Business Analytics e Ciência de Dados

Michelle Hanne Soares de Andrade
michelle.andrade@newtonpaiva.br

Conceitos

- **Business Analytics**
- **Business Inteligence**
- **Ciência de Dados**

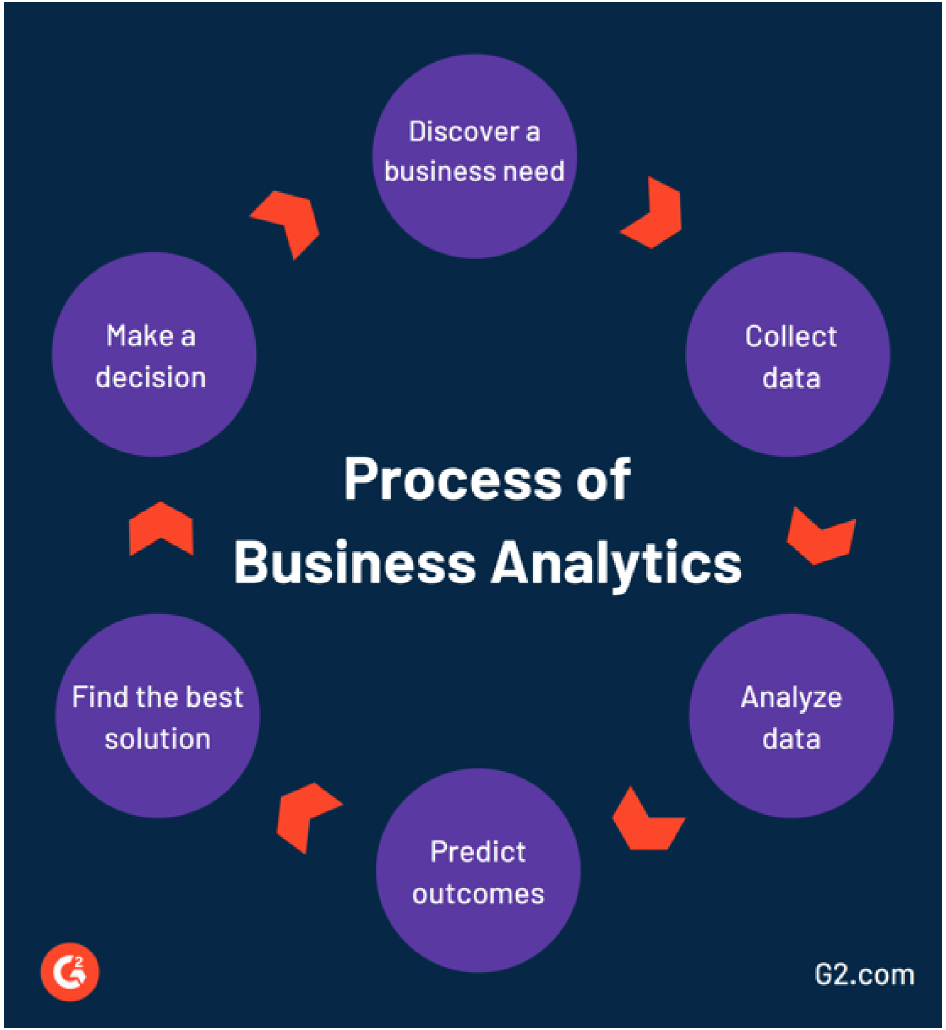
Fonte: https://www.mjvinnovation.com/pt-br/blog/business-analytics-para-negocio/

Business Analytics

Business Analytics é um conceito que descreve a exploração dos dados de uma organização, com ênfase na análise estatística. Envolve a utilização de tecnologias e métodos avançados de análise de informações das mais variadas fontes e em grandes volumes.

Fonte: https://www.mjvinnovation.com/pt-br/blog/business-analytics-para-negocio/

Business Analytics



Fonte: https://learn.g2.com/business-analytics

Métodos de Business Analytics

- **Análise descritiva:** rastreia os principais indicadores de desempenho para entender o estado atual de um negócio; No meio acadêmico está relacionada o que realmente acontece no objeto pesquisado, tomando como referência dados reais.
- **Análise preditiva:** que analisa dados de tendência para avaliar a probabilidade de resultados futuros; No meio acadêmico está relacionado aos algoritmos de *Machine Learning* que geram predições.
- **Análise prescritiva:** que usa o desempenho passado para gerar recomendações sobre como lidar com situações semelhantes no futuro. Implementação de cenários e simulações, prevendo comportamentos.

Fonte: https://www.mjvinnovation.com/pt-br/blog/business-analytics-para-negocio/

Métodos de Business Analytics

- **Ainda temos a Análise Diagnóstica -** usada principalmente para detectar as causas de um certo fenômeno ou comportamento.

