

# Conceito de Estatística

- **Estatística** – Técnicas destinadas ao estudo quantitativo de fenômenos coletivos, observáveis.
- **Unidade Estatística** – um fenômeno individual é uma unidade no conjunto que irá constituir o fenômeno coletivo
- **Dado Estatístico** - número que mede a intensidade de uma característica referente ao fenômeno coletivo em estudo.
- **Finalidade da Estatística**
  - 1. Fornecer métodos para coletar e organizar dados
  - 2. Fornecer métodos para inferir conclusões sobre um universo maior do que as observações efetivamente realizadas.

# Conceito de Estatística

- **Para inferir conclusões** – deve-se obter observações repetidas sobre um dado fenômeno, mantendo-se as mesmas condições.
- **Grau de Incerteza** - Como não é possível controlar todos os fatores que influem a observação de um fenômeno estatístico há sempre um grau de incerteza na avaliação dos resultados
- **Teoria da Probabilidade** – Estatística é, portanto, uma teoria sobre a incerteza. Por isso se baseia inteiramente na Teoria da Probabilidade.
- **Probabilidade Estatística** – são afirmações sobre a possibilidade ou a probabilidade de ocorrência de um fenômeno, desde que satisfeitas um conjunto de condições teóricas.
- **Fenômenos aleatórios** – são o objeto de estudo da estatística, e se referem a praticamente todos fenômenos na natureza.

---

# Fenômenos Aleatórios

---

- **Características Básicas**
  - Se repetem.
  - Apresentam variabilidade nas observações.
  - Não apresentam previsibilidade sobre sua variação futura..
- **Frequência de um Fenômeno Aleatório**
  - Quando as observações de um determinado fenômeno apresentam grande repetição, diz-se que existe *regularidade de frequência*.

# População e Amostra Estatística

- **População de uma Variável** – É o conjunto de todas as ocorrências ou repetições possíveis de um fenômeno aleatório. A população é o conjunto completo de dados de uma realidade.
- **Amostra** – É um subconjunto da população. Representa uma parte dos dados de uma população.
- **Levantamento de dados** – São as observações sobre uma amostra da população. Como é impossível levantar todos os dados de uma população, coletamos parte desta informação: amostra.
- **Objeto da Estatística** – Levantar dados amostrais para concluir (inferir ou generalizar) sobre as características da realidade mais ampla (população).
- **Indução Estatística** – processo pelo qual, se generaliza os dados obtidos na amostra para toda população, pelo cálculo das probabilidades.

# Amostragem

- **Seleção da Amostra** – as amostras devem se escolhidas de modo a poder aplicar a elas os cálculos de probabilidades.
- **Amostra Probabilística** – É aquela cujo processo de amostragem permite atribuir a cada elemento da amostra uma probabilidade semelhante à da população.
- **Amostragem Aleatória** – É aquela em que cada um dos elementos da população tem a mesma chance de ser selecionado no levantamento dos dados.

# Amostragem

- Esquemas de Amostragem probabilística:
  - **Amostragem casual simples com reposição:**  
Os elementos da população entram mais de uma vez na amostra
  - **Amostragem casual simples sem reposição:**  
Os elementos da população só podem entrar uma vez na amostra
  - **Amostragem sistemática:**  
seleção da amostra com base num critério: Um em cada dez.
  - **Amostragem por conglomerados:**  
a amostra é selecionada por sorteio da área de pesquisa
  - **Amostra em estágios múltiplos:**  
a amostra é selecionada por etapas: cidade/bairro/quadra
  - **Amostra estratificada:**  
A população é dividida em grupos e a amostra selecionada dentro do grupo.

# Experimento e Variável

- **Experimento** - É a observação sistemática de um fenômeno (evento aleatório) qualquer da população
- **Variável** – É o valor assumido pelo fenômeno em um experimento qualquer. A variável é, portanto, o valor que pode assumir o evento dentro de um conjunto de valores possíveis chamado *domínio* da variável.
- **Variável Qualitativa (Atributo)** – a qualidade assumida pelo fenômeno (evento aleatório) em um dado experimento.
- **Variável Quantitativa** – a medida da variação de um evento

# Experimento e Variável

- **Variável Contínua** – aquela que pode assumir qualquer valor numa escala de valores (teoricamente infinitos valores)
- **Variável Discreta** –aquela cujos valores possíveis são inteiros (contagem)
- **Variável Dependente** – assume certos valores em decorrência da variação de uma outra variável: em matemática se expressa por uma relação funcional (função)

$$y = f(x)$$

- $y$  = variável dependente
- $x$  = variável independente



# ESTADÍSTICA DESCRIPTIVA

# Amostra : Classificação e Caracterização

- Distribuição das Freqüências
- Medidas de Tendência Central
- Medidas de Variabilidade
- Medidas de Proporcionalidade ou Relativas

