



Introdução à Estatística

Profa Alcione Miranda dos Santos
Departamento de Saúde Pública – UFMA
Núcleo de Estatística e Informática – HUUFMA
email: alcione.miranda@terra.com.br

Estatística: O que é?

- Estatística pode ser pensada como a ciência de aprendizagem a partir de dados.
- Em linhas gerais, a Estatística fornece métodos que auxiliam o processo de tomada de decisão.
- A Estatística está presente em todas as áreas da ciência que envolvam a coleta e análise de dados.

Introdução à Estatística

A Estatística está compreendida em duas partes:

- **Estatística Descritiva:** Reúne um conjunto de técnicas para sumarizar os dados (**tabelas**, **gráficos**) e medidas descritivas que permitem tirar muitas informações contidas nos dados.
- **Estatística Indutiva:** Produzir afirmações sobre uma dada característica da população, na qual estamos interessados, a partir de informações colhidas de uma parte dessa população.

Conceitos Básicos de Estatística

- A finalidade da pesquisa é coletar **dados** para obter informações.
 - **Dados** – observações de uma ou mais **variáveis**.
 - **Variável** é aquilo que se deseja observar para se tirar algum tipo de conclusão, por ex., idade, sexo, peso e outras.
 - Dados usualmente provem de uma **amostra**, a qual representa uma **população** de interesse.

Conceitos Básicos de Estatística

- **População:** É o conjunto de indivíduos (ou objetos) que apresentam pelo menos uma característica em comum, cujo comportamento deseja-se analisar ou inferir.

Exemplo: Estudo sobre a ocorrência de sobrepeso em crianças de 7 a 12 anos no Município de São Luís.

- População alvo** – todas as crianças nesta faixa etária deste município.
 - População de estudo** – crianças matriculadas em escolas.
-
- **Amostra:** É um subconjunto da população.

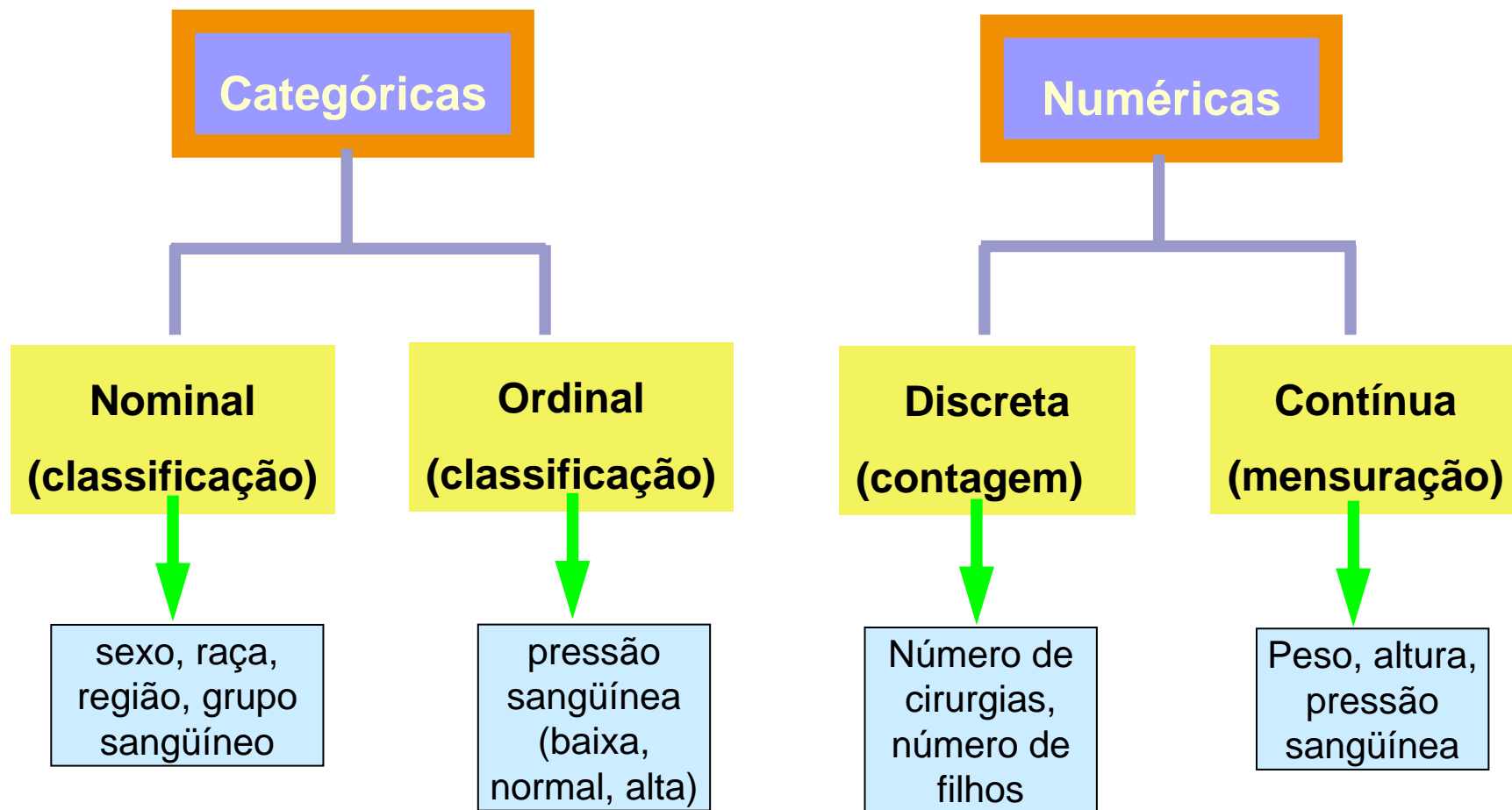
Conceitos Básicos de Estatística

- **Parâmetro:** uma medida numérica que descreve alguma característica de uma população.
 - Frequentemente desconhecido e denotado por **letras gregas**
 - **Exemplo:** **Peso médio** ao nascer de crianças que nascem no município de São Luís
- **Estatística:** uma medida numérica que descreve alguma característica de uma amostra.
 - É habitualmente representada por letras latinas
 - **Exemplo:** **Peso médio** ao nascer, calculado em uma amostra de 120.000 crianças nascidas no Município de São Luís

Tipos de Variáveis

- As variáveis podem ser categóricas (**qualitativas**) ou numéricas (**quantitativas**)
- **Variáveis qualitativas**: São características de uma população que não pode ser medidas.
 - Ordinais – **Ex**: Grau de gravidade de uma doença
 - Nominais – **Ex**: Presença de um sintoma
- **Variáveis quantitativas**: São características de uma população que pode ser quantificadas.
 - Discretas – **Ex**: Número de cirurgias
 - Contínuas – **Ex**: Idade, Pressão Arterial

Esquemáticamente



Tipos de variáveis

Classifique as variáveis apresentadas na tabela:

Idade	Sexo	Hemoglobina	Tipo de urticária	Duração
34	masc	14,2	física	curta
58	masc	14,4	física	longa
31	fem	15,1	idiopática	média
49	masc	10,9	idiopática	média
39	fem	14,4	física	longa
33	masc	14,1	física	curta
35	fem	14,0	idiopática	longa

O tipo da variável irá indicar a melhor forma para apresentação em tabelas e gráficos, em medidas de resumo e a análise estatística mais adequada.



O Papel da Estatística na Pesquisa Científica

Profa Alcione Miranda dos Santos

Departamento de Saúde Pública – UFMA

Núcleo de Estatística e Informática – HUUFMA

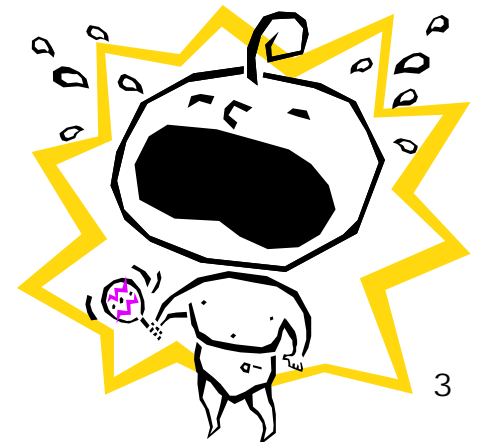
email: alcione.miranda@terra.com.br

Qual o papel da Estatística na ciência?

- O propósito da investigação é responder uma *questão científica*.
- Na ciência, são realizados estudos *experimentais* ou *observacionais*, levando à coleção de *dados* numéricos.
- O padrão de *variação* nos dados faz com que a resposta não seja óbvia.

Por que usar Estatística?

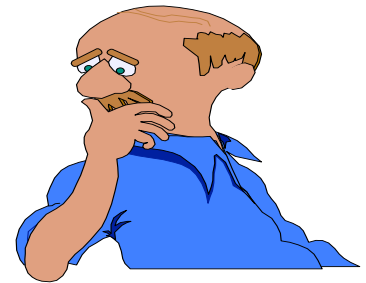
- Por que a natureza apresenta **VARIABILIDADE**:
 - Variações de indivíduo para indivíduo;
 - Variações no mesmo indivíduo;
- Segundo Pereira (1997), a estatística é a tecnologia da ciência e, portanto, a estatística deve estar presente desde o início da pesquisa.
- Sem Métodos Estatísticos, sem validade científica!



Tipos de pesquisas científicas

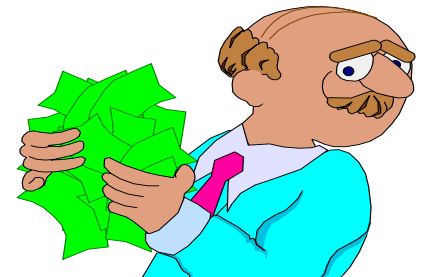
- **DE LEVANTAMENTO**

Características de interesse de uma população são levantadas (observadas ou medidas), mas sem manipulação.



- **EXPERIMENTAL**

Grupos de indivíduos (ou animais, ou objetos) são manipulados para se avaliar o efeito de diferentes tratamentos.



