



UNIVERSIDADE FEDERAL DE ALAGOAS
Instituto de Computação

PROBABILIDADE E ESTATÍSTICA

Aula I - Introdução

UNIDADE I

PROFESSOR: ESTEVAM VILAR
TURMA: ENGENHARIA DA COMPUTAÇÃO – 2018.1



APRESENTAÇÃO

Professor: Estevam Vilar

- ❑ Mestrando em Informática – Data Mining e Machine Learning
- ❑ MBA em Governança em Tecnologia da Informação
- ❑ Bacharel em Ciência da Computação
- ❑ Graduado em Telecomunicações

E-mail: estevamvilar@ic.ufal.br

Site da Disciplina: <http://sites.google.com/view/estevamvilar>

EMENTA DA DISCIPLINA

Ementa: Análise exploratória. Probabilidade. Distribuição discreta e contínua de variáveis aleatórias. Valor esperado e variância. Inferência estatística: estimação de parâmetros e teste de hipóteses. Tópicos especiais. Utilização de ferramentas computacionais.

Bibliografia:

MAGALHÃES, M.T; LIMA, A.C. **Noções de Probabilidade e Estatística**. 6a ed.. São Paulo: EDUSP, 2004.

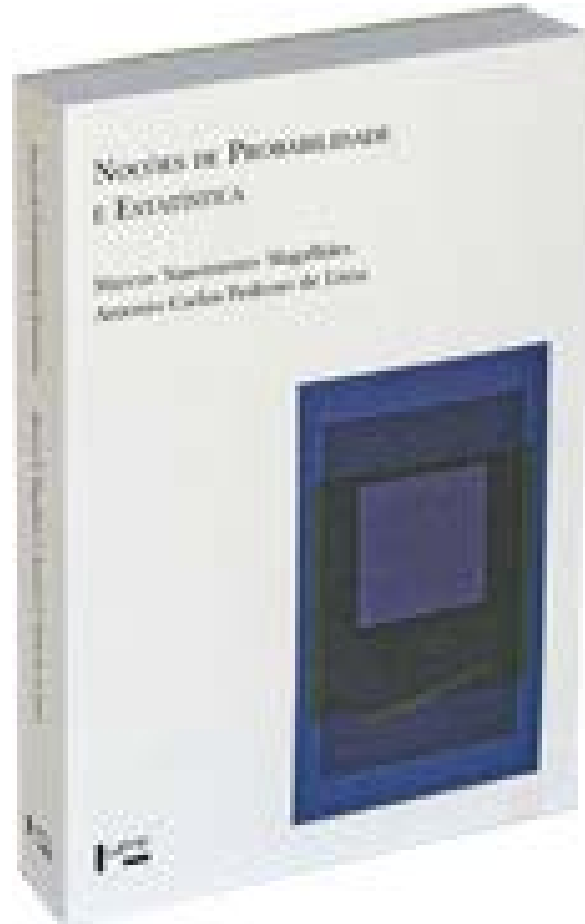
BUSSAB, W; MORETTIN, P. **Estatística Básica**. 5a ed.. São Paulo: Saraiva, 2004.

YATES, R.D; GOODMAN, D.J. **Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers**. New York: John Wiley & Sons, 2002.

DEKKING, F. M.; KRAAIKAMP, C.; LOPUHAÄ, H. P. & Meester, L. E. **A Modern Introduction to Probability and Statistics: Understanding Why and How**. Springer, 2005.

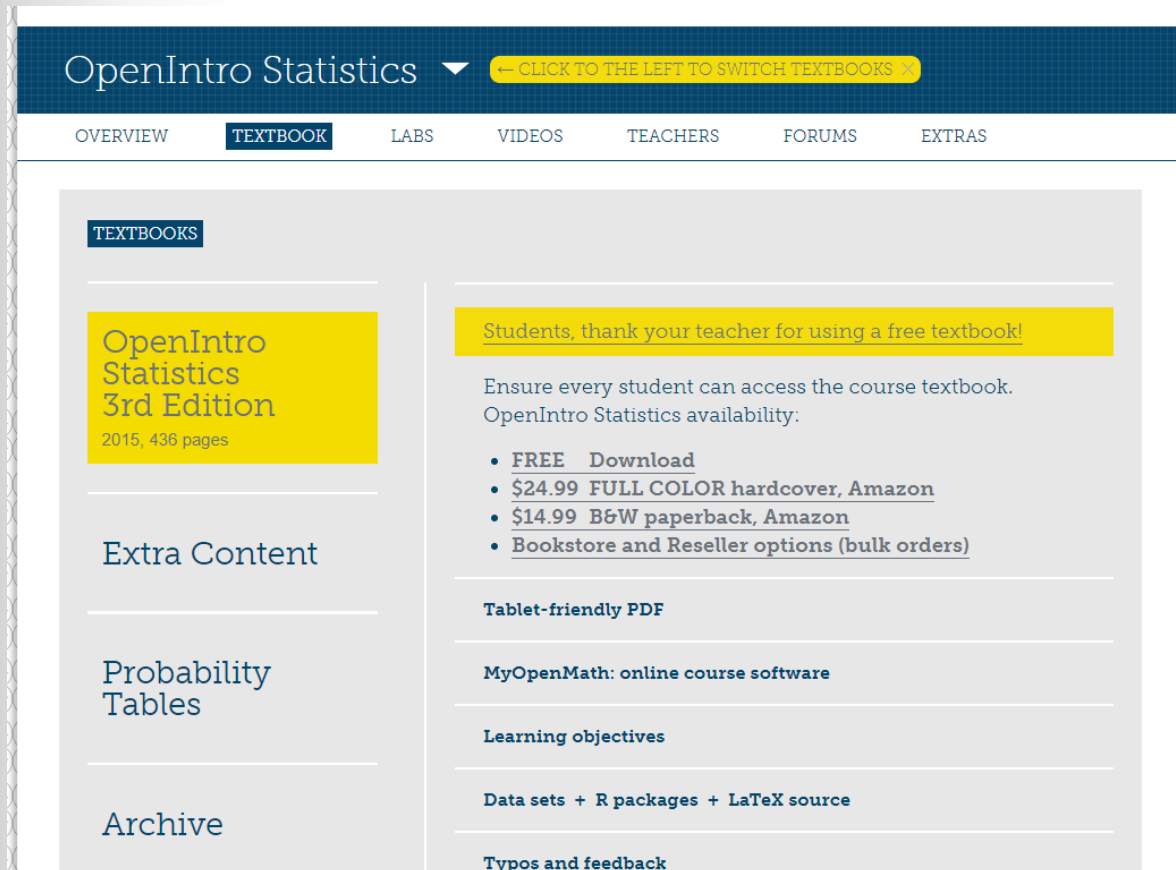
MAGALHÃES, M. N. **Probabilidade e Variáveis Aleatórias**. IME-USP, 2004. VERZANI, J. **Using R for Introductory Statistics**. Chapman & Hall/CRC, 2004. LAPPONI, J.C. **Estatística Usando Excel**. 2a ed.. São Paulo: Lapponi, 2000. DALGAARD, P. **Introductory Statistics with R**. Springer, 2002.

BIBLIOGRAFIA



MAGALHÃES, M.T; LIMA, A.C. **Noções de Probabilidade e Estatística**. 6a ed.. São Paulo: EDUSP, 2004.

BIBLIOGRAFIA



<https://www.openintro.org/stat>

EMENTA DA DISCIPLINA

R é uma linguagem e também um ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos.



ESCOPO DA DISCIPLINA

UNIDADE I

Análise exploratória.

Probabilidade.

Distribuição discreta e contínua de variáveis aleatórias.

Valor esperado e variância.

UNIDADE II

Inferência estatística: estimação de parâmetros e teste de hipóteses.

Tópicos especiais. Utilização de ferramentas computacionais.

ESCOPO DA DISCIPLINA

UNIDADE I

Análise exploratória.

Probabilidade.

Distribuição discreta e contínua de variáveis aleatórias.

Valor esperado e variância.

UNIDADE II

Inferência estatística: estimação de parâmetros e teste de hipóteses.

Tópicos especiais. Utilização de ferramentas computacionais.

CONHECENDO A TURMA...

- **Quais as suas expectativas em relação à disciplina?**
- **Qual o seu nível de conhecimento sobre Estatística?**
- **Onde você lembra de já ter utilizado estatística na sua vida?**
- **Você utiliza Estatística em algum Grupo de Pesquisa?**

INTRODUÇÃO

BREVE HISTÓRICO SOBRE PROBABILIDADE E ESTATÍSTICA

Vídeo I...

https://www.youtube.com/watch?v=r0GnS_SWU2s

INTRODUÇÃO

Estatística

- A palavra estatística lembra, à maioria das pessoas, recenseamento.
- Os censos existem há milhares de anos e constituem um esforço imenso e caro feito pelos governos, com o objetivo de conhecer seus habitantes, sua condição socioeconômica, sua cultura, religião, etc.
- Portanto, associar estatística a censo é perfeitamente correto do ponto de vista histórico, sendo interessante salientar que as palavras **estatística** e **estado** têm a mesma origem latina: *status*.

INTRODUÇÃO

Estatística

- A estatística é também comumente associada às pesquisas de opinião pública, aos vários índices governamentais, aos gráficos e às médias publicados diariamente na imprensa.
- Na realidade, entretanto, a estatística engloba muitos outros aspectos, sendo fundamental na análise de dados provenientes de quaisquer processos onde exista **variabilidade**.

INTRODUÇÃO

A Importância da Estatística

- O mundo está repleto de problemas. Para resolvermos a maioria deles, necessitamos de informações. Mas, que tipo de informação? Que quantidade de informações? Após obtê-las, que fazer com elas?
- A Estatística trabalha com essas informações, associando os dados ao problema, descobrindo como e o que coletar, assim capacitando o pesquisador (ou profissional ou cientista) a obter conclusões a partir dessas informações, de tal forma que possam ser entendidas por outras pessoas.

INTRODUÇÃO

Conceito de Estatística

“A Estatística é uma parte da Matemática que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados, viabilizando a utilização dos mesmos na tomada de decisões.”

INTRODUÇÃO

Grandes áreas da Estatística

- **Estatística Descritiva e Amostragem** – Conjunto de técnicas que objetivam coletar, organizar, apresentar, analisar e sintetizar os dados numéricos de uma população, ou amostra;
- **Estatística Inferencial** – Processo de se obter informações sobre uma população a partir de resultados observados na amostra;
- **Probabilidade** - Modelos matemáticos que explicam os fenômenos estudados pela Estatística em condições normais de experimentação.

INTRODUÇÃO

➤ Estatística Descritiva e Amostragem

É a parte mais conhecida. Quem vê o noticiário, na televisão ou nos jornais, sabe quão frequente é o uso de médias, índices e gráficos nas notícias.

Exemplo 1 - INPC (Índice Nacional de Preços ao Consumidor)

Sua construção envolve a sintetização, em um único número, dos aumentos dos produtos de uma cesta básica.

INTRODUÇÃO

➤ Estatística Descritiva e Amostragem

Exemplo 2 - Anuário Estatístico Brasileiro

O IBGE publica esse anuário apresentando, em várias tabelas, os mais diversos dados sobre o Brasil: educação, saúde, transporte, economia, cultura, etc. Embora simples, fáceis de serem entendidas, as tabelas são o produto de um processo demorado e extremamente dispendioso de coleta e apuração de dados.

Exemplo 3 - Anuário Estatístico da Embratur

A Embratur publica esse anuário apresentando, em várias tabelas e gráficos, os mais diversos dados sobre Turismo Interno e dados sobre entrada de turistas estrangeiros no Brasil.

INTRODUÇÃO

➤ Estatística Inferencial

A tomada de decisões sobre a população, com base em estudos feitos sobre os dados da amostra, constitui o problema central da *inferência estatística*.

Exemplo 1 - Suponha que a distribuição das alturas de todos os habitantes de um país possa ser representada por uma distribuição normal. Mas não conhecemos de antemão a média da distribuição. Devemos, pois, estimá-la.

INTRODUÇÃO

➤ Estatística Inferencial

Exemplo 2 - Análise financeira. Os analistas financeiros estudam dados sobre a situação da economia, visando explicar tendências dos níveis de produção e de consumo, projetando-os para o futuro.

Exemplo 3 - Ocorrência de terremotos. Os geólogos estão continuamente coletando dados sobre a ocorrência de terremotos. Gostariam de inferir quando e onde ocorrerão tremores, e qual a sua intensidade. Trata-se, sem dúvida, de uma questão complexa, que exige longa experiência geológica, além de cuidadosa aplicação de métodos estatísticos.

INTRODUÇÃO

➤ Probabilidade

O processo de generalização, que é característico do método indutivo, está associado a uma margem de incerteza.

A existência da incerteza deve-se ao fato de que a conclusão, que se pretende obter para o conjunto de todos os indivíduos analisados quanto a determinadas características comuns, baseia-se em uma parcela do total das observações.

A medida da incerteza é tratada mediante técnicas e métodos que se fundamentam na *Teoria da Probabilidade*. Essa teoria procura quantificar a incerteza existente em determinada situação.

INTRODUÇÃO

Termos Comuns:

- **População:** é uma coleção completa de todos os elementos a serem estudados.
- **Amostra:** é uma subcoleção de elementos extraídos de uma população.
- **Censo:** é uma coleção de dados relativos a todos os elementos de uma população.
- **Parâmetros:** é uma medida numérica que descreve uma característica de uma população.
- **Estatística:** é uma medida numérica que descreve uma característica de uma amostra.
- **Dados contínuos:** resultam de um número infinito de valores possíveis que podem ser associados a pontos em uma escala contínua de tal maneira que não haja lacunas.

