TRIBUNAL DE CONTAS DO ESTADO DE ESPÍRITO SANTO

TRIBUNAL DE CONTAS
DO ESTADO DO ESPÍRITO SANTO

Introdução à Estatística

Ficha técnica:

EQUIPE DE PRODUÇÃO

COORDENAÇÃO DE PRODUÇÃO DO MATERIAL PEDAGÓGICO
Avante Brasil Informática e Treinamentos Ltda.

DIRETOR DE PLANEJAMENTO
Carlos Henrique Ferraz

DIRETOR COMERCIAL
Rômulo Moura Afonso

CONTEÚDO
Avante Brasil

COORDENADORA DE EDUCAÇÃO À DISTÂNCIA
Gislene Alves

DESIGN INSTRUCIONAL
Cláudia Vasconcelos

DIAGRAMAÇÃO E PROJETO GRÁFICO
Alissom Lázaro

ÍCONES ORGANIZADORES

DEFINIÇÃO - É utilizado ao definir conceitos e significados.

SAIBA MAIS- Aprofundamento de ideias, curiosidades, links de sites e textos complementares.

REFLEXÃO - Momento para refletir sobre as questões apresentadas e aprofundar pontos relevantes.

EXEMPLO - Utilizado no momento em que exemplifica conteúdo ou ideias.

Sumário

MÓDULO 1 - INTRODUÇÃO À ESTATÍSTICA

Serão abordados desde **a introdução à estatística,** como também **s**éries estatísticas, **tabela de frequências, estatística gráfica, medidas de centralidade e de dispersão, assimetria, curtose e complementos e n**úmeros índices**.**

A estatística tem por objetivo obter, organizar e analisar dados, determinar as correlações que apresentem, tirando delas suas consequências para descrição e explicação do que passou e previsão e organização do futuro.

De acordo com a Wikipédia (dicionário interativo colaborativo), por definição a estatística é uma área do conhecimento que utiliza teorias probabilísticas para explicação de eventos, estudos e experiências.

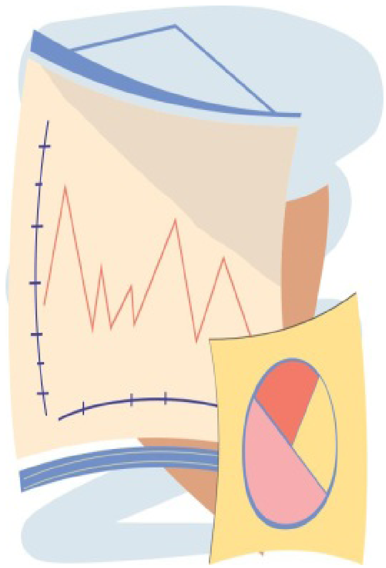
A estatística é também uma ciência e prática de desenvolvimento de conhecimento humano com o uso de dados empíricos. Baseia-se na teoria estatística, um ramo da matemática aplicada.

Na teoria estatística, a aleatoriedade e incerteza são modeladas pela teoria da probabilidade. Algumas práticas estatísticas incluem, por exemplo, o planejamento, a sumarização e a interpretação de observações.

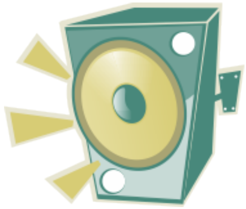
Alguns autores sugerem que a estatística é um ramo da teoria da decisão.

O termo estatística surge da expressão em Latim statisticum collegium palestra sobre os assuntos do Estado, de onde surgiu a palavra em língua italiana statista, que significa "homem de estado", ou político, e a palavra alemã Statistik, designando a análise de dados sobre o Estado.

De acordo com o Ministério do Planejamento, o que modernamente se conhece como Ciências Estatísticas ou simplesmente Estatística é um conjunto de técnicas e métodos de pesquisa que, entre outros tópicos, envolve o planejamento do experimento a ser realizado, a coleta qualificada dos dados, a inferência, o processamento, a análise e a disseminação das informações.



O objetivo da estatística é a produção da “melhor” informação possível a partir dos dados disponíveis; fornecer métodos e técnicas para lidarmos racionalmente com situações sujeitas a incertezas.

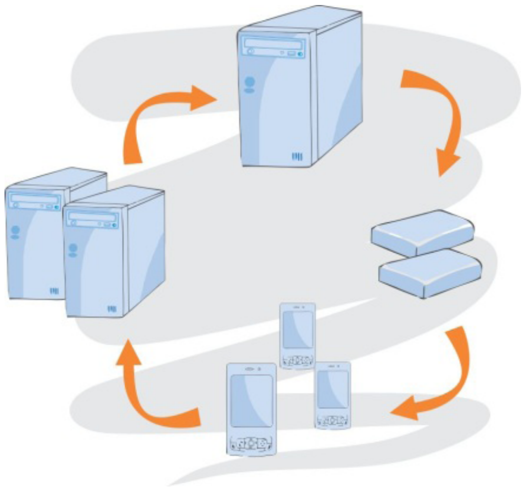


O desenvolvimento e o aperfeiçoamento de técnicas estatísticas de obtenção e análise de informações permitem o controle e o estudo adequado de fenômenos, fatos, eventos e ocorrências em diversas áreas do conhecimento.

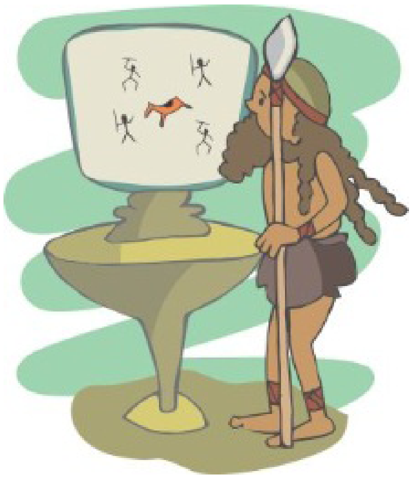
Apesar de a Estatística ser uma ciência relativamente recente na área da pesquisa, ela remonta a antiguidade, onde operações de contagem populacional já eram utilizadas para obtenção de informações sobre os habitantes, riquezas e poderio militar dos povos.

Após a idade média, os governantes na Europa Ocidental, preocupados com a difusão de doenças endêmicas que poderiam devastar populações e, também, acreditando que o tamanho da população poderia afetar o poderio militar e político de uma nação, começaram a obter e armazenar informações sobre batizados, casamentos e funerais.

Entre os séculos XVI e XVIII as nações com aspirações mercantilistas começaram a buscar o poder econômico como forma de poder político. Os governantes, por sua vez, viram a necessidade de coletar informações estatísticas referentes a variáveis econômicas, tais como: comércio exterior, produção de bens e de alimentos.



Atualmente os dados estatísticos são obtidos, classificados e armazenados em meio magnético e disponibilizados em diversos sistemas de informação acessíveis a pesquisadores, cidadãos e organizações da sociedade que, por sua vez, podem utilizá-los para o desenvolvimento de suas atividades.



A expansão no processo de obtenção, armazenamento e disseminação de informações estatísticas tem sido acompanhada pelo rápido desenvolvimento de novas técnicas e metodologias de análise de dados estatísticos.

