

ESTATÍSTICA II

Michelle Hanne Soares de Andrade

michellehanne@cefetmg.br

1º. SEMESTRE 2018

Apresentação

São 19 anos de experiência no segmento de TI, atuando na análise, coordenação e desenvolvimento de sistemas e projetos, além de atuar na área de docência do Ensino Superior. Possuo ampla formação acadêmica, mestrado e graduação em Ciência da Computação e sou especialista em Inteligência e Contrainteligência. Liderei e participei da equipe de implantação do Sistema Gestão em Saúde em Rede PBH e Sistema de Gestão de Convênios da Secretaria de Transporte e Obras Públicas de Minas Gerais. Atuei como micro empresária no desenvolvimento de sistemas para os segmentos de varejo, medicina estética e gestão comercial. Em docência, possuo mais de 10 anos de experiência em treinamentos de cursos de livre docência e mais de 5 anos no ensino superior.

Ementa

- Variável aleatória. Distribuições de probabilidades. Testes de hipóteses – média e desvio-padrão. Tabela de contingência. Aplicações com softwares e planilhas eletrônicas.

Referências

- BUSSAB, W, O; MORETTIN, P. A. **Estatística básica**. 6. ed. São Paulo: Saraiva, 2010.
- FREUND, J. E. **Estatística aplicada: economia, administração e contabilidade**. 11. ed. Porto Alegre: Bookman, 2006.
- NEUFELD, J. L. **Estatística aplicada à administração usando o Excel**. São Paulo: Pearson Prentice Hall, 2003.

Referências Complementares

- BRUNI, A. L. **Estatística aplicada à gestão empresarial**. 2. ed. São Paulo: Atlas, 2008.
- COSTA NETO, P. L. O. **Estatística**. São Paulo: Edgard Blücher, 1977.
- KAZMIER, Leonard J. **Estatística aplicada à economia e administração**. São Paulo: Makron Books, 1982.
- MEYER, P.L. **Probabilidade: aplicações à estatística**. 2. ed. Rio de Janeiro: LTC, 1984. 5 SPIEGEL, M. R.; SCHILLER, J.; SRUNIVASAN, R. A. **Probabilidade e estatística**. São Paulo: Makron Books, 1978.

Datas Importantes

- ❑ Prova 1 – 05 de Abril - 30 pontos
- ❑ Prova 2 – 10 de Maio - 30 pontos
- ❑ Prova 3 – 14 de Junho - 30 pontos
- ❑ *Pontos de Exercícios no Laboratório (Prática) – 10 pontos*
- ❑ Prova Substitutiva – 21/06*
- ❑ Exame Especial – 03/07

* Para quem perder uma das Provas ou desejar melhorar a nota. Será mantida a maior pontuação.

Noções de Métodos Estatísticos

- Do que trata a Estatística? A palavra estatística provem de “Status”.
- **A essência da ciência é a observação. A ciência que se preocupa com a organização, descrição, análise e interpretação dos dados experimentais e denominada de Estatística, um ramo da Matemática Aplicada.**

