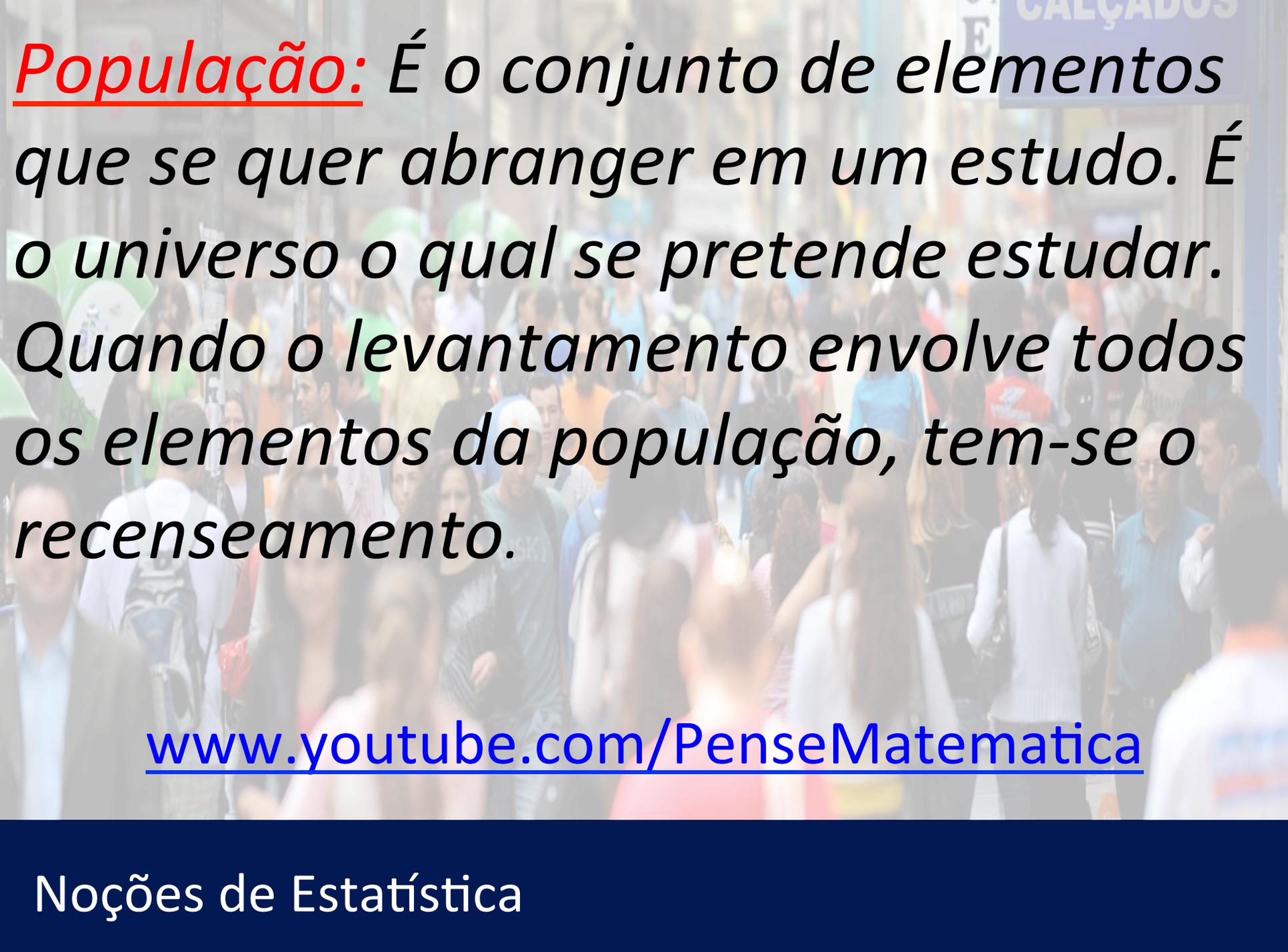


Estatística: *Parte da matemática em que se investigam os processos de obtenção, organização e análise de dados sobre uma população, e os métodos de tirar conclusões e fazer previsões com base nesses dados.*