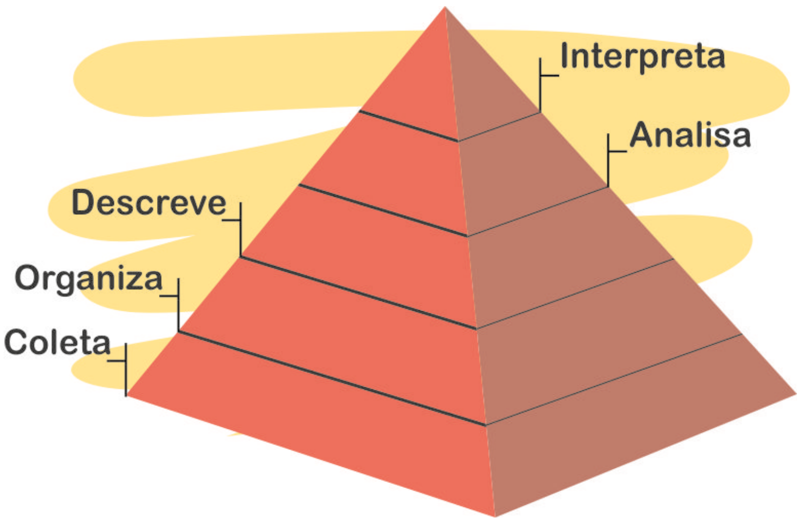
De um lado, a Estatística, basicamente, coleta, organiza e descreve os dados e, de outro, analisa e interpreta esses dados.

De acordo com a CBO (Classificação Brasileira de Ocupações), o estatístico é o profissional que trabalha nas seguintes áreas de ocupações:

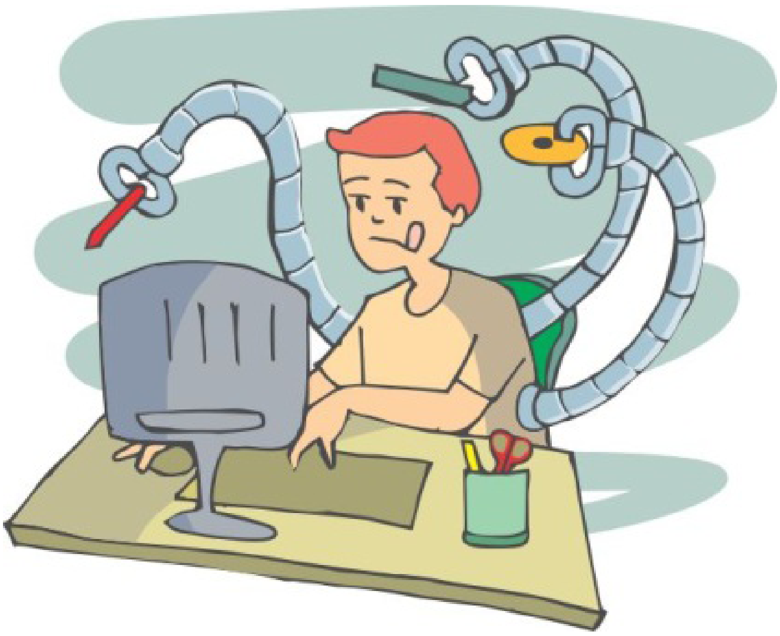
Estatístico (Amostrista, Estatístico – analista), **Estatístico** (estatística aplicada) (Bioestatístico, Demógrafo, Econometrista) e **Estatístico teórico.**

Saiba Mais sobre as Classificação Brasileira de Ocupação acessando:

http://www.mtecbo.gov.br/cbosite/pages/home.jsf



- Desenham amostras; analisam e processam dados; constroem instrumentos de coleta de dados; criam banco de dados; desenvolvem sistemas de codificação de dados; planejam pesquisa; comunicam-se oralmente e por escrito.
- Programam atividades e coordenam processos de produção petroquímica. Controlam a qualidade de insumos e produtos.
- Analisam dados estatísticos do processo produtivo; interpretam laudos de análises químicas e identificam produtos e insumos.



- Mantêm equipamentos e materiais em condições operacionais e coordenam equipe de trabalho.
- Trabalham conforme normas e procedimentos técnicos de qualidade, de segurança, de preservação ambiental e saúde.
- Usinam peças de metais ferrosos e não-ferrosos. Resinas e plásticos em máquinas cnc; preparam e ajustam máquinas de usinagem cnc.
- Ajustam ferramentas, realizam testes e controle ferramental.
- Documentam atividades, tais como preenchimento de fichas de controle de produção, resultados do controle estatístico do processo, referências das peças, atualização dos leiautes de ferramentas e ocorrências de manutenção das máquinas.
- Trabalham seguindo normas de segurança, higiene, qualidade e preservação ambiental.
- Podem programar máquinas de usinagem cnc. Classificam bobinas de fios têxteis, tecidos planos e de malhas e preparam lotes de produção conforme programação pré-estabelecida.
- Empregam ações preventivas e corretivas na produção de fios têxteis, tecidos planos e de malhas e registram dados para controle estatístico e de qualidade. Identificam necessidades de treinamento.

MÓDULO 2 - SÉRIES ESTATÍSTICAS

De acordo com Carvalho (2008, p. 12) as séries estatísticas nada mais são do que tabelas das quais expressam o resultado de um estudo estatístico. É uma forma de apresentar dados estatísticos por tabulação.

Para que tais dados sejam entendidos é importante focar em 3 (três) elementos de uma série estratégica:

1º o fato: é o fenômeno que foi investigado e cujos valores estão sendo apresentados na tabela.

2º o local: indica o âmbito geográfico ou a região onde o fato aconteceu.

3º a época: refere-se ao período, data ou tempo, quando a variável foi investigada.

Assim, uma série estatística deverá responder às seguintes questões: **O que? Quando? Onde?**

Haverá sempre um elemento que sofrerá variações enquanto outros dois permanecerão constantes.

Na série estatística é possível classificar os elementos que por variação são classificados como Histórico, Geográfico, Específico e Distribuição de Frequência:

- **Série estatística histórica -** os elementos de variação são referentes à época, permanecendo fixos o local e a descrição do fenômeno.

FESTA VEÍCULOS SA

PERÍODO	UNIDADES VENDIDAS
Jan/10	20
Fev/10	10
TOTAL	30

* Em mil unidades

- **Série estatística geográfica -** o elemento de variação é o local, permanecendo fixos o tempo e a descrição do fenômeno. Também conhecidos por séries espaciais, territoriais ou de localização.

"Denominamos **série estatística** toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou da espécie" (CRESPO, 2002, P.26).

FESTA VEÍCULOS SA
Vendas no 1º bimestre de 2010

PERÍODO	UNIDADES VENDIDAS
São Paulo	13
Brasília	17
TOTAL	30

* Em mil unidades

- **Série estatística específica -** o elemento de variação é a descrição do fenômeno, permanecendo fixos o local e o tempo. Também conhecida como série especificativa ou categórica.

FESTA VEÍCULOS SA Vendas no 1º bimestre de 2010	
PERÍODO	UNIDADES VENDIDAS
Peugeot	18
Mercedes Benz	12
TOTAL	30

* Em mil unidades

- **Série estatística de distribuição de frequências -** a maioria das provas de estatística trabalha as questões tomando por base dados apresentados sob esta forma, ou seja, dados dispostos na Distribuição de Frequências. Nesta série estatística os dados são ordenados segundo um critério de magnitude, em classes ou intervalores, permanecendo fixos o fato, o local e a época.

População Urbana do Brasil por Região 1940 a 1980 (x 1000)

Anos	REGIÕES				
	N	NE	SE	S	CO
1940	406	3381	7232	1591	271
1940	581	4745	10721	2313	424
1940	958	7517	17461	4361	1007
1940	1624	11753	28965	7303	2437
1940	3037	17567	42810	11878	5115

Fonte: Anuário Estatístico (1984)

Existem alguns sinônimos para este tipo de série estatística que devem ser cuidadosamente memorizados para o caso de uma questão teórica. São elas: cronológicas, temporais ou de marcha.