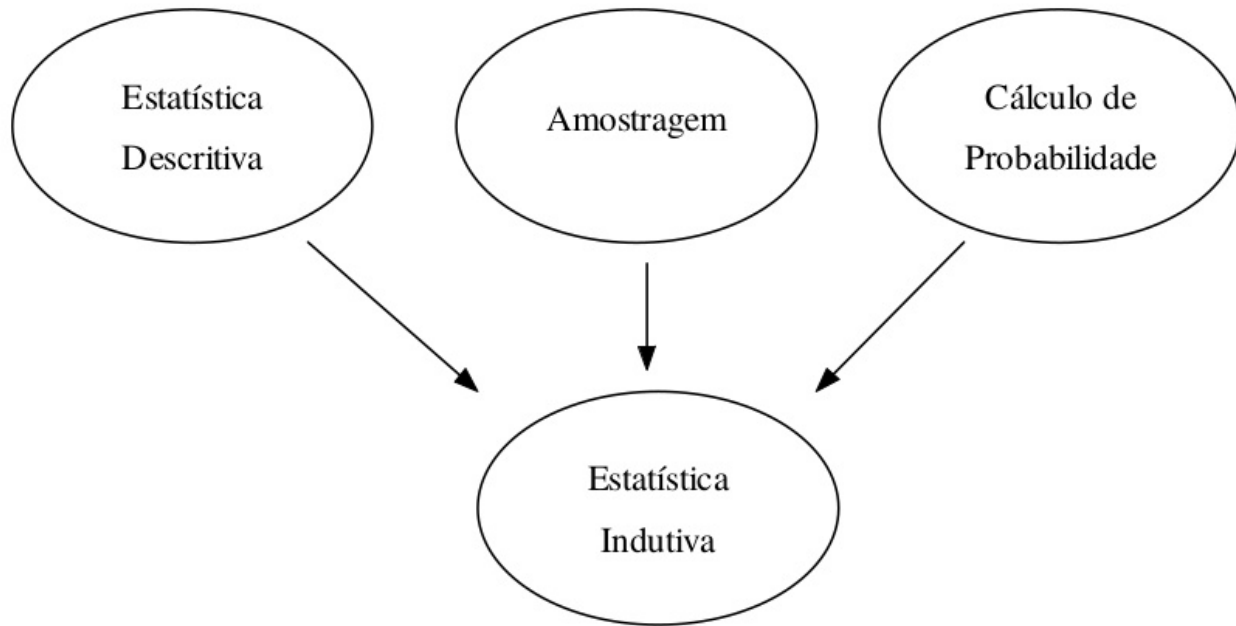
Noções de Métodos Estatísticos

O campo da estatística lida com a **coleta, apresentação, análise e uso dos dados** para **tomar decisões**, resolver problemas e planejar produtos e processos.

Qualquer tomada de **decisão e/ou conclusão** observada que seja resultado de **estudo de dados estatísticos** será tão eficiente quanto o processo utilizado para obtenção desses dados.

Noções de Métodos Estatísticos

- **Estatística Descritiva e Estatística Indutiva (ou Inferencial).**



Estatística Descritiva

A **Estatística Descritiva** Resume ou descrição das características importantes de um conjunto conhecido de dados populacionais... está relacionada com a organização e descrição de dados associada a *cálculos de médias, variâncias, estudo de gráficos, tabelas, etc.*

Estatística Descritiva

- **Com a Estatística Descritiva** entendemos melhor um conjunto de dados através de suas características.
- As três principais características são:
 - Um valor representativo do conjunto de dados. Ex.: uma media, Mediana, Moda, etc...
 - Uma medida de dispersão ou variação.
 - A natureza ou forma da distribuição dos dados: sino, uniforme, assimétrica...

Probabilidades

O Cálculo de Probabilidades é a essência dos modelos Não-Determinísticos e a corroboração de que toda inferência estatística está sujeita a erros.

Tais fenômenos precisam de um modelo matemático diferente para seu estudo. São os conhecidos modelos não-determinísticos ou probabilísticos.

Exemplo de um modelo não determinístico: O lançamento de um dado não viciado, sempre sob as mesmas condições. Não é possível afirmar com exatidão qual será a face observada a cada novo lançamento.

Amostragem

- A **Amostragem** é o ponto de partida para todo um Estudo Estatístico. Aqui pode se ter origem um problema bastante comum em Engenharia: “Análise profunda sobre dados superficiais!”
- Três métodos importantes para a coleta de dados:
 - Um estudo **retrospectivo** usando dados históricos
 - Um estudo de **observação**
 - Um **experimento planejado**

Estatística Indutiva (Inferencial)

A **Estatística Indutiva** é o objetivo básico da ciência. A ela esta associada, *Estimação de Parâmetros, Testes de Hipóteses, Modelamento, etc..* Generalização sobre uma população tomadas a partir da utilização de dados amostrais.

