade de elementos é par, devemos colocar em ordem, fazendo o rol e depois calcular a média dos dois termos centrais, pela expressão $\frac{n}{2}$ e $\frac{n}{2} + 1$. Observe abaixo:

Como temos n = 4, pois são quatro termos (7, 8, 9 e 10), substituímos na expressão, $\frac{4}{2}$ e $\frac{4}{2} + 1$, o resultado de $\frac{4}{2}$ é 2 e de $\frac{4}{2} + 1$ é 3 (2+1). Logo, os termos centrais do rol é o segundo (valor 8) e o terceiro termo (valor 9), que podemos calcular a média, fazendo $\frac{(8+9)}{2} = \frac{17}{2} = 8{,}5$, portanto a média aritmética é 8,5. Viram como é fácil?

A mediana é uma medida resistente, pois está relacionada apenas com a ordem dos valores da variável. Em outras palavras, não é sensível a valores que se encontram nos extremos (início e fim).

10.1.2 Determinações da Mediana (M_d) para dados agrupados

Para a determinação da mediana, quando os dados estão agrupados, fazemos uso de fórmulas que facilitam nosso cálculo, seja ele com ou sem intervalo.

Vejamos! O primeiro elemento é a frequência acumulada que precisamos determinar. Essa determinação pode ser feita inserindo na tabela a coluna de Fa. Como fizemos para os tipos de frequências vistos na aula oito. O segundo elemento é a classe em que a mediana se encontra. Lembre-se que classe é a mesma coisa que linha, ou seja, em uma linha da tabela possivelmente deve estar a mediana.

Para fazer essa determinação ou localização da classe da frequência acumulada que se encontra a mediana. Devemos dividir a distribuição em dois grupos, onde cada grupo (ou parte) tenha o mesmo número de elementos.

Esse valor é determinado pela fórmula, $\sum \frac{f_i}{2}$ onde temos a soma total das frequências absolutas divididas por 2 (divisão pel o meio). Para o caso de uma distribuição, porém, a ordem, a partir de qualquer um dos extremos, é dada por $\sum \frac{f_i}{2}$. Vamos analisar cada caso especifico de mediana com distribuições de frequências logo abaixo, utilizando exemplos para facilitar a sua compreensão.

10.1.3 Determinação da mediana para dados agrupados sem intervalos de classe

Calculamos a mediana para dados agrupados sem intervalos de classe quando os dados se tratam de variável discreta, contendo uma pequena variação, sendo cada valor, tomado como um único intervalo de classe, o que na estatística chama-se de intervalo degenerado.

Essa distribuição não contém intervalos de classe, como você pode observar na figura abaixo.

Distribuição de frequência genérica:

x_i	fr_i
x_1 x_2 . . . x_n	f_1 f_2 . . . f_n
	$\sum f_i = n$

Vamos determinar à mediana através do exemplo abaixo:

Partimos do pressuposto que já temos a distribuição de frequência sem intervalo, com variáveis (x_i) da quantidade de cômodos de casas ocupadas por vinte famílias (f_i) conforme abaixo:

Distribuição de frequência das famílias entrevistadas:

x_i	f_i
2	4
3	7
4	5
5	2
6	1
7	1
	$\sum = 20$

Determinamos as frequências acumuladas dessa distribuição, acrescentando uma nova coluna à tabela original. Observe:

x_i	f_i	F_i
2	4	4
3	7	11
4	5	16
5	2	18
6	1	19
7	1	20
	$\sum = 20$	

$\sum f_i = 20$, logo temos $\dfrac{20}{2} = 10$, perceba que acabamos de dividir os dados em duas partes de 10 elementos cada. Já tínhamos mencionado que isso aconteceria.

Feito isso, basta identificar a frequência acumulada imediatamente superior à metade dessa soma das frequências. Portanto:

x_i	f_i	F_i
2	4	4
3	7	11
4	5	16
5	2	18
6	1	19
7	1	20
	$\sum = 20$	

Frequência imediatamente superior a 10

Chamada de Classe Mediana

Identificada a classe mediana, a variável que se encontra nessa classe será nossa mediana, vejamos:

x_i	f_i	F_i
2	4	4
3	7	11
4	5	16
5	2	18
6	1	19
7	1	20
	$\sum = 20$	

Variável da classe mediana, que é a mediana = 3

Ou seja, a mediana dessa distribuição de frequência sem intervalo é md = 3

10.1.4 Determinação da mediana para dados agrupados com intervalos de classe

Neste caso, trata-se de distribuição com intervalos, e o problema consiste em determinar o ponto do intervalo em que está a mediana. Para tanto, temos inicialmente que determinar a classe mediana (linha da tabela) na qual se acha à mediana. Para facilitar nosso trabalho, vamos seguir os itens abaixo:

- Primeiramente determinam-se as frequências acumuladas da distribuição;
- Em segundo lugar, calcula-se $\sum \frac{f_i}{2}$;
- Em terceiro lugar, marca-se a classe correspondente à frequência acumulada imediatamente superior a $\sum \frac{f_i}{2}$.

Isso já foi visto por você quando determinamos a média no item 10.2.1. Após localizar a classe mediana, utiliza-se a fórmula abaixo para a determinação da mediana para dados agrupados com intervalos de classe.

$$M_d = l^* + \frac{\left[\frac{\sum f_i}{2} - F\,(ant)\right] . \, h^*}{f^*}$$

Onde:

$l^* \rightarrow$ limite inferior da classe mediana

$F\,(ant) \rightarrow$ frequência acumulada da classe anterior à classe mediana

$f^* \rightarrow$ frequência simples da classe mediana

$h^* \rightarrow$ amplitude do intervalo da classe mediana

Exemplo: Vamos calcular a mediana do comprimento de 40 peças de cilindro produzidas por um torneiro mecânico.

Comprimento (cm)	f_i
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
Σ	**40**

Encontramos a frequência acumulada, a nova tabela fica assim:

Comprimento (cm)	f_i	F_i
150 ⊢ 154	4	4
154 ⊢ 158	9	13
158 ⊢ 162	11	24
162 ⊢ 166	8	32
166 ⊢ 170	5	37
170 ⊢ 174	3	40
Σ	**40**	

Identificamos a classe mediana por $\sum \frac{f_i}{2} = \frac{40}{2} = 20$, e aí utilizamos a 3ª classe (ou terceira linha);

Comprimento(cm)	f_i	F_i	
150 ⊢ 154	4	4	
154 ⊢ 158	9	13	
158 ⊢ 162	**11**	**24**	← **Classe mediana**
162 ⊢ 166	8	32	
166 ⊢ 170	5	37	
170 ⊢ 174	3	40	
Σ	**40**		

Da classe mediana acima, devemos determinar os valores que compõe a fórmula $M_d = l^* + \frac{\left[\frac{\sum f_i}{2} - F\,(ant)\right] . h^*}{f^*}$, para calcular a mediana. Para facilitar, identificamos separadamente cada item da fórmula. Veja: $\sum \frac{f_i}{2} = \frac{40}{2} = 20$

$l^* \rightarrow 158$

$F\ (ant) \rightarrow 13$

$f^* \rightarrow 11$

$h^* \rightarrow 4$

Fazendo a substituição, temos:

$$M_d = 158 + \frac{[20 - 13].\ 4}{11}$$

$$M_d = 158 + \frac{[7].\ 4}{11}$$

$$M_d = 158 + \frac{28}{11}$$

$$M_d = 158 + 2{,}55$$

$$M_d = 160{,}55$$

Portanto, a mediana do comprimento dos eixos é de 160,55cm.

Quadro 10.1: Vantagens e desvantagens da utilização da mediana

Vantagens	Desvantagens
Fácil de calcular	Difícil de incluir em equações matemáticas
Não é afetada pelos valores extremos	Não utiliza todos os valores da variável
É um valor único	

Fonte: Crespo, 2007.

Empregamos a mediana quando:

- Desejamos obter o ponto que divide a distribuição em partes iguais;
- Há valores extremos que afetam de maneira acentuada a média;
- A variável em estudo é salário, renda anual e valores de bens.

Resumo

Nesta aula você viu como calcular a mediana para dados agrupados. Aprendeu as vantagens e desvantagens de utilizar essa medida de tendência central.

Atividades de aprendizagem

a) Um produto é vendido em três supermercados por R$13,00/kg, R$13,20/kg e R$13,50/kg. Determine quantos R$/kg se encontra na mediana do pagamento pelo produto.

b) O salário de 40 funcionários de um escritório está distribuído segundo o quadro abaixo. Calcule o salário mediano destes funcionários.

Salário (R$)	Número de funcionários f_i
400 ⊢ 500	12
500 ⊢ 600	15
600 ⊢ 700	8
700 ⊢ 800	3
800 ⊢ 900	1
900 ⊢ 1000	1
Total	40

c) Uma imobiliária gerencia o aluguel de residências particulares segundo o quadro abaixo. Calcule a mediana do aluguel das residências.

Aluguel (R$)	Número de casas f_i
0 ⊢ 200	30
200 ⊢ 400	52
400 ⊢ 600	28
600 ⊢ 800	7
800 ⊢ 1000	3
Total	120

d) Calcule a mediana de acidentes por dia em uma determinada esquina.

Número de acidentes por dia x_i	Número de dias f_i
0	30
1	5
2	3
3	1
4	1
Total	50

Anotações

Aula 11 – Medida central

Nessa aula conheceremos a terceira medida de posição central determinada como MODA. Veremos também como calculá-la em uma distribuição.

11.1 Moda M_o

Ela é utilizada quando desejamos obter uma medida rápida e aproximada de posição, e essa medida de posição deve ser o valor mais típico da distribuição.

Conhecendo a Moda:

Segundo Filho (2009), foi Karl Pearson quem introduziu na estatística, o conceito de moda. Ela surgiu por volta do século XIX e foi definida como sendo o valor ou valores que ocorrem com maior frequência em uma distribuição.

Logo, podemos definir a moda como sendo o valor (resultado numérico) que mais aparece na frequência nos dados de uma pesquisa. Por exemplo, na distribuição do consumo de um mesmo produto com diferentes apresentações, a moda mostra a apresentação mais consumida.

Na linguagem coloquial, as mulheres sabem perfeitamente o que isso significa! Pois a moda é algo que está em evidência, ou seja, algo que se vê bastante.

Se aparecer uma moda, dizemos que ela é unimodal, se aparecerem duas, dizemos que ela é bimodal, se forem três, trimodal e se aparecerem diversas podemos dizer que é plurimodal. Pode ser também que um conjunto de dados não exista moda, e aí dizemos que esse conjunto de dados é amodal, ou seja, quando todos os valores das variáveis apresentarem uma mesma frequência.

Fique atento para o fato de que a Moda é o elemento do conjunto que mais se repete, e não o número de vezes que ele aparece.

11.1.1 Determinação da Moda, para dados não agrupados

Como já dissemos, a moda é o valor que ocorre com maior frequência num conjunto de dados. Logo, fica fácil identificá-la. Vejamos o exemplo:

Calcular a moda dos conjuntos de valores abaixo:

a) 10, 10, 8, 6, 10

Como o número que mais aparece é o 10, a moda é 10.

b) 3, 5, 6, 7, 9

Nesse conjunto de dados, não temos moda, pois todos os elementos têm a mesma frequência.

Assim como a média e a mediana, a moda – em uma distribuição de frequência – também têm dois modos de dados agrupados para determiná-los, ou seja, com ou sem intervalos como veremos nos próximos itens.

11.1.2 Determinação da Moda para dados agrupados sem intervalos de classe

A moda é determinada pela maior frequência da distribuição. Observe o exemplo de uma distribuição de frequência sem intervalos que relaciona o número de meninas.

Número de meninas	f_i
0	2
1	6
2	10
3	12
4	4
Σ	**34**

Podemos observar que 12 é a maior frequência. Essa frequência que determina a classe modal (linha que se encontra a moda). Veja.

Número de meninas	f_i
0	2
1	6
2	10
3	12
4	4
Σ	**34**

Moda

Classe Modal

Logo, a Moda é 3.

11.2 Dados agrupados com intervalos de classe

Podemos utilizar vários modos de calcular a moda quando se tem intervalos em uma distribuição. Os mais usuais em estatística são os métodos de King e Czuber. Para facilitar, veremos a moda pelo método de King, porém esse assunto não termina aqui.

11.2.1 Método de King

$$M_o = li + \left(\frac{f_{post}}{f_{ant} + f_{post}} \right) . A$$

Onde:

li → limite inferior da classe modal

f_{ant} → frequência da classe anterior a classe modal

f_{post} → frequência da classe posterior a classe modal

A → amplitude da classe modal

Você aprendeu como calcular a moda, para fixação observe o exemplo seguinte.

A tabela abaixo indica a distribuição dos salários dos funcionários da empresa Y de logística, em reais, no ano de 2001. Vamos calcular a moda e analisar o resultado.