# Distribuição de Frequência

- **Frequência de uma variável** - é a frequência em que o evento ocorre. Em outras palavras, é a quantidade de vezes que a variável assume um certo valor.
- **Frequência para dados amostrais de variáveis contínuas:** É necessário dividir o conjunto de valores em intervalos de classe e para cada uma deles indicar a frequência dos valores observados.
- **Intervalo de Classe** – A cada intervalo de classe estão associados seus limites de classe (valores extremos) e o ponto médio.
- **Amplitude Total** – É a extensão de variação das variáveis: A diferença entre valor maior da última classe e o menor valor da primeira classe

# Distribuição de Frequência

- Ponto Médio de Intervalo de Classe = valor médio

$$\frac{\text{limite inferior} + \text{limite}}{2}$$

superior

2

- Processo de Elaboração do Histograma
  - Organizar os dados coletados em ordem crescente;
  - Determinar a amplitude total;
  - Dividir a amplitude total em um n° adequado de intervalos de preferência com a mesma amplitude;
  - N° mínimo de intervalos 5, número máximo 20;
  - Quando possível os pontos médios dos intervalos devem coincidir com os valores realmente observados

# Distribuição de Frequência

## Histograma

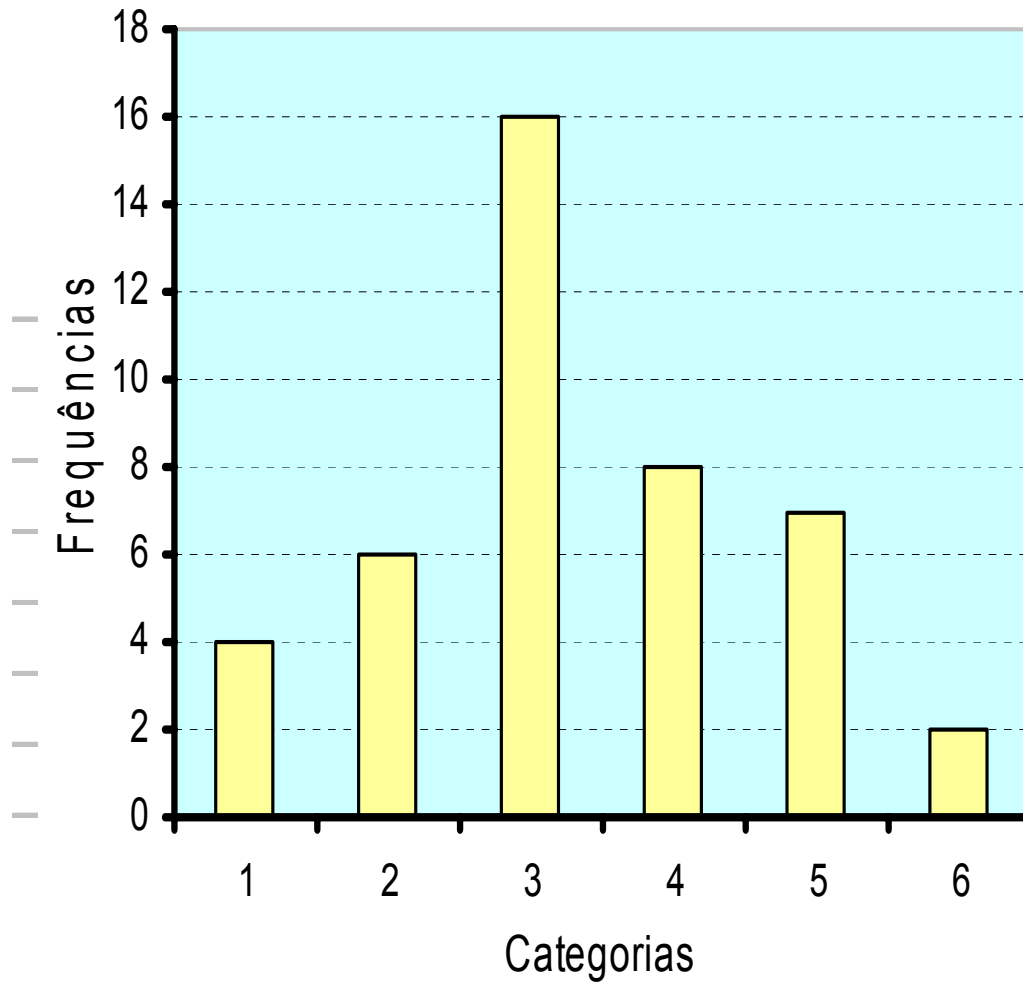
- **Frequência Absoluta** – Valor total das observações
- **Frequência Relativa** – Valor porcentual das observações
- **Frequência Acumulada** – Somatória das frequências de todos intervalos
- **Histograma:** Gráfico das distribuições das frequências de uma variável.
- **Gráfico de Barras (Histograma)** – Gráfico de retângulos, diagrama de colunas; gráfico de áreas
- **Histograma** – As frequências dos fenômenos são proporcionais à superfície de cada retângulo que as representam. Para intervalos de mesma amplitude as frequências serão proporcionais às alturas

# Distribuição de Frequência

## Histograma

- **Distribuições Simétricas e Assimétricas** - Os histogramas podem apresentar distribuição simétricas ou assimétricas
- **Polígono de Frequências** - Unindo os valores médios dos intervalos de classe, transforma-se o histograma num polígono de frequências. Pode então compara-la com uma curva teórica (Normal).