Fonte: https://www.mjvinnovation.com/pt-br/blog/business-analytics-para-negocio/

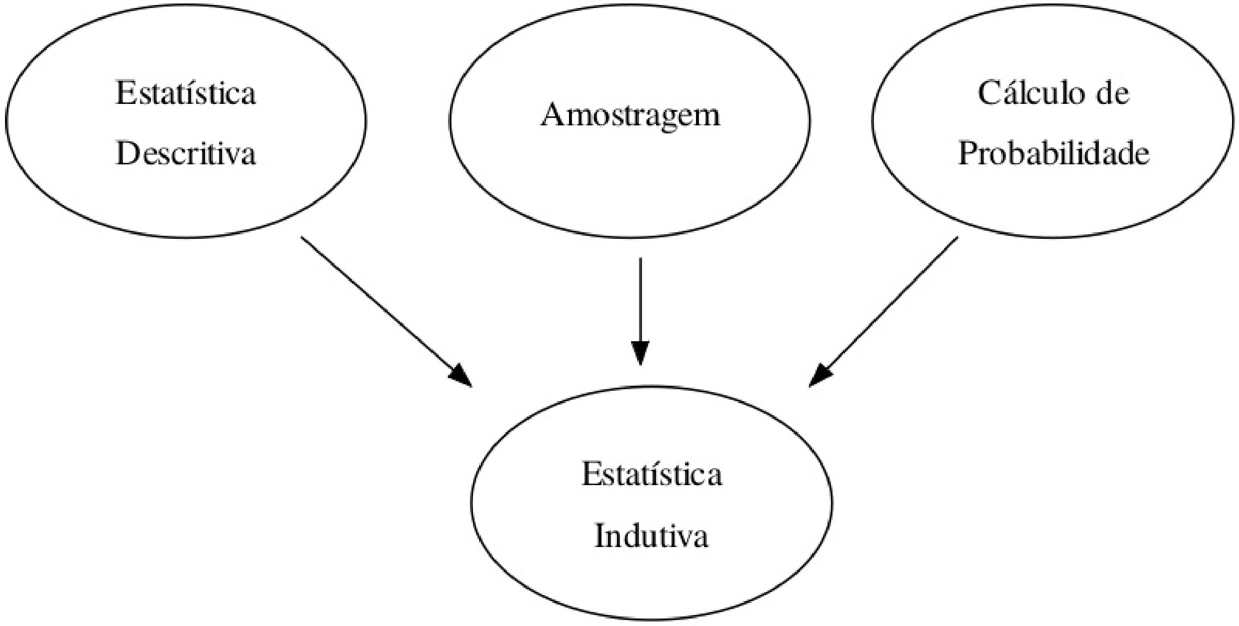
Noções de Métodos Estatísticos

O campo da estatística lida com a **coleta**, **apresentação**, **análise** e **uso dos dados** para **tomar decisões**, resolver problemas e planejar produtos e processos.

Qualquer tomada de **decisão e/ou conclusão** observada que seja resultado de **estudo de dados estatísticos** será tão eficiente quanto o processo utilizado para obtenção desses dados.

Noções de Métodos Estatísticos

- **Estatística Descritiva** e **Estatística Indutiva (ou Inferencial).**



Estatística Descritiva

A **Estatística Descritiva** Resumo ou descrição das características importantes de um conjunto conhecido de dados populacionais... está relacionada com a organização e descrição de dados associada a *cálculos de médias, variâncias, estudo de gráficos, tabelas, etc.*

Estatística Descritiva

- **Com a Estatística Descritiva** entendemos melhor um conjunto de dados através de suas características.
- As três principais características são:
  - Um valor representativo do conjunto de dados. Ex.: uma media, Mediana,
  - Moda, etc...
  - Uma medida de dispersão ou variação.
  - A natureza ou forma da distribuição dos dados: sino, uniforme, assimétrica...

Probabilidades

O Cálculo de Probabilidades é a essência dos modelos Não-Determinísticos e a corroboração de que toda inferência estatística está sujeita a erros.

Tais fenômenos precisam de um modelo matemático diferente para seu estudo. São os conhecidos modelos não-determinísticos ou probabilísticos.

Exemplo de um modelo não determinístico: O lançamento de um dado não viciado, sempre sob as mesmas condições. Não é possível afirmar com exatidão qual será a face observada a cada novo lançamento.

Amostragem

- A **Amostragem** é o ponto de partida para todo um Estudo Estatístico. Aqui pode se ter origem um problema bastante comum em Engenharia: “Analise profunda sobre dados superficiais!”
- Três métodos importantes para a coleta de dados:
  - Um estudo **retrospectivo** usando dados históricos
  - Um estudo de **observação**
  - Um **experimento planejado**

Estatística Indutiva (Inferêncial)

A **Estatística Indutiva** é o objetivo básico da ciência. A ela esta associada, *Estimação de Parâmetros, Testes de Hipóteses, Modelamento, etc.. Generalização sobre uma população tomadas a partir da utilização de dados amostrais.*

Estatística Indutiva - Exemplo

- **Variabilidade** no consumo de combustível de um veículo. Potenciais fontes de variabilidade do sistema:
  - Tipo de condições de estrada (cidade ou rodovia)
  - Mudanças nas condições do veículo ao longo do tempo (pressão dos pneus, compressão, desgaste de válvulas)
  - Marca ou categoria de gasolina utilizada
  - Condições meteorológicas
  - Modo de dirigir
  - Outros fatores
- **Estatística: fornece uma teoria para analisar estas fontes de variabilidade (quais as mais importantes?) de forma quantitativa.**

Organização de Dados Estatísticos

Planejamento de um estudo Estatístico

1- Identifique a variável (variáveis) de interesse (foco) e a população do estudo.

2- Desenvolva um plano detalhado para a coleta de dados. Se usar uma amostra, tenha certeza de que a amostra representa a população.

3- Colete adequadamente os dados.

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, indícios ou proporcionalização.

Planejamento de um estudo Estatístico

4- Apuração dos Dados: Resumo dos dados através de sua contagem e agrupamento. É a condensação e tabulação de dados.

5 - 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.

6 - 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).