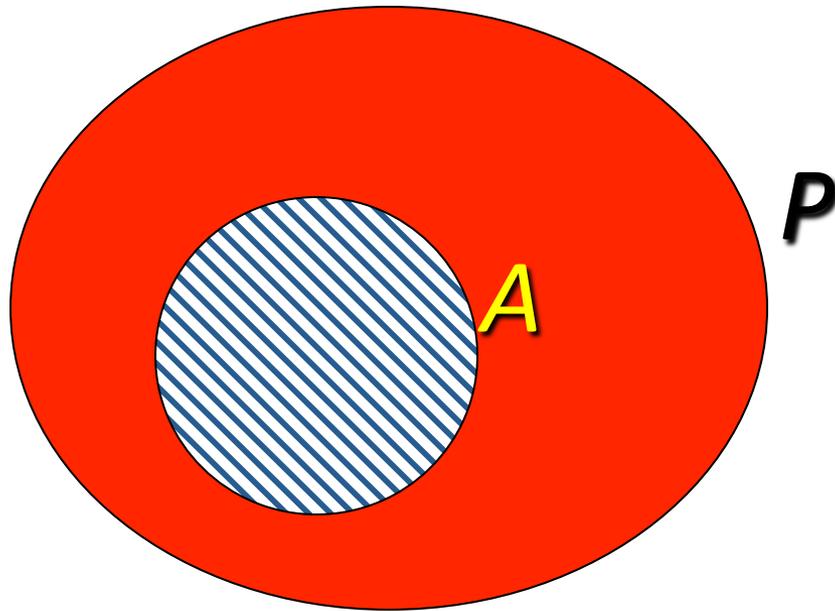
www.youtube.com/PenseMatematica



População: É o conjunto de elementos que se quer abranger em um estudo. É o universo o qual se pretende estudar. Quando o levantamento envolve todos os elementos da população, tem-se o recenseamento.

www.youtube.com/PenseMatematica

Amostra: É qualquer subconjunto da população.
É a parte da população que será observada.



P: população

A: amostra

www.youtube.com/PenseMatematica

Dados Brutos: É o conjunto de dados numéricos observados e que ainda não foram organizados.

Rol: É a representação dos dados brutos em ordem crescente ou decrescente.

Exemplo: a tabela abaixo representa notas de uma turma (dados brutos).

6	7	6	5	8
7	9	6	7	4
5	8	9	7	6
9	7	5	8	7

Colocando-se as notas em ordem crescente, constrói-se o rol:

{4,5,5,5,6,6,6,6,7,7,7,7,7,7,8,8,8,9,9,9}

||

{ $x_1, x_2, x_3, x_4, \dots, x_{18}, x_{19}, x_{20}$ }

www.youtube.com/PenseMatematica

a) Variável quantitativa: por exemplo, altura, peso, idade em anos e número de filhos, dizemos que elas são *quantitativas*, pois seus possíveis valores são números.

As variáveis quantitativas podem ser *discretas*, quando se trata de contagem (*números inteiros*), ou *contínuas*, quando se trata de medida (*números reais*).

b) Variável qualitativa representam características dos dados que estão sendo analisados, como, por exemplo: **cor do cabelo, cor da pele, feio ou bonito, alegre ou triste e assim por diante.**

Frequência Absoluta: É o número de vezes em que ocorre cada um dos valores observados.

Frequência Relativa:

É o quociente entre a frequência absoluta e o número total de valores observados.

$$F_R = \frac{F_{abs}}{N}$$

www.youtube.com/PenseMatematica

Frequência Acumulada: É a soma (ou o total corrente) de todas as frequências até o ponto presente no conjunto de dados.

Tabela de freqüências: É a apresentação ordenada dos valores observados. A tabela de freqüência para as notas do exemplo anterior é:

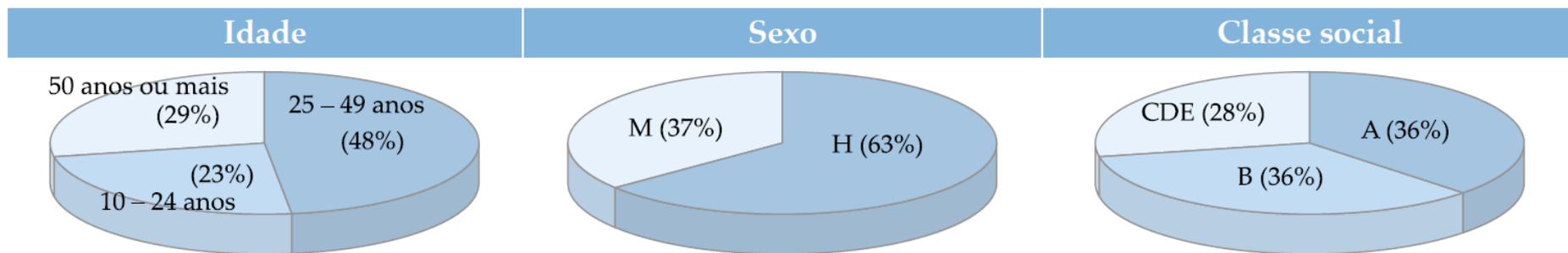
{4,5,5,5,6,6,6,6,7,7,7,7,7,7,8,8,8,9,9,9}

Nota	Freq. Absoluta	Freq. Acumulada	Freq. Relativa
4	1	1	5%
5	3	4	15%
6	4	8	20%
7	6	14	30%
8	3	17	15%
9	3	20	15%
Total	20		100%

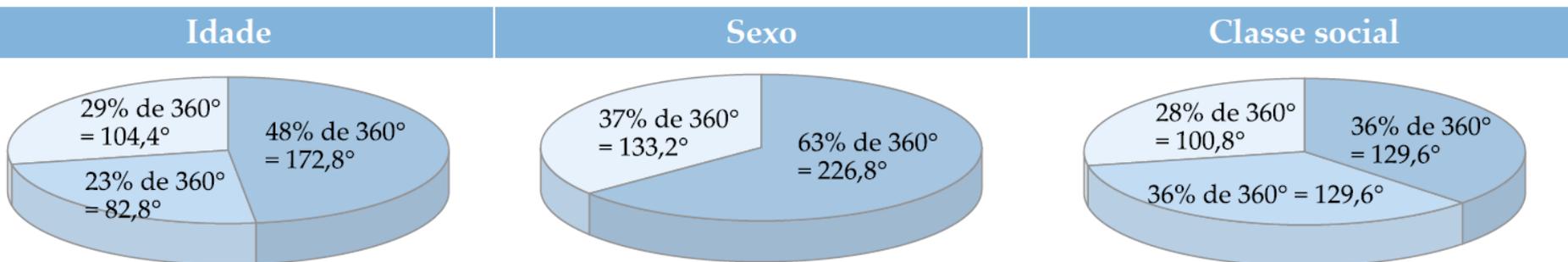
Representação Gráfica

O gráfico de setores, também conhecido como “o gráfico da pizza”, é um dos mais simples recursos gráficos, posto que consiste em um círculo cujos setores (fatias) somam 100%, sendo extremamente útil quando devemos visualizar diferentes freqüências entre algumas categorias.

Perfil dos leitores de um determinado jornal

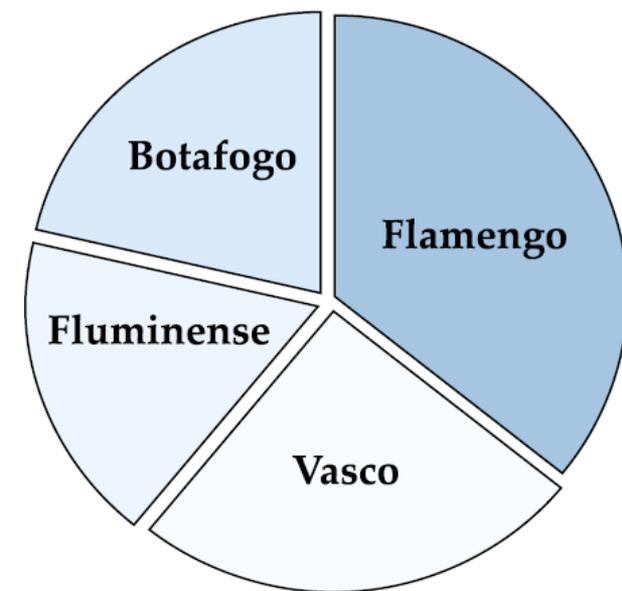


Obs.: É interessante observar que, a partir das porcentagens, que representam as freqüências relativas de cada dado, podemos chegar ao ângulo central de cada setor e vice-versa.



