ESTATÍSTICA

<u>Conteúdo</u>

Prof. Jorge Targino

ESTATÍSTICA

1. INTRODUÇÃO

ESTATÍSTICA: ramo da matemática aplicada.

ANTIGUIDADE: os povos já registravam o número de habitantes, nascimentos, óbitos. Faziam "estatísticas".

IDADE MÉDIA: as informações eram tabuladas com finalidades tributárias e bélicas.

SEC. XVI: surgem as primeiras análises sistemáticas, as primeiras tabelas e os números relativos.

SEC. XVIII: a estatística com feição científica é batizada por GODOFREDO ACHENWALL. As tabelas ficam mais completas, surgem as primeiras representações gráficas e os cálculos de probabilidades. A estatística deixa de ser uma simples tabulação de dados numéricos para se tornar ***"O estudo de como se chegar a conclusão sobre uma população, partindo da observação de partes dessa população (amostra)".***

MÉTODO ESTATÍSTICO

MÉTODO: é um meio mais eficaz para atingir determinada meta.

MÉTODOS CIENTÍFICOS: destacamos o método experimental e o método estatístico.

MÉTODO EXPERIMENTAL: consiste em **manter constante todas as causas**, menos uma, que sofre variação para se observar seus efeitos, caso existam. *Ex: Estudos da Química, Física, etc.*

MÉTODO ESTATÍSTICO: Diante da **impossibilidade de manter as causas constantes** (nas ciências sociais), admitem todas essas causas presentes **variando-as, registrando essas variações** e procurando determinar, no resultado final, **que influências cabem a cada uma delas**. Ex: *Quais as causas que definem o preço de uma mercadoria quando a sua oferta diminui?*

- Seria impossível, no momento da pesquisa, manter constantes a uniformidade dos salários, o gosto dos consumidores, nível geral de preços de outros produtos, etc.

A ESTATÍSTICA

➔ É uma parte da matemática aplicada **que fornece métodos para coleta, organização, descrição, análise e interpretação de dados** e para a **utilização dos mesmos na tomada de decisões.**

➔ A ***coleta, a organização, a descrição dos dados, o cálculo e a interpretação de coeficientes*** pertencem à **ESTATÍSTICA DESCRITIVA**, enquanto a **análise e a interpretação dos dados, associado a uma margem de incerteza**, ficam a cargo da **ESTATÍSTICA INDUTIVA** ou **INFERENCIAL**, também chamada como a medida da incerteza ou métodos que se fundamentam na teoria da probabilidade.

2. ORGANIZAÇÃO DE DADOS ESTATÍSTICOS

FASES DO MÉTODO ESTATÍSTICO

1º - DEFINIÇÃO DO PROBLEMA: Saber exatamente aquilo que se pretende pesquisar é o mesmo que definir corretamente o problema.

2º - PLANEJAMENTO: Como levantar informações? Que dados deverão ser obtidos? Qual levantamento a ser utilizado? Censitário? Por amostragem? E o cronograma de atividades? Os custos envolvidos? etc.

3º - COLETA DE DADOS: Fase operacional. É o registro sistemático de dados, com um objetivo determinado.

Dados primários: quando são publicados pela própria pessoa ou organização que os haja recolhido. *Ex: tabelas do censo demográfico do IBGE.*

Dados secundários: quando são publicados por outra organização. *Ex: quando determinado jornal publica estatísticas referentes ao censo demográfico extraídas do IBGE.*

OBS: É mais seguro trabalhar com fontes primárias. O uso da fonte secundária traz o grande risco de erros de transcrição.

Coleta Direta: quando é obtida diretamente da fonte. *Ex: Empresa que realiza uma pesquisa para saber a preferência dos consumidores pela sua marca.*

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, ou indícios.

4º - APURAÇÃO DOS DADOS: Resumo dos dados através de sua contagem e agrupamento. É a condensação e tabulação de dados.

5º - 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.

6º - 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).

DEFINIÇÕES BÁSICAS DA ESTATÍSTICA

FENÔMENO ESTATÍSTICO: é qualquer evento que se pretenda analisar, cujo estudo seja possível a aplicação do método estatístico. São divididos em três grupos:

Fenômenos de massa ou coletivo: São aqueles que não podem ser definidos por uma simples observação. A estatística dedica-se ao estudo desses fenômenos. *Ex: A natalidade na Grande Vitória, O preço médio da cerveja no Espírito Santo, etc.*

Fenômenos individuais: São aqueles que irão compor os fenômenos de massa. *Ex: cada nascimento na Grande Vitória, cada preço de cerveja no Espírito Santo, etc.*

Fenômenos de multidão: Quando as características observadas para a massa não se verificam para o particular.

DADO ESTATÍSTICO: É um **dado numérico** é a **matéria-prima sobre a qual iremos aplicar os métodos estatísticos.**

POPULAÇÃO: É o **conjunto total de elementos portadores** de, pelo menos, **uma característica comum**.

AMOSTRA: É uma **parcela representativa da população** que **É EXAMINADA** com o propósito de tirarmos conclusões sobre a essa população.

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 2º semestre 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**.

VARIÁVEL QUALITATIVA: Quando seu valor é expresso 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: Seus valores são expressos geralmente através de números inteiros não negativos. Resulta normalmente de contagens. *<u>Ex:</u> Nº de alunos presentes às aulas de estatística no semestre,a quantidade de carros no estacionamento,etc.*

VARIÁVEL CONTÍNUA: Resulta normalmente de uma **mensuração**, e a **escala numérica de seus possíveis valores corresponde ao conjunto <u>R</u>** dos números Reais, ou seja, podem assumir, teoricamente, qualquer valor entre dois limites. *<u>Ex.:</u> 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, o peso de uma carga em toneladas, etc.*

<u>Exemplos</u> **-**

. Cor dos olhos das alunas:	**qualitativa**
. Índice de liquidez nas indústrias capixabas:	**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**

<u>AMOSTRAGEM</u>

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 <u>probabilidade de cada elemento ser selecionado</u> 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 freqüentemente utilizado. É **equivalente a um sorteio lotérico**. Pode ser realizada numerando-se a população de **<u>1 a n</u>** e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer x números dessa seqüência, os quais corresponderão aos elementos pertencentes à amostra.