MÓDULO 3 - TABELA DE FREQUÊNCIAS

Frequência ou frequência absoluta é o número de vezes que um mesmo resultado acontece durante uma pesquisa. É denominado de *f*.

Uma distribuição de frequência é a apresentação dos resultados de uma pesquisa por meio de uma tabela que mostra a frequência de ocorrência de cada resultado.

De acordo com Carvalho (2008), existem seis tipos de frequências que podem estar presentes numa distribuição. Elas são divididas em duas categorias: Frequência Absoluta e Frequência Relativa.

Tabela de freqüência do Índice Antropométrico - CDC.
Índice Antropométrico - CDC

	Freqüência	Percentual
magreza	57	10,3
normal	420	76,2
risco de sobrepeso	55	10,0
sobrepeso	19	3,4
Total	551	100,0

Prevalência de obesidade em escolares da rede municipal de Governador Valadares- MG

Quando há estudo de uma variável, o pesquisador precisa conhecer o comportamento de tal variável, analisando a ocorrência de suas possíveis realizações.

Sintomas Urinários	Freqüência	
	Absoluta	Relativa %
Não apresenta	52	43
Polaciúria	15	13
Incontinência urinária aos esforços	12	10
Sensação de resíduo urinário na bexiga	10	8
Urgência míccional	09	8
Noctúria	07	6
Urge-Incontinência	06	5
Gotejamento pós-micção	04	3
Enurese noturna	04	3
Disúria	01	1
Hematúria	00	0
	100	100 %

Sintomas do trato urinário inferior em mulheres
Fonte: http://www.efdeportes.com/efd

Na **frequência absoluta** encontramos:

- **Frequência Absoluta Simples (*fi*)** - É o número de vezes que o elemento aparece na amostra ou o número de elementos pertencentes a uma classe.
- **Frequência Absoluta Acumulada Crescente – (*Fiac*)** - representa o número de observações existentes além do valor ou da classe, incluindo no cálculo as observações correspondentes a esse valor ou a essa classe. Para obter este tipo de frequência, basta somar à frequência simples absoluta da classe ou do valor individual, as frequências simples absolutas das classes ou dos valores individuais posteriores.
- **Frequência Absoluta Acumulada Decrescente - (*Fiab*)** - é a soma da frequência simples absoluta de uma classe ou de um dado valor com as frequências simples absolutas das classes ou dos valores anteriores. A expressão "abaixo de" refere-se ao fato de que as frequências a serem acumuladas correspondem aos valores menores ou anteriores ao valor ou à classe cuja frequência acumulada se deseja obter, incluindo no cálculo a frequência do valor ou da classe. É utilizada toda vez que se procura saber quantas observações existem até uma determinada classe ou valor individual.

Na **frequência relativa** temos:

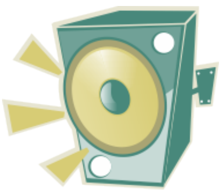
- **Frequência Relativa Simples (*Fri*)** - representa a proporção de observações de um valor individual ou de uma classe, em relação ao número total de observações. Trata-se, portanto, de um número relativo.
- **Frequência Relativa Acumulada Crescente- (*Friac*)** - igual à soma da frequência simples relativa dessa classe ou desse valor com as frequências simples relativas das classes ou dos valores posteriores.
- **Frequência Relativa Acumulada Decrescente- (*Friab*)** - é a soma da frequência simples relativa dessa classe ou desse valor com as frequências simples relativas das classes ou dos valores anteriores.

Exemplo com as frequências apresentadas.

Idade dos alunos do curso de pedagogia da UnB, no ano de 1996.

Idade	Nº de alunos (*fi*)	*fri*	*fri* (%)	*Fiab*	*Fiab* (%)	*Fiac*	*Fiac* (%)
21 - 24	7	0.23	23	7	23	30	100
24 - 27	8	0.27	27	15	50	23	77
27 - 30	1	0.03	3	16	53	15	50
30 - 33	5	0.17	17	21	70	14	47
33 - 36	9	0.30	30	30	100	9	30
TOTAL	30	1.00	100	---	---	---	---

Fonte: (dados hipotéticos)



As tabelas de frequência precisam ser vistas como simetria do conjunto. Distribuir frequência é acima de tudo visualizar a distribuição de classes.

Quando se tem alguma indicação da distribuição de uma população por argumentação probabilística ou outras, muitas vezes é possível ajustar tal distribuição teórica à distribuição de frequências obtida de uma amostra.

Por não haver na sua existência atuação isolada, a estatística torna-se o combustível para desenvolvimento dos processos de distribuição.

Sempre que um setor necessitar de distribuição, a estatística estará presente. Sempre que uma frequência necessitar de controle, a estatística estará presente. O controle dentro de uma distribuição de frequência de valores favorece a execução de um planejamento sem erros.

A estatística atua de forma permanente dos processos de distribuição da frequência. Sua aplicação na infraestrutura de uma frequência se dá a partir do momento em que o controle torna-se necessário.

Não há consolidação estrutural se não houver controle de entrada e saída. A infraestrutura é a base do processo estatístico. Sempre que houver controle de dados é possível programar o valor do investimento para construção do alicerce com uma base de qualidade e sem descontrole de investimento.

Por ser um elo de uma distribuição de frequência estatística, o cálculo implica na aplicação de uma eficiência pela integração entre elos coordenados que visam otimizar custos e garantir rapidez com qualidade nas distribuições de dados.

Quando pensamos em estatística na distribuição de frequência de valores é importante que se pense no processo de controle e distribuição tanto interna quanto externa. A estatística está sempre atuando para promover os valores dentro da distribuição de frequência e para agregar eficiência ao conjunto de processos de distribuição e crescimento de uma frequência.

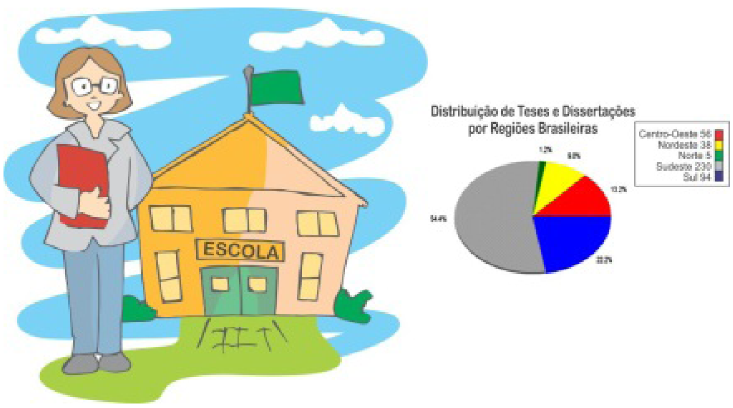
A influência da estatística é transversal (interage com as funções organizações) projetando resultados diretos nos custos, na produtividade até aos serviços ao cliente. Podemos dizer que é o fio condutor ligando diferentes setores, estando presente em todos os departamentos da distribuição de frequência de valores, direta ou indiretamente.

A Estatística é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões (CRESPO, 1995, p. 13). Ela está presente em todos os lugares como uma rede burocrática, assimétrica, modular e tangível. Onde o relacionamento entre os dados é proferido formalmente, coordenado pela unidade produtiva.

A integração da estatística em diversos processos de distribuição favoreceu a melhoria do desenvolvimento da relação entre Frequência, distribuidores e dados. Essa integração funciona de forma ordenada onde os acordos formais, a unidade produtiva, o compartilhamento das atividades de distribuição de frequência de valor, acordos, definição dos elos da distribuição de frequência de suprimentos e as relações promovem o sucesso da aplicação da estatística nos processos de distribuição integrados.