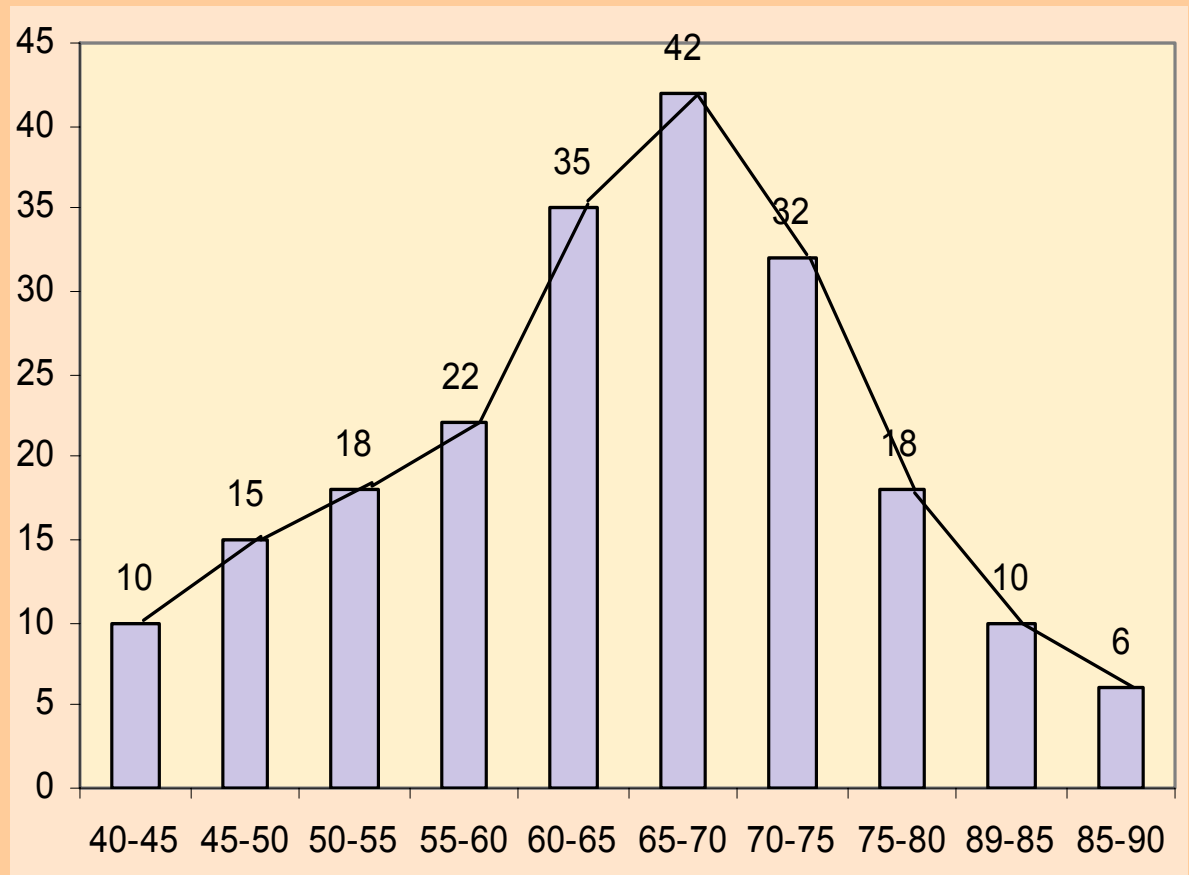
# GRÁFICO DE FREQUÊNCIAS: Histograma



Variáveis	Frequência
1	4
2	6
3	16
4	8
5	7
6	2

# HISTOGRAMA E POLÍGONO DE FREQUÊNCIA

<b>Pesos (<math>x_1</math>)</b>	<b>Nº alunos (<math>f_1</math>)</b>
40-45	10
45-50	15
50-55	18
55-60	22
60-65	35
65-70	42
70-75	32
75-80	18
85-90	10
85-90	6
<b>Total</b>	<b>208</b>

