



profuncionário

Curso Técnico de Formação para
os Funcionários da Educação

Estatística aplicada à educação

**TÉCNICO EM
GESTÃO ESCOLAR**

Brasília – 2007

Governo Federal

Ministro da Educação

Secretária de Educação Básica

Diretor do Departamento de Articulação e Desenvolvimento dos Sistemas de Ensino

Universidade de Brasília – UnB

Reitor

Timothy Martin Mulholland

Vice-Reitor

Edgar Nobuo Mamiya

Coordenação Pedagógica do Profuncionário

Bernardo Kipnis – FE/UnB

Dante Diniz Bessa – Cead/UnB

Francisco das Chagas Firmino do Nascimento – SEE-DF

João Antônio Cabral de Monlevade – FE/UnB

Maria Abádia da Silva – FE/UnB

Tânia Mara Piccinini Soares – MEC

Centro de Educação a Distância – Cead/UnB

Diretor – Sylvio Quezado de Magalhães

Coordenação Executiva – Ricardo de Sagebin

Coordenação Pedagógica – Tânia Schmitt

Unidade de Pedagogia

Gestão da Unidade Pedagógica – Ana Luísa Nepomuceno

Gestora Pedagógica – Juliana C. Jungmann

Gestão da Unidade Produção – Rossana M. F. Beraldo

Designer Educacional – Luciana Kury

Revisão – Laetícia Jensen Eble

Editoração – Raimunda Dias

Ilustração – Rodrigo Mafra

Unidade de Apoio Acadêmico e Logístico

Gerente da Unidade – Lourdinéia Martins da Silva Cardoso

Gestora do Projeto – Diva Peres Gomes Portela



profuncionário

**Curso Técnico de Formação para
os Funcionários da Educação**

Dados Internacionais de Catalogação na Publicação (CIP)

Brasil. Ministério da Educação. Secretaria de Educação Básica.

M488e Medeiros, Carlos Augusto de.

Estatística aplicada à educação. / Carlos Augusto de Medeiros. – Brasília : Universidade de Brasília, 2007.

130 p. : il.

ISBN 978-85-230-0990-8

1. Conceitos matemáticos: razões e proporções. 2. Distribuição de frequência: dados brutos e rol. 3. Medidas de resumo: medidas de tendência central (média, média aritmética ponderada, mediana e moda). I. Título. II. Universidade de Brasília. Centro de Educação a Distância.

CDU 519.2:37(81)

Apresentação

Sou professor! Não há outra atividade profissional em minha vida. Iniciei minha carreira há, aproximadamente, 15 anos, como professor de Matemática, no Ensino Fundamental, na Rede Pública de Ensino do Distrito Federal.

Nos últimos 5 anos, tenho me dedicado à docência no nível superior, atuando em cursos de Formação para Docentes, basicamente, com componentes como Metodologia Científica; Metodologia da Pesquisa; Métodos e Técnicas de Pesquisa; Organização da Educação Brasileira e Planejamento e Políticas Educacionais.

Fiquei muito feliz com o convite para escrever este Módulo de “Estatística aplicada à Educação”. É bem verdade que, como professor de Matemática, sei por experiência própria que trabalhar com cálculos repele mais do que atrai o leitor. Mas, também, da forma como têm sido trabalhadas as ciências exatas nas escolas, não é de se estranhar.

Foi nesse contexto que resolvi apresentar aos “Funcionários da Educação” uma ferramenta valiosa, fincada na Matemática, que auxilia na interpretação da realidade. Sem ela, nossas ações se pautam por bases outras que não a ciência. E isso implica acertar, algumas vezes, mas errar, outras tantas vezes.

É claro que não há receita segura para o acerto, isso todos sabemos. Mas existem ferramentas que, por força do nosso percurso individual, vão sendo oferecidas a alguns poucos que se tornam detentores dos saberes e isso não posso aceitar. Dentre essas ferramentas, a Estatística figura como (quem sabe!) uma dessas que, se não observada, confina nossas ações ao campo da “sorte”.

Mas ainda assim, reconhecendo sua importância, é preciso lidar com as resistências e limitações de todos nós, com o “traquejo algébrico”, isto é, com números, números e números.

Pois bem, estava ciente disso tudo quando escrevi esse Módulo. Tudo que escrevi buscou responder à seguinte pergunta: o que da Estatística Básica pode ser oferecido aos “Funcionários da Educação” de modo que os auxiliem em suas atividades diuturnas, caminhando no sentido de uma educação de qualidade?

Com isso em mente, procurei colocar em um prato da balança aquilo que efetivamente poderia contribuir para alcançar a tão sonhada “qualidade da educação” e, no outro prato, metodologias e procedimentos de resolução, com os fundamentos para aqueles que desejarem se aprofundar no futuro, pautados em estratégias que levem aos resultados.

Por isso, caro leitor, algumas vezes é possível que você tenha que recorrer a recursos externos para a melhor compreensão dos conteúdos. Mas se isso acontecer, serão poucas vezes, já que me empenhei para consolidar os conteúdos no interior deste Módulo.

As fórmulas, leitor, deixe que as calculadoras e as planilhas eletrônicas resolvam. A nós cabe, contudo, saber o que representam os resultados, bem como de que maneira organizar os dados para que cheguemos a eles. A nós compete identificar as ferramentas que contribuem para dar mais qualidade às nossas atividades profissionais.

Transformar dados em informação: esse é o desafio!

Objetivo do Módulo

Refletir a partir da Estatística Básica sobre as ferramentas consolidadas pelo uso e pela ciência, disponíveis a todos, que auxiliam na tomada de decisão.

Ementa

Conceitos matemáticos: razões e proporções; grandezas e medidas; regra de três simples; porcentagem; coeficientes, taxas e índices; sistema de coordenadas cartesianas; arredondamento. Variáveis, tabelas e gráficos: população e amostra; estatística descritiva e estatística indutiva ou inferencial; variáveis; tabelas; gráficos: diagramas, cartogramas e pictogramas. Distribuição de frequência: dados brutos e rol; distribuição de frequência: gráficos de uma distribuição; curvas de frequência. Medidas de resumo: medidas de tendência central (média, média aritmética ponderada, mediana e moda); medidas de dispersão (dispersão e variação, desvio padrão e coeficiente de variação); medidas de posição (quartis, decis e percentis).

Mensagem do autor

Amigos e amigas Trabalhadores da Educação!

Durante 14 anos, dediquei-me à docência da matemática, na Educação Básica, na rede pública de ensino do Distrito Federal. Após a conclusão do mestrado, tenho me dedicado à docência em cursos de formação docente, em instituições de ensino privadas.

Foi com muito prazer que recebi e prontamente aceitei o desafio para escrever sobre Estatística aplicada à Educação como etapa da formação promovida pelo Profucionário. A idéia que me motivou, do início ao fim dessa jornada, foi a de traduzir, na medida do possível, a linguagem rígida da matemática de modo a permitir a aproximação aos recursos da Estatística (e da Matemática) como ferramenta para o desenvolvimento de suas atividades profissionais.

Tenho claro que a dificuldade com a disciplina é historicamente significativa. E mais ainda, que essa dificuldade se torna instrumento de exclusão social e não o contrário. Por isso, empenhei-me para que vocês pudessem perceber que, com um pouco de disciplina, é possível fazer uso dos recursos da Estatística para a melhoria das condições de trabalho, bem como para a melhoria dos resultados desse trabalho na educação.

É assim que admito que, em alguns momentos (poucos, eu espero!), vocês se sentirão cansados de tantos números. Mas não desanimem! A alguém que porventura abra o Módulo de Estatística aplicada à Educação em alguma página aleatoriamente, poderá parecer algo bem difícil. Mas, na verdade, esforcei-me para construir degraus que conduzam do mais simples ao mais complexo, por isso, é preciso ter claro que cada página lida prepara-os para as páginas seguintes.

Penso que nossa luta reside na seguinte questão: a quem interessa a Estatística? Se ao final da leitura vocês chegarem à conclusão de que interessa a todos nós e que, por isso, não pode se limitar a um número determinado de especialistas, então, alcancei meu objetivo maior: contribuir diretamente para sua formação e, conseqüentemente, para uma educação pública de qualidade.

Forte abraço,

Carlos Augusto de Medeiros

Lista de Figuras

- Figura 1: Estatística: Pirâmide da definição **18**
- Figura 2: Razão: Comparação **24**
- Figura 3: Razão: Exercício **25**
- Figura 4: Razão: Representação **25**
- Figura 5: Proporções: Conceito **26**
- Figura 6: Razões: Proporções: Escala **27**
- Figura 7: Razões e Proporções: Exercício **27**
- Figura 8: Grandezas **28**
- Figura 9: Medida de Comprimento: Segmento de reta **29**
- Figura 10: Regra de Três: Exercício **31**
- Figura 11: Coeficiente e Taxa **34**
- Figura 12: Sistema de Coordenadas Cartesianas: Origem **37**
- Figura 13: Sistema de Coordenadas Cartesianas: Eixos **38**
- Figura 14: Sistema de Coordenadas Cartesianas: Pontos **38**
- Figura 15: Sistema de Coordenadas Cartesianas: Exercício **39**
- Figura 16: Arredondamento de Números **40**
- Figura 17: Arredondamento: Fluxograma **40**
- Figura 18: Estatística Dedutiva e Estatística Indutiva: Fluxograma **46**
- Figura 19: Variáveis: Definições **48**
- Figura 20: Pictograma: Exemplo **61**
- Figura 21: Modelo de Histograma **69**
- Figura 22: Polígono de Freqüência: Esboço **70**
- Figura 23: Curvas de Freqüência **76**
- Figura 24: Média Aritmética: Exemplo **83**
- Figura 25: Linha Mediana **92**
- Figura 26: Curvas Modais **95**
- Figura 27: Média, Mediana, Moda: Curva Simétrica **96**
- Figura 28: Média, Mediana, Moda: Curva Assimétrica **96**

Figura 29: Desvio Padrão: Gráficos: Exercício **101**

Figura 30: Quartis: Representação **111**

Figura 31: Tabela de Freqüência: Ilustração **115**

Figura 32: Exercício: Quartis **117**

Figura 33: Exercício: Quartis: Freqüência Acumulada Anterior **118**

Lista de Fórmulas

Fórmula 1: Média Aritmética **81**

Fórmula 2: Média Aritmética Ponderada **85**

Fórmula 3: Mediana **91**

Fórmula 4: Desvio Padrão: Dados Não-Agrupados: **99**

Fórmula 5: Desvio Padrão: Dados Agrupados **102**

Fórmula 6: Coeficiente de Variação **106**

Fórmula 7: Medidas de Posição: Dados Não-Agrupados: Quartil **112**

Fórmula 8: Medidas de Posição: Quartil **112**

Fórmula 9: Medidas de Posição: Dados Não-Agrupados: Decil **122**

Fórmula 10: Medidas de Posição: Dados Não-Agrupados: Percentil **122**

Fórmula 11: Medidas de Posição: Percentil **123**

Lista de Gráficos

Gráfico 1: Nº de matrículas no Ensino Médio: Brasil: Urbano **53**

Gráfico 2: Matrículas na pré-escola: Brasil: 1999-2004 **56**

Gráfico 3: Evolução das matrículas na creche: Brasil: 1999-2004 **56**

Gráfico 4: Evolução das matrículas na educação infantil: creche e pré-escola: Brasil: 1999-2004 **57**

Gráfico 5: Usuários de transporte público do Estado: 1ª a 4ª séries: Brasil: área urbana **59**

Gráfico 6: O despovoamento da Amazônia **60**

Gráfico 7: Exercício: Polígono de Freqüência **74**

Gráfico 8: Mediana **93**

Lista de Quadros

Quadro 1: As fases de desenvolvimento da Estatística **17**

Quadro 2: Tipos de variáveis **49**

Quadro 3: Níveis de medidas **80**

Quadro 4: Quartil e Percentil: Fórmula Geral: Comparação **124**

Lista de Tabelas

Tabela 1: População: Brasil **32**

Tabela 2: Aprovação: Ensino Fundamental: Brasil: 2005 **35**

Tabela 3: Função Docente: Educação Básica: Brasil: 2005 **36**

Tabela 4: Aprovação: Ensino Fundamental: Rural: Brasil: 2005 **37**

Tabela 5: População Escolar: Sexo **44**

Tabela 6: Cálculo da amostragem proporcional estratificada **45**

Tabela 7: População Mundial: Série Histórica **51**

Tabela 8: Matrículas no Ensino Fundamental de 5ª a 8ª série: Diurno: Brasil **52**

Tabela 9: Número de matrículas na pré-escola **52**

Tabela 10: N° de matrículas no Ensino Médio: Brasil: Urbano **53**

Tabela 11: Matrículas na Educação Infantil: Brasil **55**

Tabela 12: Usuários de transporte público do Estado: 1ª a 4ª séries: Brasil: área urbana **57**

Tabela 13: Pictograma: Exercício **61**

Tabela 14: Exemplo de Tabela Primitiva **64**

Tabela 15: Exemplo de Rol **65**

Tabela 16: Exemplo de Tabela de Freqüência **66**

Tabela 17: Exemplo de Tabela de Distribuição de Freqüência **66**

Tabela 18: Exemplo de Tabela de Distribuição de Freqüência **68**

Tabela 19: Exercício: Tabela Primitiva **71**

Tabela 20: Exercício: Rol **72**

Tabela 21: Exercício: Tabela de Freqüência **72**

Tabela 22: Exercício: Tabela de Freqüência com intervalos de classe **74**

Tabela 23: Série Histórica: Exercício **84**

Tabela 24: Distribuição de Frequência: Exercício **85**

Tabela 25: Distribuição de Frequência: Exercício: Ponderação **86**

Tabela 26: Distribuição de Frequência: Exercício: Ponderação: Ponto Médio **87**

Tabela 27: Vítimas de Acidentes de Trânsito, por 10.000 veículos, em 2002 **88**

Tabela 28: Distribuição de Frequência: Exercício: Mediana: Frequência Acumulada **91**

Tabela 29: Desvio Padrão: Exercício **100**

Tabela 30: Desvio Padrão: Dados Agrupados: Sem Intervalos de Classe: Exercício **102**

Tabela 31: Desvio Padrão: Exercício: Continuação **103**

Tabela 32: Desvio Padrão: Dados Agrupados: Com Intervalos de Classe: Exercício **104**

Tabela 33: Desvio Padrão: Exercício: Continuação **105**

Tabela 34: Distribuição de Frequência: Exercício: Quartis **113**

Tabela 35: Medidas de Posição: Quartis: Exercício: Tabela-Resposta **113**

Tabela 36: Medidas de Posição: Quartis: Exercício: Tabela-Resposta: Preenchimento: 2ª etapa **114**

Tabela 37: Distribuição de Frequência: Exercício: Quartis: Primeiro Quartil **117**

Tabela 38: Medidas de Posição: Quartis: Exercício: Tabela-Resposta: Preenchimento: 3ª etapa **118**

Tabela 39: Medidas de Posição: Quartis: Exercício: Tabela-Resposta: Preenchimento: 4ª etapa **119**

Tabela 40: Medidas de Posição: Quartis: Exercício: Tabela-Resposta: Preenchimento: 5ª etapa **119**

Tabela 41: Medidas de Posição: Quartis: Exercício: Tabela-Resposta: Preenchimento: 6ª etapa **120**

Tabela 42: Exercício: Quartis **121**

Tabela 43: Medidas de Posição: Percentil: Tabela-Resposta **123**

Tabela 44: Medidas de Posição: Percentis: Exercício: Tabela-Resposta: Preenchida **124**

Sumário

UNIDADE 1 – Introdução ao estudo da estatística **15**

UNIDADE 2 – Conceitos matemáticos **23**

UNIDADE 3 – Variáveis, tabelas e gráficos **43**

UNIDADE 4 – Distribuição de frequência **63**

UNIDADE 5 – Medidas de resumo **79**

CONSIDERAÇÕES FINAIS 126

REFERÊNCIAS 127

APÊNDICE: Respostas dos exercícios Pratique! **130**

1

Introdução ao estudo da estatística

SAIBA MAIS



*A população mundial está estimada hoje em mais de seis bilhões e meio de habitantes (6.600.000.000). Para daqui a trinta anos está estimada uma população de mais de oito bilhões e meio de habitantes no planeta (8.547.874.779).
Fonte: U.S. CENSUS Bureau, 2006.*

Você sabe quantas pessoas existem na sua casa? Com certeza. Mas em toda a sua família, você sabe? Bem... Quantas pessoas existem na sua rua? E no seu bairro? E na sua cidade? E no seu estado? E no Brasil? E no **mundo**, afinal? Bem, pode ser que você considere essas preocupações bastante exageradas, mas nem sempre o mundo foi tão populoso.

Se pararmos para pensar na população mundial de um tempo atrás, digamos, no século XV, veremos que a quantidade de pessoas era bem menor. Se voltássemos à Grécia Antiga, menor ainda. Pois bem, esse crescimento acelerado de habitantes foi verificado no mundo moderno, com a sociedade de massas. A partir daí, a Estatística se tornou, juntamente com a ciência da economia, a ciência social por excelência.¹ Por quê? Porque lidamos com grandes números.

A Estatística ou métodos estatísticos, como é chamada algumas vezes, nasceu com os negócios do Estado, daí seu nome. Mas, hoje, sua influência pode ser encontrada nas mais diversas atividades: agricultura, biologia, comércio, química, comunicações, economia, educação, medicina, ciências políticas e muitas outras.²

A Estatística se interessa pelos métodos científicos para coleta, organização, resumo, apresentação e análise de dados, bem como na obtenção de conclusões válidas e na tomada de decisões razoáveis baseadas em tais análises. Algumas vezes, o termo **Estatística** é empregado para designar os próprios dados ou números, por exemplo, estatística de empregos, de acidentes etc.³

Se a Estatística ganha importância com a moderna sociedade de massas, como vimos, não significa que, antes disso, não existissem preocupações com os cálculos de grandes números.

Na história, vemos que a palavra Estatística apareceu pela primeira vez no século XVIII e foi sugerida pelo alemão Gottfried Achemmel (1719-1772); palavra esta que deriva de *statu* (estado, em latim). Como se pode perceber, Estatística é um nome que deriva de Estado; de fato, na origem, as atividades da Estatística eram, basicamente, atividades de Estado. Mas hoje isso mudou bastante.

16

Estatística é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados. Ela é dividida em:

- 1) *Estatística Descritiva: parte da Estatística que apenas coleta, descreve, organiza e apresenta os dados. Nela não são tiradas conclusões.*
- 2) *Estatística Indutiva ou Inferência: analisa os dados e obtém as conclusões.*

1 ARENDT (2005, p. 51).

2 SPIEGEL (1975, Prefácio).

3 SPIEGEL (1975, p. 1).

O primeiro levantamento estatístico de que se tem conhecimento se deve a **Heródoto** e se refere a um estudo da riqueza da população do Egito, cuja finalidade era averiguar quais eram os recursos humanos e econômicos disponíveis para a construção das pirâmides, isso no ano de 3050 a. C. No ano de 2238 a. C., o **Imperador Chinês Yao** ordenou a realização de uma Estatística com fins industriais e comerciais. No ano de 1400 a. C., o famoso faraó egípcio **Ramsés II** ordenou um levantamento das terras do Egito. Existem ainda, outros casos de Estatísticas no período antigo⁴ da civilização.

Em períodos mais recentes, podemos sintetizar as preocupações com a Estatística em quatro fases:

Primeira Fase	Pepino, no ano de 758, e Carlos Magno, em 762, realizaram estatísticas sobre as terras que eram propriedade da Igreja. Essas foram as únicas estatísticas importantes desde a queda do Império Romano.
Segunda Fase	Na Inglaterra, no século XVII, já se analisavam grupos de observações numéricas referentes à saúde pública, nascimentos, mortes e comércio. Destacam-se, nesse período, John Graunt (1620-1674) e William Petty (1623-1687) que procuraram leis quantitativas para traduzir fenômenos sociais e políticos.
Terceira Fase	Também no século XVII, inicia-se o desenvolvimento do Cálculo das Probabilidades que, juntamente com os conhecimentos estatísticos, redimensionou a Estatística. Nessa fase, destacam-se: Fermat (1601-1665), Pascal (1623-1662) e Huygens (1629-1695).
Quarta Fase	No século XIX, inicia-se a última fase do desenvolvimento da Estatística, alargando e interligando os conhecimentos adquiridos nas três fases anteriores. Nesta fase, a Estatística não se limita apenas ao estudo da Demografia e da Economia, como antes; agora, o seu campo de aplicação se estende à análise de dados em Biologia, Medicina, Física, Psicologia, Indústria, Comércio, Meteorologia, Educação etc., e ainda, a domínios aparentemente desligados, como Estrutura de Linguagem e estudo de Formas Literárias. Destacam-se, no período, Ronald Fisher (1890-1962) e Karl Pearson (1857-1936).

Fonte: História da Estatística (2006)

Quadro 1: As fases de desenvolvimento da Estatística

Como se vê, a Estatística possui sua história na História do homem. Nessa última fase, com a Estatística consolidada, as

4 Podemos considerar os períodos da História com alguns marcos cronológicos: 1) **Pré-História**: até 4000 a. C., período do surgimento da escrita; 2) **Idade Antiga**: do aparecimento da escrita e das primeiras civilizações, por volta de 4000 a. C., até a queda de Roma, em 476 d. C.; 3) **Idade Média**: da queda de Roma até a tomada de Constantinopla pelos turcos otomanos, em 1453; 4) **Idade Moderna**: da queda de Constantinopla até a tomada da Bastilha, em 1789 (Revolução Francesa); 5) **Idade Contemporânea**: da tomada da Bastilha aos dias atuais.



"Heródoto (gr. Ηρόδοτος) é o mais importante dos historiadores gregos mais antigos. Foi o primeiro prosador a reunir diversas narrativas históricas ou quase-históricas em um relato coerente e vivo e é, por isso, considerado o pai da História."



"Yao era descendente do Imperador Amarelo, o primeiro antepassado dos chineses e bem respeitado por sua inteligência e caridade. Aos 16 anos de idade, Yao foi eleito como líder da tribo. Segundo registros históricos, Yao fundou seu país em Pingyang, como capital (atual cidade de Linfen, na Província de Shanxi ao norte da China). Até hoje pode-se encontrar nesta cidade o Templo de Yao, que foi construído durante a Dinastia Jun (265 a.C. - 420 d.C.) e o Túmulo de Yao construído na Dinastia Tang (618 d.C. - 907 d.C.)." (OS IMPERADORES Yao e Yun, 2006).



"[...] Filho e neto de guerreiros, Ramsés II assumiu o poder com 25 anos, em 1290 a.C., e desde o início de seu reinado o jovem general lançou-se em um esforço militar inédito. O Egito já havia sido o maior império do mundo cerca de 200 anos antes e, sob a batuta de Tutmosés III (a quem seu avô, Ramsés I, servira como general), havia controlado a Palestina e a Mesopotâmia."

Mas, agora, essas regiões haviam se rebelado, algumas estavam sob domínio hitita e as fronteiras do império ameaçavam ruir. Em sua primeira campanha militar, com apenas 10 anos e ao lado do pai, Sethi I, participou da retomada do litoral do Líbano. “A expansão atribuída a Ramsés começou com Sethi, que saneou a economia, abriu novas minas de ouro e criou as condições para que o filho recuperasse o terreno perdido”, diz a historiadora francesa Bernadette Menu, autora de Ramsés II, o Soberano dos Soberanos [...]” (ARANHA, 2006).



tabelas tornaram-se mais complexas, surgiram as representações gráficas e o cálculo de probabilidades. Desde essa época, a Estatística deixou de ser a simples catalogação de dados numéricos coletivos e se tornou o estudo de como chegar a conclusões sobre o *tudo*, partindo da observação e análise de *partes* desse tudo.⁵ Essa é sua maior riqueza.

Para tanto, seu ponto de partida são os **dados**, os quais são expressões numéricas de observações que se fazem de elementos com, pelo menos, uma característica comum.⁶ Por isso,

A Estatística é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões (CRESPO, 1995, p. 13).

De um lado, a Estatística, basicamente, *coleta, organiza e descreve* os dados e, de outro, *analisa e interpreta* esses dados.⁷ Veja a Figura 1, abaixo:

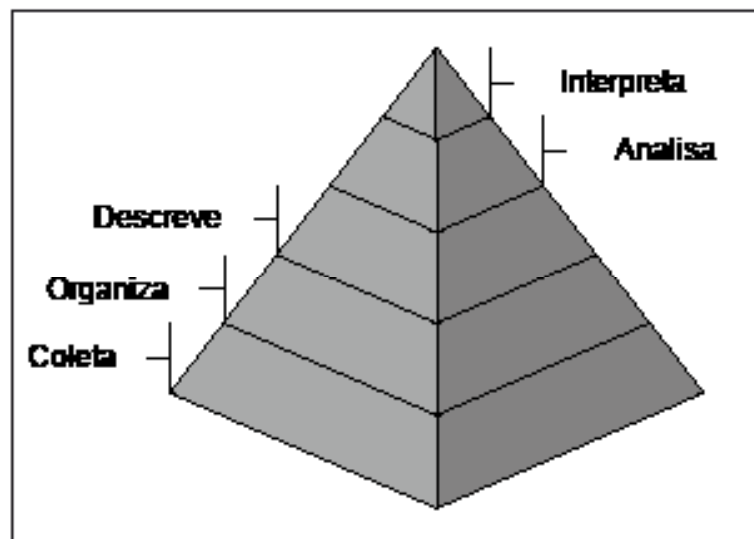


Figura 1: Estatística: Pirâmide da definição

A “Pirâmide da definição” da Estatística nos revela que no topo, isto é, o mais importante é interpretar. Normalmente,

⁵ CRESPO (1995, p. 11).

⁶ CRESPO (1995, p. 13).

⁷ Ver Seção 2: Estatística Descritiva e Estatística Indutiva, p. 42.

as pessoas limitam o termo Estatística à organização e descrição dos dados, desconhecendo, portanto, o que ela oferece de mais importante: “[...] **o aspecto essencial da Estatística é o de proporcionar métodos inferenciais, que permitam conclusões que transcendam os dados obtidos inicialmente.**” (CRESPO, 1995, p. 13, grifo do autor).

É por meio da análise e interpretação dos *dados estatísticos* que é possível o conhecimento de uma realidade, de seus problemas, bem como, a formulação de soluções apropriadas por meio de um planejamento objetivo da ação⁸, para além dos “achismos” e “casuismos” comuns.

Parece evidente, a partir da “Pirâmide”, acima, que as etapas da Estatística devem obedecer às fases da base para o topo, ou seja:

1) Coleta de Dados.

Após a definição do problema a ser estudado e o estabelecimento do planejamento do trabalho (forma de coleta dos dados, cronograma das atividades, custos envolvidos, levantamento das informações disponíveis, delineamento da amostra etc.), o passo seguinte é o da coleta de dados, que consiste na *busca ou compilação dos dados das variáveis*, componentes do fenômeno a ser estudado⁹.

A coleta de dados poderá ser realizada de maneira *direta* ou *indireta*. A coleta será *direta* quando os dados forem obtidos de fonte primária, isto é, sobre elementos informativos de registro obrigatório, como, por exemplo, elementos pertinentes aos prontuários dos alunos de uma escola. A coleta será *indireta* quando é proveniente de elementos já conhecidos (coleta direta)¹⁰.

2) Crítica dos dados.

À procura de falhas e imperfeições, os dados devem ser cuidadosamente criticados, a fim de não incorrermos em erros grosseiros que possam influenciar nos resultados.¹¹

3) Apuração dos dados.

Criticados os dados, agora, eles devem ser processados, isto é, mediante algum critério de classificação, eles serão objeto de operações matemáticas.

8 CRESPO (1995, p. 13).

9 CLEMENTE (2003, p. 4).

10 CRESPO (1995, p. 14).

11 CRESPO (1995, p. 14).



Conheça mais sobre a história da estatística no Brasil no site: <http://www.redeabe.org.br/>

4) Exposição ou apresentação dos dados.

Os dados devem ser apresentados sob a forma de tabelas ou gráficos, a fim de tornar mais fácil o exame daquilo que está sendo estudado.

5) Análise dos resultados.

Todas as fases anteriores se limitam à descrição. A análise dos resultados obtidos tem por base a *indução* ou a *inferência* com o intuito de tirarmos *conclusões* e fazermos *previsões*. Desse modo, buscamos atingir o fim último da Estatística, qual seja: tirar conclusões sobre o *todo* a partir de informações fornecidas por *parte* representativa do todo.¹²

Diante de tudo isso, podemos afirmar que



A Estatística está interessada nos métodos científicos para coleta, organização, resumo, apresentação e análise de dados bem como na obtenção de conclusões válidas e na tomada de decisões razoáveis baseadas em tais análises. (SPIEGEL, 1975, p. 1, grifo nosso).



Resulta claro que a Estatística é uma valiosa ferramenta nas tentativas humanas de interpretação da realidade. Privilegiadamente útil para o exame de fenômenos de massa, teria a Estatística utilização na educação?