Pesquisas de levantamento

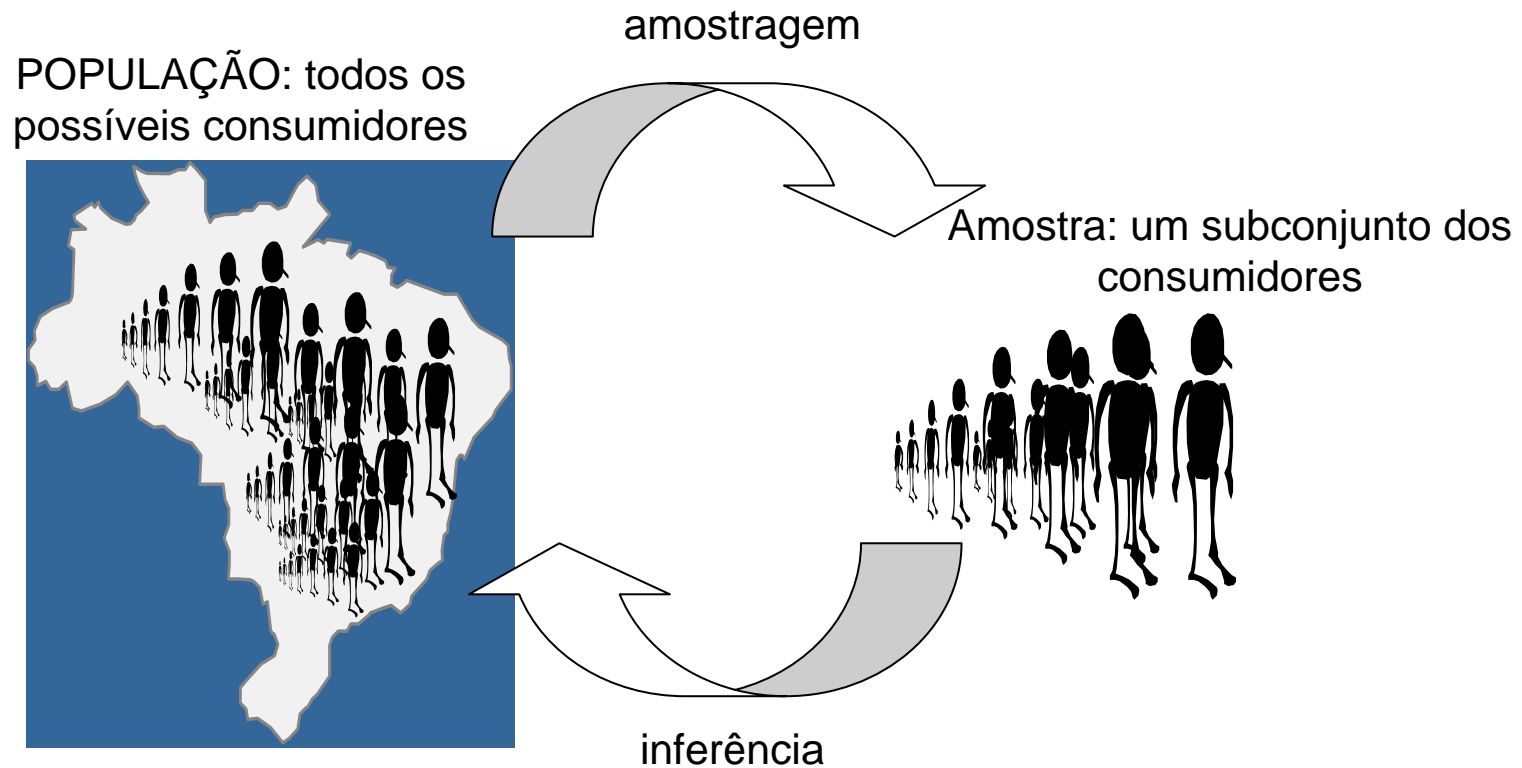
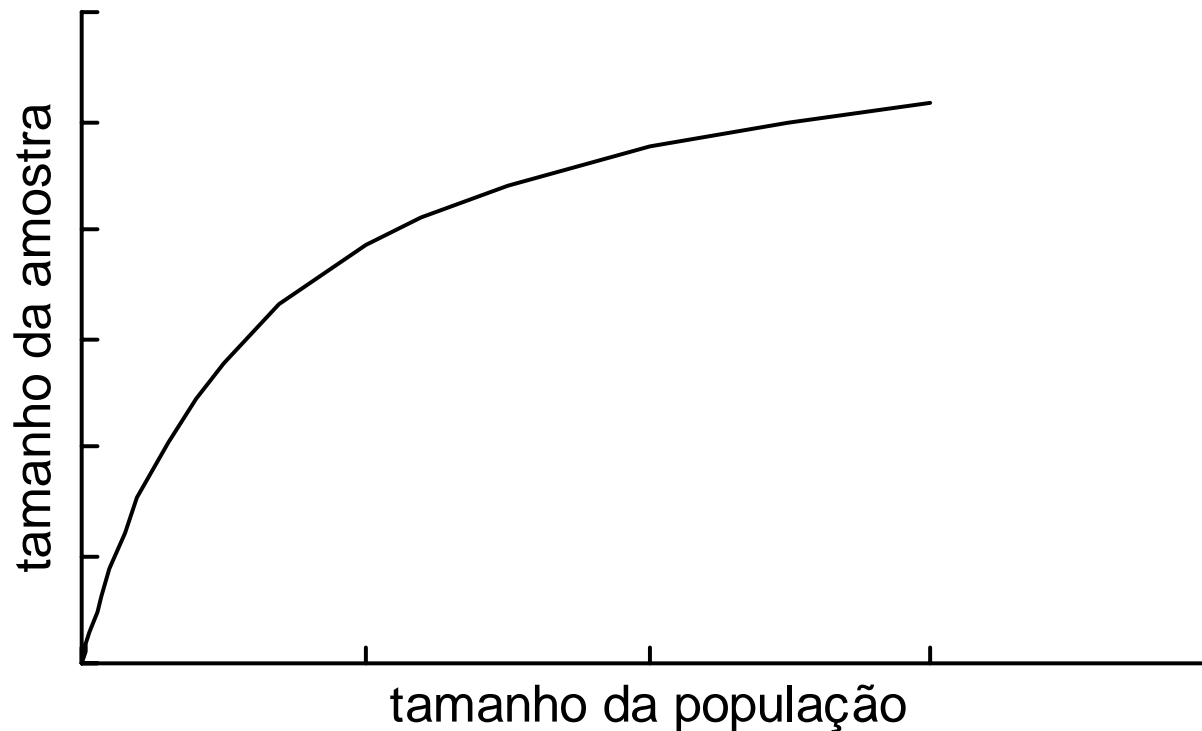


Ilustração de um levantamento por amostragem

Amostragem

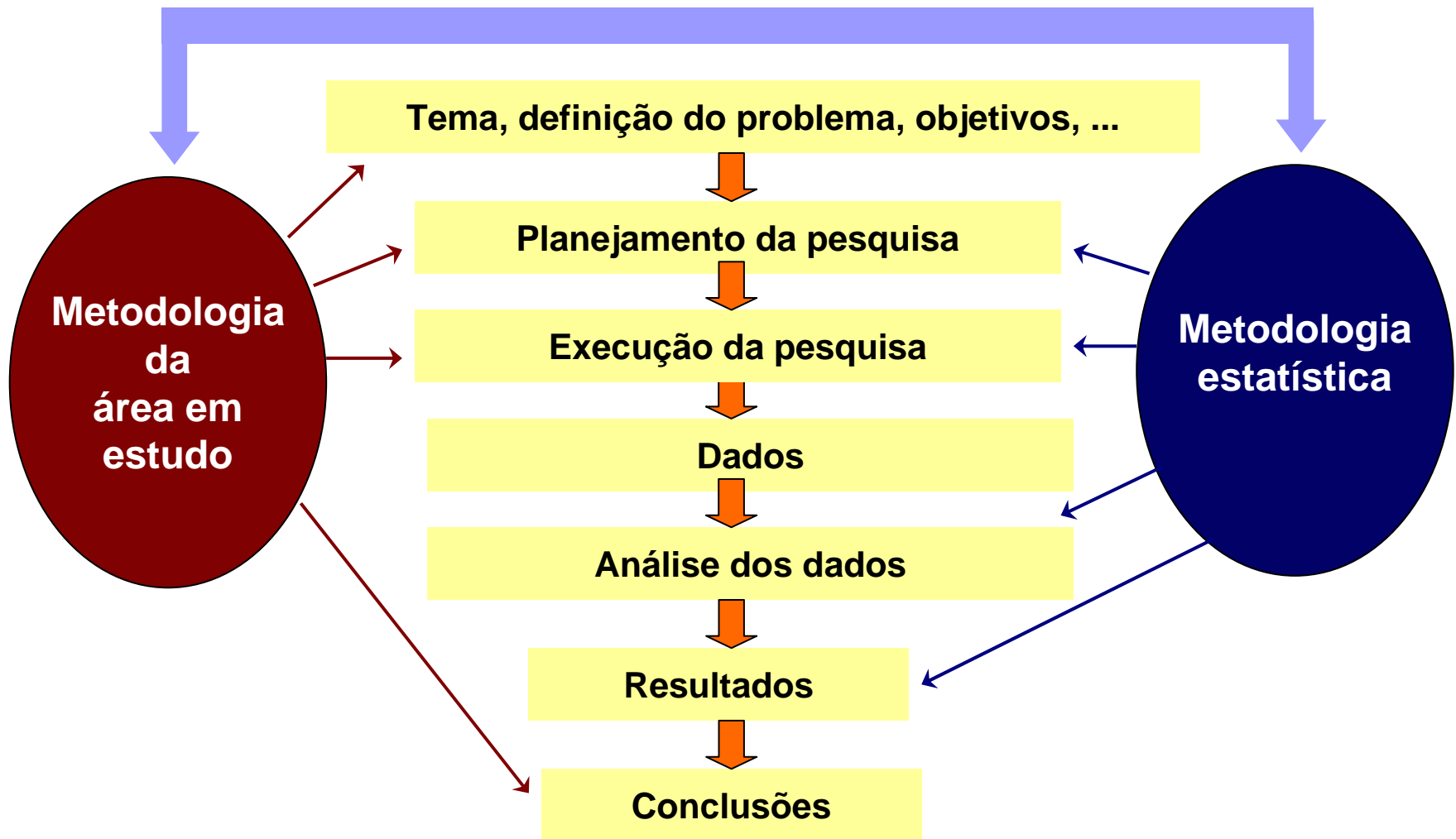
- Representatividade da amostra
- Tamanho da amostra
- Aleatoriedade da amostra
 - Garantir que TODOS os elementos da população tenham chance de pertencer à amostra.
 - Sorteio NÃO VICIADO.
 - Única forma de poder generalizar estatisticamente os resultados para a população.

Censo ou amostragem



Relação entre tamanho da população e tamanho da amostra para garantir determinada margem de erro

Etapas usuais de uma pesquisa científica



Fases de uma Pesquisa

- Definição do problema
- Planejamento
- Coleta dos dados
- Apuração dos dados
- Apresentação dos dados
- Análise e interpretação dos dados

Definição do Problema

- Formular corretamente o problema.
- Definir a população a ser estudada.
- Quais variáveis serão observadas?
- Quais hipóteses serão avaliadas?
- Determinar o que se pretende investigar.
- Estudos realizados (*revisão da literatura*).

Planejamento da Pesquisa

- Nesta fase, são definidos:
 - Objetivos a serem alcançados
 - Bibliografia, materiais, impressos, equipamentos a serem utilizados
 - Tipo de levantamento (censo ou amostragem)
 - Pessoal que vai ser envolvido no trabalho
 - Locais de trabalho
 - Cronograma da execução

Perguntas que precisam ser respondidas no planejamento de uma pesquisa

■ O quê?

- características a serem observadas → VARIÁVEIS

■ Quem?

- os elementos a serem pesquisados →
POPULAÇÃO

■ Como?

- o instrumento de coleta de dados →
QUESTIONÁRIO / ENTREVISTA ESTRUTURADA

Coleta dos dados

- Definir o instrumento de pesquisa:
 - Prontuários
 - Protocolos
 - Questionários

- Tipos de informações:
 - Primárias
 - Secundárias

Apuração dos Dados

- Consiste em resumir os dados, através de contagem ou agrupamento.
- Freqüentemente, exige um programa computacional.
- Por exemplo, **Acess**, **Excel** ou **Epi Info**.
- As variáveis categóricas devem ser codificadas.
- Codificar dados ausentes.

Exemplo de Banco de Dados

Paciente	Sexo	Peso	Tipo de Tratamento	Nº de Convulsões	Classificação da Doença
1	M	89,79	A	1	Leve
2	F	64,20	A	3	Severa
3	M	91,00	B	2	Moderada
4	F	51,68	A	0	Moderada
5	F	48,52	B	1	Leve
...					
58	M	71,00	B	0	Severa
59	M	78,80	A	2	Leve
60	F	71,00	B	3	Moderada

Análise e Interpretação dos Dados

- Tirar conclusões que auxiliem o pesquisador.
- Necessidade de um programa estatístico (STATA, EPIINFO, BIOESTAT, SAS, SPSS, MINITAB)
- Análise estatística:
 - Estatística Descritiva: tabelas ou gráficos, média, mediana, desvio padrão.
 - Estatística Indutiva: testes estatísticos.



Apresentação dos dados

Profa Alcione Miranda dos Santos

Departamento de Saúde Pública – UFMA

Núcleo de Estatística e Informática – HUUFMA

email: alcione.miranda@terra.com.br

Análise Descritiva dos Dados

- A análise descritiva consiste basicamente na **organização** e **descrição** dos dados.
- Elementos básicos: **tabelas**, **gráficos** e **medidas numéricas**.
- Começaremos a análise com apenas uma variável em estudo.

Tabela de Frequências

- Forma de representação da frequência de cada valor distinto da variável em estudo.
- Juntamente com as frequências simples, a tabela poderá ainda incluir:
 - Frequências relativas
 - Frequências acumuladas
 - Frequências relativas acumuladas.

Tabela de Freqüências

- **Freqüência relativa:** porcentagem relativa à freqüência.
- **Freqüência acumulada:** número de vezes que uma variável assume um valor inferior ou igual a esse valor.
- **Freqüência relativa acumulada:** porcentagem relativa à freqüência acumulada.

Tabela de Freqüências

- **Exemplo:** Consideremos a seguinte tabela

Nome	Sexo	Nome	Sexo
Paula	F	Gonçalo	M
Manuel	M	Pedro	M
Carla	F	Cristina	F
Maria	F	Sofia	F
João	M	Susana	F

Temos,

Sexo Masculino:

Frequência absoluta : **4**

Frequência relativa: 4 em 10 = **40%**

Sexo Feminino: Frequência absoluta : **6**

Frequência relativa: 6 em 10 = **60%**

Tabela de Freqüências

- Assim a **tabela de freqüências** da variável Sexo, para o exemplo anterior, será:

Sexo	Freq. Simples (n)	Freq. Relativa (%)
Feminino	6	60
Masculino	4	40
Total	10	100

Elementos essenciais de uma tabela

- **Título:** uma indicação que antecede a tabela e explique tudo referente a tabela.
- **Cabeçalho:** colocado na parte superior da tabela, especificando o conteúdo das colunas.
- **Corpo:** corresponde ao conjunto de colunas e de linhas que contêm informações sobre o fenômeno estudado.

Elementos complementares da tabela

- **Fonte:** é a indicação do órgão ou entidade responsável pelo fornecimento dos dados ou pela sua elaboração. É colocada no rodapé da tabela.
- **Notas:** são informações destinadas a esclarecer o conteúdo das tabelas, ou indicar a metodologia adotada na coleta ou preparo dos dados.
- **Chamadas:** são informações de natureza específica referindo-se a um item específico da tabela, colocado no rodapé da página.

Elementos essenciais de um gráfico

- Todo gráfico deve ter **título**, **escala** e **fonte** de dados, de forma a dispensar qualquer esclarecimento adicional.
- A numeração dos gráficos é feita utilizando-se algarismos arábicos.
- As escalas devem crescer da esquerda para a direita e de baixo para cima.
- As distâncias que indicam as unidades devem ser rigorosamente uniformes.