Salários (R$)	f_i
500 ├── 700	90
700 ├── 900	54
900 ├── 1100	34
1100 ├── 1300	11
1300 ├── 1500	4
Σ	

Primeiramente identificamos a maior frequência da distribuição, que nesse caso é 90. Logo, a classe modal é a primeira linha, ou seja, é nessa linha que iremos retirar os dados para aplicar na fórmula de King. Veja:

Salários (R$)	f_i
500 ⊢ 700	90
700 ⊢ 900	54
900 ⊢ 1100	34
1100 ⊢ 1300	11
1300 ⊢ 1500	4
Σ	

Logo:

$li \rightarrow 500$

$f_{ant} \rightarrow 0$

$f_{post} \rightarrow 54$

$A \rightarrow 200$

Substituindo na fórmula, temos:

$$M_o = 500 + \left(\frac{54}{0 + 54}\right) . 200$$

$$M_o = 500 + \left(\frac{54}{54}\right) . 200$$

$$M_o = 500 + (1) . 200$$

$$M_o = 500 + 200$$

$$M_o = 700$$

Tabela 11.1: Vantagens e desvantagens da utilização da moda

Vantagens	Desvantagens
Fácil de calcular	Pode estar afastada do centro dos valores
Não é afetada pelos valores extremos	Não utiliza todos os valores da variável
É um valor único	Difícil de incluir em equações matemáticas
	A variável pode ter mais de uma moda
	Algumas variáveis não têm moda

Fonte: Crespo (2007)

Resumo

Moda é a frequência que mais aparece em uma distribuição de frequência. Ela pode ser unimodal, amodal, bimodal e polimodal. Utilizamos o método de King para calculá-la.

Atividades de aprendizagem

1. Calcular a moda para os exercícios listados no item 5.1.1. Média. (Fonte: Adaptado da prova de AFRF 2002).

O atributo do tipo contínuo X, observado como um inteiro, numa amostra de tamanho 100, obtida de uma população de 1.000 indivíduos, produziu a seguinte tabela de frequência:

Xi	Frequência (f)
29,5 ├── 39,5	4
39,5 ├── 49,5	8
49,5 ├── 59,5	14
59,5 ├── 69,5	20
69,5 ├── 79,5	26
79,5 ├── 89,5	18
89,5 ├── 99,5	10

Calcule o valor modal do atributo X, pelo método de King.

2. Os dados seguintes, ordenados do menor para o maior, foram obtidos de uma amostra aleatória, de 50 preços (Xi) de ações, tomadas numa bolsa de valores internacional. A unidade monetária é o dólar americano. (Adaptado da prova AFRF (1998)

4, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 12, 12, 13, 13, 14, 15, 15, 15, 16, 16, 18, 23

Com base nestes dados, determine o preço modal.

Para cada um dos conjuntos abaixo, determine o valor da Moda, das duas maneiras distintas, utilizando o método de King.

3. Trabalhe a Distribuição abaixo:

Xi	fi
0 ⊢ 10	3
10 ⊢ 20	5
20 ⊢ 30	8
30 ⊢ 40	4
40 ⊢ 50	2

4. Trabalhe a Distribuição abaixo:

Xi	fi
0 ⊢ 15	4
15 ⊢ 30	7
30 ⊢ 45	11
45 ⊢ 60	9
60 ⊢ 75	5
75 ⊢ 90	2

5. Trabalhe a Distribuição abaixo:

Xi	fi
0	7
7	11
14	15
21	9
28	3

Aula 12 – Separatrizes

Nessa aula iremos estudar os quartis, que dividem um conjunto de dados em 4 partes iguais.

12.1 Os Quartis

São valores que ocupam lugares específicos em uma série ordenada. Outro fator importante, é que o seu cálculo é semelhante ao da mediana, substituindo na fórmula somente a posição do elemento a ser estudado.

Porém vamos acompanhar a definição: denominamos Quartis os valores de uma série que a dividem em quatro partes iguais uma série de dados.

Portanto, há três quartis importantes a se determinar:

- o primeiro quartil (Q1) – valor situado na primeira parte da divisão, de tal modo, que na série representa um quarto da parte (25%) dos dados. É a menor das partes, já que as outras somadas representam (75%) dos dados.

- o segundo quartil (Q2) –é a própria (coincide) mediana. O valor encontrado divide a série na metade (50%). Podemos afirmar que 50% dos dados são menores do que ele e a outra metade restante (50%) são maiores.

- o terceiro quartil (Q3) – valor situado de tal modo na série que as três quartas partes (75%) dos dados são menores do que ele e a quarta parte restante (25%) são maiores.

Quando os dados são agrupados, e queremos determinar os quartis, usamos a mesma técnica do cálculo da mediana, bastando substituir, na fórmula da mediana, $\Sigma\frac{f_i}{2}$, por $k\Sigma\frac{f_i}{4}$ (onde k é o número de ordem do quartil).

Assim, temos:

$$Q_1 = l^* + \frac{\left[\Sigma\frac{f_i}{4} - F\,(ant)\right] . h^*}{f^*}$$

$$Q_2 = l^* + \frac{\left[\sum \frac{f_i}{2} - F\,(ant)\right] . h^*}{f^*}$$

$$Q_3 = l^* + \frac{\left[3\sum \frac{f_i}{4} - F\,(ant)\right] . h^*}{f^*}$$

Observe o exercício resolvido.

Dada a distribuição abaixo, calcular Q1, Q2 e Q3.

Classes	fi
7 ⊢ 17	6
17 ⊢ 27	15
27 ⊢ 37	20
37 ⊢ 47	10
47 ⊢ 57	5
Total	56

Calculando o primeiro quartil, pela expressão $Q_1 = l^* + \frac{\left[3\sum \frac{f_i}{4} - F\,(ant)\right] . h^*}{f^*}$, temos que encontrar qual é a classe (linha da distribuição) que iremos retirar os dados. Para isso, lembre-se que calculávamos a classe em que poderia estar a mediana por $\sum \frac{f_i}{2}$ na frequência acumulada. Para determinar o quartil usamos $k\sum \frac{f_i}{4}$ (onde k é o número de ordem do quartil). Então temos:

Classes	fi	F
7 ⊢ 17	6	6
17 ⊢ 27	15	21
27 ⊢ 37	20	41
37 ⊢ 47	10	51
47 ⊢ 57	5	56
Total	56	

Logo:

$k\sum \frac{f_i}{4} = k\sum \frac{56}{4} = 14$, que está na segunda classe, ou linha.

$$Q_1 = l^* + \frac{\left[\sum \frac{f_i}{4} - F\,(ant)\right] . h^*}{f^*}$$

Substituindo na fórmula temos:

$$Q_1 = 17 + \frac{[14 - 6] \cdot 10}{15}$$

$$Q_1 = 17 + \frac{[8] \cdot 10}{15}$$

$$Q_1 = 17 + \frac{80}{15}$$

$$Q_1 = 17 + 5{,}333$$

$$Q_1 = 22{,}333$$

Para o segundo quartil ou mediana, observe o cálculo da classe:

$$k\sum \frac{f_i}{2} = k\sum \frac{56}{2} = 28$$

Q_2, temos:

$$Q_2 = 27 + \frac{[28 - 21] \cdot 10}{20}$$

$$Q_2 = 27 + \frac{[7] \cdot 10}{20}$$

$$Q_2 = 27 + \frac{70}{20}$$

$$Q_2 = 27 + 3{,}5$$

$$Q_2 = 30{,}5$$

E para finalizar o terceiro quartil:

$$k\sum 3 \frac{f_i}{4} = k\sum 3 \frac{56}{4} = 42$$

Logo :

Q_3, temos:

$$Q_3 = 37 + \frac{[42 - 41] \cdot 10}{10}$$

$$Q_3 = 37 + \frac{[1] \cdot 10}{10}$$

$$Q_3 = 37 + \frac{10}{10}$$

$$Q_3 = 37 + 1$$

$$Q_3 = 38$$

Portanto:

$Q_1 = 32{,}333$; $Q_2 = 30{,}5$ e $Q_3 = 38$

Resumo

Nesta aula, você aprendeu que Quartil é a separatriz que divide um conjunto de dados em quatro partes iguais.

Anotações

Aula 13 – Separatrizes II

Nessa aula iremos aprender como calcular o decil.

13.1 O Decil

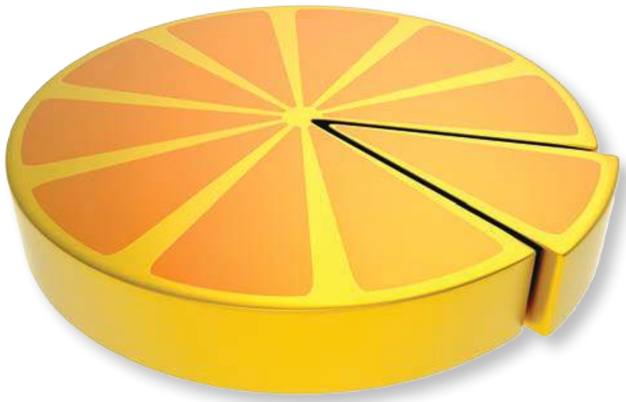


Figura 13.1: Decil
Fonte: ©maksymmo/Shutterstock

Você deve estar se perguntando quando que eu uso o Decil na prática. Dividimos um conjunto de dados em dez partes iguais. Encontra-se o valor do decil procedendo-se como no caso dos quartis. Eles também são valores que ocupam lugares específicos em uma série ordenada. Outro fator importante, é que o seu cálculo é idêntico ao do quartil, trocando apenas o fator $k\Sigma\frac{f_i}{10}$.

Assim, temos o primeiro decil determinado por:

$$D_1 = l^* + \frac{\left[\Sigma\frac{f_i}{10} - F\,(ant)\right].\,h^*}{f^*}$$

E o quinto decil que é idêntico ao segundo quartil, que é a mediana dos dados.

$$D_5 = l^* + \frac{\left[\Sigma\, 5\frac{f_i}{10} - F\,(ant)\right].\,h^*}{f^*} = Q_2 = l^* + \frac{\left[\Sigma\frac{f_i}{2} - F\,(ant)\right].\,h^*}{f^*}$$

Você deve estar se perguntando e se eu quisesse calcular o 2º, 7º ou o 9º decil?

Bastaria substituir o elemento $k\Sigma\frac{f_i}{10}$ da fórmula por $k\Sigma\, 2\frac{f_i}{10}$, ou $k\Sigma 7\frac{f_i}{10}$, ou $k\Sigma 9\frac{f_i}{10}$.

Exercício resolvido:

Dada a distribuição abaixo, calcule D1 e D8.

Classes	fi
7 ⊢ 17	6
17 ⊢ 27	15
27 ⊢ 37	20
37 ⊢ 47	10
47 ⊢ 57	5
Total	56

Calculando o primeiro decil, pela expressão $D_1 = l^* + \dfrac{\left[\sum \dfrac{f_i}{10} - F\,(\text{ant})\right].\, h^*}{f^*}$,

temos que achar qual é a classe (linha da distribuição) que iremos retirar os dados. Calculando a classe por $\sum \dfrac{f_i}{10}$ na frequência acumulada, para determinar qual a linha da distribuição que iremos utilizar. Teremos:

Classes	fi	F
7 ⊢ 17	6	6
17 ⊢ 27	15	21
27 ⊢ 37	20	41
37 ⊢ 47	10	51
47 ⊢ 57	5	56
Total	56	

Logo,

$k\sum \dfrac{f_i}{10} = k\sum \dfrac{56}{10} = 5{,}6$, que esta na primeira classe (linha).

Substituindo na fórmula $Q_1 = l^* + \dfrac{\left[\sum \dfrac{f_i}{10} - F\,(\text{ant})\right].\, h^*}{f^*}$ temos:

$Q_1 = 7 + \dfrac{[5{,}6 - 0].\,10}{6}$

$Q_1 = 7 + \dfrac{[5{,}6].\,10}{6}$

$Q_1 = 7 + \dfrac{56}{6}$

$Q_1 = 7 + 9{,}333$

$Q_1 = 16{,}333$

Assim temos o primeiro decil da distribuição.

Para encontrar os demais decil, devemos proceder do mesmo modo, porém, sempre trocando o numerador do fator.

Calculando o 8º decil.

$$k\sum 8\frac{f_i}{10} = k\sum 8\frac{56}{10} = 44{,}8$$

D_8, temos

$$D_8 = 37 + \frac{[44{,}8 - 41] \,.\, 10}{10}$$

$$D_8 = 37 + \frac{[3{,}8] \,.\, 10}{10}$$

$$D_8 = 37 + \frac{38}{10}$$

$$D_8 = 37 + 3{,}85$$

$$D_8 = 40{,}8$$

Resumo

O decil divide em dez partes iguais os dados de uma distribuição. Vimos também como calcular os decis para dados agrupados com intervalo.

Anotações

Aula 14 – Separatrizes III

Nessa aula você saberá o que é percentil, como também aprenderá a calcular o percentil.

14.1 O Percentil

Denominamos **percentis** os noventa e nove valores que separam uma série em 100 partes iguais.

Para determinarmos os percentis usamos a mesma técnica do cálculo da mediana, bastando substituir, na fórmula da mediana, $\frac{\Sigma f_i}{2}$, por k $\frac{\Sigma f_i}{100}$ (onde k é o número de ordem do percentil).

Assim, para determinar o 27º (vigésimo sétimo) percentil, utilizamos a fórmula:

$$P_{27} = l^* + \frac{\left[27 . \frac{\Sigma f_i}{100} - F\ (ant)\right] . h^*}{f^*}$$

Para calcular o 91º (nonagésimo primeiro) percentil, basta substituir o valor 27.

Vamos visualizar através de um exemplo:

Suponhamos que quiséssemos calcular o 55º percentil da distribuição da aula 13.

Classes	fi
7 ⊢— 17	6
17 ⊢— 27	15
27 ⊢— 37	20
37 ⊢— 47	10
47 ⊢— 57	5
Total	56

Já sabemos que temos que calcular a frequência acumulada e a classe em que se encontra o 55º percentil. A tabela abaixo mostra como fica esse cálculo:

Classes	fi	F
7 ⊢ 17	6	6
17 ⊢ 27	15	21
27 ⊢ 37	20	41
37 ⊢ 47	10	51
47 ⊢ 57	5	56
Total	56	

A classe é calculada por $k\Sigma \frac{f_i}{100} = 55\Sigma \frac{56}{100} = 55 \,.\, 0{,}56 = 30{,}8$, que está no intervalo da 3º (terceira) classe (linha).

