

Statistics for Business and Economics

Newbold



Capítulo 1 – Aula 02

Descrevendo dados: Graficamente



Objetivos

After completing this chapter, you should be able to:

- Explicar como as decisões geralmente são baseadas em informações incompletas
- Explicar as principais definições:
 - População vs. Amostra
 - Parâmetro vs. Estatística
 - Estatística Descritiva vs. Inferencial
- Descrever amostragem aleatória
- Explicar a diferença entre estatística descritiva e inferencial
- Identificar tipos de dados e níveis de medição



Objetivos

(cont.)

Depois de concluir este capítulo, você deverá ser capaz de:

- Criar e interpretar gráficos para descrever variáveis categóricas: distribuição de frequência, gráfico de barras, gráfico de pizza
- Criar um gráfico de linhas para descrever dados de séries temporais
- Criar e interpretar gráficos para descrever variáveis numéricas: distribuição de frequência, histograma
- Construir e interpretar gráficos para descrever relações entre variáveis: Gráfico de dispersão, tabela cruzada
- Descrever formas apropriadas e inadequadas de exibir dados graficamente

Lidando com a Incerteza

Decisões são baseadas em informações incompletas

Considerar:

O mercado de trabalho estará forte quando eu me formar?

O preço das ações do Apple será maior em seis meses do que é agora?

As taxas de juros permanecerão baixas pelo resto do ano se o déficit orçamentário federal for tão alto quanto o previsto?



Lidando com a Incerteza

(cont.)

Números e dados são usados para auxiliar as tomadas de decisões

A estatística é uma ferramenta para ajudar a processar, resumir, analisar e interpretar dados

Definições Chave

- Uma **população** é uma coleção de todos os itens de interesse ou sob investigação e N representa o tamanho da população
- Uma **amostra** é um subconjunto observado da população e n representa o tamanho da amostra
- Um **parâmetro** é uma característica específica de uma população
- Uma **estatística** é uma característica específica de uma amostra



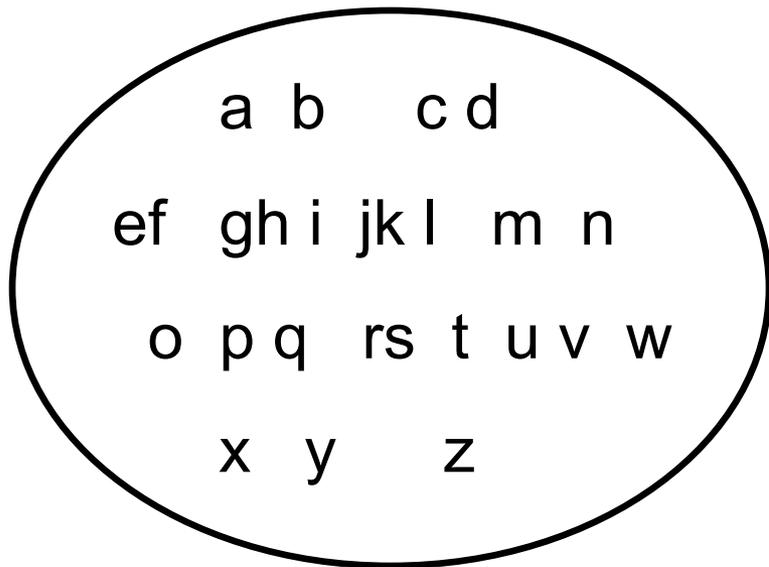
Exemplos de Variáveis

- Altura dos moradores de uma cidade do sexo masculino
- Eleitores de um certo partido político numa região
- Número de defeitos de fabricação numa linha de produção medido mensalmente.



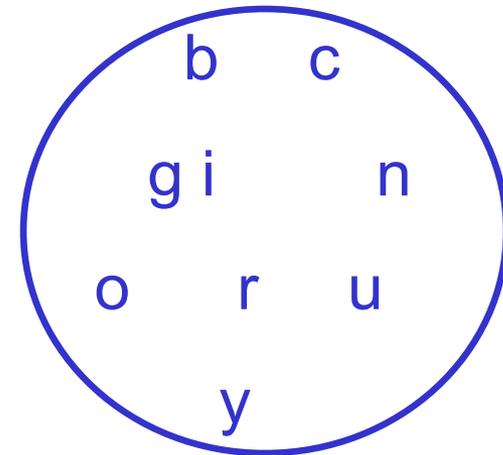
População vs. Amostra

População



Os valores calculados usando dados populacionais são chamados de **parâmetros**

Amostra



Os valores calculados a partir de dados de amostra são chamados de **estatísticas**



Exemplos de Populações

- Nomes de todos os eleitores registrados Brasil
- Notícias de todas as famílias que vivem em Salvador
- Retornos anuais de todas as ações negociadas na Bolsa de Valores de São Paulo
- Médias de notas de todos os alunos da sua universidade de São Paulo



Amostragem aleatória

Amostragem aleatória simples é um procedimento em que

- Cada membro da população é escolhido por acaso,
- Cada membro da população provavelmente será escolhido,
- Toda amostra possível de n objetos é provável que seja escolhida

A amostra resultante é chamada de **amostra aleatória**



Estatística Descritiva e Inferencial

Dois ramos das estatísticas:

- Estatísticas Descritivas

Procedimentos gráficos e numéricos para
sumariar e processar dados

- Estatística inferencial

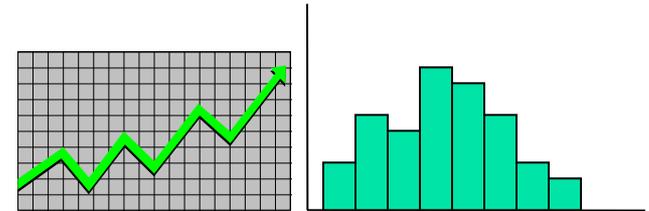
Usa dados para fazer previsões, previsões e
estimativas para auxiliar a tomada de
decisão

Estatísticas Descritivas

- Coletar dados
- e.g., Survey



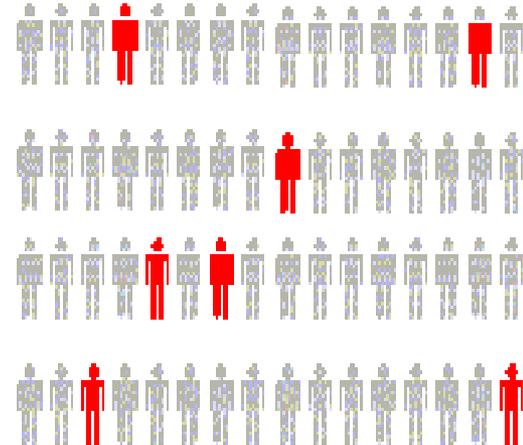
- Apresentar dados
 - e.g., Tabelas e gráficos