Variável Qualitativa

- Podemos sumarizar a variável em:
 - **Tabelas** – usando contagens ou porcentagens
 - **Gráfico de Barras** ou **Gráfico de Setores**

Tabelas: Variável Qualitativa

**Tabela 1. Tipo de parto em nascidos vivos de parto único.
São Luís- MA, 1997/98**

Tipo de parto	Freqüências	%
Vaginal	1619	66,27
Cesáreo	824	33,73
Total	2443	100,00

Fonte: Silva et al. (2001)

Tabelas: Variável qualitativa

TABELA 2 – Número e porcentagem de causas de morte de residentes em São Luís, no período de 10 de agosto a 31 de dezembro de 2005.

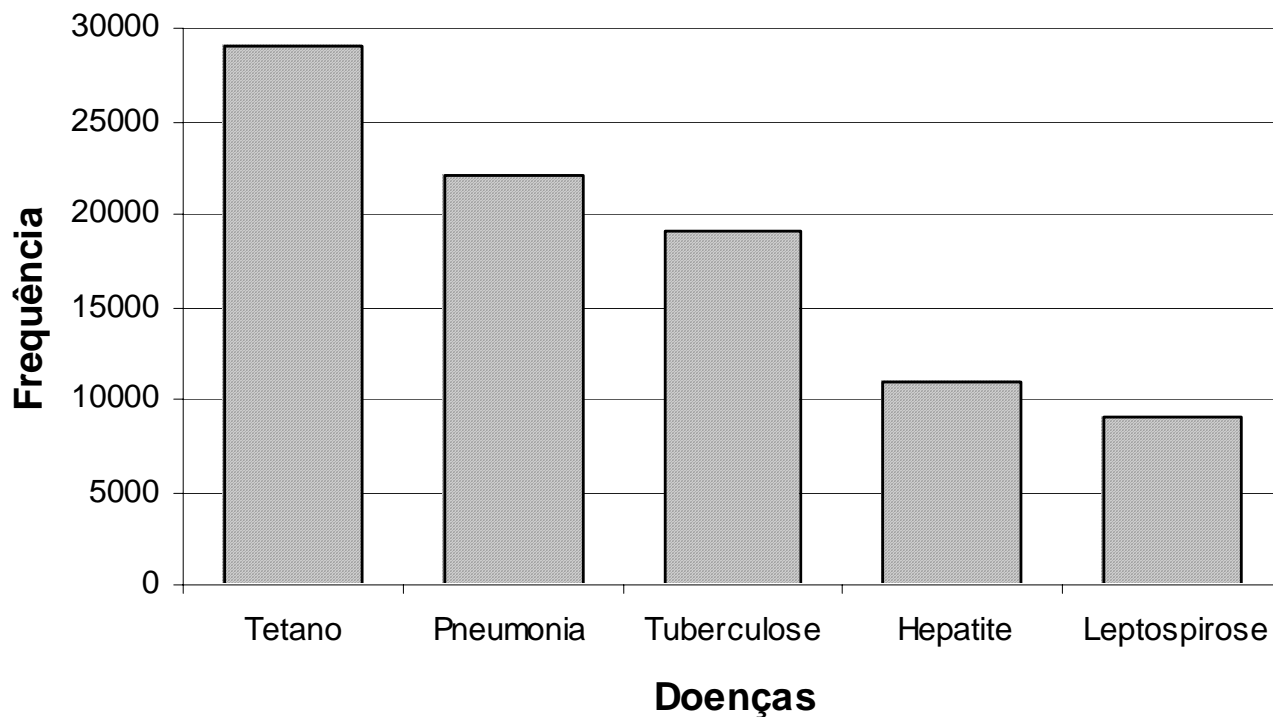
CAUSAS DA MORTE	Frequência	%
Doenças do ap. circulatório	281	33,5
Neoplasias	115	13,7
Causas externas	92	11,0
Doenças do ap. respiratório	87	10,4
Doenças das glând. endócrinas	56	6,7
Doenças do ap. digestivo	54	6,4
Doenças e infec. e parasitárias	46	5,5
Afecções do per. Perinatal	26	3,1
Demais grupos	82	9,8
TOTAL	839	100,0

Fonte: Desconhecida

Gráficos: Variável Qualitativa

Gráfico de Barras

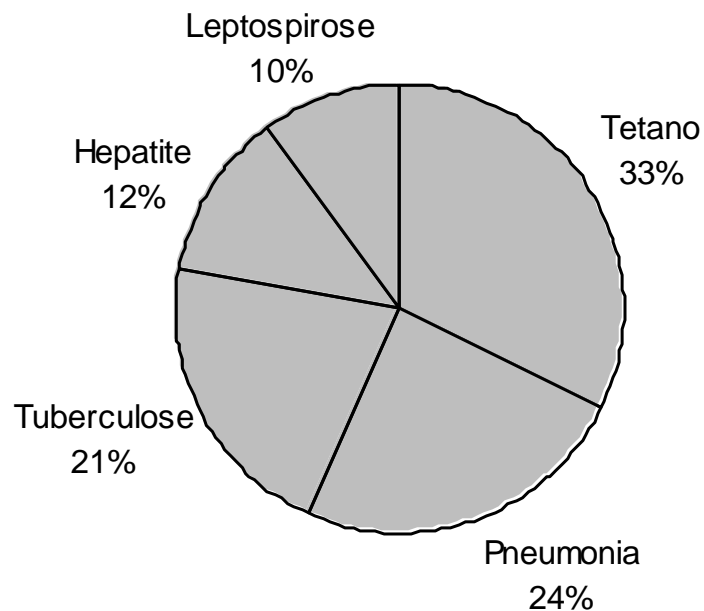
Figura 1: Dados sobre as doenças mais comuns ocorridas no Estado de São Paulo



Gráficos: Variável Qualitativa

Gráfico de Setores

FIGURA 2: Dados sobre as doenças mais comuns ocorridas no Estado de São Paulo



Variável Quantitativa

- Podemos sumarizar a variável em:
 - Tabelas de Freqüências
 - Histograma ou Polígono de Freqüências
 - Gráficos de linhas
 - Box plot

Tabela de Frequências

TABELA 2: Tempo de Internação (em dias) de 160 pacientes no Hospital X

Tempo de Internação (dias)	Nº de pacientes (f_i)
10 --- 20	38
20 --- 30	45
30 --- 40	30
40 --- 50	22
50 --- 60	10
60 --- 70	15
Total	160

Fonte: Divisão de Estatística (Março-1990)

Determinação das classes de uma tabela de frequências

- Critério para determinar a quantidade de classes:

$$k = 1 + 3,3 \log(n)$$

- Amplitude das classes

$$a = \frac{\text{maior valor} - \text{menor valor}}{\text{número de classes}}$$

Exemplo

- Considere os seguintes dados, referentes ao peso de 30 crianças com sete anos, em kg:

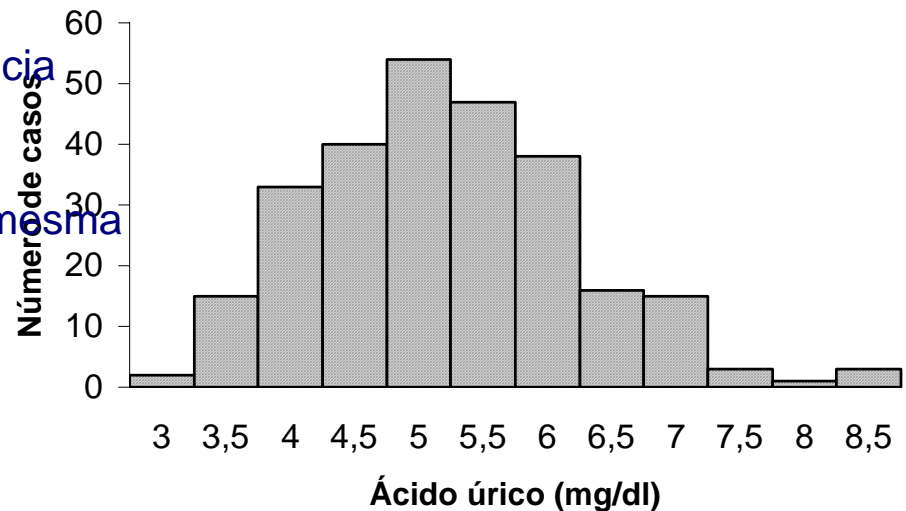
13,00	13,63	14,10	14,10	14,70	15,35	15,54	16,00	16,00	16,30
17,40	17,40	17,70	17,70	17,90	17,90	18,20	18,35	19,10	19,30
19,50	19,70	20,00	20,32	20,50	21,45	21,50	22,00	22,25	24,00

Construa uma tabela de freqüências para os dados acima.

Histograma

- ❑ Representação gráfica da distribuição das frequências absolutas ou relativas
- ❑ Normalmente utilizado para variáveis contínuas.
- ❑ Características:
 - as barras devem estar todas juntas;
 - cada barra representa a frequência de um intervalo de valores;
 - os intervalos devem ter todos a mesma amplitude.

FIGURA 3: Ácido úrico dos pacientes internados no Hospital X



Polígono de Freqüências

FIGURA 4: Ácido úrico dos pacientes internados no Hospital X

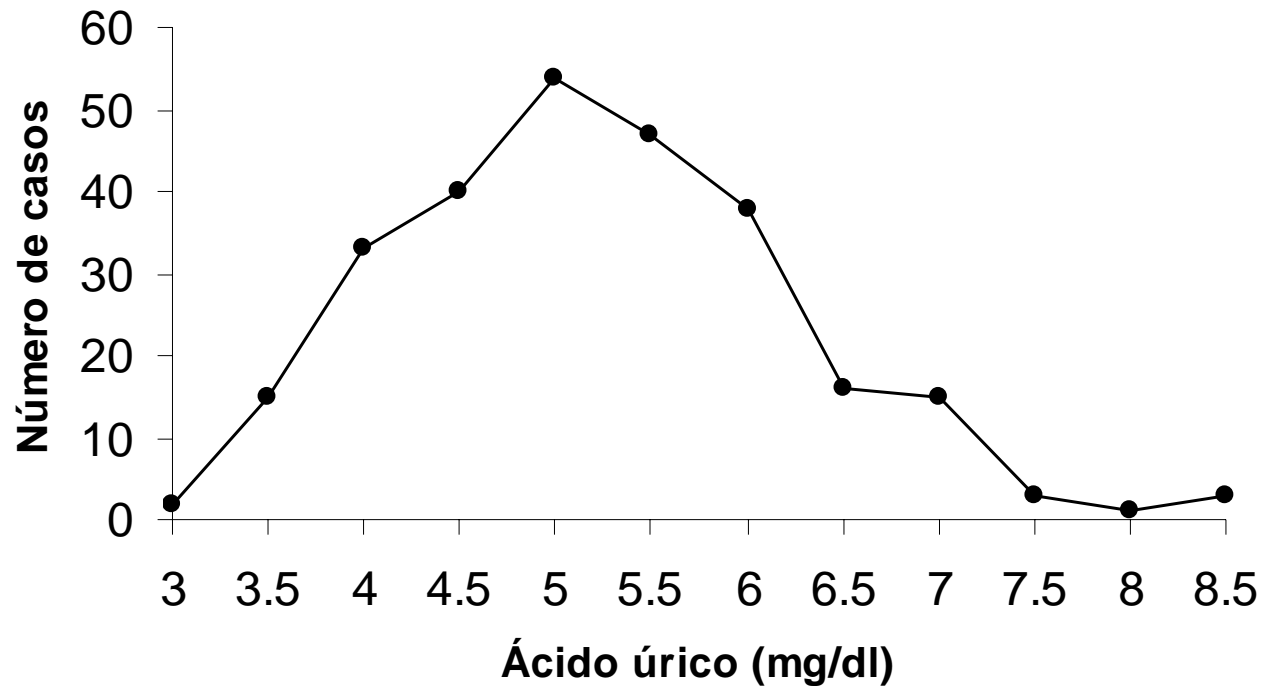
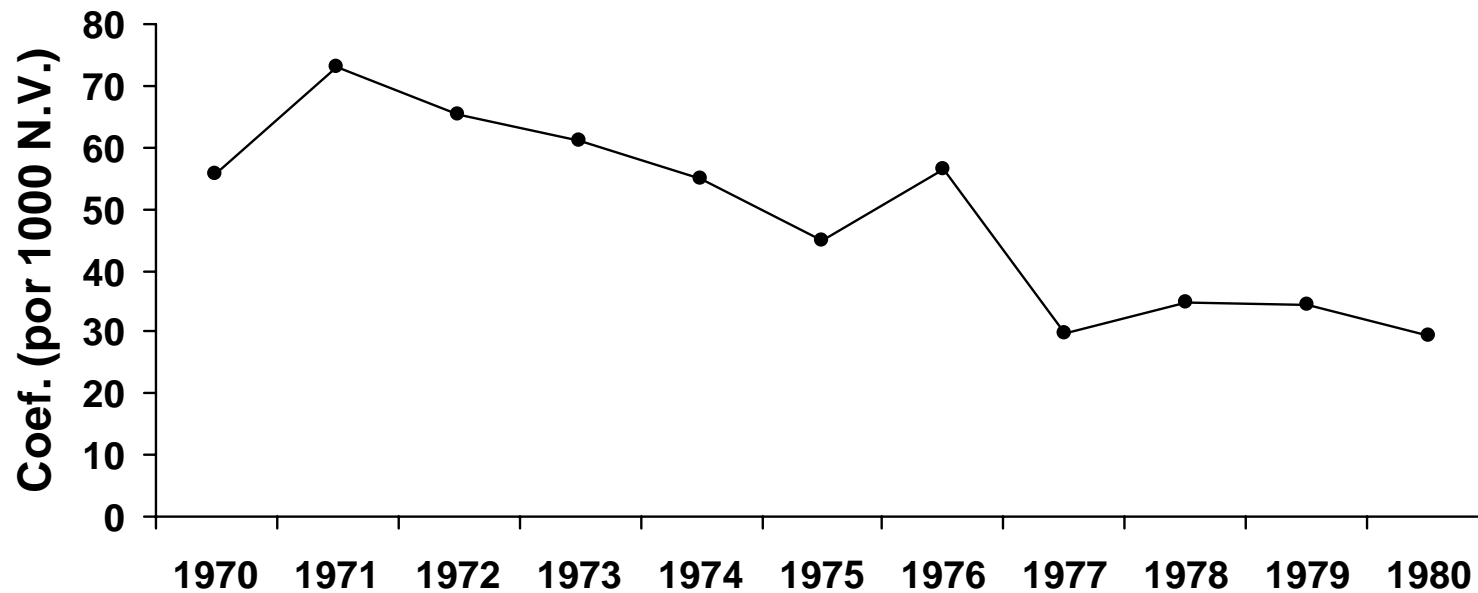