INTRODUÇÃO

Fases do Método Estatístico:

- **Definir** as variáveis que você deseja estudar, no intuito de solucionar um problema ou atender a um objetivo.
- **Coletar** os dados correspondentes às referidas variáveis, a partir das fontes apropriadas.
- **Organizar** os dados coletados por meio do desenvolvimento de tabelas.
- **Visualizar** os dados por meio do desenvolvimento de gráficos.
- **Analisar** os dados coletados de modo a tirar conclusões e apresentar os referidos resultados.

INTRODUÇÃO

Definir Dados - Tipos de Variáveis:

- **Variáveis categóricas** (também conhecidas como Variáveis qualitativas) apresentam valores que podem somente ser posicionados em categorias tais como sim e não.
- **Variáveis numéricas** (também conhecidas como Variáveis quantitativas) apresentam valores que representam quantidades.

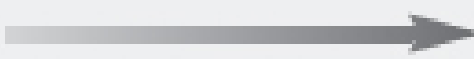
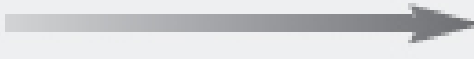
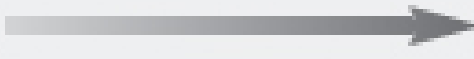
Variáveis numéricas podem ser, ainda, identificadas como variáveis discretas ou variáveis contínuas.

1. **Variáveis discretas** apresentam valores numéricos que surgem a partir de um processo de contagem.

2. **Variáveis contínuas** produzem respostas numéricas que surgem a partir de um processo de medição.

INTRODUÇÃO

Definir Dados - Tipos de Variáveis:

Pergunta	Respostas		Tipo de Dado
Você atualmente tem um perfil no Facebook?	<input type="checkbox"/> Sim <input type="checkbox"/> Não		Categórico
Quantas mensagens de texto você enviou nos últimos três dias?	_____		Numérico (discreto)
Quanto tempo foi necessário para que você baixasse a atualização para o aplicativo mais recente de seu aparelho de telefonia móvel?	_____ segundos		Numérico (contínuo)

INTRODUÇÃO

Definir Dados - Tipos de Variáveis:

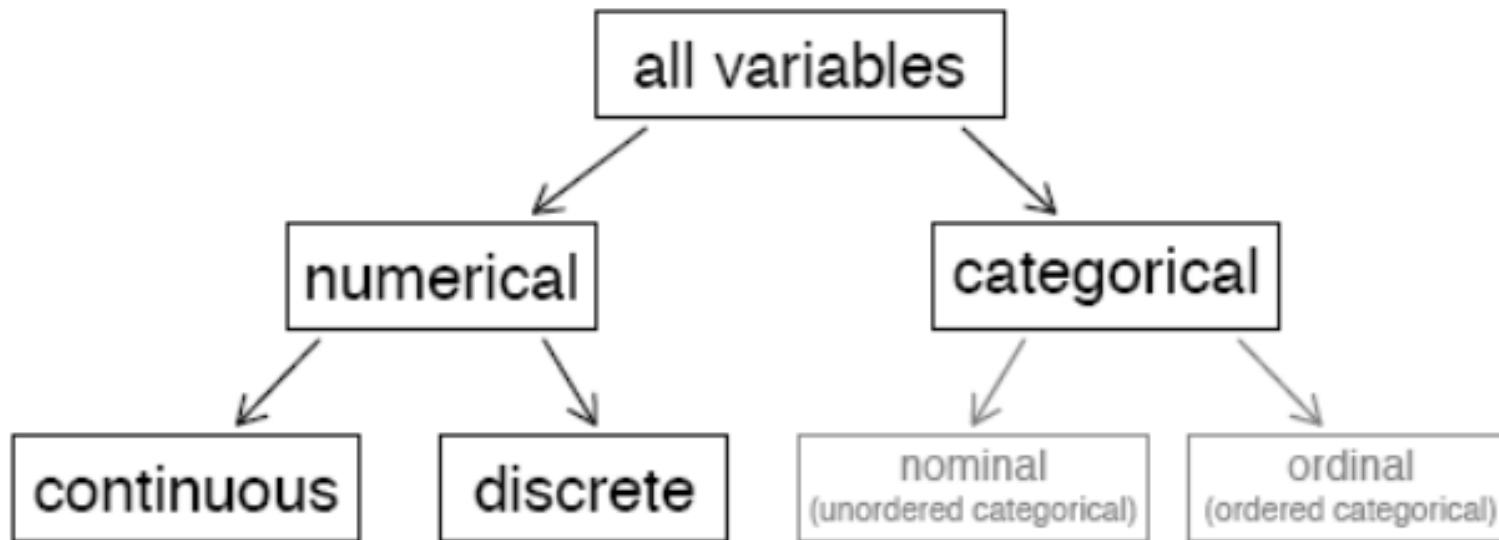
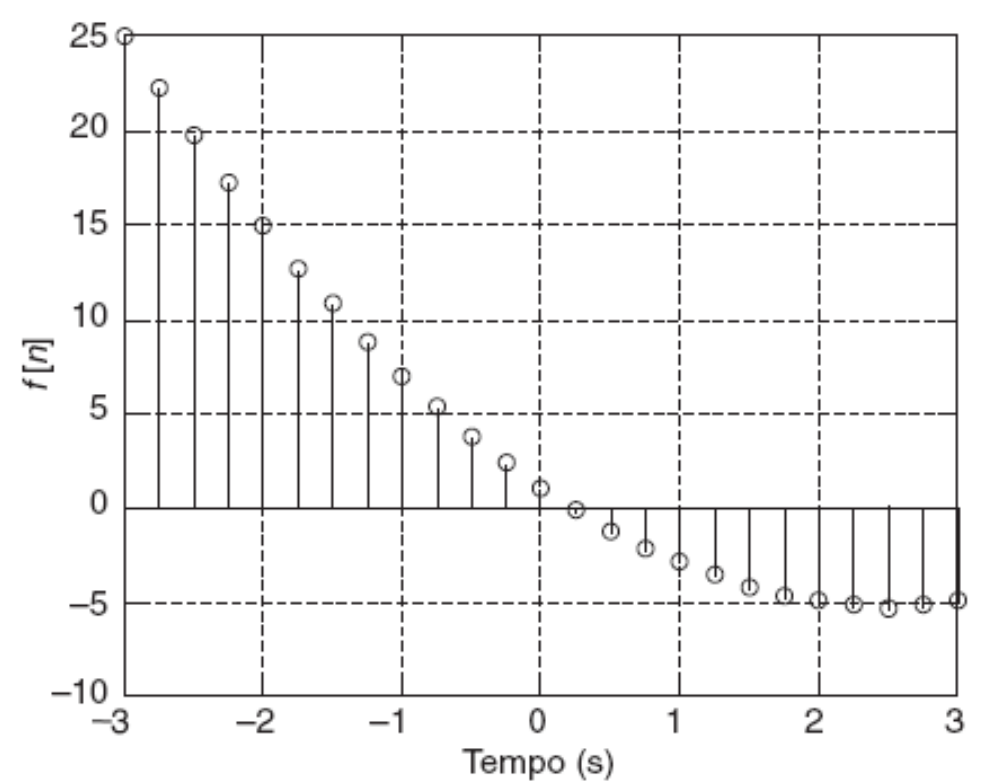
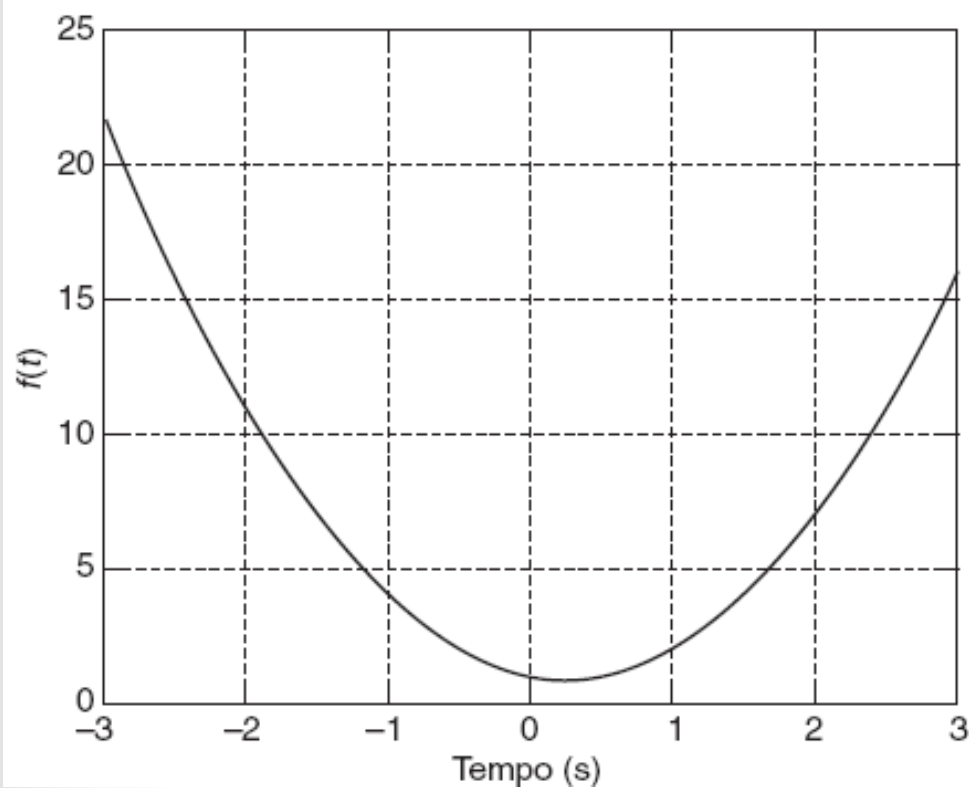


Figure 1.7: Breakdown of variables into their respective types.

INTRODUÇÃO

Definir Dados - Tipos de Variáveis: Numéricas Contínuas x Discretas



INTRODUÇÃO

Atividade:

- Explore o Site <https://www.openintro.org>
- Faça o Download do Livro *OpenIntro Statistics 3rd Edition*
- Faça a Leitura do Capítulo 1 do livro.
- Faça Download e Instale o R.

VISUALIZAÇÃO DE DADOS

VISUALIZAÇÃO DE DADOS

Séries Estatísticas

- Define-se série estatística como toda e qualquer coleção de dados estatísticos referidos a uma mesma ordem de classificação: ***quantitativa***.
- No sentido mais amplo, série é uma sucessão de números referidos a qualquer variável.
- Em outros termos, a palavra série é usada normalmente para designar um conjunto de dados dispostos de acordo com um caráter variável, residindo a qualidade serial na disposição desses valores, e não em uma disposição temporal ou espacial de indivíduos.

VISUALIZAÇÃO DE DADOS

Séries Estatísticas

Tabela é um quadro que resume um conjunto de observações.

As tabelas servem para apresentar séries estatísticas. Conforme varie um dos elementos

da série, podemos classificá-la em:

- ***Cronológicas*** - *Tempo* (fator temporal ou cronológico) – a que época refere-se o fenômeno analisado;
- ***Geográficas*** - *Local* (fator espacial ou geográfico) – onde o fenômeno acontece;
- ***Específicas*** - *Fenômeno* (espécie do fato ou fator especificativo) – o que é descrito.

VISUALIZAÇÃO DE DADOS

Série Cronológica

*Tabela 1.1 - Operadora WKX – Venda de bilhetes aéreos
– Mercado Interno – 1995*

<i>Meses</i>	<i>Vendas (em milhares de reais)</i>
Janeiro	2300
Fevereiro	1800
Março	2200
Abril	2210
Maio	2360
Junho	2600
Julho	2690
Agosto	3050
Setembro	3500
Outubro	3440
Novembro	3100
Dezembro	2760
TOTAL ANUAL	31510

Fonte: Departamento de Análise de Mercado

VISUALIZAÇÃO DE DADOS

Série Geográfica

Tabela 1.2 – Operadora WKX - Vendas por Unidade da Federação – 1995

<i>Unidades da Federação</i>	<i>Vendas (em milhares de reais)</i>
Minas Gerais	4000
Paraná	2230
Rio Grande do Sul	6470
Rio de Janeiro	8300
São Paulo	10090
Outros	420
TOTAL BRASIL	31510

Fonte: Departamento de Análise de Mercado

VISUALIZAÇÃO DE DADOS

Série Específica - Também chamada de *série categórica* ou *série por categoria*,

*Tabela 1.3.- Operadora WKX -
Venda de bilhetes aéreos por Linha - 1995*

<i>Linha do Produto</i>	<i>Vendas (em milhares de reais)</i>
Linha A	6450
Linha B	9310
Linha C	15750
TODAS AS LINHAS	31510

Fonte: Departamento de Análise de Mercado

VISUALIZAÇÃO DE DADOS

Série Específica

Tabela 1.4. Número de empregados das várias classes de salários no estado de São Paulo – 1998

<i>Classes de Salários (R\$)</i>	<i>Número de Empregados</i>
Até 80	41 326
De 80 a 119	123 236
De 120 a 159	428 904
De 160 a 199	324 437
De 200 a 399	787 304
De 400 a 599	266 002
De 600 a 799	102 375
De 800 a 999	56 170
1000 e mais	103 788
TOTAL	2 233 542

Fonte: Serviço de Estatística da Previdência e Trabalho

VISUALIZAÇÃO DE DADOS

Séries Estatísticas

Tabelas Compostas (ou de dupla entrada)

- As tabelas apresentadas anteriormente são tabelas estatísticas simples, onde apenas uma série está representada.
- É comum, todavia, haver necessidade de apresentar, em uma única tabela, mais do que uma série.
- Quando as séries aparecem conjugadas, tem-se uma tabela de dupla entrada. Em uma tabela desse tipo são criadas duas ordens de classificação: uma horizontal (linha) e uma vertical (coluna).