Bem, naturalmente, a Estatística como qualquer outra ciência, eu suponho, aplica-se à educação, na medida em que lidamos com grandes quantidades. A despeito do que possa ser considerado grande quantidade, não restam dúvidas quanto à sua fértil aplicação no campo educacional, como ferramenta para a formulação de planos, programas e projetos nos sistemas de ensino, bem como, no interior da própria escola.

Vamos supor que você, amigo Trabalhador da Educação, esteja desconfiado que os alunos estejam chegando muito

¹² CRESPO (1995, p. 15).

atrasados para o início das aulas. Estar desconfiado é um importante início, mas ainda é insuficiente para a tomada de alguma decisão que reverta esse quadro. Por isso, com os recursos da Estatística, você poderia, por exemplo, *coletar dados* sobre o comportamento de toda a escola, com um simples questionário, perguntando aos alunos (ou melhor, a uma *parcela* da escola¹³) sobre quantas vezes eles chegaram atrasados no último mês: a) de 0 a 2; b) de 3 a 5; c) mais de 6.

Observe que a partir desses dados, você pode analisar se essa desconfiança condiz com a realidade e que medidas, caso necessário, devem ser tomadas. Esse é um pequeno exemplo das infinitas possibilidades que a Estatística nos possibilita.

Nesse sentido, recorrer aos ensinamentos da Estatística implica, necessariamente, em melhorar a qualidade dos nossos serviços.

Talvez, o uso constante da matemática assuste alguns de nós. Eu compreendo que a matemática tem sido considerada uma ciência que promove a exclusão social, em virtude de sua ainda rígida forma de trabalho nos bancos escolares. No entanto, ainda assim, não posso concordar que, de maneira definitiva, ela sentencie a população à completa ignorância, como se só a alguns fosse permitida sua apropriação.

Pensando nisso, esforcei-me para que esse Módulo tornasse a Estatística (e a matemática) acessível a todos, explicando fundamentos, apresentando fórmulas e metodologias apropriadas para as resoluções, tudo isso porque, o que nos interessa são análises consistentes que levem à melhoria de nossas ações.

Nosso estudo inicia na **Unidade II: Conceitos Matemáticos** com uma breve retomada daqueles conceitos matemáticos que diretamente condicionam o aprendizado da Estatística. Assim, na *seção 1*, estudaremos um pouco as razões e as proporções; na *seção 2*, estudaremos medidas e grandezas, com enfoque na chamada regra de três simples; depois, na *seção 3*, retomaremos o conceito de porcentagem; na *seção 4*, veremos uma aplicação direta do conceito de porcentagem em coeficientes, taxas e índices; com a *seção 5*, retomaremos o importante sistema de coordenadas cartesianas e encerraremos, na *seção 6*, com uma técnica de arredondamento de números.

13 Ver Unidade 3: Variáveis, Tabelas e Gráficos, Seção 1: População e Amostra, p. 40.

Depois, na **Unidade III: Variáveis, Tabelas e Gráficos** estudaremos na *seção 1*, população e amostra; na *seção 2*, examinaremos mais detidamente os conceitos de Estatística Indutiva e Estatística Dedutiva; na *seção 3*, aprenderemos sobre variáveis; nas *seções 4 e 5*, veremos como apresentar de maneira prática nossos dados por meio de tabelas e gráficos, respectivamente.

Na **Unidade IV: Distribuição de Freqüência** estudaremos a organização dos dados. Primeiro, na *seção 1*, identificaremos dados brutos e dados organizados (rol); depois, na *seção 2*, veremos uma especificidade da organização dos dados – a chamada distribuição de freqüência; a seguir, na *seção 3*, propomos um exercício completo envolvendo os conteúdos da Unidade de estudo; por fim, na *seção 4*, apenas para conhecimento, apresentaremos alguns tipos de curvas possíveis, muito utilizadas em apresentações de dados organizados com essa natureza específica – distribuição de freqüência.

Na nossa última etapa de estudo, **Unidade V: Medidas de Resumo** exploraremos com maior aproximação os recursos da Estatística, por meio da *seção 1*, introdução, onde apontaremos algumas ressalvas desse estudo; depois, na *seção 2*, trabalharemos, de fato, com médias e medidas chamadas de tendência central (média aritmética, mediana e moda); a seguir, na *seção 3*, trabalharemos com medidas de outra natureza chamadas de medidas de dispersão (desvio padrão e coeficiente de variação), mas igualmente úteis para a tomada de decisões; por último, na *seção 4*, estudaremos as chamadas medidas de posição (quartis, decis e percentis).

Lembro, ainda, que, ao longo dos nossos estudos, existem, aqui e ali, algumas atividades propostas para você exercitar um pouco (Pratique!) e, no final do Módulo, você encontrará as respostas dessas atividades.

Desejo a todas e a todos um bom estudo!

2

Conceitos matemáticos



Veja mais sobre frações no site da Wikipedia: <http://pt.wikipedia.org/wiki/Fra%C3%A7%C3%A3o>



Uma divisão nada mais é do que uma simplificação de frações. Observe que $10 \div 5$ é o mesmo que $\frac{10}{5}$.
Essa divisão é fácil: $\frac{10}{5} = 2$

Antes de adentrarmos ao mundo da Estatística, alguns conceitos são convenientes resgatar da matemática. Nosso objetivo será o de tão somente lembrá-los, por isso, não nos deteremos muito tempo neles. A idéia é que como para o estudo da Estatística eles são pressupostos, ou seja, sem eles é impossível compreender a proposta da Estatística, pode ser útil retomá-los, sem exagerarmos a dose. Nesse sentido, retomaremos os conceitos de razão e proporção; a seguir, grandezas e medidas; depois, porcentagem; e ainda, coeficientes, taxas e índices; enfim, sistema de coordenadas cartesianas.

Boa leitura!

Seção 1: Razões e Proporções

Chamamos de **razão** a uma maneira de *comparar* quantidades. Por exemplo, se um determinado conjunto **A** possui 10 elementos e, outro conjunto **B** possui 5 elementos, podemos comparar esses conjuntos. Veja Figura 2, abaixo:

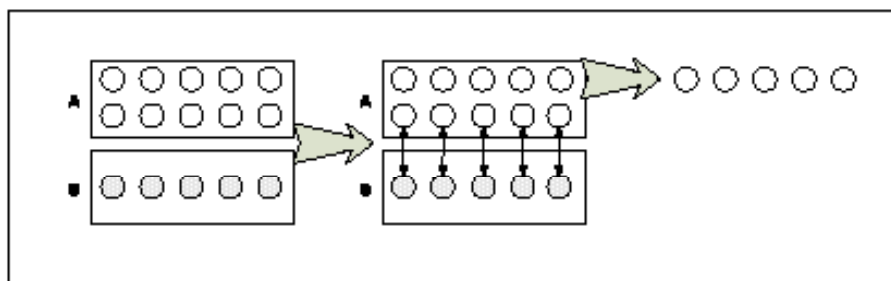


Figura 2: Razão: Comparação

Você reparou que para cada elemento do conjunto **B** existe um elemento do conjunto **A**? Reparou, ainda, que sobraram 5 elementos do conjunto **A**? Pois bem, a comparação dos conjuntos **A** e **B**, da Figura 2, acima, indica que:

$$\frac{10}{5} = 10 \div 5 = 2$$

Dizemos que a comparação dos 10 elementos do conjunto **A** com os 5 elementos do conjunto **B** é a **razão de 10 para 5**. De outra forma, para os 5 elementos de **B** existem 5 elementos mais 5 elementos de **A**, existem, portanto, 2 vezes elementos em **A** comparados a **B**.

Vejamos outro exemplo: Suponha que você possua R\$ 2,00 e eu R\$ 8,00. Qual a razão do que você possui para o que eu possuo?

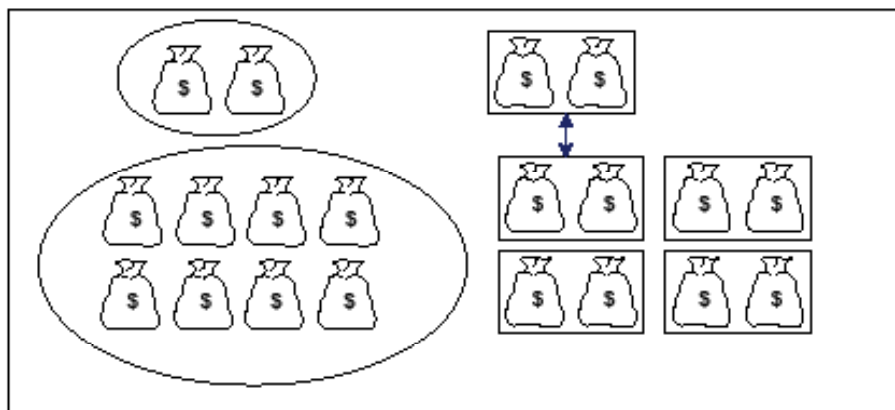


Figura 3: Razão: Exercício

Observe que se você possui R\$ 2,00 e eu possuo R\$ 8,00, dizemos que eu possuo 4 vezes aquilo que você possui ou

$$\frac{2}{8} = \frac{1}{4}$$



Desse modo, dizemos que **2 está para 8** ou **1 está para 4**. A Figura 4, abaixo, talvez ajude a compreender que $\frac{2}{8}$ representa a mesma porção que $\frac{1}{4}$. Quando isso ocorre, dizemos que as razões são **semelhantes**.

Sempre que temos razões semelhantes, é preferível usar a mais simples, a qual, em matemática, chama-se razão irredutível.

25

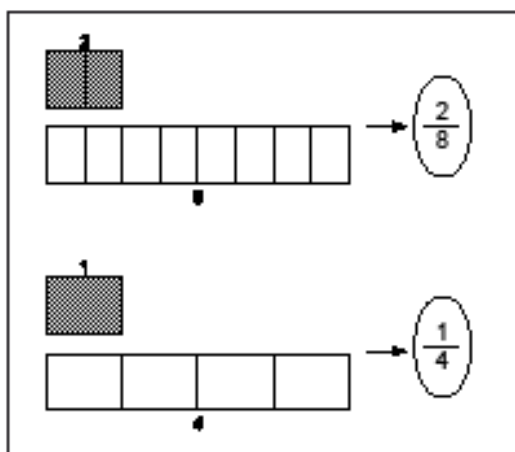


Figura 4: Razão: Representação

Proporções, por sua vez, são também comparações. Mas são comparações entre duas razões. Veja Figura 5, abaixo:

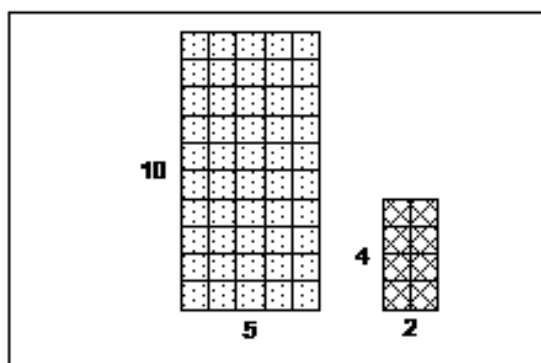


Figura 5: Proporções: Conceito

Observe que na Figura 5, acima, temos dois desenhos. O primeiro desenho é *proporcional* ao segundo. Por quê? Vamos representar o primeiro desenho por meio de uma razão: $5 \div 10 = \frac{5}{10} = \frac{1}{2}$, ou seja, **1 está para 2**. O segundo desenho pode ser representado como $2 \div 4 = \frac{2}{4} = \frac{1}{2}$, isto é, **1 está para 2**. Você notou? Quando duas razões são iguais, estamos diante de uma proporção:

$$\frac{5}{10} = \frac{2}{4},$$

dizemos que: **5 está para 10 assim como 2 está para 4**.

Um bom uso das *razões* e *proporções* é com mapas, plantas e maquetes. Veja a planta de um bairro de uma cidade, abaixo:

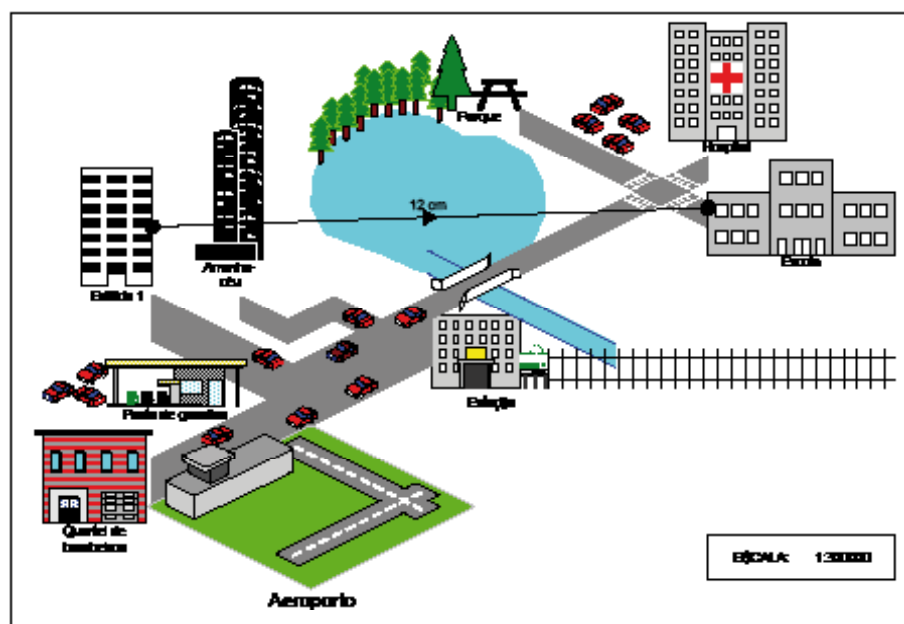


Figura 6: Razões: Proporções: Escala

A Figura 6 anterior apresenta o mapa de um bairro em escala. Isso significa que a escala do mapa indica a razão entre as *distâncias representadas* e as *distâncias reais*. Isto é, a escala 1:300000 indica que cada *cm* no desenho corresponde a 300.000 *cm* reais. Veja:

$$Escala = \frac{\text{distância no desenho}}{\text{distância real}}$$

Assim, supondo que você vá em linha reta do “Edifício 1” até a “Escola” e a distância no desenho é de 12 cm, qual a distância real? Fácil:

Solução:

$$\frac{1}{300.000} = \frac{12}{x} \Rightarrow x = 12 \times 300.000 = 3.600.000$$

$$x = 3.600.000 \text{ cm}$$

$$x = 36 \text{ km}$$

Logo, a distância real é de 36 Km.

Verifique quais figuras, abaixo são proporcionais, sabendo que as medidas estão em milímetros (mm).

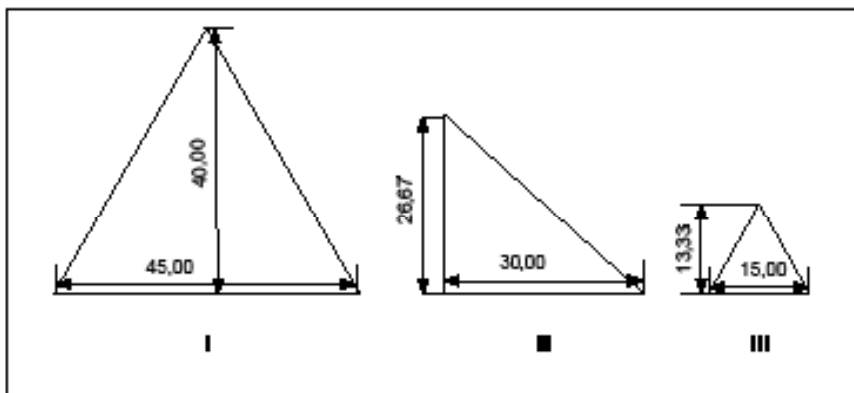


Figura 7: Razões e Proporções: Exercício

Seção 2: Grandezas e Medidas

O professor Dante¹⁴ inicia sua aula sobre **grandezas** e **medidas** fazendo algumas perguntas, como por exemplo:



- Qual é a sua altura?
- Qual será a temperatura máxima hoje?
- Qual é a sua massa?
- Quanto tempo dura seu trabalho?

O professor mostra que para responder a essas perguntas é preciso usar **medidas**. Para isso, precisamos usar *instrumentos*, bem como reconhecer as **grandezas**. Veja:



“Não se esqueça: em uma medida, deve sempre aparecer o número acompanhado da unidade de medida usada: 5 palmos, 10 cm etc.” (DANTE, 2003, p. 112).

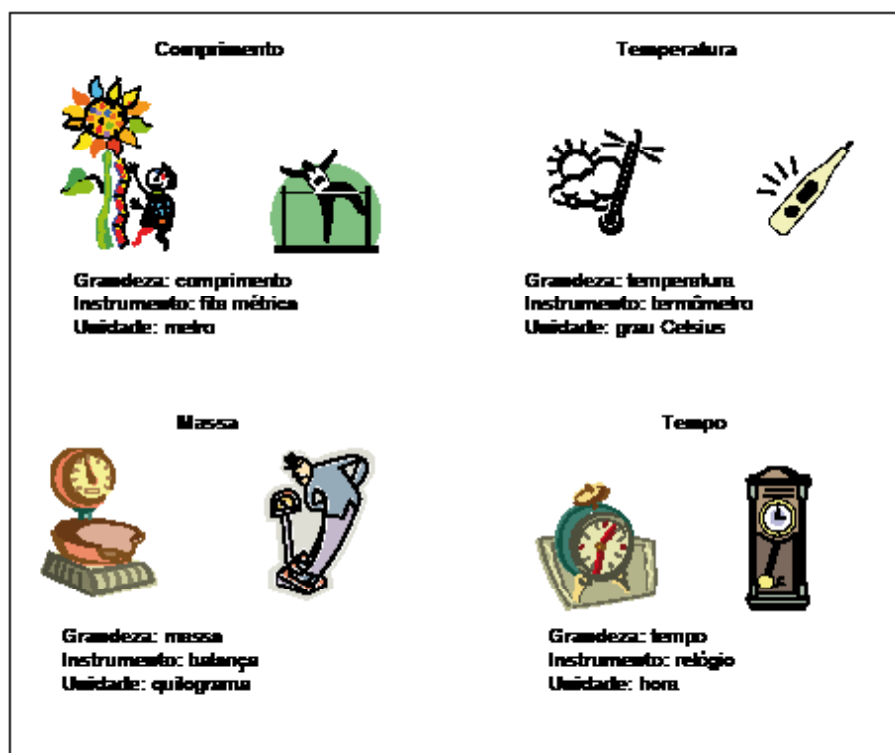


Figura 8: Grandezas



“Em Matemática, entende-se por grandeza tudo que é suscetível a aumento ou diminuição. Assim, podemos falar em grandezas como: tempo, velocidade, peso, número de pessoas, número de objetos etc.” (PARENTE; CARIBÉ, 1996, p. 44).

Medir é comparar **grandezas** de mesmo tipo. Professores de matemática adoram dizer: “– não se pode somar laranjas com limões!”. Eles têm razão: só podemos operar com grandezas iguais. Isso quer dizer que não posso somar 2 horas com 2 Km, pois, as grandezas são diferentes (no primeiro caso, a grandeza é *tempo*; no segundo, *comprimento*).

¹⁴ DANTE (2003, p. 111).

Quando eu tomo a medida do comprimento de uma mesa, por exemplo, eu digo: a mesa possui 1 metro de comprimento. Isso quer dizer que eu comparei a unidade *metro* com o comprimento da *mesa*. Observe a Figura 9, abaixo:



Figura 9: Medida de Comprimento: Segmento de reta

O segmento de reta AB mede 5 cm; podemos dizer que o segmento AB é igual a 5 unidades de medida cm; ou ainda, $\overline{AB} = 5 \text{ cm}$. Quando se mede uma **grandeza** sempre se compara com um padrão de referência estabelecido. Por exemplo, “dizer que uma corda tem 30 metros de comprimento é dizer que ela é 30 vezes maior do que um objeto cujo comprimento foi definido como sendo um metro”.¹⁵

Duas grandezas são ditas diretamente proporcionais quando o aumento do valor de uma leva ao aumento do valor da outra e são inversamente proporcionais quando, ao contrário, o aumento de uma leva à diminuição de outra. Para resolvermos problemas envolvendo grandezas direta ou inversamente proporcionais, recorreremos à regra de três.



Regra de Três Simples

Quando colocamos gasolina em um automóvel, o preço que pagamos é **diretamente proporcional** ao volume de gasolina colocado. Observe que se o preço do litro de gasolina custa R\$ 2,59, é possível saber quanto custará para encher um tanque de 55 litros. Veja:

Litros de gasolina	Preço (R\$)
1	2,59
55	x



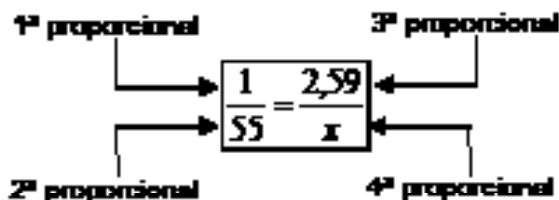
Conheça mais sobre regra de três simples no site:
<http://www.somatematica.com.br/fundam/regra3s.php>

¹⁵ SEARS; ZEMANSKY; YOUNG (1985, p. 3).

Note que conhecemos três números e queremos conhecer um número: x . Esse quarto número é conhecido como *quarta proporcional* e, para encontrá-lo, utilizamos o procedimento conhecido como **regra de três**.

Solucionando nosso problema, temos que:

Solução:



Utilizando a regra de três:

$$x = 55 \times 2,59$$

$$x = 142,45$$

Na prática, multiplicamos em X, isto é, 1ª proporcional vezes 4ª proporcional é igual a 2ª proporcional vezes 3ª proporcional.



Então, para encher um tanque de 55 litros, gastarei R\$ 142,45.

Você notou que a regra de três nada mais é do que uma proporção?

Para o caso de grandezas **inversamente proporcionais**, é preciso tomar um pequeno cuidado na hora de montar a proporção. O restante é igual ao caso anterior. Um problema clássico desse tipo é o dos pedreiros construindo um muro: 3 pedreiros trabalhando constroem um muro em 10 dias. Em quantos dias 6 pedreiros construiriam o mesmo muro trabalhando no mesmo ritmo? Vamos responder:

Número de pedreiros	Tempo (em dias)
3	10
6	x

Observe que utilizamos duas setas: uma para o número de pedreiros e outra para o tempo. A seta para cima indica que o número de pedreiros aumentou (de 3 para 6); a seta para baixo indica que o tempo diminuiu (de 10 para x). Veja que mesmo eu não sabendo, ainda, quanto tempo será, eu posso garantir que o tempo será menor do que 10 dias, se com 3 pedreiros eu preciso de 10 dias, com mais pedreiros eu precisarei de menos de 10 dias, não é mesmo? Quando as setas estão orientadas para sentidos diferentes, estamos diante de grandezas *inversamente proporcionais*. Na prática, isso mudará nossa proporção:



É preciso estar sempre atento às grandezas: se são diretamente ou inversamente proporcionais.

Solução:

$$\frac{3}{6} = \frac{x}{10}$$

Note que a segunda razão foi invertida.

Então,

$$6x = 3 \times 10$$

$$x = \frac{30}{6}$$

$$x = 5$$

Aumentando o número de pedreiros de 3 para 6, o muro seria construído em 5 dias.

Sabendo que a altura da mulher é de 1,60m, quanto mede seu cachorro?

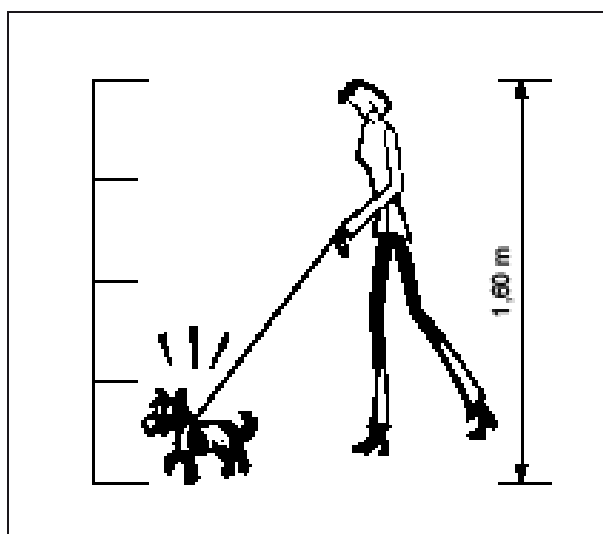


Figura 10: Regra de Três: Exercício

Seção 3: Porcentagem



Porcentagem é uma razão com o denominador sempre igual a 100.

Desse modo, $\frac{25}{100}$, por exemplo, é uma *porcentagem* e pode ser expressa como 25% (*vinete e cinco por cento*).

Na prática, calculamos as porcentagens em diversas situações. Suponha que meu salário seja de R\$ 400,00 e eu receberei um aumento de 12%. Quanto passarei a receber?

Solução:

$$12\% \text{ de } 400 = \frac{12 \times 400}{100} = 48$$

Passarei a receber, portanto, R\$ 400,00 + R\$ 48,00 = R\$ 448,00.

Sempre vemos nos supermercados o uso das porcentagens. Por exemplo: um produto de R\$32,00 está com desconto de 7%. Por quanto ele está sendo vendido?

Solução:

$$7\% \text{ de } 32 = \frac{7 \times 32}{100} = 2,24. \text{ Então,}$$

$$32,00 - 2,24 = 29,76$$

Logo, o produto está sendo vendido a R\$ 29,76.

Vamos realizar um outro tipo de exercício muito comum, com o uso de porcentagens. A Tabela 1, abaixo, apresenta a população total brasileira, por sexo. Pergunta-se: qual a porcentagem de mulheres na população total brasileira?

Tabela 1: População: Brasil

População residente, por sexo			
Grupos por idade	Total	Homens	Mulheres
Total	169 872 856	83 602 317	86 270 539

Fonte: IBGE, Censo 2000

Para responder a essa pergunta, tenho que ter clareza de que a população total brasileira corresponde a 100%. Assim,

$$100\% = 169.872.856$$

O que quero descobrir é qual a porcentagem desse total que corresponde a 86.270.539. Veja:

Porcentagem	População
100	169.872.856
x	86.270.539

Para resolver o problema, usaremos o conceito de proporções, assim:

$$\frac{100}{x} = \frac{169.872.856}{86.270.539} \Rightarrow 169.872.856x = 100 \times 86.270.539$$

$$x = \frac{8.627.053.900}{169.872.856} = 50,78\%$$

Assim, no Brasil, a população de mulheres corresponde a 50,78% da população total.

Sabendo que a população total brasileira é de 169.872.856 e que a população brasileira em idade escolar é de 30.502.425, pergunta-se: qual o percentual de brasileiros em idade escolar? Em outras palavras, quantos por cento da população total brasileira está em idade escolar? Registre a atividade em seu memorial.*

*Fonte: IBGE, Censo Demográfico 2000



33

Seção 4: Coeficientes, taxas e índices

Coeficiente, outro importante conceito matemático que queremos resgatar, também é o resultado de uma *divisão* de uma quantidade por outra. Por exemplo, se numa escola com 400 alunos, 80 ficaram reprovados, então, o *coeficiente de reprovação* foi de 0,2, porque

$$\text{número de reprovados} \div \text{número de alunos} = 0,2.$$



“Os coeficientes são razões entre o número de ocorrências e o número total (número de ocorrências e número de não-ocorrências).” (CRESPO, 1995, p. 34).



"As taxas são os coeficientes multiplicados por uma potência de 10 (10, 100, 1.000 etc.) para tornar o resultado mais inteligível." (CRESPO, 1995, p. 35).

Para facilitar os cálculos, é comum transformarmos o coeficiente em **taxa**. Para isso, basta multiplicarmos o coeficiente por 10, 100, 1000 ou qualquer outra **potência de 10**. Normalmente, usamos 100. Observe:

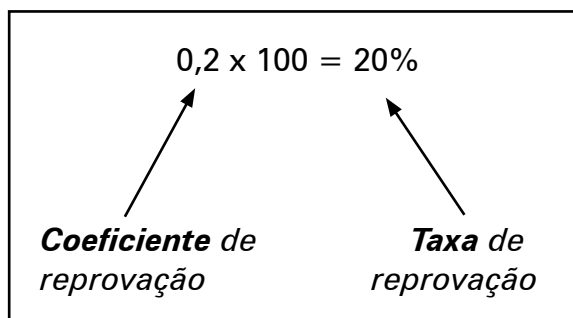


Figura 11: Coeficiente e Taxa

Nosso *coeficiente de reprovação* (**0,2**) multiplicado por 100 é igual à *taxa* de 20%, pois, $0,2 \times 100 = 20\%$. Mas o que isso significa? Significa que de cada 100 alunos, 20 ficaram reprovados.