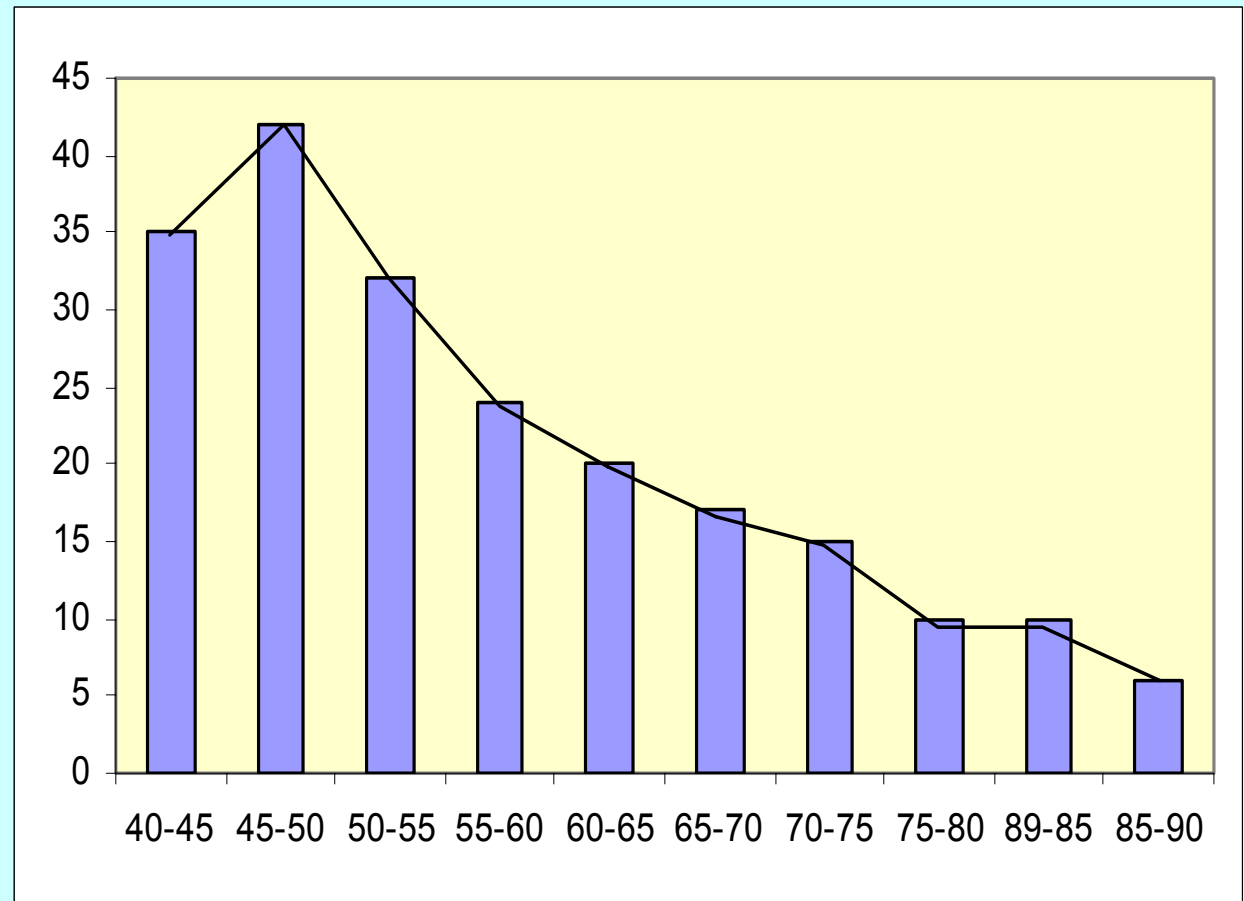




# HISTOGRAMA E POLÍGONO DE FREQUÊNCIA

## Assimétrico à esquerda

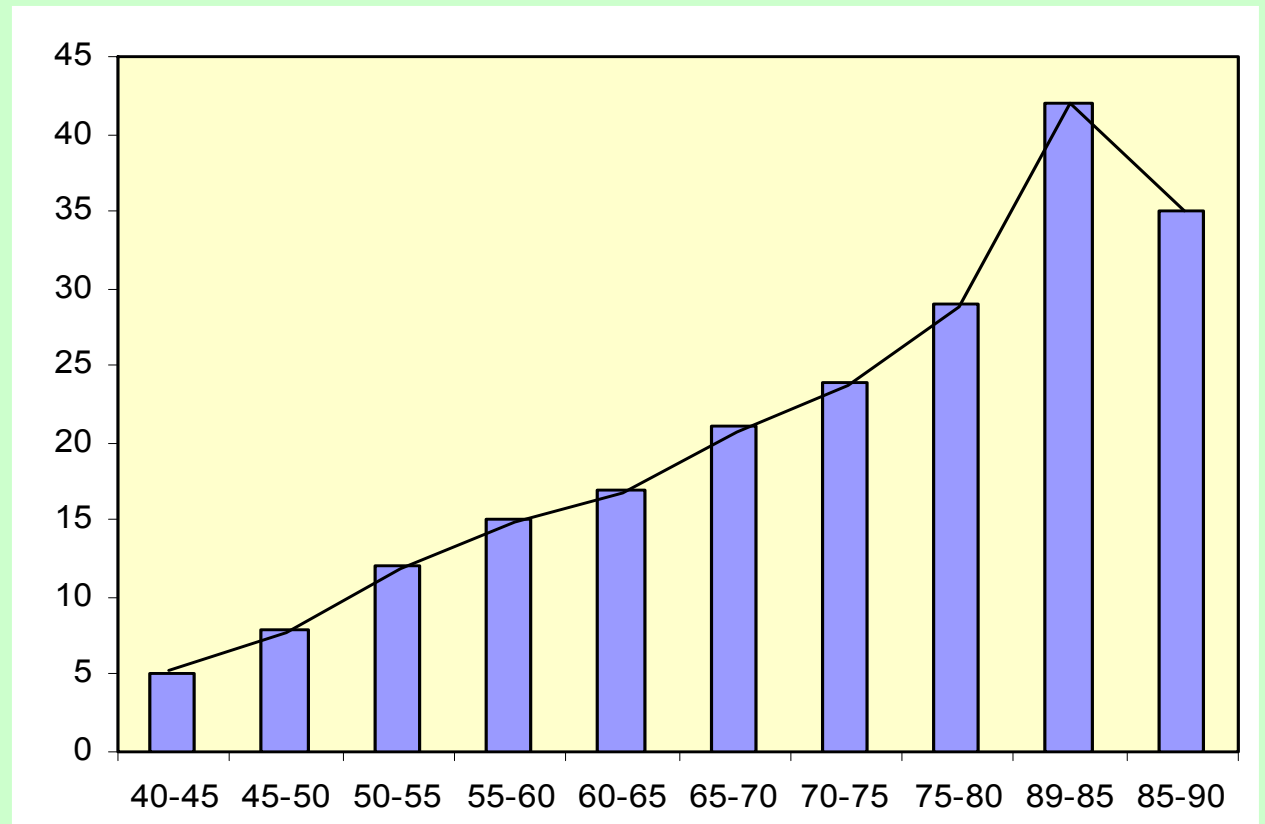
Pesos ( $x_1$ )	Nº alunos ( $f_1$ )
40-45	35
45-50	42
50-55	32
55-60	24
60-65	20
65-70	17
70-75	15
75-80	10
80-85	10
85-90	6
Total	208



# HISTOGRAMA E POLÍGONO DE FREQUÊNCIA

## Assimétrico à direita

Pesos ( $x_1$ )	Nº alunos ( $f_1$ )
40-45	5
45-50	8
50-55	12
55-60	15
60-65	17
65-70	21
70-75	24
75-80	29
89-85	42
85-90	35
Total	173



# Medidas de Tendência Central

- **Valores Centrais ou Médias de uma Amostra** – Valores que indicam posição de centralidade, ou o ponto central da distribuição.
- **Média Aritmética Simples** – Quociente da soma dos valores observados, pelo número total de valores.

$$\alpha = \frac{\sum x_i}{n} \quad i = 1, \dots, n$$

# Medidas de Tendência Central

- **Média Aritmética Ponderada** - Quando há valores que se repetem mais que outros.
- $$\alpha = \frac{\sum x_i \cdot f_i}{\sum f_i}$$
- Ex: 
$$\alpha = \frac{48.2 + 51.3 + 55.5 + 58.4 + 60.1}{15} = 54,4$$