VISUALIZAÇÃO DE DADOS

Séries Estatísticas

Tabelas Compostas - Série específico-temporal

*A) Tabela 1.5 – População economicamente ativa
por setor de atividades – Brasil*

Setor	População (1 000 Hab.)		
	1940	1950	1960
Primário	8 968	10 255	12 163
Secundário	1 414	2 347	2 962
Terciário	3 620	4 516	7 525

Fonte: IPEA

VISUALIZAÇÃO DE DADOS

Séries Estatísticas

Tabelas Compostas - Série geográfico-temporal

B) Tabela 1.6 – População Indígena Brasileira

<i>Unidade de Produção</i>	<i>Produção</i>		
	1937	1938	1939
Acre	5 007	4 765	4 727
Amazonas	6 858	5 998	5 631
Pará	4 945	4 223	4 500
Mato Grosso	1 327	1 285	1 235
Outros Estados	333	539	337

Fonte: Anuário Estatístico do Brasil – IBGE - (Dados alterados para melhor compreensão)

VISUALIZAÇÃO DE DADOS

Séries Estatísticas

A Tabela Abaixo constitui uma Série Estatística???

Tabela 1.8 – Situação dos espetáculos cinematográficos no Brasil – 1967

<i>Especificação</i>	<i>Dados Numéricos</i>
Número de cinemas	2 488
Lotação dos cinemas	1 722 348
Sessões por dia	3 933
Filmes de longa metragem	131 330 488
Meia-entrada	89 581 234

Fonte: Anuário Estatístico do Brasil – IBGE

VISUALIZAÇÃO DE DADOS

Séries Estatísticas

Atenção!!! Nem sempre uma tabela representa uma série estatística. Por vezes, os dados reunidos não revelam uniformidade, sendo meramente um aglomerado de informações gerais sobre determinado assunto, as quais, embora úteis, não apresentam a consistência necessária para se configurar uma série estatística.

A Tabela exibida apresenta resumos de dados, mas não representa uma série estatística.

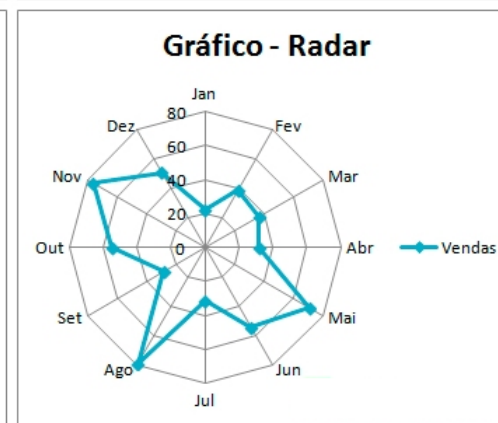
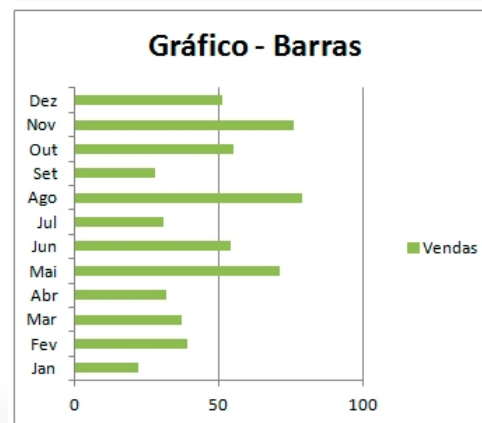
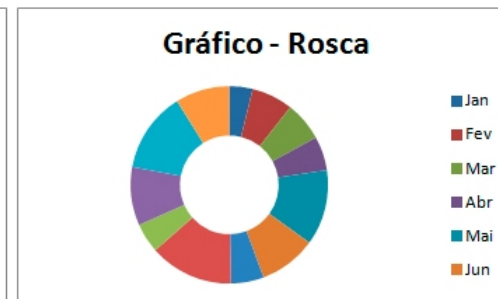
VISUALIZAÇÃO DE DADOS

Como apresentar os dados da melhor forma???



VISUALIZAÇÃO DE DADOS

Gráficos!!!



VISUALIZAÇÃO DE DADOS

Gráficos

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

VISUALIZAÇÃO DE DADOS

Gráficos

- **Simplicidade** – o gráfico deve ser destituído de detalhes de importância secundária, assim como de traços desnecessários que possam levar o observador a uma análise morosa ou sujeita a erros.
- **Clareza** – o gráfico deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.
- **Veracidade** – o gráfico deve expressar a verdade sobre o fenômeno em estudo.

VISUALIZAÇÃO DE DADOS

Gráficos – Diretrizes para a Construção

- O título do gráfico deve ser o mais claro e completo possível. Quando necessário, deve-se acrescentar subtítulos;
- A orientação geral dos gráficos deve ser da esquerda para a direita;
- As quantidades devem ser representadas por grandezas lineares;
- Sempre que possível, a escala vertical há de ser escolhida de modo a aparecer a linha 0 (zero);
- Só devem ser incluídas no desenho as coordenadas indispensáveis para guiar o olhar do leitor ao longo da leitura. Um tracejado muito cerrado dificulta o exame do gráfico;
- A escala horizontal deve ser lida da esquerda para a direita, e a vertical de baixo para cima;
- Os títulos e marcações do gráfico devem ser dispostos de maneira que sejam facilmente lidos, partindo da margem horizontal inferior ou da margem esquerda.

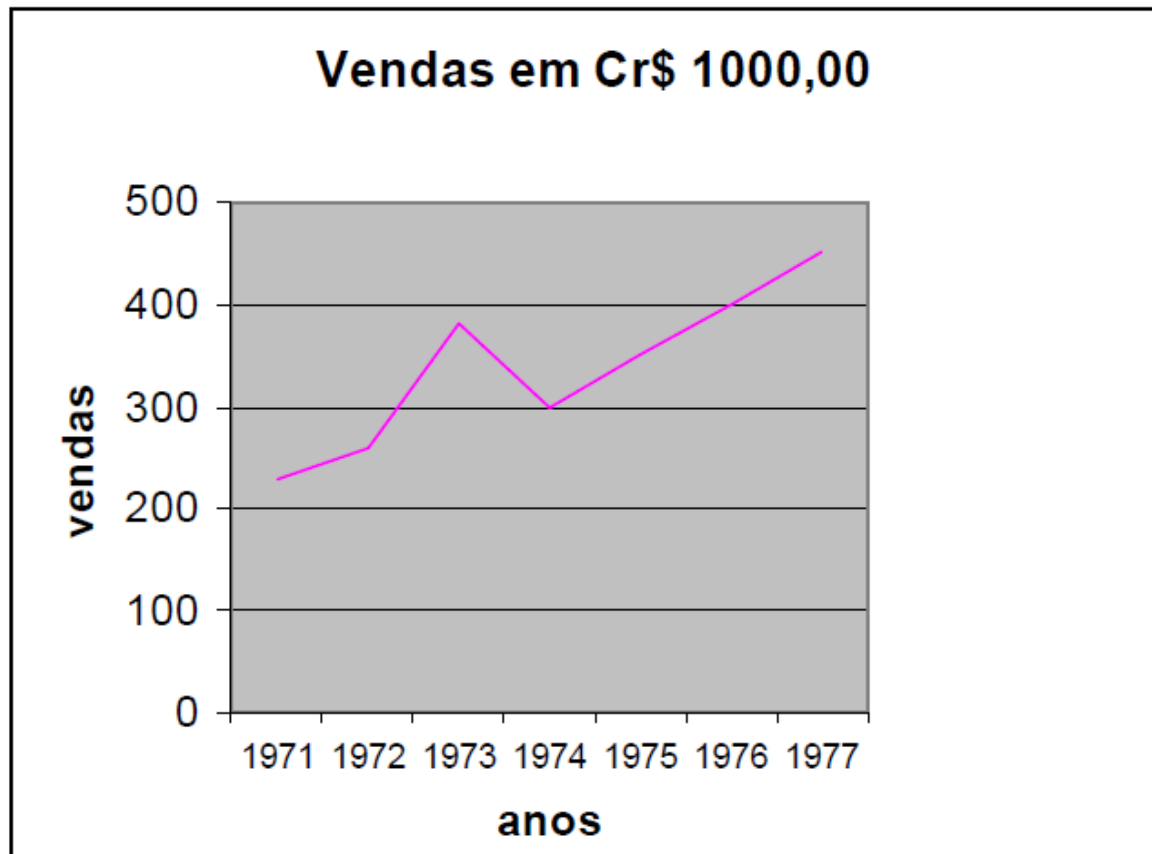
VISUALIZAÇÃO DE DADOS

Gráficos – Leitura e interpretação de um gráfico:

- Declarar qual o fenômeno ou fenômenos representados, a região considerada, o período de tempo, a fonte dos dados, etc;
- Examinar o tipo de gráfico escolhido, verificar se é o mais adequado, criticar a sua execução, no conjunto e nos detalhes;
- Analisar cada fenômeno separadamente, fazendo notar os pontos mais em evidência, o máximo e o mínimo, assim como as mudanças mais bruscas;
- Investigar se há uma “tendência geral” crescente ou decrescente ou, então, se o fato exposto é estacionário;
- Procurar descobrir a existência de possíveis ciclos periódicos, qual o período aproximado, etc.