- Sumariar dados
 - e.g., Média amostral = $\frac{\sum X_i}{n}$

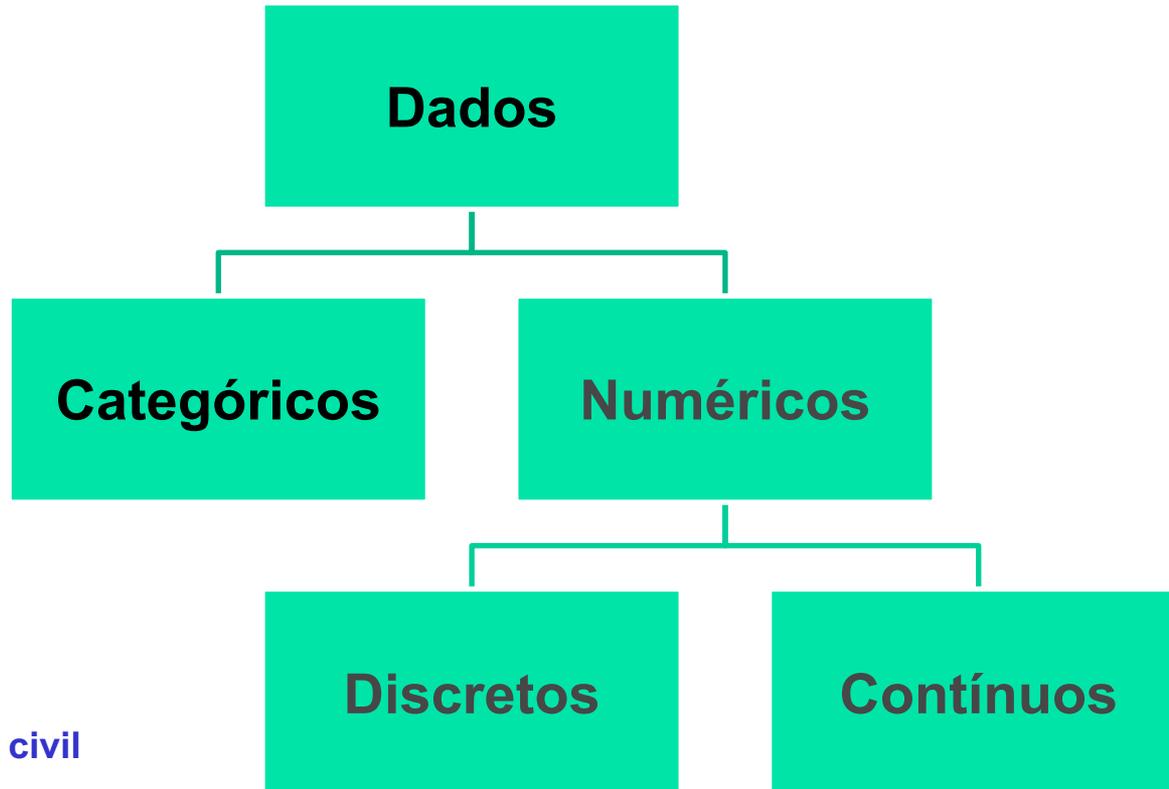
Inferência Estatística

- Estimativa
- Por exemplo, estimar o peso médio da população usando o peso médio da amostra
- Testando hipóteses
- Por exemplo, teste a alegação de que o peso médio da população é de 140 libras



Inferência é o processo de tirar conclusões ou tomar decisões sobre uma população com base em resultados de sua amostra

Tipos de Dados



Exemplos:

- Estado civil
- Você está registrado para votar?
- Cor dos olhos

(Categorias ou grupos definidos)

Exemplos :

- Número de crianças
- Defeitos por hora

(Itens contados)

Exemplos :

- Peso
- Voltagem

(Características medidas)

Níveis de medição

Diferenças entre as medições, existe zero verdadeiro

Dados de Proporção

Dados quantitativos

Diferenças entre as medidas, mas nenhum zero verdadeiro

Intervalo de Dados

Categorias ordenadas (classificações, ordem ou escala)

Dados Ordenados

Dados qualitativos

Categorias (sem ordenação ou direção)

Dados Nominais

Apresentação gráfica de dados

- Os dados em forma bruta geralmente não são fáceis de usar para a tomada de decisão
- Algum tipo de organização é necessário
- Tabelas
- Gráficos
- O tipo de gráfico a ser usado depende da variável que está sendo resumida

Apresentação gráfica de dados

(cont)

- Techniques reviewed in this chapter:

Variáveis Categóricas

- Distribuição de frequência
- Gráfico de barras
- Gráfico de pizza
- Diagrama de Pareto

Variáveis Numéricas

- Gráfico de linha
- Distribuição de frequência
- Histograma e Ogivo
- Exibição de caule e folha
- Plotagem de dispersão

Tabelas e gráficos para variáveis categóricas

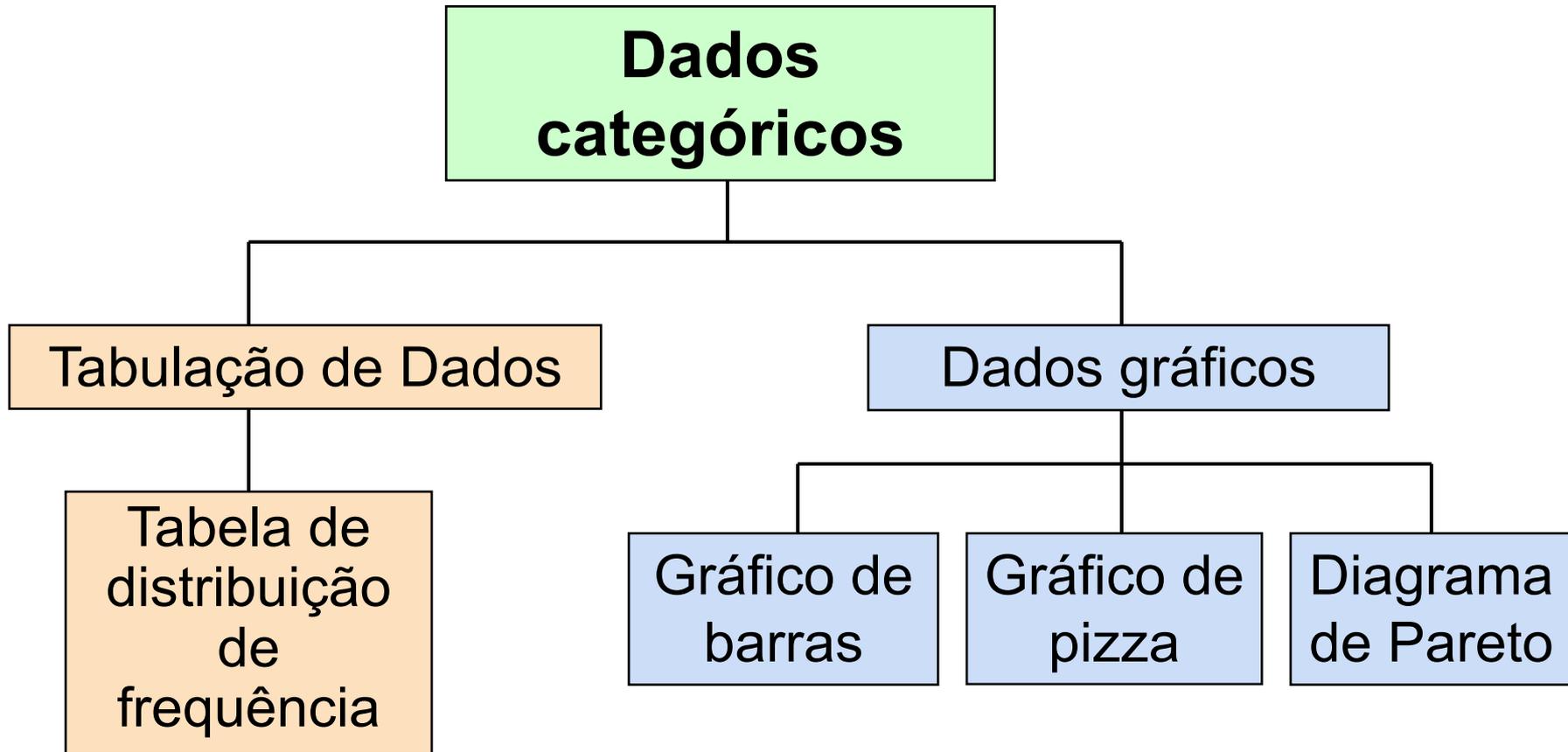




Tabela de distribuição de frequência

Resumindo dados por categoria

Exemplo: pacientes hospitalizados por unidade

Unidade hospitalar	Número de pacientes
Cuidados cardíacos	1,052
Emergência	2,245
Cuidados Intensivos.	340
Maternidade	552
Cirurgia	4,630

(Variáveis são categóricas)