- **Utilização:** média de cálculo mais fácil. Valor médio significativo por incluir todos os valores. Usada em estatística para o cálculo do desvio padrão.

# Medidas de Tendência Central

- **Mediana** – Medida de posição central. A mediana é o valor que ocupa a posição central (meio) da distribuição.
- Série de valores com **nº ímpar** de termos
  - $Mediana = n + 1 / 2$
  - Nº de termos 7 (mediana é o 4º termo)
  - $Md = 7 + 1 = 8 / 2 = 4$
  - Ex: 5, 7, 8, 11, 12, 13, 14 (7 termos) →  $Md = 11$
- Série de valores com **nº par** de termos
  - $Mediana = n + 1 / 2$  e
  - $Mediana = n / 2$
  - Nº de termos 8 (mediana entre o 4º e 5º termo)
  - $Md = 8 / 2 = 4$
  - $Md = 8 / 2 + 1 = 5$
  - Ex: 5, 7, 8, 11, 12; 13, 14, 15; (8 termos) →  $Md = 11 + 12 / 2 = 11,5$
- **Utilização**: usada quando a distribuição apresenta resultados extremos muito discrepantes. A mediana não sofre a influência de valores extremos

# Medidas de Tendência Central

- **Moda** – Valor dominante de uma distribuição. Aquele que numa série de valores se apresenta com a maior frequência. Um conjunto de valores pode apresentar mais de uma moda: *plurimodal*.