Ex: A tabela abaixo relaciona a preferência pelo time de futebol em relação a 560 pessoas entrevistadas, em que, para cada time, podemos utilizar a proporção entre a frequência relativa e o setor do gráfico.

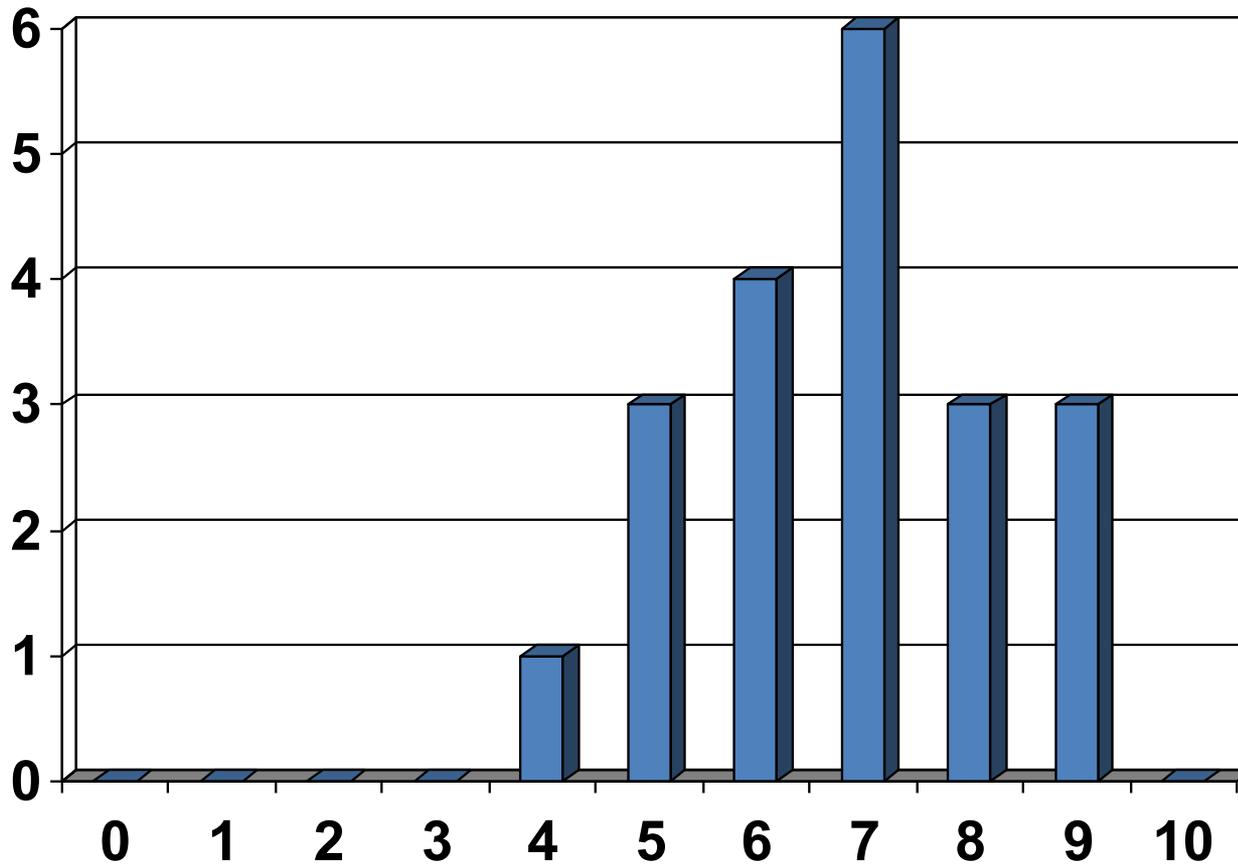
Time	Frequência absoluta	Frequência relativa $\left(\frac{f_{abs}}{N}\right)$	Porcentagem
Flamengo	200	$\frac{5}{14}$	35,71..%
Botafogo	120	$\frac{3}{14}$	21,42..%
Fluminense	100	$\frac{5}{28}$	17,85..%
Vasco	140	$\frac{1}{4}$	25%
Total	560	1	100%