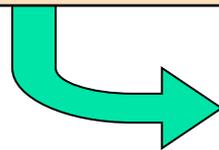
Gráficos de barras e pizza



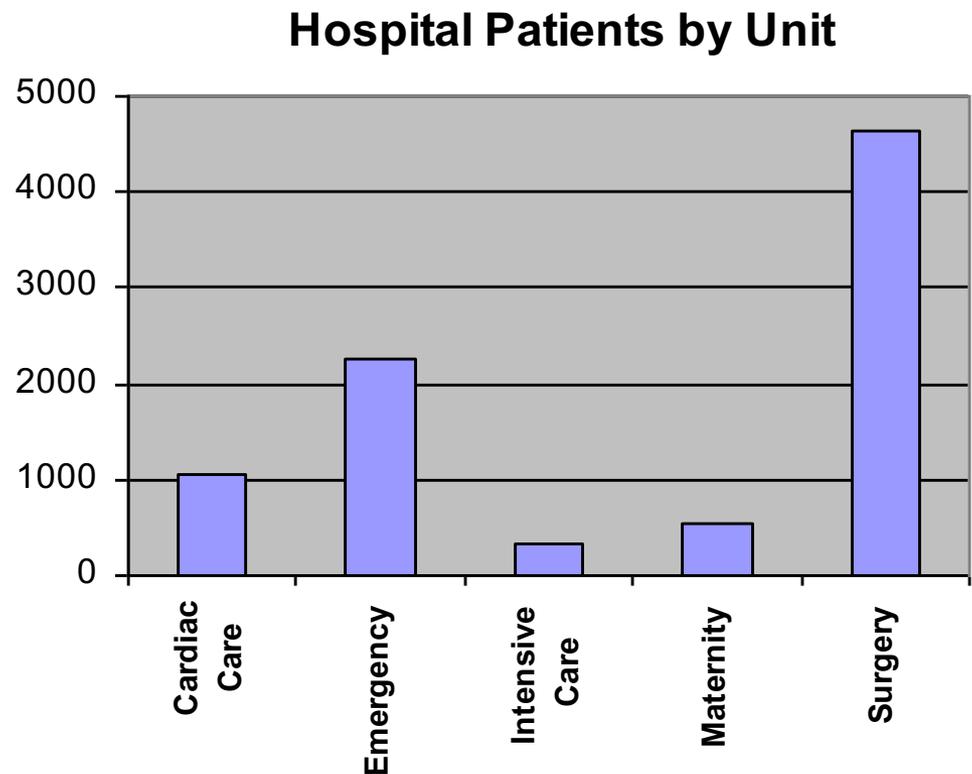
- Os gráficos de barras e os gráficos de pizza são frequentemente usados para dados qualitativos (categoria)
- A altura da barra ou tamanho da fatia de torta mostra a frequência ou porcentagem para cada categoria

Exemplo de gráfico de barras

Unidade hospitalar	Número de pacientes
Cuidados cardíacos	1,052
Emergência	2,245
Cuidados Intensivos.	340
Maternidade	552
Cirurgia	4,630



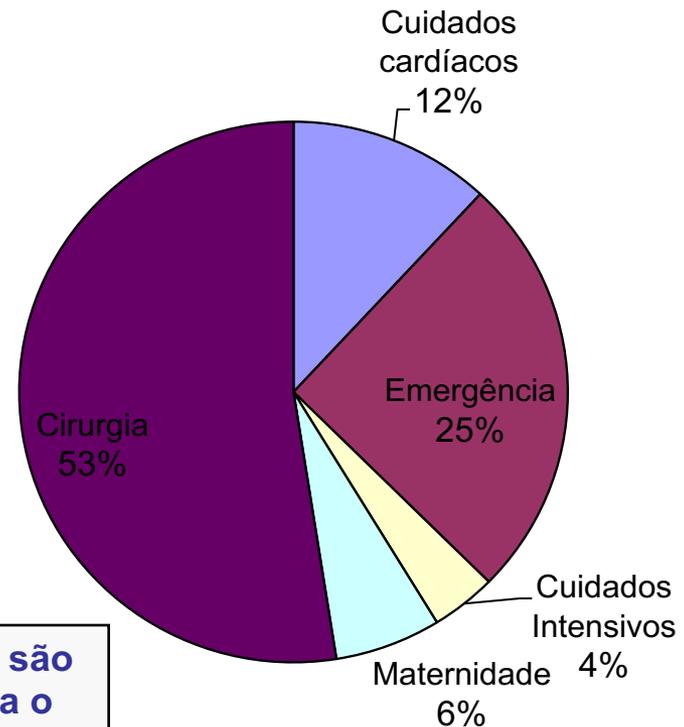
Number of patients per year



Exemplo de gráfico de Pizza

Unidade Hospitalar	Núm. de Pacientes	% do Total
Cuidados cardíacos	1,052	11.93
Emergência	2,245	25.46
Cuidados Intensivos	340	3.86
Maternidade	552	6.26
Cirurgia	4,630	52.50

Pacientes de hospital por unidade



(As porcentagens são arredondadas para o porcentagem mais próximo)



Diagrama de Pareto

- Usado para retratar dados categóricos
- Um gráfico de barras, em que as categorias são mostradas em ordem decrescente de frequência
- Um polígono cumulativo é frequentemente mostrado no mesmo gráfico
- Usado para separar os "poucos vitais" dos "muitos triviais"



Exemplo de Diagrama de Pareto

Exemplo: 400 itens defeituosos são examinados

De acordo com a causa do defeito:

Fonte de Erro de fabricação	Número de defeitos
Baixo alinhamento	34
Falha na pintura	223
Solda ruim	25
Partes faltantes	78
Case rachada	19
Curto circuito elétrico	21
Total	400



Exemplo de Diagrama de Pareto

(cont.)

Etapa 1: classificar por causa de defeito, em ordem decrescente

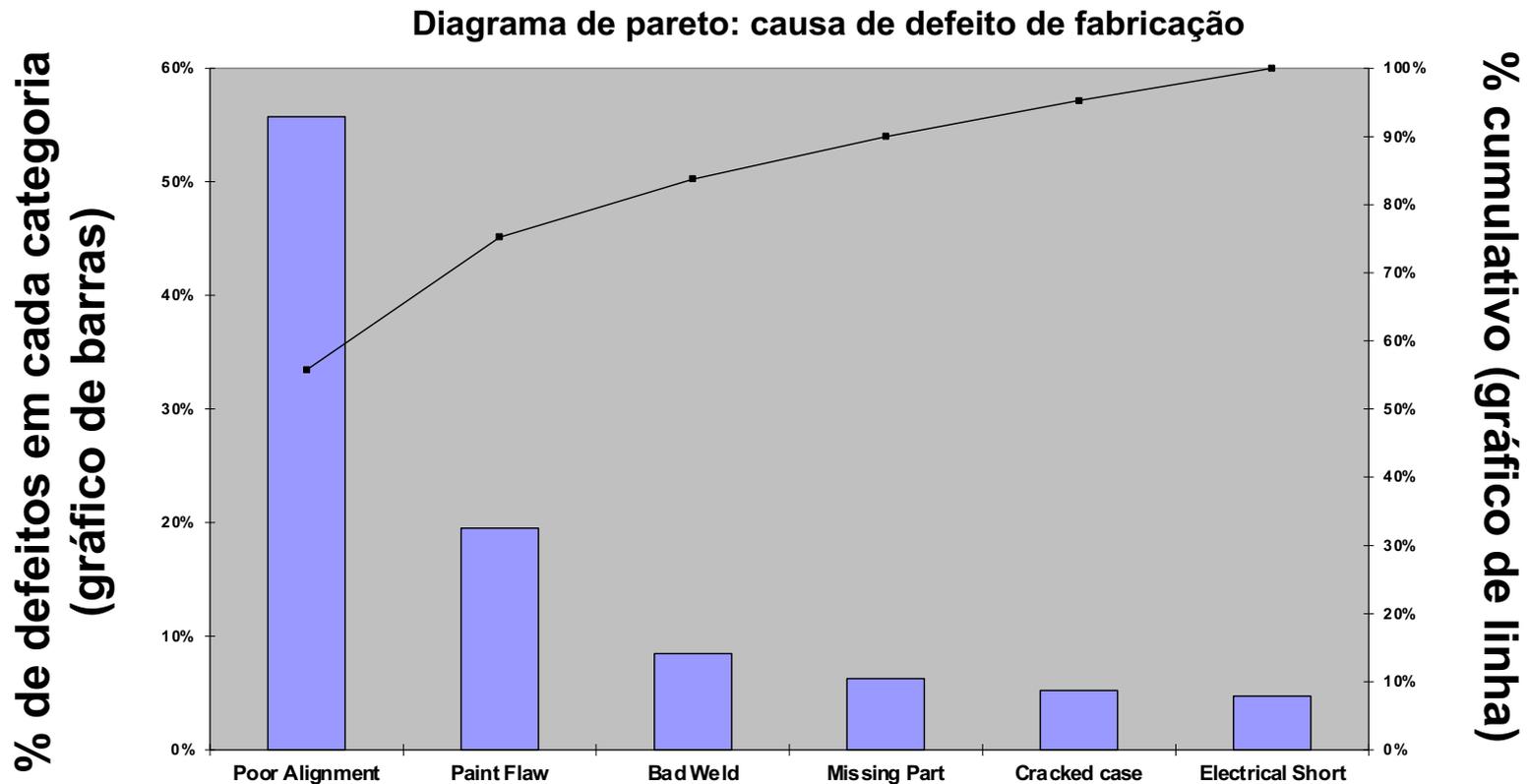
Etapa 2: determine % em cada categoria

Fonte de Erro de fabricação	Número de defeitos	% do total de defeitos
Baixo alinhamento	223	55.75
Falha na pintura	78	19.50
Solda ruim	34	8.50
Partes faltantes	25	6.25
Case rachada	21	5.25
Curto circuito elétrico	19	4.75
Total	400	100%