Estatística Indutiva - Exemplo

- **Variabilidade** no consumo de combustível de um veículo. Potenciais fontes de variabilidade do sistema:
 - Tipo de condições de estrada (cidade ou rodovia)
 - Mudanças nas condições do veículo ao longo do tempo (pressão dos pneus, compressão, desgaste de válvulas)
 - Marca ou categoria de gasolina utilizada
 - Condições meteorológicas
 - Modo de dirigir
 - Outros fatores
- **Estatística: fornece uma teoria para analisar estas fontes de variabilidade (quais as mais importantes?) de forma quantitativa.**



Organização de Dados Estatísticos

Planejamento de um estudo Estatístico

- 1- Identifique a variável (variáveis) de interesse (foco) e a população do estudo.
- 2- Desenvolva um plano detalhado para a coleta de dados. Se usar uma amostra, tenha certeza de que a amostra representa a população.
- 3- Colete adequadamente os dados.

coleta contínua: registros de nascimento, óbitos, casamentos;

coleta periódica: recenseamento demográfico, censo industrial;

coleta ocasional: registro de casos de dengue.

coleta Indireta: É feita por deduções a partir dos elementos conseguidos pela coleta direta, por analogia, por avaliação, indícios ou proporcionalização.

Planejamento de um estudo Estatístico

4- Apuração dos Dados: Resumo dos dados através de sua contagem e agrupamento. É a condensação e tabulação de dados.

5 - Apresentação dos Dados: Há duas formas de apresentação, que não se excluem mutuamente. A *apresentação tabular*, ou seja é uma apresentação numérica dos dados em linhas e colunas distribuídas de modo ordenado, segundo regras práticas fixadas pelo Conselho Nacional de Estatística. A *apresentação gráfica* dos dados numéricos constitui uma apresentação geométrica permitindo uma visão rápida e clara do fenômeno.

6 - Análise e Interpretação dos Dados: A última fase do trabalho estatístico é a mais importante e delicada. Está ligada essencialmente ao cálculo de medidas e coeficientes, cuja finalidade principal é descrever o fenômeno (estatística descritiva).

Principais Conceitos em Estatística

- **População:**

- E todo o conjunto de elementos que possuam ao menos uma característica comum observável.

- **Amostra:**

- E uma parte da população que será avaliada por um critério comum.

- **Dados estatísticos:**

- São os valores associados as variáveis de pesquisas.

- **Frequências:**

- O numero de vezes em que a variável ocorre e chamado **frequência absoluta** e é indicado por n_i .
- Definimos **frequência relativa** (f_i) como a razão entre a frequência absoluta (n_i) e o numero total de observações (n) , ou seja:

$$f_i = n_i/n$$

Principais Conceitos em Estatística

- **Parâmetros:**

- São valores singulares que existem na população e que servem para caracterizá-la. Para definirmos um parâmetro devemos examinar toda a população. *Ex: Os alunos do CEFET têm em média 1,70 metros de estatura.*

- **Estimativa:**

- É um valor aproximado do parâmetro e é calculado com o uso da amostra.

- **Atributo:**

- Quando os dados estatísticos apresentam um caráter qualitativo, o levantamento e os estudos necessários ao tratamento desses dados são designados genericamente de estatística de atributo.

- **Variável:**

- É o conjunto de resultados possíveis de um fenômeno.

Principais Conceitos em Estatística

- **Variável Qualitativa:**

- Quando seus valores são expressos por **atributos**: sexo, cor da pele, etc.

- **Variável Quantitativa:**

- Quando os **dados são de caráter nitidamente quantitativo**, e o conjunto dos resultados **possui uma estrutura numérica**, trata-se portanto da estatística de variável e se dividem em :
 - **Variável discreta ou descontínua**
 - **Variável contínua**

Principais Conceitos em Estatística

- **Variável Discreta ou Descontínua:** Seus valores são expressos geralmente através de números inteiros não negativos. Resulta normalmente de contagens.
 - *Ex:* N^o de alunos presentes às aulas de introdução à estatística econômica no 1^o semestre de 1997: mar = 18 , abr = 30 , mai = 35 , jun = 36.
- **Variável Contínua:** Resulta normalmente de uma **mensuração**, e a **escala numérica de seus possíveis valores corresponde ao conjunto \mathbb{R}** dos números Reais, ou seja, podem assumir, teoricamente, qualquer valor entre dois limites.
 - *Ex.:* Quando você vai medir a temperatura de seu corpo com um termômetro de mercúrio o que ocorre é o seguinte: O filete de mercúrio, ao dilatar-se, passará por todas as temperaturas intermediárias até chegar na temperatura atual do seu corpo.