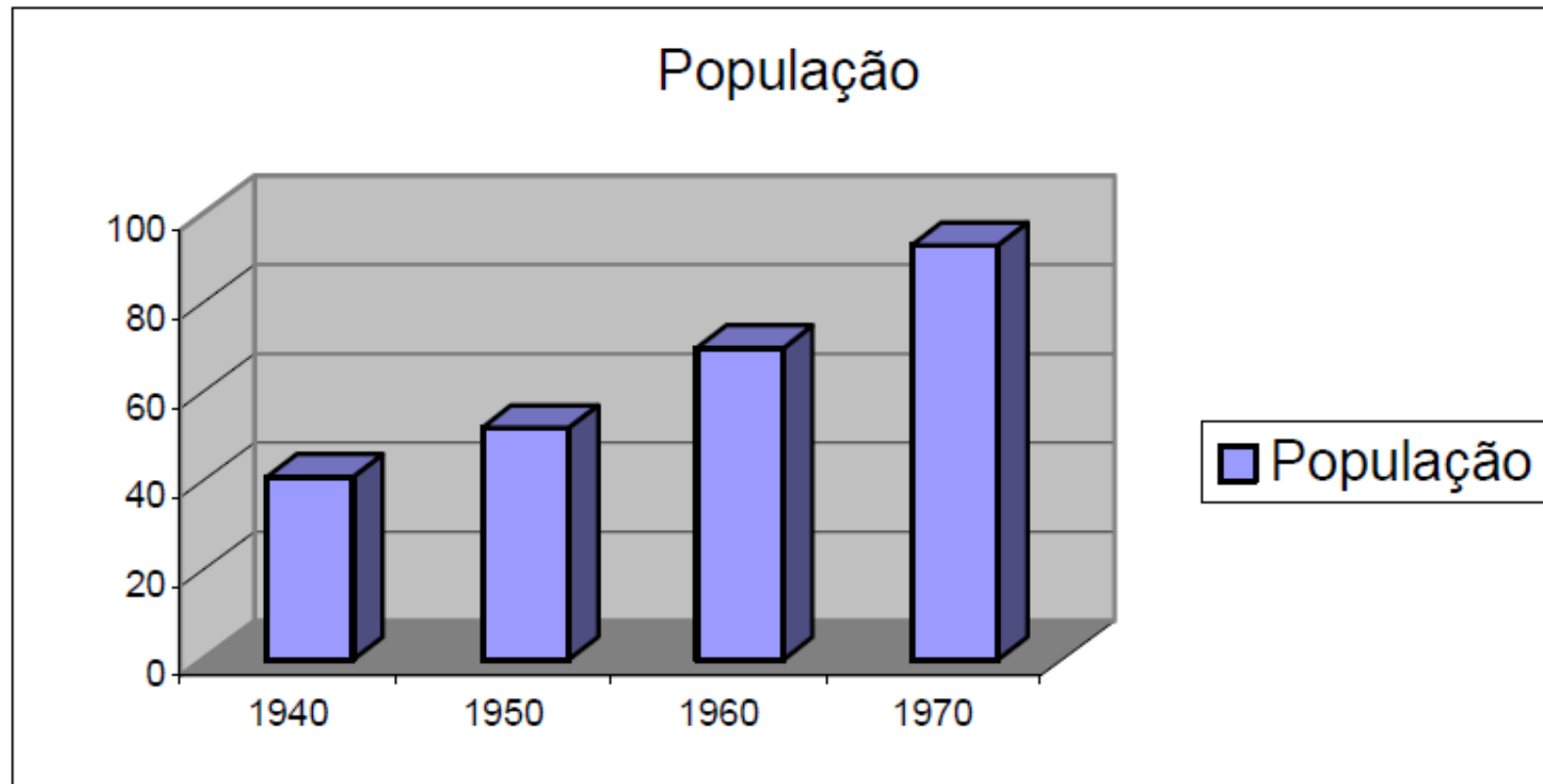
VISUALIZAÇÃO DE DADOS

Gráficos – Gráfico de Linhas



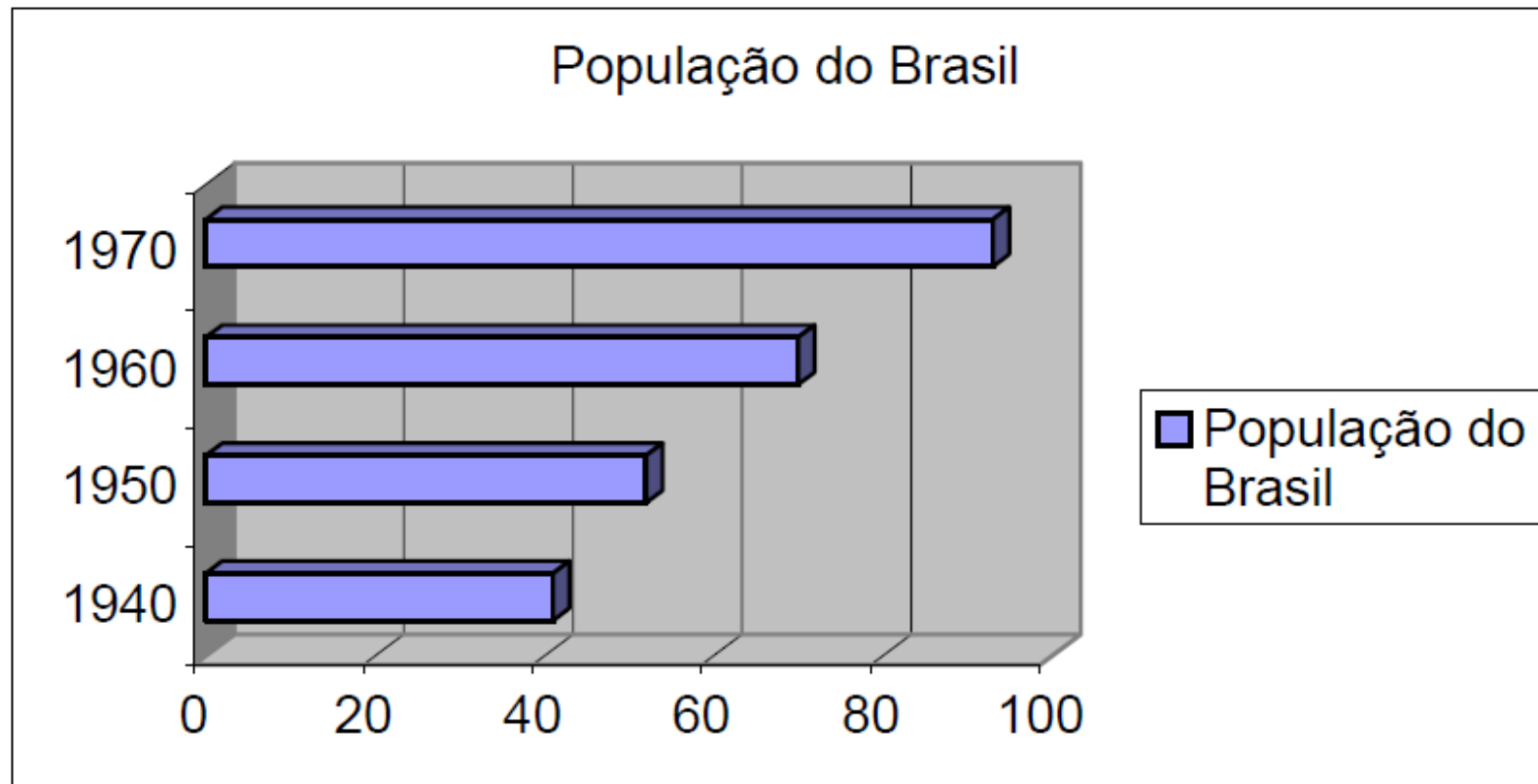
VISUALIZAÇÃO DE DADOS

Gráficos – Gráfico de Colunas



VISUALIZAÇÃO DE DADOS

Gráficos – Gráfico em Barras



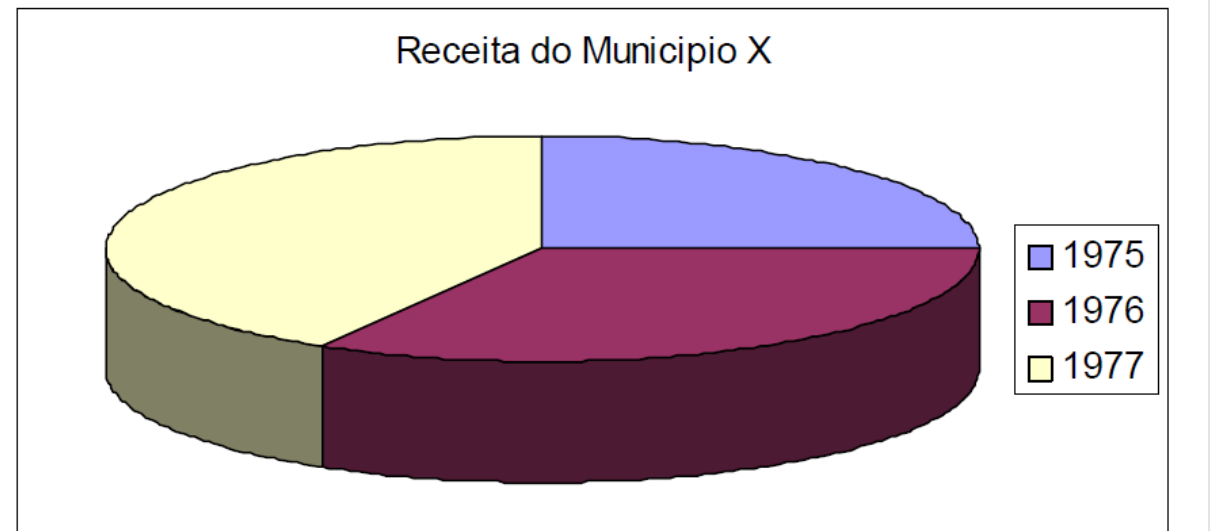
VISUALIZAÇÃO DE DADOS

Gráficos – Gráfico em Setores ou “Pizza”

Muito usado para apresentação de dados Qualitativos

Receita (em R\$ 1.000.000,00) do Município X de 1975-77

Anos	Receita (em R\$ 1.000.000,00)
1975	90
1976	120
1977	
Total	

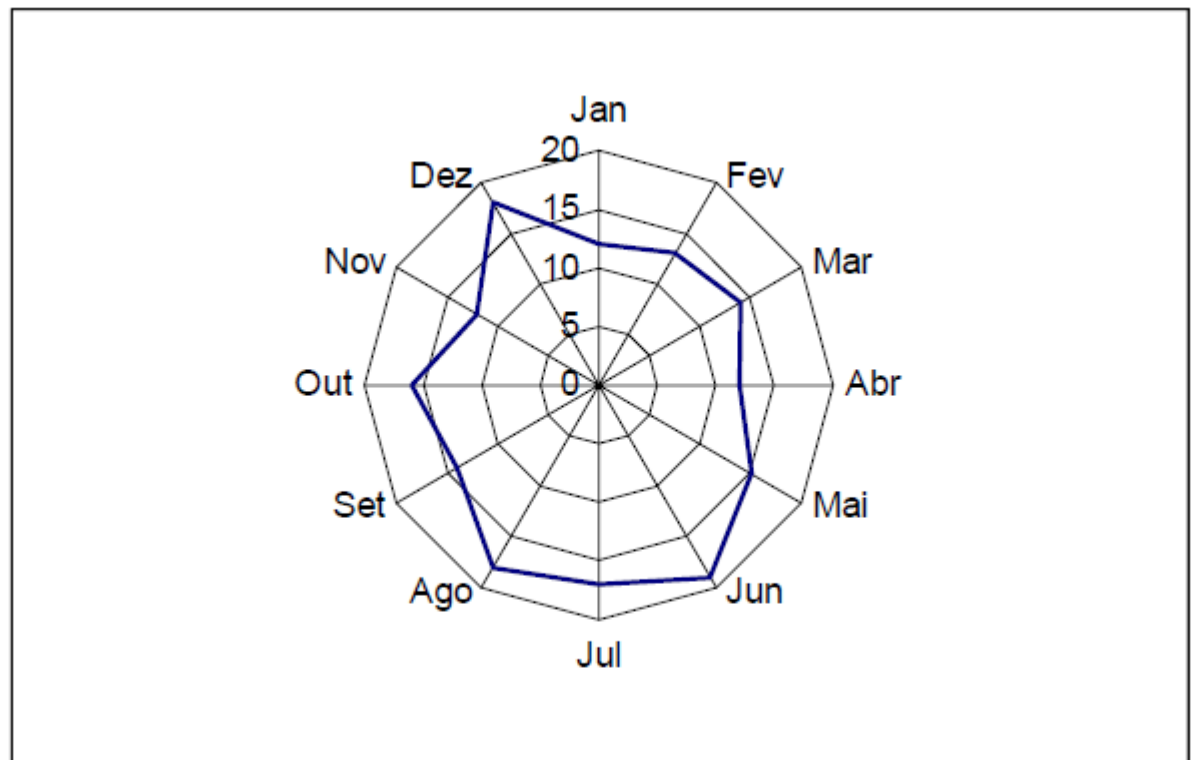


VISUALIZAÇÃO DE DADOS

Gráficos – Gráfico de Radar, Gráfico de Teia, Gráfico de Aranha, Gráfico de Estrela, Polígono Irregular, Gráfico Polar, ou Diagrama Kiviat.

Movimento Mensal de Compras de uma agencia em 1972

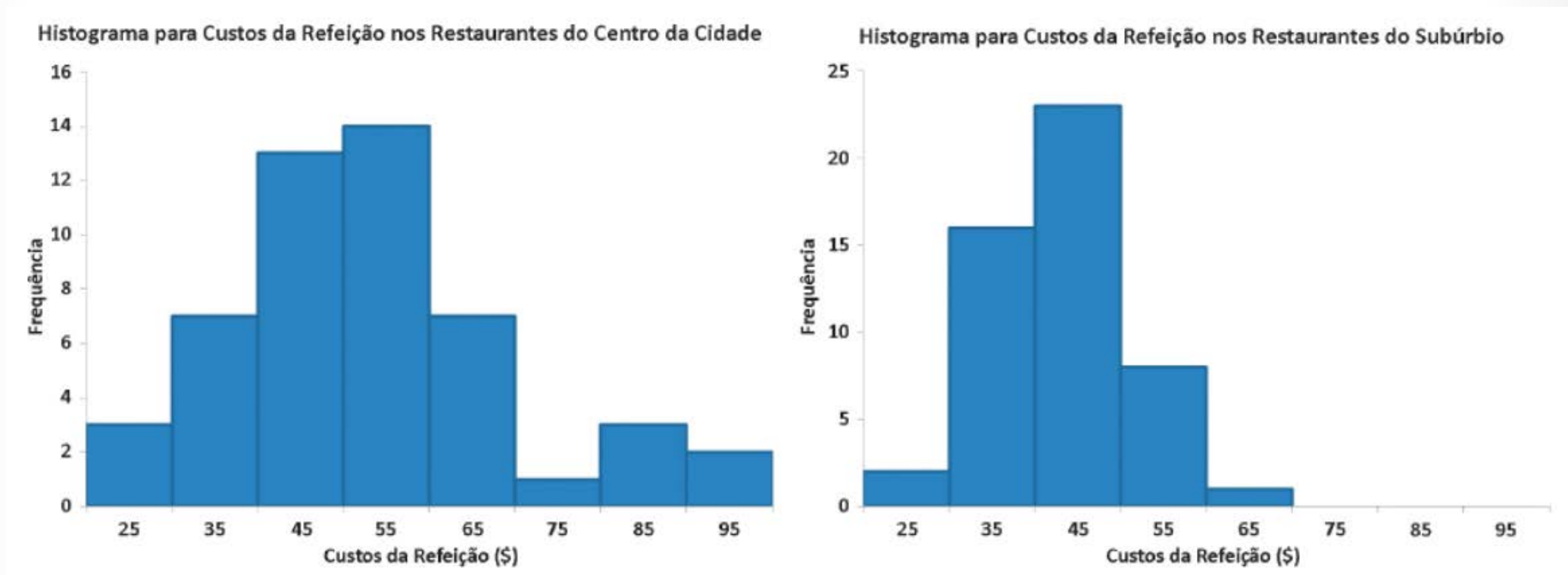
Meses	Valores (R\$1.000,00)
Janeiro	12
Fevereiro	13
Março	14
Abril	12
Maio	15
Junho	19
Julho	17
Agosto	18
Setembro	14
Outubro	16
Novembro	12
Dezembro	18