Exemplo de Diagrama de Pareto

(cont)

Etapa 3: Mostrar resultados graficamente



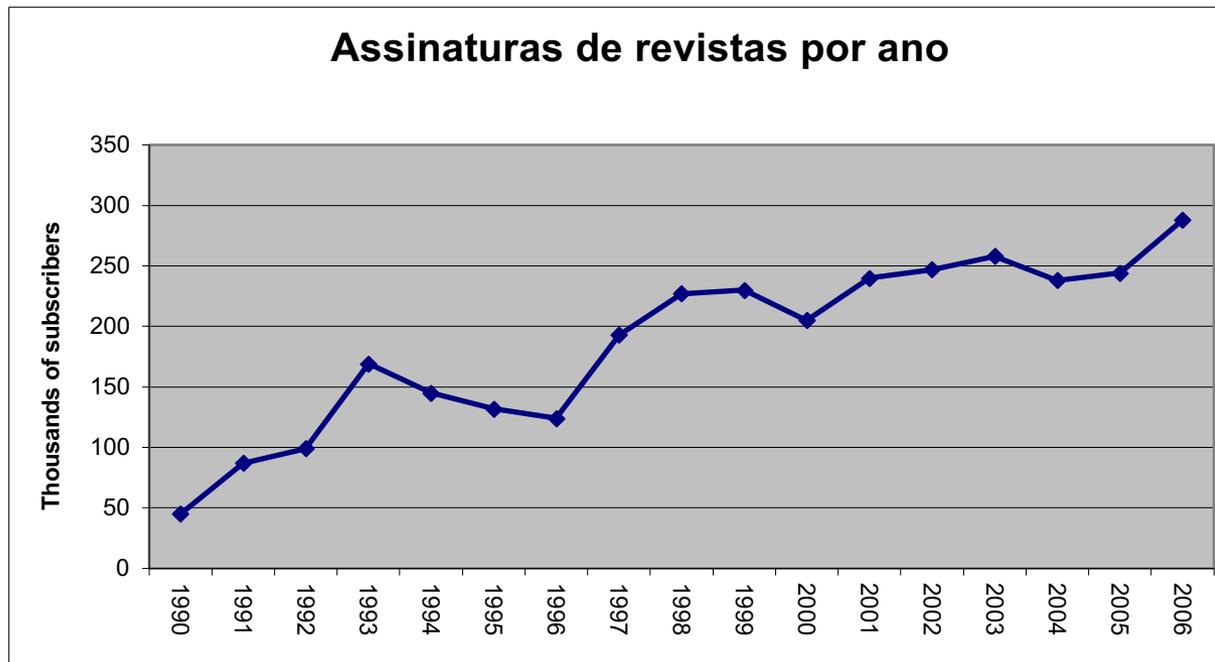
Gráficos para dados de séries temporais

1.4

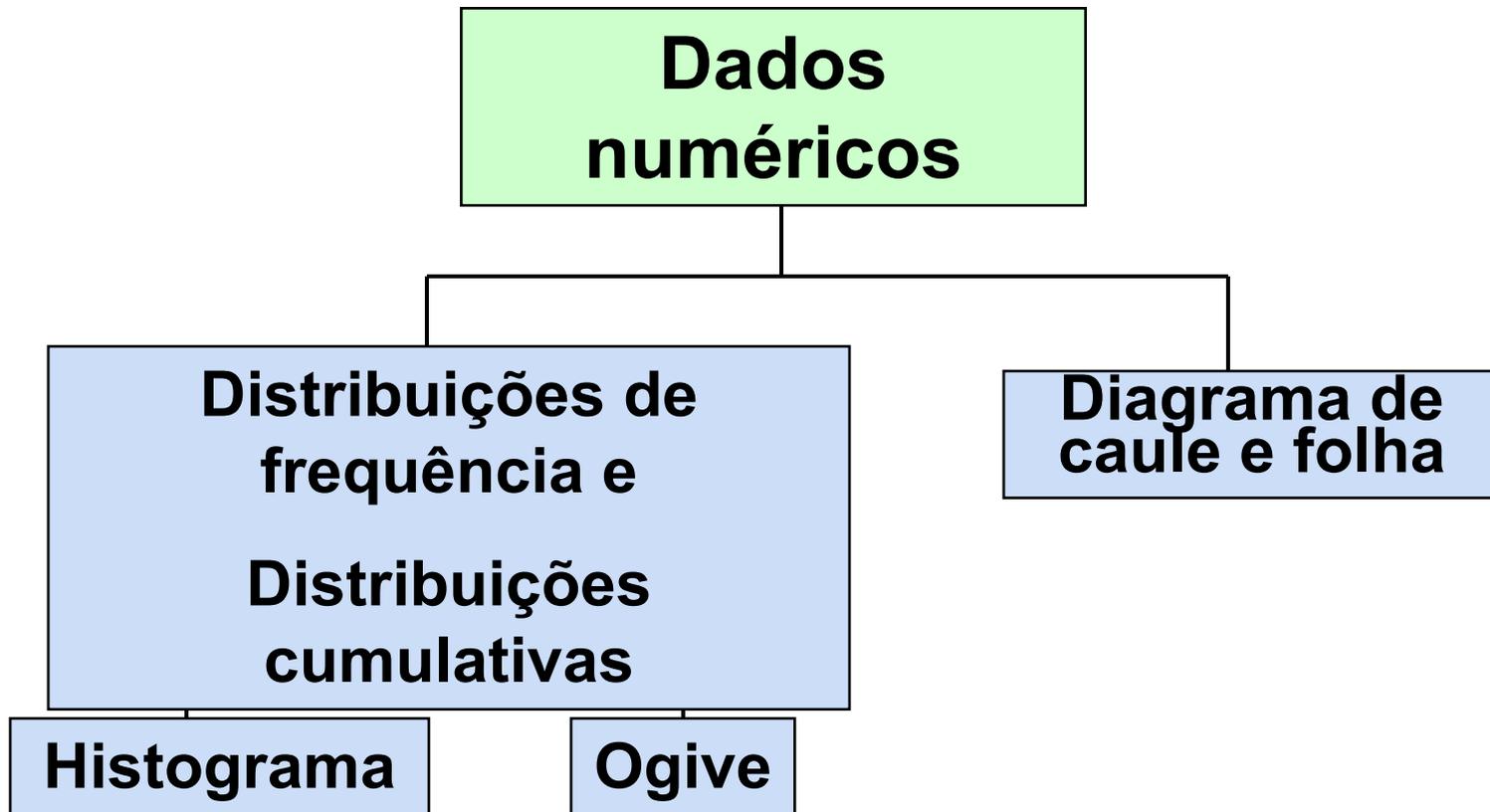
- Um gráfico de linha (gráfico de séries temporais) é usado para mostrar os valores de uma variável ao longo do tempo
- O tempo é medido no eixo horizontal
- A variável de interesse é medida no eixo vertical



Exemplo de gráfico de linha



Gráficos para descrever variáveis numéricas





Distribuição de frequência

O que é uma distribuição de frequência?

Uma distribuição de frequência é uma lista ou uma tabela...

contendo grupos de classe (categorias ou faixas nas quais os dados são classificados) ...

e as frequências correspondentes com as quais os dados se enquadram em cada classe ou categoria



Por que usar distribuições de frequência?