<u>Ex:</u> 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.

OBS: quando o número de elementos da amostra é muito grande, esse tipo de sorteio torna-se muito trabalhoso. Neste caso utiliza-se uma Tabela de números aleatórios, construída de modo que os algarismos de 0 a 9 são distribuídos ao acaso nas linhas e colunas.

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

Numeramos então os alunos de 01 a 90, sendo 01 a 54 meninos e 55 a 90, meninas e procedemos ao sorteio casual com urna ou tabela de números aleatórios.

AMOSTRAGEM SISTEMÁTICA:

➔ Quando os elementos da população já se acham ordenados, não há necessidade de construir o sistema de referência. São exemplos os prontuários médicos de um hospital, os prédios de uma rua, etc. Nestes casos, a seleção dos elementos que constituirão a amostra pode ser feita por um sistema imposto pelo pesquisador.

Ex: Suponhamos uma rua com 900 casas, das quais desejamos obter uma amostra formada por 50 casas para uma pesquisa de opinião. Podemos, neste caso, usar o seguinte procedimento: como 900/50 = 18, escolhemos por sorteio casual um número de 01 a 18, o qual indicaria o primeiro elemento sorteado para a amostra; os demais elementos seriam periodicamente considerados de 18 em 18. Assim, suponhamos que o número sorteado fosse 4 a amostra seria: 4ª casa, 22ª casa, 40ª casa, 58ª casa, 76ª casa, etc.

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 quarteirões, famílias, organizações, agências, edifícios etc.

Ex: *Num levantamento da população de determinada cidade, podemos dispor do mapa indicando cada quarteirão e não dispor de uma relação atualizada dos seus moradores. Pode-se, então, colher uma amostra dos quarteirões e fazer a contagem completa de todos os que residem naqueles quarteirões sorteados.*

MÉTODOS NÃO PROBABILÍSITCOS

➔ 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

➔ Trata-se de uma **amostra formada por aqueles elementos que vão aparecendo**, que são possíveis de se obter até completar o número de elementos da amostra. **Geralmente utilizada em pesquisas de opinião**, em que os entrevistados são acidentalmente escolhidos.

Ex: *Pesquisas de opinião em praças públicas, ruas de grandes cidades*;

AMOSTRAGEM INTENCIONAL

➔ De acordo com determinado critério, **é escolhido intencionalmente um grupo de elementos que irão compor a amostra**. O investigador se dirige intencionalmente a grupos de elementos dos quais deseja saber a opinião.

Ex: *Numa pesquisa sobre preferência por determinado cosmético, o pesquisador se dirige a um grande salão de beleza e entrevista as pessoas que ali se encontram.*

AMOSTRAGEM POR QUOTAS

➔ Um dos métodos de **amostragem mais comumente usados em levantamentos de mercado** e em **prévias eleitorais**. Ele abrange três fases:

1ª - classificação da população em termos de propriedades que se sabe, ou presume, serem relevantes para a característica a ser estudada;

2ª - determinação da proporção da população para cada característica, com base na constituição conhecida, presumida ou estimada, da população;

3ª - fixação de quotas para cada entrevistador a quem tocará a responsabilidade de selecionar entrevistados, de modo que a amostra total observada ou entrevistada contenha a proporção e cada classe tal como determinada na 2ª fase.

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.*

*A primeira tarefa é descobrir as proporções (porcentagens) dessas características na população. Imagina-se que haja 47% de homens e 53% de mulheres na população. Logo, uma amostra de 50 pessoas deverá ter 23 homens e 27 mulheres. Então o pesquisador receberá uma "quota" para entrevistar 27 mulheres. A consideração de várias categorias exigirá uma **composição amostral** que atenda ao **n** determinado e às proporções populacionais estipuladas.*

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

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.

a) **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**.

Uma loja de departamentos

Vendas no 1º bimestre de 2007

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

b) **Série Geográfica**: Apresenta como **elemento variável o fator geográfico**. A época e o fato (espécie) são elementos fixos. Também é chamada de espacial, territorial ou de localização.

Uma Grande rede de lojas.

Vendas no 1º bimestre de 1996

FILIAIS	UNIDADES VENDIDAS
São Paulo	13000
Rio de Janeiro	17000
TOTAL	30000

c) **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 1996

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**.

ABC VEÍCULOS LTDA.

Vendas no 1º bimestre de 2007

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

GRÁFICOS ESTATÍSTICOS

➔ 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.

Gráficos de informação: São gráficos destinados principalmente ao público em geral, objetivando proporcionar uma visualização rápida e clara. São gráficos tipicamente expositivos, dispensando comentários explicativos adicionais. As legendas podem ser omitidas, desde que as informações desejadas estejam presentes.

Gráficos de análise: São gráficos que se prestam melhor ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser também informativos. Os gráficos de análise freqüentemente vêm acompanhados de uma tabela estatística. Inclui-se, muitas vezes um texto explicativo, chamando a atenção do leitor para os pontos principais revelados pelo gráfico.

- **Uso indevido de Gráficos:** Podem trazer uma idéia falsa dos dados que estão sendo analisados, chegando mesmo a confundir o leitor. Trata-se, na realidade, de um problema de construção de escalas.

Classificação dos gráficos: Diagramas, Estereogramas, Pictogramas e Cartogramas.

1 - DIAGRAMAS:

➔ São gráficos geométricos dispostos em duas dimensões. São os mais usados na representação de séries estatísticas. Eles podem ser:

1.1- **Gráficos em barras horizontais**.

1.2- **Gráficos em barras verticais (colunas).**

- Quando as legendas não são breves usam-se de preferência os gráficos em barras horizontais. Nesses gráficos os retângulos têm a mesma base e as alturas são proporcionais aos respectivos dados.
  - A ordem a ser observada é a cronológica, se a série for histórica, e a

    Decrescente, se for geográfica ou categórica.