Substituindo na fórmula $P_{55} = l^* + \frac{\left[55 \,.\, \frac{\Sigma f_i}{100} - F\,(ant)\right] .\, h^*}{f^*}$ temos:

$$P_{55} = 27 + \frac{[30{,}8 - 21] \,.\, 10}{20}$$

$$P_{55} = 27 + \frac{[9{,}8] \,.\, 10}{20}$$

$$P_{55} = 27 + \frac{98}{20}$$

$$P_{55} = 27 + 4{,}9$$

$$P_{55} = 31{,}9$$

Portanto, o percentil 55 está com 31,9 dos elementos da distribuição.

Resumo

Você aprendeu que o percentil divide em cem partes iguais os dados de uma distribuição. Fazemos o cálculo de qualquer decil utilizando a fórmula:

$$P_{nn} = l^* + \frac{\left[nn.\, \frac{\Sigma f_i}{100} - F\,(ant)\right] .\, h^*}{f^*}$$

Atividades de aprendizagem

1. Calcule a média, a mediana, a moda, o 3º quartil e o 30º percentil para as seguintes distribuições:

a) As companhias de seguro pesquisam continuamente as idades e as causas de morte para ajustar as atualizações de suas apólices. Os dados se baseiam no estudo levado a efeito pela revista Time sobre as pessoas que morreram vitimadas por armas de fogo durante uma semana.Com base na tabela abaixo calcule o que se pede:

Idade na morte	Frequência
16 ⊢ 26	22
26 ⊢ 36	10
36 ⊢ 46	6
46 ⊢ 56	2
56 ⊢ 66	4
66 ⊢ 76	5
76 ⊢ 86	1

b) A distribuição de frequência é referente à quantidade de locação de casas pelo seu respectivo preço de aluguel, de uma determinado região de Curitiba. Calcule a média, mediana, moda, o terceiro quartil e 30 percentil.

Aluguel (x R$100)	Quantidade de casas
4 ⊢ 6	18
6 ⊢ 8	25
8 ⊢ 10	32
10 ⊢ 12	40
12 ⊢ 14	30
14 ⊢ 16	18
16 ⊢ 18	12

2. Das distribuições abaixo, calcule: média, mediana e moda.

- primeiro e terceiro quartis
- 10º, 23º e 90º percentis

a)

Notas	f_i
0 ⊢ 2	5
2 ⊢ 4	8
4 ⊢ 6	14
6 ⊢ 8	10
8 ⊢ 10	7
	44

b)

Estaturas	f_i
150 ⊢ 158	5
158 ⊢ 166	12
166 ⊢ 174	18
174 ⊢ 182	27
182 ⊢ 190	8
	70

3. Elabore o histograma de frequência simples, e localize as medidas calculadas nos exercícios acima.

Aula 15 – Medidas de dispersão

Nesta aula estudaremos a importância das medidas de dispersão, amplitude e desvio médio, bem como o seu cálculo.

15.1 Medidas de afastamento

As medidas de dispersão ou de afastamento são medidas estatísticas utilizadas para verificar o quanto os valores encontrados numa pesquisa estão dispersos ou afastados em relação à média ou em relação à mediana.

Em muitas situações os cálculos determinados pelas medidas de tendência central, ou a determinação de um valor específico de um conjunto qualquer analisado, não é suficiente para caracterizar uma distribuição.

Imagine a produção semanal de uma fábrica, baseada no quadro abaixo:

DIAS TURNOS	SEGUNDA Matutino	TERÇA Noturno	QUARTA	QUINTA	SEXTA	TOTAL	MÉDIA SEMANAL
I	150	150	150	150	150	750	150
II	70	130	150	180	220	750	150
III	15	67	117	251	300	750	150

Na produção média diária, não tem como identificar o grau de relacionamento entre as variáveis, já que a média semanal nos três turnos é idêntica a 150 peças por semana. Porém podemos criticar as variáveis em estudo, fazendo as seguintes perguntas: A produção é idêntica (ou homogênea)? A produção média semanal é suficiente para uma análise estatística? A produção diária dos três turnos é compatível com a produção média semanal?

Tais informações são obtidas através do estudo das medidas de dispersão. Elas permitem a análise de até que ponto estes valores apresentam oscilações para mais ou para menos, em relação a uma medida de posição fixada. Determina também o grau de homogeneidade ou heterogeneidade que existe entre os valores que compõem o conjunto.

Exemplificando:

Duas pessoas se submeteram a um teste:

Situação a) as duas pessoas tiraram nota 6,0;

Situação b) uma pessoa tirou 2,0 e a outra 10,0.

Nos dois casos a média é igual a 6,0. Todavia, na situação “a”, elas se concentraram sobre a média; e na situação “b” elas se dispersaram em torno da mesma. Isto quer dizer que a média é mais significativa em “a” do que em “b”. Ainda, em “a” existe uma homogeneidade nos conhecimentos adquiridos; em “b” heterogeneidade.

Para avaliar o grau de dispersão são utilizadas várias medidas, porém as mais utilizadas são:

- amplitude total ou intervalo total;
- desvio médio;
- variância;
- desvio padrão.

15.1 Amplitude total

Amplitude total ou intervalo total é a diferença entre o maior e o menor valor de uma série de dados. Para dados soltos ou não agrupados, é só calcular a diferença. Veja:

Exemplo:

No conjunto de números {4 , 6 , 8 , 9 , 12 , 17, 20}

A = 20 – 4

A = 16.

Para dados agrupados sem intervalo, utilizamos a expressão:

AT = x(máx) – x(mín)

Exemplo: Considere a distribuição de frequência sem intervalos abaixo:

X	f
1	3
2	7
3	13
4	6
5	2

A amplitude total é AT = 5 – 1 = 4

15.1.1 Amplitude total para dados agrupados com intervalo

Basta calcular a diferença entre o limite superior da última classe pelo limite inferior da primeira classe, vejamos:

AT = L(máx) – 1(mín)

Exemplo:

Dada a distribuição de frequência das alturas de 70 pessoas, conforme abaixo:

Estaturas	f_i
150 ⊢ 158	5
158 ⊢ 166	12
166 ⊢ 174	18
174 ⊢ 182	27
182 ⊢ 190	8
	70

Podemos calcular a amplitude total:

AT = L(máx) – 1(mín)

AT = 190 – 150

AT = 40

Como vimos, o intervalo ou amplitude total é uma medida fácil de calcular. Todavia, é instável, pois considera somente os valores extremos, não sendo afetada pela dispersão dos valores internos. É apenas uma indicação aproximada da dispersão.

15.2 Desvio médio

É a dispersão calculada em relação a todos os valores, sem exceção. É calculado através da fórmula:

$$D_m = \frac{\Sigma|x - \overline{X}| \cdot f}{n}$$

Onde:

$X \rightarrow$ é a variável em observação

$n \rightarrow$ é o número total de elementos em observação

$f \rightarrow$ é a frequência

$\overline{X} \rightarrow$ é a média aritmética.

Exemplo:

Calcular o desvio médio do grupo de valores: 4, 6, 8, 9, 10, 11.

1º) Cálculo da média:

$$\overline{X} = \frac{4 + 6 + 8 + 9 + 10 + 11}{6} \rightarrow \overline{X} = \frac{48}{6} \rightarrow \overline{X} = 8$$

2º) Cálculo dos módulos dos desvios em relação à média:

X	X-X	IX-XI
4	4 – 8 = - 4	4
6	6 – 8 = - 2	2
8	8 – 8 = 0	0
9	9 – 8 = 1	1
10	10 – 8 = 2	2
11	11 – 8 = 3	3
∑	0	12

3º) Cálculo do desvio médio:

$$D_m = \frac{\Sigma|x - \overline{X}| \cdot f}{n} \rightarrow D_m = \frac{12}{6} \rightarrow D_m = 2$$

Para calcular o desvio médio quando os dados estão agrupados em uma distribuição de frequência com intervalos de classe, devemos substituir o X da fórmula pelo ponto médio do intervalo de classe.

Resumo

- As medidas de dispersão ou de afastamento são medidas utilizadas para verificar o quanto os valores encontrados numa pesquisa estão dispersos ou afastados em relação à média ou em relação à mediana.

- Desvio médio é uma medida da dispersão dos dados em relação à média de uma sequência, ou seja, é o "afastamento" dos dados em relação a essa média.

Anotações

Aula 16 – Variância e desvio padrão

Nessa aula, você conhecerá as medidas de dispersão mais utilizadas. Irá aprender como calcular a variância e o desvio padrão, para dados agrupados.

16.1 Variância e desvio padrão

A variância e o desvio padrão são índices de variabilidade estáveis, pois leva em consideração a totalidade dos valores da variável em estudo, evitando assim falhas e tornando-se métodos mais utilizados.

A **variância** baseia-se nos desvios em torno da média aritmética, porém determinando a **média aritmética dos quadrados dos desvios**. Ela é representada por s^2, na fórmula abaixo:

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{\sum f_i} = \frac{\sum (x_i - \overline{x})^2}{n}$$

Quando nosso interesse não se restringe à descrição dos dados, mas, partindo da amostra, visamos tirar inferências válidas para a respectiva população, convém efetuar uma modificação, que consiste em usar o divisor n-1 em lugar de n.

A variância é uma medida que tem pouca importância como estatística descritiva, uma vez que sua unidade de medida é o quadrado da unidade de medida dos valores da variável. Todavia, é extremamente importante na inferência estatística e na combinação de amostras.

Por isso mesmo, imaginou-se uma nova medida que tem utilidade e interpretação práticas, denominada **desvio padrão**, definida como a raiz quadrada positiva da variância. Assim, se a variância de um determinado conjunto de valores for igual a 81, o desvio padrão será igual a 9.

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$$

16.1.1 Propriedades da variância e do desvio padrão

- A variância e o desvio padrão são sempre números positivos.

- Se os valores de uma variável forem iguais a variância, então o desvio padrão será igual a zero.

- A variância e o desvio padrão são afetados pelos valores extremos.

Para fins práticos, a fórmula do desvio padrão pode ser reorganizada para dados soltos ou agrupados com ou sem intervalos.

16.1.2 Dados não agrupados

Para dados soltos ou não agrupados, ou seja, se não estão organizados em tabelas, podemos fazer uso da fórmula abaixo:

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

Onde:

n → significa a quantidade de elementos da observação;

$\mathbf{x_1}$ → indica o ienésimo termo da observação;

$\mathbf{x_1^2}$ → indica o ienésimo termo da observação elevado ao quadrado;

∑ → indica o somatório das variáveis em observação.

Para facilitar o entendimento, tomemos como exemplo, o conjunto de valores da variável x: 40, 45, 48, 52, 54, 62 e 70.

Organizando em uma tabela, temos os dados:

x_i	x_i^2
40	
45	
48	
52	
54	
62	
70	
$\sum x_i$	

Observe que o termo x_i , indica o valor de cada variável em cada linha da tabela, e o elemento x_i^2 , significa que devemos calcular o quadrado de cada termo. Completando a tabela temos:

x_i	x_i^2
40	1600
45	2025
48	2304
52	2704
54	2916
62	3844
70	4900
$\sum x_i$	20293

Substituindo os dados na fórmula $s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$, fica:

$$s = \sqrt{\frac{20293}{7} - \left(\frac{371}{7}\right)^2}$$

$$s = \sqrt{2899 - (53)^2}$$

$$s = \sqrt{2899 - 2809}$$

$$s = \sqrt{90}$$

$$s = 9{,}48$$

Sendo assim a dispersão dos dados acima, no início do exercício, tem um desvio para mais e para menos de 9,4863.

16.1.3 Dados agrupados

a) Sem intervalo de classe

Neste caso temos a presença de frequências e devemos levá-las em consideração. Utilizaremos a técnica de acompanharmos a resolução detalhada de um exercício.

Porém a fórmula é outra, vejamos:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

Vamos partir da tabela abaixo:

x_i	f_i
0	2
1	6
2	12
3	7
4	3
∑fi	30

Na fórmula $s = \sqrt{\frac{\sum f_i x_i^2}{\sum f} - \frac{(\sum f_i x_i)^2}{\sum f}}$, temos os somatórios do produto da frequência pelo elemento observado ($f_i x_i$). Para facilitar nosso cálculo vamos acrescentar duas novas colunas para fazer esse produto. Veja:

x_i	f_i	$f_i x_i$	$f_i x_i^2$
0	2	0	0
1	6	6	6
2	12	24	48
3	7	21	63
4	3	12	48
∑	30	63	165

Substituindo, fica:

$$s = \sqrt{\frac{165}{30} - \left(\frac{63}{30}\right)^2}$$

$$s = \sqrt{5{,}5 - (2{,}1)^2}$$

$$s = \sqrt{5{,}5 - 4{,}41}$$

$$s = \sqrt{1{,}09}$$

$$s = 1{,}044$$

b) Com intervalo de classe

Fazemos uso do ponto médio (Pm) no lugar da variável $x_{i.}$

Sendo assim a fórmula fica: $s = \sqrt{\frac{\sum f_i P_m{}^2}{\sum f} - \left(\frac{\sum f_i P_m}{\sum f}\right)^2}$.

Fazendo uso da fórmula acima e aplicando na tabela abaixo, que determina a distribuição de frequência da altura de 40 pessoas, consegue-se perceber que o cálculo é semelhante ao agrupamento sem intervalos.