- Uma distribuição de frequência é uma maneira de resumir dados
- A distribuição condensa os dados brutos em uma forma mais útil ...
- e permite uma interpretação visual rápida dos dados



Intervalos de classe e limites de classe

- Cada agrupamento de classe tem a mesma largura
- Determinar a largura de cada intervalo por
- $w = \text{largura do intervalo} = \frac{\text{maior número} - \text{menor número}}{\text{número de intervalos desejados}}$
- Use pelo menos 5, mas não mais que 15-20 intervalos
- Os intervalos nunca se sobrepõem
- Arredondar a largura do intervalo para obter terminais de intervalo desejáveis



Exemplo de distribuição de frequência

Exemplo: um fabricante de isolamento seleciona aleatoriamente 20 dias de inverno e registra a alta temperatura diária

**24, 35, 17, 21, 24, 37, 26, 46, 58, 30,
32, 13, 12, 38, 41, 43, 44, 27, 53, 27**

Exemplo de distribuição de frequência

(cont)

- Classificar dados brutos em ordem crescente:
- 12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
- Localizar intervalo: $58 - 12 = 46$
- Selecionar Número de classes: 5 (geralmente entre 5 e 15)
- Largura do intervalo de computação: 10 ($46/5$ e depois arredondado)
- Determinar os limites do intervalo: 10, mas menos de 20; 20, mas menos de 30 ;. . . ; 60, mas menos de 70
- Conte observações e atribua a classes

Exemplo de distribuição de frequência

(cont)

Dados na série ordenada :

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Intervalo	Frequência	Frequência Relativa	Percentual
10 mas menos de 20	3	.15	15
20 mas menos de 30	6	.30	30
30 mas menos de 40	5	.25	25
40 mas menos de 50	4	.20	20
50 mas menos de 60	2	.10	10
Total	20	1.00	100

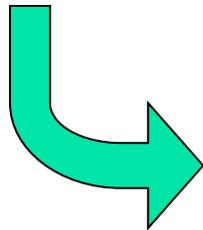


Histograma

- Um gráfico dos dados em uma distribuição de frequência é chamado de **histograma**
- Os pontos finais do intervalo são mostrados no eixo horizontal
- o eixo vertical é frequência, frequência relativa ou porcentagem
- Barras com as alturas apropriadas são usadas para representar o número de observações dentro de cada classe

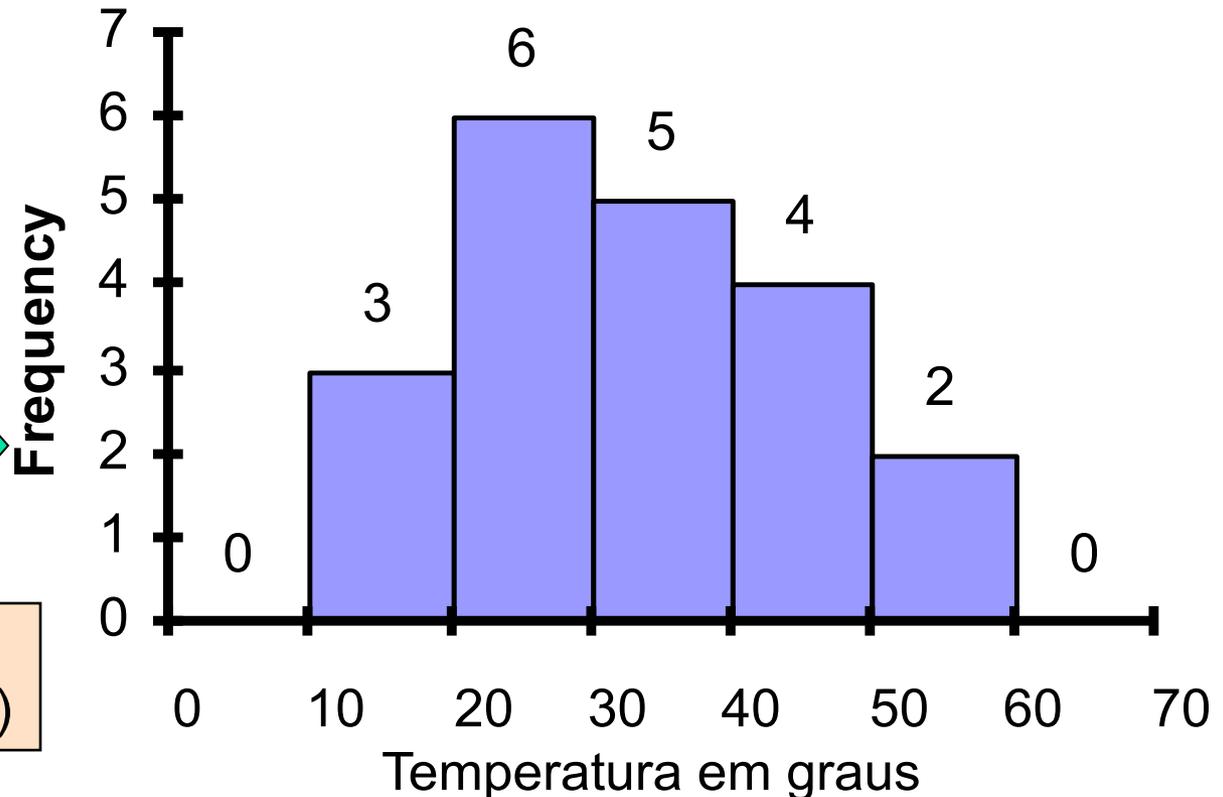
Exemplo de Histograma

Interval	Frequency
10 mas menos de 20	3
20 mas menos de 30	6
30 mas menos de 40	5
40 mas menos de 50	4
50 mas menos de 60	2



(Sem espaços entre as barras)