Observe como é fácil comprovar isso. Vamos agrupar os 400 alunos em grupos de 100. Assim, teríamos 4 grupos de 100 alunos. Cada grupo possui 20 reprovados. Logo, 20 vezes 4 é igual a 80 alunos reprovados. Bem, isso mostra que nosso coeficiente de reprovação (20%) está correto.



"Os índices são razões entre duas grandezas tais que uma não inclua a outra." (CRESPO, 1995, p. 34).

Como se vê *coeficiente* e *taxa* são conceitos muito parecidos. A única diferença é a multiplicação do *coeficiente* pela *potência de 10* que dará a *taxa*.

O conceito de **índice**, por sua vez, não é muito diferente, senão por uma única razão: dividimos grandezas diferentes. Observe que no nosso exemplo, o *coeficiente de reprovação* é 0,2 e a *taxa de reprovação* é de 20%; nos dois exemplos estamos tratando do número de *alunos*. Assim,

$$\text{Coeficiente de reprovação} = \text{n}^\circ \text{ de alunos reprovados} \div \text{n}^\circ \text{ total de alunos}$$

Mas suponha que queiramos saber a relação entre o número de *alunos* reprovados e o número de alunos reprovados em *matemática*. Nesse caso, estamos diante de duas grandezas diferentes. Assim, essa comparação de grandezas diferentes chama-se **índice** (por exemplo, *índice de reprovados por disciplina*).

Vamos realizar um exercício. Veja a Tabela 2, abaixo:

Tabela 2: Aprovação: Ensino Fundamental: Brasil: 2005

Unidade da Federação	Alunos aprovados no Ensino Fundamental				
	Total				
	Total	Federal	Estadual	Municipal	Privada
Brasil	26.368.619	23.172	9.752.502	13.434.669	3.158.276

Fonte: Censo Escolar 2005

Essa Tabela apresenta o total de alunos aprovados no ensino fundamental brasileiro, por dependência administrativa. Vamos calcular *coeficiente* e *taxa* utilizando essa Tabela.

Primeiro: qual é o coeficiente de aprovação no ensino fundamental dos alunos que freqüentam escolas da rede municipal?



Para responder a essa pergunta faremos a seguinte divisão:

$$\text{coeficiente de aprovação da rede municipal} = \frac{\text{total de aprovados na rede municipal}}{\text{total de aprovados no Brasil}}$$

Assim,

$$\text{coeficiente de aprovação da rede municipal} = \frac{13.434.669}{26.368.619} = 0,5$$

Isso tem algum significado muito importante para a educação? Pouco provável, a não ser pelo fato de que o *coeficiente* de 0,5 (que representa uma *taxa* de $0,5 \times 100 = 50\%$) corresponde a dizer que de cada 100 alunos aprovados no país, 50 são da rede municipal.

Veja que trabalhamos com *coeficiente* e *taxa* no exemplo acima. Agora, para trabalharmos com *índice*, precisaremos comparar grandezas diferentes. Relembrando, se você ainda tiver dúvidas sobre grandezas, retome a Seção 2: Grandezas e Medidas, desta Unidade.

Vamos supor que queiramos estabelecer o *índice de densidade professor-aluno aprovado no ensino fundamental na rede municipal de ensino*. Precisaremos, portanto, da Tabela 3, abaixo.

Tabela 3: Função Docente: Educação Básica: Brasil: 2005

Unidade da Federação	Funções Docentes Exercendo Atividades em Sala de Aula				
	Total	Federal	Estadual	Municipal	Privada
Brasil	2.589.688	14.980	940.039	1.110.132	524.537

Fonte: Censo Escolar 2005

Nesse caso, estamos diante de duas grandezas diferentes: professores e alunos. Assim,

$$\text{índice de densidade professor – aluno da rede municipal} = \frac{1.110.132}{13.434.669} = 0,08$$



Observe que um *índice* também pode ser transformado em *taxa*.

Isso representa uma *taxa* de $0,08 \times 100 = 8\%$; ou seja, para cada 100 alunos aprovados na rede municipal, há 8 professores.



Calcule o coeficiente de aprovação no Ensino Fundamental da rede privada, da zona rural brasileira utilizando a Tabela 4, abaixo. Depois, transforme esse coeficiente em taxa.

Registre os resultados em seu memorial.

Tabela 4: Aprovação: Ensino Fundamental: Rural: Brasil: 2005

Unidade da Federação	Alunos aprovados no Ensino Fundamental				
	Rural				
	Total	Federal	Estadual	Municipal	Privada
Brasil	4.085.448	499	499.117	3.553.931	31.901

Fonte: Censo Escolar 2005

Seção 5: Sistema de Coordenadas Cartesianas

Os professores Jakubo e Lellis (1995) contam uma história bastante interessante sobre o famoso filósofo e matemático francês **René Descartes**:



Famoso por ter proferido a frase “penso, logo existo”, Descartes (1596-1658) escreveu o Discurso do Método, em 1637, que irá marcar profundamente a realização da ciência no mundo. O nome cartesiano vem do nome do seu autor, Descartes.



Figura 12: Sistema de Coordenadas Cartesianas: Origem

“Dizem que ele estava descansando na cama, quando viu uma mosca pousada na parede. A mosca voou, mas Descartes ficou pensando. Como poderia explicar a uma outra pessoa qual era a posição exata da mosca na parede?” (JAKUBOVIC; LELLIS, 1995, p. 210).

Esse teria sido o início do **sistema de coordenadas cartesianas**. Descartes imaginou duas retas: uma horizontal e outra vertical. Se ele marcasse números nessas retas, ficaria fácil localizar a mosca. Veja Figura 13, abaixo:

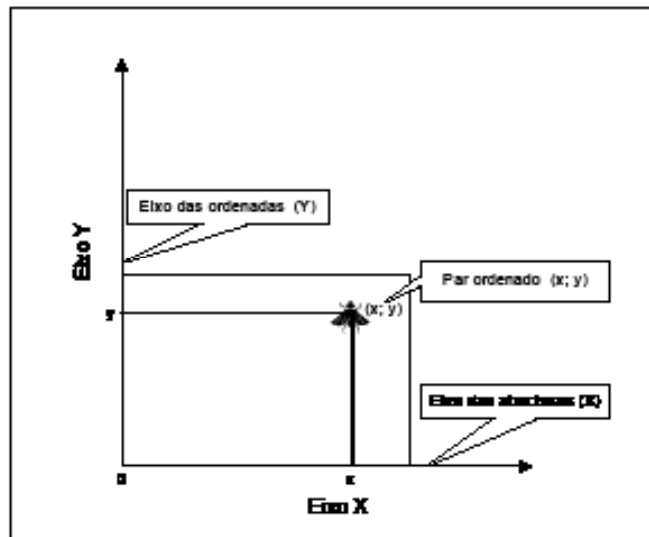


Figura 13: Sistema de Coordenadas Cartesianas: Eixos

Dessa forma, para localizar um ponto em um plano, usamos:¹⁶

- As retas numeradas **x** e **y** chamam-se **eixos cartesianos**: o eixo **x** é horizontal, o eixo **y** é vertical;
- O plano com esses eixos chama-se **plano cartesiano**;
- Os pares ordenados são as **coordenadas cartesianas** do ponto;
- O ponto correspondente à **origem** é o par ordenado (0; 0).

Veja a Figura 14, abaixo:

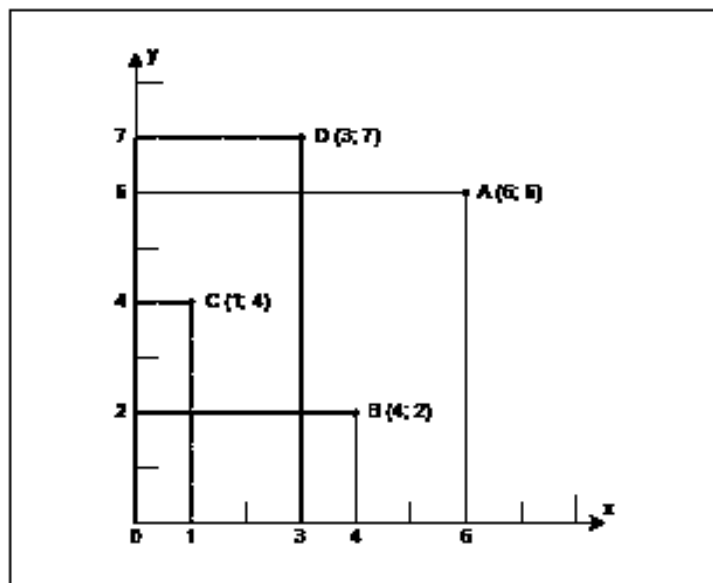


Figura 14: Sistema de Coordenadas Cartesianas: Pontos

¹⁶ JAKUBOVIC; LELLIS (1995, p. 211).

De maneira mais completa, podemos localizar qualquer ponto no plano: o ponto A se encontra em (6; 6), isto é, x é 6 e y vale 6; o ponto B (4; 2); e assim por diante. Viu? Na prática, usamos o sistema de coordenadas cartesianas em diversas situações diferentes quando queremos localizar um ponto em um plano. Veja a Figura 15, abaixo:

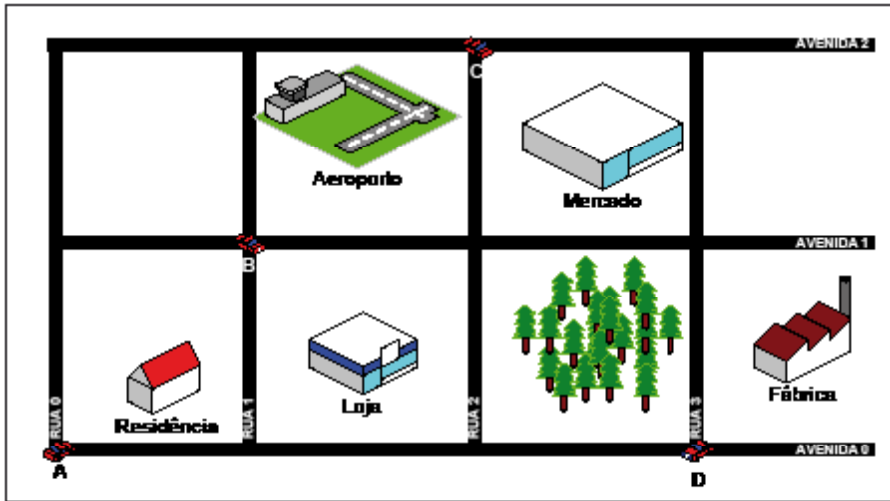


Figura 15: Sistema de Coordenadas Cartesianas: Exercício

Como localizar o carro **B**, por exemplo? Claro! O carro **B** está na Rua 1 com a Avenida 1, ou seja, **B** (Rua 1; Avenida 1). O carro **A** está na origem de nosso sistema; as Ruas indicam o primeiro número do par ordenado (x) e as Avenidas o segundo número (y). Desse modo, **A** (Rua 0; Avenida 0); o carro **C** está na Rua 2, Avenida 3, isto é, **C** (Rua 2; Avenida 3). Pronto!

Na Figura 15, acima, identifique todos os cruzamentos que não possuem carros.



Seção 6: Arredondamento

Com essa Seção 6 encerramos nossa Unidade II.

Entendemos por arredondamento de dados a técnica utilizada para suprimir unidades inferiores, isto é, arredondar um número significa reduzir a quantidade de algarismos após a vírgula.



Um número apresenta uma parte inteira e uma parte fracionária. Veja:



Na matemática, muitas vezes, deparamo-nos com situações onde o cálculo nunca dá certo se não transformarmos esse número em fração.

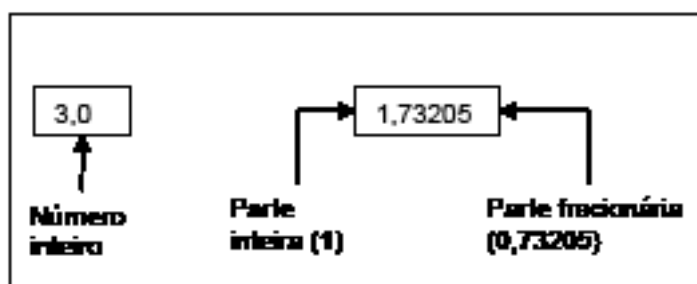


Figura 16: Arredondamento de Números

Às vezes, queremos trabalhar com números com, digamos, uma casa decimal, mas o que fazer quando o resultado encontrado for um número com muito mais casas depois da vírgula? A rigor, na Estatística, precisamos seguir um critério rígido de arredondamento a fim de não comprometermos os resultados.

Por exemplo, suponha que queiramos trabalhar com duas casas decimais e nosso resultado foi 1,1417. Como fazer?



Conforme a Resolução nº 886/66 do IBGE, o arredondamento é realizado da seguinte maneira:

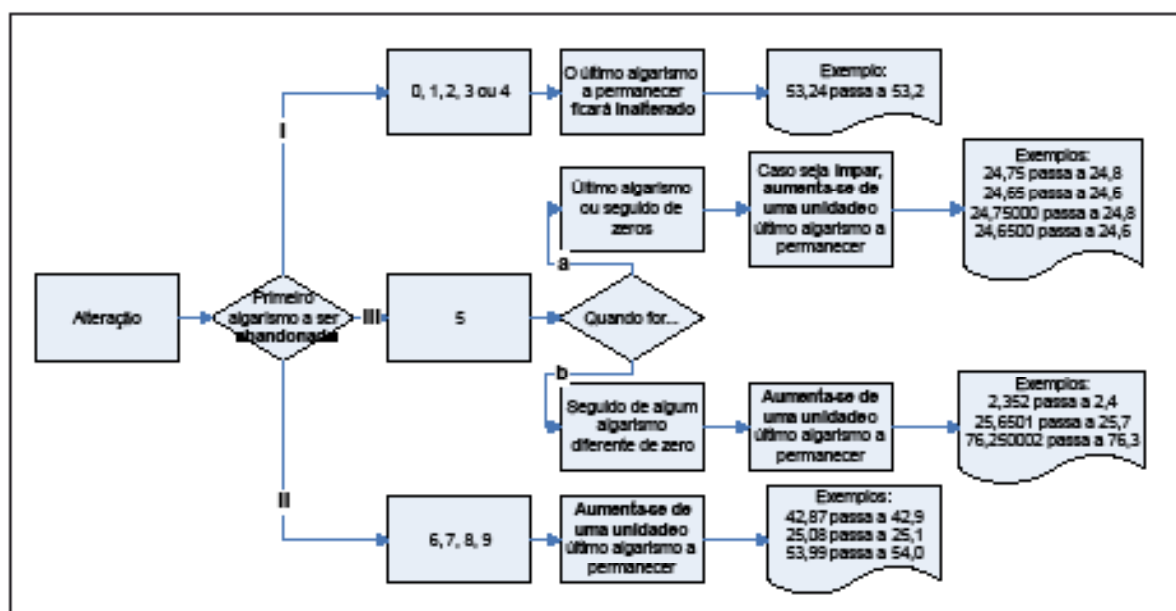


Figura 17: Arredondamento: Fluxograma
Fonte: Adaptado de: CRESPO (1995, p. 174)

Caso haja necessidade de alteração, nossa atenção deve recair sobre o primeiro algarismo a ser abandonado. Teremos três caminhos possíveis:

- 1) Seguimos o primeiro caminho (I) quando o primeiro algarismo a ser abandonado for 0, 1, 2, 3 ou 4. Nesse caso, o algarismo a permanecer ficará **sem alteração**. Por exemplo, 4,84 passa a 4,8;
- 2) Seguimos o segundo caminho (II) quando o primeiro algarismo a ser abandonado for 6, 7, 8 ou 9. Nesse caso, o último algarismo a permanecer será **aumentado de um**. Por exemplo, 4,87 passa a 4,9;
- 3) Quando o primeiro algarismo a ser abandonado for 5, seguimos o III caminho. Nesse caso, temos que prestar muita atenção, pois, o caminho se divide em dois percursos:
 - a) Quando o número a ser abandonado for 5 e ele for o último ou seguido de zeros, **aumentaremos uma unidade apenas quando o último algarismo a permanecer for ímpar**. Por exemplo: 5,85 passa a 5,8;
 - b) Quando o número a ser abandonado for 5 seguido de algum número diferente de zero, **aumenta-se uma unidade** ao algarismo a permanecer. Por exemplo, 8,55000000002 passa a 8,6.



Observe que o último algarismo a permanecer é 8 (par). Nesse caso, não sofrerá alteração.



Observe que o último algarismo a permanecer é 5 e o primeiro a ser abandonado também é 5. O último algarismo a permanecer (5) foi aumentado de 1 porque havia, após o algarismo a ser abandonado (5) um algarismo diferente de zero.

41

Casos de arredondamento não são difíceis, mas requerem muita prática até compreendermos bem os processos. Não há outra alternativa.

Ressalto que, em nosso Módulo, simplesmente abandonamos a parte fracionária sem todo esse rigor. Por isso, esteja à vontade para fazer correções às respostas, caso você julgue pertinente.



1) Arredonde cada um dos dados abaixo, deixando-os com apenas uma casa decimal (CRESPO, 1995, p. 174):

$$2,38 =$$

$$4,24 =$$

$$6,829 =$$

$$24,65 =$$

$$328,35 =$$

$$5,550 =$$

$$0,351 =$$

$$2,97 =$$

$$89,99 =$$

2) Arredonde cada um dos valores abaixo para o centésimo mais próximo (CRESPO, 1995, p. 174):

$$46,727 =$$

$$253,65 =$$

$$28,255 =$$

$$123,842 =$$

$$299,951 =$$

$$37,485 =$$

3

Variáveis, tabelas e gráficos



44 Para que as conclusões sejam válidas é preciso observar alguns critérios; quem estuda esses critérios é a estatística indutiva ou inferência estatística. Dizemos inferência quando queremos nos referir a uma conclusão sobre uma população a partir do exame da amostra dessa população.

Nessa Unidade III, nosso objetivo é estudar algumas maneiras de organização e exposição dos dados de um fenômeno sob estudo. Para isso, é preciso compreender o significado de população e amostra (*seção 1*); a seguir, na *seção 2*, retomaremos a distinção já iniciada nesse estudo, entre a Estatística voltada para a descrição (Estatística Descritiva) e a voltada para interpretação (Estatística Indutiva ou Inferencial); na *seção 3*, aprenderemos sobre como trabalhar com os fenômenos a partir de sua representação numérica conseguida com a aplicação do conceito de variável; depois, na *seção 4*, iremos formalizar a exposição dos dados em uma Tabela, como forte recurso visual da Estatística; para, enfim, na *seção 5*, reconhecermos os gráficos como poderosas ferramentas para rápida e eficiente compreensão do comportamento da(s) variável(eis) em estudo.

Boa leitura!

Seção 1: População e Amostra

Ao examinar um grupo qualquer, considerando *todos* os seus elementos, estamos tratando da **população** ou **universo**. Nem sempre isso é possível. Nesse caso, examinamos uma pequena parte chamada **amostra**.

Uma população pode ser *finita* (isto é, possuir fim) ou *infinita* (não possuir fim). Por exemplo, a população dos alunos de sua escola é finita e a população constituída de todos os resultados (cara ou coroa) em sucessivos lances de uma moeda é infinita.

Se uma *amostra* é representativa de uma população, podemos obter conclusões importantes sobre a população. Mas também, podemos analisar e descrever um certo grupo sem tirar conclusões ou inferências sobre um grupo maior, nesse caso, a parte da Estatística que se preocupa com isso é a chamada estatística descritiva ou estatística dedutiva.

Vamos realizar um exercício. Observe a Tabela 5, abaixo.

Tabela 5: População Escolar: Sexo

Escolas	Nº de Estudantes	
	Masculino	Feminino
A	80	95
B	102	120
C	110	92
D	134	228
E	150	130
F	300	290

Fonte: Adaptado de CRESPO (1995, p. 24).

Essa Tabela se refere à população escolar, por sexo e por escola, de uma determinada localização. Um exercício interessante é retirar uma amostra, digamos, de 10% da população. Bem, para isso, precisaremos considerar escola por escola.



Muitas vezes, a população se divide em subpopulações chamadas estratos. A amostragem proporcional estratificada considera os estratos para a amostra, de maneira análoga à Tabela 6, ao lado.

Tabela 6: Cálculo da amostragem proporcional estratificada

Escolas	População	10%	Amostra
A	M = 80	$\frac{10 \times 80}{100} = 8$	8
	F = 95	$\frac{10 \times 95}{100} = 9,5$	9
B	M = 102	$\frac{10 \times 102}{100} = 10,2$	10
	F = 120	$\frac{10 \times 120}{100} = 12$	12
C	M = 110	$\frac{10 \times 110}{100} = 11$	11
	F = 92	$\frac{10 \times 92}{100} = 9,2$	9
D			
E			
F			

Procedendo assim, temos que na escola A, devemos considerar 8 alunos e 9 alunas; na escola B, 10 alunos e 12 alunas; na escola C, 11 alunos e 9 alunas.

Complete a Tabela 6, acima, e registre o resultado em seu memorial.



Seção 2: Estatística Descritiva e Estatística Indutiva ou Inferencial

Como já afirmamos, a Estatística interessa-se pelo tratamento de fenômenos por meio de métodos científicos capazes de auxiliar a tomada de decisões.



O principal objetivo da Estatística é tirar conclusões sobre o todo (população), a partir de informações fornecidas por parte representativa do todo (amostra).

O primeiro passo consiste em coletar, criticar, apurar e expor os dados.¹⁷ Essas são etapas da **Estatística Descritiva**. Observe que cumpridas essas etapas, ainda não é possível tirar conclusões muito seguras, mas é possível, por exemplo, conhecer a realidade da escola, bem como conhecer seus problemas.

O passo seguinte consiste na **Estatística Indutiva** ou **Inferencial**. Basicamente, nessa etapa, ocorre a análise e a interpretação do fenômeno em estudo, com o intuito de tirar conclusões e fazer previsões.¹⁸ Agora, é possível formular soluções consistentes sobre os problemas levantados de uma dada realidade.

A Estatística, portanto, começa com a **descrição** para, só depois, chegar a **conclusões**. Veja:

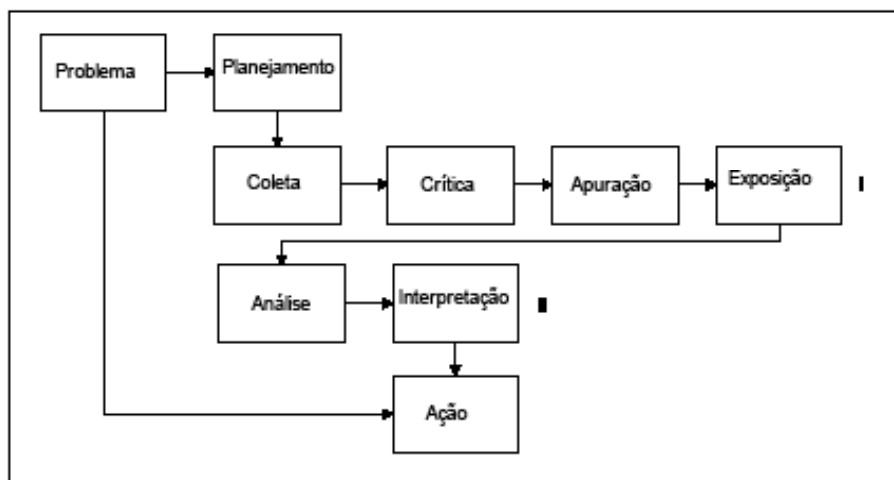


Figura 18: Estatística Dedutiva e Estatística Indutiva: Fluxograma

¹⁷ Ver Unidade 1: Introdução ao Estudo da Estatística, p. 11.

¹⁸ CRESPO (1995, p. 15).

A Figura acima revela que o ponto de partida é um **problema**. Seria muito bom se pudéssemos pegar o “atalho” e do “problema” fôssemos, imediatamente, para a “ação”. Embora alguns gestores (do setor público e do setor privado) ajam assim, isso não é muito seguro. O interessante é observar as duas etapas (**I e II**), a fim de garantir um mínimo de segurança de que estamos no caminho correto para a solução do problema evidenciado.

Dessa maneira, uma vez identificado onde se deseja atuar, o passo seguinte é o do **planejamento** (Que recursos possuo? Que métodos de coleta de dados irei utilizar? Que tempo posuo? Qual o universo? Qual a amostra? etc.). Feitas as escolhas, entramos na **Etapa I**: Estatística Descritiva.

Nessa etapa I, todos os passos devem ser observados: coleta, crítica, apuração e exposição dos dados. Só depois disso, estamos preparados para a **Etapa II**: Estatística Indutiva ou Inferencial. Nessa etapa da solução do problema, podemos tirar conclusões e fazer algumas previsões com maiores chances de acertar do que se pegássemos o “atalho”.

A propósito, essa é talvez a maior contribuição da Estatística para nossas atividades no ambiente de trabalho: apresentar-se como uma poderosa ferramenta para a solução de problemas.

Seção 3: Variáveis

Se consideramos o fenômeno “sexo”, haveria, pois, dois resultados possíveis: masculino ou feminino. O fenômeno “total de filhos” também possui um número determinado: 0, 1, 2, 3... Mas o fenômeno “estatura” apresenta uma situação diferente: 1m64cm, 1m58cm, 1m75cm...

Chamamos de **variável** o conjunto de resultados possíveis de um fenômeno¹⁹. A variável pode ser *qualitativa* (masculino-feminino) ou *quantitativa* (expressa por números: salários, idade etc.).

A *variável quantitativa* pode ser **contínua** ou **discreta**. Por exemplo, o número de crianças de uma família pode ser 0, 1, 2, 3... Mas, jamais, pode ser 2,5 ou 3,842. Chamamos essa variável de *discreta*. Já a altura de um indivíduo pode ser 1,65m,

¹⁹ CRESPO (1995, p. 17).

1,662m ou 1,6722m, conforme a precisão da medida, e é uma variável contínua.²⁰ Assim,



Uma variável quantitativa que pode assumir, teoricamente, qualquer valor entre dois limites recebe o nome de variável contínua; uma variável que só pode assumir valores pertencentes a um conjunto enumerável recebe o nome de variável discreta.²¹

Veja:

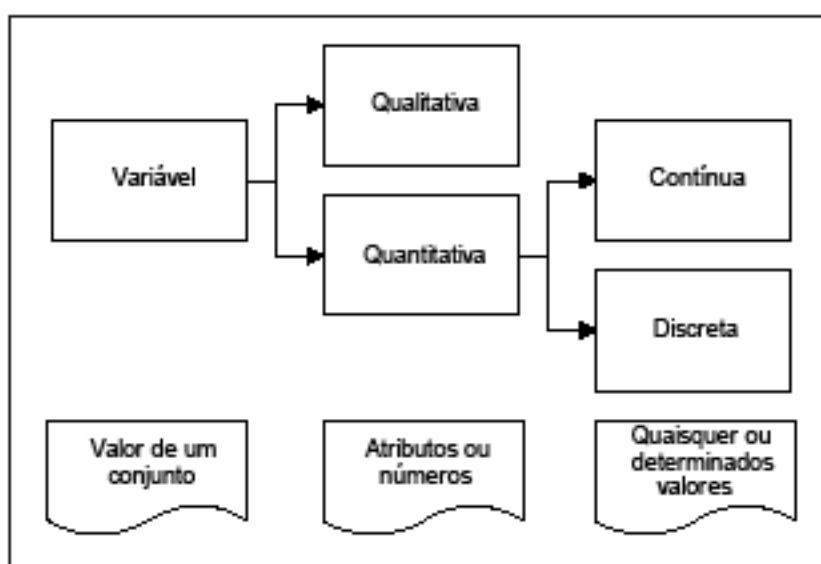


Figura 19: Variáveis: Definições

Explicando melhor, a Figura acima mostra que **variável** corresponde aos resultados possíveis de um conjunto. Será **variável qualitativa**, quando seus valores forem expressos por atributos (qualidades), como, por exemplo, sexo, cor da pele etc. e será **variável quantitativa** quando seus valores forem expressos por números. Nesse último caso, variável quantitativa, poderá ser **discreta**, quando assumir, apenas, um dos valores do conjunto como, por exemplo, o número de alunos de uma escola. Será uma variável quantitativa **contínua**, quando puder assumir qualquer valor entre dois limites, por exemplo, peso, estatura etc.²²

²⁰ SPIEGEL (1975, p. 2).

²¹ CRESPO (1995); SPIEGEL (1975).

²² CRESPO (1995).

De modo geral, as **medições** dão origem a variáveis quantitativas contínuas e as **contagens** ou **numerações**, a variáveis discretas.²³ Além disso, é comum designar as letras x , y e z para representar as variáveis. Por exemplo:

“Sejam 2, 3, 5 e 8 todos os resultados possíveis de um dado fenômeno. Fazendo uso da letra x para indicar a variável relativa ao fenômeno considerado, temos: $x \in \{2, 3, 5, 8\}$ ”.²⁴ Isso significa que x pertence ao conjunto.

Vamos realizar um exercício? Complete o Quadro 2, abaixo, classificando as variáveis em qualitativas ou quantitativas (contínuas ou discretas).