Principais Conceitos em Estatística

- **População:**
  - E todo o conjunto de elementos que possuam ao menos uma característica comum observável.
- **Amostra:**
  - E uma parte da população que será avaliada por um critério comum.
- **Dados estatísticos:**
  - São os valores associados as variáveis de pesquisas.
- **Frequências:**
  - O numero de vezes em que a variável ocorre e chamado **frequência absoluta** e é indicado por n_i.
  - Definimos **frequência relativa** (f_i) como a razão entre a frequência absoluta (n_i) e o numero total de observações (n) , ou seja:

$$f_i = n_i/n$$

Principais Conceitos em Estatística

- **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 CEFET 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.

Principais Conceitos em Estatística

- **Variável Qualitativa:**
  - Quando seu valores são expressos 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**
    - **Variável contínua**

Principais Conceitos em Estatística

- **Variável Discreta ou Descontínua**: Seus valores são expressos geralmente através de números inteiros não negativos. Resulta normalmente de contagens.
    - *Ex: Nº de alunos presentes às aulas de introdução à estatística econômica no 1º semestre de 1997: mar = 18 , abr = 30 , mai = 35 , jun = 36.*
- **Variável Contínua**: Resulta normalmente de uma **mensuração**, e a **escala numérica de seus possíveis valores corresponde ao conjunto R** dos números Reais, ou seja, podem assumir, teoricamente, qualquer valor entre dois limites.
    - *Ex.: 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.*

Exemplos de Variáveis

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

Revisão Geral Estatística I

Resumo – População e Amostra

- **População e Amostra:** Ao examinar um grupo qualquer, considerando todos os seus elementos, estamos tratando da **população** ou **universo**. Nem sempre isso e possível. Nesse caso, examinamos uma pequena parte chamada **amostra**.
- Uma **população** pode ser finita ou infinita. Por exemplo:
  - a população dos alunos de sua escola é finita e a população constituída de todos os resultados (cara ou coroa) em sucessivos lances de uma moeda e infinita.
- Se uma **amostra** é representativa de uma população, podemos obter conclusões importantes sobre a população.

Medidas de Posição

- Indicam alguma tendência central ou comportamento esperado com respeito extraídos de uma amostra qualquer.
- Dentre as principais medidas, destacam-se: **media, mediana e moda.**

Medidas de Posição

- Sejam $x_1, x_2, \dots x_n$ n observações de um fenômeno aleatório qualquer. Denominamos media aritmética da amostra.

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

Média Aritmética

- Pode ser interpretada como o centro de massa (baricentro das observações).

- Caso os dados estejam organizados em uma tabela de frequências, isto é, se para cada xi foi verificado um número fi relativo a sua frequência, então

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}.$$

Exemplo

Uma pesquisa avaliou a idade dos 25 alunos de uma turma de engenharia. Os dados foram organizados na seguinte tabela de frequências:

Idade	**freq.**
19	3
20	5
21	1
22	8
23	4
24	1
25	0
26	3
Total	25

A média aritmética pode ser obtida diretamente do dispositivo (1):

$$\overline{x} = \frac{\sum_{i=1}^{8} x_i f_i}{\sum_{i=1}^{8} f_i} = \frac{548}{25} = 21,92.$$

Mediana - dados discretos ou não agrupados

- Caso os dados de interesse sejam discretos ou, sejam contínuos e ainda não tenham passado por um agrupamento, então o calculo da mediana é mais simples.
- Basta observar a paridade do número n de observações:
  - **se n e ímpar, então a mediana e exatamente o valor central das observações.**

$$x_{\left(\frac{n+1}{2}\right)};$$

  - caso contrario, a mediana e dada pela media aritmética dos dois valores centrais

$$\frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}.$$

Mediana - dados discretos ou não agrupados

- Exemplos:
  - No exemplo dos salários, temos a seguinte ordenação: R$1000,00, R$1800,00, R$2500,00, R$2500,00, R$20000. Como o número de observações e ímpar, temos que a mediana e dada pelo elemento **central x(3) = 2500.**
  - No exemplo das classes sociais, temos a seguinte ordenação: B, B, B, B, B, M, M, M, M, M, M, M, M, M, M, A, A, A, A, A, A, A, A, A, A,A, A, A, A, A. Como o numero de observações é par, segue que a mediana é dada pelo ponto médio de x(15) e x(16). A mediana então ficaria entre os valores M e A.

Moda (Mo)

- Moda de dados agrupados sem intervalo de classe:

Nº DE MENINOS	f_i
0	2
1	6
2	10
3	12
4	4
	$\Sigma = 34$

Mo = 3,

(pois 3 tem frequência 12)

Moda (Mo)

- Moda de dados agrupados com intervalo de classe:
- A classe de maior frequência é chamada classe modal. A moda será, então, o ponto médio desta classe.

$$\boxed{Mo = \frac{\ell^* + L^*}{2}}$$

i	ESTATURAS (cm)	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11 ←
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\Sigma = 40$

$$Mo = \frac{\ell^* + L^*}{2}$$

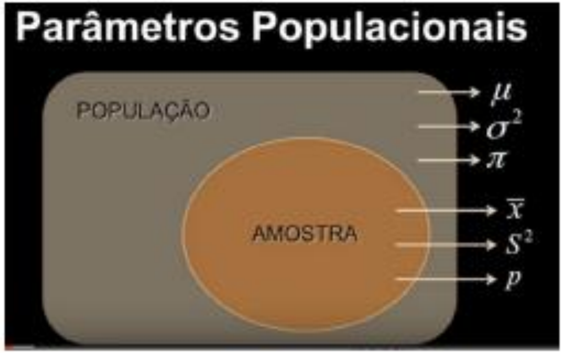
$$Mo = \frac{158 + 162}{2} = \frac{320}{2} = 160$$

$$Mo = 160 \text{ cm}$$

Resumo

Parâmetro (População)	
Valor médio	μ
Desvio padrão	σ
Proporção	p
Correlação	ρ

Estatística (Amostra)	
Média	$\bar{x}$
Desvio padrão	s
Proporção	$\hat{p}$
Correlação	r