MÓDULO 4 - ESTATÍSTICA GRÁFICA

Gráfico é a tentativa de se expressar visualmente estatísticas simplificadas, matemáticas ou não de dados ou valores obtidos, assim facilitando a compreensão. São recursos visuais muito utilizados para facilitar a leitura e a compreensão de informações sobre fenômenos e processos naturais, sociais e econômicos.



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

Assim, por meio da apresentação de dados estatísticos, a informação direta e objetiva do fenômeno de análise é passada.

> Saber desenhar um gráfico é secundário, o importante é saber ler os resultados apresentados para que as informações sejam precisas.

Os gráficos podem ser apresentados da seguinte forma:

GRÁFICO DE COLUNAS:

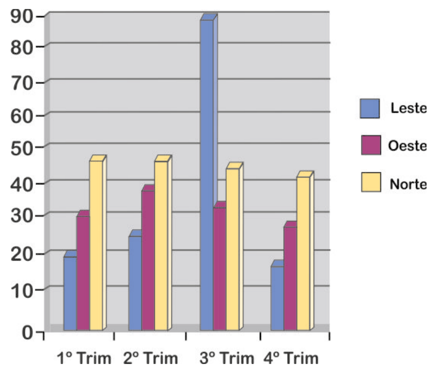


Gráfico de Coluna normal

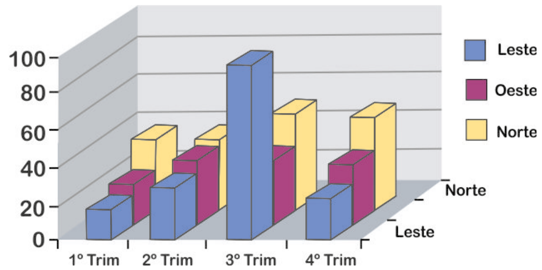


Gráfico de Coluna em 3D

O Gráfico de Colunas é usado para séries temporais, específicas ou geográficas.

O GRÁFICO EM BARRAS:

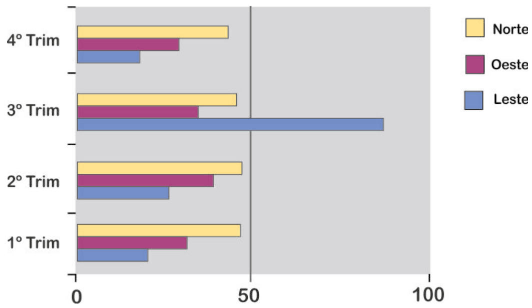


Gráfico de Barras normal

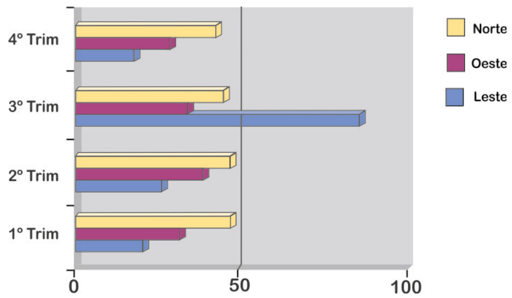
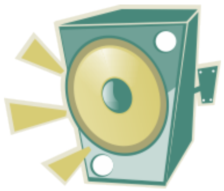


Gráfico de Barras Normal em 3D



O Gráfico de Barras é a representação de uma série estatística normalmente usada para séries geográficas ou também na representação de séries específicas.

O GRÁFICO EM LINHAS:

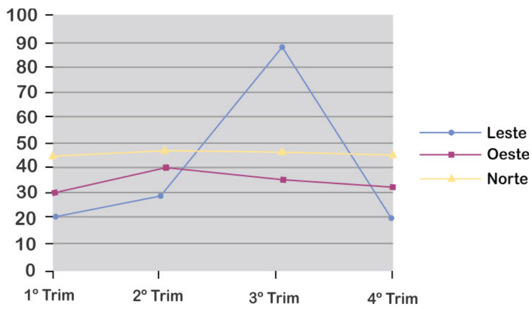


Gráfico de Linhas Normal

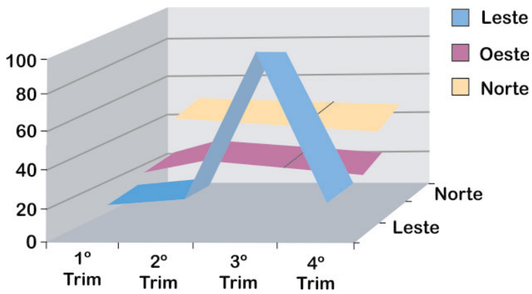


Gráfico de Linhas 3D

O Gráfico de Linhas é utilizado em séries temporais.

JÁ O GRÁFICO EM SETORES:

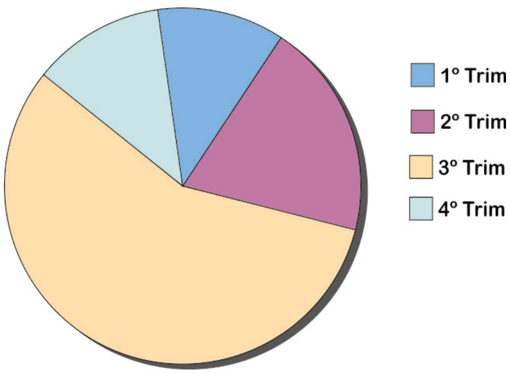


Gráfico em Pizza Normal

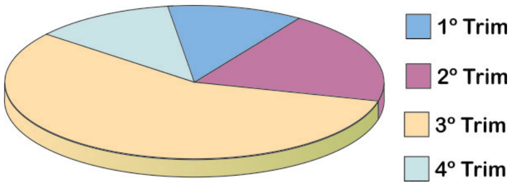
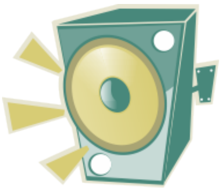


Gráfico em Pizza 3D

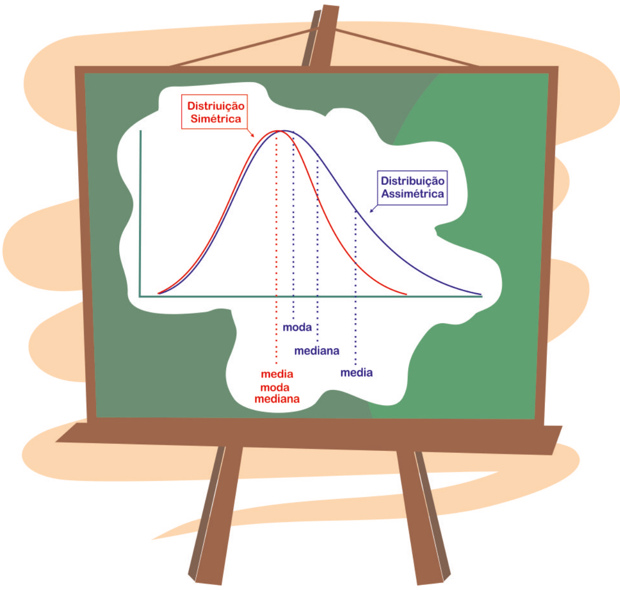


O Gráfico em Setores é utilizado para comparar proporções. Onde a classe é o setor circular e o ângulo é o tamanho da amostra.

MÓDULO 5 - MEDIDAS DE CENTRALIDADE

De acordo com Silva (2011, p. 18), as medidas de centralidade sintetizam os valores e indicam o centro da distribuição de frequências.

A média e mediana indicam o centro da distribuição de frequência, enquanto a moda indica a região de maior concentração de frequências na distribuição.



