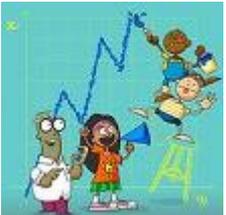


Aula 4

- Leia o material didático com muita atenção;
- Use uma calculadora e refaça os cálculos de todos os exemplos;
- Sempre que tiver dúvidas recorra ao tutor a distância.

Introdução à Estatística

<p>Sobre o IBGE</p> <p>O que é Estatística?</p> 	<p>Em nosso cotidiano, usamos a estatística para sabermos os índices de inflação ou de emprego e desemprego, por exemplo. Mas você sabe o que é a Estatística?</p> <p>A Estatística é uma ciência que cuida da coleta de dados, que são organizados, estudados e então utilizados para um determinado objetivo. No caso do IBGE, a estatística é importante para informar sobre a realidade do Brasil através de números.</p>
---	---

Fonte: http://www.ibge.gov.br/7a12/sobre_ibge/default.php?id_tema_menu=5

A ciência não se encarrega apenas de descrever fatos, enunciar leis e apresentar novas descobertas, mas principalmente de ensinar o método científico, o qual exige organização dos dados, análise e tomada de decisões em condições de incerteza. A maioria dos fenômenos que estudamos são dinâmicos, pois nada é definitivo, todos os estudos que realizamos são considerados verdadeiros até sejam questionados ou corrigidos.

A estatística é a ciência que se ocupa de coletar, organizar, analisar e interpretar dados a fim de tomar decisões. Sua grande importância reside no fato de ter um vasto campo de aplicação, seja na administração, educação, medicina, engenharia, política, entre outros.

Todos os dias somos bombardeados com dados estatísticos, seja pela televisão, jornais, rádio ou Internet. Basta à divulgação de pesquisas sobre taxas de desemprego, índice de popularidade de governantes, índice de audiência de programas de televisão para que algumas pessoas já contestem sua validade. É claro que pesquisas feitas de forma incorreta, assim como dados mal coletados, podem levar a decisões equivocadas. Quando um estudo estatístico é feito de forma criteriosa, seus resultados permitem obter conclusões confiáveis.

No passado, o grande obstáculo para o uso da Estatística eram os cálculos que exigiam horas de trabalho cansativo. Com o avanço da tecnologia este obstáculo praticamente desapareceu. Só que não basta ter um computador ou uma calculadora que elabore gráficos, tabelas e realize cálculos, é necessário que tenhamos conhecimento para compreendê-los, bem como, capacidade para construí-los. Portanto, é importante o conhecimento de conceitos básicos de estatística para saber utilizar de forma adequada as diferentes técnicas estatísticas.

Campos da Estatística

A estatística está dividida em três campos: Estatística Descritiva, Probabilidade e Estatística Inferencial.

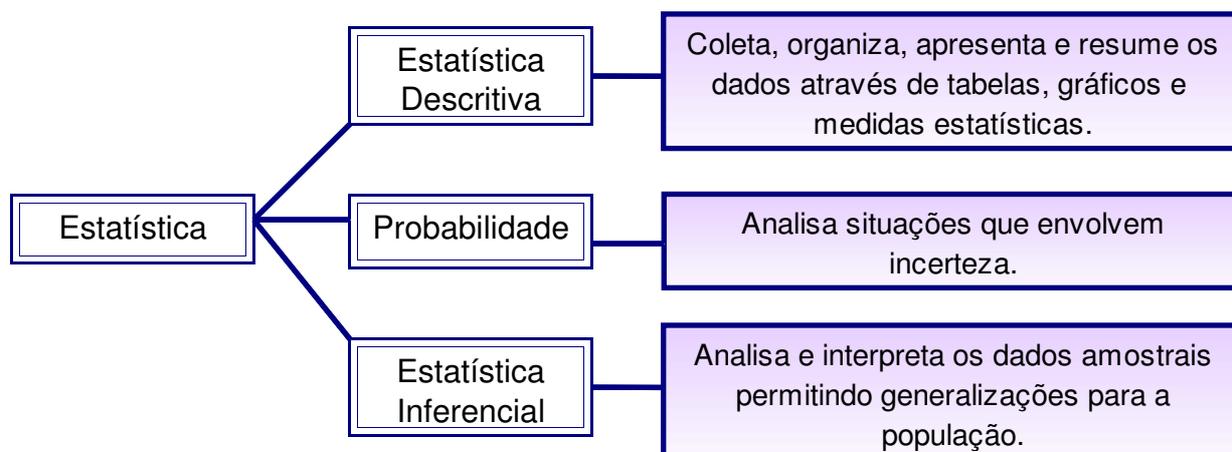
A **Estatística Descritiva** é a parte da Estatística que tem por objetivo a coleta, a organização, a apresentação dos dados em tabelas ou gráficos, assim como a obtenção de algumas medidas como média, proporção, desvio padrão que auxiliam na descrição de um conjunto de dados.

Embora a estatística descritiva seja amplamente usada em várias áreas do conhecimento ela não vai além da informação fornecida pelos dados, nenhuma inferência é feita além dos próprios dados. Por exemplo, se cinco alunos de uma turma obtêm as seguintes notas numa avaliação: 8, 9, 7, 7, 9. É possível, através da estatística descritiva, resumir esta informação calculando a média das notas destes cinco alunos, a média é 8, mas não podemos concluir que a média de toda a turma é 8.

A **Probabilidade** é uma ferramenta importante para a análise de qualquer situação que envolva incerteza. Como as amostras são formadas por apenas alguns elementos da população, apresentam certa variabilidade, o que gera incertezas com relação às conclusões que podemos chegar sobre as características de uma população. As afirmações estatísticas são sempre afirmações sobre a probabilidade de ocorrência de certo fenômeno aleatório, a partir de um conjunto de condições teóricas satisfeitas. Dessa forma, a probabilidade é muito importante na estatística inferencial.

A **Estatística Inferencial** é um conjunto de técnicas que obtém informações sobre uma população a partir de resultados observados numa amostra, permitindo conclusões e generalizações que vão além dos dados coletados.

Resumindo...



População e Amostra

Num estudo estatístico não interessa analisar unidades individuais, mas sim grupos ou conjuntos, porque seu objetivo é o estudo de toda a população. Em estatística, **população** é definida como o conjunto de indivíduos, objetos ou produtos que apresentam, pelo menos, uma característica em comum. Assim, nossa população pode ser tanto toda a população de uma cidade, como todas as peças produzidas por uma fábrica. Uma população pode ser pequena, grande ou infinita. Sendo o número de observações em uma população finita designada pela letra **N**.

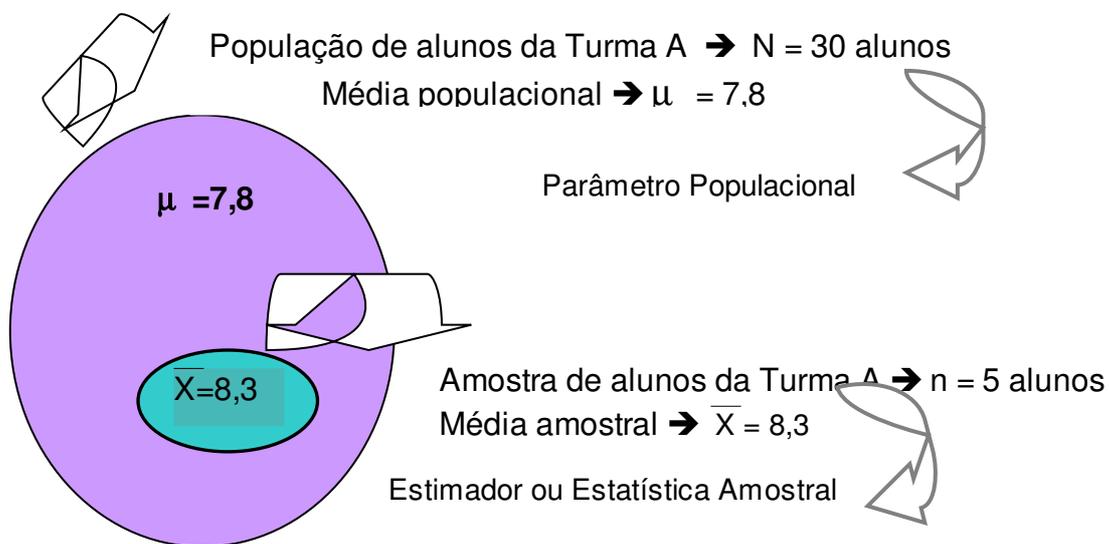
Apesar de algumas populações terem um número finito de elementos muitas vezes é inviável fazer uma análise de todos eles, seja pela distribuição geográfica, pelo custo elevado ou pelo tempo que levaria para realizar o estudo. Por isso, geralmente os pesquisadores trabalham com amostras que consistem de uma parte da população selecionada para o estudo.