VISUALIZAÇÃO DE DADOS

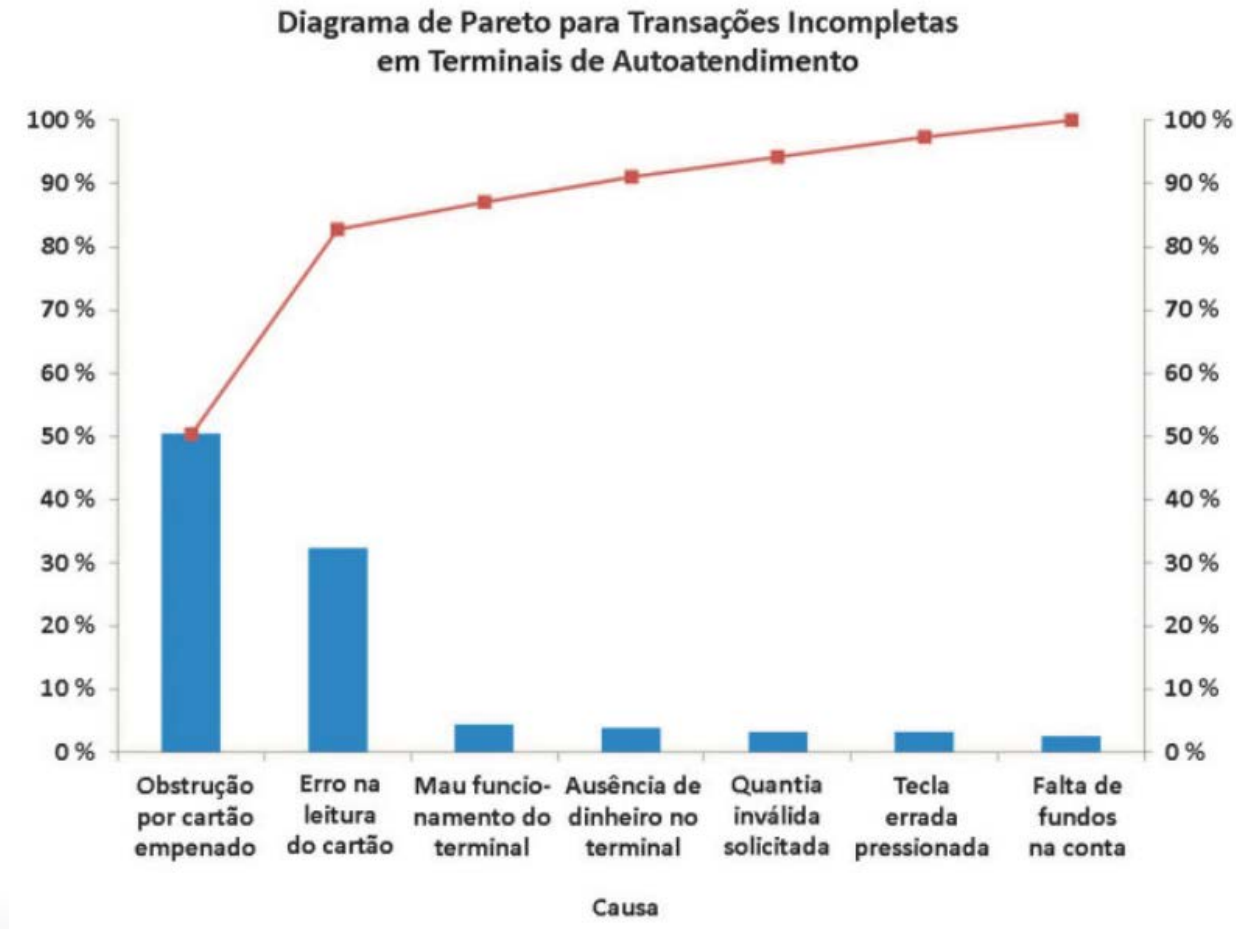
Outros Gráficos conhecidos: Histograma

Usado em Distribuição em Frequência



VISUALIZAÇÃO DE DADOS

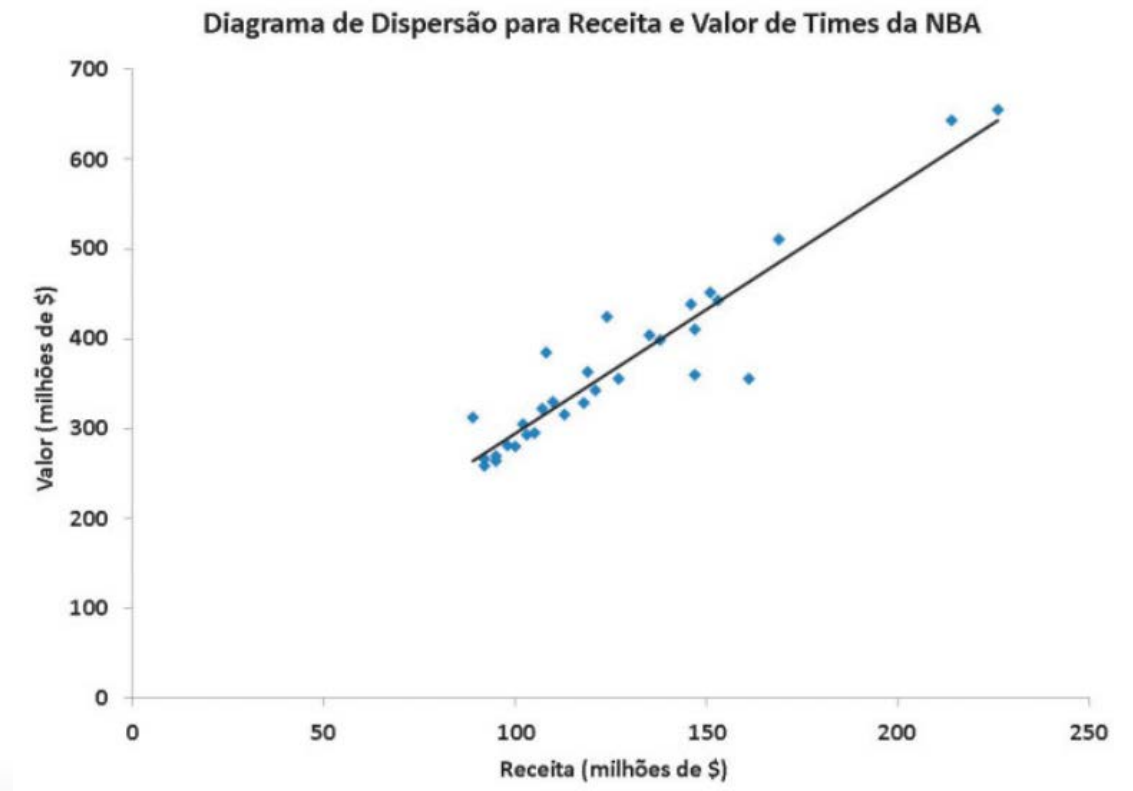
Outros Gráficos conhecidos: Diagrama de Pareto



VISUALIZAÇÃO DE DADOS

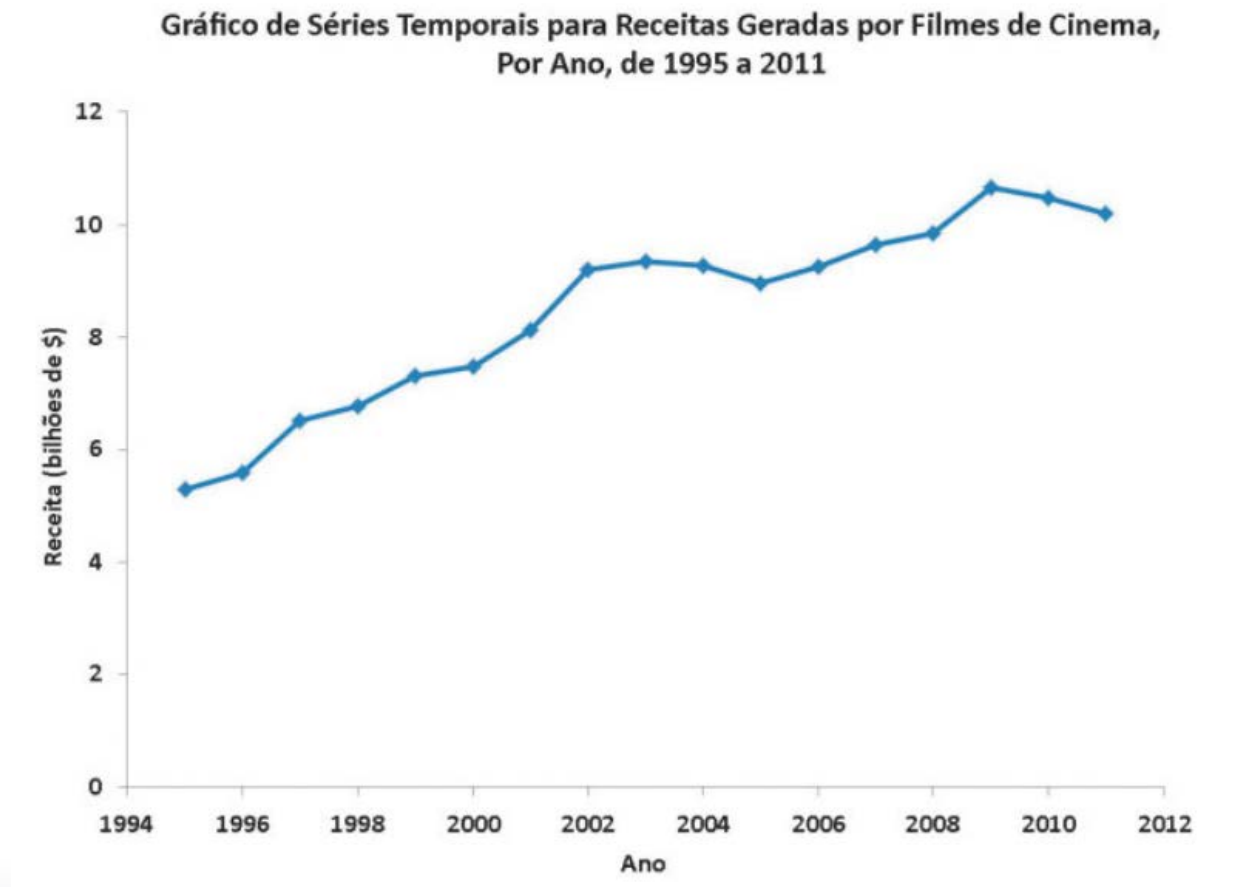
Outros Gráficos conhecidos: Diagrama de Dispersão

Usado em Correlações e Regressão Linear



VISUALIZAÇÃO DE DADOS

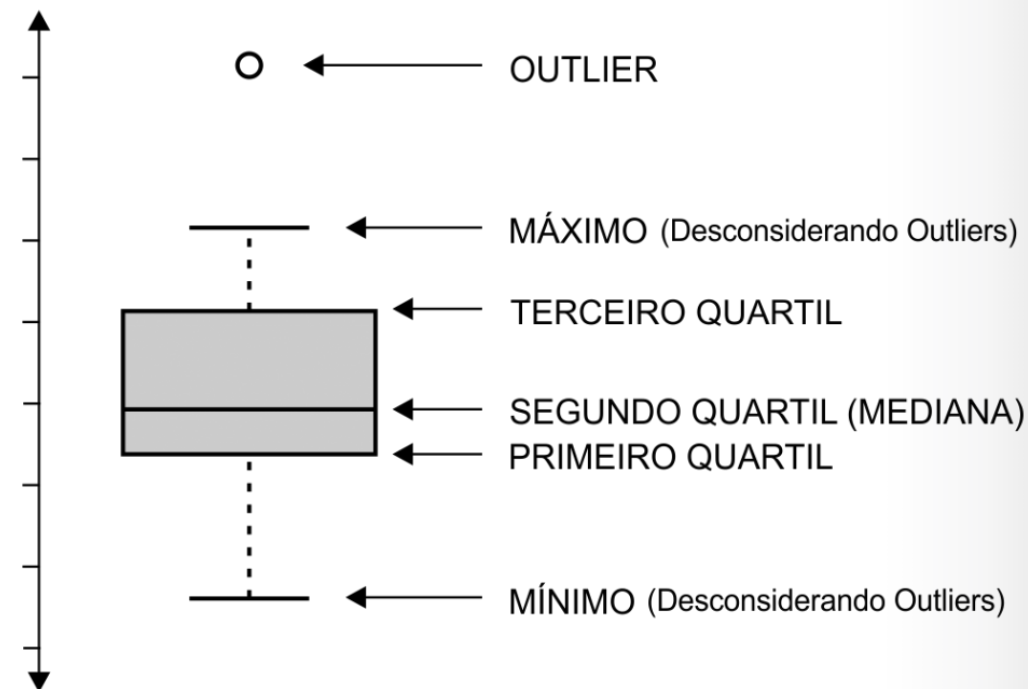
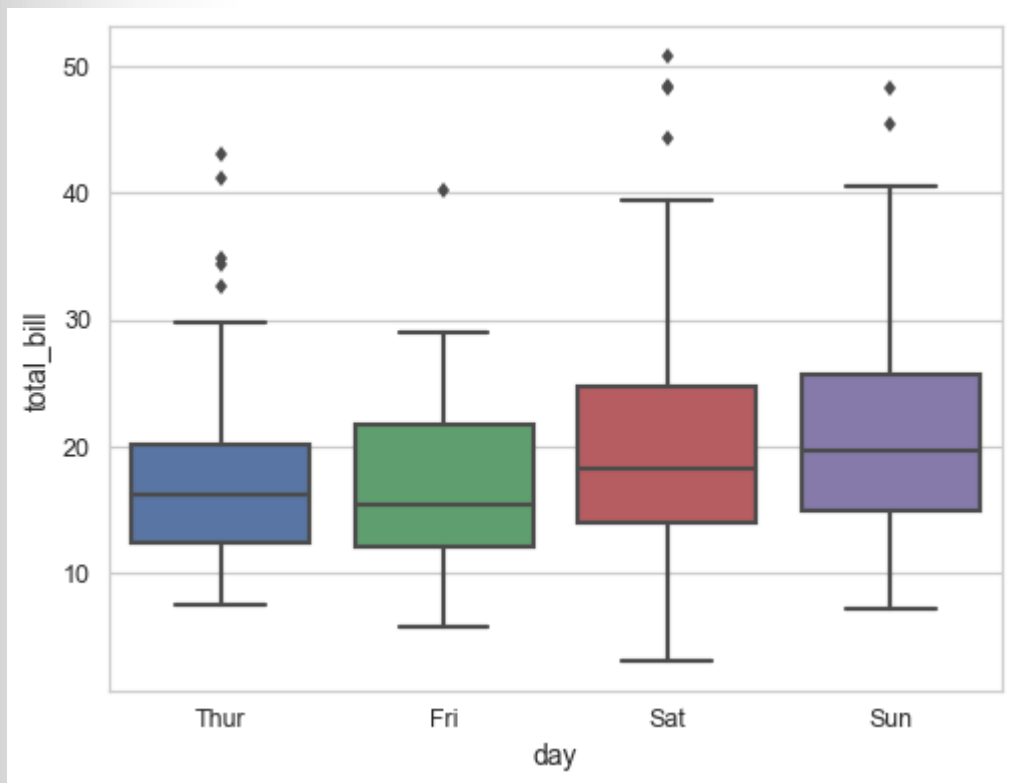
Outros Gráficos conhecidos: Séries Temporais