Estaturas	f_i
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
	$\sum f = 40$

De modo análogo, criaremos novas colunas para facilitar o cálculo e a determinação dos novos elementos da fórmula, que é $\mathbf{f_iP_m^{\ 2}}$ e $\mathbf{\sum f_iP_m}$.

Estaturas	f_i	P_m	f_iP_m	$f_iP_m^{\ 2}$
150 ⊢ 154	4	152	608	92.416
154 ⊢ 158	9	156	1404	219.024
158 ⊢ 162	11	160	1760	281.600
162 ⊢ 166	8	164	1312	215.168
166 ⊢ 170	5	168	840	141.120
170 ⊢ 174	3	172	516	88.752
∑f	**40**		**6440**	**1.038.080**

Substituindo $s = \sqrt{\dfrac{\sum f_iP_m^{\ 2}}{\sum f} - \left(\dfrac{\sum f_iP_m}{\sum f}\right)^2}$ em temos:

$$s = \sqrt{\frac{1.038.080}{40} - \left(\frac{6440}{40}\right)^2}$$

$$s = \sqrt{\frac{1.038.080}{40} - 161^2}$$

$$s = \sqrt{25.952 - 25.921}$$

$$s = \sqrt{31}$$

$$s = 5,567$$

Resumo

A variância e o desvio padrão são medidas de dispersão, que se utilizam dos quadrados dos desvios. Variância é uma medida extremamente importante na inferência estatística. O desvio padrão é muito utilizado na estatística descritiva, principalmente quando utilizamos a média aritmética.

Atividades de aprendizagem

1. Calcule o desvio padrão dos seguintes conjuntos de dados:

a) 20 14 15 19 21 22 18

b) 17,9 22,5 13,3 16,8 15,4 14,2

2. Um departamento de produção usa o procedimento de amostragem para testar a qualidade de itens recém produzidos. O departamento emprega a seguinte regra de decisão em uma estação de inspeção: se uma amostra de 14 tem uma variância de mais que 0,005, a linha de produção precisa ser paralisada para reparos. Suponha que os seguintes dados tenham sido coletados:

a) 3,43 3,45 3,43 3,48 3,52 3,50 3,39

b) 3,48 3,41 3,38 3,49 3,45 3,51 3,51

A linha de produção deve mesmo ser paralisada? Por quê?

3. Os dados abaixo se referem ao número de dias exigidos para preencher pedidos de compra para duas empresas distintas A e B:

Empresa A – 11, 10, 9, 10, 11, 11, 10, 11, 10, 10

Empresa B – 8, 10, 13, 7, 10, 11, 10, 7, 15, 12

Com base nos valores de desvio-padrão calculados, determine qual das empresas fornece tempos de entrega mais constantes e confiáveis.

4. Dada a distribuição relativa a 100 lançamentos de 5 moedas simultaneamente, calcule o desvio padrão.

Número de caras	Frequência
0	4
1	14
2	34
3	29
4	16
5	3
Total	100

5. Em um levantamento entre os assinantes da revista *Fortune* a seguinte pergunta foi realizada: "Quantas das últimas quatro edições você leu ou folheou?". A seguinte distribuição de frequência sintetiza 500 respostas.

Edição lida	Frequência
0	15
1	10
2	40
3	85
4	350
Total	500

a) Qual é o número médio de edições lidas por um assinante da *Fortune*?

b) Qual é o desvio padrão do número de edições lidas?

6. Calcule o desvio padrão da distribuição:

Classes	Frequência
$2 \vdash 6$	5
$6 \vdash 10$	12
$10 \vdash 14$	21
$14 \vdash 18$	15
$18 \vdash 22$	7
Total	60

7. Um posto de gasolina registrou a seguinte distribuição de frequência para o número de litros de gasolina vendidos por carro em uma amostra de 680 carros.

Gasolina (litros)	Frequência
0 ⊢ 5	74
5 ⊢ 10	192
10 ⊢ 15	280
15 ⊢ 20	105
20 ⊢ 25	23
25 ⊢ 30	6
Total	680

- Calcule a média, a variância e o desvio padrão para esses dados agrupados.

Anotações

Aula 17 – Coeficiente de variação

Nessa aula estudaremos o coeficiente de variação e o seu cálculo.

17.1 Coeficiente de variação

É uma medida de afastamento relativa, que vai indicar qual é a relação percentual entre o desvio padrão e a média dos dados observados Ela serve como um termo de comparação entre duas ou mais situações diferentes.

O desvio padrão sozinho não nos diz muita coisa. Assim, um desvio padrão de duas unidades pode ser considerado pequeno ou grande, depende da série estudada, ou seja, para uma série de valores, cuja média é igual a 200. Porém, se a média for igual a 20, o mesmo não pode ser dito.

Além disso, o fato do desvio padrão ser expresso na mesma unidade dos dados limita o seu emprego quando desejamos comparar duas ou mais séries de valores, relativamente a sua dispersão, quando expressas em unidades diferentes.

Ela pode ser calculada pela expressão abaixo:

$$CV = \frac{S}{\overline{X}} \,.\, 100$$

Exemplo:

	Média	Desvio padrão
Estaturas	175cm	5cm
Pesos	68kg	2kg

Temos,

$$CV_E = \frac{S}{\overline{X}} \,.\, 100 = \frac{5}{175} \,.\, 100 = 2{,}85\%$$

$$CV_P = \frac{S}{\overline{X}} \,.\, 100 = \frac{2}{68} \,.\, 100 = 2{,}94\%$$

Logo, neste grupo de indivíduos, os pesos apresentam maior grau de dispersão do que as estaturas.

Resumo

Vimos nesta aula a relação entre o coeficiente de variação, na comparação de duas medidas.

Atividades de aprendizagem

1. Dez alunos foram submetidos a um teste de estatística e de matemática, obtendo as seguintes notas:

Aluno	Matemática (X)	Estatística (Y)
A	7	6
B	6	5
C	9	10
D	10	9
E	3	2
F	4	3
G	8	9
H	7	5
I	6	6
J	2	3

a) Determinar o coeficiente de Pearson.

2. Foi realizado uma pesquisa entre 8 famílias de uma certa região de Curitiba.

Famílias	Renda (XR$100,00)	Poupança (XR$1000,00)	N° de filhos
A	8	1	5
B	12	2	6
C	20	6	3
D	25	5	1
E	30	10	2
F	33	9	2
G	40	15	1
H	42	22	0

a) Determinar o coeficiente de correlação linear entre a renda familiar e a poupança das 8 famílias.

b) Determinar o coeficiente de correlação linear entre a renda familiar e o número de filhos das 8 famílias.

c) Determinar o coeficiente de correlação linear entre a poupança e o número de filhos.

3. Determinar os coeficientes de correlação de Pearson para todos os exercícios de aplicação do item Regressão linear simples.

Anotações

Aula 18 – Probabilidade

Nesta aula, você saberá o que é probabilidade. Estudaremos a chance de ocorrência de um evento em um experimento aleatório, os conceitos dos elementos da probabilidade e efetuaremos o seu cálculo.

18.1 Definição

Podemos definir probabilidade como sendo uma ciência que estuda a chance de um acontecimento ocorrer em relação a situações de incerteza.

Você deve estar se perguntando, como assim?

Em muitas situações do dia a dia, você ou as empresas, tomam decisões o tempo todo. Algumas dessas decisões são tomadas com bases em fatos e dados. Só que às vezes esses fatos não são precisos ou não admitem certeza.

Quando a situação é pessoal utilizamos nossa intuição ou a nossa "sorte", porém em se tratando de empresa, cujo objetivo é o lucro e não prejuízo, não **é permitido** aos gestores arriscarem perder dinheiro, usando tão somente a intuição. Sendo assim, as empresas, fazem uso da probabilidade.

Para entender um pouco mais, vamos aprofundar nosso estudo sobre esse assunto. Lembre-se que os fenômenos estudados em Estatística são fenômenos cujo resultado, mesmo em condições normais de experimentação, varia de uma observação para outra, dificultando a previsão de um resultado futuro.

18.2 Elementos da probabilidade

Para nossa compreensão é melhor nos familiarizarmos com alguns termos. Então, vamos a eles.

- Experimento aleatório

É a denominação dada para uma experiência. Por exemplo, quando você lança uma moeda para tirar cara ou coroa; ou quando você retira uma carta de um baralho, querendo que ela seja de ouro. Esse tipo de experimento, que pode ser repetido indefinidamente, sob as mesmas condições, pode apresentar variações de resultados. Note que não é possível afirmar, qual será o resultado sem que o experimento tenha sido realizado.

- Espaço amostral (S)

São todos os resultados possíveis do experimento aleatório. O espaço amostral representa o universo de todos os possíveis eventos. No caso do lançamento da moeda, entendo como espaço amostral os resultados possíveis, que são cara ou coroa.

Se pensarmos no baralho, seriam todas as treze cartas de ouro do baralho, veja: S = { As, 2,3,4,5,6,7,8,9,10,V,D,R }. No c**á**lculo probabilístico sempre utilizamos a letra **S** para representar o conjunto solução de todos os resultados, bem como o conjunto dos números n(S) = 13.

- Evento

É o resultado que queremos que aconteça na experiência. É qualquer conjunto de resultados de um experimento. Os eventos são indicados por letras maiúsculas: A, B, etc.

Os eventos podem ser divididos em:

- **Simples** – É um evento formado por um único elemento do espaço amostral.

- **Composto** – É aquele que possui mais de um evento.

- **Certo** – É o próprio espaço amostral S.

- **Impossível** – É o evento que não tem possibilidade de acontecer. É simbolizado pelo conjunto vazio ∅.

18.3 Calculando a probabilidade

A probabilidade matemática de um acontecimento é a razão entre o número de elemento dos eventos pelo número de elementos do espaço amostral, desde que haja rigorosa equipossibilidade entre todos os casos.

Assim, calculamos a probabilidade por:

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

Exemplos:

1º Exemplo – Qual é a probabilidade de se obter uma cara em uma única jogada de moeda?

Resolução:

Primeiramente, calculamos o espaço amostral ao lançarmos a moeda; podemos encontrar os seguintes resultados:

S = { cara, coroa } $\rightarrow$ n(S) = 2

Agora vamos determinar o número de vezes que o evento pode ocorrer, lembre-se que o evento é o que queremos que aconteça, logo:

A = { cara } $\rightarrow$ n(A) = 1, então substituindo na fórmula temos:

$$P(A) = \frac{n(A)}{n(S)} \rightarrow P(\text{Cara}) = \frac{1}{2} \text{ ou } 50\%$$

Podemos então concluir que temos 50% de chances de lançar uma moeda e ela dar cara.

2º Exemplo – Qual é a probabilidade de se obter uma única cara em um lançamento de três moedas?

Resolução:

Primeiramente calculamos o espaço amostral:

S = {uma única cara} $\rightarrow$ n(S) = 8, pois:

S = {KKK, KKC, KCK, KCC, CKK, CKC, CCK, CCC}

O evento é o que queremos, logo queremos uma única cara, portanto:

A = { cara } $\rightarrow$ A = {KKC, KCK, CKK} $\rightarrow$ n(A) = { 3 }, logo:

$P(A) = \frac{n(A)}{n(S)} \rightarrow P(\text{unicacara}) = \frac{3}{8}$ ou 37,5%

Ou seja, a chance de sair uma única cara no lançamento de três moedas é de 37,5%.

18.4 Calculando outras probabilidades

Existem outros formatos de acontecimentos probabilísticos, que são:

- **Acontecimentos mutuamente exclusivos**

Diz-se que dois eventos são mutuamente exclusivos quando, na ocorrência de um, o outro não pode ocorrer. Utilizamos a fórmula:

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

Exemplo: Qual a probabilidade de - ao retirar uma carta de um baralho de 52 cartas – obtermos um ás ou uma dama.

$P(B \cup C) = P(B) + P(C)$

$P(B \cup C) = \frac{4}{52} + \frac{4}{52}$

$P(B \cup C) = \frac{8}{52}$

Resposta: A probabilidade é de $\frac{8}{52}$ ou 0,1538 ou 15,38%.

- **Eventos não mutuamente exclusivos**

Diz-se que dois eventos **NÃO** são mutuamente exclusivos quando podem ocorrer simultaneamente.

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

Exemplo: Ao se retirar uma carta de um baralho comum, com 52 cartas, qual a probabilidade de obtermos um ás ou uma carta de espadas?

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

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

$$P(B \cup C) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

$$P(B \cup C) = \frac{17}{52}$$

Resposta: A probabilidade é de $\frac{17}{52}$ ou 0,3269 ou 32,69%

- **Probabilidade condicional**

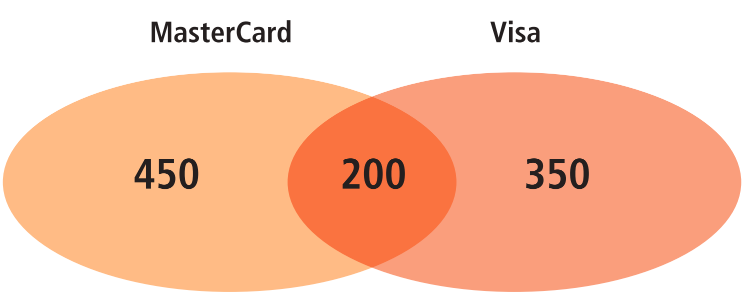
Dados dois eventos A e B de um espaço amostral S. A probabilidade do evento A ocorrer quando B já tiver ocorrido é chamada de probabilidade condicional de A em relação a B.