Neste caso, toma-se uma parte da população escolhida para o estudo, chamada **amostra**, com **n** elementos, com as mesmas características da população. O objetivo de selecionar uma amostra é de obter conclusões que possam ser generalizadas para a população, isto é, possam ser inferidas.

Exemplo: Uma pesquisa pretende estudar a opinião de estudantes com relação à mudança do sistema de avaliação em uma universidade. Neste caso, a população é formada por todos os estudantes matriculados nesta instituição. Para compor a amostra os pesquisadores entrevistaram 200 estudantes.

Denominamos **parâmetro** uma descrição numérica de uma característica da população e **estatística amostral ou estimador** uma descrição numérica de uma característica da amostra, que será usada no processo de estimação de um parâmetro populacional.

Exemplo: Um professor calcula a média das notas obtidas entre os 30 alunos da Turma A. Se considerarmos a Turma A como a população a ser considerada a média obtida é um parâmetro populacional. No entanto, se o professor seleciona 5 alunos entre os 30 da turma e calcula a média das notas destes 5 alunos, a média obtida é considerada uma estatística amostral ou um estimador.



Como, em geral, os parâmetros são desconhecidos, pois dificilmente é possível fazer uma análise de toda a população, as estatísticas amostrais ou estimadores são usados para se ter uma idéia do valor dos parâmetros populacionais. Antes da coleta dos dados é preciso definir o método de amostragem mais adequado para escolher os elementos da população que irão compor a amostra.

Amostragem

Muitas vezes, dadas as limitações impostas pelos custos da pesquisa ou urgência dos resultados, o pesquisador precisa restringir as suas observações a uma amostra da população a ser estudada. É muito importante que os elementos que irão compor essa amostra sejam coletados de forma adequada. Quando um estudo é feito com dados imprecisos os resultados são questionáveis.

Existem dois tipos de métodos de amostragem: os **métodos não probabilísticos** e os **métodos probabilísticos**. A Estatística só lida com esses últimos, pois nos métodos não probabilísticos não podemos determinar a probabilidade que cada elemento da população tem de pertencer à amostra, o que impossibilita as generalizações. Nos métodos não probabilísticos a escolha dos indivíduos ou elementos que compõem a amostra depende em parte do julgamento do pesquisador ou do entrevistador. Por exemplo, o professor de um curso de treinamento precisa selecionar um aluno para participar de um teste a fim de verificar se o método utilizado no treinamento é adequado. Ao designá-lo o professor leva em consideração, mesmo que inconscientemente, as diversas características de cada aluno, logo nem todos os alunos que compõem a população terão a mesma probabilidade de pertencer a amostra.

Já os métodos probabilísticos exigem que cada elemento ou indivíduo da população possua uma probabilidade conhecida (não nula) de ser selecionado para compor a amostra. Neste caso, os resultados obtidos podem ser generalizados para a população. Retomando o exemplo anterior, supondo que no curso de treinamento tenham 8 alunos, cada aluno do curso terá uma probabilidade igual a $1/8$ de ser selecionado para realizar o teste.

Para a escolha do método de amostragem, o pesquisador deve levar em conta:

- o tipo de pesquisa;
- a acessibilidade e disponibilidade dos elementos da população;
- a representatividade necessária para atingir os objetivos da pesquisa;
- a disponibilidade de tempo para sua realização;
- os recursos financeiros e humanos disponíveis entre outros fatores.

Lembre-se, não existe técnica estatística capaz de salvar uma amostra mal coletada e nenhuma inferência estatística poderá ser realizada com base nesta amostra.

Dentre **os métodos de amostragem probabilísticos** podemos citar a amostragem aleatória simples, amostragem sistemática, amostragem estratificada e amostragem por agrupamento ou conglomerados.

Amostragem Aleatória Simples

A **amostragem aleatória simples** (AAS) é a maneira mais fácil de selecionarmos uma amostra. Na AAS todos os elementos da população têm a mesma probabilidade de serem selecionados para compor a amostra. Eles podem ser selecionados por sorteio ou através de

uma tabela de números aleatórios. Atualmente, é possível tomar uma amostra com números aleatórios gerados por meio de calculadoras ou computadores.

Exemplo 1: O professor deseja selecionar cinco alunos para apresentar o trabalho realizado. Escreve o nome dos 30 alunos em pedaços de papel e seleciona os alunos por sorteio.

Exemplo 2: O dono de uma locadora de filmes deseja fazer uma pesquisa de satisfação entre seus clientes. A amostra será composta por 50 clientes. Como a locadora tem 3.000 sócios fica inviável colocar o nome de cada um num pedaço de papel e realizar o sorteio. Como os sócios estão organizados por números, o proprietário da locadora utiliza um computador e gera 50 números aleatoriamente de 1 a 3000. Os clientes com o número gerado pelo computador irão compor a amostra.

Amostragem Sistemática

A **amostragem sistemática** é utilizada quando a população apresenta-se organizada segundo algum critério, de modo tal que cada um de seus elementos possa ser unicamente identificado pela posição (p.ex. fichas cadastrais, lista telefônica,...). Neste caso, a amostragem é realizada por intervalos fixos. Seleciona-se, aleatoriamente, o primeiro elemento que deve estar entre 1 e o fator de sistematização (s), depois escolhe-se os membros da amostra a intervalos regulares. O fator de sistematização é obtido pela divisão do número

de elementos da população (N) pelo número de elementos da amostra (n), $s = \frac{N}{n}$ onde s é arredondado para o número inteiro mais próximo.

Exemplo: Se uma amostra sistemática com 30 elementos for selecionada de uma população de 600 indivíduos, o fator de sistematização será de $600/30 = 20$ intervalos. Um número entre 1 e 20 (s) será sorteado aleatoriamente entre os primeiros indivíduos da população. Supondo que tenha sido sorteado o número 7. O sétimo elemento será o primeiro elemento da amostra, as seleções subseqüentes seriam 27, 47, 67, 87, ... , 567 e 587.

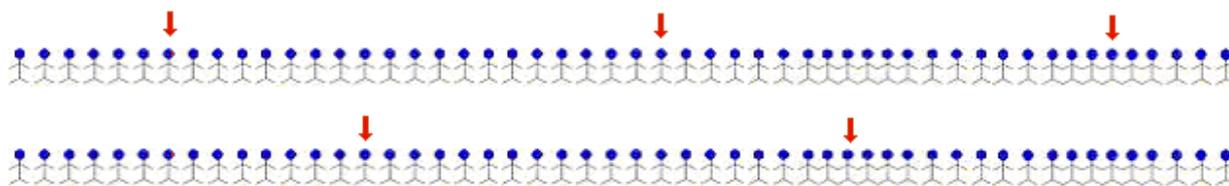


Figura 4.1 - Ilustração parcial do exemplo de amostragem sistemática

Como o primeiro elemento selecionado é uma escolha aleatória, considera-se que uma amostra sistemática usualmente tem as mesmas propriedades da amostragem aleatória simples. Esta hipótese é aplicável quando a lista da população é uma disposição aleatória da população.

Amostragem Estratificada

A **amostragem estratificada** é indicada quando a população encontra-se dividida em grupos distintos (população heterogênea). Dependendo dos objetivos do estudo a população será dividida em dois ou mais subgrupos, denominados estratos, que compartilham uma característica comum, como sexo, grau de instrução ou classe social. Depois que uma população é dividida em estratos apropriados, podemos fazer uma amostra aleatória simples em cada estrato. O número de elementos de cada estrato que constituirão a amostra é calculado com base em duas informações: (1) o tamanho que deve ser a amostra total e (2) como a amostra total deve ser alocada entre os estratos. As amostras dentro de cada estrato podem ser proporcionais ou desproporcionais ao tamanho do estrato em relação à população.

Exemplo: Uma escola é formada por 2.000 indivíduos entre professores, estudantes e funcionários. Na tabela 4.1 é apresentado o número de indivíduos em cada um destes estratos, proporcional ao seu número na população, considerando uma amostra com 5% dos elementos da população.

Tabela 4.1 – Amostragem estratificada

Estratos	População	Amostra
Professores	200	10
Funcionários	300	15
Estudantes	1.500	75
Total	2.000	100

Amostragem por Agrupamento ou Conglomerado

Na **amostragem por agrupamento** ou **conglomerado** os elementos da população são divididos em grupos, de forma que cada grupo seja representativo da população total. Uma amostra aleatória simples dos grupos é então obtida, e todos os elementos dentro de cada grupo são analisados. Podemos citar como agrupamentos agências, quarteirões, edifícios ou bairros. Quando a população estiver dispersa por uma extensa área geográfica, a amostragem por agrupamento possibilita obter um tamanho de amostra maior com um custo significativamente mais baixo.