Exemplos de Variáveis

- Cor dos olhos das alunas: **qualitativa**
- Índice de liquidez nas indústrias: **quantitativa contínua**
- Produção de café no Brasil: **quantitativa contínua**
- Número de defeitos em aparelhos de TV: **quantitativa discreta**
- Comprimento dos pregos produzidos por uma empresa: **quantitativa contínua**
- O ponto obtido em cada jogada de um dado: **quantitativa discreta**



Revisão Geral Estatística I

Métodos Probabilísticos

- 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 ser selecionado será **$1/N$** .
- Trata-se do método que garante cientificamente a aplicação das técnicas estatísticas de inferências. Somente com base em amostragens probabilísticas é que se podem realizar inferências ou induções sobre a população a partir do conhecimento da amostra.
- **É uma técnica especial para recolher amostras, que garantem, tanto quanto possível, o acaso na escolha.**

Amostragem Casual ou Aleatória Simples

- É o processo mais elementar e frequentemente utilizado.
- É **equivalente a um sorteio lotérico**. Pode ser realizada numerando-se a população de 1 a n e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, x números dessa sequência, os quais corresponderão aos elementos pertencentes à amostra.
 - *Ex: Vamos obter uma amostra, de 10%, representativa para a pesquisa da estatura de 90 alunos de uma escola:*
 - *1º - numeramos os alunos de 1 a 90.*
 - *2º - escrevemos os números dos alunos, de 1 a 90, em pedaços iguais de papel, colocamos na urna e após mistura retiramos, um a um, nove números que formarão a amostra.*

Amostragem Proporcional Estratificada

- Quando a população se divide em estratos (sub-populações), convém que o sorteio dos elementos da amostra leve em consideração tais estratos, daí obtemos os elementos da amostra proporcional ao número de elementos desses estratos.
 - Ex: *Vamos obter uma amostra proporcional estratificada, de 10%, do exemplo anterior, supondo, que, dos 90 alunos, 54 sejam meninos e 36 sejam meninas. São portanto dois estratos (sexo masculino e sexo feminino). Logo, temos:*

SEXO	POPULAÇÃO	10 %	AMOSTRA
MASC.	54	5,4	5
FEMIN.	36	3,6	4
Total	90	9,0	9

Amostragem por Conglomerados (ou Agrupamentos)

- **Algumas populações não permitem, ou tornam extremamente difícil que se identifiquem seus elementos.** Não obstante isso, pode ser relativamente fácil **identificar alguns subgrupos da população.** Em tais casos, uma amostra aleatória simples desses subgrupos (conglomerados) pode se colhida, e uma contagem completa deve ser feita para o conglomerado sorteado. Agrupamentos típicos são bairros, famílias, organizações, agências, edifícios etc.
 - *Ex: Num levantamento da população de determinada cidade, podemos dispor do mapa indicando cada bairro e não dispor de uma relação atualizada dos seus moradores. Pode-se, então, colher uma amostra dos bairros e fazer a contagem completa de todos os que residem naqueles bairros sorteados.*

Métodos Não Probabilísticos

- São amostragens em que há uma **escolha deliberada dos elementos da amostra**. Não é possível generalizar os resultados das pesquisas para a população, pois as ***amostras não-probabilísticas não garantem a representatividade da população***.
 - **Amostragem Acidental:** Ex: *Pesquisas de opinião em praças públicas, ruas de grandes cidades;*
 - **Amostragem Intencional:** é escolhido intencionalmente um grupo de elementos que irão compor a amostra.
 - **Amostragem por Quotas:** Ex: *Numa pesquisa sobre o "trabalho das mulheres na atualidade", provavelmente se terá interesse em considerar: a divisão cidade e campo, a habitação, o número de filhos, a idade dos filhos, a renda média, as faixas etárias etc.*