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

Fonte: extraído do http://www.matematicadidatica.com.br/ProbabilidadeCondicional.aspx

Uma pesquisa realizada entre 1.000 consumidores, registrou que 650 deles trabalham com cartões de crédito da bandeira MasterCard; que 550 trabalham com cartões de crédito da bandeira VISA e que 200 trabalham com cartões de crédito de ambas as bandeiras. Qual a probabilidade de ao escolhermos deste grupo uma pessoa que utiliza a bandeira VISA, ser também um dos consumidores que utilizam cartões de crédito da bandeira MasterCard?

Observe a figura abaixo e a compare com as informações do enunciado. Este exemplo ajudará na resolução de outros problemas:



Onde:

$$\begin{cases} n(A) = 450 + 200 = 650 \\ n(B) = 350 + 200 = 550 \\ n(A \cap B) = 200 \\ n(s) = 1000 \end{cases}$$

A probabilidade procurada é dada pela fórmula:

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

Como mencionamos, a probabilidade da intersecção é a razão do seu número de elementos, para o número de elementos do espaço amostral, então a fórmula acima pode ser reduzida a:

$$P\,(A \cap B) = \frac{n(A \cap B)}{n(S)}$$

O número de pessoas que utilizam as duas bandeiras, ou seja, a quantidade de elementos da intersecção é igual a **200**, já o número de consumidores que utilizam ao menos a bandeira **VISA** é **550**, portanto:

$$P\,(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{200}{550} = \frac{4}{11}$$

Logo, a probabilidade de escolher uma pessoa que utiliza a bandeira VISA, ser também usuária da bandeira MasterCard é $\frac{4}{11}$ ou 36,36%.

Resumo

- Espaço amostral é o conjunto de todos os resultados possíveis de ocorrer.

- Experimentos aleatórios é um experimento em que ocorrendo nas mesmas condições podem apresentar resultados diferentes a cada ocorrência.

- Evento é qualquer subconjunto de um espaço amostral.

- Classificação de Eventos se dividem em Evento Simples, Evento Certo e Evento Impossível.

Atividades de aprendizagem

1. Um dado é lançado. Qual é a probabilidade de obtermos um número divisor de 6?

2. Uma moeda é lançada 4 vezes. Qual é a probabilidade de obtermos ao menos uma coroa?

3. Em uma caixa há 2 fichas amarelas, 5 azuis e 7 verdes. Se retirarmos uma única ficha, qual a probabilidade dela ser verde ou amarela?

4. De uma sacola contendo 15 bolas numeradas de 1 a 15 retira-se uma bola. Qual é a probabilidade desta bola ser divisível por 3 ou divisível por 4?

5. Em uma escola de idiomas com 2.000 alunos, 500 fazem o curso de inglês, 300 fazem o curso de espanhol e 200 cursam ambos os cursos. Selecionando-se um estudante do curso de inglês, qual a probabilidade dele também estar cursando o curso de espanhol?

Aula 19 – Medidas de assimetria e curtose

Nesta aula iremos estudar os tipos de curtose de uma distribuição, identificando-a através de cálculos matemáticos.

19.1 Assimetria

A assimetria que também pode ser denominada de medidas de enviesamento, indicam o grau de deformação de uma curva de frequência.

Para Filho (2007, p.73), podemos definir a curtose como: "... o grau de deformação (assimetria) e o grau de a*chat*amento ou afilamento (curtose) da curva de frequências ou do histograma".

Para dados que estão agrupados, além do histograma, que é o grafico de uma distribuição de frequência, nós representamos também por uma curva de frequência; as diferenças entre os valores da média, da mediana e da moda são indicadores da forma da curva em termos de assimetria.

Sendo assim, como já vimos na aula sobre gráficos, podemos formar um polígono de frequência que descreve as curvas abaixo. Vejamos primeiramente as medidas de assimetria e seus tipos:

- **Primeira medida de assimetria**

É a curva chamada de simétrica ou padrão – quando a curva tem um formato de sino, bem simétrica em relação ao eixo que é determinado pela média. Nesse eixo as medidas de mediana e moda são iguais a média aritmética. Confira na **figura 19.1**.

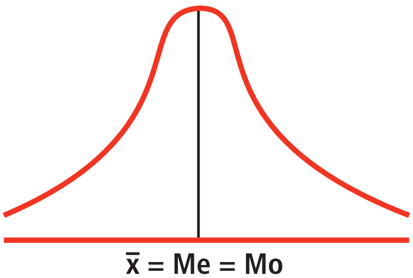


Figura 19.1: Curva normal
Fonte: (FILHO, 2007)

- **Segunda curva assimétrica**

Essa curva é chamada de **assimétrica à direita** ou positiva, pois como podemos observar ela tem uma cauda à esquerda. Suas medidas estão numa desigualdade, de ordem igual a:

Mo<Me<x

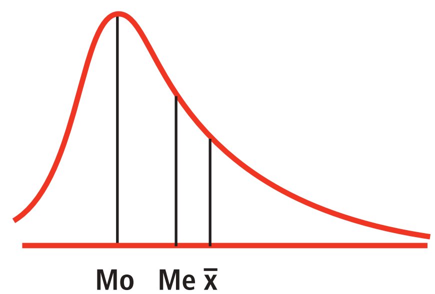


Figura 19.2: Curva assimétrica à direita ou assimétrica positiva
Fonte: (FILHO, 2007)

- **Terceira curva assimetrica**

Denominada de **assimétrica à esquerda** ou **negativa**, ela tem sua cauda no sentido à direita e estabelece a relação de desigualdade x<me<mo. Observe a **figura 19.3**.

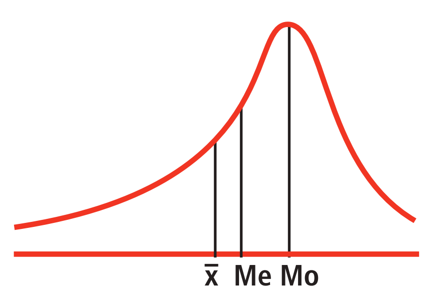


Figura 19.3: Curva assimétrica à esquerda ou curva assimétrica negativa
Fonte: (FILHO, 2007)

Como você pode ver estabelecemos algumas relações de grandeza, tais como:

- Na distribuição assimétrica positiva, podemos observar que a moda é maior que a mediana que é maior que a média, ou seja: Mo < Me < X

Observe:

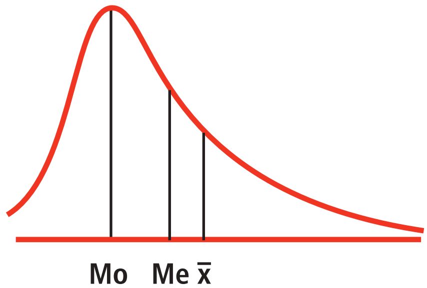


Figura 19.4: Distribuição assimétrica positiva
Fonte: (FILHO, 2007)

19.1 Coeficientes de assimetria

Para verificar se uma curva é simétrica ou assimétrica, positiva ou negativa, não precisamos ficar desenhando o polígono. Fazemos uso da matemática, utilizando equações que determinam qual o formato da curva. Essas equações algébricas utilizam as medidas de tendência central e de dispersão para determinar o chamado coeficiente de assimetria. O primeiro coeficiente foi descoberto por Cal Pearson, que pode ser obtido através de duas fórmulas, como vocês podem ver:

1º) Coeficiente de Pearson

$$Sk = \frac{(\bar{x} - Mo)}{dp}$$

2º) Coeficiente de Pearson

$$Sk = \frac{3.(\bar{x} - Me)}{dp}$$

Quando:

$Sk = 0 \rightarrow$ distribuição simétrica, ou seja ela é normal.

Se $Sk > 0 \rightarrow$ então a distribuição é assimétrica positiva.Porém se:

$Sk < 0 \rightarrow$ a distribuição é dita assimétrica negativa.

Exemplo:

Calcular o coeficiente de assimetria e classificar a curva de uma distribuição que possui:

a) $\bar{x}$ = 6; Md = 6 e Dp = 1.

Substituindo temos:

$$Sk = \frac{3.(\bar{x} - Me)}{dp}$$

$$Sk = \frac{3.(6 - 6)}{1}$$

$$Sk = \frac{3.(0)}{1}$$

$$Sk = \frac{0}{1}$$

$$Sk = 0$$

Logo a curva é normal, ou seja, simétrica e tem o formato abaixo:

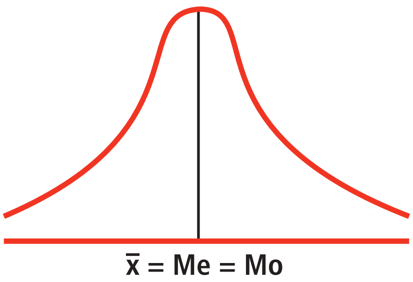


Figura 19.5: Curva simétrica
Fonte: (FILHO, 2007)

b) $\bar{x}$ = 88; Md = 82 e Dp = 40.

$$Sk = \frac{3.(\bar{x} - Me)}{dp}$$

$$Sk = \frac{3.(88 - 82)}{40}$$

$$Sk = \frac{3.(6)}{40}$$

$$Sk = \frac{18}{40}$$

$$Sk = 0,45$$

Sendo Sk = 0,45, podemos saber que a curva é assimétrica positiva, como na **figura 19.6**.

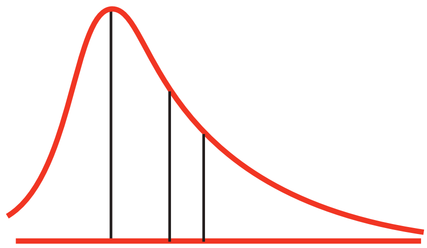


Figura 19.6: Curva assimétrica à direita ou assimétrica positiva
Fonte: (FILHO, 2007)

19.2 Medidas de curtose

Indicam o grau de achatamento de uma curva de frequência, ou do histograma correspondente a ela. Esse comportamento é dado pela concentração dos valores em relação à moda.

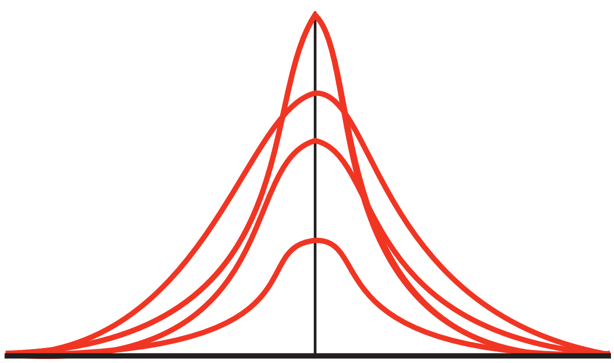


Figura 19.7: Medida de curtose
Fonte: http://dc99.4shared.com/

Com a medida assimétrica, podemos determinar a medida de curtose por uma fórmula. Veja.

$$K = \frac{Q_3 - Q_1}{2 \cdot (P_{90} - P_{10})}$$

Quando:

$K = 0{,}263 \rightarrow$ distribuição normal (mesocúrtica)

$K < 0{,}263 \rightarrow$ distribuição alongada (leptocúrtica)

$K > 0{,}263 \rightarrow$ distribuição achatada (platicúrtica)

Exemplo:

Com os dados abaixo, calcule o coeficiente de curtose.

$Q_3 = 6{,}75$ $\quad P_{90} = 8{,}20$ $\quad Q_1 = 3{,}25$ $\quad P_{10} = 1{,}80$

Solução:

Substituindo, temos:

$$K = \frac{Q_3 - Q_1}{2 \,.\, (P_{90} - P_{10})}$$

$$K = \frac{6{,}75 - 3{,}25}{2 \,.\, (8{,}20 - 1{,}80)}$$

$$K = \frac{3{,}50}{2 \,.\, (6{,}40)}$$

$$K = \frac{3{,}50}{12{,}80}$$

$$K = 0{,}2734$$

Portanto, a curva é chamada de platicúrtica, pois 0,273 é maior que 0,263.

Resumo

- Assimetria mede o grau de deformação de uma curva de frequência. Tipos de assimetria: simétrica ou normal, assimétrica positiva ou negativa.

- Curtose mede o grau de achatamento de uma distribuição de frequência. Tipos de curtose: mesocúrtica, leptocúrtica e platicúrtica.

Atividades de aprendizagem

1. Os dados abaixo se referem à renda nominal de 60 famílias (valores em R$)

400	350	370	375	399	405	360	408	430	385
390	390	385	360	397	400	406	440	415	410
400	410	382	340	360	370	380	370	413	390
400	390	355	375	427	397	413	430	357	340
350	410	420	360	403	382	390	425	420	404
420	410	411	404	397	420	404	421	440	380

Pede-se:

a) montar uma tabela de distribuição por frequência usando a fórmula de "Sturges";

b) a média aritmética: (R$ 395,75)

c) a mediana: (R$ 400,00)

d) a moda de Karl Pearson: (R$ 408,50)

e) o primeiro e o terceiro quartis; o décimo e o nonagésimo percentis; (R$378,33; R$412,50; R$360,00; R$426,25)

f) o desvio médio; (R$19,75)

g) desvio padrão; (R$24,09)

h) o percentual de famílias cujas rendas situam-se entre a C ± 1 DM; (56,67%)

i) o número de famílias cujas rendas situam-se entre a C ±1 S (suponha que a distribuição seja normal). (± 41 famílias)

j) coeficiente de variação; (6,09%)

---

k) o coeficiente de assimetria de Pearson (primeiro e segundo). Faça a curva e determine o seu comportamento; (– 0,529; Assimetria Negativa)

---

l) determinar o grau de achatamento da curva, identificando-a pelo nome e pela curva; (0,258; Leptocúrtica)

Trabalhe sempre com duas casas depois da vírgula, fazendo arredondamento estatístico, onde for possível.