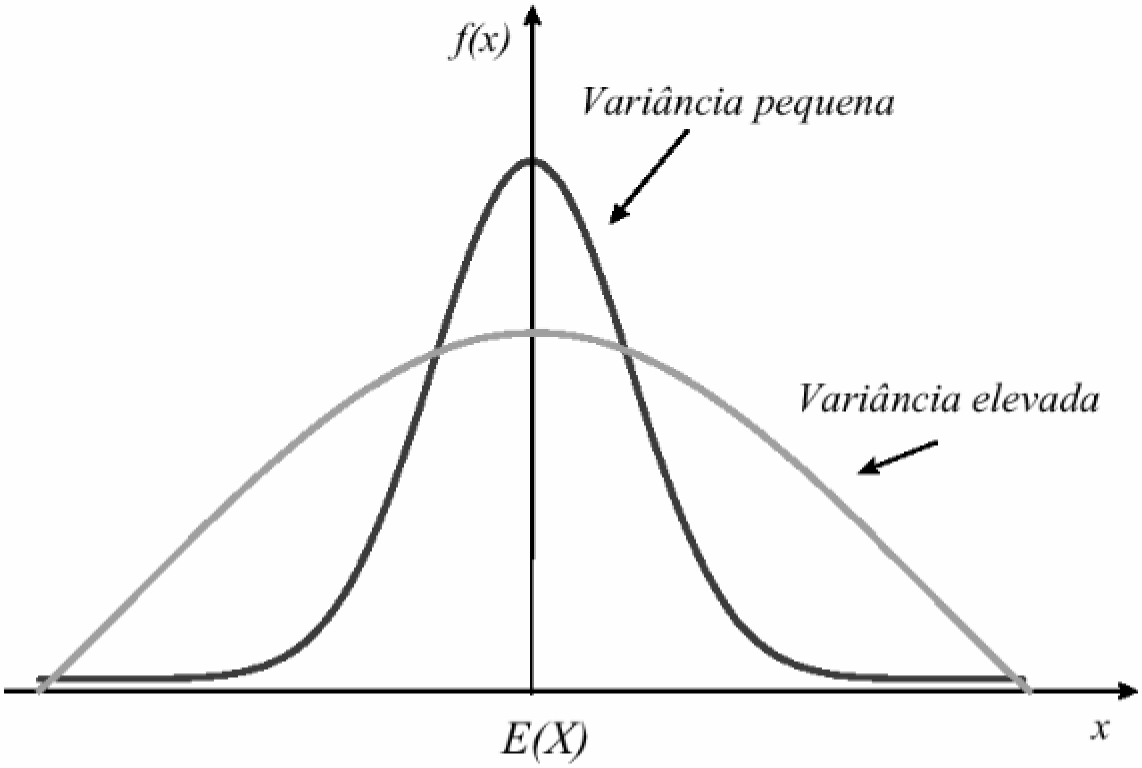
Resumo - Desvio Padrão (S) x Variância (S^2)

- **Desvio Padrão (S) x Variância (S^2)**
  - O desvio padrão é a medida mais usada na comparação de diferenças entre conjuntos de dados, por ter grande precisão. O desvio padrão determina a dispersão dos valores em relação à média e é calculado por meio da raiz quadrada da variância.

Resumo - Variância (S^2)

- **Variância (S^2)**
  - Sendo a variância calculada a partir dos quadrados dos desvios, ela é um número em unidade quadrada em relação a variável em questão, o que, sob o ponto de vista prático é um inconveniente; por isso, tem pouca utilidade na estatística descritiva, **mas é extremamente importante na inferência estatística e em combinações de amostras.**

Resumo – Variância (S^2)



Desvio Padrão

- O conceito de variância é bastante rico, contudo, deve ser utilizado com cautela já que trata do problema original em escala quadrática.
- O desvio padrão surge como uma alternativa para corrigir este detalhe e assim facilitar a análise dos resultados.
- Tal medida e dada pela raiz quadrada da variância amostral:

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \overline{x}^2}.$$

Desvio Padrão

Em geral, a variância amostral possui propriedades matemáticas melhores, enquanto o desvio padrão oferece interpretações mais razoáveis.

Desvio Padrão

Desvio Padrão (S)

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

x_i são os valores da variável

média

Exemplo:

Calcular o desvio padrão

da população

representada por:

{ - 4 , -3 , -2 , 3 , 5}

Xi	$\bar{X}$	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
- 4	- 0,2	- 3,8	14,44
- 3	- 0,2	- 2,8	7,84
- 2	- 0,2	- 1,8	3,24
3	- 0,2	3,2	10,24
5	- 0,2	5,2	27,04
		Σ =	62,8

Sabemos que n = 5 e 62,8 / 5 = 12,56.

A raiz quadrada de 12,56 é o desvio padrão = **3,54**

Resumo – Desvio Padrão

- **O desvio padrão é uma medida que só pode assumir valores não negativos e quanto maior for, maior será a dispersão dos dados.**
- Algumas propriedades do desvio padrão, que resultam imediatamente da definição, são:
  - O desvio padrão é sempre não negativo e será tanto maior, quanta maior a variabilidade entre os dados.
- **Se S = 0, então não existe variabilidade, isto é, os dados são todos iguais.**

Resumo – Exercício

- Tendo por base uma amostra da altura de uma parcela da população apresentada na Tabela 5.2, determinar:

  a) A variância das alturas;

  b) O desvio-padrão das alturas.