Séries Estatísticas

- **Tabela:** É um quadro que resume um conjunto de dados dispostos segundo **linhas** e **colunas** de maneira sistemática.
- De acordo com a Resolução 886 do IBGE, nas casas ou células da tabela devemos colocar :
 - um traço horizontal (-) quando o valor é **zero**;
 - três pontos (...) quando não temos os dados;
 - zero (**0**) quando **o valor é muito pequeno** para ser expresso pela unidade utilizada;
 - um ponto de interrogação (?) quando temos dúvida quanto à exatidão de determinado valor.
- **Obs:** O lado direito e esquerdo de uma tabela oficial deve ser aberto.

Séries Estatísticas

- **Série Estatística:** É qualquer tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da **época**, do **local** ou da **espécie**.
- **Séries Homógradas:** São aquelas em que a variável descrita apresenta variação discreta ou descontínua. Podem ser do tipo temporal, geográfica ou específica.
 - **Série Temporal:** Identifica-se pelo **caráter variável do fator cronológico**. O local e a espécie (fenômeno) são elementos fixos. Esta série também é chamada de **histórica ou evolutiva**.

ABC VEÍCULOS LTDA. - Vendas no 1º bimestre de 2016

PERÍODO	UNIDADES VENDIDAS
JAN/96	20000
FEV/96	10000
TOTAL	30000

Séries Estatísticas

- **Série Específica:** O caráter variável é apenas o fato ou espécie. Também é chamada de série **categórica**

ABC VEÍCULOS LTDA. - Vendas no 1º bimestre de 2016

MARCA	UNIDADES VENDIDAS *
FIAT	18000
GM	12000
TOTAL	30000

- **Séries Conjugadas:** Também chamadas de **tabelas de dupla entrada**. São apropriadas à apresentação de duas ou mais séries de maneira conjugada, havendo duas ordens de classificação: uma horizontal e outra vertical. O exemplo abaixo é de uma série **geográfica-temporal**.

Séries Estatísticas

ABC VEÍCULOS LTDA. - Vendas no 1º bimestre de 2016

FILIAIS	Janeiro/96	Fevereiro/96
São Paulo	10000	3000
Rio de Janeiro	12000	5000
TOTAL	22000	8000



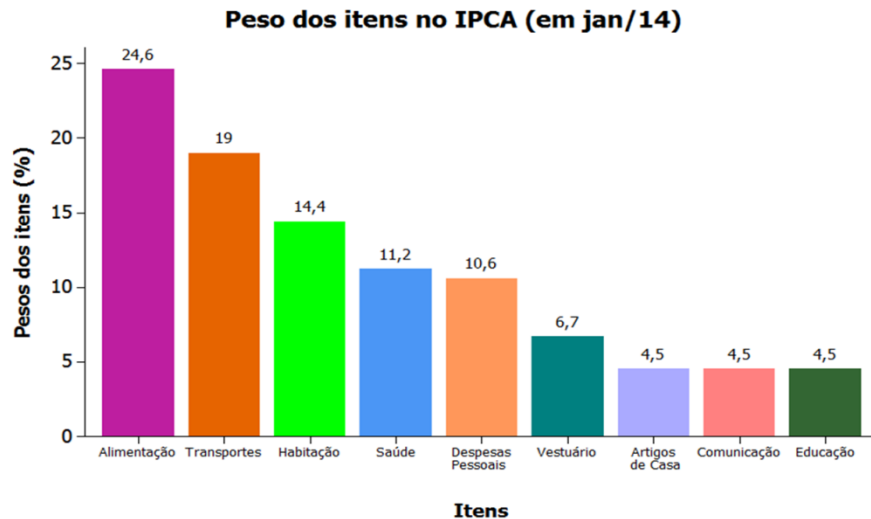
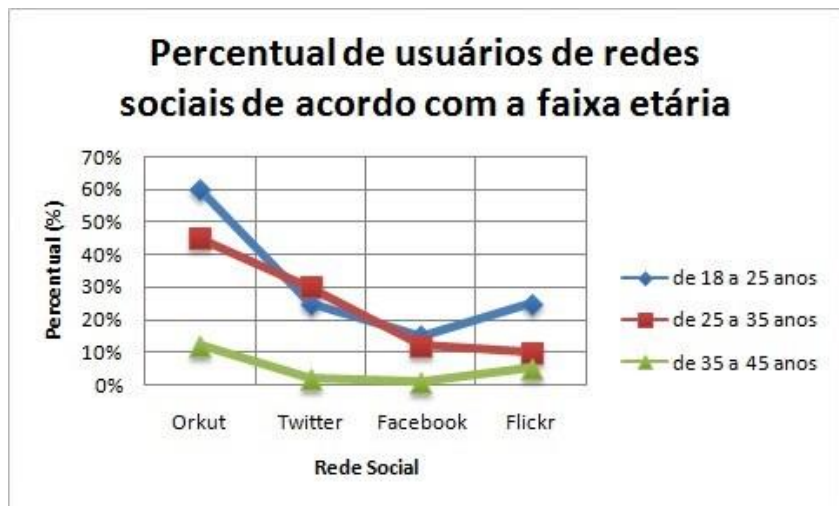
Representação Gráfica de uma Distribuição