2. Em um ensaio para o estudo da distribuição de um atributo financeiro (X), foram examinados 200 itens de natureza contábil do balanço de uma empresa. Esse exercício produziu a tabela de frequência abaixo. A coluna Classes representa intervalos de valores de X em reais e a coluna P representa a frequência relativa acumulada. Não existem observações coincidentes com os extremos das classes. (Fonte: Questão extraída do AFRF-2002.1)

Classes	P(%)
70 ⊢ 90	5
90 ⊢ 110	15
110 ⊢ 130	40
130 ⊢ 150	70
150 ⊢ 170	85
170 ⊢ 190	95
190 ⊢ 210	100

Entende-se por curtose de uma distribuição seu grau de achatamento em geral medido em relação à distribuição normal. Uma medida de curtose é dada pelo quociente k = Q / (P90-P10), onde Q é a metade da distância interquartílica e P90 e P10 representam os percentis de 90% e 10%, respectivamente. Assinale a opção que dá o valor da curtose k para a distribuição de X.

a) 0,263

b) 0,250

c) 0,300

d) 0,242

e) 0,000

Anotações

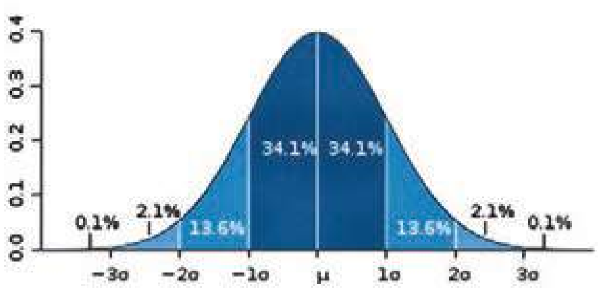
Aula 20 – Curva normal

Nesta aula iremos estudar a curva normal, e calcular a probabilidade de determinar a sua ocorrência.

20.1 Distribuição normal

A distribuição normal de probabilidade é uma distribuição de comportamento normal, com probabilidade contínua que é simétrica e mesocúrtica. A curva que representa a distribuição normal de probabilidade é descrita como tendo o formato de sino e é também conhecida como **Curva de Gauss**.

Observe abaixo a curva representativa da distribuição normal de probabilidade:



A área em azul escuro está a menos de um desvio padrão (σ) da média. Em uma distribuição normal, isto representa cerca de 68% do conjunto, enquanto dois desvios padrões desde a média (azul médio e escuro) representam cerca de 95%, e três desvios padrões (azul claro, médio e escuro) cobrem cerca de 99.7%. Este fato é conhecido como *regra 68-95-99.7*, ou a *regra empírica*, ou a *regra dos 3-sigmas*.

Figura 20.2: curva representativa da distribuição normal de probabilidade
Fonte: http://pt.wikipedia.org/

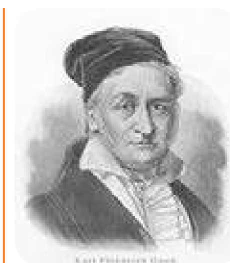


Figura 20.1: Karl Friedrich Gauss
Fonte: www.sil.si.edu

Para saber mais sobre Karl Friedrich Gauss acesse: http://www.educ.fc.ul.pt/docentes/opombo/seminario/gauss/gauss.htm

Este tipo de distribuição é importante na inferência estatística por três razões:

a) as medidas produzidas em diversos processos aleatórios seguem este tipo de distribuição;

b) probabilidades normais podem ser usadas frequentemente como aproximações de outras distribuições (binomial e de Poisson);

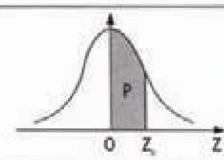
c) as distribuições amostrais de estatísticas, tais como a média, frequentemente seguem a distribuição normal independente da distribuição da população.

Representando as médias por l e o desvio padrão por S, qualquer combinação destes valores gera uma distribuição normal de probabilidade diferente (todas simétricas e mesocúrticas. As tabelas de probabilidades normal são baseadas em uma distribuição normal de probabilidades com l = 0 e S = 1. Vejamos:

Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segundo decimal de Z_c: 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

Figura 20.3: Distribuição normal padronizada
Fonte: http://www.engenheirodepetroleo.com.br/

Qualquer conjunto de valores X, normalmente distribuídos, pode ser convertido em valores normais **padronizados** através da fórmula:

$$z = \frac{X - \lambda}{S}$$

Os valores de z indicam as proporções de área para vários intervalos, com fronteira começando sempre na média.

Exemplo: Sabe-se que a vida útil de determinado componente eletrônico segue uma distribuição normal com média l = 2.000 horas e desvio padrão S = 200 horas. Qual é a probabilidade de um componente eletrônico, escolhido aleatoriamente, dure entre 2.000 e 2.400 horas?

Cálculo do valor padronizado:

$$z_1 = \frac{X - \lambda}{S} \rightarrow z_1 = \frac{2.400 - 2.000}{200} \rightarrow z_1 = \frac{400}{200} \rightarrow z_1 = 2$$

Portanto, o valor padronizado correspondente a 2.400 horas é o valor z = 2,00.

Visualizando, na curva, o que deve ser calculado:

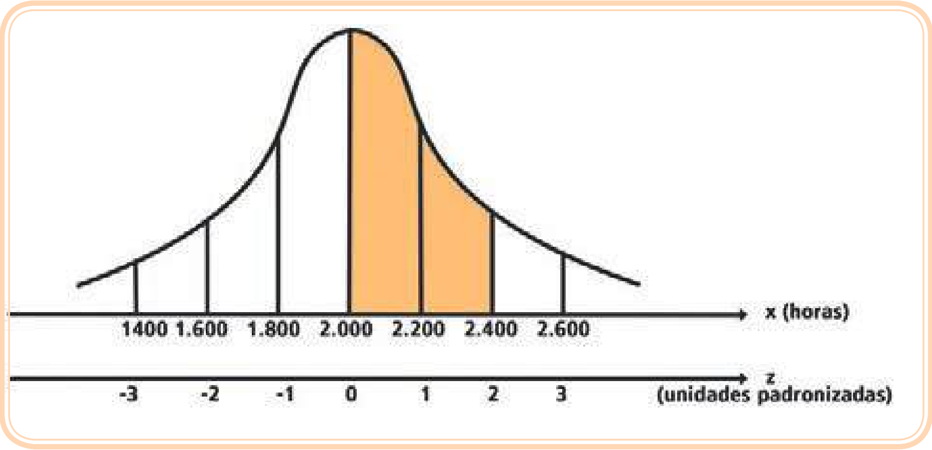


Figura 20.4: Cálculo do valor padronizado
Fonte: Castanheira, 2010

Procurando na tabela z = 2,00, encontramos o número 0,4772.

Então, a probabilidade procurada pode ser escrita da seguinte forma:

P (2.000 $\leq$ z $\leq$ 2.400) = 0,4772 ou P (2.000 $\leq$ z $\leq$ 2.400) = 47,72%.

Referências

ANDERSON, David R. **Estatística aplicada à administração e economia**. 2ª edição. São Paulo: Pioneira Thomson Learning, 2003.

CASTANHEIRA, Nelson Pereira. **Estatística aplicada a todos os níveis**. Curitiba: IBPEX, 2010.

CRESPO, Antônio Arnot. **Estatística Fácil**. 17 ed. São Paulo: Saraiva, 1999.

FILHO, Jair Croce. **Apostila de Estatística básica, 2007**. Disponível em: http://lia.uncisal.edu.br/ensino/pdf2/Apostila_Estatistica_II.pdf

LAPPONI, Juan Carlos. **Estatística usando Excel**. São Paulo: Lapponi Treinamento, 2000.

MACEDO, Luiz Roberto Dias de. **Dados numéricos na empresa: interpretação e análise.** Curitiba: IBPEX, 2004.

MARTINS, Gilberto de Andrade. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2008.

MILANI, Giusepe, et. Al. **Estatística Aplicada**. São Paulo: Atlas, 1995.

SILVA, Ermes Medeiros da, e outros. **Estatística para os cursos de economia, administração e ciências contábeis**. 3ªedição. São Paulo: Atlas, 1999.

STEVENSON, William J. **Estatística Aplicada à Administração**. São Paulo: Harper e Row do Brasil, 1981.

PASQUALI, Luiz. **A Curva Normal**. Disponível em: <http://www.psiambiental.

net/pdf/PasqCap03.pdf>. Acesso em: 06 de Dezembro de 2011.

PEREIRA, Paulo Henrique. **Noções de estatística: com exercícios para administração e ciências humanas.** São Paulo: Papirus, 2004.

WIKIPEDIA. Desenvolvido pela Wikipédia Foundation. Apresenta conteúdo enciclopédico. [S.I], 2011. Disponível em: <http://pt.wikipedia.org/wiki/>. Acesso em: 8 de outubro de 2011.

Referências das figuras

Figura 1.1: Análise estatística
Fonte: http://www.decifrando.com/wp-content/uploads/2010/10/estatistica.jpg

Figura 1.2: Godofredo Achenwall
Fonte: http://cdn.dipity.com/uploads/events/6adb31d1855709f428a5c1499902af9e_1M.png

Figura 1.3: O mundo logístico
Fonte: http://www.tocadacotia.com/tecnologia/tecnologia-em-logistica

Figura 2.1: Métodos – DMAIC / 6 SIGMA
Fonte: http://www.estatcamp.com.br/empresa/4-dmaic

Figura 3.1: População
Fonte: ©Andrea Danti/Shutterstock

Figura 3.2: População e Amostra
Fonte: elaborado pelo autor.

Figura 3.3: Variáveis quantitativas e qualitativas.
Fonte: http://img20.imageshack.us/img20/1919/se100pessoas.jpg

Figura 4.1: Números aleatórios
Fonte: http://tic4-39188.blogspot.com/2008/12/c8-tabela-de-nmeros-aleatrios.html

Figura 4.2: Amostra estratificada
Fonte: Adaptado de www.mbi.com.br/mbi/biblioteca/tutoriais/estratificacao

Figura 5.1: Componentes de uma tabela
Fonte: Adaptado de Estatística fácil, 2002.

Figura 6.1: Gráfico em linha ou em curva
Fonte: MDIC/SECEX

Figura 6.2: Gráfico em colunas ou em barras
Fonte: MDCI / SECEX

Figura 6.3: Gráfico em setores
Fonte: MDIC/SECEX

Figura 6.4: Histograma
Fonte: http://mathforum.forumais.com/t15-poligno-de-frequencia

Figura 8.1: Limites de classe
Fonte: Elaborado pelo autor

Figura 8.2: Cálculo das frequências
Fonte: Elaborado pelo autor

Figura 13.1: Decil
Fonte: ©maksymmo/Shutterstock

Figura 19.1: Curva normal
Fonte: Filho (2007)

Figura 19.2: Curva assimétrica à direita ou assimétrica positiva
Fonte: Filho, 2007

Figura 19.3: Curva assimétrica à esquerda ou curva assimétrica negativa
Fonte: Filho (2007)

Figura 19.4: Curva assimétrica à direita ou assimétrica positiva
Fonte: Filho (2007)

Figura 19.5: Curva simétrica
Fonte: Filho (2007)

Figura 19.6: Curva assimétrica à direita ou assimétrica positiva
Fonte: Filho (2007)

Figura 19.7: Medida de curtose
Fonte: http://dc99.4shared.com/doc/6eUysS5d/preview.html

Figura 20.1 Karl Friedrich Gauss
Fonte: http://www.sil.si.edu/digitalcollections/hst/scientific-identity/CF/by_name_display_results.cfm?scientist=Gauss,%20 Carl%20Friedrich
Figura 20.2: curva representativa da distribuição normal de probabilidade
Fonte: http://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_normal

Figura 20.3: Distribuição normal padronizada
Fonte: http://www.engenheirodepetroleo.com.br/

Figura 20.4: Cálculo do valor padronizado
Fonte: Castanheira, 2010

Tabela 5.3: Série geográfica
Fonte: FGV Management

Tabela 5.4: Séries específicas
Fonte: Elaborado pelo autor

Tabela 5.5: Distribuição de frequência com intervalos
Fonte: Elaborado pelo autor

Tabela 7.1: Distribuição de frequência das notas dos alunos
Fonte: Elaborado pelo autor

Tabela 7.2: distribuição de frequência com intervalos e suas respectivas frequências.
Fonte: Elaborado pelo autor

Tabela 8.1: Tabela com poucos dados e sem intervalos
Fonte: Dados fictícios do autor

Tabela 8.2: Tabela com muitos dados e intervalos
Fonte: Dados fictícios do autor

Tabela 11.1: Vantagens e desvantagens da utilização da moda
Fonte: Crespo (2007)

Quadro 9.1: Vantagens e desvantagens da média aritmética
Fonte: Crespo, 2007

Quadro 10.1: Vantagens e desvantagens da utilização da mediana
Fonte: Crespo, 2007

Atividades autoinstrutivas

1. Dentre as alternativas abaixo assinale a que está INCORRETA.

a) População é o conjunto de elementos que desejamos observar para obter determinada informação.

b) Amostra é o subconjunto de elementos retirados da população que se está observando, para obter determinada informação.

c) Estatística Descritiva ou Dedutiva é a parte da Estatística referente à coleta e à tabulação dos dados.

d) Estatística Indutiva ou Inferencial é a parte da Estatística referente às conclusões sobre as fontes de dados, consistindo, assim, em obter e generalizar conclusões.

e) Rol é a relação dos resultados obtidos em uma pesquisa e que foram transcritos aleatoriamente, ou seja, fora de ordem numérica.