Ex: 48, 49, 50, 50 50, 55, 58, 59, 60 → **M = 50**

Ex: 4, 5, 6, 4, 5, 7, 4, 8, 5, 10 → **M = 4 e 5**  
**(plurimodal)**

# Mediadas de Variabilidade

- Índices que indicam o grau de concentração ou dispersão de uma distribuição em torno da média.
- Principais índices de variabilidade:
  - **Amplitude total**
  - **Desvio médio**
  - **Variância**
  - **Desvio padrão**
- **Amplitude Total** (Intervalo Total) - É a diferença entre o maior e o menor valor de uma série.
- Ex: 48, 49, 50, 50 50, 55, 58, 59, 60 →  $A = 60 - 48 = 12$

# Medidas de Variabilidade

- **Desvio Médio** – Média aritmética dos afastamentos (ou desvios), tomados em valor absoluto, entre cada valor e a média aritmética.

- $$DM = \frac{\sum |d_i \cdot f_i|}{\sum f_i}$$
 sendo  $d_i = |x_i - \alpha|$   
sendo  $\alpha =$  média aritmética

Ex: 
$$dm = \frac{(48 - 54,4).2 + (51-54,4).3 + (55-54,4).5 + (58-54,4).4 + (60-54,4).1}{15}$$

$$dm = \frac{12,8 + 10,2 + 3,0 + 14,4 + 5,6}{15}$$

$$dm = 3,07$$

Utilização: Indica em média o quanto os valores se afastam da média considerada no entendimento da curva de Gaus



# Medidas de Variabilidade

- **Variância** – Considerando-se o conjunto de dados, cada dado isolado pode ter um desvio (dispersão) em relação à média, isto é; uma diferença entre o valor individual e a média do conjunto de dados. Para se avaliar o grau de dispersão do conjunto de dados utiliza-se a variância que é a soma dos quadrados dos desvios dividido pelo tamanho da amostra, menos 1.

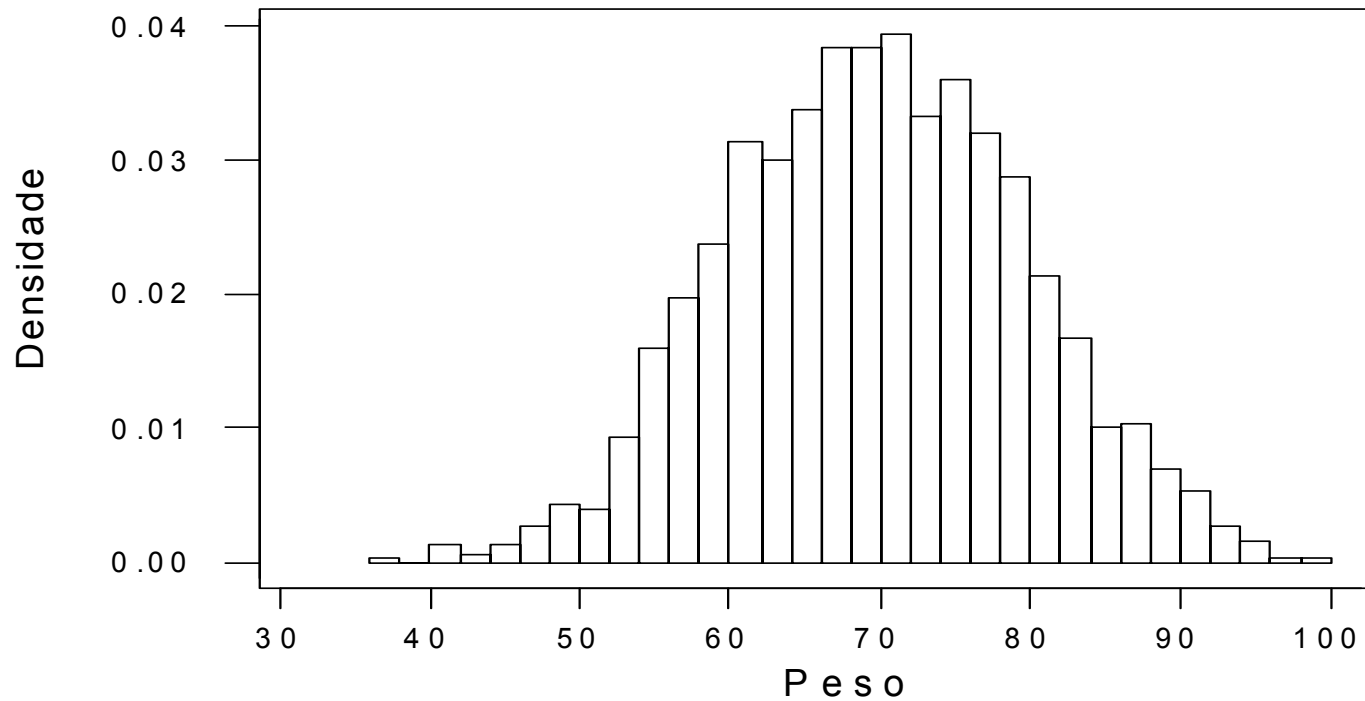
- $$s^2 = \frac{\sum (x_i - \alpha)^2 \cdot f_i}{\sum f_i - 1}$$