Gráficos

- São representações visuais dos dados estatísticos que devem corresponder, mas nunca substituir as tabelas estatísticas.
- Características: Uso de escalas, sistema de coordenadas, simplicidade, clareza e veracidade.
- **Classificação dos gráficos**: Diagramas, Estereogramas, Pictogramas e Cartogramas.

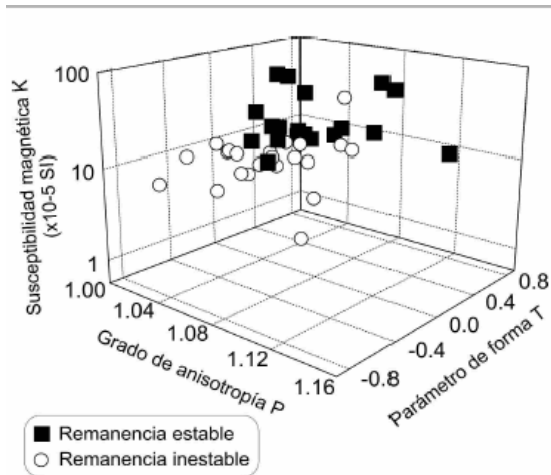
Gráficos

- **Diagramas:** São gráficos geométricos dispostos em duas dimensões. São os mais usados na representação de séries estatísticas. Eles podem ser: Barras, Linhas, Colunas e Setores

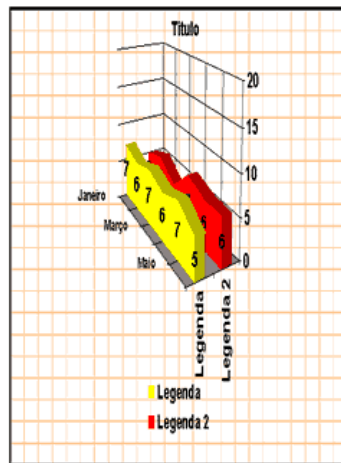


Gráficos

- **Estereogramas**: São gráficos geométricos dispostos em três dimensões, pois representam volume. São usados nas representações gráficas das tabelas de dupla entrada. Em alguns casos este tipo de gráfico fica difícil de ser interpretado dada a pequena precisão que oferecem.