Universo	Variável
Alunos de uma escola.	Cor dos cabelos – <i>Variável qualitativa.</i>
Casais residentes em uma cidade.	Número de filhos – <i>Variável quantitativa discreta.</i>
As jogadas de um dado.	O ponto obtido em cada jogada –
Peças produzidas por certa máquina.	Número de peças produzidas por hora –
Peças produzidas por certa máquina.	Diâmetro externo –

Quadro 2: Tipos de variáveis
Fonte: Adaptado de CRESPO (1995, p. 18).

Classifique as variáveis abaixo em (1) variável qualitativa, (2) variável quantitativa discreta e (3) variável quantitativa contínua, relacionando as duas colunas



23 CRESPO (1995, p. 18).
24 CRESPO (1995, p. 18).



As tabelas apresentam informações tratadas estatisticamente, conforme IBGE (1993) (BRASIL, 2002).

Coluna 1	Coluna 2
() População: alunos de uma cidade Variável: cor dos olhos	(1) variável qualitativa
() P: estação meteorológica de uma cidade V: precipitação pluviométrica durante um ano	(2) variável quantitativa discreta
() P: Bolsa de Valores de São Paulo V: número de ações negociadas	(3) variável quantitativa contínua
() P: funcionários de uma empresa V: salários	
() P: pregos produzidos por uma máquina V: comprimento	
() P: casais residentes em uma cidade V: sexo dos filhos	
() P: propriedades agrícolas V: produção de algodão	
() P: segmentos de reta V: comprimento	
() P: bibliotecas da cidade de São Paulo V: número de volumes	
() P: aparelhos produzidos em uma linha de montagem V: número de defeitos por unidade	
() P: indústrias de uma cidade V: índice de liquidez	

Fonte: Adaptado de CRESPO (1995, p. 18-19).

Seção 4: Tabelas

Uma das preocupações da estatística, como já vimos, é analisar dados, para isso, é preciso compreender o comportamento deles. E isto, a estatística consegue apresentando valores em **tabelas** e **gráficos**, que irão fornecer informações rápidas e seguras a respeito das variáveis em estudo.

Até aqui, em nosso estudo, lidamos com tabelas e quadros, qual a diferença? Quadros apresentam informações não numéricas, isto é, informações que não são objeto de tratamento numérico. Diferentemente, as tabelas são numéricas e servem para cálculos.

As tabelas são muito úteis para a construção de **séries estatísticas**. Denominamos série estatística toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou da espécie (CRESPO, 1995, p. 26).

Por exemplo:

1 → **Tabela 7: População Mundial:**

Série Histórica

5

6

2

Ano	População
2002	6.229.629.168
2003	6.303.112.453
2004	6.376.863.118
2005	6.451.058.790
2006	6.525.486.603

3

4

Fonte: U.S. CENSUS (2006)

A Tabela 7, acima, apresenta:

- 1) **Título:** Conjunto de informações, o mais completo possível. Responde a perguntas como: o quê? Quando? Onde? No nosso exemplo: **Tabela 7: População Mundial: Série Histórica.**
- 2) **Cabeçalho:** Parte superior da tabela que especifica o conteúdo das linhas. No nosso exemplo: **Ano e População.**
- 3) **Linhas:** Retas imaginárias que facilitam a leitura, no sentido horizontal, de dados que se inscrevem nos seus cruzamentos com as colunas. Por exemplo, no ano de 2002 havia 6.229.629.168 de habitantes no planeta.
- 4) **Casa ou célula:** Espaço destinado a um só número. Por exemplo, 6.525.486.603 é um número que ocupa uma casa ou célula.
- 5) **Coluna indicadora:** Parte da tabela que especifica o conteúdo das linhas. No nosso exemplo, a coluna indicadora é a do **Ano** (2002 a 2006).
- 6) **Coluna numérica:** Parte da tabela que contém os dados apresentados. Em nosso exemplo, a coluna numérica é a da **População.**

Agora que conhecemos a constituição de uma tabela simples, vamos estudar uma série estatística. Observe a Tabela 8, abaixo:

Tabela 8: Matrículas no Ensino Fundamental de 5ª a 8ª série:
Diurno: Brasil

Unidade da Federação	Matrículas no Ensino Fundamental de 5ª a 8ª série Diurno				
	Total	Federal	Estadual	Municipal	Privada
Brasil	13.629.874	18.183	7.386.348	4.664.840	1.560.503

Fonte: MEC/Inep

O *título* da tabela é “Matrículas no Ensino Fundamental de 5ª a 8ª série: Diurno: Brasil”. Observe que, pelo título, é possível apreender diversas informações, tais como: a tabela se refere a matrículas no Ensino Fundamental de 5ª a 8ª série; na tabela encontraremos dados referentes ao ensino diurno; e se refere ao Brasil como um todo, não a um estado da federação em particular. Mas, apenas pelo título não é possível saber todo o conteúdo (como por exemplo, não sabemos se encontraremos dados do sistema privado de ensino), mas ele já nos informa muito. Agora...



Identifique os demais componentes da Tabela 8: Matrículas no Ensino Fundamental de 5ª a 8ª série: Diurno: Brasil (acima).



Conjugando duas ou mais séries em uma única tabela, obtemos uma tabela de dupla entrada. Em uma tabela desse tipo ficam criadas duas ordens de classificação: uma horizontal (linha) e uma vertical (coluna) (CRESPOL, 1995, p. 28).

Algumas vezes, é necessário apresentar em uma única tabela a variação de valores de mais de uma variável, isto é, fazer a *conjugação* de duas ou mais séries. Tabelas contendo *série geográfica* e *série histórica* são muito comuns no campo da educação. Vamos trabalhar com uma tabela parecida com a anterior. Observe a Tabela 9, abaixo:

Tabela 9: Número de matrículas na pré-escola

Unidade da Federação	Matrículas na Pré-Escola		
	2002	2003	2004
Acre	21.737	21.682	23.148
Alagoas	57.671	57.981	73.741
Distrito Federal	71.985	76.926	81.786
São Paulo	1.276.434	1.325.507	1.391.238

Fonte: MEC/Inep (2006)

Essa é uma típica tabela conjugada de dupla entrada. Observe que ela possui uma série histórica (2002, 2003 e 2004) e uma série geográfica (Acre, Alagoas, Distrito Federal e São Paulo). Podemos dizer que a *horizontal* (linha) e a *vertical* (coluna) formam duas ordens de classificação. Por exemplo, no Distrito Federal (*linha horizontal* – série geográfica), o número total de alunos matriculados na pré-escola variou no período de 2002 a 2004 (*colunas verticais* – série histórica). Sem dúvida, estamos diante de uma tabela conjugada de dupla entrada.



Séries compostas de três ou mais entradas podem existir, mas são raras devido a dificuldade de representação.

Visite o sítio do Inep e procure a Tabela de Matrícula no Ensino Fundamental de 5ª a 8ª série (ou outra Tabela qualquer) do seu município e identifique os componentes dessa tabela. Monte duas tabelas: uma simples e uma de dupla entrada.



Conheça o sítio do INEP : <http://www.inep.gov.br>

Seção 5: Gráficos

Observe a comparação abaixo, sobre a exposição dos mesmos dados por estratégias diferentes: Tabela e Gráfico.

Tabela 10: Nº de matrículas no Ensino Médio: Brasil: Urbano

Unidade da Federação	Matrículas no Ensino Médio				
	Total	Federal	Estadual	Municipal	Privada
Brasil	8.824.397	56.464	7.528.326	149.917	1.089.690

Fonte: Censo Escolar 2005

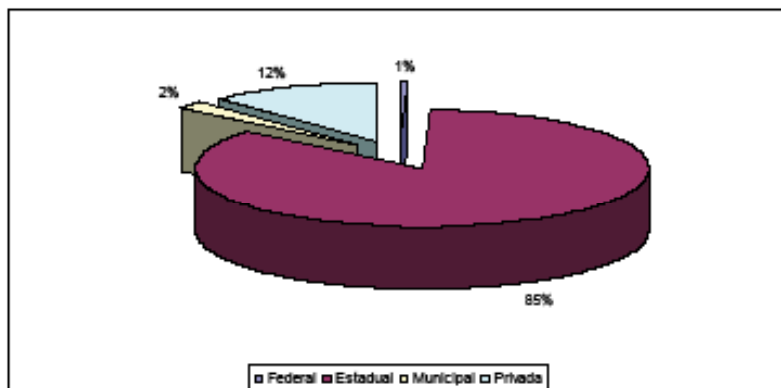


Gráfico 1: Nº de matrículas no Ensino Médio: Brasil: Urbano

Fonte: Censo Escolar 2005



Tanto a Tabela 10, quanto o Gráfico 1, acima, possuem a mesma finalidade: sintetizar os valores que a variável “matrículas no Ensino Médio brasileiro, urbano” pode assumir, para que tenhamos uma visão global da variação dessa variável. Ambos, Tabela e Gráfico, são maneiras válidas de apresentação dos dados de tal forma que podemos, de maneira clara, explorá-los.

Na comparação acima, por exemplo, vemos com mais clareza e mais rapidamente no Gráfico 1 que a maioria dos alunos do Ensino Médio brasileiro encontra-se na rede estadual de ensino. Essa é a finalidade da disposição dos dados quer seja em Tabelas ou em Gráficos: apresentar de maneira simples, com eficiência e rigor, os dados de um conjunto em estudo. Como já vimos muito sobre Tabelas, iremos nos concentrar, agora, em Gráficos.

Por definição:



O gráfico estatístico é uma forma de apresentação dos dados estatísticos, cujo objetivo é o de produzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão que as séries. (CRESPO, 1995, p. 38).

Um Gráfico estabelece uma relação entre os termos de uma série e determinada figura geométrica, como no nosso Gráfico 1, acima, no qual a série estatística (Tabela 10) foi apresentada na forma de gráfico de “pizza”.

Mas atenção: “uma das formas mais eficazes de transmitir uma informação com certo rigor é usando gráficos. No entanto, um gráfico que não seja claro pode confundir o leitor”²⁵. Por isso, a representação gráfica de um fenômeno deverá obedecer a certos critérios fundamentais:²⁶

- 1) Simplicidade;
- 2) Clareza;
- 3) Veracidade (o gráfico deve expressar a verdade sobre o fenômeno).

²⁵ PEREIRA (2004, p. 51)

²⁶ CRESPO (1995, p. 38).

Os principais tipos de gráficos são: **diagramas, cartogramas e pictogramas.**

Diagramas

Os **diagramas**, normalmente, possuem duas dimensões, onde fazemos uso do sistema de coordenadas cartesianas²⁷. Podem ser dos seguintes tipos: **gráfico em linha ou em curva; gráfico em colunas ou em barras; gráfico em colunas ou em barras múltiplas; gráfico em setores.**

Vejamos um exemplo de **gráfico em linha**. Consideremos a seguinte série histórica apresentada na Tabela abaixo:

Tabela 11: Matrículas na Educação Infantil: Brasil

Modalidade	Matrículas na Educação Infantil: Brasil.					
	1999	2000	2001	2002	2003	2004
Creche	831.978	916.864	1.093.347	1.152.511	1.237.558	1.348.237
Pré-Escola	4.235.278	4.421.332	4.818.803	4.977.847	5.155.676	5.555.525

Fonte: MEC/Inep

Vamos construir o gráfico em linha, por exemplo, do número de alunos matriculados na Pré-Escola, no período considerado. Para isso, precisaremos montar o sistema de coordenadas cartesianas. É muito simples, como já vimos, nesse sistema, para cada ano do eixo x , encontraremos uma quantidade de matrículas correspondente y , formando, assim, o par ordenado $(x; y)$. Em 1999, temos 4.235.278 matrículas, formando o par ordenado $(1999; 4.235.278)$; em 2000, o par ordenado será $(2000; 4.421.332)$; e assim sucessivamente. Pronto, a tarefa está realizada! Veja o resultado, abaixo.

²⁷ Ver Unidade 2: Conceitos Matemáticos, Seção 5: Sistema de Coordenadas Cartesianas, p. 33.

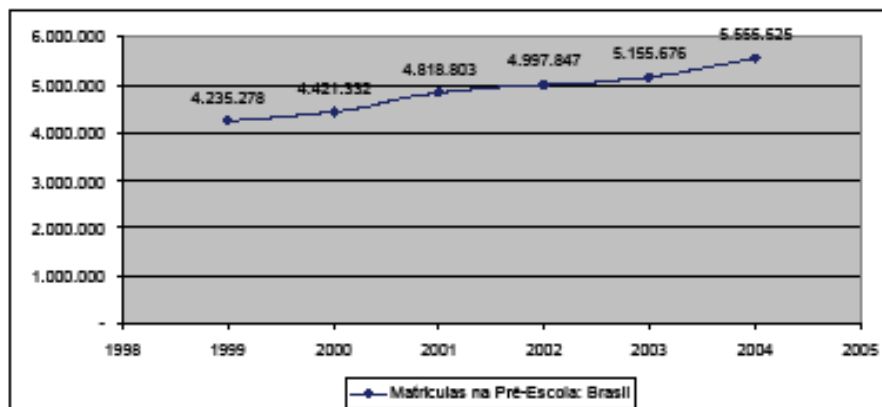


Gráfico 2: Matrículas na Pré-Escola: Brasil: 1999-2004

Fonte: MEC/Inep

Considerando ainda a série estatística representada pela Tabela 11, acima, realizaremos, agora, outra representação gráfica: o **gráfico em barras**. Nesse tipo de gráfico, a representação será em forma de *retângulos*, dispostos *horizontalmente* (em barras). Poderíamos também, dispor a série histórica verticalmente, então, teríamos um gráfico em colunas.

Vamos representar desta vez, a evolução das matrículas na Creche. Dessa vez, o eixo x será representado pelo número de matrículas na Creche e o período está representado no eixo y. Veja como fica o gráfico:

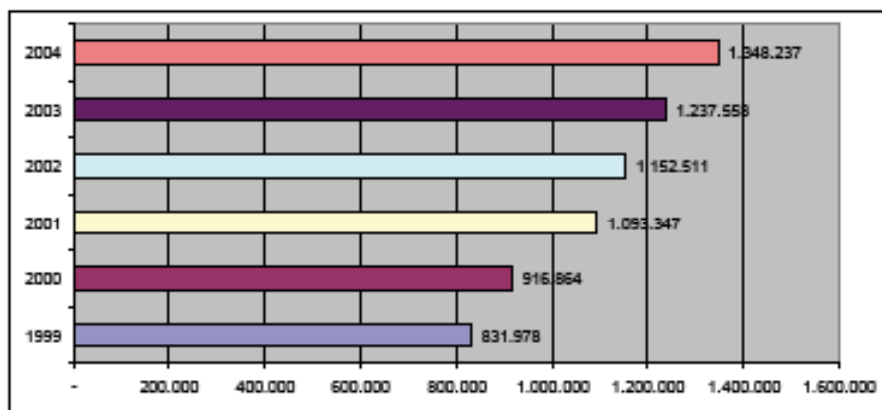


Gráfico 3: Evolução das matrículas na creche: Brasil: 1999-2004

Fonte: MEC/Inep

Vamos juntar as duas informações, a evolução das matrículas na Creche e na Pré-Escola, em um só gráfico? Para isso, iremos considerar, novamente, a série estatística representada pela Tabela 11. Observe o resultado:

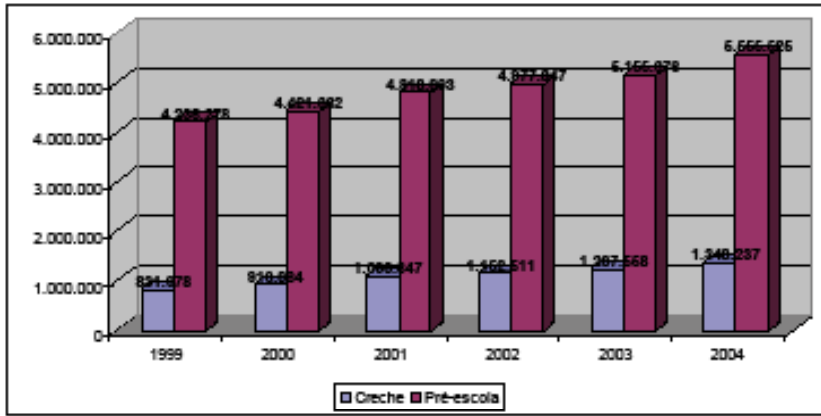


Gráfico 4: Evolução das matrículas na educação infantil: creche e pré-escola: Brasil: 1999-2004

Fonte: MEC/Inep

O Gráfico 4, acima, é um exemplo de **gráfico em colunas** ou **barras múltiplas**. Nele, podemos comparar, rapidamente e com clareza, a evolução das matrículas na educação infantil brasileira, na Creche e na Pré-Escola, ao mesmo tempo.

Como você já notou, as diversas representações gráficas servem para apresentar os dados com rigor metodológico e de maneira clara; seus usos dependem da finalidade da exposição. Às vezes, podemos utilizar diversas representações gráficas, mas, algumas vezes, existem representações ideais para os dados a serem expostos. É assim que, por exemplo, o **gráfico em setores** é empregado sempre que desejamos ressaltar a *participação do dado no total*, dessa maneira, ele serve para mostrar proporções relativas; o total é representado pelo círculo, que fica dividido em tantos setores quantas são as partes.²⁸

Vejamos na prática: considere a seguinte série estatística:

Tabela 12: Usuários de transporte público do estado: 1ª a 4ª séries: Brasil: área urbana

Unidade da Federação	Alunos do Ensino Fundamental de 1ª a 4ª séries, área urbana, que utilizam transporte escolar do poder público estadual e municipal Área urbana				
	Total	Federal	Estadual	Municipal	Privada
Brasil	447.847	324	81.482	363.994	2.047

Fonte: Censo Escolar 2005

28 CRESPO (1995); PEREIRA (2004).

A Tabela 12, acima, apresenta os alunos de 1ª a 4ª séries do ensino fundamental que freqüentam escolas urbanas e fazem uso do transporte público oferecido pelo Poder Público estadual e/ou municipal, de acordo com a dependência administrativa (Federal, Estadual, Municipal e Privada). Para trabalharmos com setores, precisaremos estabelecer as proporções para cada esfera administrativa. Assim,

Solução:

Para encontrar as proporções de cada dependência administrativa, usaremos o procedimento da regra de três simples:²⁹

1) Encontrando a porção da **esfera federal**:

1ª etapa: preparando a regra de três

Alunos	%
447.847	100
324	x

2ª etapa: montando a proporção

$$\frac{447.847}{324} = \frac{100}{x}$$

3ª etapa: resolvendo a equação

$$447.847 \times x = 324 \times 100 \Rightarrow x = \frac{32.400}{447.847} = 0,072\%$$

2) Encontrando a porção da **esfera estadual**:

1ª etapa: preparando a regra de três

Alunos	%
447.847	100
81.482	x

2ª etapa: montando a proporção

$$\frac{447.847}{324} = \frac{100}{x}$$

²⁹ Ver Unidade 2: Conceitos Matemáticos, Seção 2: Grandezas e Medidas, Regra de três simples, p. 25.

3ª etapa: resolvendo a equação

$$447.847 \times x = 81.482 \times 100 \Rightarrow x = \frac{8.148.200}{447.847} = 18,19\%$$

Viu como é fácil? Agora é a sua vez!

Continue o exercício e encontre as porções municipal e privada.



Após encontrar as proporções de cada esfera administrativa (federal, estadual, municipal e privada), basta, agora, construir o *gráfico em setores*. Veja o resultado abaixo:

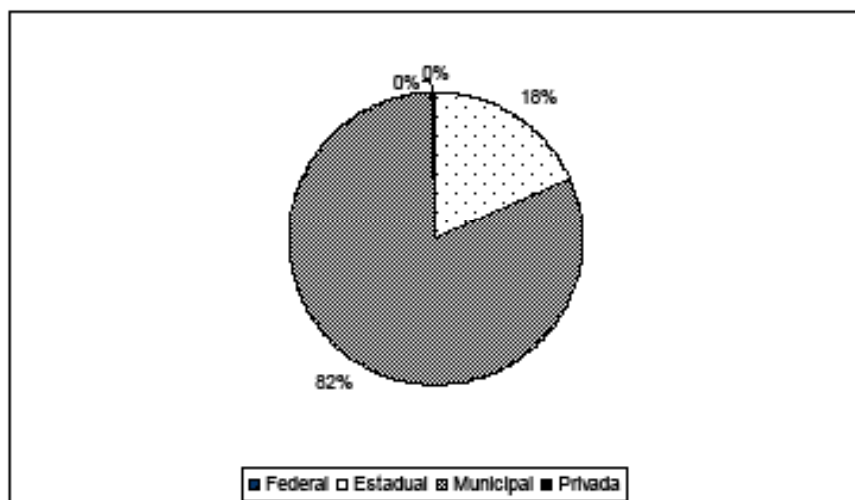


Gráfico 5: Usuários de transporte público do estado: 1ª a 4ª séries: Brasil: área urbana
Fonte: Censo Escolar 2005

Observe como é interessante a comparação das *partes* com o *todo*. No nosso exemplo, o *gráfico em setores* apresenta, com inigualável clareza, que as participações federal e privada são insignificantes (tanto que nem aparecem) e a participação municipal é esmagadora. Convenhamos, essa demonstração é mais interessante que a série estatística na forma de tabela, não é mesmo?

Cartogramas



*Cartogramas são representações sobre uma **carta geográfica**. Eles são muito úteis quando queremos relacionar dados estatísticos com áreas geográficas ou políticas. Essas representações são muito úteis para expressarem população e densidade.³⁰*

Vejamos um exemplo:

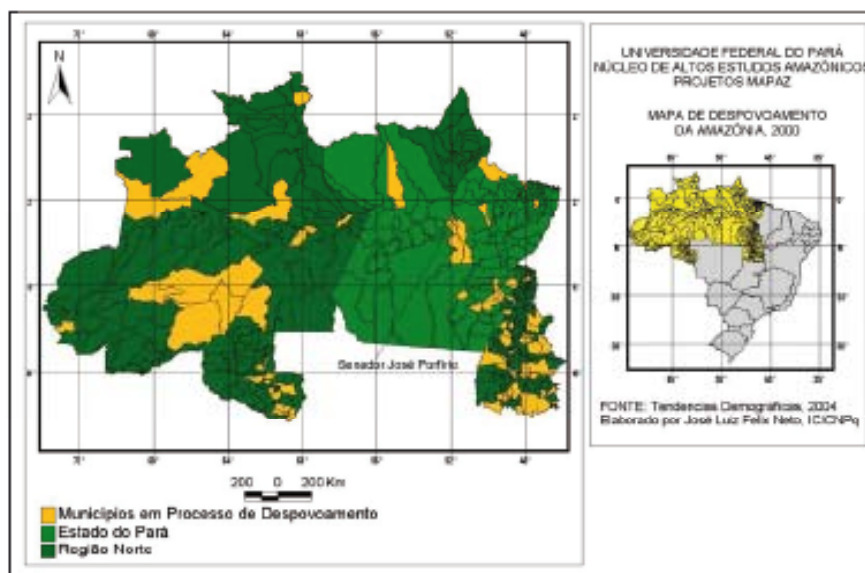


Gráfico 6: O despovoamento da Amazônia

Fonte: FELIX NETO (2006, p. 5).

Observe que o Gráfico 6, acima é uma apresentação agradável aos olhos e de fácil interpretação também. Esse é o objetivo.

Pictogramas



*Os **pictogramas** são os processos gráficos de maior aceitação pública por sua forma atraente e sugestiva.³¹*

Em sua representação encontram-se figuras, desenhos etc. Seja a série estatística abaixo:

³⁰ CRESPO (1995, p. 46).

³¹ CRESPO (1995, p. 48).

Tabela 13: Pictograma: Exercício

Vítimas Fatais							
Local	Idade (anos)						Igno- rado
	0 a 9	10 a 12	13 a 17	18 a 29	30 a 59	60 e mais	
Brasil	808	307	891	5006	6950	1666	3249

Fonte: Adaptado de Anuário Estatístico de Acidentes de Trânsito (2002)

A Tabela acima, revela o número de vítimas fatais em acidentes de trânsito no Brasil, no ano de 2002. Em forma de pictograma, poderia ser assim representada:

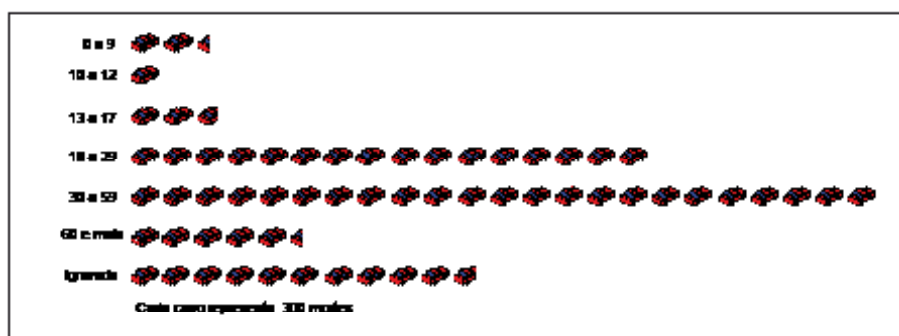


Figura 20: Pictograma: Exemplo

Observe que os carros são representativos para a série estatística de vítimas fatais em acidentes de trânsito. Naturalmente, “na confecção de gráficos pictóricos temos que utilizar muita criatividade, procurando obter uma otimização na união da arte com a técnica” (CRESPO, 1995, p. 49).

Procure, em jornais, revistas, livros e outros, um exemplo de cada representação gráfica estudada, isto é, um gráfico em setores (em forma de “pizza”), um gráfico em linha, um gráfico em barras, um gráfico em colunas múltiplas, um cartograma e, por fim, um pictograma. Recorte ou tire uma cópia (se possível) e cole em seu memorial.



4

Distribuição de frequência

O objetivo desta Unidade é partir dos *dados brutos*, isto é, desorganizados, para uma apresentação formal. Nesse percurso, *seção 1*, destacaremos a diferença entre *tabela primitiva* e *rol*, bem como a importância do resumo dos dados por meio de uma técnica que agrupa as repetições, chamadas de *freqüência* (*seção 2*). Voltaremos às Tabelas e Gráficos, na *seção 3*, porque, agora, aparecerá algo novo: os *dados agrupados*. Em função disso, as Tabelas apresentarão diferenças das anteriores e os Gráficos assumem formatos já consagrados pelo uso (histograma e polígono de frequência).

Boa leitura!

Seção 1: Dados Brutos e Rol

Na Unidade anterior, trabalhamos com exposição de dados. Mas, infelizmente, os dados, raramente, apresentam-se organizados. Por exemplo, vamos supor que um professor entregue as notas de seus alunos, conforme a Tabela 14, abaixo:

Tabela 14: Exemplo de Tabela Primitiva

Notas de 40 alunos de uma disciplina									
8,0	5,0	3,0	3,5	4,0	10,0	5,6	3,0	2,5	1,5
9,5	7,5	6,3	6,6	7,8	4,0	2,5	5,0	7,0	8,0
10,0	9,8	9,7	3,5	3,8	5,0	3,7	4,9	5,4	6,8
6,3	7,8	8,5	6,6	9,9	10,0	2,6	2,9	5,2	8,8

Observe que, nessa Tabela, as notas não estão numericamente organizadas. Esse tipo de tabela denomina-se **Tabela primitiva**.³² Partindo dessa Tabela, é difícil identificar o comportamento das notas, isto é: onde se concentram? Qual a maior? Qual a menor? Quantos alunos estão abaixo ou acima de uma determinada nota?

Esses dados estão, de fato, desorganizados, por isso, vamos organizá-los. A maneira mais simples é realizando uma ordenação (crescente ou decrescente). Após essa ordenação dos dados, a Tabela recebe o nome de **rol**. Veja como fica:

³² CRESPO (1995, p. 54).

Tabela 15: Exemplo de Rol

Notas de 40 alunos de uma disciplina									
1,5	2,9	3,5	4,0	5,0	6,3	6,8	7,8	8,8	9,9
2,5	3,0	3,7	4,9	5,2	6,3	7,0	8,0	9,5	10,0
2,5	3,0	3,8	5,0	5,4	6,6	7,5	8,0	9,7	10,0
2,6	3,5	4,0	5,0	5,6	6,6	7,8	8,5	9,8	10,0

De fato, com os dados assim organizados, podemos saber, com facilidade, qual a menor nota (1,5) e qual a maior (10,0). E também, podemos encontrar a **amplitude** de variação, isto é, a diferença entre o *maior* valor e o *menor* valor: $10,0 - 1,5 = 8,5$. Além dessas informações, com um pequeno esforço, podemos ainda identificar que as notas se concentram em dois valores (5,0 e 10,0) e que 6,0 é o valor que divide as notas. Convém destacar que os **dados** são úteis, apenas, se conseguirmos transformá-los em **informação**. Mais à frente, discutiremos essas medidas.