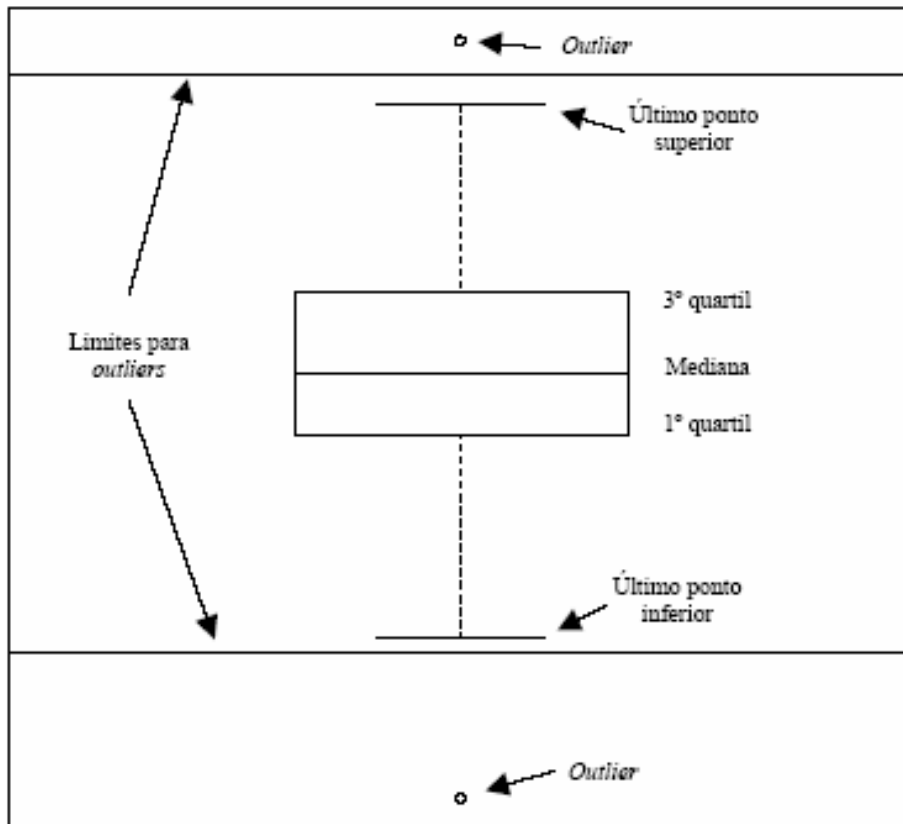
Gráfico de linhas

Mortalidade Infantil, São Caetano do Sul (SP), 1970-80



Box-Plot

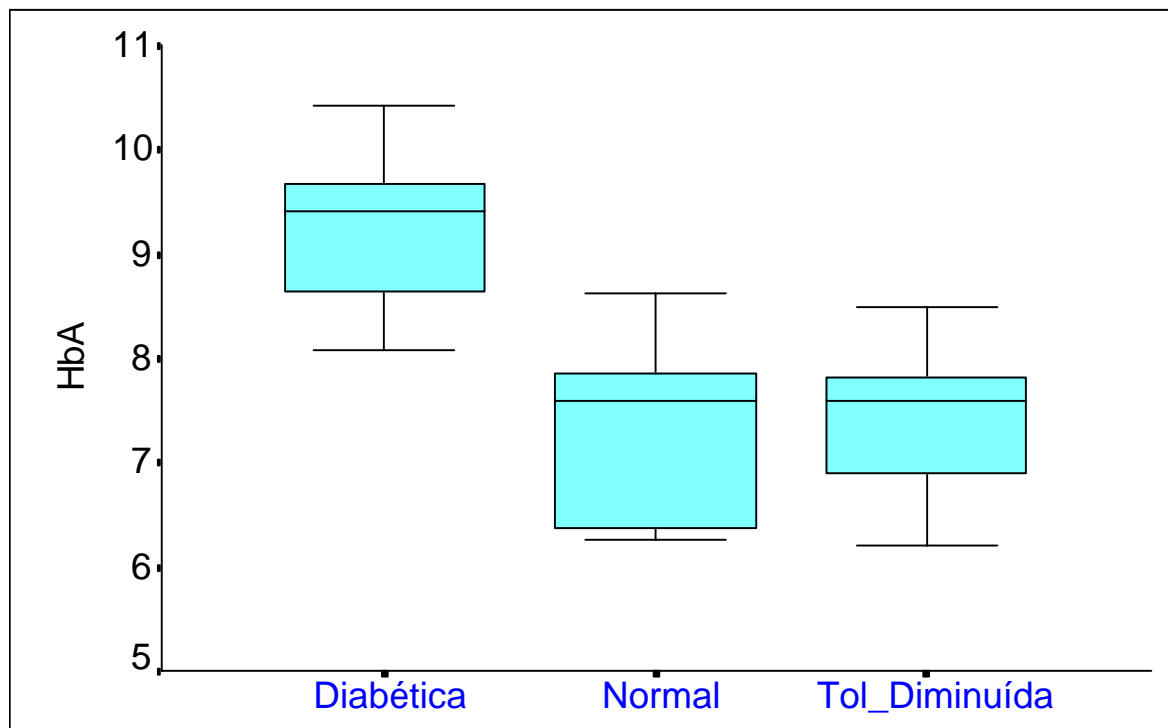
Representação gráfica de cinco medidas: **mínimo**, **quartil inferior**, **mediana**, **quartil superior**, **máximo**



$$\text{limite1} = Q1 - 1,5 \cdot (Q3 - Q1)$$
$$\text{limite2} = Q3 + 1,5 \cdot (Q3 - Q1)$$

Exemplo Box-Plot

Figura 1: Boxplot do nível de Hemoglobina glicosilada, segundo grupo de gestantes.



Análise Bivariada

- Muitas vezes queremos verificar se há uma relação entre duas variáveis (se as variáveis são dependentes ou não).
- Podemos construir tabelas de frequência com dupla entrada. Essas tabelas de dados cruzados são conhecidas por **tabelas de contingência**, e são utilizadas para estudar a relação entre duas variáveis categóricas.

Tabela de Contingência

TABELA 4. Tipo de parto segundo categoria de internação em nascidos vivos de parto único. São Luís - MA, 1997/98

Tipo de parto	Categoria de internação					
	Pública		Privada		Total	
	f	%	f	%	f	%
Cesáreo	572	26,31	252	93,68	824	33,73
Vaginal	1602	73,69	17	6,32	1619	66,27
Total	2174	100,00	269	100,00	2443	100,00

Fonte: Silva et al (2001)

Gráficos: Duas Variáveis Qualitativas

Gráfico de barras

FIGURA 5: Óbitos por acidentes, segundo tipo e sexo.

Município de São Paulo, 1980.

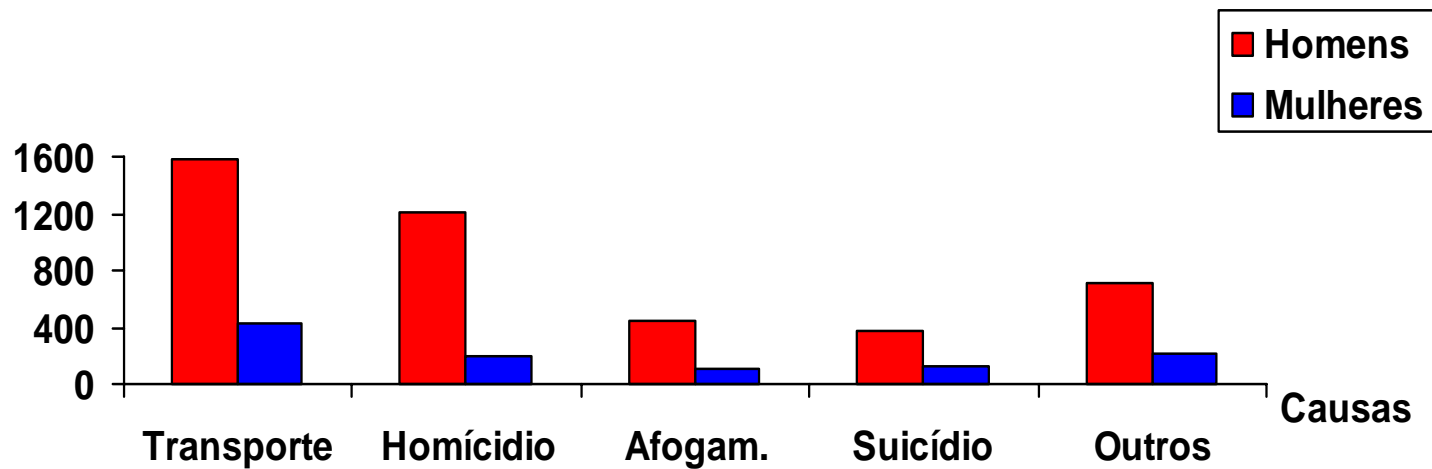
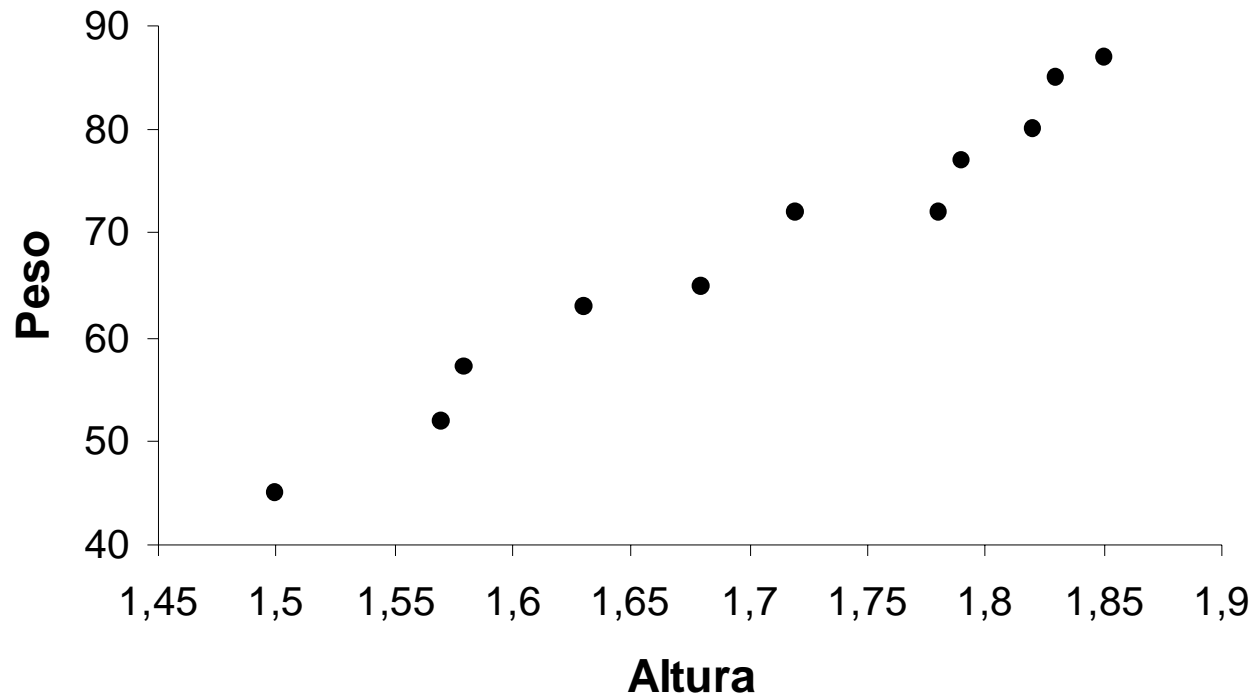


Gráfico: Duas Variáveis Quantitativas

Gráfico de Dispersão





Medidas Descritivas

Profa Alcione Miranda dos Santos
Departamento de Saúde Pública – UFMA
Núcleo de Estatística e Informática – HUUFMA
email: alcione.miranda@terra.com.br



Medidas Descritivas

- Medidas de Tendência Central
- Medidas Separatrizes
- Medidas de Dispersão ou Variabilidade

Medidas de Tendência Central

- Servem para termos uma idéia acerca dos valores médios da variável em estudo.
- São usados para sintetizar em um único número os dados observados.
- São exemplos de medidas de tendência central: **Média, Moda e Mediana.**
- A escolha de qual medida usar, depende...