VISUALIZAÇÃO DE DADOS

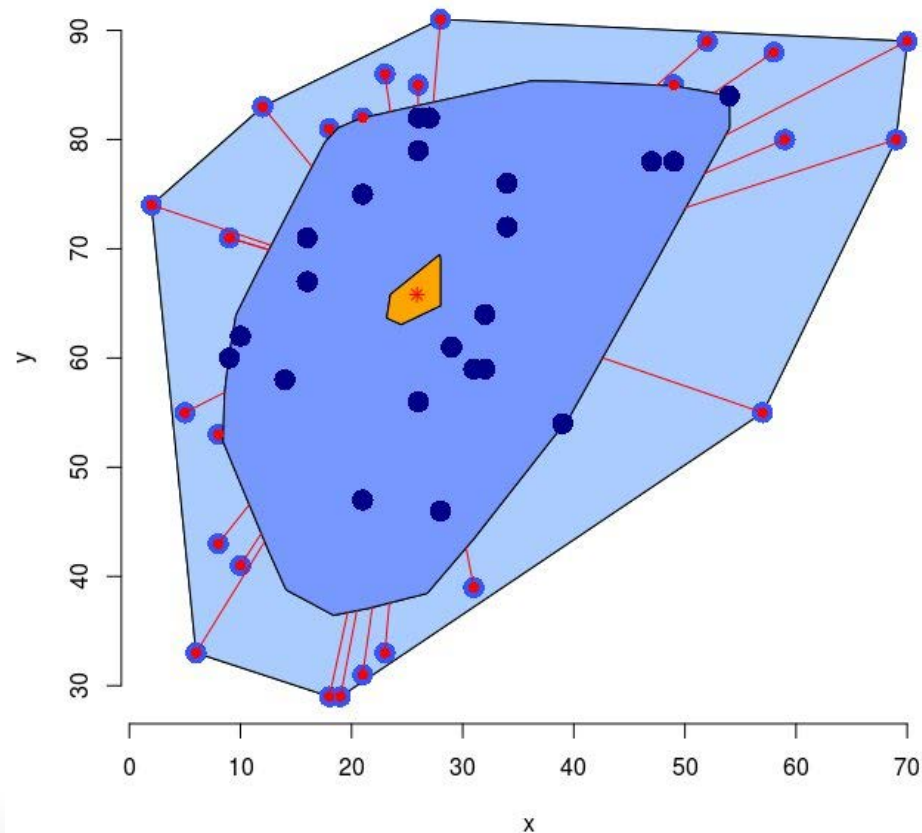
Outros Gráficos conhecidos: *Boxplot* ou Gráfico de Caixa

Usado na Estatística Descritiva para representar variações de uma variável



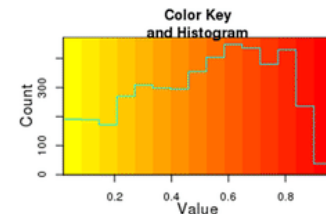
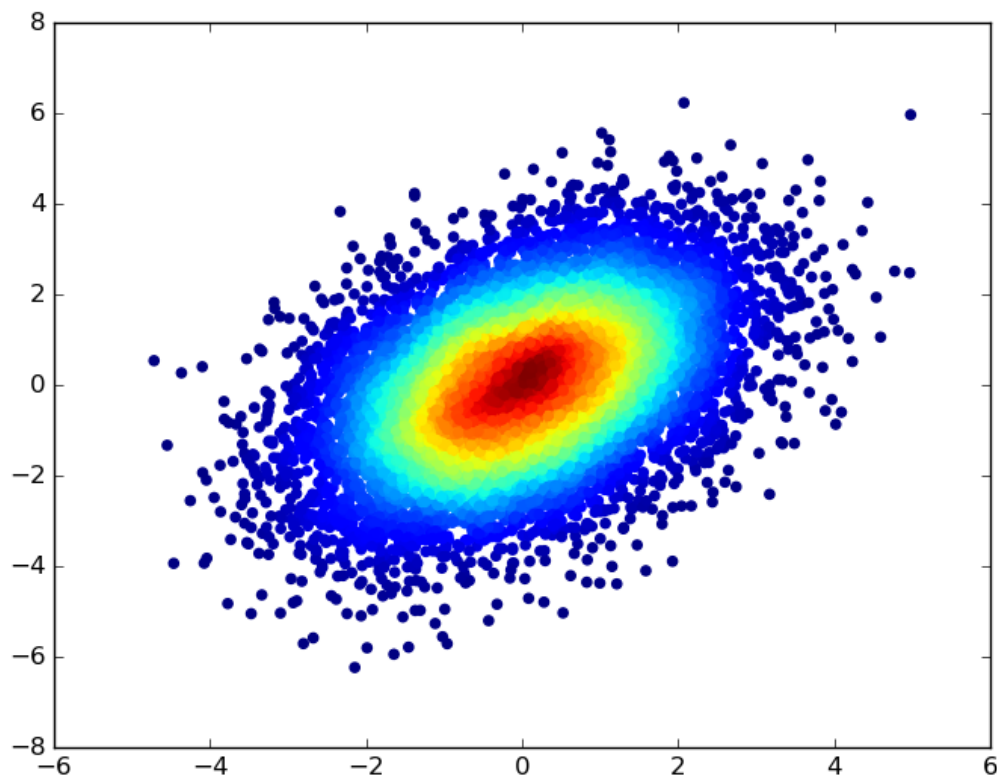
VISUALIZAÇÃO DE DADOS

Outros Gráficos conhecidos: *Bagplot*



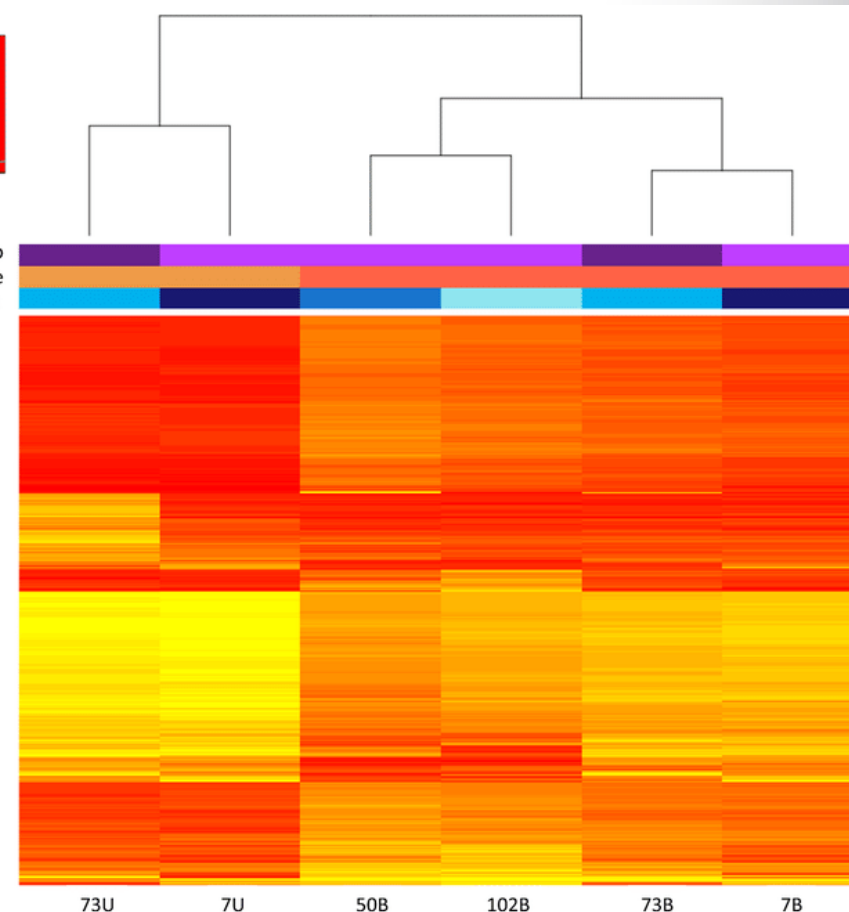
VISUALIZAÇÃO DE DADOS

Outros Gráficos conhecidos: *HeatMap*



Disease Group
Sample Type
Patient

- Group 3
- Group 4
- Urine
- Buffy coat
- Patient 102
- Patient 73
- Patient 50
- Patient 7



AMOSTRAGEM

AMOSTRAGEM

Amostragem - Definição:

- É o processo de obtenção de **amostras**, que são uma pequena parte de uma população à qual se pretende estudar.
- É a parte inicial de qualquer estudo estatístico.
- Consiste na escolha criteriosa dos elementos a serem submetidos ao estudo.
- Geralmente, as pesquisas são realizadas através de estudo dos elementos que compõem uma amostra, extraída da população que se pretende analisar.

AMOSTRAGEM

Amostragem – Conceitos Fundamentais:

O conceito de população é intuitivo; trata-se do conjunto de indivíduos ou objetos que apresentam em comum determinadas características definidas para o estudo.

- **Amostra** - é um subconjunto da população.
- **Amostragem** - são procedimentos para extração de amostras que representem bem a população.
- **Riscos** - é a margem de erro motivado pelo fato de investigarmos parcialmente (amostras) o universo (população).
- **População-alvo** - é a população sobre a qual vamos fazer inferências baseadas na amostra.

AMOSTRAGEM

Amostragem – Conceitos Fundamentais:

Quanto à extração dos elementos, as amostras podem ser:

- **Com reposição** - quando um elemento sorteado puder ser sorteado novamente;
- **Sem reposição** - quando o elemento sorteado só puder figurar uma única vez na amostra.

AMOSTRAGEM

Amostragem – Conceitos Fundamentais:

- O método de **amostragem probabilística** exige que cada elemento da população possua determinada probabilidade de ser selecionado. Normalmente, possuem a mesma probabilidade.
- Assim, se N for o tamanho da população, a probabilidade de cada elemento será $1/N$.
- Somente com base em amostragens probabilísticas pode-se realizar inferências sobre a população, a partir dos parâmetros estudados na amostra. São elas:
 - **Amostragem Aleatória Simples;**
 - **Amostragem Aleatória Estratificada;**
 - **Amostragem Sistemática;**
 - **Amostragem por Conglomerado.**