1.3- **Gráficos em colunas superpostas**.

- Eles diferem dos gráficos em barras ou colunas convencionais apenas pelo fato de apresentar cada barra ou coluna segmentada em partes componentes. Servem para representar comparativamente dois ou mais atributos.

1.4- **Gráficos em linhas ou lineares**.

São freqüentemente usados para representação de séries cronológicas com um grande número de períodos de tempo. As linhas são mais eficientes do que as colunas, quando existem intensas flutuações nas séries ou quando há necessidade de se representarem várias séries em um mesmo gráfico.

- Quando representamos, em um mesmo sistema de coordenadas, a variação de dois fenômenos, a parte interna da figura formada pelos gráficos desses fenômenos é denominada de ***área de excesso.***

1.5- **Gráficos em setores**.

- Este gráfico é construído com base em um círculo, e é empregado sempre que desejamos ressaltar a participação do dado no total. O total é representado pelo círculo, que fica dividido em tantos setores quantas são as partes. Os setores são tais que suas áreas são respectivamente proporcionais aos dados da série. O gráfico em setores só deve ser empregado quando há, no máximo, sete dados.

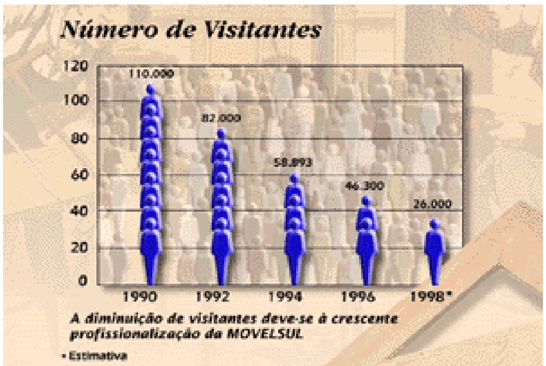
- **Obs:** As séries temporais geralmente não são representadas por este tipo de gráfico.

2 - 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.

3 - 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 auto-explicativos. A desvantagem dos pictogramas é que apenas mostram uma visão geral do fenômeno, e não de detalhes minuciosos. Veja o exemplo abaixo:



4- 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.

DISTRIBUIÇÃO DE FREQÜÊNCIA

➔ É um tipo de tabela **que condensa uma coleção de dados conforme as freqüências** (repetições de seus valores).

Tabela primitiva ou dados brutos: É uma **tabela ou relação de elementos que não foram numericamente organizados**. É difícil formarmos uma idéia exata do comportamento do grupo como um todo, a partir de dados não ordenados.

Ex : 45, 41, 42, 41, 42 43, 44, 41 ,50, 46, 50, 46, 60, 54, 52, 58, 57, 58, 60, 51

ROL: É a **tabela obtida após a ordenação dos dados** (crescente ou decrescente).

Ex : 41, 41, 41, 42, 42 43, 44, 45 ,46, 46, 50, 50, 51, 52, 54, 57, 58, 58, 60, 60

Distribuição de freqüência SEM INTERVALOS DE CLASSE: É a simples condensação dos dados conforme as repetições de seus valores. Para um **ROL** de tamanho razoável esta distribuição de freqüência é inconveniente, já que exige muito espaço. Veja exemplo abaixo:

Dados	Freqüência
41	3
42	2
43	1
44	1
45	1
46	2
50	2
51	1
52	1
54	1
57	1
58	2
60	2
Total	20

Distribuição de freqüência COM INTERVALOS DE CLASSE: Quando o tamanho da amostra é elevado, é mais racional efetuar o agrupamento dos valores em vários intervalos de classe.

Classes	Freqüências
41 |------- 45	7
45 |------- 49	3
49 |------- 53	4
53 |------- 57	1
57 |------- 61	5
Total	20

ELEMENTOS DE UMA DISTRIBUIÇÃO DE FREQÜÊNCIA (com intervalos de classe) ➔

CLASSE: são os intervalos de variação da variável e é simbolizada por i e o número total de classes simbolizada por k. Ex: na tabela anterior $k = 5$ e **49 |------- 53** é a **3ª classe**, onde $i = 3$.

A quantidade de classes é calculada através da formula de **Sturges** abaixo:

$$K \cong 1 + 3{,}22 \log N$$

onde N é quantidade de elementos da amostra.

LIMITES DE CLASSE: são os extremos de cada classe. O menor número é o limite inferior de classe (**li**) e o maior número, limite superior de classe (**Li**). *Ex: em **49 |------- 53,** **l3** = 49 e **L3** = 53. O símbolo **|-------** representa **um intervalo fechado à esquerda e aberto à direita.***

*O dado 53 do **ROL** não pertence a classe 3 e sim a classe 4 representada por **53 |------- 57**.*

AMPLITUDE DO INTERVALO DE CLASSE: é obtida através da diferença entre o limite superior e inferior da classe e é simbolizada por ***h = Ls - li***. *<u>Ex</u>: na tabela anterior **h** = 53 - 49 = 4. <u>**Obs**</u>: Na distribuição de freqüência c/ classe o **h será igual em todas as classes.***

AMPLITUDE TOTAL DA DISTRIBUIÇÃO: é a diferença entre o limite superior da última classe e o limite inferior da primeira classe. ***A = L(max) - l(min)***. *<u>Ex:</u> na tabela anterior **A** = 61 - 41= 20.*

AMPLITUDE TOTAL DA AMOSTRA (ROL): é a diferença entre o valor máximo e o valor mínimo da amostra (ROL). Onde ***A = Xmax - Xmin***. Em nosso exemplo ***A*** = 60 - 41 = 19. <u>**Obs**</u>: ***A amplitude da distribuição será sempre maior que amplitude da amostra.***

PONTO MÉDIO DE CLASSE: é o ponto que divide o intervalo de classe em duas partes iguais. *<u>Ex:</u> em **49 |------- 53** o ponto médio **x3** = (53+49)/2 = 51, ou seja **x3=(l3 + L3)/2.***

Método prático para construção de uma Distribuição de Freqüências c/ Classe ➔

1º - Organize os dados brutos em um **ROL**.