Média Amostral

- Se os dados consistem de n observações x_1, x_2, \dots, x_n , a média é dada pela soma das observações dividida pelo número de observações. Por exemplo, se os dados são $x_1=2, x_2=3, x_3=1$, então a média é $(2+3+1)/3=2$.
- A média amostral é definida por :

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Média Amostral - Exemplo

Turma A : 2 3 4 4 5 6 7 7 7 7 8

Turma B : 2 3 4 4 4 5 6 7 7 8 9

Objetivo: Obter a média de cada turma:

Turma A

$$(2+3+4+4+5+6+7+7+7+7+8) / 11 = 60/11$$

Média turma A = 5,45

Turma B

$$(2+3+4+4+4+5+6+7+7+8+9)/11 = 59/ 11$$

Média turma B = 5,36

Mediana

- Divide uma distribuição ordenada de dados em duas partes iguais.
- A mediana (M_d) é a observação central, depois de ordenada a amostra.
- Se a amostra tiver dimensão ímpar, a **mediana** coincide com a observação central.

Exemplo: Na amostra 1.2; 1.7; 2.1; 2.2; 2.4 a mediana é 2.1

- Se a amostra tiver dimensão par, a **mediana** toma o valor da média das duas observações mais centrais.

Exemplo: Na amostra 0.3; 0.7; 0.9; 1.1 a mediana é 0.8.

Mediana

- Para calcularmos a mediana é preciso ordenarmos os dados: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$.

- A mediana de um conjunto de dados é:

$$Md = x_{(n+1/2)}, \text{ se } n \text{ é ímpar}$$

$$Md = [x_{(n/2)} + x_{(n/2+1)}] / 2, \text{ se } n \text{ é par}$$

- A mediana é mais robusta que a média a erros ou a observações afastadas.

Mediana - Exemplo

Exemplo 1: Turma A : 2 3 4 4 5 6 7 7 7 8

Turma B : 2 3 4 4 4 5 6 7 7 8 9

Turma A : Mediana = 6

Turma B : Mediana = 5

Exemplo 2: Turma A : 2 3 4 4 5 6 7 7 7 8

Turma B : 2 3 4 4 4 5 6 7 8 9

Turma A : Mediana = $(5+6)/2=5,5$

Turma B : Mediana = $(4+5)/2=4,5$

Mediana - Exemplo

Caso	1	2	3	4	5	6	7	8
X_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Valores	2	4	5	5	7	9	10	30

Qual a média e a mediana ? Resposta: 6 e 5

Qual a média e a mediana ao acrescentarmos a observação 8?

Resposta: 9 e 6

Moda

- Valor que ocorre com maior frequência.
- Obtida por inspeção da tabela de distribuição de frequências.
- Ao contrário do que acontece com a mediana e a média, uma amostra pode possuir mais do que uma moda.

Moda - Exemplo

Turma A : 2 3 4 4 5 6 7 7 7 7 8

Turma B : 2 3 4 4 4 5 6 7 7 8 9

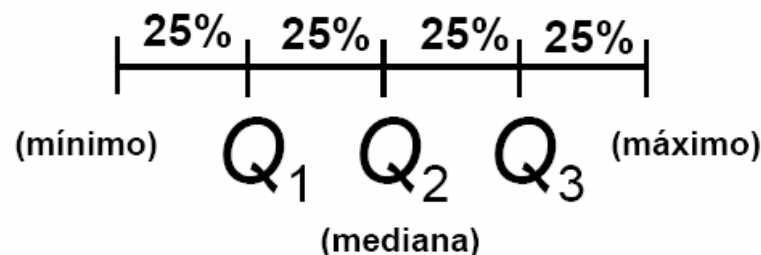
- *Moda turma A = 7*
- *Moda turma B = 4*

Medidas Separatrizes

- Medidas que separam a distribuição em partes iguais.
 - Quartis
 - Decis
 - Percentis

Quartis

- Quartis são os valores (Q_1 , Q_2 e Q_3) que dividem a amostra, depois de ordenada, em **quatro** partes iguais (ou o mais iguais possível).



- Obtendo os quartis
 - Ordena-se os dados;
 - Calcula-se a posição do quartil através da fórmula: $P_{Q_i} = i \cdot \frac{n}{4}$
 - O quartil será o valor que ocupa a posição calculada anteriormente.

Decis

- Dividem um conjunto de dados em dez partes iguais
- Encontra-se o valor do decil desejado, procedendo-se como no caso dos quartis, sendo a posição do decil, encontrada por:

$$P_{Di} = i \cdot \frac{n}{10}$$

Percentis

- Dividem um conjunto de dados em cem partes iguais
- Procede ~~e~~ como no caso dos quartis, sendo que para o cálculo da posição do percentil , a fórmula será:

$$P_{Pi} = i \cdot \frac{n}{100}$$

Medidas Sepatrizes - Exemplo

Turma A

2 3 4 4 5 5 7 7 7 8 8



Q_1

Q_2

Q_3

P_{25}

P_{50}

P_{75}

Md

Medidas de Variabilidade

- Medidas de **tendência central** são descritores insuficientes de uma amostra.
- São necessárias medidas que reflitam a **variação** dentro de um conjunto de dados (**medidas de variabilidade**).
- Essas medidas serão pequenas se os dados forem próximos e grandes se eles estiverem muito espalhados.
- Além disso, tais medidas devem permitir comparar amostras de diferentes tamanhos e determinar se uma amostra é mais variável (ou heterogênea) que a outra.

Exemplo

- Os dados abaixo referem-se aos pesos dos pacientes em dois grupos:

	<i>Grupo A</i>	<i>Grupo B</i>
	78	65
	80	69
	82	78
	85	85
	85	85
	85	93
	86	96
	88	98
<i>Soma</i>	669	669
<i>Média</i>	83,6	83,6
<i>Mediana</i>	85	85
<i>Moda</i>	85	85
<i>N</i>	8	8

Amplitude Total

- Diferença entre o maior e o menor valor do conjunto de dados.

Grupo A

$$\text{AMPLITUDE TOTAL} = 88 - 78 = 10$$

Grupo B

$$\text{AMPLITUDE TOTAL} = 98 - 65 = 33$$

$$\text{AT (grupo A)} < \text{AT (grupo B)}$$

Variância

- É um indicativo da dispersão de um conjunto de dados em relação à média.

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right)$$

- A variância populacional é denotada por σ^2 . Usualmente σ^2 é desconhecida.
- A variância amostral é denotada por S^2 .
Desvantagem - não é expressa na unidade de medida do dado original.

Desvio Padrão

- Corresponde à raiz quadrada da **variância**, tendo portanto a mesma unidade da variável que está sendo estudada. O **desvio padrão** será denotado por S .
- É a medida mais usada na comparação de diferenças entre grupos.
- Fornece um número que permite especificar quão acima ou quão abaixo da média está um determinado valor.
- Quanto maior o **desvio padrão**, maior a variabilidade dos dados.

Coeficiente de Variação

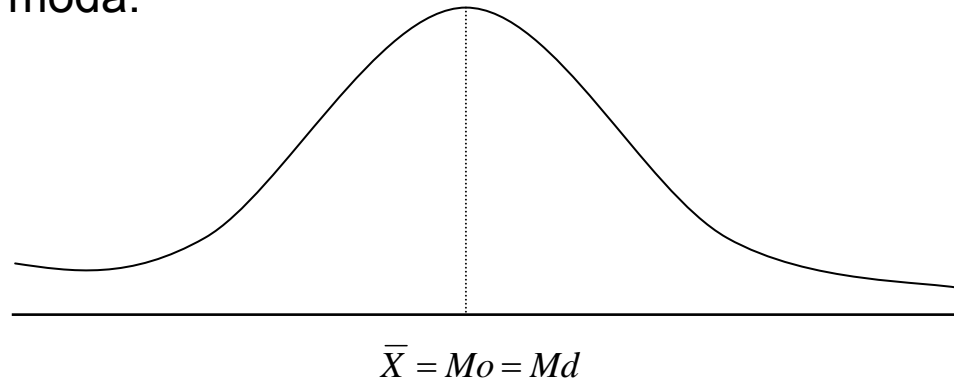
- Muitas vezes o desvio padrão pode ser considerado grande ou pequeno dependendo da ordem de grandeza da variável.
- Pode-se obter um índice relativo de dispersão:

$$CV = \frac{S}{X} \cdot 100$$

- Alguns analistas consideram:
 - Baixa dispersão: $CV \leq 15\%$
 - Média dispersão: $15\% < CV < 30\%$
 - Alta dispersão: $CV \geq 30\%$

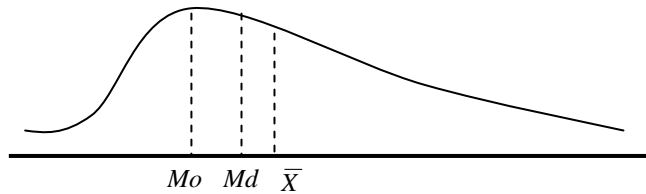
Assimetria

- Assimetria é o grau de deformação de uma curva ou distribuição de frequências.
- Em uma distribuição simétrica tem-se igualdade dos valores da média, mediana e moda.



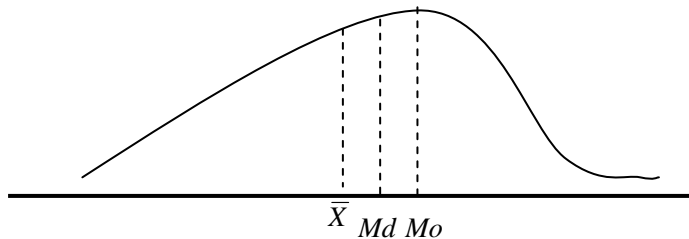
- Toda distribuição deformada é sempre assimétrica. Entretanto, a assimetria pode dar-se na cauda **esquerda** ou na **direita** da curva de frequências.

- Em uma distribuição **assimétrica positiva**, ou assimetria à direita, tem-se :



$$Mo < Md < \bar{X}$$

- Em uma distribuição **assimétrica negativa**, ou assimetria à esquerda, predominam valores inferiores à Moda.



$$\bar{X} < Md < Mo$$

- Fórmulas para o cálculo do coeficiente de assimetria:

Coeficiente de Pearson

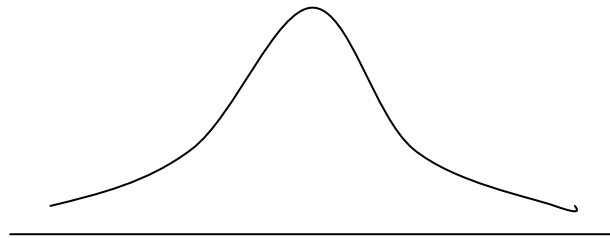
$$AS = \frac{\bar{X} - Mo}{S}$$

AS = 0 diz-se que a distribuição é simétrica

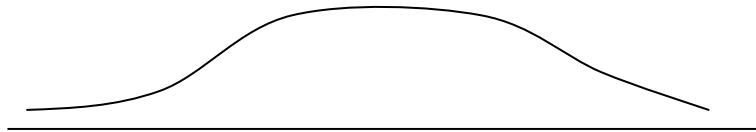
AS > 0 diz-se que a distribuição é assimétrica positiva (à direita)

Curtose

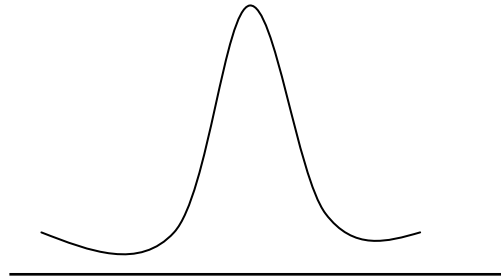
- Denomina-se curtose o grau de achatamento da distribuição.
- Uma distribuição nem chata e nem delgada, é denominada de **mesocúrtica**.



- Uma distribuição achatada denomina-se **platicúrtica**.



- Uma distribuição delgada é denominada de **leptocúrtica**.



- Para medir o grau de curtose utiliza-se o coeficiente:

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

- Se $K = 0,263$, diz-se que a curva correspondente à distribuição de frequência é **mesocúrtica**.
- Se $K > 0,263$, diz-se que a curva correspondente à distribuição de frequência é **platicúrtica**.
- Se $K < 0,263$, diz-se que a curva correspondente à distribuição de frequência é **leptocúrtica**.