São as estatísticas que representam uma série de dados orientando-nos quanto à disposição da repartição em inclusão ao eixo horizontal do gráfico da curva de frequência.

As medidas de posições mais importantes são as medidas de tendência central ou promédias.

A Medida de Centralidade Média pode ser apresentada em 3 tipos:

- Aritmética;
- Geométrica, e
- Harmônica.

A aritmética pode ser obtida somando-se os valores da variável e dividindo pelo número de observações, e é mais utilizada do que as médias geométricas e harmônicas.

Média=(soma dos dados) ÷ (número de dados)

Para representar a média existem dois símbolos: o símbolo $\overline{x}$ é utilizado para média amostral, e o símbolo μ (leia-se mi) é utilizado para média populacional.

Fórmula para média amostral:	Fórmula para média populacional:
$\overline{x}$ = (x1 + x2 + x3 + ... + xn) / n	μ = (x1 + x2 + x3 + ... + xn) / N
Sendo n = número de dados da amostra	Sendo N = número de dados da população

Número de acertos de uma prova de matemática **(Amostral):**

Número de acertos de uma prova de matemática (Amostral):

3 1 2 0 2 5 0 1 2 2 4 3 1

A média de acertos será dada por:

$\overline{x}$ = (3+1+2+0+2+5+0+1+2+2+4+3+1) / 13

$\overline{x}$ = 26 / 13 = 2.

A média de acertos da prova foi 2.

O valor é contido entre o menor e o maior valor observado, a soma dos desvios em relação à média é igual a zero, multiplicando-se ou dividindo-se todos os elementos de um conjunto de dados por uma constante, a média aritmética fica multiplicada ou dividida por essa constante e somando-se ou subtraindo-se uma constante de todos os elementos de um conjunto de dados, a média aritmética fica aumentada ou subtraída dessa constante.

Na **média geométrica**, a obtenção pode ser apresentada por multiplicação dos valores da variável e tirando a raiz de ordem *n* do produto desses números.

Digamos que tenhamos os números **1,2 e** 4, para obtermos o valor médio aritmético deste conjunto, multiplicamos os elementos e obtemos o produto **8.** Pegamos então este produto e extraímos a sua raiz cúbica, chegando ao valor médio **2.**

Extraímos a raiz cúbica, pois o conjunto é composto de **3** elementos. Se fossem ***n*** elementos, extrairíamos a raiz de índice ***n***.

Neste exemplo teríamos a seguinte solução:

$$\sqrt[3]{1.2.4} \rightarrow \sqrt[3]{8} = 2$$

Normalmente utiliza-se a **média geométrica** quando os dados estão organizados em uma progressão geométrica.

Uma das utilizações deste tipo de média é na definição de uma progressão geométrica que diz que em toda P.G., qualquer termo é média geométrica entre o seu antecedente e o seu consequente:

$$a_n = \sqrt{a_{n-1} \cdot a_{n+1}}$$

Três termos consecutivos de uma P.G.: **2, 4 e 8.** Temos então que o termo **4** é média geométrica dos termos **2** e **8.**

Vejamos:

$$\sqrt{2.8} \rightarrow \sqrt{16} = 4$$

A **média harmônica** pode ser obtida dividindo o número de observações pela soma do inverso dos valores da variável.

Média harmônica entre **3, 6** e **12**. Primeiramente é necessário calcular a média aritmética dos inversos dos valores dados:

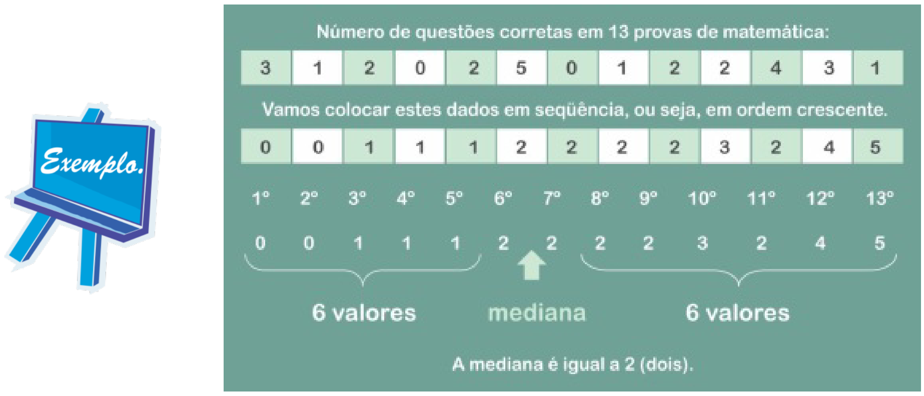
$$\frac{\frac{1}{3}+\frac{1}{6}+\frac{1}{12}}{3} = \frac{\frac{4+2+1}{12}}{3} = \frac{\frac{7}{12}}{3} = \frac{7}{12} \cdot \frac{1}{3} = \frac{7}{36}$$

$$\frac{1}{\frac{7}{36}} = \frac{1}{1} \cdot \frac{36}{7} \cong 5{,}14$$

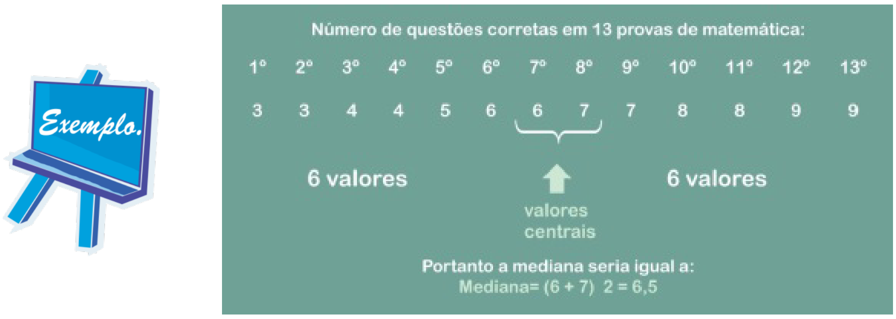
Depois, **faz-se o inverso** do resultado, tendo finalmente a **média harmônica de 3**, **6** e **12**:

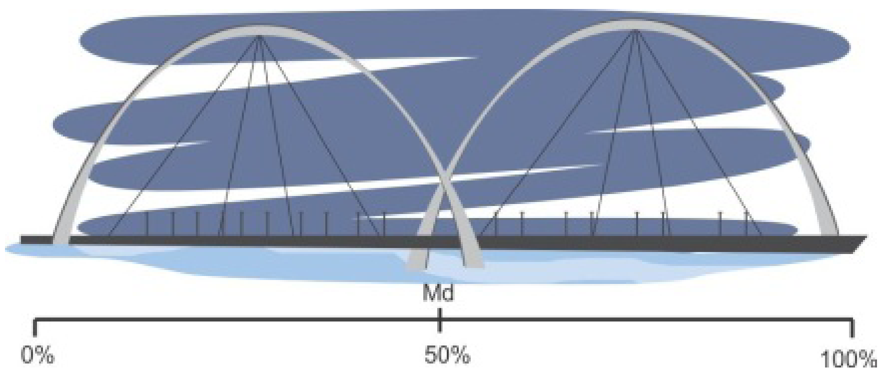
A **média harmônica** é muito útil em diversas situações práticas, em especial quando temos variáveis inversamente proporcionais, tais como velocidade e tempo ou custo médio de produtos comprados com uma quantia fixa.

A Medida de Centralidade **mediana** também visa caracterizar o centro da distribuição de frequências, mas com um critério diferente. **A mediana é o termo que ocupa a posição central de um conjunto *n* dados ordenados.**



E se o número de dados for par?