Exemplo: Uma pesquisa foi aplicada nas escolas públicas (conglomerados) de um município com o objetivo de estudar a desnutrição dos estudantes. Para tal, algumas escolas foram sorteadas aleatoriamente e todos os estudantes de cada escola participaram da pesquisa.

Quadro Resumo dos Métodos de Amostragem Probabilística

Tipo	Descrição
Amostragem Aleatória Simples	A seleção pode ser feita por uma lista aleatória de elementos ou por sorteio.
Amostragem Sistemática	População organizada sob algum critério. O primeiro valor é escolhido aleatoriamente e depois a amostragem é realizada por intervalos fixos.
Amostragem Estratificada	A população é dividida em estratos homogêneos e amostras aleatórias são selecionadas de cada estrato.
Amostragem por Agrupamento	A população é dividida em seções ou grupos e uma amostra aleatória dos grupos é obtida. Todos os elementos de cada grupo são analisados.

Variáveis e Dados Estatísticos

Definimos como **variável** a característica que cada indivíduo, objeto ou produto assume (p. ex., sexo, idade, peso) e **dados estatísticos** os diferentes valores ou categorias que cada variável pode assumir (p. ex., masculino e feminino). As variáveis podem ser **qualitativas** ou **quantitativas**.

Como as variáveis têm naturezas diferentes quanto aos possíveis valores que podem assumir, a escolha do processo a ser utilizado na análise ou descrição de dados estatísticos depende do tipo de variável considerada. Por exemplo, numa pesquisa de opinião em que é perguntado ao usuário se ele está satisfeito com o serviço recebido, se as opções de resposta forem: muito bom, bom, regular ou ruim não poderemos calcular a média, pois esta variável não é numérica e qualquer técnica estatística baseada na média não poderá ser utilizada. Portanto, conhecer os tipos de variáveis é muito importante em qualquer estudo estatístico.

Variável Qualitativa ou Categórica

As **variáveis qualitativas ou categóricas** podem ser separadas em diferentes categorias que se distinguem por alguma característica não-numérica, podendo ser do tipo nominal ou ordinal.

Em uma **variável qualitativa nominal** os dados observados são meramente classificados em várias categorias distintas, usando-se nomes, marcas ou qualidades. Nenhum cálculo matemático pode ser feito com este tipo de variável.

Exemplo 1: A variável “sexo” assume duas diferentes categorias: masculino e feminino.

Exemplo 2: A variável “disciplinas cursadas no semestre” assume diferentes categorias: Matemática, Língua Portuguesa, Ciências, História, ...

Uma **variável qualitativa ordinal** envolve dados que podem ser classificados em categorias distintas, podendo ser arranjados em ordem.

Exemplo: A variável notas assume categorias que apresentam ordem: ótimo, muito bom, bom, regular, insuficiente, muito insuficiente.

Variável Quantitativa ou Numérica

As **variáveis quantitativas ou numéricas** consistem em números que representam contagens ou medidas. Podem ser do tipo discreta ou contínua.

Em uma **variável quantitativa discreta** os dados observados assumem apenas valores inteiros (contagens).

Exemplo: Número de alunos em sala de aula.

Em uma **variável quantitativa contínua** os dados observados podem assumir qualquer valor num certo intervalo (medições).

Exemplo: A variável altura dos alunos assume valores dentro de um intervalo que pode variar de acordo com a precisão do instrumento de medida usado: 1,7 m; 1,75 m; 1,752 m, ...

Quadro Resumo dos Tipos de Variáveis

Variável	Tipo	Exemplos
Qualitativa	Nominal	Sexo: masculino e feminino Disciplina: matemática, português, história, ...
	Ordinal	Notas: A, B, C, D, E Opinião do usuário: bom, regular ou ruim
Quantitativa	Discreta	Número de peças produzidas, número de alunos, número de defeitos.
	Contínua	Altura de pessoas, temperatura do ar, peso.

Planejamento de um estudo estatístico

O objetivo de um estudo estatístico é coletar dados e então usá-los na tomada de decisões. Se o processo for planejado de forma inadequada, as conclusões obtidas deste estudo não serão confiáveis.

Um bom planejamento fornece respostas para questões como: Quais são as variáveis de interesse? Que tipo de método de amostragem será usado? Que técnicas estatísticas serão usadas para analisar os dados? Qual o tempo e recursos disponíveis para realizar a pesquisa? Desta forma, evita-se surpresas indesejáveis ao final da pesquisa por dados coletados

incorretamente, unidades de medidas não adequadas às técnicas estatísticas de interesse entre outras.

Um bom estudo estatístico depende da correta execução desde a coleta, análise e interpretação dos dados que devem seguir as seguintes etapas:

- Defina com clareza os objetivos da pesquisa, ou seja, o que se pretende estudar;
- Identifique a(s) variável(is) de interesse e a população que for objeto de estudo;
- Elabore um plano para a coleta de dados. Se for usar uma amostra, assegure-se de que ela é representativa da população;
- Colete os dados;
- Organize e apresente os dados em tabelas ou gráficos que facilitem sua visualização. Resuma os dados através de medidas estatísticas como média, desvio padrão, coeficiente de variação, etc;
- Analise e interprete os dados;
- Elabore um relatório de tal forma que seja facilmente compreendido por quem for fazer uso dos resultados desta pesquisa na tomada de decisões.

Muitas pessoas jamais precisarão desenvolver um estudo estatístico, mas, provavelmente, terão que interpretar os resultados de algum desses estudos, seja no exercício de sua cidadania, no trabalho ou na sua vida pessoal. Por isso, é importante que todos estejam familiarizados com as etapas do planejamento de um estudo estatístico, para ter condições de analisá-los com criticidade.

Apresentação e organização dos dados

A apresentação e a organização dos dados de um estudo estatístico pode ser feita por tabelas e gráficos.

Uma **tabela** é um meio bastante eficiente de apresentar um conjunto de dados, facilitando sua compreensão e interpretação. Além disso, auxilia o entendimento global e o relacionamento entre as variáveis representadas.

Uma tabela deve apresentar: título, cabeçalho, corpo e fonte.

TÍTULO

Cabeçalho →	Variável	Freqüência
Corpo →		
	Fonte	

- O título da tabela dá informação sobre o local, data e tipo de variável que está sendo tratada;
- No cabeçalho coloca-se uma variável do tipo: tempo, local ou espécie, com a quantidade ou freqüência observada;
- No corpo apresentam-se os dados;
- A fonte especifica quem coletou originalmente os dados.

A função do **gráfico** na descrição e apresentação dos dados é a de auxiliar na compreensão dos dados. Para tal, são utilizados diversos tipos de gráficos como gráfico de linhas, setores, barras verticais ou horizontais. A escolha do tipo de gráfico depende dos dados a serem apresentados. Os gráficos têm a vantagem de facilitar a interpretação rápida das principais características de um conjunto de dados.

Apresentação tabular de dados qualitativos ou categóricos

Um conjunto de dados estatísticos dispostos em tabelas em função do tempo, do local e da espécie é denominado **série estatística**. Estas podem ser classificadas em específica, temporal, geográfica dependendo do tipo de variável que ela apresenta. As séries mistas ou tabelas de dupla entrada são séries que apresentam mais de uma variável.

A **série específica** é constituída por dados obtidos nas diferentes espécies ou categorias de uma mesma variável.

Tabela 7.1 – Série estatística específica ou categórica
As Cinco Áreas do Conhecimento com o maior Número de Cursos Participantes no ENADE/2005

Área	Frequência (f_i)
Pedagogia	891
Letras	714
Computação	685
Matemática	457
Biologia	422
Total	3.169

Fonte: MEC/INEP/DEAES - ENADE 2006

A **Série temporal** apresenta as medidas de uma variável tomadas em intervalos regulares obtidos ao longo do tempo.

Tabela 7.2 – Série temporal

Evolução do Número de Vagas no Ensino Superior no Brasil 1970 - 1995

Ano	Número de Vagas
1970	145.000
1975	348.227
1980	404.814
1985	430.482
1990	502.784
1995	610.375

FONTE: MEC/INEP/SEEC

A **Série geográfica** apresenta dados provenientes de diferentes regiões geográficas ou locais.

Tabela 7.3 – Série Geográfica

Número de Cursos Participantes Segundo Região Geográfica - ENADE/2005

Região	Número de Cursos
Norte	340
Nordeste	944
Sudeste	2.528
Sul	1.133
Centro-Oeste	566
Total	5.511

Fonte: MEC/INEP/DEAES - ENADE 2005

A **série mista** apresenta duas variáveis onde os dados podem referir-se a locais, espécies ou tempos diferentes. Por apontar mais de uma característica dos dados ao mesmo tempo, a série mista exige sempre mais de uma coluna.