Distribuição Normal

Profa Alcione Miranda dos Santos

Departamento de Saúde Pública – UFMA

Núcleo de Estatística e Informática – HUUFMA

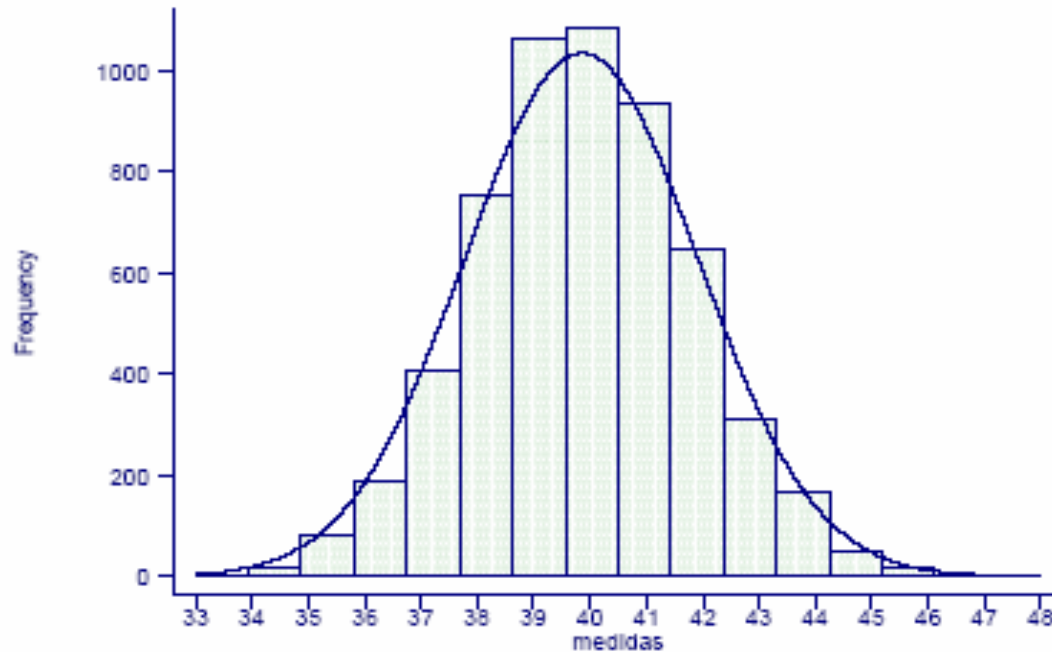
email: alcione.miranda@terra.com.br

Distribuição Normal

- Muitas variáveis estudadas na área biomédica apresentam **distribuição simétrica** (os valores centrais são mais freqüentes e os valores extremos mais raros).
- Na prática, se o coeficiente de assimetria está situado no intervalo $(-0.5, +0.5)$, considera-se a distribuição aproximadamente simétrica.
- Uma distribuição simétrica típica é a **distribuição normal**.

Exemplo: Distribuição Normal

Distribuição de medidas do tórax (polegadas) de soldados escoceses



Fonte: Daly F et al. Elements of Statistics, 1999

Distribuição Normal

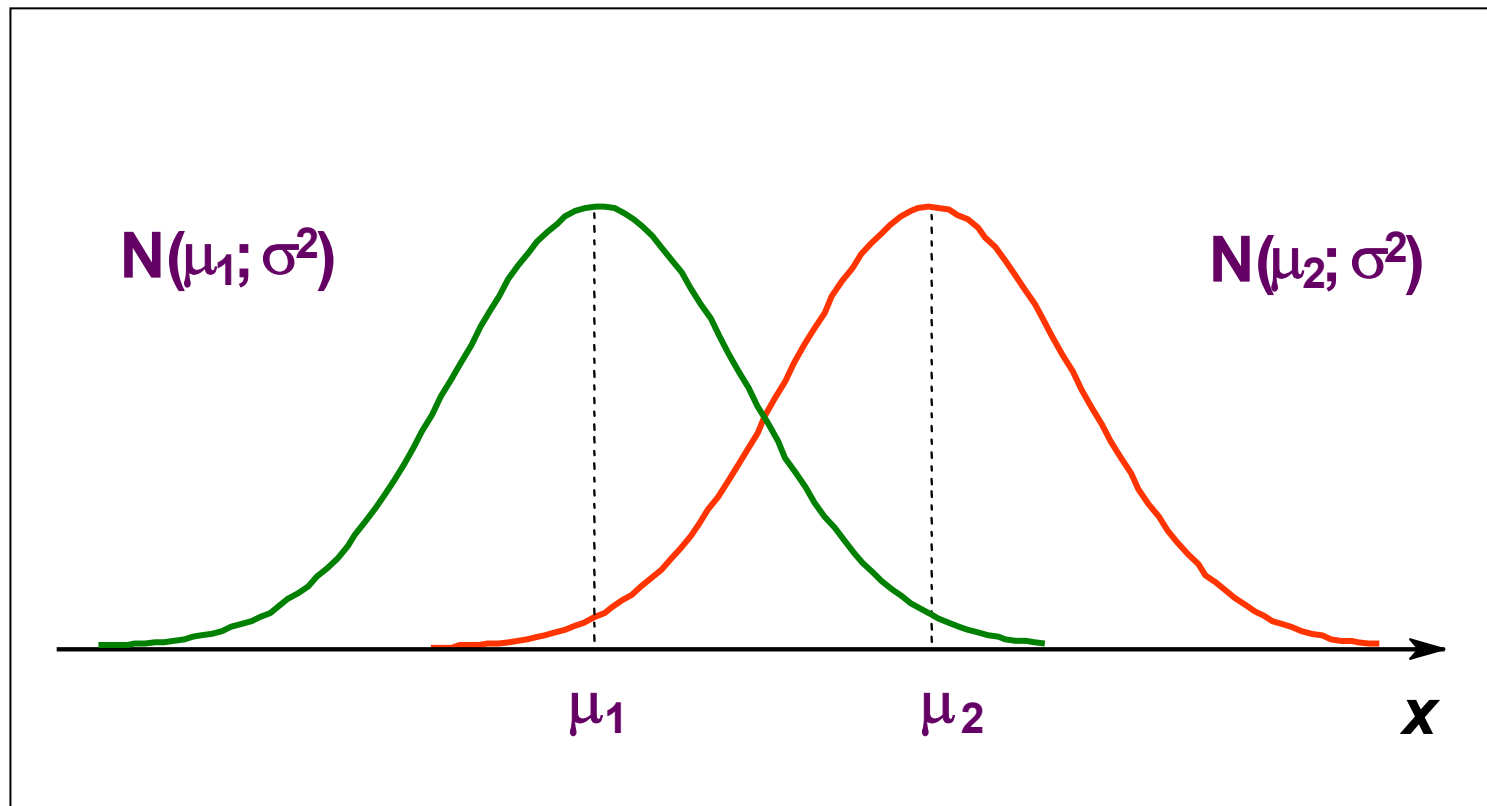
- Por que é importante que as variáveis possam ser descritas por uma distribuição normal?
- **Motivo é simples:** Se as variáveis respeitam uma distribuição normal, pode-se aplicar a grande maioria dos testes e métodos estatísticos conhecidos.
 - **tem-se maior facilidade!**
- Variáveis que não têm distribuição normal podem ser submetidas a transformações (**raiz quadrada, logaritmo**)

Propriedades da Distribuição Normal

- A distribuição é simétrica: **Média = mediana = moda.**
- Os parâmetros μ (média) e σ (desvio padrão) definem completamente uma curva normal. **Notação:** $X \sim N(\mu, \sigma^2)$
- Na distribuição normal com média μ e desvio padrão σ :
 - 68% das observações estão a menos de $\pm\sigma$ da média μ .
 - 95% das observações estão a menos de $\pm 2\sigma$ de μ .
 - 99.5% das observações estão a menos de $\pm 3\sigma$ de μ .

Propriedades da Distribuição Normal

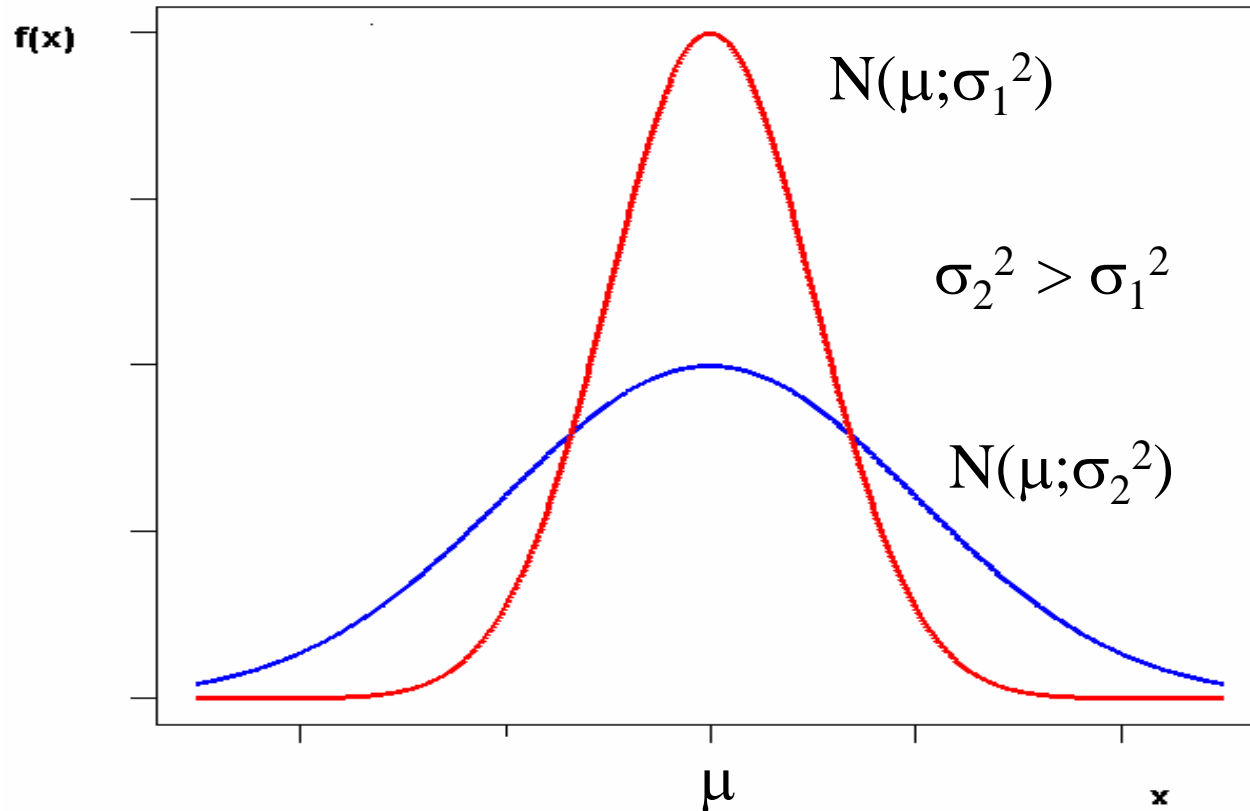
A distribuição Normal depende dos parâmetros μ e σ^2



Curvas Normais com mesma variância σ^2
mas médias diferentes ($\mu_2 > \mu_1$).

Propriedades da Distribuição Normal

Influência de σ^2 na curva Normal



**Curvas Normais com mesma média μ ,
mas com variâncias diferentes ($\sigma_2^2 > \sigma_1^2$).**

Distribuição Normal

- A distribuição normal pode ser descrita pela seguinte “função de densidade”:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \times \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, -\infty < x < +\infty$$

- A área total embaixo da curva normal é igual a 1.
- Quando temos em mãos uma variável aleatória com **distribuição normal**, nosso principal interesse é obter a probabilidade dessa variável aleatória assumir um valor em um determinado intervalo.