2. Em relação à distribuição de frequências apresentada abaixo, é CORRETO afirmar que:

IDADES	f	fa
18 ⊢ 21	9	9
21 ⊢ 24	12	21
24 ⊢ 27	12	33
27 ⊢ 30	17	50
30 ⊢ 33	16	66
33 ⊢ 36	14	80
36 ⊢ 39	11	91
39 ⊢ 42	9	100
	$\sum f = 100$	

a) A amplitude total da distribuição é igual a 3.

b) Esta é uma distribuição que possui 100 classes.

c) O limite superior da 5ª classe é $Ls_5 = 30$.

d) O ponto médio da 4ª classe é $Pm_4 = 28,5$.

e) A frequência acumulada da 6ª classe é $fa_6 = 14$.

3. Em relação à distribuição de frequência dada abaixo, podemos afirmar que:

ALTURA (em cm.)	f	fa
160 ⊢ 163	4	4
163 ⊢ 166	8	12
166 ⊢ 169	10	22
169 ⊢ 172	9	31
172 ⊢ 175	7	38
175 ⊢ 178	7	45
178 ⊢ 181	5	50
	$\sum f = 50$	

(Obs.: Efetue os cálculos com duas casas decimais)

a) A mediana é igual a Me = 168,25.

b) A moda é igual a Mo = 165,79.

c) A média é igual a $\overline{X} = 170{,}38$.

d) A frequência relativa da 4ª classe é fr_4 = 0,18 e informa que 18% dos elementos pesquisados têm altura entre 169cm e172 cm.

e) Estão corretas as alternativas c e d.

4. Dentre as alternativas dadas abaixo, assinale a que está CORRETA.

a) A amplitude total do conjunto de números – 4, – 2 , – 1, 0 , 2 , 4 , 6 , 8, é igual a A = 4.

b) A amplitude total da distribuição de frequências da questão de número 2 é igual a 24, seja calculada através dos pontos médios das classes, seja calculada pela diferença entre o limite superior da última classe e o limite inferior da primeira classe.

c) O desvio médio do conjunto de números 4 , 6 , 8 , 10 , 11 é igual a D_m = 6.

d) O desvio padrão da distribuição de frequências da questão de número 2 é igual a s = 6,26, calculado com duas casas decimais, considerando-se que a distribuição é representativa de uma amostra da população.

e) Todas as alternativas estão corretas.

5. Dentre as alternativas dadas abaixo, assinale a que está INCORRETA.

a) As medidas de assimetria, também denominadas de medidas de enviesamento, indicam o grau de deformação de uma curva de frequências.

b) Uma curva de distribuição de frequências, quanto à assimetria, pode ser classificada em simétrica, assimétrica positiva ou assimétrica negativa.

c) Curtose é o grau de achatamento ou de afilamento de uma distribuição de frequências, ou seja, do histograma correspondente.

d) Quanto ao grau de achatamento, uma curva de distribuição de frequências pode ser classificada em mesocúrtica, platicúrtica ou leptocúrtica.

e) Se o coeficiente de curtose (K) para uma determinada distribuição de frequência for igual a 0,295, pode-se afirmar que a curva é do tipo leptocúrtica.

6. João, Um empresário bem sucedido, pediu a Pedro e a Paulo que resolvessem um determinado problema que a empresa estava enfrentando. A probabilidade de que Pedro resolver o problema é igual a $\frac{1}{5}$ e a de Paulo resolver a questão é de $\frac{2}{3}$. Se ambos tentarem resolver o problema, independentemente um do outro, qual a probabilidade de que o problema seja resolvido? (Obs.: Efetue os cálculos com duas casas decimais)

a) $\frac{2}{15}$ ou 13,33%

b) $\frac{11}{15}$ ou 73,33%

c) $\frac{3}{8}$ ou 37,50%

d) $\frac{1}{4}$ ou 25,00%

e) $\frac{1}{2}$ ou 50,00%

7. Uma caixa A contém 5 peças perfeitas e 3 com algum tipo de defeito . Outra caixa B, contém 4 peças perfeitas e 2 com algum tipo de defeito. Uma peça é retirada de cada caixa aleatoriamente. Qual a probabilidade das peças retiradas serem ambas perfeitas ou ambas defeituosas? (Obs.: Efetue os cálculos com duas casas decimais)

a) $\frac{5}{12}$ ou 41,67%.

b) $\frac{1}{8}$ ou 12,50%.

c) $\frac{13}{24}$ ou 54,17%

d) $\frac{9}{14}$ ou 64,29%.

e) $\frac{5}{14}$ ou 35,71%.

8. Dentre as alternativas abaixo assinale a que está CORRETA.

a) População é o subconjunto de elementos retirados da população que se está observando, para obter determinada informação.

b) Amostra é o conjunto de elementos que desejamos observar para obter determinada informação.

c) Rol é a relação dos resultados obtidos em uma pesquisa e que foram transcritos obedecendo a uma ordem numérica, seja crescente ou decrescente.

d) Estatística Indutiva ou Inferencial é a parte da Estatística referente à coleta e à tabulação dos dados.

e) Estatística Descritiva ou Dedutiva é a parte da Estatística referente às conclusões sobre as fontes de dados, consistindo, assim, em obter e generalizar conclusões.

9. Em relação à distribuição de frequências apresentada abaixo, é CORRETO afirmar que:

IDADES	f	fa
20 ⊢ 23	8	8
23 ⊢ 26	12	20
26 ⊢ 29	13	33
29 ⊢ 32	18	51
32 ⊢ 35	15	66
35 ⊢ 38	14	80
38 ⊢ 41	11	91
41 ⊢ 44	9	100
	$\sum f = 100$	

a) A amplitude total da distribuição é igual a 24.

b) Esta é uma distribuição que possui 8 classes.

c) O limite superior da 5ª classe é $Ls_5 = 35$.

d) O ponto médio da 4ª classe é $Pm_4 = 30,5$.

e) Todas as alternativas estão corretas.

10. Em relação à distribuição de frequência dada abaixo, podemos afirmar que: (Obs.: Efetue os cálculos com duas casas decimais)

ALTURA (em cm.)	f	fa
140 ⊢ 148	3	3
148 ⊢ 156	8	11
156 ⊢ 164	10	21
164 ⊢ 172	13	34
172 ⊢ 180	9	43
180 ⊢ 188	5	48
188 ⊢ 196	2	50
	$\sum f = 50$	

a) A mediana é igual a Me = 164,66.

b) A moda é igual a Mo = 167,79.

c) A média é igual a $\bar{x} = 164,60$.

d) A frequência relativa da 4ª classe é $fr_4 = 0,26$ e informa que 26% dos elementos pesquisados têm altura entre 164cm e 172cm.

e) Estão corretas as alternativas b e d.

11. Dentre as alternativas dadas abaixo, assinale a que está CORRETA.

a) A amplitude total do conjunto de números – 6, – 3 , – 2, 0 , 4 , 7 , 9 , 11, é igual a $A_T = 20$.

b) A amplitude total da distribuição de frequências da questão de número 2 é igual a 24, seja calculada através dos pontos médios das classes, seja calculada pela diferença entre o limite superior da última classe e o limite inferior da primeira classe.

c) O desvio médio do conjunto de números 5 , 8 , 12 , 14 , 16 é igual a $D_m = 6{,}3$.

d) O desvio padrão da distribuição de frequências da questão de número 2 é igual a s = 5,90, calculado com duas casas decimais, considerando-se que a distribuição é representativa de uma amostra da população.

e) Todas as alternativas estão corretas.

12. Dentre as alternativas dadas abaixo, assinale a que está INCORRETA.

a) Em relação à assimetria, as curvas de distribuição de frequências podem ser classificadas em simétricas, assimétricas à esquerda (assimetria negativa) ou assimétricas à direita (assimetria positiva).

b) As medidas de assimetria (ou de enviesamento) indicam o grau de deformação de uma curva de distribuição de frequências.

c) As medidas de curtose indicam o grau de achatamento ou de afilamento de uma curva representativa de uma distribuição de frequências.

d) Quanto à curtose uma curva de distribuição de frequências pode ser classificada em mesocúrtica, platicúrtica ou leptocúrtica.

e) Se o coeficiente de assimetria (S_K) para uma determinada distribuição de frequência for menor do que zero, pode-se afirmar que a curva é do tipo mesocúrtica.

13. Uma urna contém 7 bolas brancas numeradas com números impares distintos e 4 bolas amarelas numeradas com números pares, também distintos. Em outra urna, contém 5 bolas brancas numeradas com números impares e 2 bolas amarelas numeradas com números pares. Uma bola é retirada de cada urna aleatoriamente. Qual a probabilidade das bolas retiradas serem ambas pares ou ambas impares? (Obs.: Efetue os cálculos com duas casas decimais)

a) $\frac{34}{77}$ ou 44,16%.

b) $\frac{43}{77}$ ou 55,84%.

c) $\frac{45}{77}$ ou 58,44%.

d) $\frac{6}{7}$ ou 85,71%.

e) $\frac{9}{16}$ ou 56,25%.

14. Uma urna contem três bolas numeradas com 1, 2 e 3. Retirando-se sucessivamente duas bolas dessa urna, obtém-se um par ordenado. O número de pares ordenados possíveis, fazendo-se extrações com reposição, é:

a) 9.

b) 6.

c) 5.

d) 8.

e) 3.

15. Uma urna contem três bolas numeradas com 1, 2 e 3. Retirando-se sucessivamente duas bolas dessa urna, obtém-se um par ordenado. O número de pares ordenados possíveis, fazendo-se extrações sem reposição, é:

a) 5.

b) 3.

c) 8.

d) 9.

e) 6

16. **Uma urna contem três bolas numeradas com 1, 2 e 3. Retirando-se simultaneamente duas bolas dessa urna, obtém-se um conjunto. O número de conjuntos possíveis é:**

a) 8.

b) 5.

c) 6.

d) 3.

e) 9.

17. **Lançando-se uma moeda usual 5 vezes, seus resultados formam uma sequência. O número de sequências possíveis é:**

a) 2.

b) 5.

c) 10.

d) 25.

e) 32.

18. **Considere o seguinte experimento aleatório: "lançar dois dados e observar os números obtidos nas faces superiores". O número de elementos do espaço amostral desse experimento é:**

a) 6.

b) 12.

c) 2.

d) 64.

e) 36 .

19. **Uma moeda é lançada três vezes. Vamos representar por n (E) o número de resultados possíveis e representar por n(A) o número de resultados que apresentam apenas duas caras. Então:**

a) n (E) = 6 e n (A) = 3.

b) n (E) = 6 e n (A) = 4.

c) c.n (E) = 8 e n (A) = 4.

d) d.n (E) = 8 e n (A) = 6.

e) n (E) = 8 e n (A) = 3.

20. **Lançando-se um dado honesto duas vezes, o número de resultados que apresentam soma 7, é:**

a) 4.

b) 5.

c) 6.

d) 7.

e) 3.

21. **Uma urna tem 20 bolas numeradas com 1, 2, 3..., 20. Sorteia-se uma bola dessa urna. Considere os seguintes eventos:**
Evento A : Ocorrência de um número primo
Evento B : Ocorrência de um divisor de 30
Nesse experimento, o número de elementos do evento $A \cup B$ é:

a) 16.

b) 15.

c) 13.

d) 14.

e) 12.

22. **Dois jogadores disputam um jogo onde é lançado, uma única vez um par de dados. O jogador A ganha se a soma dos resultados for 6 e B, se a soma for 10. Nessas condições, pode-se afirmar CORRETAMENTE que:**

a) B tem mais chance de ganhar que A.

b) A não tem chance de ganhar.

c) A tem mais chance de ganhar que B.

d) B não tem chance de ganhar.

e) Ambos têm as mesmas chances.

23. **Denomina-se espaço amostral ao conjunto formado por todos os resultados possíveis de um experimento aleatório. Se um experimento consiste em se escolherem duas pessoas, ao acaso, de uma sala contendo dez pessoas, então o número de elementos do espaço amostral é:**

a) 20.

b) 19.

c) 90.

d) 45.

e) 32.

01	02	03	04	05	06	07	08	09	10					
A	E	D	E	E	E	C	E	C	D					

24. **Em uma distribuição de frequências, verificou-se que a moda é igual a 8,0, a média é igual a 7,8 e o desvio padrão é igual a 1,0. Podemos dizer que o coeficiente de assimetria de Pearson é.**

a) 0,20.

b) -0,20.

c) 2,0.

d) -2,0.

e) 0,50.

25. **Em uma distribuição de frequências, verificou que a mediana é igual a 15,4, a média é igual a 16,00 e o desvio padrão é igual a 6,0. Podemos dizer que o coeficiente de assimetria de Pearson vale:**

a) 0,10.

b) -0,10.

c) 0,30.

d) -0,30.

e) 0,50.

26. **Observou-se que, em uma distribuição de frequência, o primeiro quartil é igual a 3, o terceiro quartil é igual a 8, o décimo centil é igual a 1,5 e o nonagésimo centil é igual a 9. Com base nesses resultados, podemos afirmar que trata-se de uma curva:**

a) mesocurtica, com K = 0,263.

b) leptocúrtica, com k = 0,233.

c) leptocúrtica, com k = 0,25.

d) platicúrtica, com k = 0,45.

e) platicúrtica com k = 0,333.