Enfim,

Dados brutos são aqueles que não foram numericamente organizados e **rol** é um arranjo de dados numéricos brutos em ordem: crescente ou decrescente. Em um rol, a diferença entre o maior e o menor número chama-se *amplitude total*.³³



Seção 2: Distribuição de Frequência

Vamos continuar estudando as notas entregues por um professor apresentada acima. Para estudarmos melhor a variável, construiremos uma Tabela apresentando os valores de maneira mais resumida. Com os dados organizados em um *rol*, identificamos que existem repetições de muitos valores. Essa repetição recebe o nome de **frequência**. Vejamos:

³³ SPIEGEL (1975, p. 43).



“Classes de frequência ou, simplesmente, classes são intervalos de variação da variável.” (CRESPO, 1995, p. 57).



A Tabela de Distribuição de Frequência é uma Tabela como outra qualquer, mas que apresenta o número de repetição dos valores ao invés de repeti-los integralmente. Por exemplo, ao invés de expor 2, 2, 2, 2 e 3, em uma Tabela de Frequência colocamos 2 (4 vezes) e 3.

Tabela 16: Exemplo de Tabela de Frequência

Notas	Frequência	Notas	Frequência	Notas	Frequência
1,5	1	5,0	3	8,0	2
2,5	2	5,2	1	8,5	1
2,6	1	5,4	1	8,8	1
2,9	1	5,6	1	9,5	1
3,0	2	6,3	2	9,7	1
3,5	2	6,6	2	9,8	1
3,7	1	6,8	1	9,9	1
3,8	1	7,0	1	10,0	3
4,0	2	7,5	1		
4,9	1	7,8	2	Total	40

Dispor os dados dessa maneira é melhor do que da forma anterior, mas ainda é inconveniente. Isso porque exige muito espaço. Uma alternativa é agrupar os dados. Para desenvolver tal tarefa, é comum, em primeiro lugar, distribuir os dados em *classes* ou *categorias* em uma Tabela. Essa Tabela receberá o nome de **Distribuição de Frequência** ou **Tabela de Frequência**.

Para construir a *tabela de frequência* das notas, consideraremos, por exemplo, quatro *classes*: da nota 0,0 até a nota 4,9 (0,0–4,9); da nota 5,0 até a nota 6,9 (5,0–6,9); da nota 7,0 até a nota 8,9 (7,0–8,9); por fim, da nota 9,0 até a nota 10,0 (9,0–10,0). Agrupando os dados dessa maneira, é comum chamá-los de **dados agrupados**. Vejamos:

Tabela 17: Exemplo de Tabela de Distribuição de Frequência

Notas de 40 alunos de uma disciplina	
Notas	Número de estudantes (frequência)
0,0 – 4,9	14
5,0 – 6,9	11
7,0 – 8,9	8
9,0 – 10,0	7
Total	40

A *distribuição de frequência*, acima, apresenta uma disposição mais amigável. Nela, podemos observar que 14 alunos tiraram

notas entre 0,0 e 4,9; 11 alunos, entre 5,0 e 6,9; 8 alunos, entre 7,0 e 8,9; 7 alunos, entre 9,0 e 10,0. Identifica-se, de imediato, a maior e a menor concentração das notas dos alunos e essa é uma informação muito interessante.

Aprofundamento: regras para a elaboração de uma distribuição de frequência

Na construção de uma **distribuição de frequência**, a determinação do número de **classes** e da **amplitude** dessas classes é sempre uma preocupação.

No nosso exemplo anterior, as classes escolhidas não foram de maneira aleatória, mas, de qualquer forma, existem regras que podem ser observadas se quisermos maior rigor no estudo de um evento.

Assim, Spiegel (1975, p. 45-46) sugere as seguintes regras gerais:

- 1) Determinam-se o maior e o menor número de dados brutos e, então, calcula-se a amplitude total do rol (diferença entre o maior e o menor daqueles números);
- 2) Divide-se a amplitude total em um número conveniente de intervalos de classe que tenham a mesma amplitude. Nem sempre isso é possível; nesse caso, usamos intervalos de classe de amplitudes diferentes. O número de intervalo de classes é normalmente entre 5 e 20, dependendo dos dados;
- 3) Os intervalos de classe são escolhidos de maneira que seus pontos médios coincidam com dados realmente observados. Isso tende a diminuir erros;
- 4) Determina-se o número de observações que caem dentro de cada intervalo de classe, isto é, calculam-se as frequências de classe.



Seguindo as regras gerais acima, que alterações teríamos no nosso exercício das notas?

Bem, primeiro, vamos calcular a diferença entre o maior e o menor número: $10,0 - 1,5 = 8,5$. Isso significa que entre a

maior nota e a menor nota há uma distância de 8,5. Essa é a amplitude total, isto é, os valores variam, no máximo, 8,5. De outra forma, a distância do menor valor para o maior valor é de 8,5. OK!

Agora, na segunda etapa das regras acima, vamos escolher o número de intervalos de classe.³⁴ Vamos tentar o menor número sugerido: 5. Se quero 5 classes e minha amplitude total é 8,5, basta dividir a amplitude total pelo número de classes escolhido para determinar os intervalos de classe. Assim,

$$\text{Intervalo de classes} = \frac{\text{amplitude total}}{\text{total de classes}} = \frac{8,5}{5} = 1,7 = 2$$

Observe que arredondamos³⁵ o valor para 2 (assim temos um número fácil de trabalhar). O que esse resultado significa? Significa que teremos cinco intervalos de amplitude 2. Desse modo, nossa nova tabela de distribuição de frequência será:

Tabela 18: Exemplo de Tabela de Distribuição de Frequência

Notas de 40 alunos de uma disciplina	
Notas	Número de estudantes (frequência)
0,0 – 2,0	1
2,1 – 4,1	12
4,2 – 6,2	7
6,3 – 8,3	11
8,4 – 10,0	9
Total 40	

Observe que alterando os intervalos de classes, as concentrações mudam.

Gráficos de uma distribuição

Graficamente, uma distribuição de frequência pode ser representada pelo **histograma** ou pelo **polígono de frequência**.

³⁴ Relembrando: no nosso exemplo utilizamos 4 intervalos: 0,0–4,9; 5,0–6,9; 7,0–8,9; 9,0–10,0.

³⁵ Ver Unidade 2: Conceitos Matemáticos, Seção 6: Arredondamento, p. 35.

Ambos os gráficos são representados no sistema cartesiano, sendo o eixo x (linha horizontal) a representação da variável e no eixo y (linha vertical) a representação das freqüências.

Histograma

Vejam os um modelo de **histograma**.

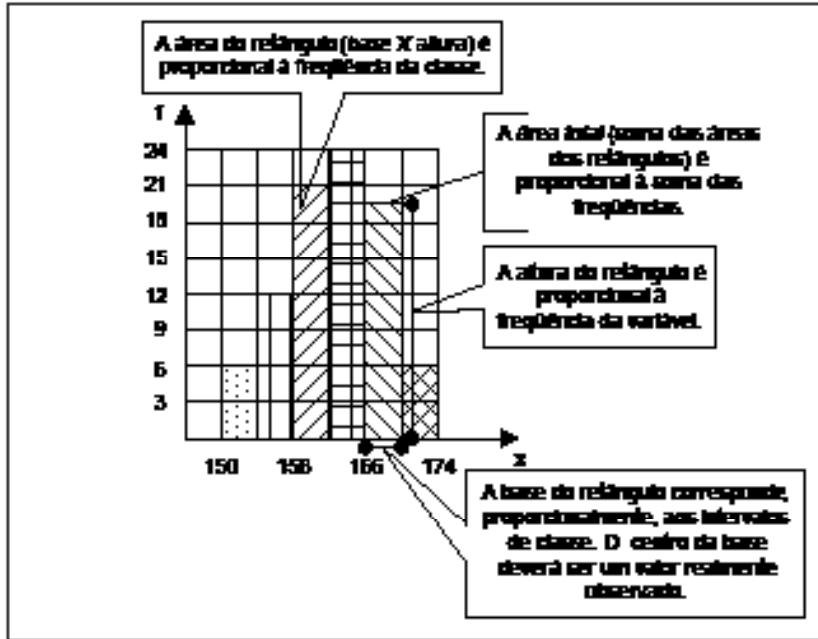


Figura 21: Modelo de Histograma

O modelo de histograma do gráfico da Figura 21, acima, revela que o **histograma** é formado por um conjunto de retângulos justapostos representados no sistema de coordenadas cartesianas, onde, o eixo x é o “eixo das variáveis” e o eixo y , o “eixo das freqüências”.

As bases dos retângulos representam os *intervalos de classe* e o ponto médio delas deverá ser um valor observado no estudo das variáveis. As alturas dos retângulos são proporcionais às freqüências das classes. Calculando a área de um retângulo, encontramos a freqüência daquele intervalo de classe e calculando a área de todos os retângulos, encontramos a soma de todas as freqüências. Formalmente,

O histograma é formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal, de tal modo que seus pontos médios coincidam com os pontos médios dos intervalos de classe (CRESPO, 1995, p. 69).



Polígono de frequência

Polígono de frequência é um gráfico de linha³⁶. Na verdade, essa representação gráfica nada mais é do que a união dos pontos de frequência das variáveis. Observe abaixo:

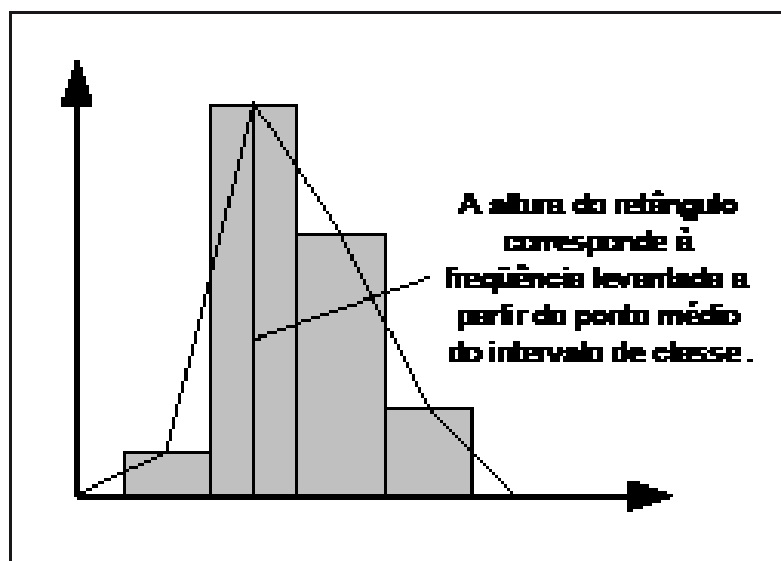


Figura 22: Polígono de Frequência: Esboço

Observando o esboço do polígono de frequência da Figura 22, acima, identificamos que a **linha** é construída a partir dos pontos médios dos topos dos retângulos de um histograma. A rigor, não precisamos construir o histograma, basta levantar uma reta a partir do ponto médio da base do triângulo (altura). Formalmente,



O polígono de frequência é um gráfico de linha, sendo as frequências marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classe (CRESPO, 1995, p. 70).

Seção 3: Um exercício completo

Vamos, agora, realizar um exercício completo sobre distribuição de frequência, envolvendo todos os fundamentos vistos até agora, incluindo a construção gráfica. Nosso problema é o seguinte:

³⁶ Ver Unidade 3: Variáveis, Tabelas e Gráficos, Seção 5: Gráficos, Diagramas, p. 49.

Ana Maria, secretária de uma grande escola, ouviu muitas conversas na secretaria. Em uma conversa dessas, ouviu uma reclamação do professor Paulo. As pessoas diziam que as notas dos seus alunos eram muito baixas; segundo a conversa, a maioria dessas notas eram abaixo da média.

Ana Maria ficou curiosa. Ela gostaria de *analisar o desempenho dos alunos do professor Paulo*, para saber se esses boatos eram verdade. Para realizar tal tarefa, ela seguiu **5 etapas**.

1ª Etapa: levantamento dos dados brutos. A primeira coisa a fazer era conseguir todas as notas dos alunos do professor Paulo. Isso foi fácil. O resultado está abaixo.

Tabela 19: Exercício: Tabela Primitiva

Notas dos alunos do professor Paulo											
5	7	7	2	0	0	3	9	8	4	8	4
1	7	9	6	7	7	1	4	0	2	1	1
3	9	7	5	6	4	9	8	6	5	4	0
8	9	3	2	9	6	8	7	4	5	4	8
3	2	8	8	0	5	3	5	1	5	9	0
9	9	3	9	8	8	7	5	8	7	0	2
7	7	1	7	7	1	7	0	6	3	2	0
2	7	8	6	2	1	6	7	4	6	9	6
5	1	7	9	2	5	9	1	8	5	2	8
7	3	0	7	8	8	6	9	7	4	8	3
5	2	5	1	8	8	8	7	4	0	3	6
2	9	8	4	8	5	8	6	5	8	6	4
2	1	1	0	3	9	0	3	8	1	2	9
1	7	4	9	0	3	8	1	2	9	7	7

Bem, como podemos notar, o professor Paulo possuía muitas turmas e, por isso, muitas notas. O levantamento inicial foi organizado em uma **Tabela primitiva**. Agora, é preciso expor esses dados em um **rol**.

2ª Etapa: construção de rol. Levantados os dados brutos, agora, é preciso organizá-los. Ana Maria realizou a tarefa colocando as notas em ordem crescente, conforme Tabela 20, abaixo.

Tabela 20: Exercício: Rol

Notas dos alunos do professor Paulo											
0	1	1	2	3	4	5	6	7	8	8	9
0	1	1	2	3	5	5	7	7	8	8	9
0	1	2	2	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9
0	1	2	3	4	5	6	7	7	8	8	9

Mesmo depois de ter organizado os dados, Ana Maria sentiu necessidade de diminuir os espaços. Essa foi a tarefa da próxima etapa.

3ª Etapa: construção da Tabela de Freqüência. Ana Maria percebeu que trabalhar com o *rol* era melhor que trabalhar com a *Tabela primitiva*. Mas, mesmo assim, sentiu necessidade de diminuir ainda mais a quantidade de dados. Para isso, ela construiu uma Tabela de Freqüência, já que percebeu que diversas notas se repetiam. Veja o resultado, abaixo:

Tabela 21: Exercício: Tabela de Freqüência

Notas dos alunos do professor Paulo	
Notas	Freqüência
0	14
1	16
2	15
3	13
4	13
5	15
6	13
7	24
8	26
9	19
10	0
Total 168	

Quando Ana Maria construiu a Tabela de Freqüência das notas dos alunos do professor Paulo ela verificou com mais clareza onde se concentravam a maioria das notas. A partir desse momento, ela já pôde dizer que as pessoas estavam enganadas, pois, embora parecesse que o professor Paulo atribuía muitas notas baixas, na verdade, as notas se concentravam entre 7, 8 e 9.

Ana Maria saiu da aparência: já pensou se ela emitisse alguma opinião com base, apenas, no levantamento inicial dos dados (Tabela Primitiva)? Bem, a chance dela fazer um julgamento equivocado seria muito grande. Mas ela ainda se sentia insegura. Portanto, ela agrupou os dados para uma análise mais apurada.

4ª Etapa: construção da Tabela de Freqüência com intervalos de classe. Quando Ana Maria decidiu agrupar ainda mais os dados, a primeira dificuldade a enfrentar foi: quantas classes e qual o intervalo delas? A primeira tarefa que realizou foi a determinação da *amplitude total de variação*, pois, a partir dela seria possível determinar os *intervalos de classes*.

Então, Ana Maria realizou a seguinte operação:

$$\text{amplitude total} = \text{nota maior} - \text{nota menor} = 9 - 0 = 9$$

De posse da amplitude total, Ana Maria decidiu que seu estudo teria 5 classes. Portanto, o *intervalo de classe* deveria ser:

$$\text{Intervalo de classes} = \frac{\text{amplitude total}}{\text{N}^{\circ} \text{ de classes}} = \frac{9}{5} = 1,8 = 2$$

Naquele momento, Ana Maria estava pronta para elaborar sua nova *Tabela de freqüência com intervalo de classes*. O resultado foi:



Ana Maria sabia que as classes, normalmente, variam de 5 a 20, conforme as regras para a elaboração de intervalos de classe.



Convém reforçar que se um intervalo é de 0 a 2 e outro intervalo é de 2 a 4, como fazer para não contar o 2 duas vezes?

A saída é considerar aquilo que na matemática se chama pontos abertos e fechados. Assim, no caso de 0 a 2, consideraremos fechado à esquerda e aberto à direita; vale dizer: o zero entra e o 2 não.

Da mesma forma, no intervalo de 2 a 4, o 2 entra e o 4 não; e assim sucessivamente.

Tabela 22: Exercício: Tabela de Frequência com intervalos de classe

Notas dos alunos do professor Paulo	
Notas	Frequência
0 a 2	30
2 a 4	28
4 a 6	28
6 a 8	37
8 a 10	45
Total 168	

Organizados os dados em uma **tabela de frequência com intervalos de classe**, Ana Maria pôde identificar, ao contrário do que as pessoas andavam conversando, que as notas se concentravam no intervalo de 8 a 10. Além disso, a segunda maior concentração das notas de seus alunos pertencia ao intervalo de 6 a 8. Os resultados do seu estudo, até aqui, demonstraram uma situação diferente do que poderia parecer à primeira vista.

Depois, para apresentar os resultados, Ana Maria construiu um gráfico.

5ª Etapa: representação gráfica. A fim de expor os dados rapidamente e com clareza, Ana Maria optou pelo **polígono de frequência**. Veja o resultado abaixo.

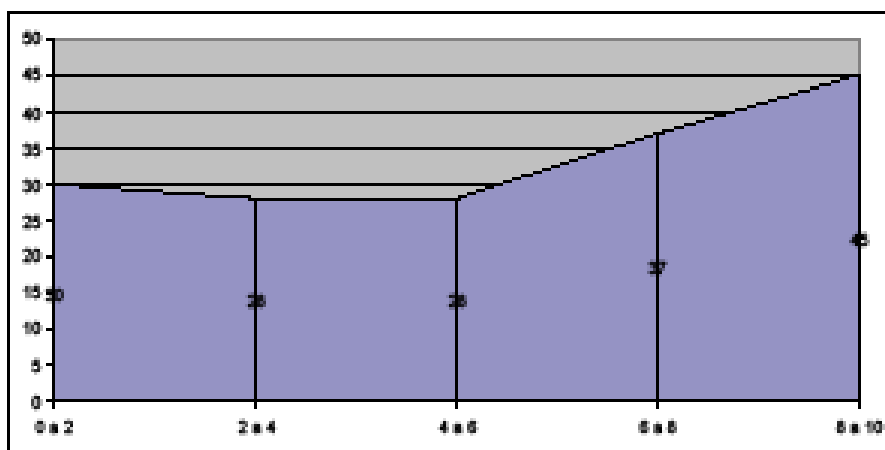


Gráfico 7: Exercício: Polígono de Frequência

Concluindo o estudo, o polígono de frequência parece demonstrar que o resultado do trabalho do professor Paulo é satisfatório, pois, há mais alunos com notas acima do intervalo de 4 a 6 do que abaixo dele. Nada mais podemos afirmar.

Chegamos ao fim do nosso exercício. Você observou que seguindo as etapas, não é difícil estudar, com rigor, um fenômeno qualquer. Que tal você realizar uma atividade parecida?

Selecione dois diários de classe e realize todas as cinco etapas do nosso exercício:

- 1) 1ª etapa: levantamento dos dados brutos;
- 2) 2ª etapa: construção do rol;
- 3) 3ª etapa: construção da Tabela de Frequência;
- 4) 4ª etapa: construção da Tabela de Frequência com Intervalos de Classe;
- 5) 5ª etapa: representação gráfica.

Sugiro que você realize a atividade com diários de professores que não estejam na escola. Caso não consiga acesso aos Diários de Classe, peça a alguém para inventar algumas notas ou invente você mesmo. Coloque os resultados em seu memorial.



“Os dados coletados podem, usualmente, ser considerados como pertencentes a uma amostra extraída de grande população. Como se dispõe de muitas observações da população, é teoricamente possível (para dados contínuos) a escolha de intervalos de classe muito pequenos e ter, até, números convenientes de observações que se situam dentro de cada classe. Assim, seria possível contar com um polígono de frequência [...] para uma grande população que tenha tantos pequenos segmentos de linha quebrada que se aproximem bastante de uma curva que será denominada curva de frequência [...]” (SPIEGEL, 1975, p. 49).

Seção 4: As Curvas de Frequência

Para completar nossa Unidade de estudo, vamos apenas tomar conhecimento de outras representações gráficas.

A tendência da análise de populações cada vez mais amplas é de que a *linha poligonal* se torne uma *curva*. Essa curva recebe o nome de **curva de frequência**. Enquanto o *polígono de frequência* nos dá a *imagem real* do fenômeno estudado, a *curva de frequência* nos dá a *imagem tendencial*.

Na prática, essas curvas aparecem de diversas formas. Observe a Figura 23, abaixo:



“A curva simétrica caracteriza-se por apresentar o valor máximo no ponto central e os pontos eqüidistantes [à mesma distância] desse ponto terem a mesma frequência.” (CRESPO, 1995, p. 74).

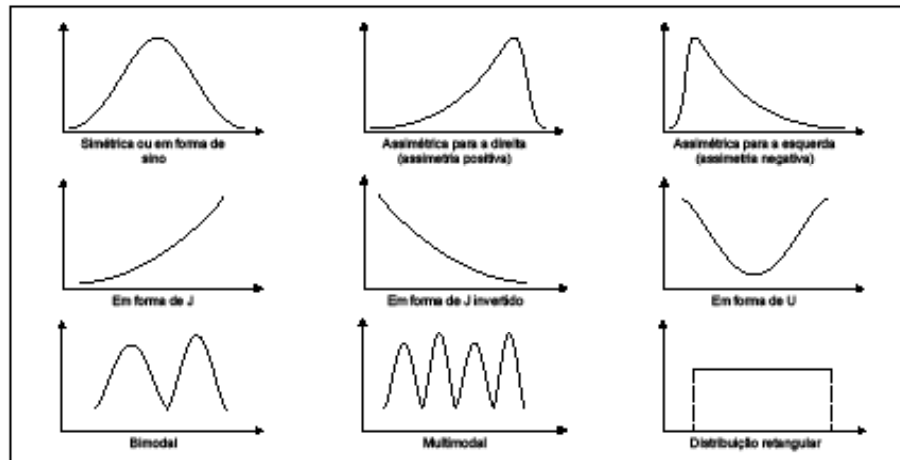


Figura 23: Curvas de Frequência

Cada curva apresenta, naturalmente, um significado diferente. A **curva simétrica** ou **em forma de sino** caracteriza-se pelo fato de apresentar um valor máximo na região central. A curva com esse comportamento simétrico é uma *curva normal*. Muitos fenômenos apresentam essa distribuição, tais como: a estatura dos adultos; o peso dos adultos; os preços relativos etc.³⁷

Alguns fenômenos apresentam uma *moderada assimetria*. Nas curvas **assimétricas** ou **desviadas**, a cauda da curva de um lado é mais longa do que do outro. Se a parte mais alongada fica à direita, chamamos a curva de **desviada para a direita** ou de **assimetria positiva**; se ocorre o contrário, a parte alongada fica à esquerda, a curva chama-se **desviada para a esquerda** ou de **assimetria negativa**.³⁸

As curvas **em forma de J** ou em **J invertido** são *extremamente assimétricas*. O ponto de máximo ocorre em uma das extremidades. São curvas típicas de fenômenos econômicos e financeiros, tais como: distribuição de vencimentos ou rendas pessoais.³⁹

Uma curva de frequência **em forma de U** possui ordenadas máximas em ambas as extremidades. Um bom exemplo de um fenômeno com esse comportamento é o da “mortalidade por idade”.⁴⁰

³⁷ CRESPO (1995, p. 74).

³⁸ SPIEGEL (1975, p. 49).

³⁹ CRESPO (1995, p. 75).

⁴⁰ CRESPO (1995, p. 75).

Tanto a curva **bimodal**, quanto a **multimodal** se referem à quantidade de pontos de máximos: a primeira, possui dois pontos de máximos; a segunda, mais de dois máximos.

Por fim, a **distribuição retangular** é uma manifestação rara. Apresenta todas as classes com a mesma freqüência. Representada em um *histograma*, todas as colunas apresentam a mesma altura e representada por um *polígono de freqüência*, reduz-se a um *segmento de reta horizontal*.⁴¹



1) Feita a coleta de dados das estaturas de 150 alunos, os resultados foram disponibilizados como abaixo (em centímetros). A partir de 145 cm, com intervalos de classe de 5 cm, exponha o resultado em uma Tabela.

159	150	170	155	162	148	168	156
150	153	165	162	155	165	153	157
159	152	155	164	146	155	163	147
152	150	149	157	167	154	157	156
151	159	163	146	157	162	157	158
152	152	146	160	171	156	147	156
154	159	166	156	152	147	166	148
152	153	177	148	162	150	150	162
159	154	148	150	160	156	148	160
153	159	161	156	156	155	178	151
161	154	156	160	146	172	158	153
150	151	147	148	157	146	163	159
155	159	167	151	163	158	167	157
160	154	158	163	158	161	148	158
153	153	161	161	147	158	157	154
174	159	168	147	151	168	169	158
151	154	147	155	155	151	151	157
155	152	164	164	158	164	164	149
163	152	169	149	164	153	--	--

Fonte: CASTRO (1964, p. 3)

2) A partir da Tabela de Distribuição de Freqüência, acima, construa o gráfico de barras que a representa.

41 CRESPO (1995, p. 76).

5

Medidas de resumen



"Numa comparação grosseira, é como se a mensuração fosse o processo de fotografar e medida, a fotografia resultante" (COSTA, 2004, p. 36).

Seção 1: Introdução

É preciso iniciar nossa última Unidade de estudo, fazendo a importante distinção entre **mensuração** e **medida**. **Mensuração** é processo do qual resulta uma **medida**; medida é valor, número resultante do processo de mensuração.⁴² Medir algo é, portanto, atribuir um número.

Há quatro níveis de medidas:

Os níveis de medidas	
Níveis	Variáveis
1º nível	Nominal , pois, apesar de expressa em números, é apenas um nome. Exemplos: número de telefone, RG, CIC, CPF etc. Esses números não são objetos de operações matemáticas.
2º nível	Ordinal , quando os itens podem ser colocados em ordem de grandeza. As notas escolares são um bom exemplo desse nível.
3º nível	Intervalar . Aqui, faz sentido quantificar. Na escala intervalar, <i>adição</i> e <i>subtração</i> são permitidas (mas <i>multiplicação</i> e <i>divisão</i> não). Escalas termométricas são um bom exemplo.
4º nível	Racional ou de razão . Nesse nível, <i>todas as operações matemáticas</i> são permitidas. Medidas tomadas com régua, fita métrica, balança, litro são bons exemplos, pois o <i>medido</i> corresponde ao <i>real</i> e não a uma correspondência.

Quadro 3: Níveis de medidas

Fonte: COSTA (2004, p. 36-40).

Pelos níveis de medidas acima, é fácil notar que um professor, ao atribuir uma nota bimestral a um aluno, está, na verdade, lidando com uma variável ordinal. Assim, ele está, apenas, indicando em uma escala, por exemplo, de 0 a 10, onde o aluno se encontra. Essa nota bimestral não é, portanto, uma medida racional, isto é, não possui a qualidade de uma medida obtida com uma fita métrica onde o resultado expressa a realidade.

Além disso, ao final do ano, os professores costumam tirar média das notas bimestrais. Isso é matematicamente sem sentido, pois, as notas não são reais, isto é, não representam a totalidade do conhecimento do aluno. Sendo assim, a

⁴² COSTA (2004, p. 36).

matemática não autoriza a operação com variáveis *ordinais*. Os professores costumam tirar média de notas. Por tradição e desconhecimento, não sabem que a Matemática não autoriza esse tipo de cálculo. Imagine que a nota de um aluno no 1º bimestre seja 5, o que isso significa? Significa que no processo de mensuração a resultante pode ser expressa pelo número 5 (medida). Isto é, numa escala de 0 a 10, o aluno pode ser colocado no posto 5. Somente isso, trata-se de uma variável ordinal, pois, pode ser colocado em uma ordem (ordem 5, na escala de 0 a 10). Não tem significado algum realizar operações com as notas do 1º e 2º bimestre para produzir uma resultante final. (COSTA, 2004).

Esse é um problema que, a meu ver, tarda em ser enfrentado. Mas fique sabendo que

“existe, hoje, embora com pouca divulgação entre nós, uma teoria capaz de dar conta dos problemas apontados: trata-se da Teoria de Resposta ao Item (TRI), extremamente complexa e fortemente dependente de conhecimentos probabilísticos. Pouco a pouco, essa teoria vai ganhando espaço, graças, entre outros fatores, à rápida evolução de recursos computacionais. Em países como Estados Unidos, Holanda e Espanha, a TRI já conta com forte adesão” (COSTA, 2004, p. 40).