Distribuição Normal Padrão

- Caso especial da distribuição Normal: $N(0,1)$.
- Para transformar uma variável de forma que tenha média 0 e desvio padrão 1 (padronização ou normalização), basta fazer o cálculo:

$$Z = \frac{X - \mu}{\sigma}$$

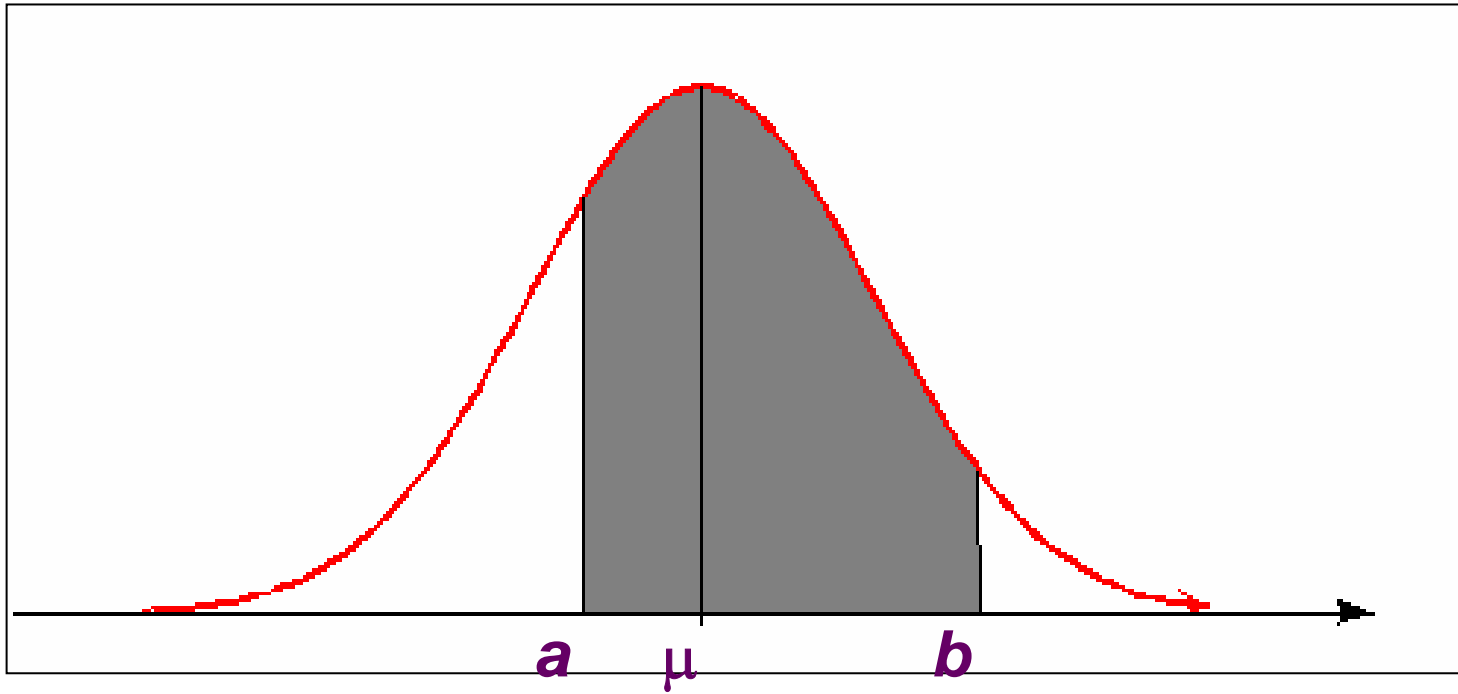
- Propriedade dessa distribuição: Podemos calcular **probabilidades** usando a tabela da distribuição normal padronizada.

Cálculo de probabilidades

$$P(a < X < b)$$



Área sob a curva e acima do eixo horizontal (x) entre a e b .



Usando escores Z para determinar probabilidades:

Se $X \sim N(\mu ; \sigma^2)$, definimos

$$Z = \frac{X - \mu}{\sigma}$$

Portanto,

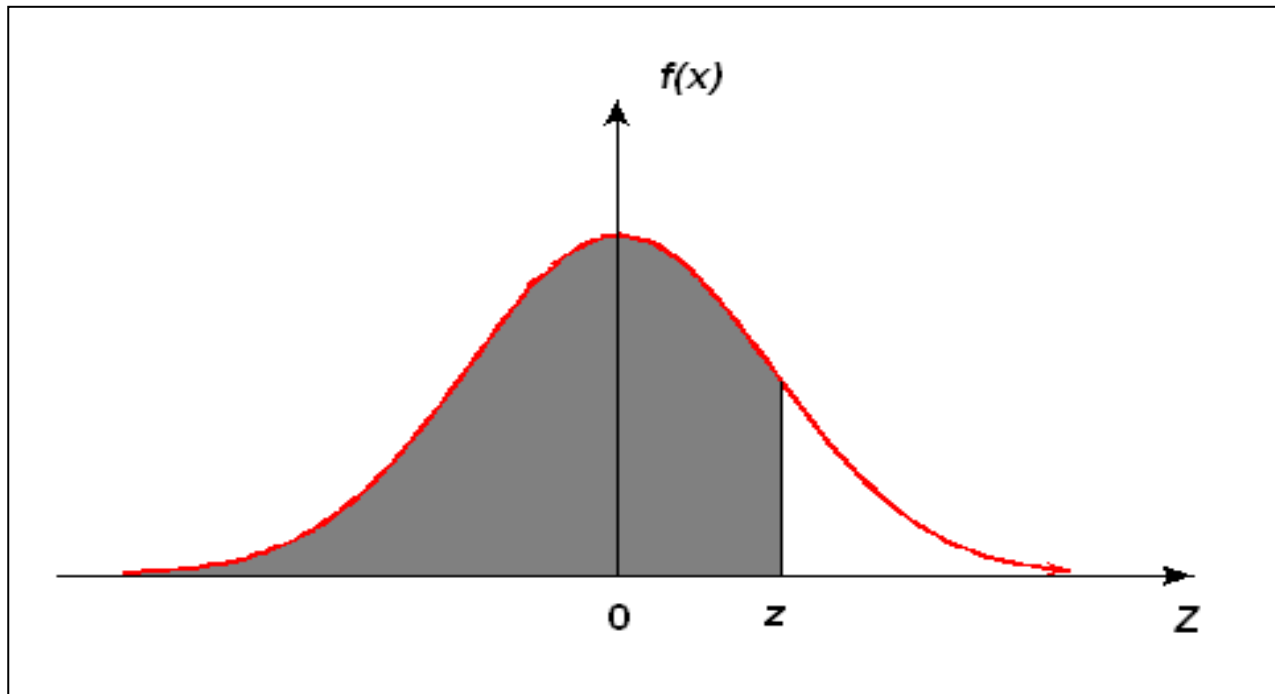
$$P(a < X < b) = P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) = P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right)$$

Exemplo: Seja $X \sim N(10 ; 64)$ ($\mu = 10$, $\sigma^2 = 64$ e $\sigma = 8$).
Calcular $P(6 \leq X \leq 12)$.

$$P(6 \leq X \leq 12) = P\left(\frac{6 - 10}{8} < \frac{X - 10}{8} < \frac{12 - 10}{8}\right) = P(-0,5 < Z < 0,25)$$

Para cálculo dessa probabilidade utilizamos a tabela normal padrão.

USO DA TABELA NORMAL PADRÃO

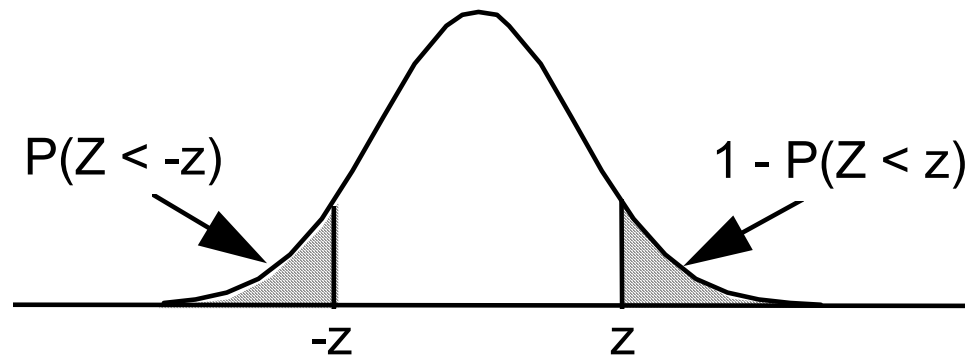


Denotamos : $A(z) = P(Z \leq z)$, para $z \geq 0$.

USO DA TABELA NORMAL PADRÃO

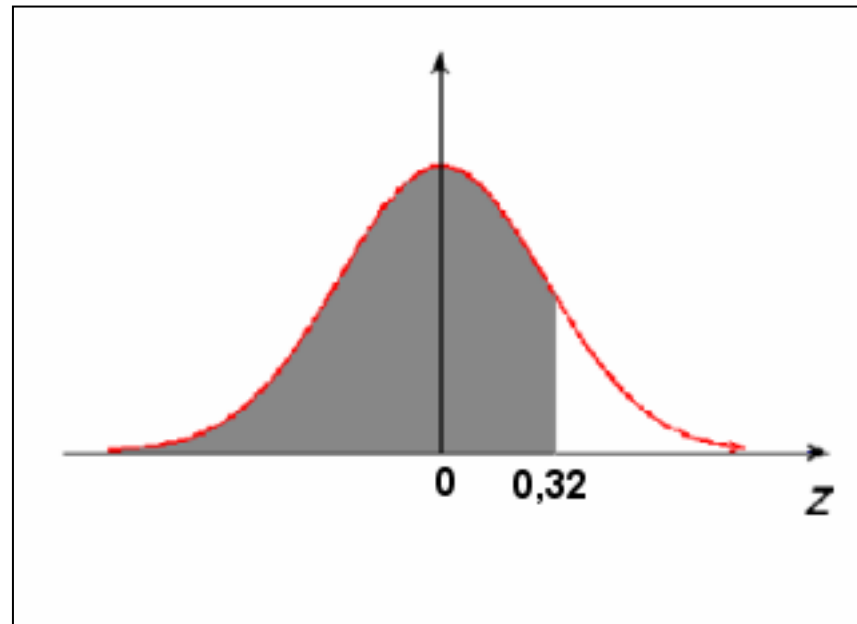
As propriedades que seguem podem ser deduzidas da simetria da densidade em relação à média 0, e são úteis na obtenção de outras áreas não tabuladas.

1. $P(Z > z) = 1 - P(Z < z)$
2. $P(Z < -z) = P(Z > z)$
3. $P(Z > -z) = P(Z < z)$.



Exemplo: Seja $Z \sim N(0; 1)$, calcular

a) $P(Z \leq 0,32)$

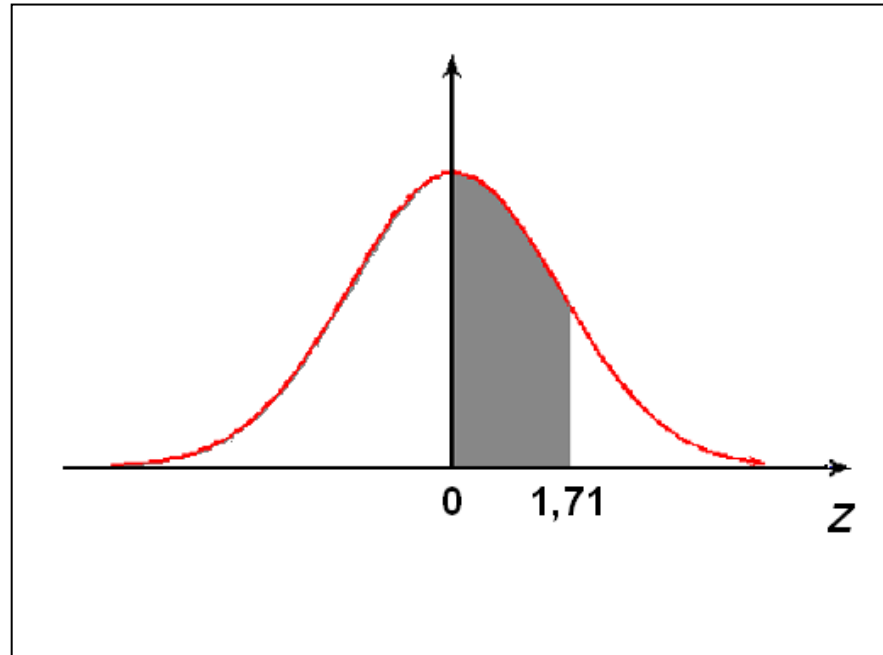


$$P(Z \leq 0,32) = A(0,32) = 0,6255.$$

Encontrando o valor na Tabela N(0;1):

z	0	1	2
0,0	0,5000	0,5039	0,5079
0,1	0,5398	0,5437	0,5477
0,2	0,5792	0,5831	0,5870
0,3	0,6179	0,6217	0,6255
•	•	•	•
•	•	•	•
•	•	•	•

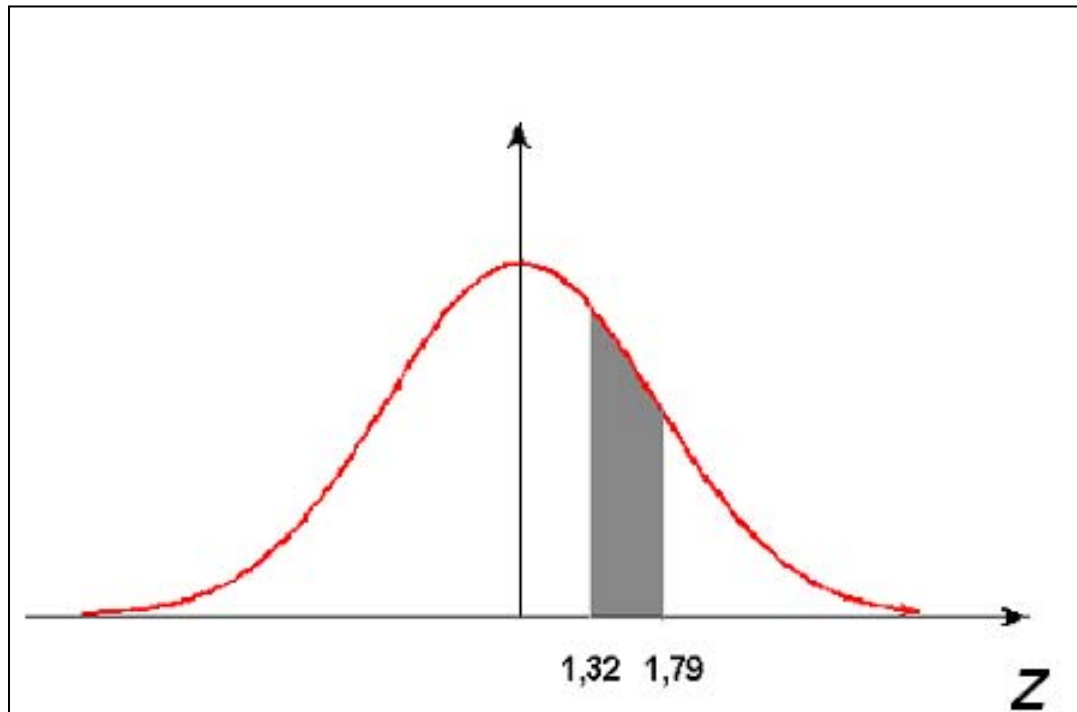
b) $P(0 < Z \leq 1,71)$



$$\begin{aligned} P(0 < Z \leq 1,71) &= P(Z \leq 1,71) - P(Z \leq 0) \\ &= A(1,71) - A(0) \\ &= 0,9564 - 0,5 = 0,4564. \end{aligned}$$