Tabela 5.2 – Estatura de uma amostra de uma população A

Altura (cm)	Nº de pessoas
150 |— 158	5
158 |— 166	18
166 |— 174	42
174 |— 182	27
182 |— 190	8
Σ	100

Fonte: Dados fictícios, apenas para fins ilustrativos.

Resumo – Exercício - Solução

Solução

a) A variância das alturas

Usando a fórmula 5.13 obtém-se o seguinte resultado:

$$s^2 = \frac{\sum X_i^2 f_i - \frac{\left(\sum X_i f_i\right)^2}{n}}{n-1}$$

Para calcular a variância, necessita-se conhecer as informações a seguir, cujos valores estão calculados na Tabela 5.3:

$$\sum X_i f_i$$

$$\sum X_i^2 f_i$$

Resumo – Exercício - Solução

Solução

a) A variância das alturas

Usando a fórmula 5.13 obtém-se o seguinte resultado:

$$s^2 = \frac{\sum X_i^2 f_i - \frac{\left(\sum X_i f_i\right)^2}{n}}{n-1}$$

Para calcular a variância, necessita-se conhecer as informações a seguir, cujos valores estão calculados na Tabela 5.3:

$\sum X_i f_i$

$\sum X_i^2 f_i$

Tabela 5.3 – Tabela auxiliar da Tabela 5.2

Altura (cm)	Nº de pessoas	X_i	$X_i f_i$	$X_i^2 f_i$
150 |— 158	5	154	770	118.580
158 |— 166	18	162	2916	472.392
166 |— 174	42	170	7140	1.213.800
174 |— 182	27	178	4806	855.468
182 |— 190	8	186	1.488	276.768
$\sum$	100		17.120	2.937.008

Fonte: Dados fictícios, apenas para fins ilustrativos.

Resumo – Exercício - Solução

Substituindo os valores na fórmula 5.13 obtém-se os seguinte resultados:

$$s^2 = \frac{2937008 - \frac{17120^2}{100}}{100 - 1}$$

$$s^2 = \frac{2937008 - 2930944}{100 - 1} = \frac{6064}{99}$$

$s^2 = 61{,}25$

Resposta:

A variância das alturas é de 61,25 cm²

b) O desvio-padrão das alturas

Solução

Extrai-se a raiz quadrada da variância.

Assim,

$$s = \sqrt{s^2}$$

$$s = \sqrt{61{,}25}$$

s = 7,8263

Resposta: As estaturas das pessoas estão dispersas em média 7,83 cm em relação à média da distribuição.

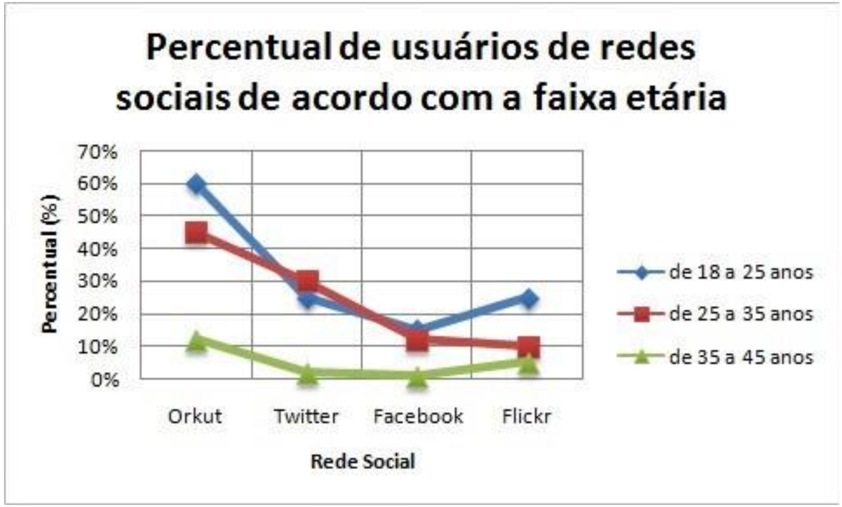
Representação Gráfica de uma Distribuição