2º - Calcule a **amplitude amostral *A***.

- No nosso exemplo: ***A*** = 60 - 41 = 19

3º - Calcule o número de classes através da "**Regra de Sturges**":

K ≅ 1 + 3,22 logN para N =20 temos K ≅ 6

Obs: Qualquer regra para determinação do nº de classes da tabela não nos levam a uma decisão final; esta vai depender na realidade de um julgamento pessoal, que deve estar ligado à natureza dos dados.

No nosso exemplo: ***n*** = 20 dados, então, a princípio, a regra sugere a adoção de 6 classes.

4º Decidido o nº de classes, calcule então a amplitude do intervalo de classe **h > A/K**

No nosso exemplo: **A/K** = 19/6 = 3,16. <u>Obs.</u>: Como **h > A/K** um valor ligeiramente superior para haver folga na última classe. Utilizaremos então ***h*** = 4

5º - Temos então o menor nº da amostra, o nº de classes e a amplitude do intervalo. Podemos montar a tabela.

No nosso exemplo: o menor nº da amostra = 41 + ***h*** = 45, logo a primeira classe será representada por **41 |------- 45**. As classes seguintes respeitarão o mesmo procedimento.

O primeiro elemento das classes seguintes sempre será formado pelo último elemento da classe anterior.

Classes
41 |----- 45
45 |----- 49
49 |----- 53
53 |----- 57
57 |----- 61

REPRESENTAÇÃO GRÁFICA DE UMA DISTRIBUIÇÃO

Histograma, Polígono de freqüência e Polígono de freqüência acumulada

➔ Em todos os gráficos acima utilizamos o primeiro quadrante do sistema de eixos coordenados cartesianos ortogonais. Na linha horizontal (eixo das abscissas) colocamos os valores da variável e na linha vertical (eixo das ordenadas), as freqüências.

Histograma: É formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal, de tal modo que seus pontos médios coincidam com os pontos médios dos intervalos de classe. A área de um histograma é proporcional à soma das freqüências simples ou absolutas.

Freqüências simples ou absoluta F_i: são os valores que realmente representam o número de dados de cada classe. A soma das freqüências simples é igual ao número total dos dados da distribuição.

Freqüências relativas f_i: são os valores das razões entre a freqüência absolutas de cada classe e a freqüência total da distribuição. A soma das **freqüências relativas é igual a 1 (100 %).**

Polígono de freqüência: é um gráfico em linha, sendo as freqüências 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.

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

Freqüência simples acumulada de uma classe F_{ac}: é o total das freqüências de todos os valores inferiores ao limite superior do intervalo de uma determinada classe.

Freqüência relativa acumulada de um classe frc: é a freqüência acumulada da classe, dividida pela freqüência total da distribuição.

CLASSE	fi	xi	fr	F_{ac}	Fr_{ac}
50 |-------- 54	4	52	0,100	4	0,100
54 |-------- 58	9	56	0,225	13	0,325
58 |-------- 62	11	60	0,275	24	0,600
62 |-------- 66	8	64	0,200	32	0,800
66 |-------- 70	5	68	0,125	37	0,925
70 |-------- 74	3	72	0,075	40	1,000
Total	40		1,000		

Fi = freqüência absoluta; **xi** = ponto médio de classe; F_{ac} = freqüência acumulada;

fr = freqüência relativa e fr_{ac} = freqüência relativa acumulada.

- **Obs:** uma distribuição de freqüência sem intervalos de classe é representada graficamente por um diagrama onde cada valor da variável é representado por um segmento de reta vertical e de comprimento proporcional à respectiva freqüência.

3. MEDIDAS DE POSIÇÃO

Introdução

➔ São as estatísticas que representam uma série de dados orientando-nos quanto à posição da distribuição em relação ao eixo horizontal do gráfico da curva de freqüência.

- As medidas de posições mais importantes são as **medidas de tendência central** (verifica-se uma tendência dos dados observados a se agruparem em torno dos valores centrais).
- As medidas de tendência centrais mais utilizadas são**: média aritmética**, **moda** e **mediana**. Outros promédios menos usados são as médias: **geométrica**, **harmônica**, **quadrática, cúbica e biquadrática**.
- As outras medidas de posição são as **separatrizes**, que englobam: a própria **mediana**, os **decis**, os **quartis** e os **percentis**.

MÉDIA ARITMÉTICA = $\overline{X}$

➔ É igual ao quociente entre a soma dos valores do conjunto e o número total dos valores.

$$\bar{X} = \frac{\sum X}{N}$$

Para dados não agrupados já

para dados agrupados temos a seguinte fórmula:

$$\bar{X} = \frac{\sum x_{i F_i}}{\sum F_i}$$

onde **xi** são os valores da variável e **n** o número de valores.

Dados não-agrupados: Quando desejamos conhecer a média dos dados não-agrupados em tabelas de freqüências, determinamos a **média aritmética simples**.

Ex: Sabendo-se que a venda diária de arroz tipo A, durante uma semana, foi de 10, 14, 13, 15, 16, 18 e 12 kilos, temos, para venda média diária na semana de:

$\bar{X}$ *= (10+14+13+15+16+18+12) / 7 = 14 kilos*

Desvio em relação à média: é a diferença entre cada elemento de um conjunto de valores e a média aritmética, ou seja:

di = Xi - $\bar{X}$

No exemplo anterior temos sete desvios: $d_1 = 10 - 14 = -4$, $d_2 = 14 - 14 = 0$, $d_3 = 13 - 14 = -1$, $d_4 = 15 - 14 = 1$, $d_5 = 16 - 14 = 2$, $d_6 = 18 - 14 = 4$ *e* $d_7 = 12 - 14 = -2$.

Propriedades da média aritmética ➔

1ª propriedade: A soma algébrica dos desvios em relação à média é nula.