A Teoria de Resposta ao Item (TRI) já possui vasta aplicação no Brasil. Consulte o endereço eletrônico abaixo, para ver a aplicação da TRI na produção de indicadores socioeconômicos. <http://www.scielo.br/pdf/pope/v25n1/24252.pdf>

81

Sem perder de vista a importante diferenciação entre **mensuração** e **medida**, passemos ao estudo das **medidas**. Em Estatística Descritiva,⁴³ alguns conceitos são fundamentais para analisarmos os dados, se quisermos uma análise responsável. As medidas podem ser divididas em:⁴⁴

- a) medidas de tendência central (média, moda e mediana);
- b) medidas de dispersão (desvio-padrão e coeficiente de variação);
- c) medidas de posição (*quartis, decís e percentis*).

Como a finalidade dessas medidas é resumir as informações, essas medidas são chamadas **medidas de resumo**.⁴⁵ Por essa

43 Ver Seção 2: Estatística Descritiva e Estatística Indutiva ou Inferencial, p. 42.

44 Segundo PEREIRA (2004, p. 11)

45 PEREIRA (2004).

razão, a *média*, por exemplo, é um valor que resume as informações de um conjunto maior de dados. Por exemplo, “quando um jornalista diz na TV que o salário médio do brasileiro é algo que gira em torno de R\$ 450,00 é porque muitos salários foram considerados, em todo o país, e o valor de R\$ 450,00 expressa esse conjunto de salários.” (PEREIRA, 2004, p. 11).

No nosso estudo, nesta **Unidade V**, enfocaremos algumas dessas medidas. Começaremos com as *medidas de tendência central*; nessa parte, *seção 2*, estudaremos a *média* e a *média aritmética ponderada*, a *mediana*, a *moda* e, por fim, a *relação entre média, mediana e moda*. Depois, na *seção 3*, estudaremos as *medidas de dispersão*, especialmente, os conceitos de *dispersão* e *variação*, *desvio padrão* e *coeficiente de variação*. Por último, na *seção 4*, estudaremos as *medidas de posição* conhecidas como *quartis*, *decis* e *percentis*.

Bom estudo a todos!

Seção 2: Medidas de Tendência Central



A média é a mais importante das medidas estatísticas.

A média é um valor típico de um conjunto de dados que tende a se localizar em um ponto central. Por essa razão, medidas com essa tendência são também denominadas **medidas de tendência central**. Vários tipos de médias podem ser definidos, sendo as mais comuns a **média aritmética**, a **média aritmética ponderada**, a **mediana** e a **moda**.⁴⁶

Média Aritmética

Para se calcular a **média aritmética**, ou simplesmente média, de um conjunto depende do tipo de dados. Para **dados não-agrupados** é muito simples. Observe o exemplo:

⁴⁶ Existem outras médias, tais como a Média Geométrica e a Média Harmônica, que não serão estudadas por nós.

As notas de um estudante em seis provas foram 8,4; 9,1; 7,2; 6,8; 8,7 e 7,8. Determinar a média aritmética das notas.

Solução:

$$\text{Média Aritmética} = \frac{8,4 + 9,1 + 7,2 + 6,8 + 8,7 + 7,8}{6} = \frac{48}{6} = 8,0$$

Figura 24: Média Aritmética: Exemplo

Fonte: Adaptado de SPIEGEL (1975, p. 80)

Observe que, na prática, o que realizamos foi *somar todas as notas* (48) e dividir pela *quantidade total de notas* (6).

Já que os números servem para “resumir” as informações, que tal diminuir a quantidade de dados por meio de fórmulas?

Estatísticos e matemáticos gostam muito de fórmulas. Isso se deve ao fato de elas “economizarem” quantidade de informações. Eles são muito práticos.

Assim, ao invés de escreverem “*média aritmética*”, na resolução de um exercício, eles utilizam a letra “*x*”, com uma barra em cima ($\bar{x}$); cada elemento do conjunto eles chamam de “ x_i ”; todos os elementos, “*n*” e, para representarem uma soma de todos os elementos de um conjunto, eles utilizam o símbolo chamado “*somatório*” ($\sum$).

Dessa maneira, a fórmula para a *média aritmética* fica assim representada:

$$\bar{x} = \frac{\sum x_i}{n}, \text{ onde } \begin{cases} \bar{x} = \text{Média Aritmética} \\ x_i = \text{Valores da Variável} \\ \sum x_i = \text{Soma Total dos Valores da Variável} \\ n = \text{Número Total de Valores} \end{cases}$$

Fórmula 1: Média Aritmética

Vamos realizar outro exercício para *dados não-agrupados* utilizando, desta vez, a Fórmula 1. Considere as aprovações na disciplina de matemática do professor João, de uma turma, nos últimos anos, representadas na série histórica abaixo:



Soma, Total ou $\sum$ são maneiras diferentes de representar a mesma coisa: a soma total.



Tabela 23: Série Histórica: Exercício

Total de aprovados em matemática – Professor João				
2001	2002	2003	2004	2005
35	38	32	40	37

Pergunta-se: qual a média aritmética dos aprovados nessa disciplina, no período considerado?

Solução:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{35+38+32+40+37}{5} = 36,4$$

Então, $\bar{x} = 36,4$.



*Você notou que não existe o número 36,4 no conjunto de dados? Quando isso acontece, dizemos que a média **não tem existência concreta**.⁴⁷ O que esse valor significa? Significa que, considerando todas as grandezas, dentro do conjunto de dados ordenados, esse valor tende a uma posição central, por isso, a média é uma medida de tendência central.*



*Vejamos, agora, como se calcula a média aritmética para **dados agrupados**. Os dados agrupados podem se apresentar sem intervalos de classe ou com intervalos de classes.⁴⁸*

*Vamos calcular a média aritmética para dados agrupados **sem intervalos de classe**. Considere a distribuição de frequência relativa a 34 famílias de quatro filhos, tomando como variável o número de filhos do sexo masculino,⁴⁹ abaixo.*

⁴⁷ CRESPO (1995, p. 80).

⁴⁸ Ver Unidade 4: Distribuição de Frequência, particularmente, a Seção 2: Distribuição de frequência e Aprofundamento: regras para a elaboração de uma distribuição de frequência, p. 61.

⁴⁹ CRESPO (1995, p. 82).

Tabela 24: Distribuição de Freqüência: Exercício

Número de filhos do sexo masculino	
Nº de meninos	Freqüência (f_i)
0	2
1	6
2	10
3	12
4	4
$\Sigma = 34$	

Fonte: CRESPO (1995, p. 82)

O levantamento foi realizado em 34 famílias, todas com 4 filhos. A coluna da esquerda, número de meninos, é a coluna indicadora. A coluna da direita, freqüência, é a coluna numérica.⁵⁰ De acordo com a Tabela de Distribuição de Freqüência, de todas as famílias em estudo, 2 famílias não possuíam meninos; 6 famílias apresentaram 1 menino; 10 famílias, 2 meninos; 12 famílias, 3 meninos e, por fim, 4 famílias possuíam 4 meninos.

Dessa forma, as freqüências são indicadoras da *intensidade* de cada valor da variável número de meninos. Esse é um caso de **ponderação**, o que nos leva a calcular a **média aritmética ponderada**, porque cada variável possui intensidade diferente.

Para o cálculo da média, precisaremos de outra Fórmula:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}, \text{ onde } \begin{cases} \bar{x} = \text{Média Aritmética} \\ x_i = \text{Valores da Variável} \\ f_i = \text{Freqüência} \\ x_i f_i = \text{Ponderação} \end{cases}$$

Fórmula 2: Média Aritmética Ponderada

50 Ver Unidade 3: Variáveis, Tabelas e Gráficos, Seção 4: Tabelas, p. 46.



Quando na Tabela aparece, por exemplo, que para 1 menino a freqüência é 6, é o mesmo que dizer que existem 1+1+1+1+1+1 meninos ou 6 vezes 1. Viu? Ponderar nada mais é do que considerar as repetições.



O modo mais prático para calcular uma média ponderada⁵¹ é construir na Tabela de Distribuição de Frequência mais uma coluna com os produtos “nº de meninos” vezes “frequência” (ou, segundo a fórmula, $x_i f_i$). Veja:

Tabela 25: Distribuição de Frequência: Exercício: Ponderação

Número de filhos do sexo masculino		
Nº de meninos	Frequência (f_i)	$x_i f_i$
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
	$\sum = 34$	$\sum = 78$

Agora ficou fácil. Temos, então, que:

$$\sum x_i f_i = 78 \text{ e } \sum f_i = 34$$

Logo, pela Fórmula 2:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \Rightarrow \bar{x} = \frac{78}{34} = 2,29 \Rightarrow \bar{x} = 2,3$$

A média de 2,3 nos indica que as famílias têm em média 2 meninos e 2 meninas, sendo que existe uma tendência geral de uma leve superioridade numérica dos meninos em relação ao número de meninas.

Por fim, vamos calcular a média aritmética para *dados agrupados com intervalos de classes*. Quando os dados são apresentados em uma distribuição de frequência, todos os valores incluídos num certo intervalo de classe são considerados coincidentes com o ponto médio do intervalo.⁵² Para o cálculo da média aritmética ponderada, utilizamos a Fórmula 2:

⁵¹ Para falar a verdade, sempre que formos aplicar uma Fórmula, construiremos tabelas de auxílio. Desse modo, identificamos os dados da Fórmula e, depois, encontramos o resultado.

⁵² SPIEGEL (1975, p. 73).

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}, \text{ onde } x_i \text{ é o ponto médio da classe.}$$



Qual o ponto médio do intervalo de 0 até 2?

A resposta é 1.

Qual é o ponto médio do intervalo de 2 a 4?

A resposta é 3.

Viu?

Ponto médio é o ponto que está no meio do intervalo.

Dessa forma, o raciocínio é o mesmo para a média aritmética ponderada sem intervalos de classe.

Vamos realizar um exercício. Você se lembra do professor Paulo? Bem, vamos retornar às notas dos alunos dele.⁵³

Tabela 26: Distribuição de Frequência: Exercício: Ponderação: Ponto Médio

Notas dos alunos do professor Paulo			
Notas	f_i	x_i	$x_i f_i$
0 a 2	30	1	30
2 a 4	28	3	84
4 a 6	28	5	140
6 a 8	37	7	259
8 a 10	45	9	405
	$\sum = 168$		$\sum = 918$

A Tabela 26, acima, recuperou a distribuição de frequência do professor Paulo, acrescentando, apenas, o ponto médio dos intervalos de classe (x_i) e a ponderação, isto é, o produto dos pontos médios pela frequência ($x_i f_i$). Bem, sabemos, portanto, que:

$$\sum x_i f_i = 918 \text{ e } \sum f_i = 168$$

Logo, utilizando a Fórmula 2 para o cálculo da média aritmética ponderada, temos que:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \Rightarrow \bar{x} = \frac{918}{168} = 5,46 \Rightarrow \bar{x} = 5,5$$

⁵³ Tabela 22, p. 70.



*O que isso indica? Indica que temos que mudar nossa opinião sobre o trabalho do professor Paulo. E por quê? Porque a “análise” que realizamos, naquele momento, nos levou a afirmar “que o resultado do trabalho do professor Paulo é satisfatório, pois, há mais alunos com notas acima do intervalo de 4 a 6 do que abaixo dele”. Você se lembra?*⁵⁴

E o que mudou de lá para cá? Bem, a média das notas do professor sendo 5,5, indica que praticamente, metade dos alunos do professor estão com notas abaixo de 5,0, com uma tendência para notas acima de 5,0. Ora, isso não parece tão satisfatório, não é mesmo? Diante disso, não é ilícito afirmar que o professor Paulo precisa rever seus processos de mensuração.⁵⁵



Calcule a média dos acidentes de trânsito, na Região Centro-Oeste, em 2002.

Tabela 27: Vítimas de Acidentes de Trânsito, por 10.000 veículos, em 2002

Unidade da Federação	Vítimas de acidentes
Distrito Federal	11.256
Brasília	6.747
Goiás	22.383
Goiânia	9.567
Mato Grosso	-
Cuiabá	-
Mato Grosso do Sul	7.346
Campo Grande	3.071

Fonte: Adaptado de Anuário Estatístico de Acidentes de Trânsito (2002)

⁵⁴ Ver p. 70.

⁵⁵ Sobre *mensuração e medida*, ver Seção 1: Introdução desta Unidade, p. 76.

Mediana e Média

Em um conjunto ordenado, o ponto central que divide esse conjunto em dois subconjuntos com o mesmo número de elementos chama-se **mediana**. Aqui, diferentemente da *média* (que nos fornece a concentração dos dados), a *mediana* nos fornece a posição que divide, exatamente, um conjunto em função da quantidade de seus elementos. Por exemplo:

Vamos considerar o conjunto dos números

3, 4, 4, 5, 6, 8, 8, 8, 10

Quem está no meio do conjunto?

6

Então, os elementos antes de 6 são:

3, 4, 4 e 5

E depois de 6:

8, 8, 8 e 10

Observe que temos a mesma quantidade de elementos antes e depois de 6. A **mediana** indica isso: o número que divide o conjunto ao meio, isto é, a quantidade antes e depois dele é a mesma. Assim,

A mediana é [...] definida como o número que se encontra no centro de uma série de números, estando estes dispostos segundo uma ordem. Em outras palavras, a mediana de um conjunto de valores, ordenados segundo uma ordem de grandeza, é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos. (CRESPO, 1995, p. 93).

Para **dados não agrupados**, como no exemplo acima, calcula-se a mediana de duas maneiras:

- 1) quando os dados forem de número ímpar, basta encontrar o *ponto central*, isto é, encontrar o valor que antes dele e depois dele, tenham o mesmo número de elementos;
- 2) quando os dados forem de número par, não haverá um ponto central. Nesse caso, calcula-se o *ponto médio* dos dois valores centrais, com a ajuda da média aritmética.



Não se esqueça que, para fazer isso, é preciso que os elementos estejam em um rol, isto é, apresentem-se em uma ordem crescente ou decrescente.

Considere o conjunto:⁵⁶ 145, 68, 1, 2, 6, 5, 4, 3, 4, 8. Vamos calcular a *média* e a *mediana* (*md*). A primeira coisa a fazer, nunca se esqueça, é colocar os elementos em ordem:

1, 2, 3, 4, 4, 5, 6, 8, 68, 145.

Efetuada os cálculos:

Média

Aplicando a Fórmula 1, temos:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3+4+4+5+6+8+68+145}{10} = 24,6$$

Mediana

Para conjunto de dados par, realizar a média dos dois pontos centrais:

$$md = \frac{4+5}{2} = 4,5$$

Observe que a *média* é muito diferente da *mediana*. Média igual a 24,6 significa que os dados do conjunto se concentram em torno desse número, isto é, “o problema da média é que *ela é afetada pelos grandes valores*” (PEREIRA, 2004, p. 19)⁵⁷. Com o cálculo da *mediana* (*md*) igual a 4,5, podemos afirmar que metade dos valores está abaixo de 4,5 e, portanto, são muito baixos.

Embora ambas as medidas sejam de tendência central (ou seja, representem pontos que tendem para o centro dos dados), no nosso caso, os valores do conjunto estão mais próximos de 4,5 do que de 24,6, não concorda? Por isso dizemos que a média leva em conta os valores e a mediana não.

⁵⁶ PEREIRA (2004, p. 20).

⁵⁷ Um exemplo dessa importante informação: dizer que a *média* dos salários de três amigos meus é de R\$ 1.900,00 não me indica quase nada, pois, eles podem receber R\$ 350,00, R\$ 350,00 e R\$ 5.000,00. O que isso prova? Prova que a média é afetada pelos grandes valores.

Se os dados estão **agrupados**, para calcular a mediana utilizamos a fórmula:

$$\sum \frac{f_i}{2}$$

Fórmula 3: Mediana

No caso de **dados agrupados sem intervalos de classe**, como é o caso da Tabela 28, abaixo, podemos utilizar um recurso que nos auxilia a calcular a *mediana*: a coluna de freqüências acumuladas (F_i). **Freqüência acumulada** nada mais é do que a soma das freqüências de cada variável. Observe que para a variável “0 menino”, temos freqüência 2, logo, a freqüência acumulada é 2; para a variável “1 menino”, temos freqüência 6, logo, a freqüência acumulada é 8, pois, 2 (freqüência acumulada anterior) + 6 (freqüência simples); para a variável “2 meninos”, temos freqüência simples igual a 10, logo, a freqüência acumulada será 8 (anterior) + 10 = 18; e assim sucessivamente. Freqüência acumulada será então, a soma das freqüências simples.



Observe que, para freqüência, utilizamos o símbolo f_i . Quando queremos nos referir à freqüência acumulada, utilizamos F_i .

Tabela 28: Distribuição de Freqüência: Exercício: Mediana:
Freqüência Acumulada

Número de filhos do sexo masculino		
Nº de meninos	f_i	F_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\sum = 34$	

Fonte: CRESPO (1995, p. 95).

Pois bem, como calcular o ponto que divide igualmente a quantidade de valores acima e abaixo dele, ou seja, como calcular a *mediana*? Para o cálculo da mediana, aplicamos a Fórmula 3. O resultado indica que a mediana será um dos valores da coluna da esquerda (0, 1, 2, 3 ou 4) correspondente à freqüência acumulada imediatamente superior.

Vamos resolver o exercício acima. Sabemos que

$$\sum f_i = 34$$

Aplicando a Fórmula 3, temos que

$$\frac{\sum f_i}{2} = \frac{34}{2} = 17$$

Pela Fórmula 3, a mediana é 17. Na Tabela existe freqüência acumulada 17? Não. Caso existisse, aquela seria a linha em se encontraria a mediana. Mas, no caso de não existir, como proceder? Simples, veja:

As freqüências acumuladas são 2, 8, 18, 30 e 34. Qual é a imediatamente superior a 17? Isso mesmo, 18. Então, vamos destacar a linha:

Linha da mediana		
Número de filhos do sexo masculino		
Nº de meninos	f_i	F_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\Sigma = 34$	

Mediana

Freqüência Acumulada imediatamente superior

Figura 25: Linha Mediana

O número 17, conseguido com a Fórmula 3, indica que a mediana pertence à linha em que esse número se encontra. Mas como não há freqüência acumulada 17, como não é possível encontrar diretamente 17 na freqüência acumulada, então, consideramos a freqüência acumulada imediatamente superior. Nesse caso, essa freqüência é o 18. Destacamos a **linha mediana**, isto é, a linha onde a nossa mediana procurada se encontra. A mediana é, portanto, 2.

Vamos explorar um pouco mais esse resultado. Observe o Gráfico 8, abaixo:

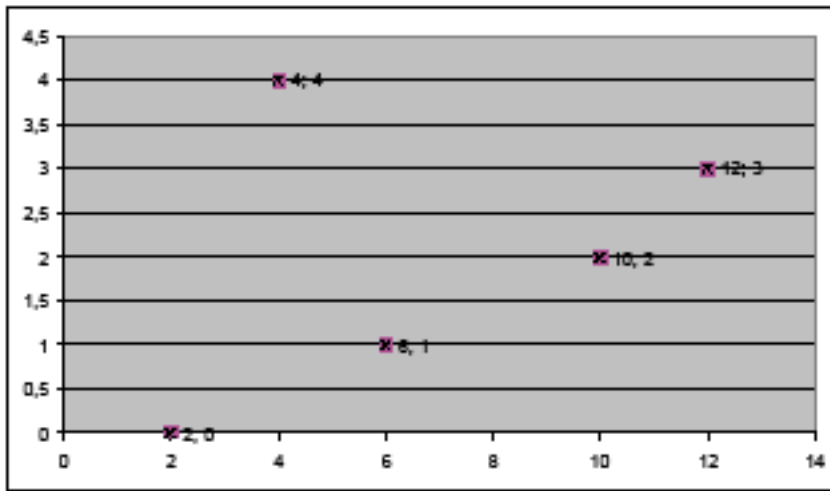


Gráfico 8: Mediana

O Gráfico 8 mostra que: duas famílias não possuem filhos meninos (2,0); 4 famílias possuem 4 meninos; seis famílias possuem 1 menino (6,1); 10 famílias possuem 2 meninos (10,2); 12 famílias possuem 3 meninos (12,3). Temos no nosso conjunto 78 meninos, por quê? Veja:

- 2 famílias não possuem meninos $\rightarrow 2 \times 0 = 0$;
- 4 famílias possuem 4 meninos $\rightarrow 4 \times 4 = 16$;
- 6 famílias possuem 1 menino $\rightarrow 6 \times 1 = 6$;
- 10 famílias possuem 2 meninos $\rightarrow 10 \times 2 = 20$;
- 12 famílias possuem 3 meninos $\rightarrow 12 \times 3 = 36$.

Logo, o total de meninos é $0 + 16 + 6 + 20 + 36 = 78$ ($\sum = 78$).

A mediana encontrada foi 2, isso significa que **as famílias que possuem dois meninos** dividem nosso conjunto de 78 meninos ao meio: metade desses meninos estão nas famílias com nenhum filho, com um filho e com dois filhos; a outra metade é composta de famílias com dois meninos, com três meninos e famílias com quatro meninos. Agora ficou mais claro que a mediana divide nosso conjunto ao meio.



Vá à Secretaria de sua escola e pegue, aleatoriamente, dados sobre 10 famílias. Calcule a média e a mediana do número de filhas meninas.

Ainda não concluímos o estudo sobre mediana. É preciso, por último, calcular a mediana de **dados agrupados em intervalos de classe**. Mas isso, faremos mais à frente.

Moda



Em um conjunto de números, chamamos de **moda** o valor que ocorre com maior frequência, isto é, o valor mais comum. É assim que podemos dizer que “o salário modal dos empregados de uma indústria é o salário mais comum, isto é, o salário recebido pelo maior número de empregados dessa indústria”. (CRESPO, 1995, p. 89). Por exemplo:⁵⁸

- a) O conjunto 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 tem *moda* 9;
- b) O conjunto 3, 5, 8, 10, 12, 15, 16 não tem moda;
- c) O conjunto 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9 tem duas modas, 4 e 7. Nesse caso é chamado *bimodal*.

Para **dados agrupados sem intervalos de classe**, é possível determinar imediatamente a moda, como nos exemplos acima. Mas, por exemplo, a Tabela 28, p. 77, indica que a moda é 3. Por quê? Porque o valor que mais se repete é aquele que possui maior frequência simples, não é mesmo?

É ainda possível encontrar a moda para **dados agrupados com intervalos de classe**, mas deixaremos esse estudo para uma outra oportunidade.

Expressões gráficas da moda



Em uma curva de frequência, o maior valor de um conjunto é chamado **moda**. Na prática, a **moda** é o valor que corresponde, no eixo das abscissas, ao ponto de ordenada máxima, em outras palavras. Veja exemplos abaixo:

⁵⁸ SPIEGEL (1975, p. 74).

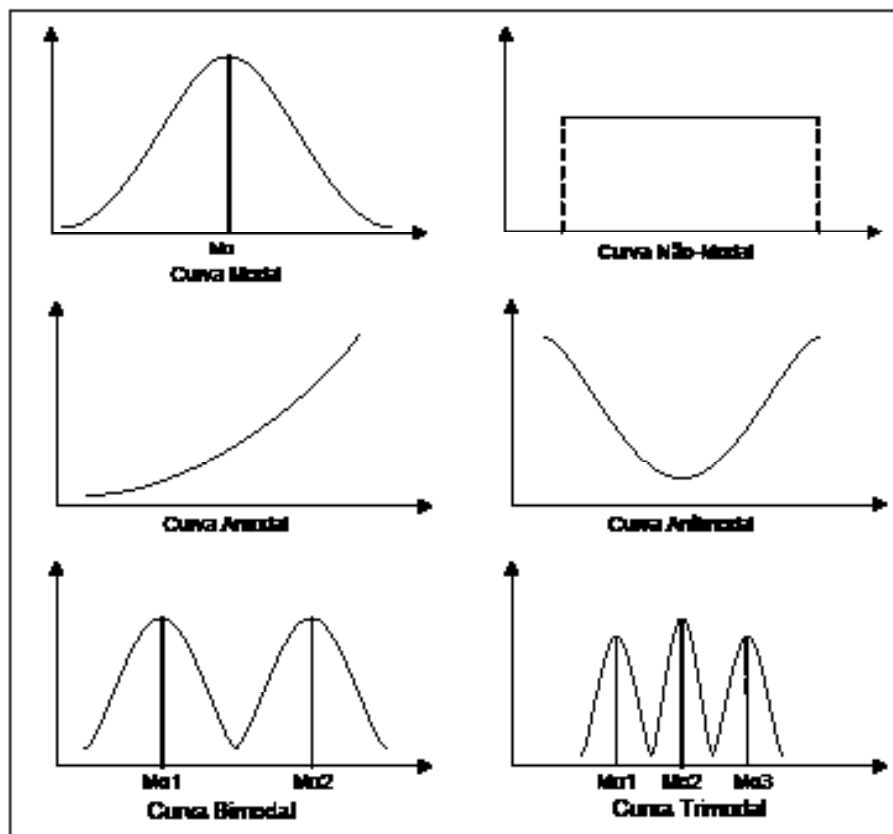


Figura 26: Curvas Modais

Vamos verificar a Curva Modal, acima (primeiro gráfico). Repare que ela possui um valor maior, mais alto no gráfico. O que isso indica? Indica que é o maior valor que o conjunto pode assumir, por isso, é a *moda* do conjunto.

Já no último gráfico – Curva Trimodal –, identificamos três valores de máximo, isto é, o conjunto possui três valores “maiores” que todos os demais, por isso, trimodal.



Conjuntos com mais de três valores máximos são chamados de polimodais.

Relação entre Média, Mediana e Moda

Em curvas simétricas, unimodais, a média ($\bar{x}$), a mediana (Md) e a Moda (Mo) coincidem. Observe:

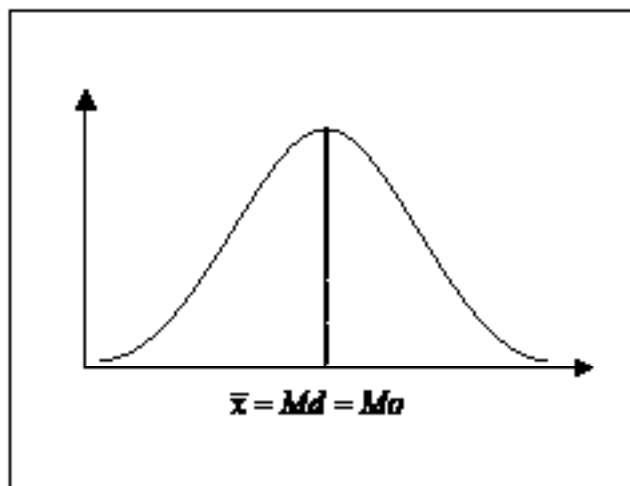


Figura 27: Média, Mediana, Moda: Curva Simétrica

Em curvas de frequência desviadas para a direita e para a esquerda, as posições são diferentes. Veja:

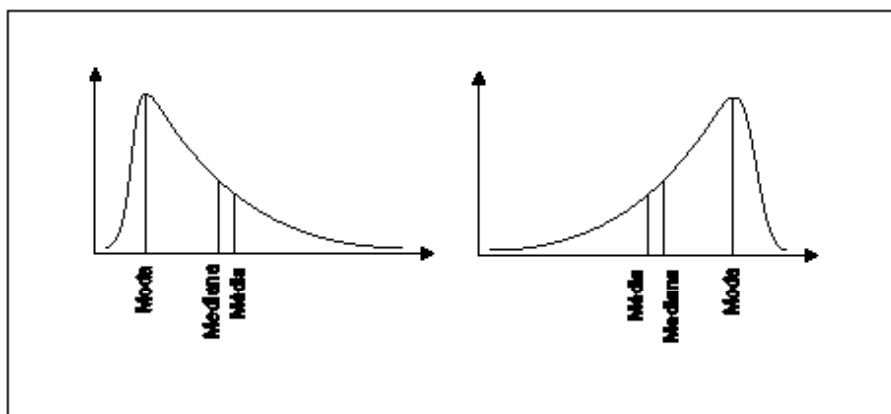


Figura 28: Média, Mediana, Moda: Curva Assimétrica



Determinar a média, a mediana e a moda dos conjuntos de números:⁵⁹

A = 7, 4, 10, 9, 15, 12, 7, 9, 7

B = 8, 11, 4, 3, 2, 5, 10, 6, 4, 1, 10, 8, 12, 6, 5, 7

(Atenção: não se esqueça de colocar os conjuntos em rol).

⁵⁹ SPIEGEL (1975, p. 105).

Seção 3: Medidas de Dispersão

Até aqui, vimos que **média**, **mediana** e **moda** são valores que podem servir de comparação, mas, fundamentalmente, fornecem a posição de qualquer elemento do conjunto. Mas para interpretar dados estatísticos, mesmo aqueles já convenientemente simplificados, é preciso conhecer a evolução desses dados.