- Exemplo: 
$$s^2 = \frac{81,92 + 34,68 + 1,80 + 51,84 + 31,36}{14}$$
$$s^2 = 14,4$$

# Medidas de Variabilidade

- **Desvio padrão** – afastamento quadrático médio ou afastamento padrão. É a raiz quadrada da variância.
- Desvio padrão dos dados isolados ponderados com frequências distintas:
- $$s^2 = \frac{\sum (x_i - \alpha)^2 \cdot f_i}{\sum f_i - 1}$$
- Exemplo: 
$$s^2 = \frac{81,92 + 34,68 + 1,80 + 51,84 + 31,36}{14}$$
$$s = \sqrt{14,4} = 3,79$$
- **Utilização**: é a medida mais usada com medida de variabilidade, principalmente quando a distribuição for normal

# O histograma de Densidades

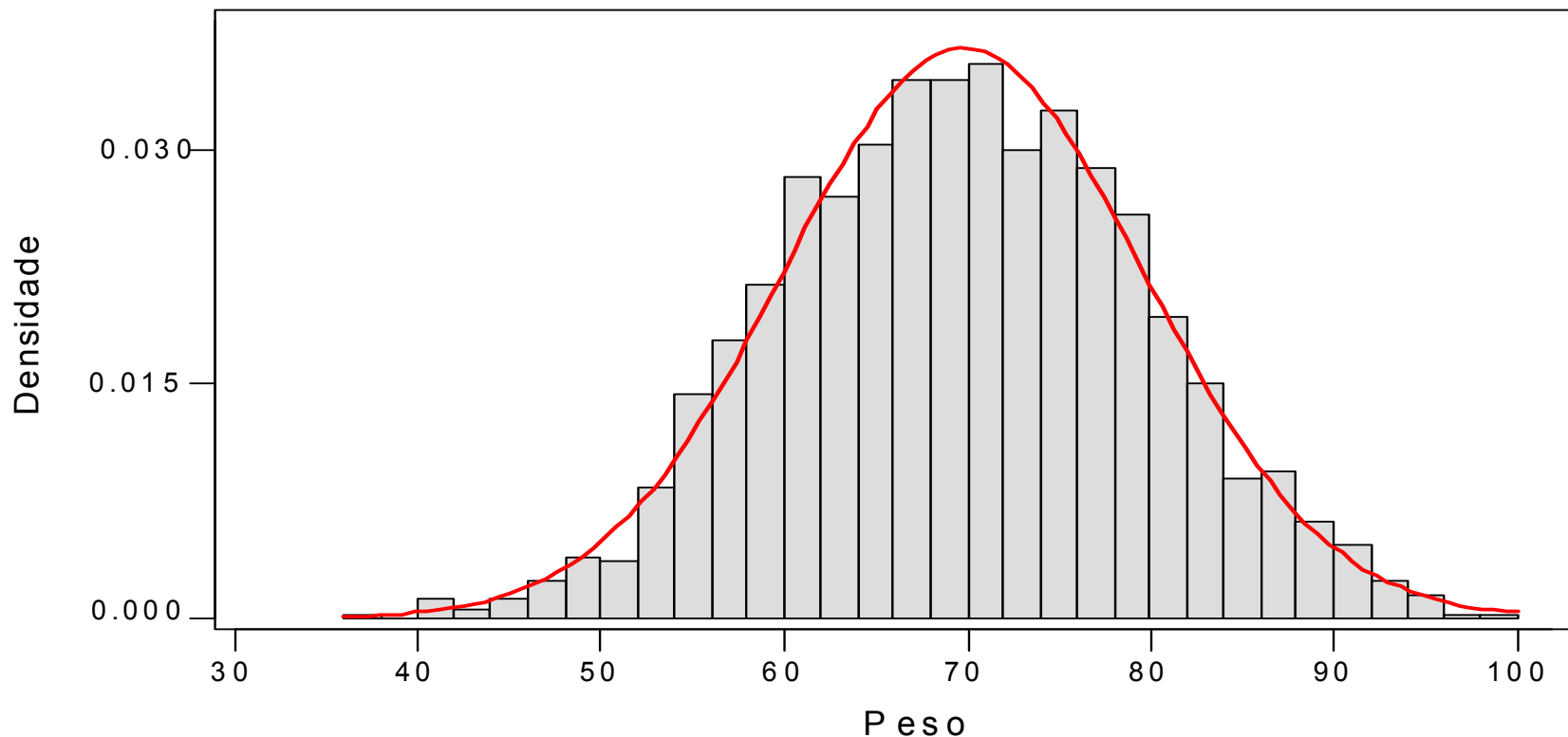


**Peso, em kg, de 1000 pessoas adultas selecionadas ao acaso em uma população**

# Distribuição dos valores da variável aleatória $X$ .

Distribuição de probabilidades de  $X$  ?