- No exemplo anterior: $d_1+d_2+d_3+d_4+d_5+d_6+d_7 = 0$

2ª propriedade: Somando-se (ou subtraindo-se) uma constante (c) a todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante.

- Se no exemplo original somarmos a constante **2** a cada um dos valores da variável temos:
- Y = 12+16+15+17+18+20+14 / 7 = 16 kilos ou

Y = $\bar{X}$ + 2 = 14 +2 = 16 kilos

3ª propriedade: Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada (ou dividida) por essa constante.

- Se no exemplo original multiplicarmos a constante **3** a cada um dos valores da variável temos:

Y = 30+42+39+45+48+54+36 / 7 = 42 kilos ou

Y = $\bar{X}$x 3 = 14 x 3 = 42 kilos

Dados agrupados:

Sem intervalos de classe ➔ Consideremos a distribuição relativa a 34 famílias de quatro filhos, tomando para variável o número de filhos do sexo masculino. Calcularemos a quantidade média de meninos por família:

Nº de meninos	freqüência = fi
0	2
1	6
2	10
3	12
4	4
total	**34**

- Como as freqüências são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva a calcular a **média aritmética ponderada**, dada pela fórmula:

$$\bar{X} = \Sigma\, xi.fi \,/\, \Sigma fi$$

xi	fi	xi.fi
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
total	**34**	**78**

onde 78 / 34 = **2,3 meninos por família**

Com intervalos de classe ➔ Neste caso, convencionamos que todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio, e determinamos a média aritmética ponderada por meio da fórmula:

$$\bar{X} = \frac{\sum X_{iF_i}}{\sum F_i}$$

onde **Xi** é o ponto médio da classe.

Ex: Calcular a estatura média de bebês conforme a tabela abaixo.

Estaturas (cm)	freqüência = **fi**	ponto médio = **xi**	**xi.fi**
50 |------------ 54	4	52	208
54 |------------ 58	9	56	504
58 |------------ 62	11	60	660
62 |------------ 66	8	64	512
66 |------------ 70	5	68	340
70 |------------ 74	3	72	216
Total	40		2.440

Aplicando a fórmula acima temos: 2.440 / 40 = 61 logo $\overline{X}$ **= 61 cm**

MÉDIA GEOMÉTRICA **=** $\overline{X}g$

➔ É a raiz n-ésima do produto de todos eles.

Média Geométrica Simples: $\overline{X}g = \sqrt[n]{X1 \, . \, X2 \, . \, ... \, . \, Xn}$ ou $\overline{X}g = (X1 \, . \, X2. \, ... \, . \, Xn)^{1/n}$

Ex.: - Calcular a média geométrica dos seguintes conjuntos de números:

a) { 10, 60, 360 } : = $\sqrt[3]{(10 * 60 * 360}$ = ***60***

b) { 2, 2, 2 } : $\sqrt[3]{2 * 2 * 2 *}$ = ***2***

c) { 1, 4, 16, 64 }: $\sqrt[4]{1 * 4 * 16 * 64}$ = ***8***

Média Geométrica Ponderada:

$\overline{X}gp = \sqrt[\sum fi]{X1^{f1} . X2^{f2} . \, ... \, . \, Xi^{fi}}$ ou $\overline{X}gp = (X1^{f1} . X2^{f2} . \, ... \, . \, Xi^{fi})^{\frac{1}{\sum fi}}$

Ex - Calcular a média geométrica dos valores da tabela abaixo:

xi	**fi**
1	2
3	4
9	2
27	1

Total	9

$$\sqrt[9]{1^2 * 3^4 * 9^2 * 27^1} = 3{,}8296$$

MÉDIA HARMÔNICA - $\overline{X}h$

➔ É o inverso da média aritmética dos inversos.

Média Harmônica Simples: (para dados não agrupados)

$$\overline{X}h = \frac{1}{\dfrac{1/X1 + 1/X2 + \ldots + 1/Xn}{n}} \quad \text{ou} \quad \overline{X}h = \frac{n}{1/X1 + 1/X2 + \ldots + 1/Xn}$$

Média Harmônica Ponderada : (para dados agrupados em tabelas de freqüências)

$$\overline{X}hp = \frac{\sum fi}{\sum fi/xi}$$

Ex.: Calcular a média harmônica dos valores da tabela abaixo:

classes	fi	xi	fi/xi
1 |--------- 3	2	2	2/2 = 1,00
3 |--------- 5	4	4	4/4 = 1,00
5 |--------- 7	8	6	8/6 = 1,33
7 |--------- 9	4	8	4/8 = 0,50
9 |--------- 11	2	10	2/10 = 0,20
total	**20**		**4,03**

Resp: 20 / 4,03 = 4,96

<u>OBS</u>: A média harmônica não aceita valores iguais a zero como dados de uma série.

A igualdade $\overline{X}g = \overline{X}h = \overline{X}$ Só ocorrerá quando todos os valores da série forem iguais.

<u>OBS:</u> Quando os valores da variável não forem muito diferentes, verifica-se aproximadamente a seguinte relação:

$$\boxed{\overline{X}g = (\overline{X} + \overline{X}h) / 2}$$

- Demonstraremos a relação acima com os seguintes dados:

z = { 10,1 ; 10,1 ; 10,2 ; 10,4 ; 10,5 }

Média aritmética = 51,3 / 5 = **10,2600**

Média geométrica= = **10,2587**

Média harmônica = 5 / 0,4874508 = **10,2574**

Comprovando a relação: 10,2600 + 10,2574 / 2 = 10,2587 = média geométrica

MODA - Mo

➔ É o **valor que ocorre com maior freqüência em uma série de valores**.

- Desse modo, o salário modal dos empregados de uma fábrica é o salário mais comum, isto é, o salário recebido pelo maior número de empregados dessa fábrica.

A Moda quando os dados não estão agrupados ➔

- A moda é facilmente reconhecida: basta, de acordo com definição, procurar o valor que mais se repete.