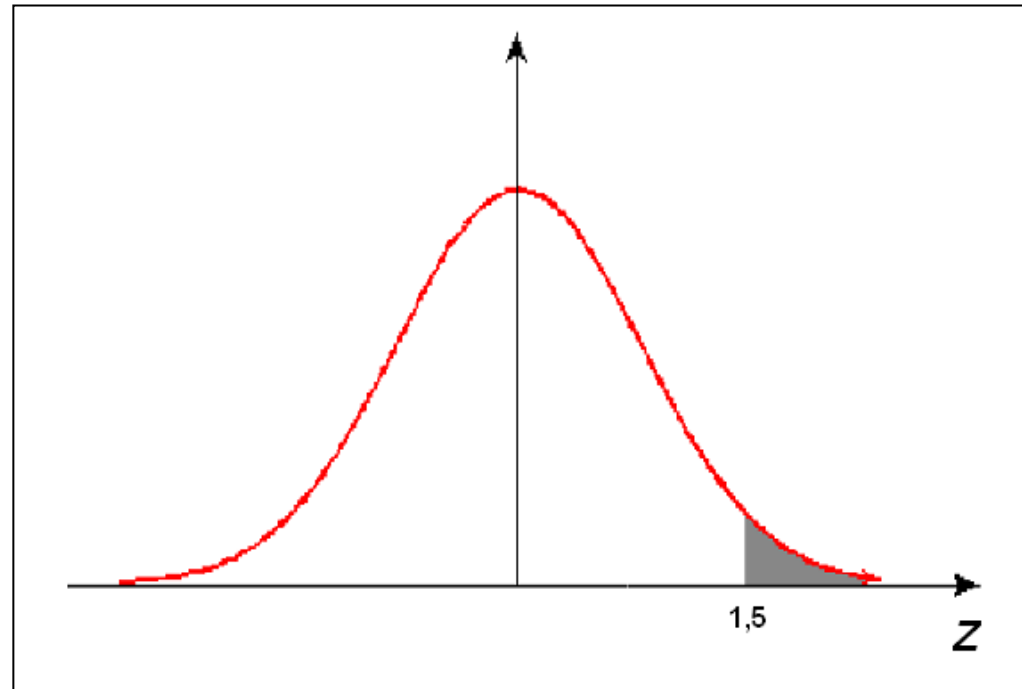
Obs.: $P(Z < 0) = P(Z > 0) = 0,5.$

c) $P(1,32 < Z \leq 1,79)$



$$\begin{aligned} P(1,32 < Z \leq 1,79) &= P(Z \leq 1,79) - P(Z \leq 1,32) = A(1,79) - A(1,32) \\ &= 0,9633 - 0,9066 = 0,0567. \end{aligned}$$

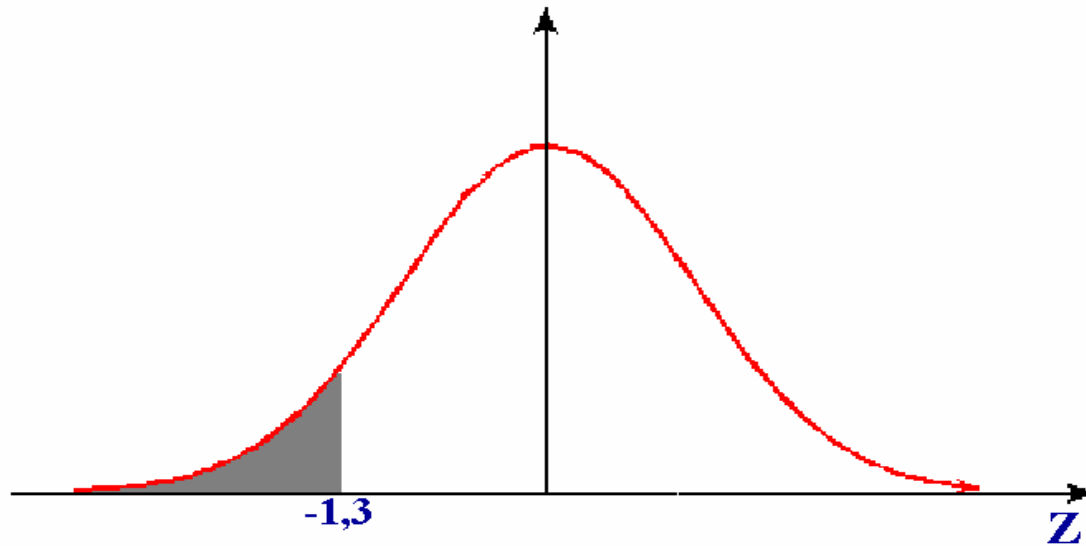
d) $P(Z \geq 1,5)$



$$P(Z > 1,5) = 1 - P(Z \leq 1,5) = 1 - A(1,5)$$

$$= 1 - 0,9332 = 0,0668.$$

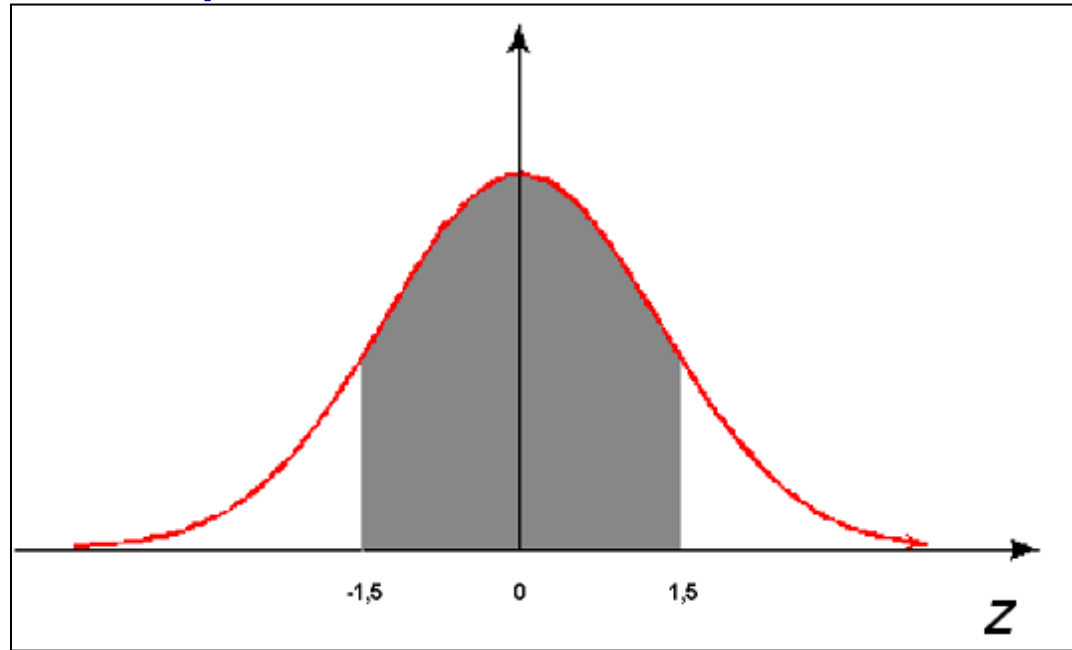
e) $P(Z \leq -1,3)$



$$\begin{aligned} P(Z \leq -1,3) &= P(Z \geq 1,3) = 1 - P(Z \leq 1,3) = 1 - A(1,3) \\ &= 1 - 0,9032 = 0,0968. \end{aligned}$$

Obs.: Pela simetria, $P(Z \leq -1,3) = P(Z \geq 1,3)$.

f) $P(-1,5 \leq Z \leq 1,5)$



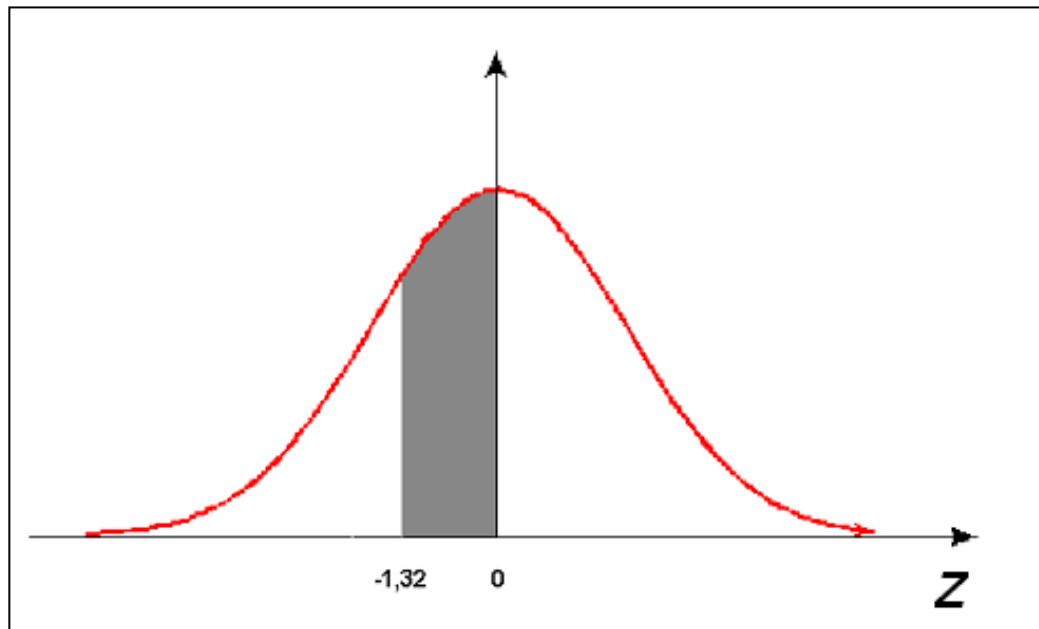
$$P(-1,5 \leq Z \leq 1,5) = P(Z \leq 1,5) - P(Z \leq -1,5)$$

$$= P(Z \leq 1,5) - P(Z \geq 1,5) = P(Z \leq 1,5) - [1 - P(Z \leq 1,5)]$$

$$= 2 \times P(Z \leq 1,5) - 1 = 2 \times A(1,5) - 1$$

$$= 2 \times 0,9332 - 1 = 0,8664.$$

g) $P(-1,32 < Z < 0)$



$$\begin{aligned} P(-1,32 < Z < 0) &= P(0 < Z < 1,32) \\ &= P(Z \leq 1,32) - P(Z \leq 0) = A(1,32) - 0,5 \\ &= 0,9066 - 0,5 = 0,4066. \end{aligned}$$

Distribuição Normal-Exemplo

- $QI \sim N(100, 225)$
 - $Z = (QI - 100) / 15 \sim N(0, 1)$
 - Qual a probabilidade que uma pessoa escolhida aleatoriamente tenha o QI superior a 135?
 $Z = (135 - 100) / 15 = 2,33$
 $P(Z > 2,33) = 0,01$ (tabela normal padrão)
 - Qual a probabilidade que uma pessoa escolhida aleatoriamente tenha o QI inferior a 90?
 $Z = (90 - 100) / 15 = -0,67$
 $P(Z < -0,67) = P(Z > 0,67) = 0,2514$
 - Lembre-se da simetria
 - Probabilidades que uma pessoa escolhida aleatoriamente tenha o QI entre dois valores também podem ser determinadas.

Faixa de Normalidade

- média aritmética \pm desvio-padrão
- corresponde à aproximadamente 68% dos indivíduos da amostra

Exemplo

- Os dados abaixo referem-se aos pesos dos pacientes em dois grupos:

	<i>Grupo A</i>	<i>Grupo B</i>
	78	65
	80	69
	82	78
	85	85
	85	85
	85	93
	86	96
	88	98
<i>Soma</i>	669	669
<i>Média</i>	83,6	83,6
<i>Mediana</i>	85	85
<i>Moda</i>	85	85
<i>N</i>	8	8

Faixa Normalidade: GRUPO A

- Limite inferior = $83,6 - 3,3 = 80,3$
- Limite superior = $83,6 + 3,3 = 86,9$



Faixa Normalidade: GRUPO B

- Limite inferior = $83,6 - 12,2 = 71,4$
- Limite superior = $83,6 + 12,2 = 95,8$