Tabela 7.4 – Série Mista

Número de Cursos Avaliados Segundo a Área - ENADE 2005			
Área	Categoria Administrativa		
	Pública	Privada	Total
Biologia	157	265	422
Física	86	41	127
Geografia	134	147	281
História	150	221	371
Letras	235	479	714
Matemática	194	263	457
Pedagogia	257	634	891
Química	102	86	188
Total	1.315	2.136	3.451

Fonte: MEC/INEP/DEAES - ENADE 2005

Tabela 7.5 – Série Mista

Média na Prova Segundo as Áreas Avaliadas - ENADE/2005		
Área	Média	
	Formação Geral	Componente Específico
Biologia	58,7	24
Física	54,2	25,4
Geografia	59,6	32,1
História	53,4	33,4
Letras	56,9	28,6
Matemática	54	24,4
Pedagogia	49,2	41,6
Química	52,4	24,5

Fonte: MEC/INEP/DEAES - ENADE 2005

Freqüência relativa e freqüência percentual

Além da freqüência simples ou quantidade de observações, a série estatística também pode apresentar a freqüência relativa ou a freqüência percentual de cada categoria.

A **freqüência relativa** de determinado valor é a razão entre a freqüência simples (f_i) e a freqüência total ($\sum f_i$).

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

Tabela 7.6 – Série específica com freqüência relativa e percentual
As Cinco Áreas do Conhecimento com o maior Número de Cursos
Participantes no ENADE/2005

Área	Freqüência simples (f_i)	Freqüência relativa (f_r)
Pedagogia	891	$891/3.169 = 0,28$
Letras	714	$714/3.169 = 0,23$
Computação	685	$685/3.169 = 0,22$
Matemática	457	$457/3.169 = 0,14$
Biologia	422	$422/3.169 = 0,13$
Total	3.169	1,00

Fonte: MEC/INEP/DEAES - ENADE 2005

Observações:

1) $\sum f_i = 891+714+685+457+422 = 3.169$

2) A soma das freqüências relativas é sempre igual a um (1): $\sum fr_i = 1$

3) O uso da freqüência relativa facilita as comparações entre mais de um conjunto de dados com diferentes número de elementos, o que com a freqüência simples seria mais difícil.

Para obter a **freqüência percentual** basta multiplicar este valor por 100.

$$f_p = f_r \times 100$$

Logo, para o curso de Matemática a freqüência percentual será:

$$f_p = 0,14 \times 100 = 14\%.$$

Tabela 7.6 – Série específica com freqüência relativa e percentual
As Cinco Áreas do Conhecimento com o maior Número de Cursos
Participantes no ENADE/2005

Área	Freqüência simples (f_i)	Freqüência relativa (f_r)	Freqüência percentual (f_p)
Pedagogia	891	0,28	28%
Letras	714	0,23	23%
Computação	685	0,22	22%
Matemática	457	0,14	14%
Biologia	422	0,13	13%
Total	3.169	1,0	100%

Fonte: MEC/INEP/DEAES - ENADE 2005

Apresentação gráfica de dados qualitativos ou categóricos

Em alguns casos as informações são mais facilmente compreendidas quando são apresentadas na forma de gráficos. Os gráficos da Figura 7.1 apresentam os dados da tabela 7.1. O **gráfico de colunas** da Figura 7.1(a) apresenta retângulos dispostos verticalmente. Os retângulos têm a mesma base e as alturas representam o número ou a freqüência observada. O mesmo conjunto de dados é apresentado na Figura 7.1(b), num **gráfico de barras**. O gráfico de colunas é mais indicado quando as legendas forem curtas e o de barras quando as legendas são extensas.

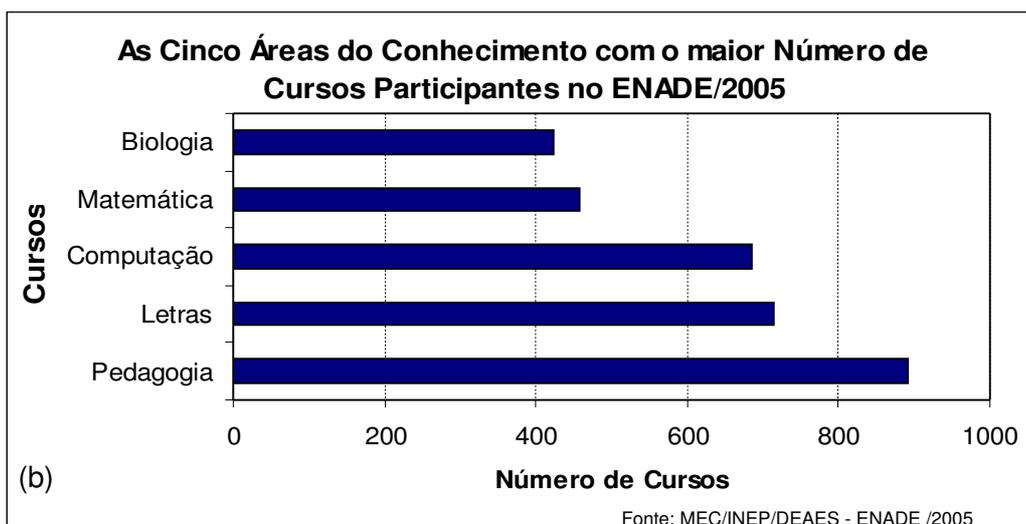
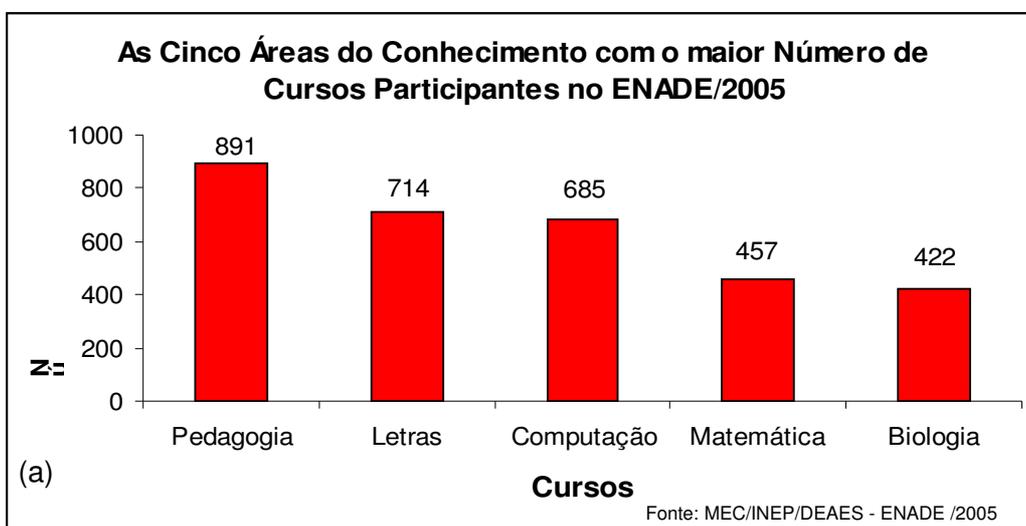


Figura 7.1 - Gráfico de colunas (a) e barras (b) para a série específica da tabela 7.1

Na representação gráfica de uma série temporal o gráfico de linhas, geralmente, é o mais indicado principalmente quando o objetivo é mostrar as observações em função do tempo em que elas foram medidas. As linhas são particularmente mais eficientes do que as colunas, quando existem intensas flutuações nos dados. Na Figura 7.2 é apresentada a série temporal da Tabela 7.2.

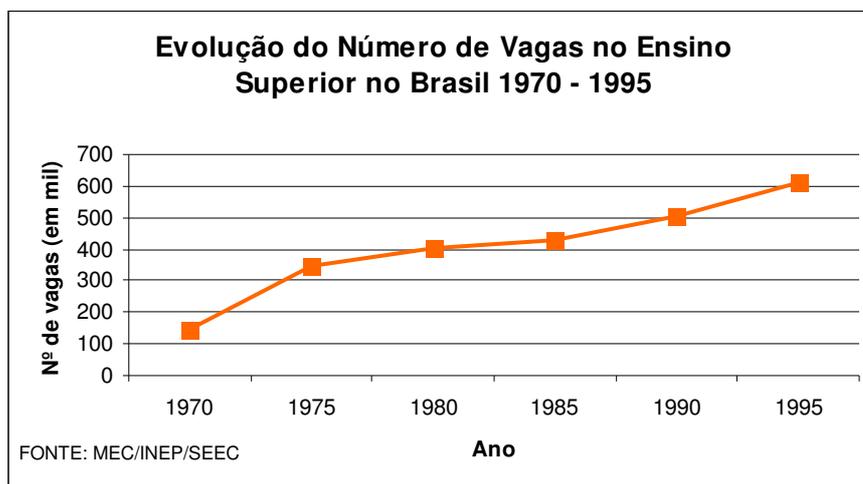


Figura 7.2 - Gráfico de linhas para a série temporal da tabela 7.2

O **gráfico de setores** ajuda a visualizar a porção que cada grupo representa no todo. O total é representado pelo círculo que fica dividido em tantos setores quantas são as partes. Por exemplo, na Figura 7.3, a fatia relativa a Região Sul corresponde a 21% do círculo. Obtemos cada setor por meio de uma regra de três simples, lembrando que o total da série corresponde a 360°.