Um exemplo clássico para a compreensão da importância das **medidas de dispersão** é o da comparação de temperaturas entre cidades⁶⁰: saber que a temperatura média de duas cidades é de 24°C não me diz muita coisa a respeito da variação dessas temperaturas.

Em uma cidade, o dia pode ter iniciado muito frio e terminado muito quente; aqui, ocorreu uma grande variação da temperatura.

Na outra cidade, o dia pode ter iniciado e terminado como 24°C; nesse caso, não haveria variação alguma de temperatura.

Viu? Embora as médias sejam importantes, elas não são suficientes para as *inferências estatísticas*, por isso, precisamos de outras medidas.

Vamos reforçar a importância das medidas de dispersão, por meio de um exercício. Consideraremos os três conjuntos abaixo, com seus respectivos valores:⁶¹

X: 70, 70, 70, 70, 70.

Y: 68, 69, 70, 71, 72.

Z: 5, 15, 50, 120, 160.

Vamos calcular a *média* das idades dos três conjuntos:

Solução:

Para calcular as médias, precisaremos da Fórmula 1, p. 71:

$$\bar{x} = \frac{\sum x_i}{n}, \text{ onde } \begin{cases} \bar{x} = \text{Média Aritmética} \\ x_i = \text{Valores da Variável} \\ n = \text{Número Total de Valores} \end{cases}$$

60 CRESPO (1995, p. 108).

61 CRESPO (1995, p. 108).

Então,

$$\text{Para X: } \bar{x} = \frac{70+70+70+70+70}{5} = \frac{350}{5} = 70$$

$$\text{Para Y: } \bar{x} = \frac{68+69+70+71+72}{5} = \frac{350}{5} = 70$$

$$\text{Para Z: } \bar{x} = \frac{5+15+50+120+160}{5} = \frac{350}{5} = 70$$

Como podemos observar, os três conjuntos possuem a mesma *média aritmética*: 70.

Mas também, podemos notar que o conjunto **X** é mais **homogêneo** do que os conjuntos **Y** e **Z**; o conjunto **Y**, por sua vez, é mais **homogêneo** que o conjunto **Z**; por fim, o conjunto **Z** é o mais **heterogêneo** de todos. Viu? Mesmo possuindo a mesma média, os conjuntos apresentam comportamentos muito diferentes. A isso chamamos de **dispersão**.

Dispersão e Variação



Dispersão (ou variabilidade) de um conjunto refere-se à maior ou menor diversificação dos valores de uma variável em torno de um valor de tendência central⁶² tomado como ponto de comparação.

No nosso exercício acima, os conjuntos **X**, **Y** e **Z** apresentam como ponto de tendência central para fins de comparação a *média*. Essa *média* é a mesma para os três conjuntos: 70. Assim, o conjunto **X** apresenta *dispersão* nula, pois não há variação dos valores do conjunto em relação a essa média; o conjunto **Y** apresenta *dispersão* menor que o conjunto **Z**; isso porque os valores de **Y** estão mais próximos que os do conjunto **Z**.

Em resumo, a estatística recorre às **medidas de dispersão** (ou **de variabilidade**) quando deseja *qualificar* os valores de uma variável, ressaltando a maior ou menor dispersão entre

⁶² Ver Seção 2: Medidas de Tendência Central, p. 78.

esses valores e a sua medida de posição.⁶³ Dessas medidas de dispersão,⁶⁴ estudaremos apenas o *desvio padrão* e o *coeficiente de variação*.

Desvio Padrão

O desvio padrão é a medida da variação, da dispersão, de um conjunto.

Assim, quanto maior for o desvio padrão, maior será a heterogeneidade entre os valores que estão sendo analisados. Isso significa, portanto, que quanto maior for o desvio padrão, maior será a variação entre os valores. Vamos entender melhor isso.

De volta aos conjuntos **X**, **Y** e **Z** acima, vimos que a *média* de todos eles era 70. Notamos, também, que os conjuntos **X** e **Y** eram mais homogêneos que o conjunto **Z**. Agora vamos calcular essa medida matematicamente, utilizando mais uma fórmula:

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}, \text{ onde } \begin{cases} s \text{ é o desvio padrão} \\ n \text{ é a soma das frequências} \end{cases}$$

Fórmula 4: Desvio Padrão: Dados Não Agrupados

Os nossos conjuntos **X**, **Y** e **Z** são de **dados não agrupados**. Vamos representá-los em Tabelas, para melhor visualização.



Conjuntos mais homogêneos apresentam desvios-padrão menores.



Muita atenção à diferença abaixo:

$$\frac{\sum x_i^2}{n} \neq \left(\frac{\sum x_i}{n}\right)^2$$

Matematicamente, os parênteses alteram tudo. Acompanhe o exercício para detectar a diferença.

⁶³ Veremos, mais adiante, as *medidas de posição*. Por ora, podemos considerar, apenas, as medidas de tendência central.

⁶⁴ A lista de medidas de dispersão é longa. Para Spiegel (1975), essas medidas são: a amplitude total; o desvio médio; a amplitude semi-interquartílica ou o desvio quartílico; o desvio-padrão; a variância; o coeficiente de variação.

Tabela 29: Desvio Padrão: Exercício

Tabela X		Tabela Y		Tabela Z	
x_i	x_i^2	x_i	x_i^2	x_i	x_i^2
70	490	68	4624	5	25
70	490	69	4761	15	225
70	490	70	4900	50	2500
70	490	71	5041	120	14400
70	490	72	5184	160	25600
$\sum = 350 \quad \sum = 24500$		$\sum = 350 \quad \sum = 24510$		$\sum = 350 \quad \sum = 42750$	

Note que cada valor do conjunto é representado por x_i e seu quadrado é x_i^2 . Sabemos que n é igual a 5, para todos os conjuntos. Agora ficou fácil calcular o desvio padrão dos três conjuntos. Vejamos:

Solução:

Aplicando a Fórmula 4, temos que:

Para o conjunto X:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 2450 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{24500}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{4900 - 4900} = 0$$

Para o conjunto Y:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 24510 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{24510}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{4902 - 4900} = \sqrt{2} = 1,4$$

Para o conjunto Z:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 42750 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{42750}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{8550 - 4900} = \sqrt{3650} = 60,4$$

Você reparou que colocando na tabela os elementos que iremos usar (x_i e x_i^2) fica mais fácil resolver o problema? Depois de todos esses cálculos, temos que:

- O desvio padrão do conjunto **X** é igual a 0. De fato, isso significa que não há variação alguma no conjunto, portanto, é um conjunto *homogêneo*;
- O desvio padrão do conjunto **Y** é igual a 1,4 e o do conjunto **Z** é igual a 60,4. Comparando-se os dois conjuntos, vemos que há uma pequena variação em **Y** (1,4) e uma alta variação em **Z** (60,4). Na prática, significa que os valores do conjunto **Y** estão mais próximos da *média*, ao passo que, em **Z**, os valores do conjunto estão muito distantes da *média*.

Graficamente, é ainda mais fácil identificar um conjunto *mais homogêneo*. Observe:

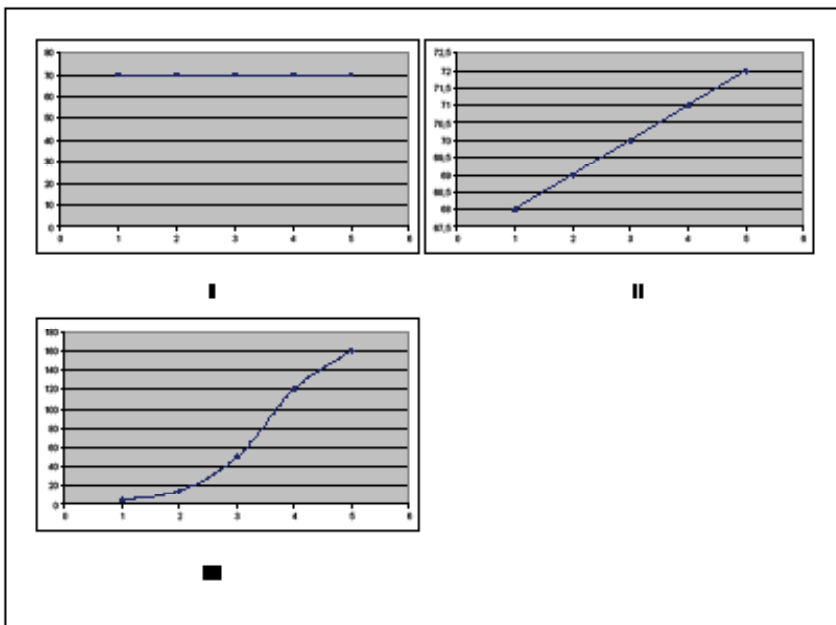


Figura 29: Desvio Padrão: Gráficos: Exercício

Você é capaz de dizer qual das três representações gráficas acima, é o conjunto **X**? E o conjunto **Y**? E o conjunto **Z**? Note que se o conjunto for homogêneo (I), o gráfico é uma linha reta paralela ao eixo x ; observe também, que quanto menos homogêneo o conjunto, a reta tenderá a ser uma curva.



Calcule o desvio padrão dos conjuntos abaixo:

$A = 12, 6, 7, 3, 15, 10, 18, 5$

$B = 9, 3, 8, 8, 9, 8, 9, 18$

Vamos fazer um exercício de cálculo do desvio padrão para conjuntos com **dados agrupados sem intervalos de classe**. Nesse caso, como temos frequências (ou seja, como os valores se repetem), vamos fazer uma pequena alteração na Fórmula.

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}, \text{ onde } \begin{cases} f_i x_i^2 \text{ é o produto do quadrado dos valores pela frequência} \\ f_i x_i \text{ é o produto dos valores pela frequência} \end{cases}$$

Fórmula 5: Desvio Padrão: Dados Agrupados

Vamos encontrar o desvio padrão da Tabela 30, abaixo.

Tabela 30: Desvio Padrão: Dados Agrupados: Sem Intervalos de Classe: Exercício

x_i	f_i
0	2
1	6
2	12
3	7
4	3
$\Sigma = 30$	

Fonte: CRESPO (1995, p. 115).

Da mesma maneira que estamos resolvendo nossos exercícios, aqui, vamos acrescentar à Tabela três colunas que serão úteis.

Tabela 31: Desvio Padrão: Exercício: Continuação

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$
0	2	0	0	0
1	6	6	1	6
2	12	24	4	48
3	7	21	9	63
4	3	12	16	48
$\Sigma = 30 \quad \Sigma = 63$				$\Sigma = 165$

Com a Tabela assim, é fácil encontrar o desvio padrão. Veja:

Sabendo que: $\begin{cases} n = \Sigma f_i = 30 \\ \Sigma f_i x_i^2 = 165 \\ \Sigma f_i x_i = 63 \end{cases}$. Então,

$$s = \sqrt{\frac{\Sigma f_i x_i^2}{n} - \left(\frac{\Sigma f_i x_i}{n}\right)^2} = \sqrt{\frac{165}{30} - \left(\frac{63}{30}\right)^2} = \sqrt{5,5 - 4,41} = \sqrt{1,09} = 1,044$$

Portanto, o desvio padrão é de 1,044.

Para encontrar o **desvio padrão** de um conjunto **com intervalos de classe**, utilizaremos o mesmo recurso de acrescentar à tabela os dados que iremos precisar na mesma Fórmula 5, acima. Como recurso didático, usaremos a mesma Fórmula para dados agrupados *sem intervalos de classe*.

Primeiro, vamos repetir a Fórmula 5:

$$s = \sqrt{\frac{\Sigma f_i x_i^2}{n} - \left(\frac{\Sigma f_i x_i}{n}\right)^2}, \text{ onde } \begin{cases} f_i x_i^2 \text{ é o produto do quadrado dos valores pela frequência} \\ f_i x_i \text{ é o produto dos valores pela frequência} \end{cases}$$

Suponha, agora, que queiramos encontrar o desvio padrão da Tabela 32, abaixo:



Relembrando: Se n é quantidade de valores por que deu 30 se os valores são 0, 1, 2, 3 e 4? Ou seja, por que n não é 5?

Simples! Porque, na verdade, a Tabela indica que temos os seguintes valores: 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4 e 4. Isso é que é a frequência (f_i). Temos, portanto, 30 valores organizados por frequências.

Tabela 32: Desvio Padrão: Dados Agrupados: Com Intervalos de Classe: Exercício

Estaturas	f_i
150–154	4
154–158	9
158–162	11
162–166	8
166–170	5
170–174	3
	$\Sigma = 40$

Fonte: CRESPO (1995, p. 116)



Por exemplo, no intervalo 150–154, os valores podem assumir de 150 cm até 154 cm: esses são os valores de mínimo e de máximo.

O que essa tabela apresenta de diferente? Os dados são agrupados com intervalos de classe. Ou seja, os valores variam de *um valor mínimo para um máximo*. Portanto, temos um problema a resolver!

A Fórmula 5, acima, é para o cálculo do desvio padrão de um conjunto de *dados agrupados sem intervalos de classe*. Isso significa que nela temos x_i e não um *intervalo de classe*, como, por exemplo, 150–154. Mas se eu tivesse **um valor** ao invés de **um intervalo de valores** (como é o caso), a Fórmula 5 poderia ser a mesma, não é verdade?

Bem, vamos utilizar um recurso para manter a mesma Fórmula: vamos encontrar um ponto, que chamaremos **ponto médio**, para cada *intervalo de classe*. Dessa maneira, teremos x_i como no exercício anterior e, assim, poderemos utilizar a mesma Fórmula.

Os demais elementos ($f_i x_i$, x_i^2 e $f_i x_i^2$) já sabemos como encontrar. Agora, vamos à solução. Nossa Tabela, com os acréscimos necessários, ficará assim:



O ponto médio é o ponto que está no meio do intervalo. Veja: O que está no meio do intervalo que varia de 150 cm a 154 cm? 152 cm é o ponto médio. Qual é o ponto médio do intervalo 154–158? É 156 cm que está no meio. E assim por diante.

Tabela 33: Desvio Padrão: Exercício: Continuação

Estaturas	f_i	x_i	$f_i x_i$	x_i^2	$f_i x_i^2$
150–154	4	152	608	23.104	92.416
154–158	9	156	1.404	24.336	219.024
158–162	11	160	1.760	25.600	281.600
162–166	8	164	1.312	26.896	215.168
166–170	5	168	840	28.224	141.120
170–174	3	172	516	29.584	88.752
	$\Sigma = 40$		$\Sigma = 6.440$		$\Sigma = 1.038.080$

Com a Tabela preenchida, vamos encontrar o desvio padrão.

Solução:

Sabendo que $\begin{cases} n = \sum f_i = 40 \\ \sum f_i x_i^2 = 1.038.080 \\ \sum f_i x_i = 6.440 \end{cases}$. Então,

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{1.038.080}{40} - \left(\frac{6.440}{40}\right)^2} = \sqrt{25.952 - 25.921} = \sqrt{31} = 5,567$$

Viu?! Acrescentando os dados que iremos necessitar para o cálculo à Tabela, tudo fica mais fácil. O desvio padrão é 5,57 cm.

Calcule o desvio padrão da distribuição abaixo:



CUSTO (R\$)	450	–	550	–	650	–	750	–	850	–	950	–	1.050	–	1.150
f_i	8		10		11		16		13		5		1		

Fonte: CRESPO (1995, p. 118).



Não se esqueça de montar a Tabela.

Estaturas	f_i	x_i	$f_i x_i$	x_i^2	$f_i x_i^2$
450–550	8				
550–650	10				
650–750	11				
750–850	16				
850–950	13				
950–1.050	5				
1.050–1.150	1				
	$\Sigma =$		$\Sigma =$		$\Sigma =$

Coeficiente de Variação

Até aqui, nossos esforços têm se voltado para caracterizar, com o maior rigor possível, a dispersão dos conjuntos. O **coeficiente de variação** é uma medida muito útil para essa intenção.

O **coeficiente de variação** (CV) está sempre relacionado ao valor médio de um conjunto porque, como já vimos, a dispersão é uma medida sempre relacionada a uma determinada média.

Sua fórmula é bastante simples:

$$CV = \frac{\text{desvio padrão}}{\text{média}} \times 100$$

De maneira mais simplificada:

$$CV = \frac{s}{\bar{x}} \times 100, \text{ onde } \begin{cases} s \text{ é o desvio padrão} \\ \bar{x} \text{ é a média} \end{cases}$$

Fórmula 6: Coeficiente de Variação

*O Coeficiente de Variação (CV) é uma medida expressa em **porcentagem**, por isso, está multiplicado por 100.*



Vamos realizar um exercício completo. Suponha que queiramos estudar a variação das idades de dois grupos,⁶⁵ abaixo relacionados:

G1: 7 7 7 7 7 7
G2: 8 9 10 11 19 22

Vamos calcular a média e o desvio padrão de **G1** e **G2**.

1) Cálculo da média: vamos utilizar a Fórmula 1: Média Aritmética, p. 72.

$$\bar{x} = \frac{\sum x_i}{n} \text{ Então,}$$

Para G1: $\bar{x} = \frac{\sum x_i}{n} = \frac{7+7+7+7+7+7}{6} = 7 \text{ anos}$

Para G2: $\bar{x} = \frac{\sum x_i}{n} = \frac{8+9+10+11+19+22}{6} = 13,1$ aproximadamente, 13 anos.

2) Cálculo do desvio padrão: Vamos utilizar a Fórmula 4: Desvio Padrão: Dados Não Agrupados, p. 95.

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

Então, antes do uso da Fórmula, como estamos fazendo sempre, vamos colocar em uma Tabela os dados que serão utilizados.

⁶⁵ PEREIRA (2004, p. 24).

G1		G2	
x_i	x_i^2	x_i	x_i^2
7	49	8	64
7	49	9	81
7	49	10	100
7	49	11	121
7	49	19	361
7	49	22	484
$\sum = 42$	$\sum = 294$	$\sum = 79$	$\sum = 1211$

Dessa forma,

Para G1:
 Sabendo que $\begin{cases} n = 6 \\ \sum x_i^2 = 294 \\ \sum x_i = 42 \end{cases}$ Então,

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{294}{6} - \left(\frac{42}{6}\right)^2} = \sqrt{49 - 49} = 0$$

Para G2:
 Sabendo que $\begin{cases} n = 6 \\ \sum x_i^2 = 1.211 \\ \sum x_i = 79 \end{cases}$ Então,

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{1.211}{6} - \left(\frac{79}{6}\right)^2} = \sqrt{201,8 - 173,3} = \sqrt{28,4} = 5,3$$

Aproximadamente, 5 anos.

Até aqui, podemos sintetizar da seguinte forma:

	G1	G2
$\bar{x}$	7	13
s	0	5

A média de idade de **G1** é de 7 anos e o desvio padrão é zero. Isso significa que, no conjunto, os valores das idades são *homogêneos* ou *sem variação*. Já em **G2**, a média das idades é de, aproximadamente, 13 anos e o desvio padrão de, aproximadamente, 5 anos. Essa variação no conjunto **G2**, pode ser medida. Para isso, vamos utilizar a Fórmula 6:

$$CV = \frac{s}{\bar{x}} \times 100 = \frac{5}{13} \times 100 = 38\%$$

Isso significa que podemos afirmar que **G2** é um grupo cujas idades variaram mais do que as idades de **G1**. E ainda, essa variação foi de 38%. Viu? A CV mede a variação.

O Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE) divulgou a seguinte informação sobre a taxa de desemprego, nas Regiões Metropolitanas e Distrito Federal, do país:

	1998	1999	2000	2001	2002	2003	2004	2004
Total	15,9	17,9	17,8	18,3	18,1	20,0	19,3	16,7

Fonte: DIEESE (2006).

Calcule a média, o desvio padrão e o coeficiente de variação da taxa de desemprego brasileira, a partir dos dados da Tabela acima.



Seção 4: Medidas de Posição

Onde se localiza o 20º elemento do grupo? Quais são as medidas que dividem o grupo em 4 partes iguais?



Respondendo a essas questões, estaremos encontrando a localização dos valores em um conjunto. Por essa razão, essas medidas são chamadas de **medidas de posição**, isto é, **indicam onde se localizam os pontos na série**.

Isso é muito útil. Por exemplo, digamos que, em uma escola, descobrimos que 25% dos alunos apresentam ausências constantes nas aulas de sexta-feira. Esse dado é significativo, pois, a partir dele, podemos criar estratégias para a correção do problema indesejado.

Pois bem, para afirmarmos essa ausência, localizamos um valor, a partir do qual sabemos o comportamento do conjunto acima e abaixo dele, essa é uma *medida de posição*.

As **medidas de posição** mais conhecidas são as de **tendência central**, isto é, são aquelas medidas que concentram valores em torno de si.⁶⁶

Outras medidas de posição, como os *quartis*, os *decis* e os *percentis*, embora sejam *medidas de posição*, possuem uma característica muito especial: separam os conjuntos em quantidades de iguais valores. Por isso, essas medidas podem ser chamadas de **separatrizes**.⁶⁷

Alguns estudiosos da estatística preferem chamar as *separatrizes* de medidas de posição e a *média*, a *mediana* e a *moda* (que também são medidas de posição), preferem chamar de *medidas de tendência central*. Os autores não concordam quanto a melhor maneira de considerá-las. Em nosso estudo, fizemos uma escolha. Optamos por chamar os *quartis*, os *decis* e os *percentis* de medidas de posição, mesmo sabendo que isso não agrada a todos.⁶⁸

Assim, nesta seção 4, estudaremos os *quartis*, os *decis* e os *percentis* que, a despeito de onde se encontram teoricamente, todos concordam com a forma de encontrá-los. E isso, no momento, é o que mais nos importa, não é mesmo?

Bom estudo para todos!

Quartis, Decis e Percentis



Quartis, Decis e Percentis são medidas de posição, isto é, semelhante às medidas de tendência central, eles nos indicam uma determinada localização em relação ao conjunto de dados sob estudo.

66 Ver Seção 2: Medidas de tendência central, p. 78.

67 Conforme Crespo (1995) prefere chamá-las.

68 Ver Seção 1: Introdução, p. 76.

Entretanto, eles separam o conjunto em 4 partes iguais (*quartis*), 10 partes iguais (*decis*) ou 100 partes iguais (*percentis*), ou seja, em partes que apresentam o mesmo número de valores. Por isso, alguns autores preferem chamar as medidas de posição *quartis*, *decis* e *percentis* de *separatrizes* (juntamente com a *mediana*).

Estudaremos essas três medidas, com especial dedicação aos *quartis*. Por isso, primeiro, veremos os *quartis* e depois, *decis* e *percentis* juntos.



"[...] Essas medidas – os quartis, os percentis e os decis – são, juntamente com a mediana, conhecidas pelo nome genérico de separatrizes."
(CRESPO, 1995, p. 101).

Quartis

Você se lembra que deixamos de calcular a **mediana** em conjuntos com **dados agrupados em intervalo de classe**?⁶⁹ Pois bem, chegou a hora de lidarmos com essa valiosa ferramenta.

Na verdade, estrategicamente, deixamos para calcular a mediana de conjuntos com essas características (dados agrupados com intervalos de classe) para esse momento, porque a mediana nada mais é do que uma particularidade no estudo dos **quartis**. Mas, vamos por partes.

Já sabemos que em um conjunto de dados ordenados, o valor médio que divide o conjunto em duas partes iguais é a **mediana**. Nessa mesma linha de raciocínio, podemos pensar em valores que dividem o conjunto em quatro partes iguais. Veja a Figura 30, abaixo:

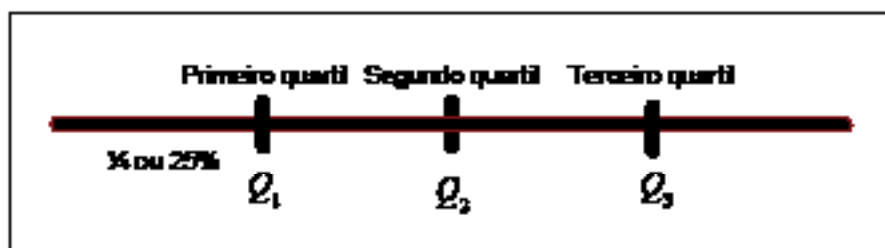


Figura 30: Quartis: Representação

Em um conjunto numérico, ocorre o mesmo que a figura acima: os *quartis* dividem o conjunto numérico em quatro partes iguais; Q_2 é o segundo *quartil* e divide o conjunto ao meio (por isso, é também a **mediana**); Q_1 divide a *metade do conjunto* em duas partes iguais, isto é, $\frac{1}{4}$ para cada lado; Q_3 é o terceiro *quartil*.



Quartis é o plural de quartil que significa $\frac{1}{4}$, isto é, um quarto.

⁶⁹ Ver Mediana e média, p. 85.

Para o cálculo dos *quartis* em conjuntos numéricos com **dados não agrupados**, basta aplicar a Fórmula, abaixo:



$$\frac{k \sum f_i}{4}, \text{ sendo } \begin{cases} k \text{ o número de ordem do quartil} \\ \sum f_i \text{ a soma total das frequências simples} \end{cases}$$

Fórmula 7: Medidas de Posição: Dados Não Agrupados: Quartil

Para **dados agrupados, com intervalos de classe**, utilizaremos outra Fórmula:



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



Onde,

k é o número de ordem do *quartil* (1, 2 ou 3);

l^* é o limite inferior da *classe mediana*;

$F(\text{ant})$ é a frequência acumulada da classe *anterior* à classe mediana;

f^* é a frequência simples da classe mediana;

h^* é a amplitude do intervalo da classe mediana.

Fórmula 8: Medidas de Posição: Quartil

Fórmulas podem até parecer assustadoras, às vezes, são mesmo. Mas não é o caso dessa última. Realizaremos um exercício, de modo prático, para mostrar *o que e como fazer* em casos como esse.

Vamos ao exercício:

Calcular o primeiro, o segundo e o terceiro *quartis* da distribuição de frequência abaixo:

Tabela 34: Distribuição de Frequência: Exercício: Quartis⁷⁰

Altura dos alunos da Turma A		
Estaturas (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
	$\Sigma = 40$	

Fonte: CRESPO (1995, p. 97)



Você notou que usamos um símbolo diferente? Bem, na verdade, é aquela mesma história de intervalo fechado e aberto.

Nesse caso, por exemplo, [150,154[indica que é um intervalo fechado em 150 e aberto em 154, isto é, trata-se de um intervalo de 150 até quase 154 (mas o 154 não entra).

Vamos resolver o problema em etapas.

1ª etapa: Construção da Tabela-Resposta. Começaremos a resolver o problema, construindo uma Tabela que nos ajudará em nossa tarefa.



Esta Tabela-Resposta será muito útil para nós. Não fique com dúvidas!

Tabela 35: Medidas de Posição: Quartis: Exercício: Tabela-Resposta

Tabela-Resposta: Quartis						
Quartil	$\frac{k \sum f_i}{4}$	l^*	$F(ant)$	h^*	f^*	Resultado
Q_1						
Q_2						
Q_3						

Essa Tabela-Resposta é uma preciosa ajuda para organizar os dados. Observe que nela constam todos os dados que serão

⁷⁰ Você se lembra que já trabalhamos com essa tabela? Ver **Tabela 32:** Desvio Padrão: Dados Agrupados: Com Intervalos de Classe: Exercício, p. 100.

utilizados pela Fórmula 8. A idéia é ir preenchendo-a, à medida que formos encontrando os valores.

2ª etapa: Posição ($\frac{k \sum f_i}{4}$). Os *quartis*, como sabemos, são valores que dividem os conjuntos em 4 partes iguais.⁷¹ O resultado encontrado com a ajuda da Fórmula 7: Medidas de Posição: Dados Não Agrupados: Quartil (p. 108), lamentavelmente, não nos fornece, de imediato, a *posição* do *quartil*, mas nos indica *em que linha de classe* ele se encontra. Vamos explicar isso melhor, mas antes, que tal encontrar a posição do primeiro, do segundo e do terceiro *quartis*?

Para isto, basta utilizarmos a Fórmula 7,⁷² vista anteriormente.

Como se pode notar, teremos três resultados, porque queremos encontrar a posição dos três *quartis*. Assim,



Solução

Sabemos que $\sum f_i = 40$ Então,

Primeiro quartil (k = 1) Segundo quartil (k = 2) Terceiro quartil (k = 3)

$$\frac{1 \times \sum f_i}{4} = \frac{40}{4} = 10 \quad \frac{2 \times \sum f_i}{4} = \frac{80}{4} = 20 \quad \frac{3 \times \sum f_i}{4} = \frac{120}{4} = 30$$

Agora volte à Tabela-Resposta e preencha a coluna " $\frac{k \sum f_i}{4}$ " com os resultados encontrados para cada *quartil*. Sua Tabela-Resposta ficará assim:

Tabela 36: Medidas de Posição: Quartis: Exercício:
Tabela-Resposta: Preenchimento: 2ª etapa

Tabela-Resposta: Quartis						
Quartil	$\frac{k \sum f_i}{4}$	l^*	$F(ant)$	h^*	f^*	Resultado
Q_1	10					
Q_2	20					
Q_3	30					

71 Ver **Figura 30:** Quartis: Representação, p. 107.

72 Ver p. 97.

Qual o significado, por exemplo, da posição 20 para Q_2 ?