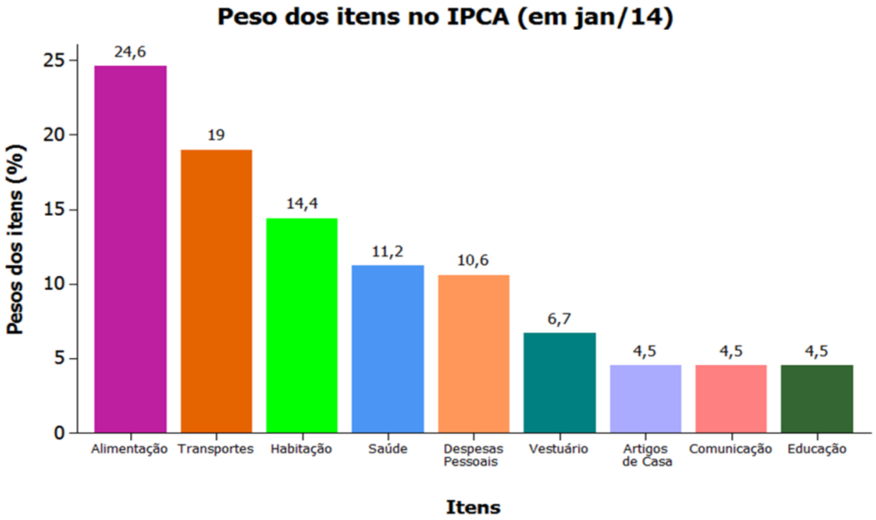
Gráficos

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

Gráficos

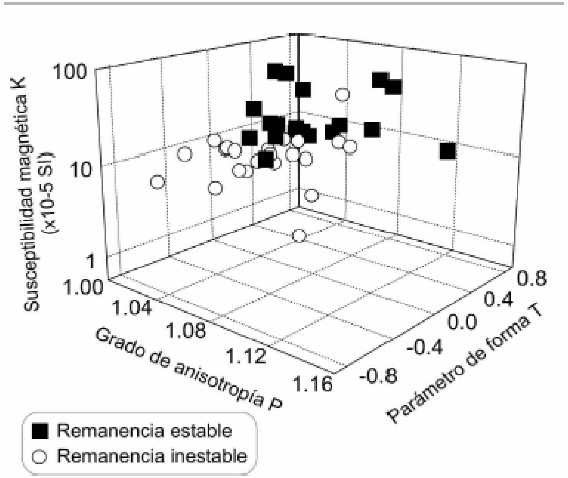
- **Diagramas:** São gráficos geométricos dispostos em duas dimensões. São os mais usados na representação de séries estatísticas. Eles podem ser: Barras, Linhas, Colunas e Setores



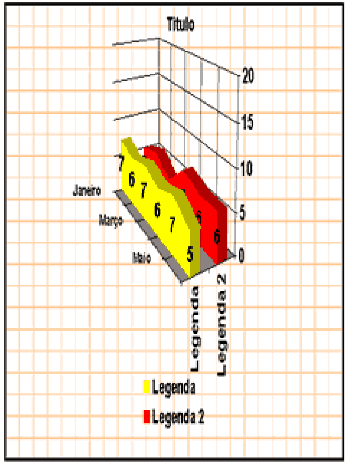


Gráficos

- **<u>Estereogramas</u>:** 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.



	Legenda	Legenda 2
Janeiro	7	5
Fevereiro	6	6
Março	7	4
Abril	6	7
Maio	7	6
Junho	5	6
Julho	6	5



Gráficos

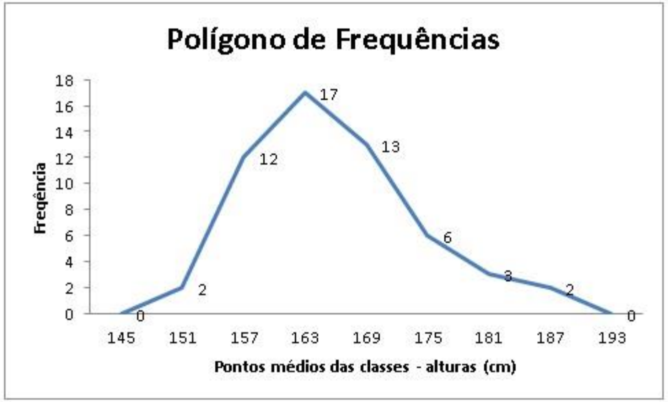
- **<u>Pictogramas</u>:** São construídos a partir de figuras representativas da intensidade do fenômeno. Este tipo de gráfico tem a vantagem de despertar a atenção do público leigo, pois sua forma é atraente e sugestiva. Os símbolos devem ser autoexplicativos. A desvantagem dos pictogramas é que apenas mostram uma visão geral do fenômeno, e não de detalhes minuciosos. Veja o exemplo abaixo:

Gráficos

- **<u>Cartogramas:</u>** 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.

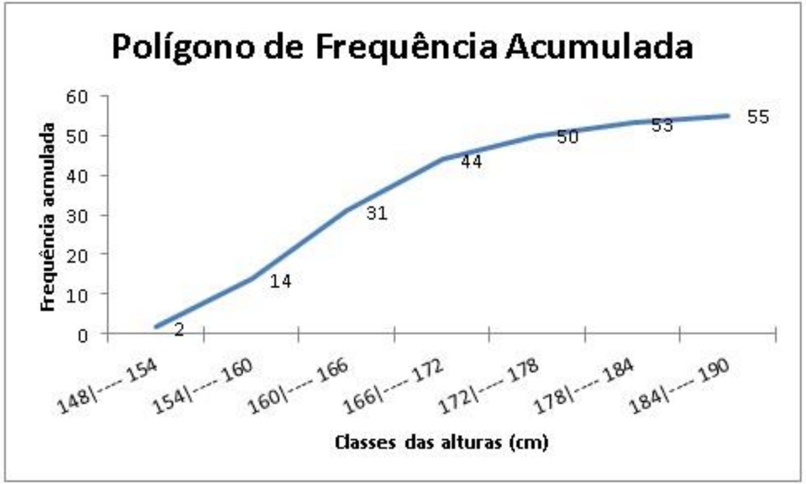
Gráficos

- **Polígono de frequência:** é um gráfico em linha, sendo as frequências 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.