Suas propriedades são destacadas por coincidir ou não com um elemento de série. Coincidir ou não com a média aritmética.

Não sofre a influência de calores extremos, pois depende da posição e não dos valores dos elementos da série.

Por causa desta característica, em muitos casos o uso da mediana é mais conveniente do que a média. É uma estatística de ordem (se refere à posição de valores em uma amostra).

Já **a moda** é o valor mais frequente em uma série de dados. **Podendo ser encontrados em séries sem moda (amodal), com 1 moda (unimodal), 2 modas (bimodal), 3 modas (trimodal) e assim por diante.**

Pesquisa de opinião sobre a Top Model Linda
B=Bom, M=Médio, R=regular e P=Péssimo
Pesquisa N°1 – B,B,B,B,M,M,R,P - a moda é B (distribuição unimodal ou modal).
Pesquisa N°2 – B,B,B,M,M,M,R,P – as modas são B e M (distribuição bimodal).
Pesquisa N°3 – B, M, R, P – não existe moda (distribuição amodal).

A moda é a única medida de tendência central que pode ser calculada para dados qualitativos e quantitativos.

Em resumo:
Média: obtém-se somando os valores de todos os dados e dividindo a soma pelo número de dados do conjunto numérico.
Moda: é o valor mais frequente de um conjunto de dados.
Mediana: após ordenar os valores por ordem crescente ou decrescente, a média é: -se a quantidade desses valores for ímpar, a mediana é valor que ocupa a posição central; -se a quantidade desses valores for par, a mediana é a média dos dois valores centrais.

Gastos com alimentação:

Números ímpares					
Meses	Jan.	Fev.	Mar.	Abr.	Mai.
Gastos em R$	25	22	35	28	35

Média: 29
25+22+35+28+35=145
145/5=29

Moda: 35

Mediana: 28
22, 25,28,35,35

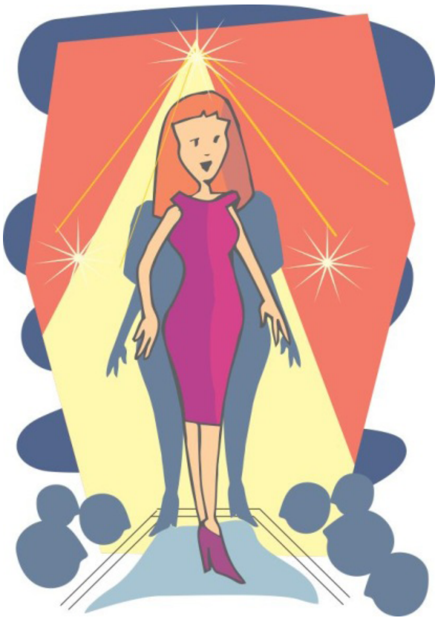
Números pares						
Meses	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.
Gastos em R$	25	22	35	28	35	33

Média: 29,67
25+22+35+28+35+33=178
178/6=29,67

Moda:35

Mediana: 30,5
22, 25,28,33,35,35
28+33=61
61/2=30,5

MÓDULO 6 - MEDIDAS DE DISPERSÃO, ASSIMETRIA E CURTOSE



As medidas de dispersão indicam se os valores estão relativamente próximos um dos outros, ou separados em torno de uma medida de posição, no caso, a média.

As principais medidas de dispersão são: **amplitude total, desvio médio, variância, desvio padrão, coeficiente de variação e desvio-quartil,** e são importantes na análise estatística de dados, uma vez que as medidas de tendência central não são suficientes para descrever completamente um conjunto de dados.

A seguir, destacaremos duas principais medidas de dispersão: amplitude total, desvio padrão e coeficiente de variação.

A amplitude total é a diferença entre o mínimo e o máximo valor da variável.

É a **única medida de dispersão que não tem na média o ponto de referência.**

Quando os dados <u>não estão agrupados</u> **a amplitude total é a diferença entre o maior e o menor valor observado**:

A = Xmáx – Xmín
At = amplitude total
Xmáx - valor máximo observado na amostra
Xmín - valor mínimo observado na amostra

Para os valores 40, 45, 48, 62 e 70 a amplitude total será:

At = 70 - 40 = 30.

Quando **os dados estão agrupados sem intervalos de classe ainda temos:**

At = x máximo - x mínimo.

xi	0	1	3	4
fi	2	6	5	3
At	4 - 0 = 4			

Com intervalos de classe a amplitude total é a diferença entre o limite superior da última classe e o limite inferior da primeira classe.

Então:

At = L máximo - L mínimo.

Classes	fi
4 |--- 6	6
6 |--- 8	2
8 |--- 10	3
At = 10 - 4 = 6	

O **desvio padrão** é a medida de dispersão mais utilizada na inferência estatística. Esta **leva em consideração a totalidade dos valores da variável em estudo**. É um **indicador de variabilidade bastante estável**. O desvio padrão baseia-se nos desvios em torno da média aritmética e a sua fórmula básica pode ser traduzida como: **a raiz quadrada da média aritmética dos quadrados dos desvios e é representada por S.**

Lembre-se: Desvio (xi – $\bar{x}$, nesse caso, é a diferença entre a média e cada valor da variável em estudo.

Lembre-se: Desvio (xi – $\bar{x}$, nesse caso, é a diferença entre a média e cada valor da variável em estudo.

$$S = \sqrt{\frac{\Sigma\,(Xi - \bar{x})^2}{n}}$$

A fórmula acima é empregada quando tratamos de uma **população de dados não-agrupados.**

*Calcular o desvio padrão da população representada por **- 4 , -3 , -2 , 3 , 5***

Calcular o desvio padrão da população representada por - 4 , -3 , -2 , 3 , 5

Xi	$\bar{x}$	$(Xi - \bar{x})$	$(Xi - \bar{x})^2$
- 4	- 0,2	- 3,8	14,44
- 3	- 0,2	- 2,8	7,84
- 2	- 0,2	- 1,8	3,24
3	- 0,2	3,2	10,24
5	- 0,2	5,2	27,04
		Σ =	62,8

Sabemos que n = 5 e 62,8 / 5 = 12,56.
A raiz quadrada de 12,56 é o desvio padrão = 3,54

Sabemos que **n** = 5 e 62,8 / 5 = 12,56.

A raiz quadrada de 12,56 é o **desvio padrão = 3,54.**

O **coeficiente de variação** tem o objetivo de analisar o grau de variabilidade de determinadas situações, através dele podemos perceber desempenhos iguais, muito próximos ou muito distantes. Ela determina de forma mais específica as possíveis variações, no intuito de não comprometer os resultados da análise.

Este coeficiente está sempre relacionado ao valor médio de um conjunto porque a dispersão é uma medida sempre relacionada a uma determinada média.

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

De maneira mais simplificada:

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

1) Suponha que queiramos estudar a variação das idades de dois grupos abaixo relacionados:

G1: 7 7 7 7 7 7

G2: 8 9 10 11 19 22

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

Cálculo da média: vamos utilizar a Fórmula: Média Aritmética

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

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

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

aproximadamente, 13 anos

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

Cálculo do desvio padrão: Vamos utilizar a Fórmula 4: Desvio Padrão: Dados Não Agrupados.

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

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

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

Dessa forma,

Para G1:

Sabemos que $\begin{cases} n = 6 \\ \sum x_i^2 = 294 \\ \sum x_i = 42 \end{cases}$ Então,

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

Para G2:

Sabemos que $\begin{cases} n = 6 \\ \sum x_i^2 = 1.211 \\ \sum x_i = 79 \end{cases}$ Então,

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

Aproximadamente, 5 anos.