Histogram: Daily High Temperature



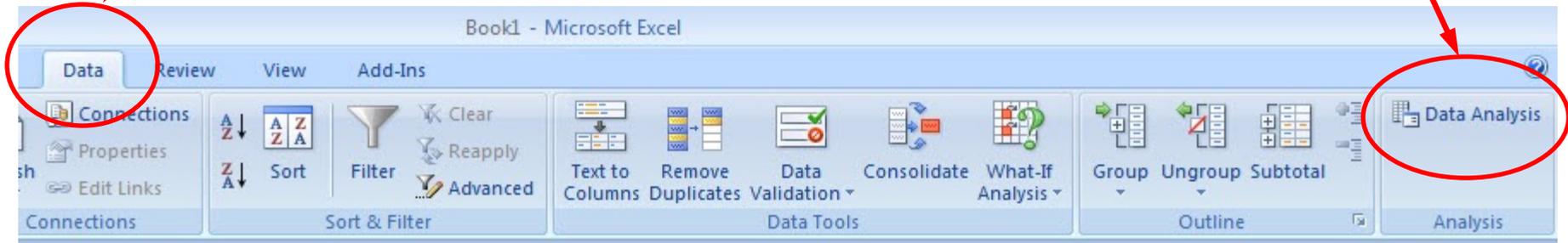
Histogramas no Excel

1

Select **Data** Tab

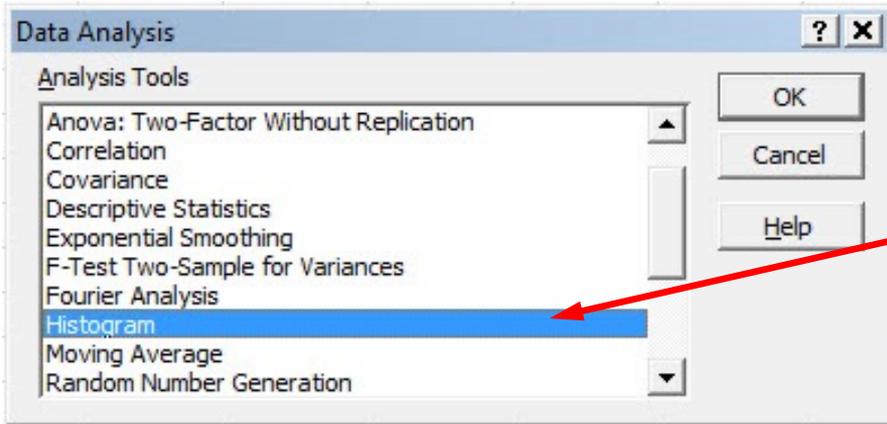
2

Click on **Data Analysis**



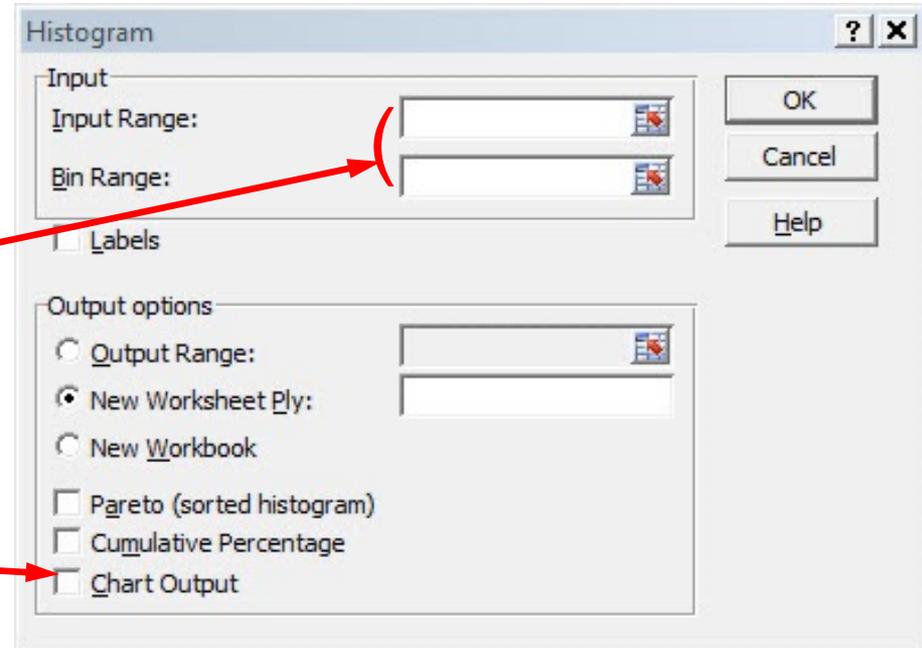
Histograms no Excel

(continued)



3

Choose Histogram



4

Input data range and bin range (bin range is a cell range containing the upper interval endpoints for each class grouping)

Select Chart Output and click "OK"



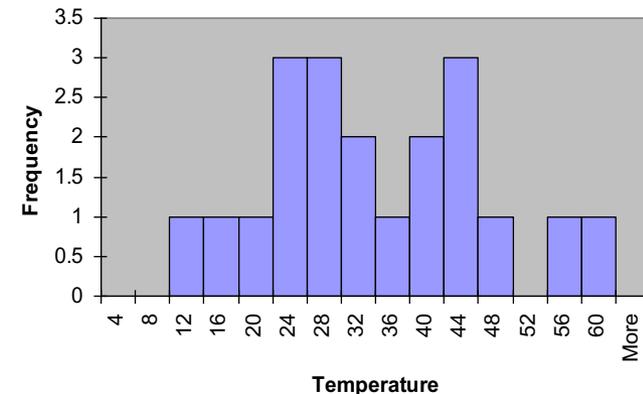
Questões para agrupar dados em intervalos

1. Qual deve ser a largura de cada intervalo?
(Quantas classes devem ser usadas?)
2. Como devem ser determinados os pontos finais dos intervalos?
 - Frequentemente respondido por tentativa e erro, sujeito ao julgamento do usuário
 - O objetivo é criar uma distribuição que não seja nem muito "irregular" nem muito "por blocos"
 - O objetivo é mostrar adequadamente o padrão de variação nos dados

Quantas classes devem ser usadas?

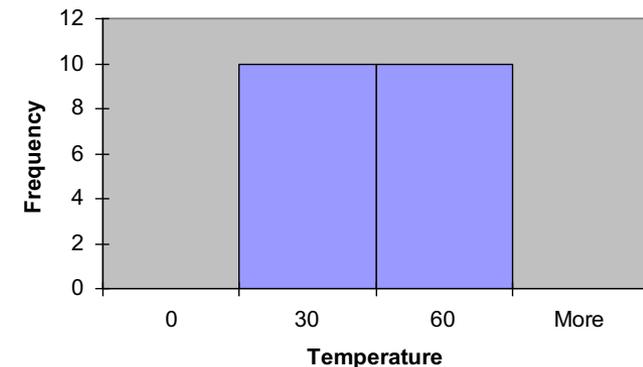
Muitos (intervalos de classe estreitos)

- pode produzir uma distribuição muito irregular com lacunas de classes vazias
- pode dar uma indicação ruim de como a frequência varia entre as classes



Poucos (Amplios intervalos de aula)

- pode comprimir muito a variação e produzir uma distribuição em blocos
- pode obscurecer importantes padrões de variação.



(X axis labels are upper class endpoints)

A distribuição de frequência acumulada

Dados na série ordenada :

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

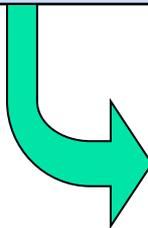
Classe	Frequência	Percentual	Frequência acumulada	Percentual acumulado
10 mas menos de 20	3	15	3	15
20 mas menos de 30	6	30	9	45
30 mas menos de 40	5	25	14	70
40 mas menos de 50	4	20	18	90
50 mas menos de 60	2	10	20	100
Total	20	100		

Gráficos acumulados de frequências

Ogive



Intervalo	Ponto final do intervalo superior	Porcentagem Acumulada
Menos de 10	10	0
10 mas menos de 20	20	15
20 mas menos de 30	30	45
30 mas menos de 40	40	70
40 mas menos de 50	50	90
50 mas menos de 60	60	100



Ogive: Daily High Temperature

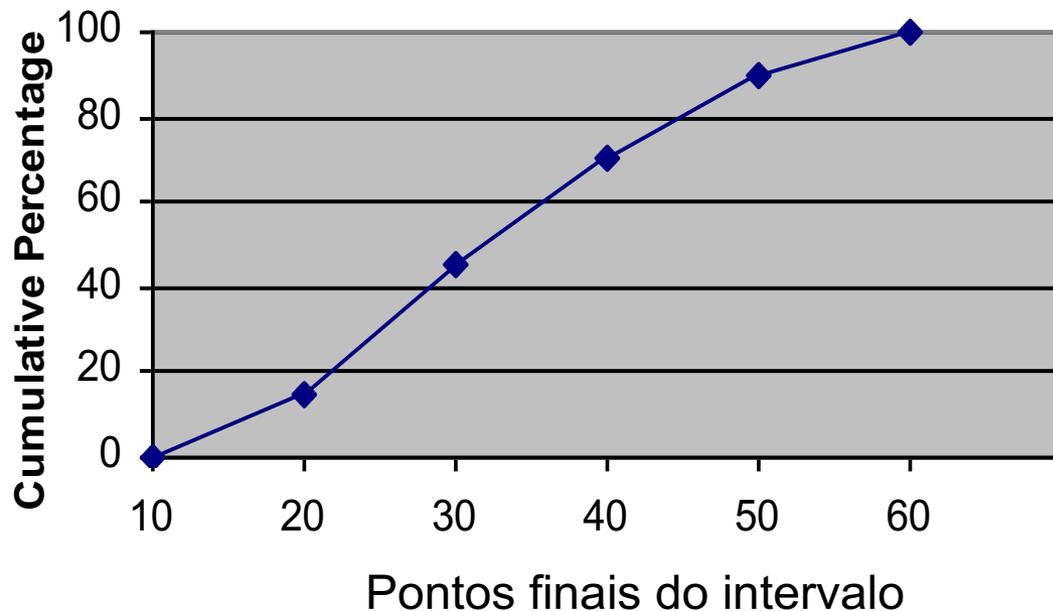




Diagrama de caule e folha

- Uma maneira simples de ver detalhes de distribuição em um conjunto de dados
- MÉTODO: Separe a série de dados classificados em dígitos iniciais (o caule) e os dígitos finais (as folhas)

Exemplo

Dados ordenados em série :

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- Aqui, use o dígito 10 para a unidade do caule:

- 21 é mostrado como
- 38 é mostrado como

Caule	Folha
2	1
3	8

Exemplo

(cont)

Dados ordenados em série :

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- **Diagrama de caule e folha completo**

Caule	Folhas
2	1 4 4 6 7 7
3	0 2 8
4	1

Usando outras unidades de caule

- Usando dígitos de 100 como caule:
 - Arredonde o algarismo de 10 para formar as folhas

- 613 seria
- 776 seria
- ...
- 1224 seria.