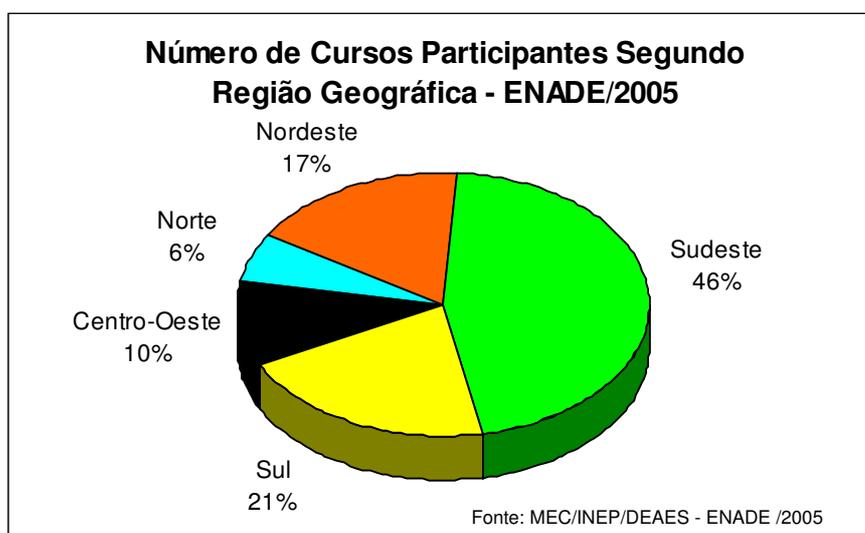


Figura 7.3 - Gráfico de setores para a série geográfica da tabela 7.3

Na representação gráfica de uma série mista podemos utilizar um gráfico de colunas, barras ou linhas múltiplas. Na Figura 7.4 é representada a serie mista da tabela 7.4 e na Figura 7.5 a série mista da tabela 7.5. A escolha do gráfico depende do tipo de informação a ser transmitida.

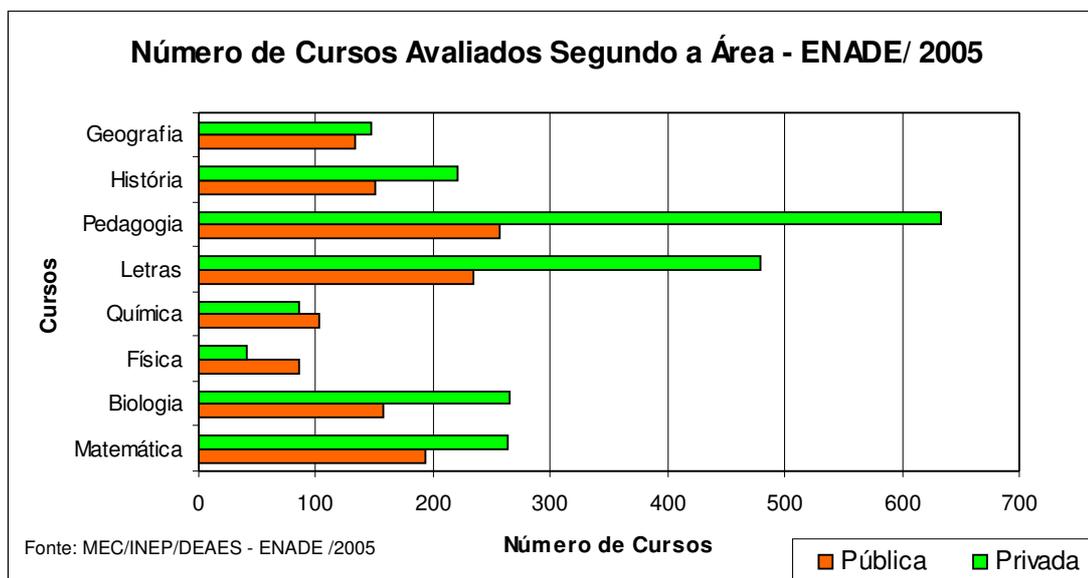


Figura 7.4 - Gráfico de barras múltiplas para a série mista da tabela 7.4

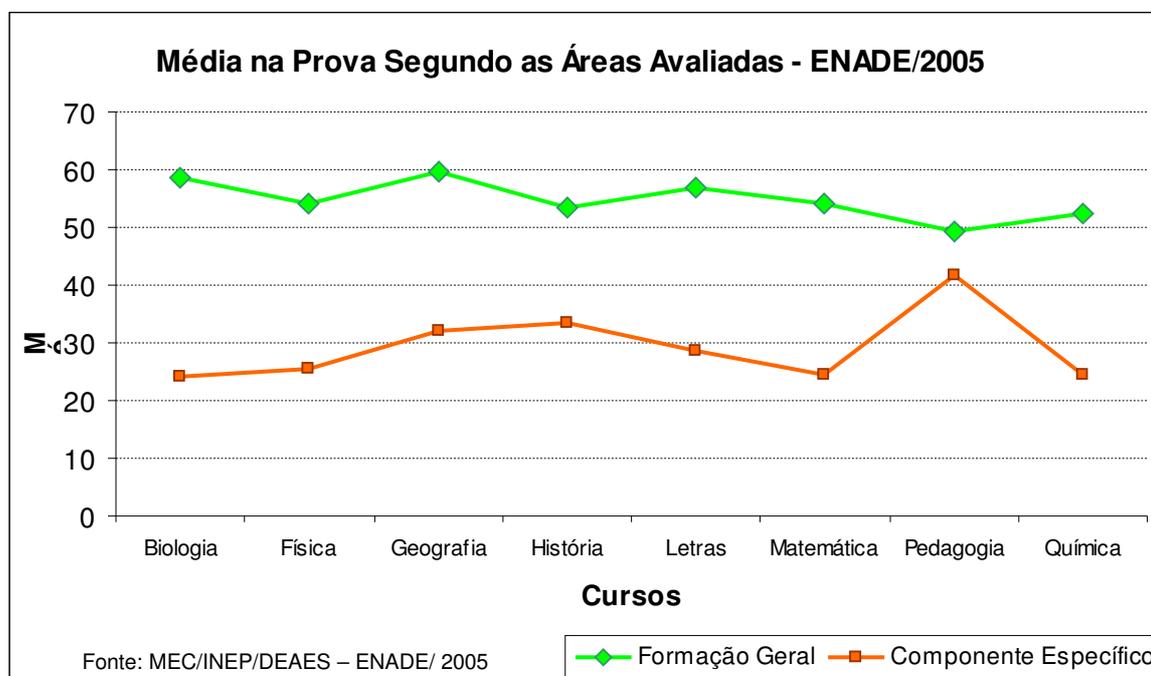


Figura 7.5 - Gráfico de linhas múltiplas para a série mista da tabela 7.5

Mais sobre gráficos

Os gráficos constituem excelente forma de apresentar dados, mas também podem ser enganosos. Na Figura 7.6 são apresentados dois gráficos que representam a evolução nas vendas de uma loja durante os meses de janeiro a junho.

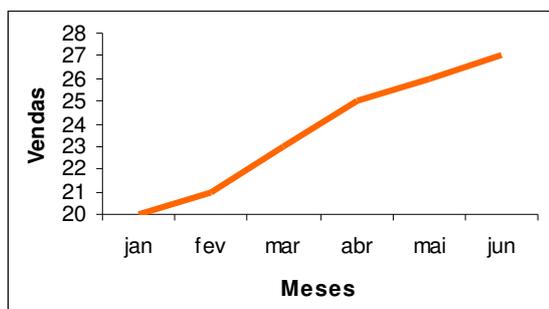


Figura 7.6 (a)

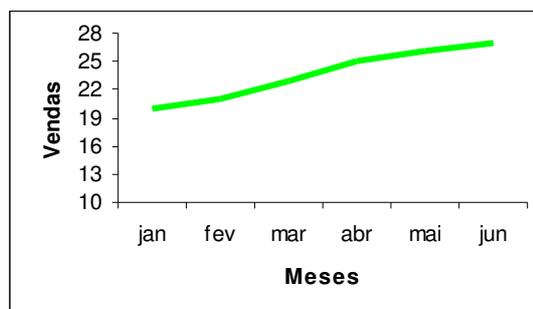


Figura 7.6 (b)

O gráfico da Figura 7.6(a) aparentemente apresenta maior evolução do que o gráfico da Figura 7.6(b), no entanto, os dois gráficos representam o mesmo conjunto de dados a diferença está na escala do eixo vertical. O primeiro gráfico inicia os valores do eixo vertical em 20 e o segundo gráfico inicia em 10 o que modifica a linha do gráfico. Portanto, preste atenção nos gráficos, porque o visual pode estar “maquiado”.