Gráficos

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



Box-Plot (Diagrama de caixa)

- O **boxplot**, ou diagrama de caixa, é um gráfico que capta importantes aspectos de um conjunto de dados através do seu resumo dos cinco números (valor mínimo, primeiro quartil, segundo quartil, terceiro quartil e valor máximo). Bem como, o centro, dispersão, desvio da simetria e identificação das observações que estão longe do centro dos dados (*outliers*).
- **O gráfico e formado por uma caixa construída paralelamente ao eixo da escala dos dados (pode ser horizontal ou vertical).**
- **Esse *box* vai desde o primeiro quartil até o terceiro quartil e nela traça-se uma linha na posição da mediana.**

Exemplo 1 Box-Plot

Ordem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Valor	3,0	3,5	4,5	5,0	5,0	5,5	6,5	6,5	6,5	7,5	7,6	7,9	8,0	8,0	9,0	9,5	10,0	15,0

A mediana divide o conjunto em duas partes, cada uma com 9 observações. A mediana será, então, a média dos dois valores centrais:

$$Q2 = \frac{6{,}5 + 7{,}5}{2} = 7{,}0$$

Ordem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Valor	3,0	3,5	4,5	5,0	5,0	5,5	6,5	6,5	6,5	7,5	7,6	7,9	8,0	8,0	9,0	9,5	10,0	15,0

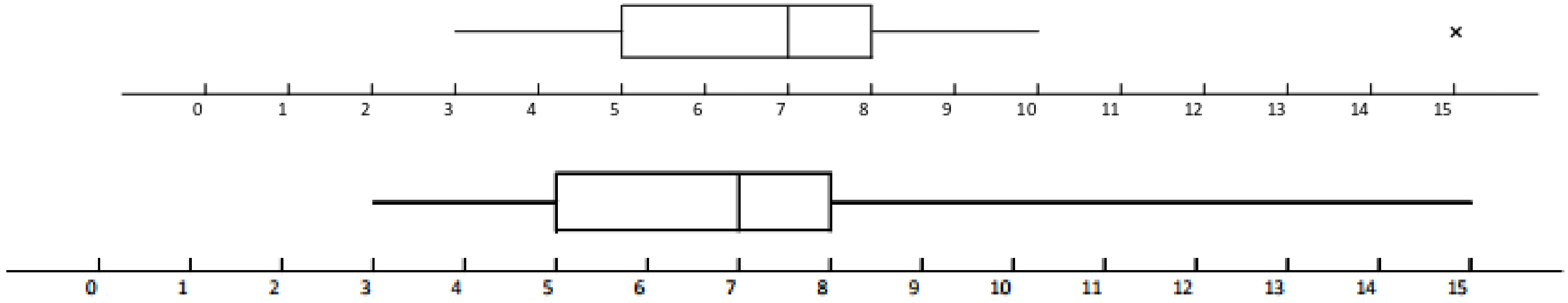
Exemplo 1 Box-Plot

- **O cálculo do primeiro e do terceiro quartis:**
  - Calcular as medianas das duas metades – o primeiro quartil é a mediana da metade inferior e o terceiro quartil é a mediana da metade superior.

$$Q1 = 5,0$$

$$Q3 = 8,0$$

Ordem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Valor	3,0	3,5	4,5	5,0	5,0	5,5	6,5	6,5	6,5	7,5	7,6	7,9	8,0	8,0	9,0	9,5	10,0	15,0



Exemplo 2 Box-Plot

Ordem	1	2	3	4	5	6	7	8	9	10	11	12	13
Valor	15	17	18	19	19	20	25,0	26	26	28	30	32	42

Ordem	1	2	3	4	5	6	7	8	9	10	11	12	13
Valor	15	17	18	19	19	20	25,0	26	26	28	30	32	42

$$Q2 = 25{,}0$$

$$Q1 = \frac{18 + 19}{2} = 18{,}5$$

$$Q3 = \frac{28 + 30}{2} = 29$$

Exemplo 3 - Dispositivo Prático

- Suponha o conjunto de números 1, 2, 3, 4 e 5.
- Organize os dados em ordem numérica (Um gráfico de frequência cumulativa facilita o trabalho, mas não é essencial).

1 - Encontre a mediana**. O número, no centro de dados → 3**

2 - Encontre os quartis superiores e inferiores, encontrando as medianas dos números maiores do que a mediana (quartil superior) e os números inferiores a mediana (quartil inferior).