	Caule	Folha
→	6	1
→	7	8
→	12	2

Usando outras unidades de caule

(cont)

- Usando dígitos de 100 como caule:
A exibição completa de caule e folhas:

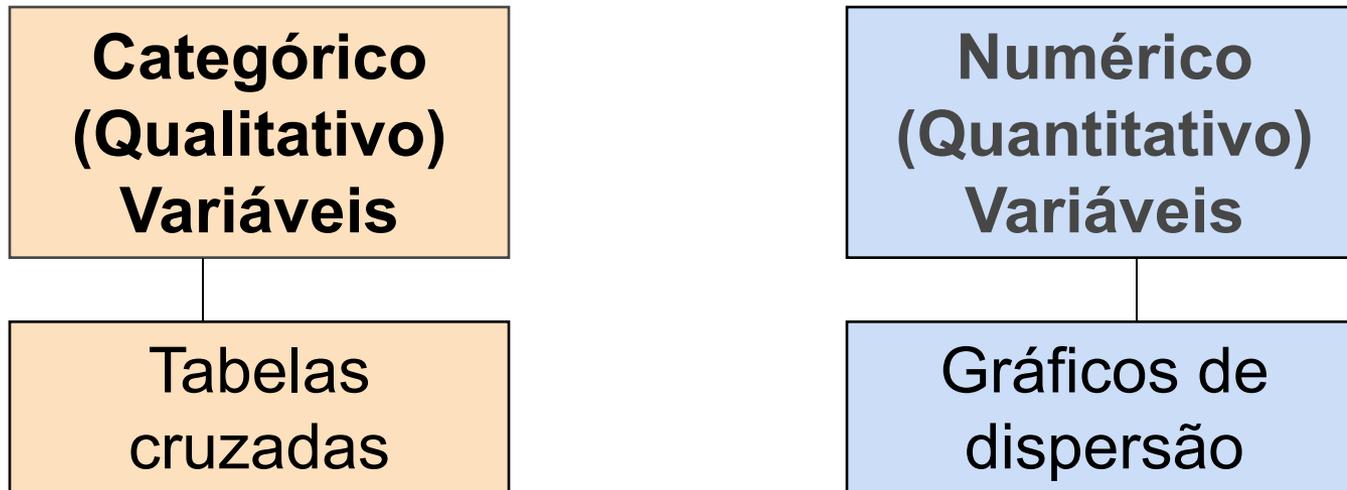
Dados:
613, 632, 658, 717, 722, 750, 776, 827, 841, 859, 863, 891, 894, 906, 928, 933, 955, 982, 1034, 1047, 1056, 1140, 1169, 1224



Caule	Folhas
6	1 3 6
7	2 2 5 8
8	3 4 6 6 9 9
9	1 3 3 6 8
10	3 5 6
11	4 7
12	2

Relações entre Variáveis

- Os gráficos ilustrados até agora envolveram apenas uma única variável
- Quando existem duas variáveis, outras técnicas são usadas:





Gráficos de dispersão

- Diagramas de dispersão são usados para observações pareadas tiradas de duas variáveis numéricas
- O diagrama de dispersão:
 - uma variável é medida no eixo vertical e a outra variável é medida no eixo horizontal

Exemplo de Gráfico de Dispersão

Volume diário	Custo diário
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200

Cost per Day vs. Production Volume

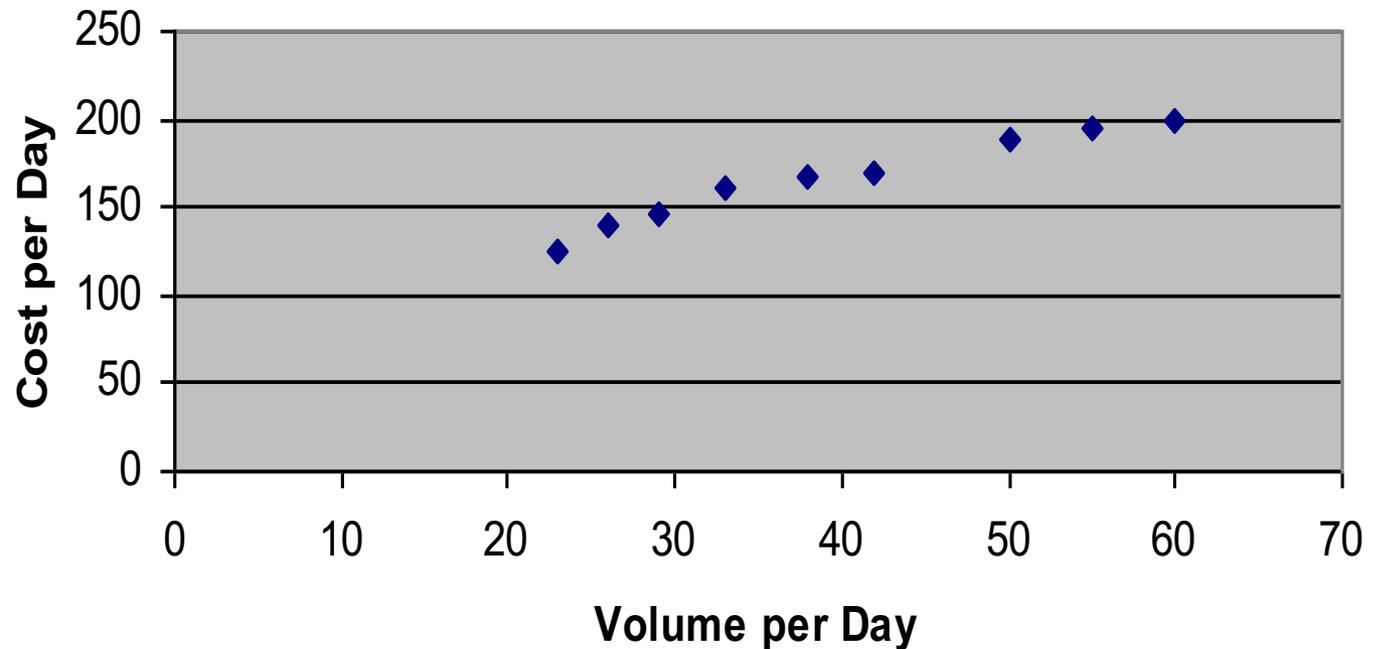
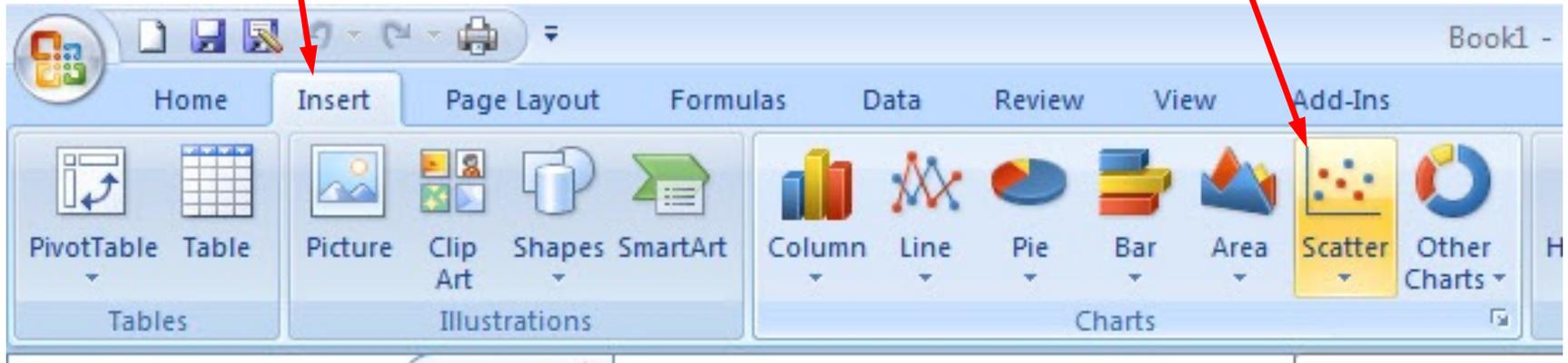


Gráfico de Dispersão em Excel

1 Select the **Insert** tab

2 Select **Scatter** type from the Charts section



3 When prompted, enter the data range, desired legend, and desired destination to complete the scatter diagram



Tabelas cruzadas

- Tabelas cruzadas (ou tabelas de contingência) listam o número de observações para cada combinação de valores para duas variáveis categóricas ou ordinais
- Se houver **r** categorias para a primeira variável (linhas) e **c** categorias para a segunda variável (colunas), a tabela é chamada de tabela cruzada **r x c**.