Até aqui, podemos sintetizar da seguinte forma:

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

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

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

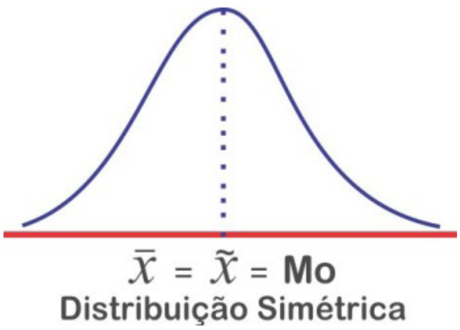
Isso significa que podemos afirmar que **G2** é um grupo cujas idades variaram mais do que as idades de **G1**. E ainda, essa variação foi de 38%. ***A CV* mede a variação.**

Na **medida assimetria** temos:

($\bar{X}$) Média
($\tilde{X}$) Mediana
(Mo) Moda

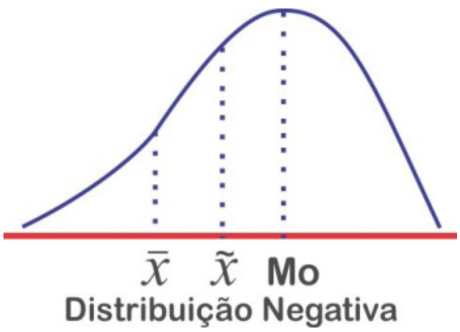
Uma distribuição com classes é **simétrica** quando:

Média = Mediana = Moda



Uma distribuição com classes é **Assimétrica à esquerda ou negativa** quando:

Média < Mediana < Moda



Assimétrica à direita ou positiva quando:

Média > Mediana > Moda

Coeficiente de assimetria: por ser absoluta, apresenta a mesma deficiência do desvio padrão; não permite a possibilidade de comparação entre as medidas de duas distribuições. Por isso, o enfoque será dado ao coeficiente de assimetria de Person:

As = 3 (Média - Mediana) / Desvio Padrão

Escalas de assimetria:

| AS | < 0,15 $\rightarrow$ assimetria pequena;

0,15 < | AS | < 1 $\rightarrow$ assimetria moderada;

| AS | > 1 $\rightarrow$ assimetria elevada.

- Se **AS = - 0,49** $\rightarrow$ a assimetria é considerada moderada e negativa.
- Se **AS = 0,75** $\rightarrow$ a assimetria é considerada moderada e positiva.

A **medida de curtose** é denominada pelo grau de achatamento de uma distribuição em relação a uma distribuição padrão, denominada curva normal (curva correspondente a uma distribuição teórica de probabilidade).

Quando a distribuição apresenta uma curva de frequência mais fechada que a normal (ou mais aguda ou afilada em sua parte superior), ela recebe o nome de leptocúrtica.

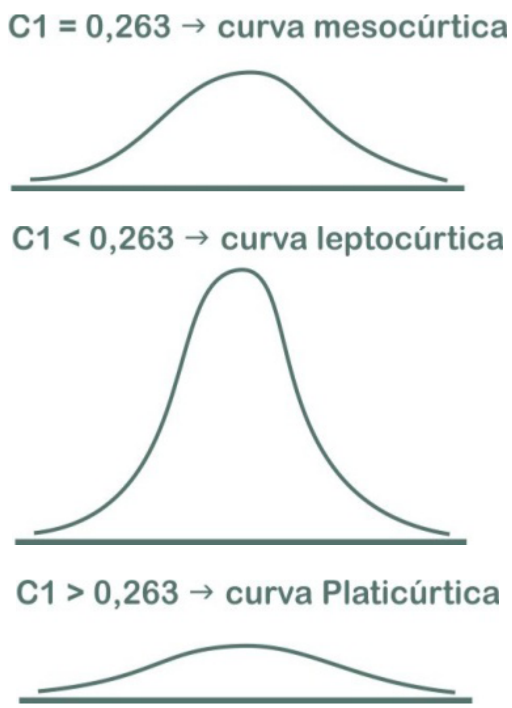
Quando a distribuição apresenta uma curva de frequência mais aberta que a normal (ou mais achatada em sua parte superior), ela recebe o nome de platicúrtica. A curva normal, que é a nossa base referencial, recebe o nome de mesocúrtica.

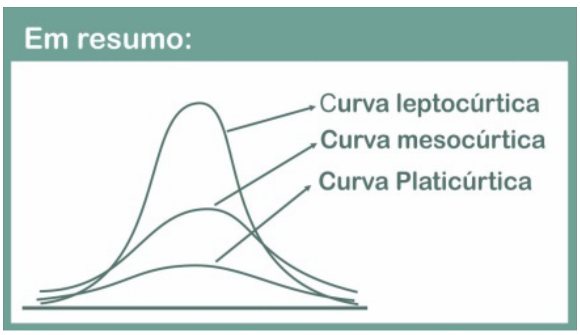
Coeficiente de curtose:

$$C1 = (Q3 - Q1) / 2(P90 - P10)$$

- Este coeficiente é conhecido como **percentílico de curtose**.

- Relativamente a **curva normal**, temos:





MÓDULO 7 - NÚMEROS ÍNDICES

De acordo com Castanheira (2008), os números índices são medidas estatísticas frequentemente usadas para comparar grupos de variáveis relacionadas entre si e para obter um quadro simples e resumido das mudanças significativas ocorridas ao longo do tempo.

De acordo com Willian J. Stevenson, os números-índices são usados para indicar variações relativas em quantidades, preços ou valores de um artigo, durante dado período de tempo. (p. 396, 1995)

São expressos em termos percentuais e, também, têm certas características em comum, sendo uma delas, as razões de quantidade no período corrente para as quantidades no período-base.

É importante ressaltar que os números índices são destituídos de qualquer significado se não forem especificadas as datas a que se referem. Normalmente, os números índices são expressos de forma percentual.

Os índices mais usados geralmente são os que apresentam variações de preço, de quantidade ou de valor, mas também são frequentemente usados nas ciências físicas, químicas, naturais e sociais.

Na **estatística** o número índice é sinônimo de variação relativa na variável de interesse; em **economia** encontra-se nos índices de preços, nas quantidades e valor dos bens, no de custo de vida, de (des)emprego, etc; em **administração pode ser observado na** velocidade de vendas, na lucratividade, no endividamento; já na **administração pública** permitem avaliar a qualidade de vida, a permanência ou evasão escolar, o nível de criminalidade e o padrão de saúde das populações.

Os principais índices utilizados no Brasil são:

IPC – Índice de Preços ao Consumidor: IPC/Fipe, ICV/Dieese, INPC- Índice Nacional de Preços ao Consumidor e IPCA – Índice de Preços ao Consumidor Ampliado; IPA – Índice de Preços por Atacado; INCC – Índice Nacional da Construção Civil; IGP – Índice Geral de Preços.

As três classificações de números-índices administrativos e econômicos são: **o índice de preço, o de quantidade e o de valor.**

Dividem-se em números-índices **simples,** quando um só produto está em jogo e, números-índices **composto** quando envolver um grupo de artigos. **No momento apresentaremos apenas o simples.**

Os Números-índices simples avaliam a variação relativa de um único item ou variável econômica entre dois períodos de tempo. Calcula-se como a razão do preço, quantidade ou valor em dado período e o correspondente preço, quantidade ou valor num período-base. Sua principal limitação é que eles se referem apenas a itens isolados, enquanto que, frequentemente, necessitamos sintetizar variações para um grupo de itens. Por exemplo, se uma pessoa percebe que o preço de um produto atualmente é o quíntuplo do que custava há dois anos, está fazendo uso de certo tipo de número índice comparativo. Quando um só produto está em jogo, o índice é dito índice simples, enquanto que uma comparação que envolva um grupo de artigos é chamada de índice composto.