Gráficos de barras (Histogramas)

www.youtube.com/PenseMatematica

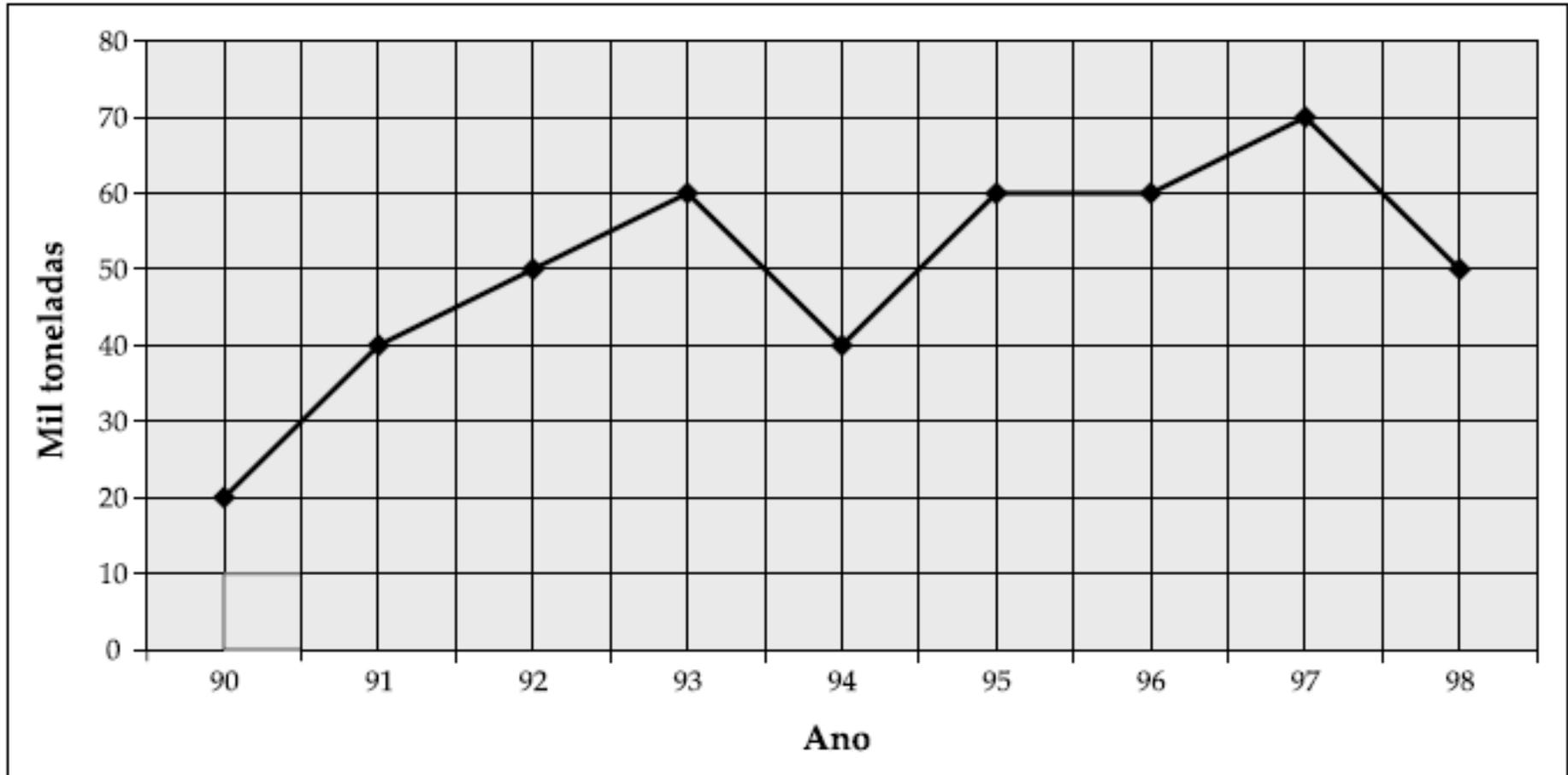
O gráfico de setores ilustra, de modo fácil e rápido, dados que podem ser subdivididos em algumas categorias. Comparativamente, o **gráfico de barras pode “acomodar” qualquer quantidade de categorias de qualquer nível de mensuração e, por isso, é muito mais usado em pesquisas.**



Gráficos de linha

www.youtube.com/PenseMatematica

O gráfico de linha, tal como o polígono de frequência, é útil para mostrar a evolução contínua da análise, em que é importante destacar que a declividade dos segmentos indica a intensidade do crescimento ou do decréscimo.