Exemplo de Tabela cruzada

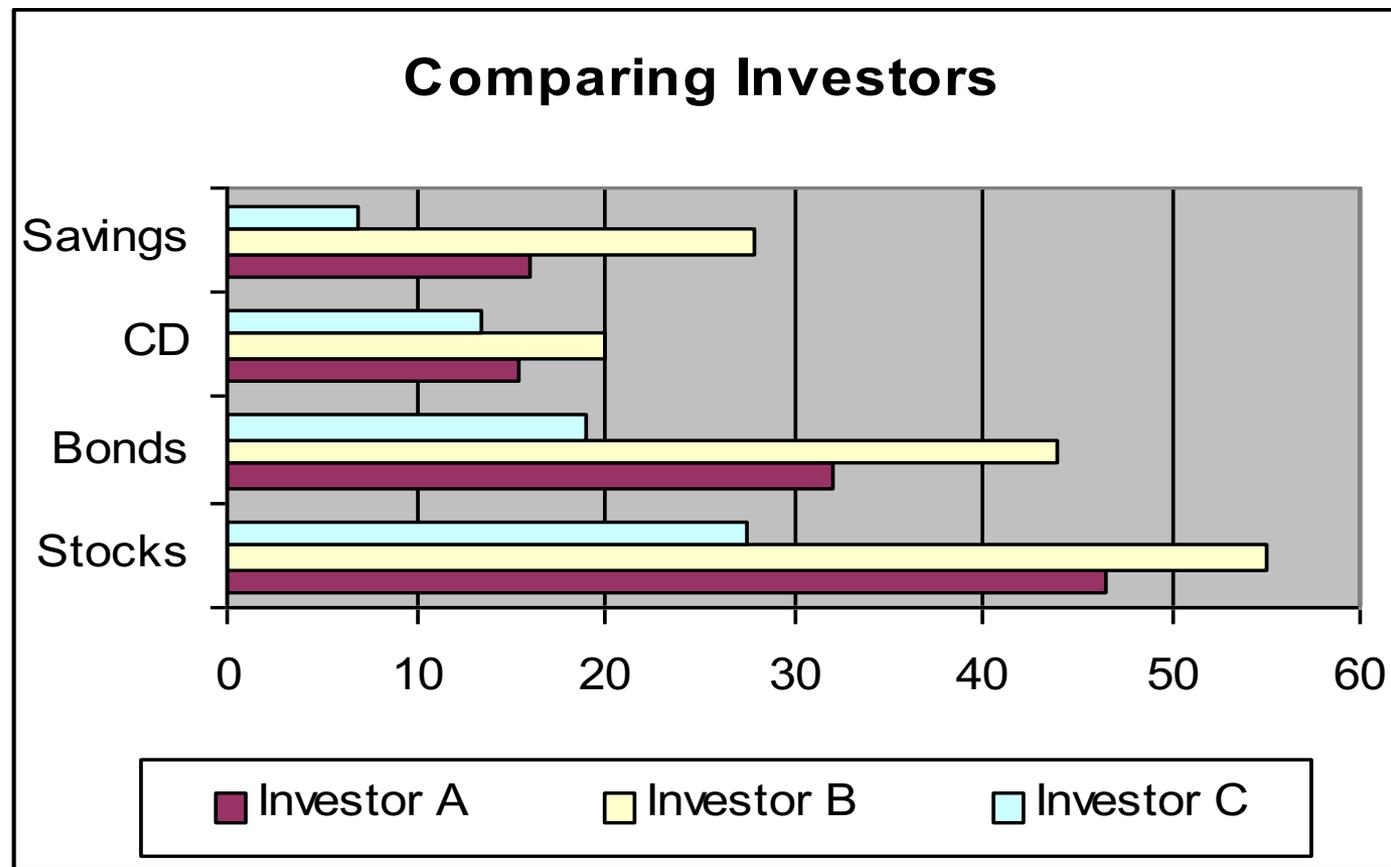
- Tabela cruzada 4 x 3 para opções de investimento por investidor (valores em US\$ 1.000)

Categoria	Investidor A	Investidor B	Investidor C	Total
Ações	46.5	55	27.5	129
Títulos	32.0	44	19.0	95
CD	15.5	20	13.5	49
Poupança	16.0	28	7.0	51
Total	110.0	147	67.0	324

Representação gráfica de dados categóricos multivariados

(cont)

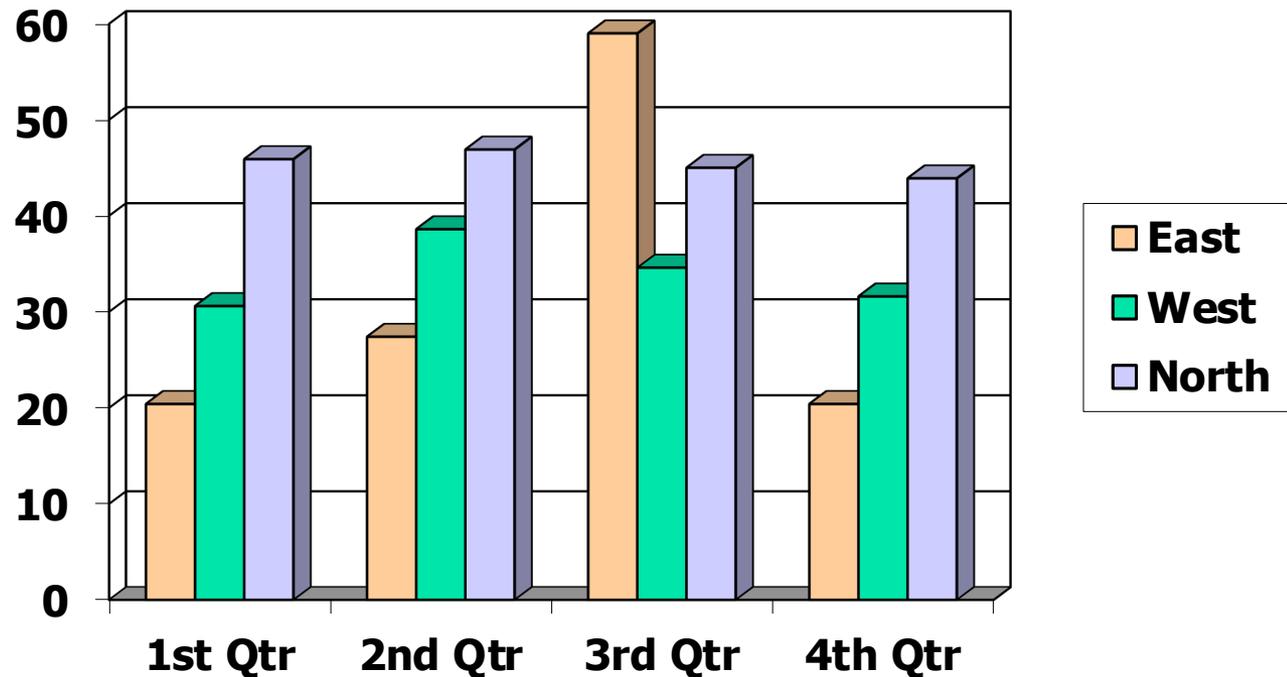
- Gráficos de barras lado a lado



Exemplo de Gráficos de barras lado a lado

- Sales by quarter for three sales territories:

	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr
East	20.4	27.4	59	20.4
West	30.6	38.6	34.6	31.6
North	45.9	46.9	45	43.9



Como evitar erros de apresentação de dados?

1.7

Metas para uma apresentação de dados eficaz:

Apresentar dados para exibir informações essenciais

Comunicar ideias complexas com clareza e precisão

Evite distorções que possam transmitir a mensagem errada

Evite:

(cont)

- Larguras de intervalo de histograma desiguais
- Comprimir ou distorcer o eixo vertical
- Não fornecendo nenhum ponto zero no eixo vertical
- Falha ao fornecer uma base relativa na comparação de dados entre grupos