Gráfico em escala logarítmica

Todos os gráficos até aqui apresentados foram feitos em escala aritmética, mas nem sempre essa escala é a mais adequada. A escala logarítmica é usada na representação de valores em que a amplitude total é tão grande que a representação em escala aritmética torna-se pouco prática ou até inviável devido ao tamanho necessário para a escala. Na escala logarítmica as proporções entre as grandezas são alteradas, corrigindo as deformações traçadas na escala aritmética.

Exemplo: A tabela abaixo apresenta a comparação da evolução da produção agrícola entre o Estado A e o Estado B. Podemos verificar pela tabela abaixo que a produção do Estado A evoluiu proporcionalmente a produção do Estado B, ou seja, do ano de 1980 para 1985 a produção dobrou nos dois Estados. De 1985 para 1990 e de 1990 para 1995, quando a produção dobra num Estado também dobrou no outro. No ano de 1995 para 2000 a produção triplica nos dois Estados ocorrendo o mesmo de 2000 a 2005.

Tabela 7.6 Comparação da Evolução da produção Agrícola entre dois Estados

Anos	Estado A (t)	Estado B (t)
1980	900	200
1985	1.800	400
1990	3.600	800
1995	7.200	1.600
2000	21.600	4.800
2005	64.800	14.400

Fonte: Fictícia

Para representar os dados acima utilizaremos um gráfico de linhas múltiplas, uma linha para cada Estado. Como a evolução da produção foi proporcional nos dois Estados espera-se que as duas linhas sejam paralelas. Usaremos primeiro o gráfico com escala aritmética, Figura 7.7.

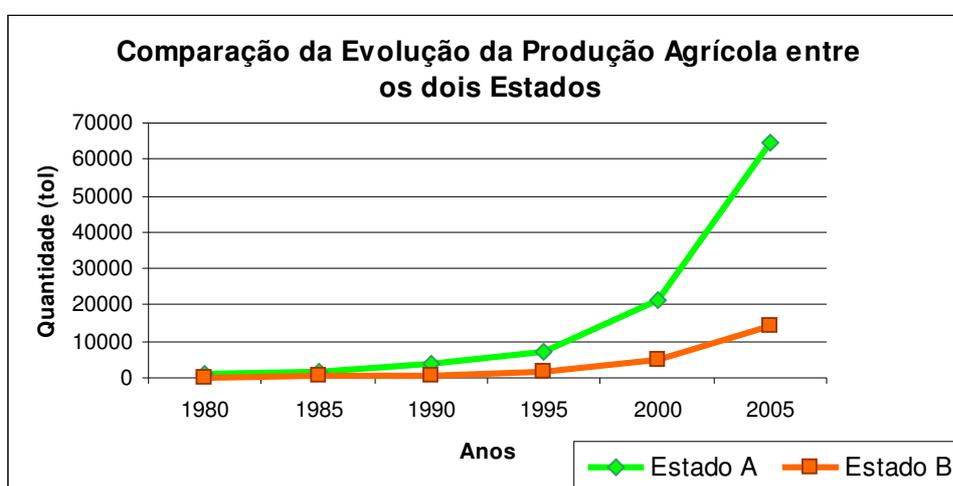


Figura 7.7 - Gráfico de linhas múltiplas em escala aritmética para os dados da Tabela 7.6

A observação do gráfico deixa evidente a impressão de que a produção de café do Estado A cresceu de forma mais acentuada do que a produção agrícola do Estado B. Ocorre que tal impressão é absolutamente errônea. Em face do exposto, verificamos que a escala aritmética não se presta para acompanhar a evolução relativa em dados com amplitude total muito grande, ou seja, a diferença entre o menor valor (produção do Estado B = 200 toneladas) e o maior valor (produção do Estado A = 64.800 toneladas), sendo mais apropriada à escala logarítmica apresentada no gráfico abaixo.

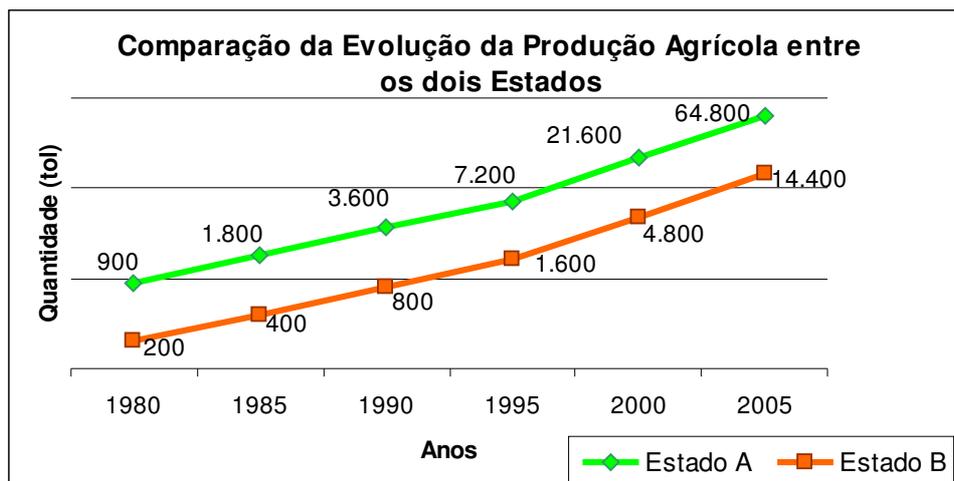


Figura 7.8 - Gráfico de linhas múltiplas em escala logarítmica para os dados da Tabela 7.6

O gráfico em escala logarítmica, Figura 7.8, retrata exatamente o que foi observado na Tabela 7.6, a produção entre os dois Estados é proporcional, quando a produção dobra em um Estado dobra no outro, quando a produção triplica em um Estado, também triplica no outro.

Apresentação tabular de dados quantitativos ou numéricos

Uma distribuição de freqüência é uma série estatística na qual os dados são organizados em grupos de classes ou categorias convenientemente estabelecidas. Na apresentação dos dados numéricos as distribuições de freqüência podem ser divididas em dois tipos: distribuição de freqüência sem intervalos de classe e distribuição de freqüência com intervalos de classe.

Quando os dados quantitativos estão listados na ordem que foram coletados sem nenhum tipo de ordenação, são denominados **dados brutos**. Na tabela abaixo são apresentados os dados brutos das idades de 40 alunos da disciplina de Probabilidade e Estatística, FURG-1998.

Dados brutos:

18	19	19	19	20	20	21	19	18	22
21	22	20	18	19	19	18	21	23	20
19	20	20	21	22	20	19	19	18	21
23	18	24	19	20	19	22	20	19	21

Observando a tabela acima, verifica-se que é difícil determinar se os valores estão concentrados ou dispersos, e qual é o menor ou o maior valor. Mas se os dados forem

organizados em ordem crescente ou decrescente, em **rol**, fica mais fácil ter uma idéia do comportamento dos valores.

Dados em rol:

18 18 18 18 18 18 19 19 19 19
19 19 19 19 19 19 19 19 20 20
20 20 20 20 20 20 20 21 21 21
21 21 21 22 22 22 22 23 23 24

Se um conjunto de dados consiste em relativamente poucos valores, muitos dos quais estão repetidos, como no exemplo das idades, simplesmente contamos quantas vezes cada um desses valores ocorre e então apresentamos o resultado em uma tabela.

Tabela das Idades dos Alunos

Idade	Freqüência
18	6
19	12
20	9
21	6
22	4
23	2
24	1
Total	40

Acrescentando título e fonte à tabela teremos uma distribuição de freqüência sem intervalos de classe. Neste tipo de distribuição de freqüência não há perda de informação, pois sabemos exatamente o número de alunos e a idade de cada uma dessas classes.