27. **Com o objetivo de estudar a eficácia de um regime alimentar para tratamento de diabetes foram recolhidas 12 amostras de sangue em diabéticos e analisada a quantidade de açúcar. Obtiveram-se os seguintesresultados (em mg/100ml):**

187.45	**187.57**	**187.37**	**187.49**	**187.58**	**187.37**
187.46	**187.62**	**187.47**	**187.53**	**187.39**	**187.46**

Com base nesses dados podemos afirmar que:

a) a média dos pesos é 60.

b) a moda é 60.

c) a amplitude total é 55.

d) a mediana é 62,08.

e) a mediana é 60.

28. **Leia e marque uma das alternativas.**

Os pesos em kg de uma amostra de 25 alunos do curso de Psicologia da UFPB em 2005 foram: 35,40, 43, 45, 47, 49, 50, 50, 55, 58, 59, 60, 60, 64, 65, 65, 70, 70, 72, 75, 80, 80, 80, 85 e 95. Com base nessa informação, podemos concluir que:

a) a média dos pesos é 60.

b) a moda é 60.

c) a amplitude total é 55.

d) a mediana é 62,08.

e) a mediana é 60.

29. **Leia o enunciado e marque a alternativa CORRETA.**
Os dados abaixo representam os escores resultantes da aplicação de um teste para verificar o quociente de inteligência (Q.I.) dos candidatos a vagas de emprego de uma determinada empresa do setor logístico, atuante no Paraná.

82 84 80 90 80 82 80 82 76 86 76
88 90 78 78 78 88 78 84 80

Com base nesses escores, podemos afirmar que a moda é:

a) 78.

b) 80.

c) 84.

d) 90.

e) 82.

30. **Leia o enunciado e marque a alternativa CORRETA.**
Com o objetivo de analisar o tempo de permanência dos estudantes da UFPR na Biblioteca Central, pesquisou-se em certo dia, um grupo de 50 estudantes. O resultado da pesquisa encontra-se na tabela a seguir:

Tempo de permanência (min)	Nº de estudantes
50 ⊢ 70	5
70 ⊢ 90	12
90 ⊢ 110	13
110 ⊢ 130	15
130 ⊢ 150	5
	Total 50

Com base na tabela, podemos afirmar que:

a) O número de classes dessa distribuição de frequências é 6.

b) Que os pontos médios dessa distribuição são: 60, 80, 100, 120.

c) Que a frequência acumulada da terceira classe é 30.

d) Que a frequência absoluta da ultima classe é 50.

e) Que a moda dessa distribuição é 15.

31. **As companhias de seguro pesquisam continuamente as idades na morte e as causas de morte. Os dados se baseiam no estudo levado a efeito pela revista Time sobre as pessoas que morreram vitimadas por armas de fogo durante uma semana, conforme tabela abaixo.**

Idade na morte	Frequência
16 ⊢ 26	22
26 ⊢ 36	10
36 ⊢ 46	6
46 ⊢ 56	2
56 ⊢ 66	4
66 ⊢ 76	5
76 ⊢ 86	1

O primeiro decil dessa distribuição é:

a) 22.

b) 5.

c) 18,27.

d) 9,55.

e) 76.

32. **A distribuição abaixo revela a quantidade de casas alugadas em função do valor do aluguel.**

Aluguel (x R$100)	Quantidade de casas
4 ⊢ 6	18
6 ⊢ 8	25
8 ⊢ 10	32
10 ⊢ 12	40
12 ⊢ 14	30
14 ⊢ 16	18
16 ⊢ 18	12

Podemos afirmar que o terceiro quarto (quartil) dessa distribuição é:

a) 32.

b) 75.

c) 43,75.

d) 13,51.

e) 131,25.

33. **Da distribuição abaixo, podemos afirmar que o 54º percentil é:**

NOTAS	f_i
0 ⊢ 2	5
2 ⊢ 4	8
4 ⊢ 6	14
6 ⊢ 8	10
8 ⊢ 10	7
	44

a) 4,4.

b) 0,44.

c) 23,76.

d) 5,54.

e) 7,54.

34. **Na distribuição de frequência abaixo, temos a estaturas de 70 pessoas,**

Estaturas	f_i
150 ⊢ 158	5
158 ⊢ 166	12
166 ⊢ 174	18
174 ⊢ 182	27
182 ⊢ 190	8
	70

Podemos afirmar que:

a) a amplitude total é 40.

b) a maior frequência é 70.

c) que a frequência acumulada é da quarta classe é 35.

d) que a amplitude do intervalo é 7.

e) que a mediana é 35.

35. **Com base na distribuição de frequência abaixo, resultante dos salários de 70 funcionários.**

Salários (R$)	f_i
500 ⊢ 700	18
700 ⊢ 900	31
900 ⊢ 1100	15
1100 ⊢ 1300	3
1300 ⊢ 1500	1
1500 ⊢ 1700	1
1700 ⊢ 1900	1
	70

Podemos afirmar que nessa distribuição o intervalo usado é:

a) Aberto a esquerda.

b) Fechado a esquerda.

c) Aberto.

d) Fechado.

e) Aberto a esquerda e a direita.

36. **Os seguintes dados referem-se ao salário (em R$) de 40 funcionários da empresa GIS.**

1638	2456	1715	1330	1398	1230	3153	1275
2154	704	2000	2622	1319	3050	3502	1154
1271	1894	780	1137	3296	2169	2578	654
2415	1104	756	2634	802	2846	1158	486
1317	516	1351	585	1440	588	1243	1674

a) A media aritmética dos salários é R$1.650,00.

b) A mediana tem como valor o R$1.340,50.

c) A amplitude total é R$2.499,00.

d) A moda dos salários é R$1.650,00.

e) A media aritmética é acima de R$1.650,00.

37. **Leia o enunciado e responda marcando uma alternativa (adaptado de Crespo, 2007)**
Ao nascer, os bebês são pesados e medidos, para se saber se estão dentro das tabelas de peso e altura esperados. Estas duas variáveis são:

a) Qualitativas.

b) Ambas discretas.

c) Ambas contínuas.

d) Contínua e discreta, respectivamente.

e) Discreta e contínua, respectivamente.

38. Leia o enunciado e responda marcando uma alternativa (adaptado de Crespo, 2007).

A parcela da população convenientemente escolhida para representa-la é chamada de:

a) Variável.

b) Rol.

c) Amostra.

d) Dados brutos.

e) Nada podemos afirmar, porque a informação é incompleta.

39. Leia o enunciado e responda marcando uma alternativa (adaptado de Crespo, 2007).

Em uma empresa, os custos estão separados por setor. 70% desse custo vão para o setor produtivo, 12% para o setor logístico e manutenção e 18% para os setores administrativos, comercial e rh. O gráfico que melhor representa essa situação é:

a) O de linha.

b) O de barras.

c) O de setores.

d) O histograma.

e) O de Pareto.

40. Leia o enunciado e responda marcando uma alternativa (adaptado de Crespo, 2007).

Um conjunto de 100 peças de calibradores, retirados de um estoque de pneus, constitui:

a) um rol.

b) uma tabela.

c) uma relação de dados brutos.

d) uma distribuição de frequência.

e) uma população.

41. Dado o conjunto de números A = {8, 4, 6, 9, 10, 5}, determine o desvio médio dos elementos desse conjunto, supondo que esses valores correspondem a uma amostra.

a) 28.

b) 2,3664.

c) 7.

d) 2,8.

e) 2.

42. Podemos afirmar que a mediana do conjunto de números da questão anterior, é:

a) 7.

b) 6.

c) 9.

d) 8.

e) 7,5.

43. Com base na distribuição de frequência abaixo, resultante dos salários de 70 funcionários.

Salários (R$)	f_i
500 ⊢ 700	18
700 ⊢ 900	31
900 ⊢ 1100	15
1100 ⊢ 1300	3
1300 ⊢ 1500	1
1500 ⊢ 1700	1
1700 ⊢ 1900	1
	70

Podemos afirmar que:

a) sua curva tem comportamento normal.

b) que ela é assimétrica positiva.

c) que ela é assimétrica negativa.

d) ela tem uma curtose placurtica.

e) ela tem um comportamento leptocurtico.

44. **As notas de um candidato nas provas de um concurso foram: 8,4; 9,1;7,2; 6,8; 8,7 e 7,2. A nota média, a nota mediana e a nota modal dessealuno são respectivamente:**

a) 7,9; 7,8; 7,2.

b) 7,1; 7,8; 7,9.

c) 7,8; 7,8; 7,9.

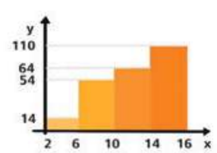
d) 7,2; 7,8; 7,9.

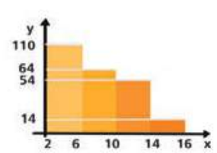
e) 7,8; 7,9; 7,2.

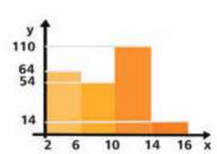
45. **(Exercício adaptado da livro ADM/Estatística IFPR) Os dados da tabela abaixo, foram gerados pelo departamento de qualidade de uma empresa. Ela representa a distribuição de frequência, de 242 peças defeituosas, retiradas do setor de produção do lote nº 2896/2009.**

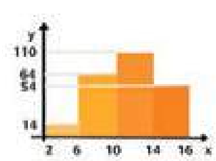
	X: classe	y: número de peças com defeito
I	2 ⊢ 6	14
II	6 ⊢ 8	54
III	10 ⊢ 14	110
IV	14 ⊢ 16	64

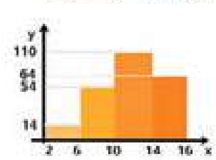
O histograma que melhor se adapta aos dados da tabela e:

a)


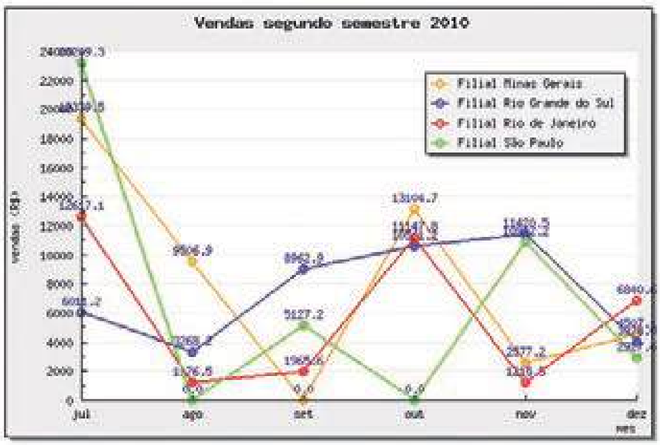
b)


c)


d)


e)


46. **Numa grande empresa, o departamento comercial elaborou um gráfico, referente às vendas do segundo semestre de 2010, de suas quatro filiais.**



Fonte: http://www.adianti.com.br/bkrpt

Pela análise do gráfico, e CORRETO afirmar que:

a) O número de vendas das quatro filiais no mês de setembro passou de 20.000,00.

b) A filial de Minas Gerais superou no mês de novembro a filial de São Paulo.

c) As vendas, da filial do Rio Grande do Sul, no mês de julho foram nula.

d) O gráfico de linha do departamento de vendas está claro, e mostra a evolução das vendas das filiais.

e) No mês de outubro de 2010, as filiais de Rio Grande do Sul e Rio de Janeiro, obtiveram as mesmas receitas de vendas.

47. **Analise a série abaixo e marque a alternativa que corresponde ao tipo de série.**

A serie representa as Aplicações (em milhões R$) do ultimo quadrimestre de 2008, do banco PoupeMais.

Mês	Aplicações
Setembro	20,3
Outubro	22,2
Novembro	23,1
Dezembro	21,0

a) cronológica.

b) geográfica.

c) especifica.

d) distribuição de frequência.

e) nenhuma das anteriores.

As questões de número 48 e 49 devem ser respondidas com base na Distribuição abaixo, referente aos salários de 70 empregados de uma empresa.

Salários (R$)	f_i
500 ├── 700	18
700 ├── 900	31
900 ├── 1100	15
1100 ├── 1300	3
1300 ├── 1500	1
1500 ├── 1700	1
1700 ├── 1900	1
	70

48. O limite superior da quinta classe é:

a) 1300.

b) 1900.

c) 1500.

d) 1100.

e) 1450.

49. As frequências acumulada da terceira e da quinta classe são, respectivamente:

a) 15 e 3.

b) 3 e 1.

c) 64 e 67.

d) 64 e 68.

e) 64 e 66.

50. A tabela a seguir representa a distribuição de frequências dos salários de um grupo de 50 empregados de uma empresa, no mês de julho de 2006.

Número de classe	Salário do mês em R$	Número de empregados
1	1000 ├── 2000	20
2	2000 ├── 3000	18
3	3000 ├── 4000	9
4	4000 ├── 5000	3

O salario médio desses empregados, nesse mês, foi de:

a) R$2 637,00.

b) R$2 520,00.

c) R$2 500,00.

d) R$2 420,00.

e) R$2 400,00.

Currículo do professor-autor

Marcos Antonio Barbosa

Graduado em Matemática pela Universidade Tuiuti do Paraná - UTP, é especialista em Educação Matemática e mestre em Educação, pela Pontifícia Universidade Católica do Paraná - PUCPR. Atualmente é professor e coordenador da gestão de polos/EAD do Instituto Federal do Paraná. Tem experiência na área de Educação, com ênfase em Educação Matemática e Educação a Distância, atuando principalmente nos seguintes temas: didática; prática de ensino, coordenação de cursos de graduação e especialização em EaD, gestão de polos em EaD.