*Ex: Na série { 7 , 8 , 9 , 10 , 10 , 10 , 11 , 12 } a moda é igual a **10.***

- Há séries nas quais não exista valor modal, isto é, nas quais nenhum valor apareça mais vezes que outros.

*Ex: { 3 , 5 , 8 , 10 , 12 } **não apresenta moda**. A série é **amodal**.*

- Em outros casos, **pode haver dois ou mais valores de concentração**. Dizemos, então, que a série tem dois ou mais valores modais.

*Ex: { 2 , 3 , 4 , 4 , 4 , 5 , 6 , 7 , 7 , 7 , 8 , 9 } apresenta duas modas**: 4** e **7**. A série **bimodal**.*

A Moda quando os dados estão agrupados ➔

a) Sem intervalos de classe: Uma vez agrupados os dados, é possível determinar imediatamente a moda: **basta fixar o valor da variável de maior freqüência.**

Ex: Qual a temperatura mais comum medida no mês abaixo:

Temperaturas	Freqüência
0º C	3
1º C	9
2º C	12

3º C	6

*Resp: **2º C** é a temperatura modal, pois é a de maior freqüência.*

b) Com intervalos de classe: A classe que apresenta a maior freqüência é denominada **classe modal.** Pela definição, podemos afirmar que a **moda**, neste caso, é o valor dominante que está **compreendido entre os limites da classe modal**. O método mais simples para o cálculo da moda consiste em tomar o ponto médio da classe modal. Damos a esse valor a denominação de **moda bruta**.

Mo = (l* + L*) / 2

onde **l*** = *limite inferior da classe modal* e **L* =** *limite superior da classe modal.*

Ex: Calcule a estatura modal conforme a tabela abaixo.

Classes (em cm)	Freqüência
54 |------------ 58	9
58 |------------ 62	11
62 |------------ 66	8
66 |------------ 70	5

Resposta: a classe modal é **58|-------- 62**, pois é a de maior freqüência**. l* = 58** e **L* = 62**

Mo = (58+62) / 2 = 60 cm (este valor é estimado, pois não conhecemos o valor real da moda).

Método mais elaborado pela fórmula de CZUBER: **Mo =** $L_{mo} + \frac{\Delta_1}{\Delta_1 + \Delta_2} h$

L_{mo} =limite inferior da classe modal

Δ_1= freqüência da classe modal - freqüência da **classe anterior** à da classe modal

Δ_2= freqüência da classe modal - freqüência da **classe posterior** à da classe modal

h = amplitude da classe modal

Mo = 58 + ((11-9) / ((11-9) + (11 – 8)) x 4 ➔ **Mo = 59,6**

Obs: A **moda** é utilizada quando desejamos obter uma medida rápida e aproximada de posição ou quando a medida de posição deva ser o valor mais típico da distribuição. Já a **média aritmética** é a medida de posição que possui a maior estabilidade.

MEDIANA - Md

➔ A **mediana de um conjunto de valores**, dispostos segundo uma ordem (crescente ou decrescente), é o valor situado de tal forma no conjunto **que o separa em dois subconjuntos de mesmo número de elementos.**

A mediana em dados não-agrupados ➔

Dada uma série de valores como, por exemplo: { 5, 2, 6, 13, 9, 15, 10 }

De acordo com a definição de mediana, o primeiro passo a ser dado é o da ordenação (crescente ou decrescente) dos valores: { 2, 5, 6, 9, 10, 13, 15 }

O valor que divide a série acima em duas partes iguais é igual a **9**, logo a **Md = 9**.

Método prático para o cálculo da Mediana:

➔ **Se a série dada tiver número ímpar de termos:** O valor mediano será o termo de ordem dado pela fórmula e será o termo central da amostra:

$$\boxed{(n + 1) / 2}$$

Ex: Calcule a mediana da série { 1, 3, 0, 0, 2, 4, 1, 2, 5 }

1º - ordenar a série { 0, 0, 1, 1, 2, 2, 3, 4, 5 }

n = 9 logo (n + 1)/2 é dado por (9+1) / 2 = 5, ou seja, o 5º elemento da série ordenada será a mediana

*A mediana será o **5º elemento = 2 que é o elemento central***

Se a série dada tiver número par de termos: O valor mediano será o termo de ordem dado pela fórmula e será a média dos dois elementos centrais:

$$\boxed{[(n/2) + (n/2 + 1)] / 2}$$

Obs: n/2 e (n/2 + 1) serão termos de ordem e devem ser substituídos pelo valor correspondente.

Ex: Calcule a mediana da série { 1, 3, 0, 0, 2, 4, 1, 3, 5, 6 }

1º - ordenar a série { 0, 0, 1, 1, 2, 3, 3, 4, 5, 6 } *n = 10 logo a fórmula ficará: [(10/2) + (10/2 + 1)] / 2*

[(5 + 6)] / 2 será na realidade (5º termo+ 6º termo) / 2

5º termo = 2

6º termo = 3

A mediana será = (2+3) / 2 ou seja, ***Md = 2,5*** *. A mediana no exemplo será a média aritmética do 5º e 6º termos da série os quais são os dois elementos centrais.*

Notas:

- Quando o número de elementos da série estatística for ímpar, haverá coincidência da mediana com um dos elementos da série.

- Quando o número de elementos da série estatística for par, nunca haverá coincidência da mediana com um dos elementos da série. **A mediana será sempre a média aritmética dos 2 elementos centrais da série.**

- Em uma série **a mediana, a média e a moda** não têm, necessariamente, o mesmo valor.

- A **mediana, depende da posição e não dos valores dos elementos** na série ordenada. Essa é uma da diferenças marcantes entre **mediana** e **média (que se deixa influenciar, e muito, pelos valores extremos).** Vejamos:

Em { 5, 7, 10, 13, 15 } a **média = 10** e **a mediana = 10**

Em { 5, 7, 10, 13, 65 } a **média = 20** e a **mediana = 10**

- isto é, a média do segundo conjunto de valores é maior do que a do primeiro, por influência dos valores extremos, ao passo que a mediana permanece a mesma.