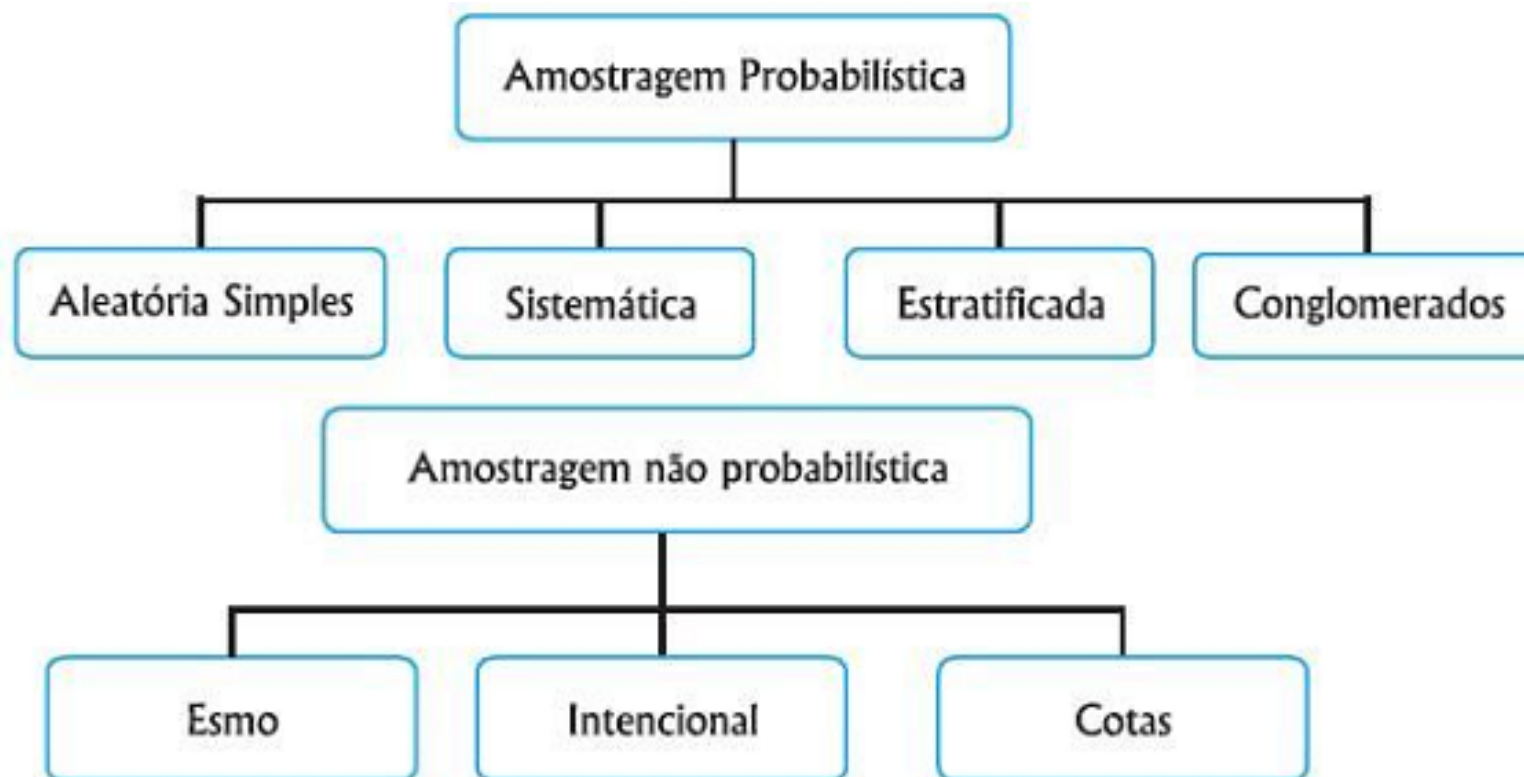
AMOSTRAGEM

Amostragem – Conceitos Fundamentais:

- Os **métodos não probabilísticos** são amostragens em que há uma escolha deliberada dos elementos que compõem a amostra.
- Não se pode generalizar os resultados das pesquisas para a população, uma vez que as amostras não probabilísticas não garantem a representatividade da população. São elas:
 - **Amostragem Acidental;**
 - **Amostragem Intencional;**
 - **Amostragem por Quotas.**

AMOSTRAGEM

Amostragem – Conceitos Fundamentais



Quadro resumo amostragem

AMOSTRAGEM

Amostragem – Probabilísticas:

- **Amostragem Aleatória Simples**: É um processo para selecionar amostras de tamanho “ n ” dentre as “ N ” unidades em que foi dividida a população.
- Sendo a amostragem realizada sem reposição, que é o caso mais comum, existem (N,n) possíveis amostras, todas igualmente prováveis.
- Na prática, a amostra aleatória simples é escolhida unidade por unidade.

Exemplo: um sorteio no qual se colocam todos os números misturados dentro de uma urna. As unidades correspondentes aos números escolhidos formarão a amostra.

AMOSTRAGEM

Amostragem – Probabilísticas:

- **Amostragem Aleatória Estratificada**: Uma amostra estratificada é obtida separando-se as unidades da população em grupos não superpostos chamados estratos, e selecionando-se independentemente uma amostra aleatória simples de cada estrato.
- Existem dois tipos de amostragem estratificada:
 - **De igual tamanho (uniforme);**
 - **Proporcional.**

AMOSTRAGEM

Amostragem – Probabilísticas:

Amostragem Aleatória Estratificada – Uniforme x Proporcional

- * Estrato 1 - População mexicana menor de 19 anos: 42,4 milhões (41,0%)
- * Estrato 2 - População mexicana de 20 a 44 anos: 37,6 milhões (36,3%)
- * Estrato 3 - População mexicana maior de 44 anos: 23,5 milhões (22,7%)

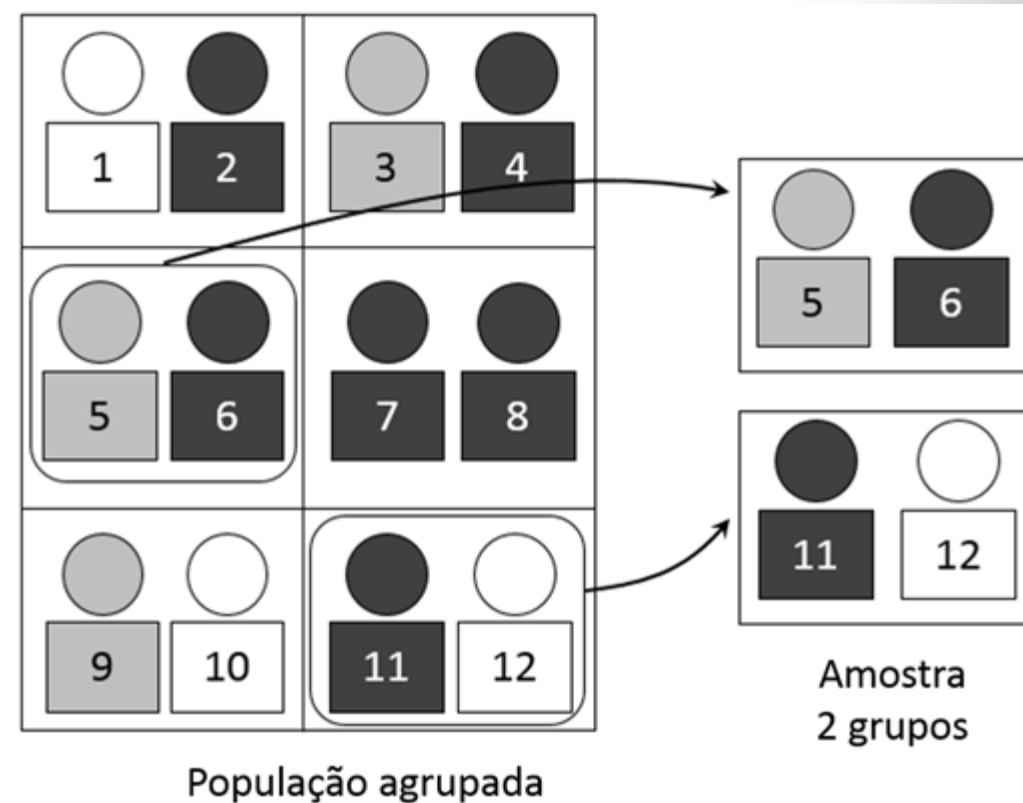
Estrato	População	Proporção	Amostra
1	42,4M	41,0%	334
2	37,6M	36,3%	333
3	23,5M	22,7%	333

Estrato	População	Proporção	Amostra
1	42,4M	41,0%	410
2	37,6M	36,3%	363
3	23,5M	22,7%	227

AMOSTRAGEM

Amostragem – Probabilísticas:

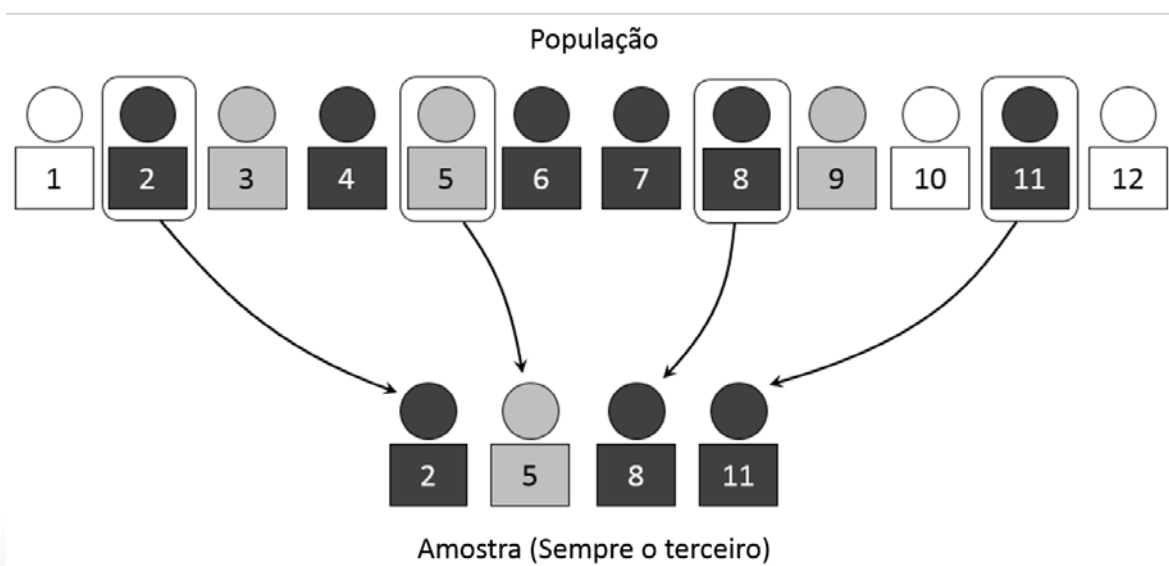
- **Amostragem por Conglomerado:** É uma técnica que explora existência de grupos (*clusters*) na população.
- Esses grupos representam adequadamente a população total em relação a característica que queremos medir.
- Em outras palavras, estes grupos contêm variabilidade da população inteira.
- Pode-se selecionar apenas alguns desses conglomerados para realizar o estudo



AMOSTRAGEM

Amostragem – Probabilísticas:

- **Amostragem por Conglomerado:** É uma técnica que requer certo controle do marco amostral entre os indivíduos selecionados junto com a probabilidade que sejam selecionados
- Consiste em escolher um indivíduo inicialmente de forma aleatória entre a população e, posteriormente, selecionar para amostra cada enésimo indivíduo disponível no marco amostral.



AMOSTRAGEM

Amostragem

Sistemática:

Diretrizes para calcular as amostras:

1º - Estabelecer o intervalo de amostragem K :

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

OBS: Para valores de $K=N/n$, arredondar para o valor inteiro menor.

2º - Iniciar aleatoriamente a composição da amostra.

$b \rightarrow$ início (n° de ordem inicial sorteado na TNA).

OBS: $0 < b \leq K$

3º - Composição da Amostra:

1º item $\rightarrow b$

2º item $\rightarrow b + K$

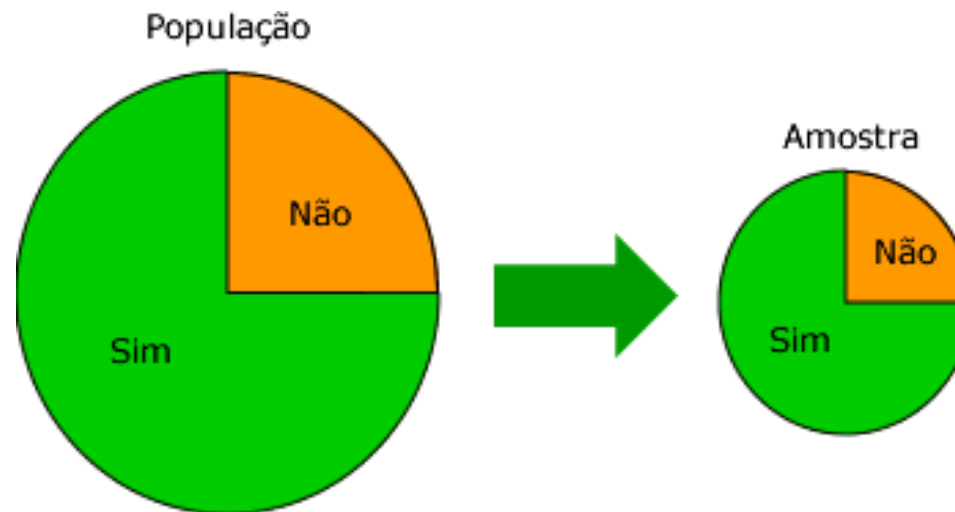
3º item $\rightarrow b + 2K$

AMOSTRAGEM

Amostragem – Não Probabilísticas:

➤ **Amostragem Acidental** - É formada por elementos que vão aparecendo, que são possíveis de se obter até completar o número de elementos da amostra.

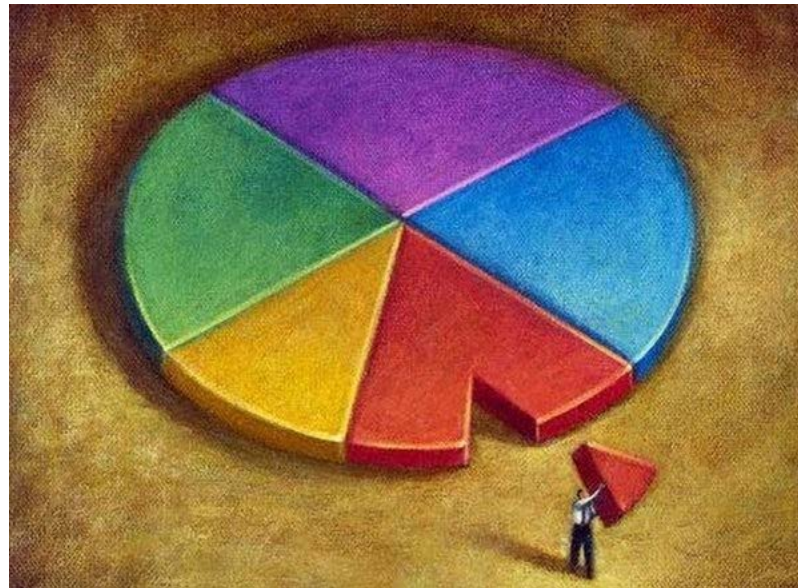
Ex: Pesquisa de opinião, em que os entrevistados são acidentalmente escolhidos.