O segundo *quartil*, sabemos, divide o conjunto em duas partes iguais. Não sabemos ainda, que valor é esse; mas o resultado 20 nos indica a linha (ou classe) em que ele se encontra. Vamos entender melhor isso. Veja a Figura 31, abaixo:

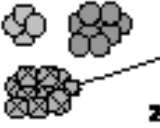
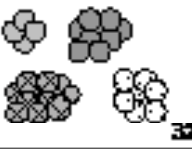
	Freqüência	Freqüência Acumulada
 [150,154]	 4	 4
 [154,158]	 9	 13
 [158,162]	 11	 24
 [162,166]	 8	 32
 [166,170]	 5	 37
 [170,174]	 3	 40

Figura 31: Tabela de Freqüência: Ilustração⁷³

A Figura 31, acima, representa a Tabela de Distribuição do nosso exercício, mas construída de maneira mais amigável. Vamos entendê-la por meio de uma metáfora: **a reunião das esferas**.

Em um planeta distante, os habitantes eram esferas. Existiam somente **6 tipos de esferas** com tamanhos (estaturas) que va-

⁷³ Agradeço ao amigo e professor de Estatística, Adolfo Dani, pela seguinte consideração: é preciso tomar cuidado para não pensar que todos os elementos do intervalo de classe tenham o mesmo tamanho, como as esferas parecem sugerir. Eu posso ter, por exemplo, no intervalo 150–154, alguns elementos com 150 cm, outros com 151 cm, outros com 152 cm e, portanto, eles podem não possuir a mesma altura. É verdade! Ainda assim, mantive a metáfora da “reunião das esferas”, pois, ela é feliz em seu objetivo central: mostrar a *posição* em uma distribuição com freqüência acumulada. Mas estamos atentos!



Assim, por exemplo:

- a esfera ● possuía estatura entre 166 cm e 170 cm.
- a esfera ○ possuía estatura entre 162 cm e 166 cm.



Intervalo de classe de 150 cm a 154 cm e frequência igual a 4.



Intervalo de classe de 154 cm a 158 cm e frequência igual a 9.

riavam, de acordo com a primeira coluna da figura 31, acima. Todas as esferas foram convidadas para uma reunião. Assim, as esferas foram chegando para o encontro por ordem de tamanho: primeiro, chegaram **4 esferas do tipo ○**; depois, **9 esferas do tipo ●**; a seguir, chegaram 11 esferas do tipo ●; assim, tipo por tipo, as esferas foram se reunindo até todas as 40 estarem presentes.

Pergunta-se: qual foi a esfera que chegou em 20º lugar?

Para responder a essa questão, basta analisarmos a terceira coluna (frequência acumulada). Repare que primeiro chegaram 4 esferas do tipo ○; depois chegaram mais 9 esferas do tipo ●. Até agora, portanto, chegaram 13 esferas, então, ainda não chegou a 20ª esfera.

Logo depois, chegaram 11 esferas do tipo ●. Como elas entraram todas juntas e rapidamente, ninguém se deu conta de que já haviam 24 esferas reunidas. Portanto, ninguém viu quem chegou em 20º lugar, mas todos sabiam que a esfera procurada já havia chegado, estava presente e só poderia ser do tipo ●.

Viu? Essa metáfora da reunião das esferas nos ensina que: em uma tabela de Distribuição de Frequência com dados agrupados em intervalos de classe, para localizarmos uma determinada posição, temos que primeiro encontrar a linha (ou a classe) onde ela se encontra.

Já fizemos um exercício semelhante quando estudamos *mediana*, você se lembra?⁷⁴ Dissemos que:

- 1) se o valor encontrado existir na linha da Frequência Acumulada (no nosso exercício esse valor é 20), então, esta será a classe *quartil* (a linha que estou procurando);
- 2) caso o valor não exista, a classe *quartil* será aquela que contiver a Frequência Acumulada, imediatamente, superior. No nosso caso, não existe a Frequência acumulada 20, portanto, a imediatamente superior é 24. Essa é a linha que estamos procurando.

Voltando agora, ao nosso exercício, sabemos que o segundo *quartil* se encontra na posição 20. Então, ele só pode estar na 3ª linha da Tabela de Distribuição de Frequência.

74 Ver Figura 25, p. 88.

Encontre as linhas em que se encontram o primeiro e o terceiro quartis.



Se você se concentrou na atividade, então, você conseguiu encontrar as linhas de classe dos *quartis*, conforme apresentado na Figura 32, abaixo:

Altura dos alunos da Turma A		
Estaturas (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
	$\Sigma = 40$	



Figura 32: Exercício: Quartis

3ª etapa: limite inferior da classe (l^*). Uma vez descobertas as classes do primeiro, segundo e terceiro *quartis*, essa etapa é rápida. Vamos destacar a linha de classe do primeiro *quartil*:

Tabela 37: Distribuição de Frequência: Exercício: Quartis: Primeiro Quartil

Altura dos alunos da Turma A		
Estaturas (cm)	f_i	F_i
[154,158[	9	13

Na linha de classe de Q_1 , as estaturas variam de 154 cm a 158 cm: o limite inferior (l^*), isto é, o menor valor é 154. Na linha de classe de Q_2 , o limite inferior da classe é 158. E para Q_3 , $l^* = 162$.

Pronto! Agora, vamos transportar os resultados para a Tabela-Resposta. Sua tabela ficará assim:

Tabela 38: Medidas de Posição: Quartis: Exercício:
Tabela-Resposta: Preenchimento: 3ª etapa

Tabela-Resposta: Quartis						
Quartil	$\frac{k \sum f_i}{4}$	l^*	$F(ant)$	h^*	f^*	Resultado
Q_1	10	154				
Q_2	20	158				
Q_3	30	162				

4ª etapa: Frequência Acumulada Anterior – $F(ant)$. Já sabemos que a frequência acumulada é a terceira coluna de nossa Tabela de Distribuição de Frequência. Para encontrar a $F(ant)$, uma vez determinada a linha de Q_1 , basta observarmos a frequência acumulada da **linha de cima**. Para Q_1 , a frequência acumulada anterior será 4. Veja:

Altura dos alunos da Turma A		
Estaturas (cm)	f_i	F_i
[150, 154[	4	4
[154, 158[	9	13

← Q_1

Frequência imediatamente anterior – $F(ant)$

Figura 33: Exercício: Quartis: Frequência Acumulada Anterior

Consultando nossa Tabela de Distribuição de Freqüência, o resultado é imediato: $Q_1 = 4$; $Q_2 = 13$ e $Q_3 = 24$. Vamos, agora, transportar os dados para nossa Tabela-Resposta:

Tabela 39: Medidas de Posição: Quartis: Exercício:
Tabela-Resposta: Preenchimento: 4ª etapa

Tabela-Resposta: Quartis						
Quartil	$\frac{k \sum f_i}{4}$	l^*	$F(ant)$	h^*	f^*	Resultado
Q_1	10	154	4			
Q_2	20	158	13			
Q_3	30	162	24			

5ª etapa: Amplitude do Intervalo (h^*). A determinação da **amplitude do intervalo de classe** também é imediata. Localizada a linha *quartil*, basta subtrair o *maior* valor do *menor* valor do intervalo de classe.

Desse modo, como Q_1 pertence à 2ª linha e o intervalo de classes é $[154, 158[$, a **amplitude do intervalo** será dada por $158 - 154 = 4$. Efetuando o mesmo cálculo para Q_2 e Q_3 encontraremos o mesmo resultado. Transportando esses resultados para a Tabela-Resposta, temos:

Tabela 40: Medidas de Posição: Quartis: Exercício:
Tabela-Resposta: Preenchimento: 5ª etapa

Tabela-Resposta: Quartis						
Quartil	$\frac{k \sum f_i}{4}$	l^*	$F(ant)$	h^*	f^*	Resultado
Q_1	10	154	4	4		
Q_2	20	158	13	4		
Q_3	30	162	24	4		

6ª etapa: frequência simples (f^*). Determinamos na 2ª etapa, a posição, isto é a linha de classe que os *quartis* ocupam na distribuição dos dados (chamamos esse linha de *classe quartil*). Consultando essa Tabela de Distribuição, basta identificarmos a frequência simples de cada classe *quartil*. Assim, teremos: 9, 11 e 8, respectivamente para Q_1 , Q_2 e Q_3 . Lançando na Tabela-Resposta, teremos:

Tabela 41: Medidas de Posição: Quartis: Exercício:
Tabela-Resposta: Preenchimento: 6ª etapa

Tabela-Resposta: Quartis						
Quartil	$\frac{k \sum f_i}{4}$	l^*	$F(ant)$	h^*	f^*	Resultado
Q_1	10	154	4	4	9	
Q_2	20	158	13	4	11	
Q_3	30	162	24	4	8	

7ª etapa. Resultado. Chegamos à última etapa. Passo a passo, fomos encontrando todos os dados que precisamos para a utilização da Tabela 35: Medidas de Posição: Quartis: Exercício: Tabela-Resposta, p. 98. Consultando a Tabela-Resposta, basta substituírmos os valores e pronto!

Vamos aos cálculos:

Solução:

Primeiro Quartil.

$$Q_1 = l^* + \frac{\left[\frac{k \sum f_i}{4} - F(ant) \right] h^*}{f^*} = 154 + \frac{[10 - 4] \times 4}{9} = 156,66$$

Segundo Quartil.

$$Q_2 = l^* + \frac{\left[\frac{k \sum f_i}{4} - F(ant) \right] h^*}{f^*} = 158 + \frac{[20 - 13] \times 4}{11} = 160,54$$

Terceiro Quartil.

$$Q_3 = l^* + \frac{\left[\frac{k \sum f_i}{4} - F(ant) \right] h^*}{f^*} = 162 + \frac{[30 - 24] \times 4}{8} = 165$$

Pronto, determinamos nossos *quartis*.

Calcule o primeiro, o segundo e o terceiro quartis da Distribuição de Freqüência, abaixo:



Tabela 42: Exercício: Quartis

Custos R\$	f_i	F_i
[450,550[	8	8
[550,650[	10	18
[650,750[	11	29
[750,850[	16	45
[850,950[	13	58
[950,1050[	5	63
[1050,1150[	1	64
$\Sigma = 64$		

Fonte: CRESPO (1995, p. 103)

Não deixe de preencher a Tabela-Resposta:

Tabela-Resposta						
Quartil	$\frac{k \sum f_i}{4}$	l^*	$F(ant)$	h^*	f^*	Resultado
Q_1						
Q_2						
Q_3						



Decis e Percentis

Decis e **Percentis** são encontrados de maneira análoga aos *quartis*. Se *quartis* dividem o conjunto de dados em 4 partes iguais; *decis* dividem o conjunto em 10 partes e *percentis* em 100 partes. Se podemos encontrar 3 *quartis* (Q_1 , Q_2 e Q_3), podemos encontrar 9 *decis* (D_1 , D_2 , D_3 , ..., D_9) e 99 *percentis* (P_1 , P_2 , P_3 , ..., P_9).

Para encontrar as posições dos *decis* e dos *percentis* utilizamos fórmulas semelhantes às da *mediana* e dos *quartis* para *dados não agrupados*. Veja:



ATENÇÃO

$$\frac{k \sum f_i}{10}$$

Fórmula 9: Medidas de Posição: Dados Não Agrupados: Decil



ATENÇÃO

$$\frac{k \sum f_i}{100}$$

Fórmula 10: Medidas de Posição: Dados Não Agrupados: Percentil

Um exemplo será o suficiente para mostrar que *quartis*, *decis* e *percentis* são calculados da mesma maneira. Vamos a ele:

Considerando a Tabela de Distribuição de Frequência utilizada no exercício de *quartis* (abaixo, reproduzida), calcule o *oitavo percentil*.

Altura dos alunos da Turma A		
Estaturas (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
$\Sigma = 40$		

Solução:

Obedecendo as etapas, construiremos a Tabela-Resposta, antes de mais nada.

Tabela 43: Medidas de Posição: Percentil: Tabela-Resposta

Tabela-Resposta: Percentis

Percentil	$\frac{k \sum f_i}{100}$	l^*	$F(ant)$	h^*	f^*	Resultado
P_8						

Observe duas mudanças na nossa Tabela-Resposta:

- 1) Aparece "Percentil", na primeira coluna (ao invés de "quartil");
- 2) Aparece " $\frac{k \sum f_i}{100}$ ", na segunda coluna (ao invés de " $\frac{k \sum f_i}{4}$ "). Isso se deve ao fato de querermos o *percentil* e não o *quartil*, como antes.



Quartil, vem de 1/4, por isso, divide-se por 4; percentil vem de 1/100, por isso divide-se por 100.

123

Da mesma forma, nossa Fórmula Geral será alterada:

$$P = l^* + \frac{\left[\frac{k \sum f_i}{100} - F(ant) \right] h^*}{f^*}$$



Fórmula 11: Medidas de Posição: Percentil

Comparando-se as Fórmulas do *Quartil* e do *Percentil*, temos que:

Comparação: Fórmula Geral: Quartil e Percentil	
Quartil	Percentil
$Q = l^* + \frac{\left[\frac{k \sum f_i}{4} - F(ant) \right] h^*}{f^*}$	$P = l^* + \frac{\left[\frac{k \sum f_i}{100} - F(ant) \right] h^*}{f^*}$

Quadro 4: Quartil e Percentil: Fórmula Geral: Comparação

Observe com a comparação acima que se trata apenas de uma adaptação, mas as fórmulas são as mesmas. Como já dissemos, são apenas duas alterações: de Q passou a P (isto é, de quartil passou a percentil) e de 4 passou para 100 (isto é, divisão do quartil – 4 – e divisão do percentil – 100).

Vamos então, encontrar a *classe percentil*:

$$\frac{k \sum f_i}{100} = \frac{8 \times 40}{100} = 3,2$$

Logo,

Como não existe na coluna de Frequência Acumulada o valor 3,2, o valor imediatamente acima dele é 4. Portanto, nosso percentil (P_8) encontra-se na 1ª linha (ou classe).

Preenchendo toda a Tabela-Resposta, encontramos:



Encontre os demais valores da Tabela-Resposta.

Tabela 44: Medidas de Posição: Percentis: Exercício:
Tabela-Resposta: Preenchida

Tabela-Resposta: Percentis						
Percentil	$\frac{k \sum f_i}{100}$	l^*	$F(ant)$	h^*	f^*	Resultado
P_8	3,2					

Após o preenchimento da Tabela-Resposta com os dados que estão faltando, efetuaremos o cálculo com a Fórmula 11:

$$P = l^* + \frac{\left[\frac{k \sum f_i}{100} - F(\text{ant}) \right] h^*}{f^*}$$

$$P_8 = 150 + \frac{[3,2 - 0] \times 4}{4} = 153,2$$

Encontramos, portanto, $P_8 = 153,2$. Significa que 8% possuem estatura inferior a 153,2%.

Viu? Tão simples quanto o cálculo do *quartil*, basta um pouco de disciplina e atenção.

Encontre o 1º e o 9º decis da Tabela de Distribuição de Frequência acima ("Altura dos Alunos da Turma A").



CONSIDERAÇÕES FINAIS

Ufa! Chegamos ao final.

Aqui, não poderia deixar de agradecer pela oportunidade que tive de apresentar aos colegas de profissão – Trabalhadores em Educação – minhas opiniões sobre a Estatística. Aproveito também para agradecer a todos os colegas que fizeram a leitura prévia do texto contribuindo, assim, para torná-lo melhor.

Ao longo de todo esse Módulo, fixei-me, principalmente, em um objetivo: desmistificar a matemática. Considerando que a Estatística é uma aplicação da matemática, procurei mostrar a vocês, colegas da educação, que, com certa disciplina, é possível fazer uso da Estatística, mesmo com alguma dificuldade na matemática. Por isso, após o chamamento para o estudo (na Introdução), demos a partida para a jornada, apresentando, brevemente, aqueles conceitos principais da matemática, sem os quais seria impossível a compreensão da Estatística.

Depois, mergulhamos na Estatística Descritiva, isso significa que passamos a olhar com atenção tabelas e gráficos tão presentes em nossas vidas. Nosso objetivo foi apresentar ao leitor metodologias de organização e exposição de dados como ferramenta para a leitura da realidade.

Com foco ainda na Estatística Descritiva, no momento seguinte, buscamos aprimorar a organização e exposição de dados a partir de modelos já consagrados pelo uso.

Depois, mudamos de foco. Passamos a manipular os dados, vale dizer: saímos da organização e exposição para a manipulação de dados. Nesse momento do estudo, procuramos organizar informações já manipuladas por todos nós, em nossas atividades profissionais, mas que mereciam atenção especial. A partir desse instante, adentramos ao mundo da Estatística Inferencial, pois já podemos propor soluções a alguns problemas que nos afligem há muito, em nosso trabalho.

Uma última palavra: se o leitor, de alguma forma, em qualquer nível ou intensidade, em poucos setores de atuação, em síntese, por menor que seja a contribuição desse estudo, se ele agregou qualidade a suas atividades profissionais, então, esse Módulo foi vitorioso.

Certo da importância da Formação Inicial na vida de todo profissional e, especialmente, na vida do profissional de Educação, parablenho a todas e a todos pelo esforço!

Muito Obrigado!

ARANHA, Carla. Ramsés II, o faraó guerreiro. *Aventuras na História: para viajar no tempo*, 11. ed. São Paulo: Editora Abril S.A., 2006. Disponível em: <http://historia.abril.com.br/edicoes/11/capa/conteudo_historia_42675.shtml>. Acesso em: 30 ago. 06.

ARENDT, Hanna. *A condição humana*. 10. ed. Tradução de Roberto Raposo, posfácio de Celso Lafer. Rio de Janeiro: Forense Universitária, 2005.

BRASIL. MEC. INEP. *Censo Escolar 2003*. Disponível em: <<http://www.inep.gov.br/basica/censo/Escolar/resultados.htm>>. Acesso em: 10 ago. 2006.

_____. *Censo Escolar 2004*. Disponível em: <<http://www.inep.gov.br/basica/censo/Escolar/resultados.htm>>. Acesso em: 10 ago. 2006.

_____. *Censo Escolar 2005*. Disponível em: <<http://www.inep.gov.br/basica/censo/Escolar/resultados.htm>>. Acesso em: 10 ago. 2006.

_____. *Resultados do Censo Escolar*. Disponível em: <<http://www.inep.gov.br/basica/censo/Escolar/resultados.htm>>. Acesso em: 10 ago. 2006.

_____. *NBR 14724: Informação e documentação: trabalhos acadêmicos: apresentação*. Rio de Janeiro, ago. 2002.

_____. *Emenda 1*. Rio de Janeiro, 30 dez. 2005.

_____. Ministério das Cidades. Departamento Nacional de Trânsito – Denatran. Coordenação-Geral de Informatização e Estatística – CGIE. *Anuário Estatístico de Acidentes de Trânsito 2002*. Disponível em: <<http://www.denatran.gov.br/acidentes.htm>>. Acesso em: 31 ago. 06.

CASTRO, Lauro Sodré Viveiros de. *Exercícios de Estatística*. 9. ed. Rio de Janeiro: Editora Científica, 1964.

CLEMENTE, Rosana Giovanni Pires. *Apostila de Estatística*, Taubaté. Universidade de Taubaté, 2003.

COSTA, Sérgio Francisco. *Estatística aplicada à pesquisa em educação*. Brasília: Plano Editora, 2004.

CRESPO, Antônio Arnot. *Estatística fácil*. 13. ed. São Paulo: Saraiva, 1995.

DANTE, Luiz Roberto. *Tudo é matemática*. 6ª série. São Paulo: Ática, 2003.

DEPARTAMENTO Intersindical de Estatística e Estudos Socio-econômicos (DIEESE). *Pesquisas e Análises do Dieese*. Disponível em: <<http://turandot.dieese.org.br/icv/TabelaPed?tabela=5>>. Acesso em: 28 ago. 06.

DESCARTES, René. *Discurso do método; As paixões da alma; Meditações; Objeções e respostas*. Introdução de Gilles-Gaston Granger. Prefácio e notas de Gerard Lebrun. Tradução J. Guinsburg e Bento Prado Júnior. 5. ed. São Paulo: Nova Cultural, 1991. (Os pensadores).

HERÓDOTO. Disponível em: <<http://greciantiga.org>>. Acesso em: 30 ago. 06. ISSN 91679-5709.

HISTÓRIA da Estatística. Disponível em: <<http://www.estatisticapr.hpg.ig.com.br/historia.html>>. Acesso em: 03 ago. 2006.

JAKUBOVIC, José; LELLIS, Marcelo. *Matemática na medida certa*, 6ª série. 3. ed. São Paulo: Scipione, 1995.

JOHNSTON, Ian. Section Four: The Beginnings of Modern Probability Theory. In: _____ *And Still We Evolve: A Handbook on the History of Modern Science*. Disponível em: <<http://www.mala.bc.ca/~johnstoi/darwin/sect4.htm>>. Acesso em: 04 ago. 2006.

OS IMPERADORES Yao e Yun. Disponível em: <<http://po.chinabroadcast.cn/chinaabc/chapter17/chapter170104.htm>>. Acesso em: 30 ago. 06.

PARENTE, Eduardo Afonso de Medeiros; CARIBÉ, Roberto. *Matemática comercial e financeira*: 145 exercícios resolvidos, 262 exercícios propostos com respostas, 99 testes de concursos e vestibulares. São Paulo: FTD, 1996.

PEREIRA, Paulo Henrique. *Noções de estatística*: com exercícios para administração e ciências humanas (dirigidos a pedagogia e turismo). Campinas: Papirus, 2004.

SEARS, Francis; ZEMANSKY, Mark W.; YOUNG, Hugh D. *Física 1*: mecânica da particular e dos corpos rígidos. 2. ed. rev. tradução de Jean Pierre von der Weid. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora Ltda., 1985.

SOARES, Tufi Machado. *Utilização da Teoria da Resposta ao Item na produção de Indicadores Sócio-econômicos*. Dispo-

nível em: <<http://www.scielo.br/pdf/pope/v25n1/24252.pdf>>. Acesso em: 13 ago. 07.

SPIEGEL, Murray Ralph. *Estatística*: resumo da teoria, 875 problemas resolvidos, 619 problemas propostos. Tradução de Pedro Cosentino. ed. rev. por Carlos José Pereira de Lucena. São Paulo: McGraw-Hill do Brasil, 1975.

U.S. CENSUS Bureau. Disponível em: <<http://www.census.gov>>. Acesso em: 04 ago. 2006.

FELIX NETO, José Luiz. *O despovoamento do território amazônico*. Disponível em: <http://www.abep.nepo.unicamp.br/encontro2006/docspdf/ABEP2006_434.pdf>. Acesso em: 12 ago. 2006.

Apêndice: Respostas dos exercícios Pratique!

Unidade II

(p. 23)

I é proporcional a II que é proporcional a III. Os três são proporcionais, pois, $\frac{45}{40} = \frac{30}{26,67} = \frac{15}{13,33} = 1,125$.

(p. 27)

Altura do cão (cm)	Altura da mulher (cm)
1	4
x	160

$$\frac{1}{x} = \frac{4}{160} \Rightarrow 4x = 160 \Rightarrow x = \frac{160}{4} \Rightarrow x = 40 \text{ cm.}$$

O cão mede 40 cm.

(p. 29)

População	%
169 872 856	100
30 940 542	x

$$\frac{169872856}{30940542} = \frac{100}{x} \Rightarrow 169872856x = 30940542 \times 100 \Rightarrow x = \frac{3094054200}{169872856} \Rightarrow x = 18,21\%$$

No Brasil, a população em idade escolar (dos 6 aos 14 anos), corresponde a 18,21% da população total.

(p. 32)

Coeficiente=0,007 / Taxa = 0,7% ou 7%. (Repare que o símbolo mudou. Significa que o denominador é 1000. Nesse caso, lemos: sete por mil).

(p. 35)

(Rua 0; Avenida 1) / (Rua 0; Avenida 2) / (Rua 1; Avenida 0) /
(Rua 1; Avenida 2) / (Rua 2; Avenida 0) / (Rua 2; Avenida 1) /
(Rua 3; Avenida 1) / (Rua 3; Avenida 2)

(p. 37)

- 1) a. 2,4 / b. 24,6 / c. 0,4 / d. 4,2 / e. 328,4 / f. 3,0 / g. 6,8 / h. 5,6 / i. 90,0
- 2) a. 46,73 / b. 123,84 / c. 253,65 / d. 299,95 / e. 28,26 / f. 37,48

Unidade III

(p. 41)

Escolas	População	10%	Amostra
D	M = 134	$\frac{10 \times 134}{100} = 13,4$	13
	F = 228	$\frac{10 \times 228}{100} = 22,8$	23
E	M = 150	$\frac{10 \times 150}{100} = 15$	15
	F = 130	$\frac{10 \times 130}{100} = 13$	13
F	M = 300	$\frac{10 \times 300}{100} = 30$	30
	F = 290	$\frac{10 \times 290}{100} = 29$	29

(p. 42)

Universo	Variável
As jogadas de um dado.	O ponto obtido em cada jogada – Variável quantitativa discreta.
Peças produzidas por certa máquina.	Número de peças produzidas por hora – Variável quantitativa discreta.
Peças produzidas por certa máquina.	Diâmetro externo – Variável quantitativa contínua.

(p. 45)

1/3/2/2/3/1/3/3/2/2/3

(p. 48)

Cabeçalho: Unidade da Federação / Matrículas no Ensino Fundamental de 5ª a 8ª série, Diurno, Total, Federal, Estadual, Municipal e Privada.

Linha: Brasil / 13.629.874 / 18.183 / 7.386.348 / 4.664.840 / 1.560.503.

Casa ou célula: cinco casas: 13.629.874 / 18.183 / 7.386.348 / 4.664.840 / 1.560.503.

Coluna indicadora: Unidade da Federação / Brasil.

Coluna numérica: são cinco: 1ª Total - 13.629.874 / 2ª Federal - 18.183 / 3ª Estadual - 7.386.348 / 4ª Municipal - 4.664.840 / 5ª Privada - 1.560.503.

(p. 55)

Esfera municipal=81,27%

Esfera privada=0,45%

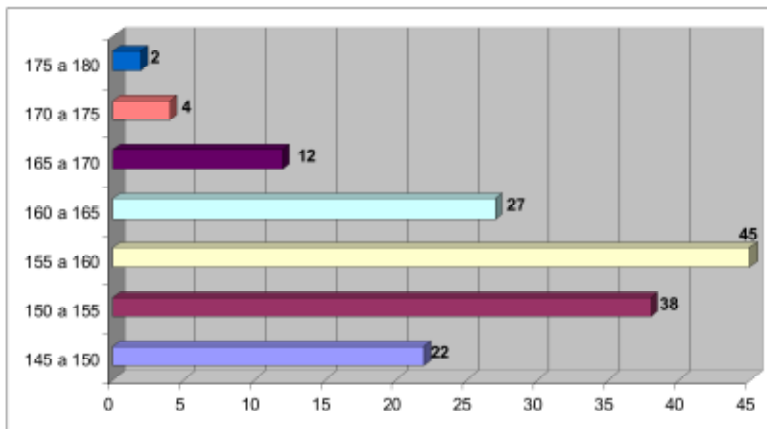
Unidade IV

(p. 73)

1)

Estatura: 150 alunos	
Estatura (cm)	f_i
145 a 150	22
150 a 155	38
155 a 160	45
160 a 165	27
165 a 170	12
170 a 175	4
175 a 180	2
Total 150	

2)



Unidade V

(p. 84)

$$\bar{x} = 7546,25$$

(p. 92)

Conjunto A: média = 8,9 / mediana = 9 / moda = 7

Conjunto B: média = 6,4 / mediana = 6 / moda = 4, 5, 6, 8 e 10 (5 modas; polimodal).

(p. 98)

Conjunto A: 4,87

Conjunto B: 3,87

Note que o Conjunto B apresenta dispersão menor que o Conjunto A

(p. 101)

$s = R\$ 154,00$

(p. 105)

$\bar{x} = 18\%$; $s = 1,22$; $CV = 6,78\%$

(p. 117)

$Q_1 = 630$, $Q_2 = 768$ e $Q_3 = 873$. Significa que 25% do custo varia de R\$ 450,00 a próximo de R\$ 630,00; 50% é menor que R\$ 768,00 e 75%, menor que R\$ 873,00.

(p. 120)

Tabela-Resposta: Percentis

Percentil	$\frac{k \sum f_i}{100}$	l^*	$F(ant)$	h^*	f^*	Resultado
P_8	3,2	150	0	4	4	

(p. 121)

$D_1 = 154$ e $D_9 = 169,2$. Significa que 10% possuem estatura inferior a 154 cm e 90%, inferior a 169,2 cm. Ou ainda, apenas 10% possuem altura superior a 169,2 cm.