A mediana em dados agrupados ➔

a) Sem intervalos de classe: Neste caso, é o bastante identificar a freqüência acumulada imediatamente superior à metade da soma das freqüências. A **mediana** será aquele valor da variável que corresponde a tal freqüência acumulada.

Ex.: conforme tabela abaixo:

Variável xi	Freqüência fi	Freqüência acumulada
0	2	2
1	6	8
2	9	17

3	13	30
4	5	35
total	**35**	

- Quando o somatório das freqüências for **ímpar** o valor mediano será o termo de ordem dado pela fórmula:

$$\frac{\sum fi + 1}{2}$$

Como o somatório das freqüências = 35 a fórmula ficará: (35+1) / 2 = **18º termo = 3**

- Quando o somatório das freqüências for **par** o valor mediano será o termo de ordem dado pela fórmula:

$$\frac{(\sum fi/2)+(\sum fi/2+1)}{2}$$

<u>Ex</u>: Calcule Mediana da tabela abaixo:

Variável xi	Freqüência fi	Freqüência acumulada
12	1	1
14	2	3
15	1	4
16	2	6
17	1	7
20	1	8
total	**8**	

- Aplicando fórmula acima teremos:[(8/2)+ (8/2+1)]/2 = (4º termo + 5º termo) / 2 = (15 + 16) / 2 = **15,5**

b) Com intervalos de classe: Devemos seguir os seguintes passos:

1º) Calculamos $\sum fi / 2$; ou N/2

2º) Marcamos a classe correspondente à freqüência acumulada imediatamente superior à $\sum fi / 2$. Tal classe será a classe mediana

3º) Calculamos a Mediana pela seguinte fórmula:

$$Md = L_{md} + \frac{\left(\frac{N}{2} - \sum F\right)h}{F_{md}}$$

L_{md} = é o **limite inferior** da **classe mediana**.

$\sum F$ = é a **freqüência acumulada** da classe anterior à **classe mediana**.

F_{md} = é a **freqüência absoluta** da **classe mediana.**

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

<u>*Ex*</u>:

classes	freqüência = fi	Freqüência acumulada
50 |------------ 54	4	4
54 |------------ 58	9	13
58 |------------ 62	**11**	**24**
62 |------------ 66	8	32
66 |------------ 70	5	37
70 |------------ 74	3	40
total	**40**	

$\sum fi / 2$ = 40 / 2 = 20 logo a classe mediana será **58 |---------- 62**

L_{md} = 58 $\sum F$ = 13 F_{md} = 11 h = 4

Substituindo esses valores na fórmula, obtemos:

Md = 58 + [(20 - 13) x 4] / 11 = 58 + 28/11 = **60,54**

OBS: Esta mediana é estimada, pois não temos os 40 valores da distribuição.

Emprego da Mediana

- Quando desejamos obter o ponto que divide a distribuição em duas partes iguais.
- Quando há valores extremos que afetam de maneira acentuada a média aritmética.
- Quando a variável em estudo é salário.

SEPARATRIZES

➔ Além das medidas de posição que estudamos, há outras que, consideradas individualmente, **não são medidas de tendência central**, mas estão ligadas à mediana relativamente à sua característica de separar a série em duas partes que apresentam o mesmo número de valores.

Essas medidas - **os quartis, os decis e os percentis** - são, juntamente com a **mediana**, conhecidas pelo nome genérico de **separatrizes**.

QUARTIS - Q

➔ Denominamos **quartis** os valores de uma série que a **dividem em quatro partes iguais**. Precisamos portanto de **3 quartis (Q1 , Q2 e Q3)** para dividir a série em quatro partes iguais.

Obs: O **quartil 2 (Q2) SEMPRE SERÁ IGUAL A MEDIANA DA SÉRIE**.

Quartis em dados não agrupados ➔

➔ O método mais prático é utilizar **o princípio do cálculo da mediana** para os **3 quartis**. Na realidade **serão calculadas " 3 medianas " em uma mesma série**.

*Ex 1: Calcule os **quartis** da série: { 5, 2, 6, 9, 10, 13, 15 }*

- O primeiro passo a ser dado é o da ordenação (crescente ou decrescente) dos valores: { 2, 5, 6, 9, 10, 13, 15 }

*- O valor que divide a série acima em duas partes iguais é igual a **9**, logo a **Md = 9 que será = Q2 = 9***

*- Temos agora **{2, 5, 6 }** e **{10, 13, 15** } como sendo os dois grupos de valores iguais proporcionados pela mediana (**quartil 2**). Para o cálculo do **quartil 1 e 3** basta **calcular as medianas das partes iguais provenientes da verdadeira Mediana da série (quartil 2).***

*Logo em **{ 2, 5, 6 }** a mediana é = **5** . Ou seja: será o **quartil 1 = Q1 = 5***

*em **{10, 13, 15 }** a mediana é =**13** . Ou seja: será o **quartil 3 = Q = 13***

*Ex 2: Calcule os **quartis** da série: { 1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13 }*

- *A série já está ordenada, então calcularemos o **Quartil 2 = Md = (5+6)/2 = 5,5***
-

*- O **quartil 1** será a mediana da série à esquerda de Md **: { 1, 1, 2, 3, 5, 5 }***

Q1 = (2+3)/2 = 2,5

*- O **quartil 3** será a mediana da série à direita de Md : **{6, 7, 9, 9, 10, 13 }***

Q3 = (9+9)/2 = 9

Quartis para dados agrupados em classes ➔

➔ Usamos a mesma técnica do cálculo da mediana, bastando substituir, na fórmula da mediana,N/4 se for o primeiro quartil.

$$Q_1 = L_{Q1} + \frac{\left(\frac{N}{4} - \sum F\right)h}{F_{Q1}}$$ onde

L_{Q1} = é o **limite inferior** da **classe** onde esta o primeiro quartil.

$\sum F$ = é a **freqüência acumulada** da classe anterior à **classe onde esta o primeiro quartil**

F_{Q1} = é a **freqüência absoluta** da **classe onde esta o primeiro quartil.**

h = é a amplitude do intervalo da **classe da classe onde esta o primeiro quartil**.