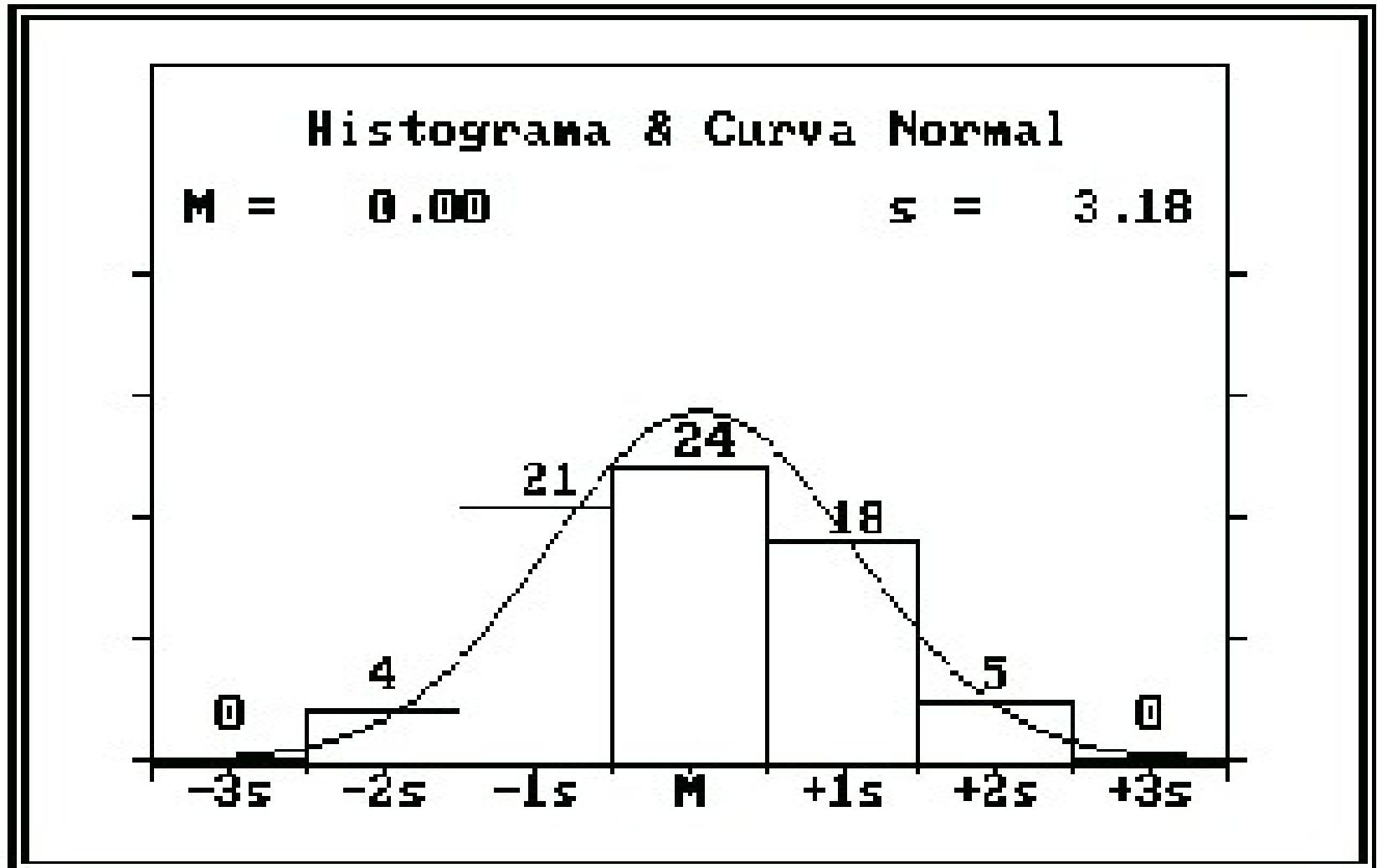
A curva contínua da figura denomina-se ***Curva Normal***



**$X$ : peso, em kg, de uma pessoa adulta escolhida ao acaso da população.**

# Curva de Gauss

## Distribuição Normal



# Curva de Gauss

## Distribuição Normal

$$-1s \rightarrow 1s$$

68,27 % dos casos estão incluídos entre  $M-1s$  e  $M+1s$

$$-2s \rightarrow 2s$$

95,45 % dos casos estão incluídos entre  $M-2s$  e  $M+2s$

$$-3s \rightarrow 3s$$

99,73 % dos casos estão incluídos entre  $M-3s$  e  $M+3s$

# Curva de Gauss

## Distribuição Normal