**Tabela 7.7 – Distribuição de freqüência sem intervalos de classe
Idade dos Alunos da Disciplina de Probabilidade e Estatística FURG-1998**

classes	Idade	Freqüência
1	18	6
2	19	12
3	20	8
4	21	4
5	22	3
6	23	2
7	24	1
	Total	36

Fonte: Sala de aula

Em uma distribuição de freqüência as classes representam as linhas da tabela. Na tabela 7.7 há sete classes de variação. A idade de 20 anos está na terceira classe da distribuição de freqüência. As classes são representadas pela letra i , sendo $i = 1,2,3,\dots,k$ (onde k é o número total de classes da distribuição).

Quando um conjunto de dados possui muitos valores com poucas repetições, a construção de uma distribuição de freqüência sem intervalos de classes pode ficar muito extensa, sendo mais conveniente agrupar os valores em intervalos de classe, o que resultará em uma distribuição de freqüência com intervalos de classe.

Exemplo: Idade de 40 alunos

17	18	18	18	18	19	19	19	20	20
20	21	22	22	22	23	23	24	24	24
24	25	25	26	26	26	27	28	28	29
29	30	30	31	33	34	35	39	42	46

Organizando os dados em uma tabela temos:

Idade	f_i	Idade	f_i	Idade	f_i
17	1	24	4	31	1
18	4	25	2	33	1
19	3	26	2	34	1
20	3	27	1	35	1
21	1	28	2	39	1
22	3	29	2	42	1
23	2	30	2	46	1

Para montar uma distribuição de freqüência com intervalos de classe precisamos realizar os seguintes cálculos:

- Amplitude total (AT): é a diferença entre o maior e o menor valor do conjunto de dados.

$$AT = V_{\max} - V_{\min}$$

$$AT = 46 - 17 = 29 \quad \rightarrow \quad AT = 29$$

- Determinar o número de classes, geralmente usa-se de 5 a 15 classes. Menos de 5 há muita perda de informação e mais do que 15 a tabela fica muito extensa dificultando a interpretação. Para ter uma idéia do número de classes mais adequado pode-se usar a regra de Sturges:

$$K = 1 + 3,3 \log N$$

sendo k o número de classes e N o número de valores a distribuir.

$$k = 1 + 3,3 \cdot \log 40 = 1 + 3,3 \cdot 1,60206 = 6,2868 \rightarrow k = 6,29$$

- Calcular a amplitude de cada classe (h) dividindo a amplitude total (AT) pelo número de classes (k). A amplitude de classe é definida como o intervalo entre o limite inferior e o limite superior de cada classe. O valor de h pode ser arredondado para um número inteiro para facilitar a análise e interpretação dos dados.

$$h = \frac{AT}{k} = \frac{V_{\max} - V_{\min}}{k}$$

$$h = \frac{AT}{k} = \frac{29}{6,29} = 4,61 \rightarrow h \cong 5$$

- Contar a freqüência de ocorrência de cada valor que pertence à classe.

Geralmente, inicia-se a primeira classe pelo menor valor do conjunto de dados, somando-se o valor da amplitude de classe para encontrar o limite superior, mas pode-se iniciar com outro valor mais conveniente.

Observação: Limites de classe são os extremos de cada classe. O menor número é o limite inferior da classe (L_i) e o maior número é o limite superior da classe (L_s).

Pelos cálculos a distribuição de freqüência do exemplo terá em torno de seis classes com a amplitude de cada classe igual a 5, ou seja, se começarmos com o menor valor a primeira classe conterà todas as idades maiores ou iguais a 17 e menores do que 22 ($17+5$). O valor 22 é o limite superior da primeira classe. O limite superior não será contado nesta classe e sim na próxima por isso o sinal entre os números na distribuição é fechado à esquerda e aberto a direita.

Tabela 7.8 – Distribuição de freqüência com intervalos de classe

Distribuição de Freqüência das Idades dos Alunos	
Idades	Freqüências
17 -- 22	12
22 -- 27	14
27 -- 32	8
32 -- 37	3
37 -- 42	1
42 -- 47	2
Total	40

Fonte: Fictícia

Algumas distribuições de freqüência são montadas utilizando classes do tipo “menos de” ou “mais de”, denominadas classes abertas, usadas, geralmente, para reduzir o número de classes de uma distribuição quando o conjunto de dados possui alguns valores muito menores ou muito maiores do que os demais. A desvantagem no uso de distribuições com classes abertas é a impossibilidade de calcular medidas estatísticas como, por exemplo, média e desvio padrão.

Tabela 7.9 - Distribuição de freqüência com classes abertas

Distribuição de Freqüência das Idades dos Alunos	
Idades	Freqüências
Menos de 20	8
20 -- 25	13
25 -- 30	10
30 -- 35	5
35 -- 40	2
Mais de 40	2
Total	40

Fonte: Fictícia

Existem outras formas de se montar uma distribuição de freqüência. O importante é escolher a melhor maneira de descrever o conjunto de dados facilitando a compreensão e análise da variável sob estudo sem muita perda de informação.

Quando os dados são quantitativos, além da freqüência simples, relativa e percentual de cada classe, também podemos calcular a freqüência acumulada.

Freqüência acumulada

A freqüência acumulada ajuda a estabelecer pontos de corte em uma distribuição de freqüência. No exemplo das idades, tabela 7.8, podemos dizer que dos 40 alunos 26 tem menos do que 27 anos. Para calcular a freqüência acumulada simplesmente somamos as freqüências simples, partindo do topo da distribuição.

- **Freqüência acumulada (Fa):** é a soma da freqüência simples da classe com as freqüências simples das classes anteriores. A freqüência acumulada da última classe é igual ao número total de observações.

Exemplo: A freqüência acumulada da terceira classe da tabela 7.8 é dada pela soma das freqüências simples da primeira, segunda e terceira classe.

$$\begin{aligned}F_3 &= f_1 + f_2 + f_3 \\F_3 &= 12 + 14 + 8 = 34\end{aligned}$$

Da mesma forma, também podemos calcular a freqüência relativa ou percentual da freqüência acumulada de forma semelhante ao realizado no cálculo da freqüência relativa e percentual da freqüência simples.

- **Freqüência acumulada relativa (Far):** é a freqüência acumulada da classe, dividida pela freqüência total do conjunto de dados:

$$Far_i = \frac{F_i}{\sum f_i}$$

Exemplo: Calcule a freqüência acumulada relativa da terceira classe.

$$Fr_3 = \frac{F_3}{\sum f_i} \Rightarrow Fr_3 = \frac{8}{40} = 0,2 \Rightarrow Fr_3 = 0,2$$

- **Freqüência acumulada percentual (Fap_i):** é a freqüência acumulada relativa da classe multiplicada por 100.

Exemplo: Calcule a freqüência acumulada percentual da terceira classe.

$$Fap_i = Fr_i \times 100 = 0,2 \times 100 = 20\%$$

Tabela 7.10 – Tipos de freqüência

Distribuição de Freqüência das Idades dos Alunos				
Idades	f_i	$f \%$	F_a	F_{ap}
17 -- 22	12	30%	12	30%
22 -- 27	14	35%	26	65%
27 -- 32	8	20%	34	85%
32 -- 37	3	7,5%	37	92,5%
37 -- 42	1	2,5%	38	95%
42 -- 47	2	5%	40	100%
Total	40	100%	-	-

Fonte: Fictícia

Na distribuição das idades dos alunos podemos verificar que oito alunos tem 27 anos ou mais e menos de 32 anos o que representa 20% dos alunos da turma. Continuando a interpretar a mesma classe ainda podemos dizer que 34 alunos têm menos de 32 anos, representando 85% dos alunos.

Apresentação gráfica de dados quantitativos ou numéricos

As distribuições de freqüência podem ser representadas graficamente pelo **histograma** e pelo **polígono de freqüência**.

Quando a distribuição de freqüência é **sem intervalos** de classe o **histograma** é um conjunto de hastes ou colunas, Figura 7.9(a), representadas em um sistema de coordenadas cartesianas que tem por base os valores da distribuição e por altura a freqüências simples, relativa ou percentual correspondente.

Quando a distribuição de freqüência é **com intervalos** de classe o **histograma** é formado por um conjunto de retângulos justapostos, Figura 7.9(b), cujas bases representam os intervalos de classe e cujas alturas representam as freqüências simples, relativas ou percentuais de cada intervalo de classe.