3N/4 se for o terceiro quartil

$$Q_3 = L_{Q3} + \frac{\left(\frac{3N}{4} - \sum F\right)h}{F_{Q3}}$$ onde

L_{Q3} = é o **limite inferior** da **classe** onde esta o terceiro quartil.

$\sum F$ = é a **freqüência acumulada** da classe anterior à **classe onde esta o terceiro quartil**

F_{Q3} = é a **freqüência absoluta** da **classe onde esta o terceiro quartil.**

h = é a amplitude do intervalo da **classe da classe onde esta o terceiro quartil**

<u>Ex 3</u> - Calcule os quartis da tabela abaixo:

classes	freqüência = fi	Freqüência acumulada
50 |------------ 54	4	4
54 |------------ 58	9	13

58 |------------ 62	**11**	**24**
62 |------------ 66	8	32
66 |------------ 70	5	37
70 |------------ 74	3	40
total	**40**	

- O **quartil 2 = Md** , logo:

$\sum fi / 2$ = 40 / 2 = 20 logo a classe mediana será **58 |---------- 62**

L_{md} = 58 $\sum F$ = 13 F_{md} = 11 h = 4

- Substituindo esses valores na fórmula para a mediana a qual também é denominada, de segundo quartil obtemos:

Md = 58 + [(20 - 13) x 4] / 11 = 58 + 28/11 = **60,54 = Q2**

- O primeiro **quartil:** N/ **4** = 10

Q1 = 54 + [(10 - 4) x 4] / 9 = 54 + 2,66 = **56,66 = Q1**

- O terceiro **quartil 3**: **3. N** / **4** = 30

Q3 = 62 + [(30 -24) x 4] / 8 = 62 + 3 = **65 = Q3**

<u>DECIS</u> - D

➔ A definição dos **decis** obedece ao mesmo princípio dos **quartis**, com a modificação da porcentagem de valores que ficam aquém e além do **decil** que se pretende calcular. A fórmula básica será : i **.N** / **10** onde i é o número de ordem do **decil** a ser calculado. Indicamos os **decis : D1, D2, ... , D9**. Deste modo precisamos de **9 decis** para **dividirmos uma série em 10 partes iguais**.

- De especial interesse é o **quinto decil,** que ***divide o conjunto em duas partes iguais***. Assim sendo,o **QUINTO DECIL É IGUAL AO SEGUNDO QUARTIL**, que por sua vez **É IGUAL À MEDIANA**.

Para **D5** temos : **5.N** / **10 = N** / **2**

*<u>Ex:</u> Calcule o **3º decil** da tabela anterior com classes.*

***I = 3** onde 3 **.N** / **10 =** 3 x 40 / 10 = **12**.*

Este resultado corresponde a 2ª classe.

D3** = 54 + [(12 - 4) x 4] / 9 = 54 + 3,55 = **57,55 = D3

<u>PERCENTIL ou CENTIL</u>

➔ Denominamos **percentis ou centis** como sendo os noventa e nove valores que separam uma série em 100 partes iguais. Indicamos: P1, P2, ... , P99. É evidente que **P50 = Md** ; **P25 = Q1** e **P75 = Q3.**

- O cálculo de um **centil** segue a **mesma técnica do cálculo da mediana**, porém a fórmula será: i.N/ **100** onde i é o número de ordem do **centil** a ser calculado. Portanto utilizaremos a seguinte fórmula:
- $$P_i = L_{Pi} + \frac{\left(\frac{iN}{100} - \sum F\right)h}{F_{Pi}}$$ onde
- L_{Pi} = é o **limite inferior** da **classe** onde esta o i-ésimo percentil.
- $\sum F$ = é a **freqüência acumulada** da classe anterior à **classe onde esta o i percentil**
- F_{Pi} = é a **freqüência absoluta** da **classe onde esta o i percentil.**
- **h** **= é a amplitude** do intervalo da **classe da classe onde esta o i percentil**.

Dispersão ou Variabilidade: É a maior ou menor diversificação dos valores de uma variável em torno de um valor de tendência central (média ou mediana) tomado como ponto de comparação.

- A **média** - ainda que considerada como um número que tem a faculdade de representar uma série de valores - não pode por si mesma, destacar o grau de homogeneidade ou heterogeneidade que existe entre os valores que compõem o conjunto.
  - **Consideremos os seguintes conjuntos de valores das variáveis X, Y e Z:**

    X = { 70, 70, 70, 70, 70 }

    Y = { 68, 69, 70 ,71 ,72 }

    Z = { 5, 15, 50, 120, 160 }

    - Observamos então que os três conjuntos apresentam a mesma média aritmética = 350/5 **= 70**

- Entretanto, é fácil notar que o **conjunto X é mais homogêneo** que os conjuntos **Y** e **Z**, já ***que todos os valores são iguais à média***. O **conjunto Y**, por sua vez, é mais homogêneo que o **conjunto Z**, pois há menor diversificação entre cada um de seus valores e a média representativa.

  Concluímos então que o **conjunto X** apresenta **DISPERSÃO NULA** e que o **conjunto Y** apresenta uma **DISPERSÃO MENOR** que o **conjunto Z**.

X = { 70, 70, 70, 70, 70 }

Y = { 68, 69, 70 ,71 ,72 }

Z = { 5, 15, 50, 120, 160 }

- Observamos então que os três conjuntos apresentam a mesma média aritmética = 350/5 **= 70**

• Entretanto, é fácil notar que o **conjunto X é mais homogêneo** que os conjuntos **Y** e **Z**, já ***que todos os valores são iguais à média***. O **conjunto Y**, por sua vez, é mais homogêneo que o **conjunto Z**, pois há menor diversificação entre cada um de seus valores e a média representativa.

Concluímos então que o **conjunto X** apresenta **DISPERSÃO NULA** e que o **conjunto Y** apresenta uma **DISPERSÃO MENOR** que o **conjunto Z**.