Você sabia que os índices podem ser usados também em áreas como **Engenharia, Física, Medicina** (índices de fertilidade, natalidade, morbidez, mortalidade, etc.), nas chamadas **ciências do comportamento** (psicologia, sociologia, etc.) e em **educação** (quociente de inteligência, coeficiente de aprovação, etc.). Fonte: Livro: «Estatística Geral e Aplicada» Autor: Giuseppe Milone

Para os números índices temos:

- **relativo de preço =** $\frac{P_n}{P_0}.100$
- **relativo de quantidade =** $\frac{q_n}{q_0}.100$
- **relativo de valor =** $\frac{P_n \cdot q_n}{P_0 \cdot q_0}.100$

- é o preço de um item no ano-base - P_0
- é o preço de um item em determinado ano - P_n
- é a quantidade de um item no ano-base – Q_0
- é a quantidade de um item em determinado ano - Q_n
- é o valor na época-base – V_0
- é o valor na época-atual - V_t

A empresa Cristal S.A, em 1992 vendeu 300 unidades do produto "Boina", cobrando 20 reais por peça e, em 1993, vendeu 450 unidades do mesmo produto, cobrando 25 reais por peça. Determinar os relativos de preço, quantidade e valor em 1993, tomando como base 1992.

Solução:

É usual a notação 1992=100, para denotar que 1992 é o ano base.

- Relativo de preço – $P_{92}93$=. 100=125
- Relativo de quantidade – $Q_{92}93$=.100=150
- Relativo de Valor – $V_{92}93$=. 100=187,5

Observe que houve:

- um aumento de 25 % no preço, em relação ao ano base;
- um aumento de 50 % na quantidade, e
- um aumento de 87,5 % no valor.

O aumento do valor é, portanto, o aumento acumulado do aumento de preço pelo aumento de quantidade, ou seja, o produto dos índices de preço e quantidade é o índice de valor:

(1,25 x 1,5 = 1,875).

Relativo de Preço: Número índice mais simples. Relacionando-se o preço de um determinado bem ou serviço numa época atual "*t*" (chamada época atual, "hoje") e "*0*" de uma época (chamada época básica, ou simplesmente base "ontem"), obteremos o relativo de preço. Considerando P_t = preço numa época atual e P_0 = preço na época base, tem-se:

$$\frac{\frac{1}{7}}{\frac{1}{36}} = \frac{1}{1} \cdot \frac{36}{7} \cong 5,14$$

ou na Forma percentual

Relativo de Preço: $P_{0,t} = \dfrac{P_t}{P_0}$ $\qquad P_{0,t} = \left(\dfrac{P_t}{P_0}\right) \text{x}100$

Variação percentual

$$\Delta\%P = \left(\frac{P_t}{P_0} - 1\right) \text{x}100$$

Relativo de Preço: O preço da camiseta Paulus em 2004 era de R$ 25,00 e em 2005 subiu para R$ 27,50. Tomando como base o ano de 2004, determine o preço do relativo em 2005.

Solução: O ano tomado como base corresponderá sempre ao índice igual a 100.

$$p_{(04,05)} = \left[\left(\frac{p_{2005}}{p_{2004}}\right) \text{x}100\right] \rightarrow p_{(04,05)} = \left(\frac{p_{2005}}{p_{2004}}\right) \text{x}100$$

$$p_{(04,05)} = \left[\frac{27,50}{25,00}\right] \text{x}100 \rightarrow p_{(04,05)} = [1,1\text{x}100]$$

$$\rightarrow p_{(04,05)} = 110$$

Relativo de Quantidade: Para se comparar os preços de bens, pode-se fazê-lo em relação à quantidade. Considerando Q_t = quantidade de um bem ou serviço numa época atual e Q_0 = quantidade do mesmo bem ou serviço na época $_0$ (básica).

A quantidade relativa será o seguinte:

relativo de quantidade: $q_{0,t} = \dfrac{q_t}{q_0}$

ou termos percentuais

$$q_{0,t} = \left(\frac{q_t}{q_0}\right) \text{x}100$$

Relativo de Quantidade: A produção da empresa Visual em 2004 era de 40.000 camisolas e 45.000 em 2005. A quantidade relativa, tomando como base o ano de 2004, será:

$$q_{(04,05)} = \left[\left(\frac{q_{2005}}{q_{2004}}\right) x\, 100\right] \rightarrow q_{(04,05)} = \left(\frac{q_{2005}}{q_{2004}}\right) x\, 100$$

$$q_{(04,05)} = \left[\frac{45000}{40000}\right] x\, 100 \rightarrow q_{(04,05)} = [1{,}125 x 100]$$

$$\rightarrow q_{(04,05)} = 112{,}5$$

Relativo de Valor: Se P é o preço de um determinado bem ou serviço em um determinado período, e Q a quantidade consumida desse mesmo bem ou serviço no mesmo período, o produto P x Q é denominado valor total de produção ou consumo. Se P_t e Q_t respectivamente o preço e quantidade de um bem ou serviço numa época atual (t) e P_0 e Q_0, preço e a quantidade do mesmo bem ou serviço na época base ($_0$). Define-se:

$$v_{(0,t)} = \frac{v_t}{v_0} = \frac{p_t\, x\, q_t}{p_0\, x\, q_0} \rightarrow p_{(0,t)}\, x\, q_{(0,t)}\ v_{(2003,2002)} = \left[\left(\frac{600 \, x\, 4000}{500 \, x\, 3000}\right) x\, 100\right] \rightarrow$$

$$v_{(2003,2002)} = [1{,}6]\, x\, 100 \rightarrow v_{(2003,2002)} = 160$$

Relativo de Valor: A empresa Look vendeu em 2002, 3000 unidades de um secador de cabelos ao preço unitário de R$ 500,00. Em 2003 vendeu 4000 unidades do mesmo produto ao preço de R$ 600. O valor relativo das vendas em 2003 será:

Relativo de valor:

$$v_{0,t} = \frac{p_t \ x\ q_t}{p_0 \ x\ q_0}$$

ou em termos percentuais

$$v_{0,t} = \left(\frac{p_t\, x\, q_t}{p_0\, x\, q_0}\right) x100$$

REFERÊNCIAS BIBLIOGRÁFICAS

BRAGA, Luis Paulo Vieira. Compreendendo probabilidade e estatística. Rio de Janeiro: E-papers, 2010.

CARVALHO, Sérgio. Estatística simplificada. Rio de Janeiro: Elsevier, 2008.

CASTANHEIRA, Nelson Pereira. Métodos quantitativos. Curitiba: IBpex, 2008.

LARSON,R. e FARBER,B., Estatística Aplicada, 2ª ed.,São Paulo:Prentice Hall, 2004.

LEVIN Jack, e FOX, James A. Estatística para Ciências Sociais, 9a. Ed., São Paulo Prentice Hall, 2004.

LEVINE, D.M., BERENSON, M.L. s STEPHAN, D. Estatística: Teoria e Aplicações. 3a. ed. Rio de Janeiro: LTC, 2005.

MARIANO, Fabrício. Noções de estatística para concurso. Rio de Janeiro: Elsevier, 2010.

MARTINS, G.A. Estatística Geral e Aplicada. 2a. Ed. São Paulo: Editora Atlas. 2002.

SILVA, André Luiz Carvalhal. Introdução à análise de dados. Rio de Janeiro: E-papers, 2011.

STEVENSON, W. J. Estatística Aplicada à Administração. São Paulo, Harbra, 1986.