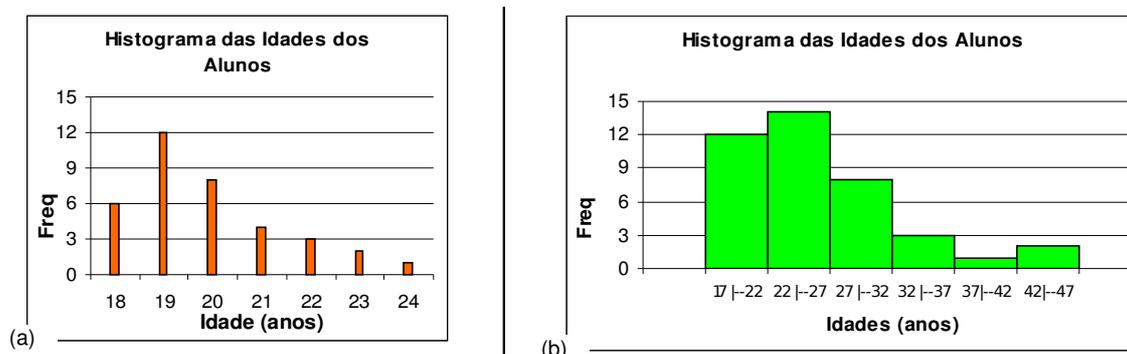


Figura 7.9 - (a) Histograma da distribuição sem intervalo de classes das idades dos alunos da tabela 7.7. (b) histograma da distribuição com intervalos de classe das idades dos alunos, da tabela 7.8.

Ao analisar a representação gráfica de um conjunto de dados numéricos é possível observar seus aspectos principais, como forma, valores máximos e mínimos, picos mais destacados, valores extremos.

O **polígono de frequência** é um gráfico em linha, sendo as frequências marcadas sobre os pontos médios dos intervalos de classe e unidas por segmentos de retas. Uma vez que o gráfico deve começar e terminar sobre o eixo horizontal, cria-se uma classe antes da primeira classe da distribuição e marca-se seu ponto médio. Faça o mesmo no final da distribuição. Estende-se a linha do gráfico até estes dois últimos pontos médios calculados.

Observação: O ponto médio de uma classe é a metade da soma entre os limites inferior e superior da classe.

$$\text{Ponto médio} = \frac{\text{limite inferior da classe} + \text{limite superior da classe}}{2}$$

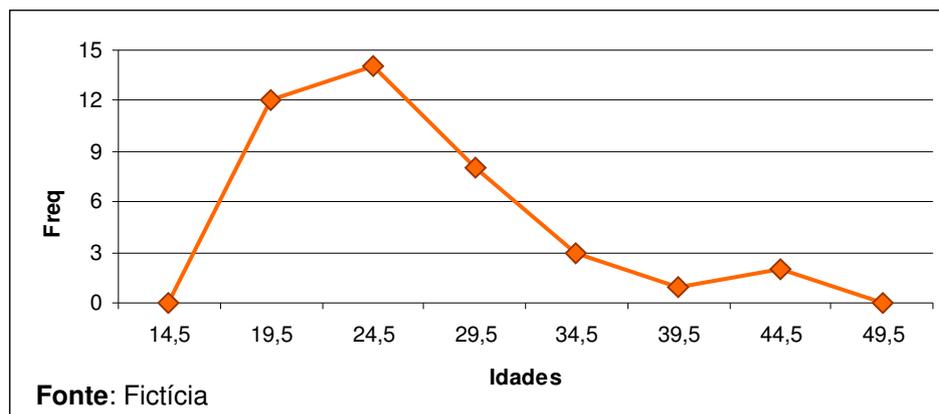


Figura 7.11 - Polígono de frequência das idades da tabela 2.8.

O polígono de frequência é utilizado também quando duas ou mais distribuições de frequência são comparadas. Neste caso, os histogramas não são apropriados, pois a sobreposição das colunas de um histograma e outro causaria dificuldades na interpretação. A utilização da distribuição de frequência relativa ou da distribuição de percentagem torna-se essencial sempre que um conjunto de dados está sendo comparado com outro conjunto de dados, especialmente se o número de observações em cada conjunto difere entre si.

Exemplo: Uma pesquisa de mercado pretende comparar os salários dos empregados de duas empresas os quais estão apresentados nas tabelas 7.11 e 7.12.

Tabela 7.11 – Distribuição de freqüência dos salários dos empregados da empresa A
 Distribuição de Freqüência dos Salários dos
 Empregados da Empresa A

Classe	Salário	Freqüência simples
1	400 -- 600	2
2	600 -- 800	4
3	800 -- 1000	9
4	1000 -- 1200	14
5	1200 -- 1400	21
6	1400 -- 1600	16
7	1600 -- 1800	13
8	1800 -- 2000	8
9	2000 -- 2200	3
	Total	90

Fonte: Ficitícia

Tabela 7.12 – Distribuição de freqüência dos salários dos empregados da empresa B
 Distribuição de Freqüência dos Salários
 dos Empregados da Empresa B

Classe	Idades	Freqüência simples
1	400 -- 600	1
2	600 -- 800	2
3	800 -- 1000	5
4	1000 -- 1200	39
5	1200 -- 1400	62
6	1400 -- 1600	37
7	1600 -- 1800	8
8	1800 -- 2000	5
9	2000 -- 2200	1
	Total	160

Fonte: Ficitícia

Para construir o polígono de freqüência calcula-se o ponto médio e a freqüência relativa de cada classe. Para finalizar o gráfico é necessário criar uma classe anterior a primeira classe e uma classe posterior a última classe da distribuição de freqüência com freqüência simples igual a zero, como indicado pelas setas nas tabelas abaixo.

Tabela 7.11 – Distribuição de freqüência dos salários dos empregados da empresa A
Distribuição de Freqüência dos Salários dos Empregados

Classe	Salário	Ponto Médio	Freqüência simples	Freqüência relativa
-	200 -- 400	300	0	0
1	400 -- 600	500	2	0,02
2	600 -- 800	700	4	0,04
3	800 -- 1000	900	9	0,10
4	1000 -- 1200	1100	14	0,16
5	1200 -- 1400	1300	21	0,23
6	1400 -- 1600	1500	16	0,18
7	1600 -- 1800	1700	13	0,14
8	1800 -- 2000	1900	8	0,09
9	2000 -- 2200	2100	3	0,03
-	2200 -- 2400	2300	0	0
Total		-	90	1,0

Fonte: Fictícia

Tabela 7.12 – Distribuição de freqüência dos salários dos empregados da empresa B
Distribuição de Freqüência dos Salários dos Empregados

Classe	Salário	Ponto Médio	Freqüência simples	Freqüência relativa
-	200 -- 400	100	0	0
1	400 -- 600	300	1	0,01
2	600 -- 800	500	2	0,01
3	800 -- 1000	700	5	0,03
4	1000 -- 1200	900	39	0,24
5	1200 -- 1400	1100	62	0,39
6	1400 -- 1600	1300	37	0,23
7	1600 -- 1800	1500	8	0,05
8	1800 -- 2000	1700	5	0,03
9	2000 -- 2200	1900	1	0,01
-	2200 -- 2400	2100	0	0
Total		-	160	1,0

Fonte: Fictícia

Montando o gráfico:

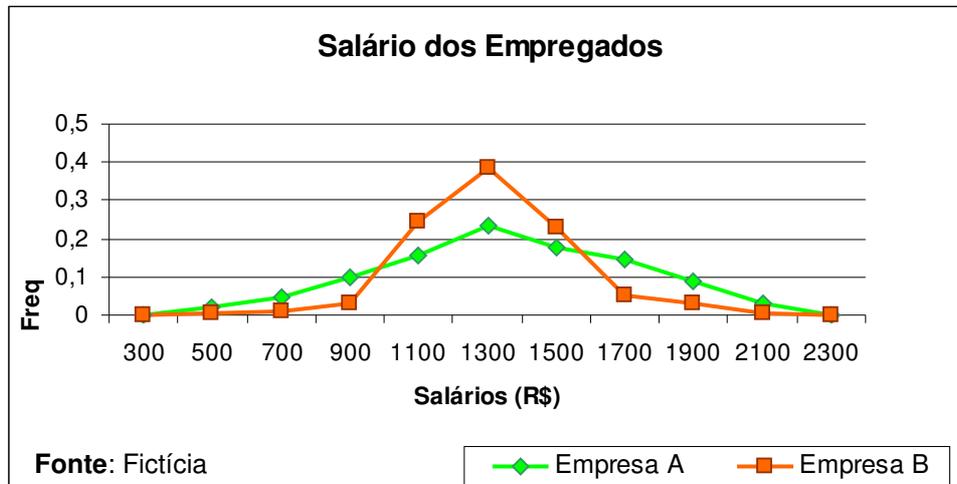


Figura 7.12 - Polígono de frequência relativa dos salários dos empregados.

Na Figura 7.12 podemos verificar que as empresas apresentam diferentes distribuições salariais. Na Empresa B os salários estão mais concentrados variando de R\$ 900,00 a R\$ 1.700,00, sendo que 39% dos funcionários recebem em torno de R\$ 1.300,00. Já na empresa A, a variação de faixa salarial é mais ampla variando de R\$ 700,00 a R\$ 2.100,00.