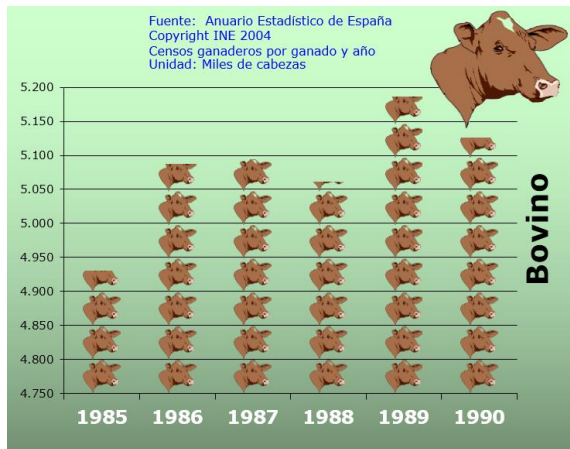


	Legenda	Legenda 2
Janeiro	7	5
Fevereiro	6	6
Março	7	4
Abril	6	7
Maió	7	6
Junho	5	6
Julho	6	5



Gráficos

- **Pictogramas**: São construídos a partir de figuras representativas da intensidade do fenômeno. Este tipo de gráfico tem a vantagem de despertar a atenção do público leigo, pois sua forma é atraente e sugestiva. Os símbolos devem ser autoexplicativos. A desvantagem dos pictogramas é que apenas mostram uma visão geral do fenômeno, e não de detalhes minuciosos. Veja o exemplo abaixo:



Gráficos

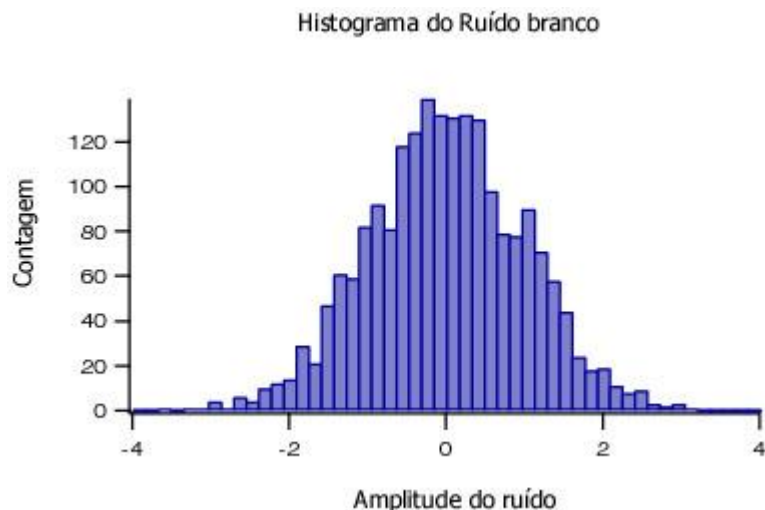
- **Cartogramas**: São ilustrações relativas a cartas geográficas (mapas). O objetivo desse gráfico é o de figurar os dados estatísticos diretamente relacionados com áreas geográficas ou políticas.



Gráficos

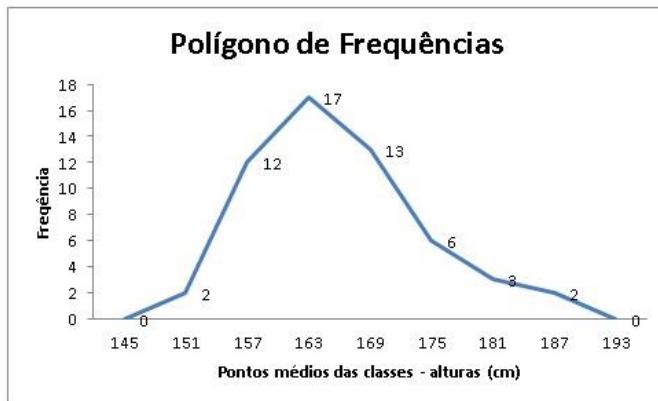
■ Histograma:

- 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. A área de um histograma é proporcional à soma das frequências simples ou absolutas.



Gráficos

- **Polígono de frequência:** é um gráfico em linha, sendo as frequências marcadas sobre perpendiculares ao eixo horizontal, levantadas 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.



Gráficos

- **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.