Medidas de Centralidade

www.youtube.com/PenseMatematica

A medida de centralidade é um número que está representando todo o conjunto de dados; nas pesquisas tal número é conhecido como medida de tendência central, que pode ser encontrado a partir da **média aritmética, da moda ou da mediana**, e o uso de cada uma delas é mais conveniente de acordo com o nível de mensuração, o aspecto ou forma da distribuição de dados e o objetivo da pesquisa.

Média Aritmética: É o quociente do somatório dos valores observados pelo número de valores.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \Rightarrow \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Média Aritmética Ponderada: Se, a cada valor observado x_i , associarmos uma frequência ou peso f_i , temos:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n} \Rightarrow \bar{x} = \frac{\sum_{i=1}^n x_i \cdot f_i}{\sum_{i=1}^n f_i}$$

www.youtube.com/PenseMatematica

Moda (Mo): É o valor observado de maior frequência.

www.youtube.com/PenseMatematica

Mediana (Md): É o valor que ocupa a posição central no rol. A mediana divide o rol em dois subconjuntos com igual número de elementos. Se o número de elementos do rol for ímpar, a mediana será o valor central do rol. Se o número de elementos do rol for par, a mediana será a média aritmética dos dois valores centrais do rol. No exemplo das notas, a mediana é a média aritmética dos dois valores centrais.

Medidas de Dispersão

www.youtube.com/PenseMatematica

Amplitude total

É uma medida de dispersão muito rápida e, ao mesmo tempo, **muito imprecisa**, pois consiste simplesmente em verificar a diferença entre o maior valor e o menor valor obtido na coleta de dados.

Pessoas	Peso (kg)
Pedro	30
Eduardo	15
Adriana	55
Susi	52
Renata	60
Anne Lise	53
Abelardo	75
Alexandre	20
Diogo	40

Na tabela ao lado, temos o peso das pessoas de um determinado grupo analisado e podemos verificar que a amplitude total foi de:

$$AT = 75 - 15 = 60$$

Desvio médio

O desvio deve ser empregado como a diferença do elemento analisado em relação à média, ou seja, o quanto o elemento se afasta da média da seqüência. Daí, é importante perceber que essa diferença deve ser necessariamente trabalhada em módulo, pois não tem sentido a distância negativa. E o desvio médio, então, passa a ser encontrado a partir da média aritmética de todos os desvios.

$$DM = \frac{\sum_{i=1}^n |\bar{x} - x_i|}{n}$$

www.youtube.com/PenseMatematica

Variância

A variância é uma medida de dispersão muito parecida com o desvio médio, a única diferença em relação a este é que, na variância, ao invés de trabalharmos em módulo as diferenças entre cada elemento e a média, tomamos os quadrados das diferenças. Isso se dá pelo fato de que, elevando cada diferença ao quadrado, continuamos trabalhando com números não negativos, como também pelo fato de que, em procedimentos estatísticos mais avançados, tal método facilita futuras manipulações algébricas.

$$Var = \frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n}$$

www.youtube.com/PenseMatematica

Desvio-padrão

www.youtube.com/PenseMatematica

Para entendermos o procedimento para o cálculo do desvio-padrão, é interessante percebermos que, no cálculo da variância, tal como vimos no tópico anterior, cometemos um “erro técnico” que será corrigido pelo desvio-padrão, ou seja, no momento em que elevamos ao quadrado as dispersões (diferenças) de cada elemento em relação à média, automaticamente alteramos a unidade de trabalho. Por exemplo: se estivermos trabalhando com a coleta das alturas, em metro, das pessoas de uma determinada comunidade, a unidade da variância encontrada será o m² (metro quadrado), que representa áreas. E é aí que entra o desvio-padrão, ou seja, extraindo a raiz quadrada da variância.

$$\textit{Desvio-padrão } \sigma = \sqrt{\textit{var}}$$

Quanto mais próximo de zero é o desvio padrão, mais homogênea (regular) é a amostra.

Candidatos que obtém menor desvio padrão são considerados mais regulares.