AMOSTRAGEM

Amostragem – Não Probabilísticas:

- Amostragem Intencional - É formada por elementos escolhidos por determinado critério, ou seja, escolhe-se intencionalmente um grupo de elementos que irão compor a amostra.



AMOSTRAGEM

Amostragem – Não Probabilísticas:

- Amostragem por Cotas - Classificação da população em termos de propriedades que se sabe serem relevantes para a característica a ser estudada.
- Determinação da proporção da população para cada característica com base na constituição conhecida, ou estimada, da população.



DISTRIBUIÇÃO EM FREQUÊNCIA

DISTRIBUIÇÃO EM FREQUÊNCIA

- Nem sempre é possível compreender o significado contido numa amostragem por simples inspeção visual dos dados numéricos coletados. Entretanto, entendemos que o sucesso de uma decisão dependerá da nossa habilidade em compreender as informações contidas nesses dados.
- Uma forma eficiente de visualização de dados para determinados estudos é através das ***distribuições de frequências*** e dos **histogramas**.

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Elementos da Distribuição em Frequência:

Amplitude total (A) - é a diferença entre o maior e o menor número do rol.

Exemplo: Estatura de 40 alunos do Colégio A em cm. (Dados ordenados em ordem crescente, por colunas)

150	154	155	157	160	161	162	164	166	169
151	155	156	158	160	161	162	164	167	170
152	155	156	158	160	161	163	164	168	172
153	155	156	160	160	161	163	165	168	173

$$A = 173 - 150 = 23$$

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Elementos da Distribuição em Frequência:

Número de classes (K) e Classe (i) – não existe regra fixa para se determinar o número de classes. Podemos utilizar:

- A Regra de Sturges, que nos dá o número de classes em função do número de valores da variável:

$$K = 1 + 3,3 \cdot \log n, \text{ onde } n \text{ é o número de itens que compõe a amostra;}$$

- Ou

$$K = 5 \text{ para } n \leq 25 \text{ e } k \cong \sqrt{n}, \text{ para } n > 25.$$

Exemplo: considerando o exemplo anterior $n=40$

- Pela formula de Sturges: $K = 1 + 3,3 \log 40 = 6,28 \rightarrow K=6$
- Adotando $K = \sqrt{n}$, temos $k = \sqrt{40} = 6,3 \rightarrow K=6$

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Elementos da Distribuição em Frequência:

Amplitude de um intervalo de classe (h) – ou simplesmente intervalo de classe é a medida do intervalo que define a classe.

$$h = A / K$$

Exemplo, considerando o exemplo anterior:

$$H = 23 / 6 = 3,83 \rightarrow h = 4$$

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Elementos da Distribuição em Frequência:

Limites de Classe – denominamos limites de classe os extremos de cada classe. Assim temos:

- limite inferior (l_{inf}) e
- limite superior (L_{sup})

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Elementos da Distribuição em Frequência:

Exemplo: Do exemplo anterior, temos:

<i>i</i>	<i>Classes</i>	<i>n</i>	
1	150 — 154	4	
2	154 — 158	9	
3	158 — 162	11	
4	162 — 166	8	
5	166 — 170	5	
6	170 — 174	3	
	Σ	40	

Na segunda classe, temos:

- $L_2 = 158$
- $l_2 = 154$

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Gráficos Representativos da Distribuição em Frequência:

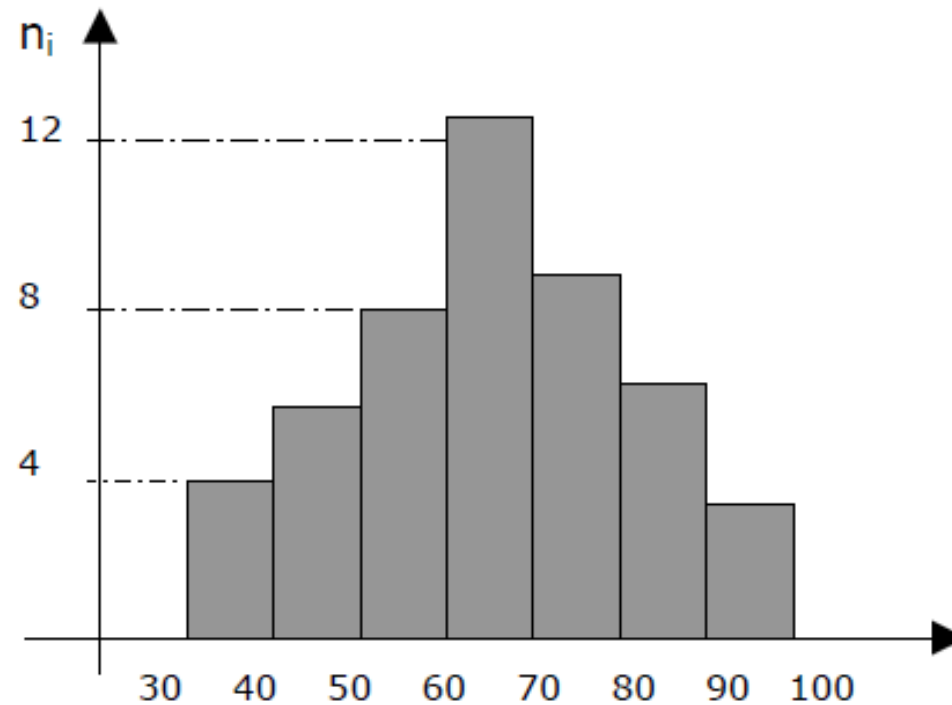
Histograma – é formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal, de tal modo que seus pontos médios coincidem com os pontos médios dos intervalos de classe.

- As larguras dos retângulos são iguais às amplitudes dos intervalos de classe.
- As alturas dos retângulos devem ser proporcionais às frequências das classes, sendo igual a amplitude dos intervalos.

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Gráficos Representativos da Distribuição em Frequência:

Histograma



DISTRIBUIÇÃO EM FREQUÊNCIA

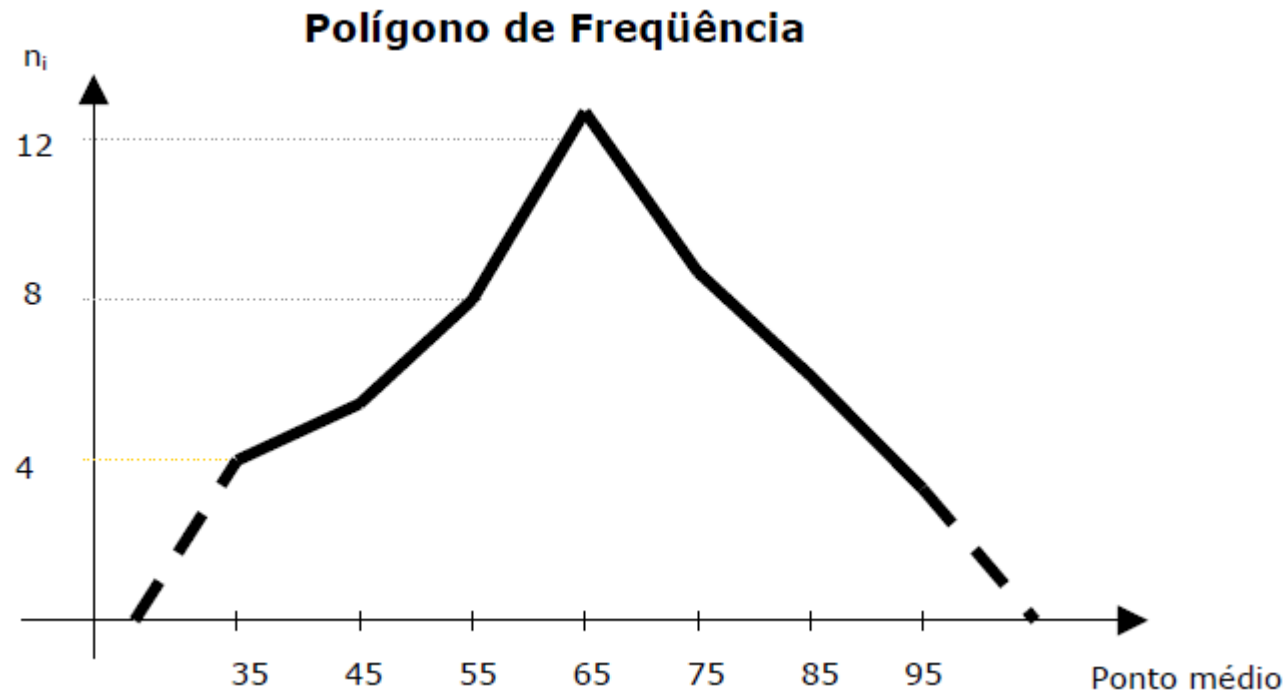
➤ Gráficos Representativos da Distribuição em Frequência:

Polígono de frequência – é um gráfico em linha, sendo as frequências marcadas sobre perpendiculares ao eixo horizontal, levantada pelos pontos médios dos intervalos de classe.

Para realmente obtermos um polígono (linha fechada), devemos completar a figura, ligando os extremos da linha obtida aos pontos médios da classe anterior à primeira e da posterior à última, da distribuição.

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Gráficos Representativos da Distribuição em Frequência:



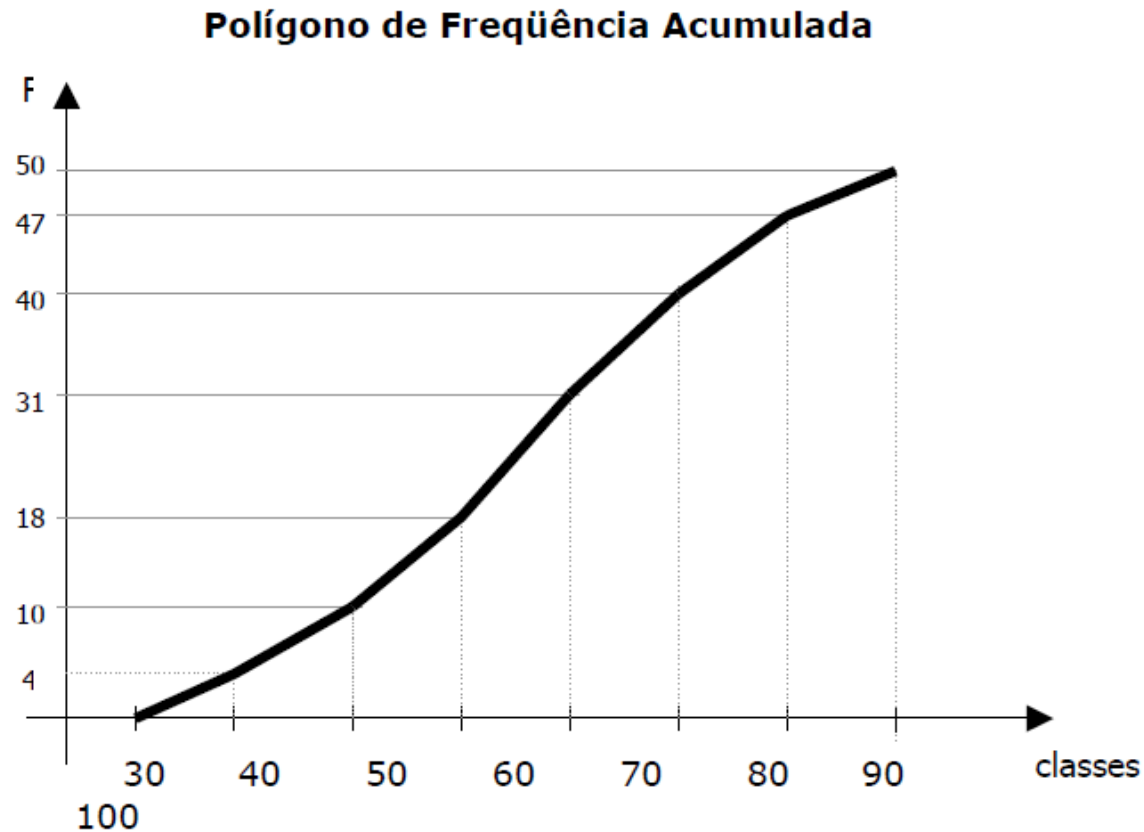
DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Gráficos Representativos da Distribuição em Frequência:

Polígono de frequência acumulada – é traçado marcando-se as frequências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.

DISTRIBUIÇÃO EM FREQUÊNCIA

➤ Gráficos Representativos da Distribuição em Frequência:



Dúvidas???

E-mail: estevamvilar@ic.ufal.br

Site da Disciplina: <http://sites.google.com/view/estevamvilar>

1. Materiais das Aulas
2. Links Recomendados
3. Discussões