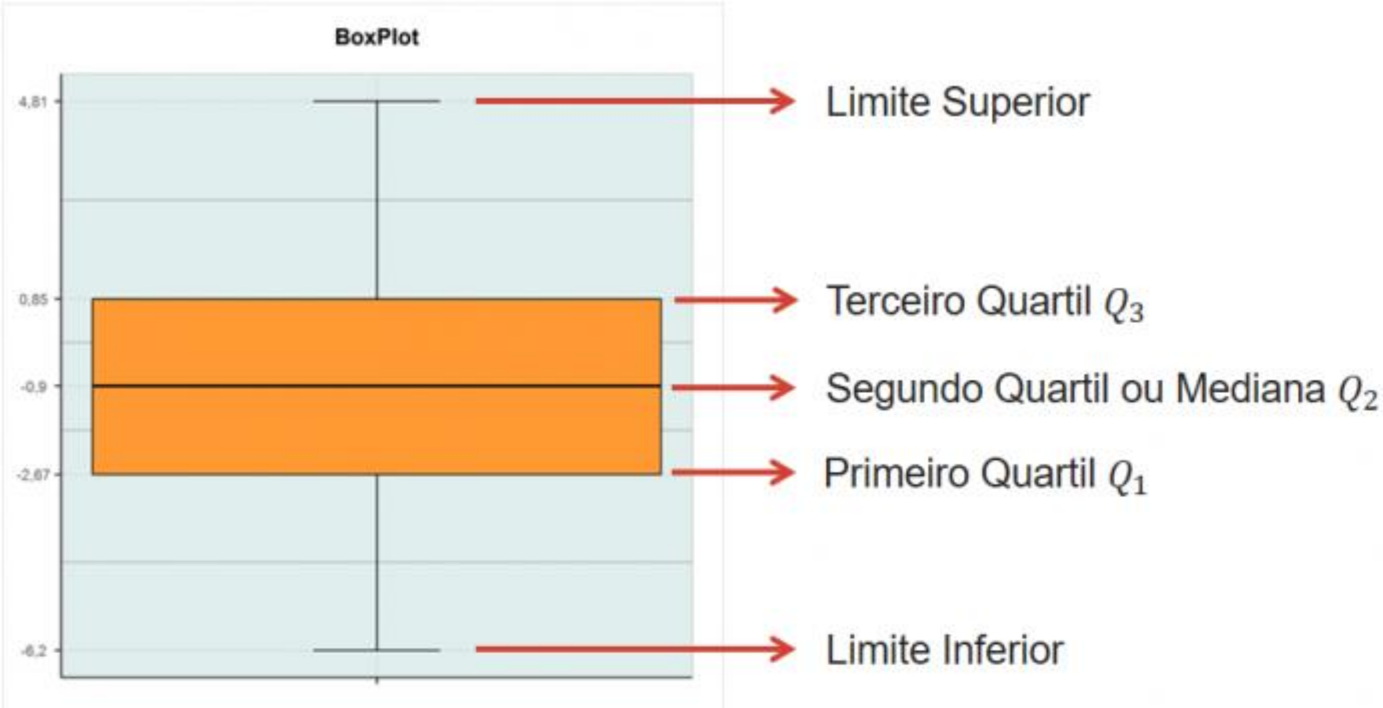
3 - Marque seus valores discrepantes, ou os extremos. Estes são os pedaços maiores e menores de dados e devem ser marcados com um ponto (ou uma pequena linha vertical) praticamente em direção ao centro da sua caixa. Neste caso, o extremo inferior é 1 e o extremo superior é 5.

Exemplo 3 - Dispositivo Prático

- Desenhe uma linha guia. Esta deve ser longa o suficiente para conter todos os seus dados. Os números devem ser colocados no gráfico em intervalos regulares.
- Marque as suas medianas e quartis. Desenhe uma linha a partir desses pontos na altura que você deseja que sua caixa tenha. Conecte os topos para fazer a caixa.

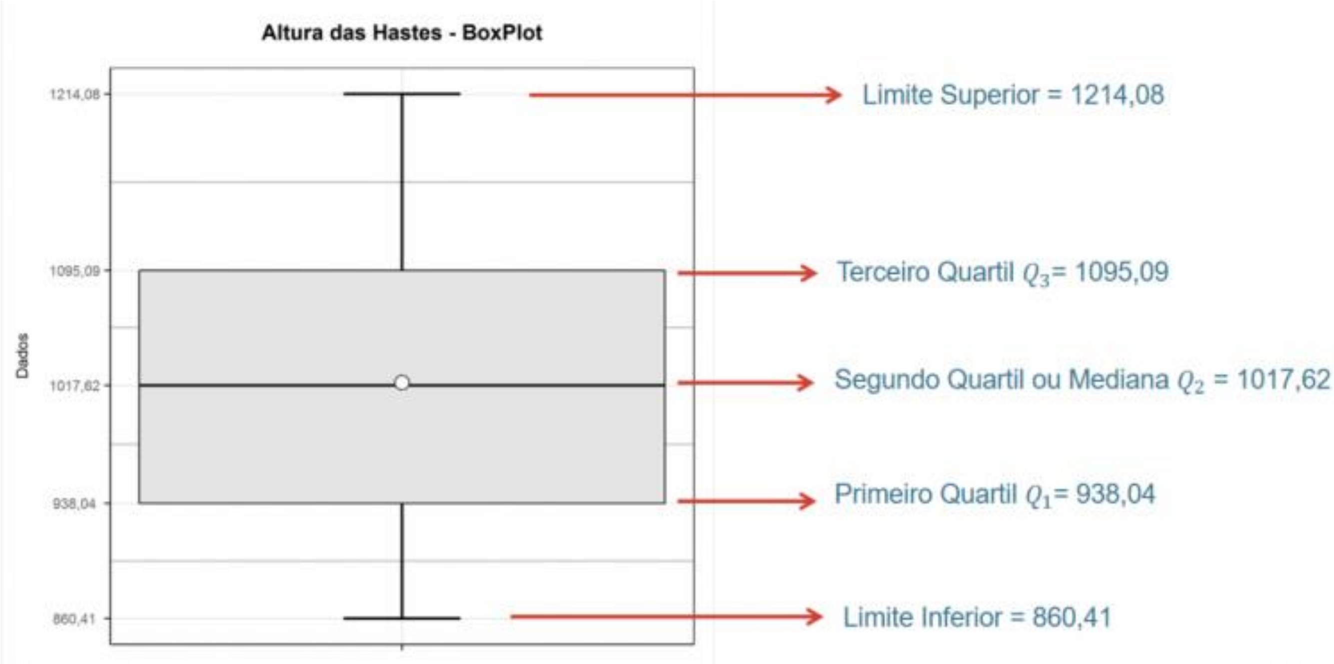
Box-Plot

- Exemplo



Box-Plot – Exemplo 1

- Na Tabela a seguir temos as medidas da altura de 20 hastes. Faça o Boxplot correspondente.



Dados da usinagem			
903,88	1036,92	1098,04	1011,26
1020,70	915,38	1014,53	1097,79
934,52	1214,08	993,45	1120,19
860,41	1039,19	950,38	941,83
936,78	1086,98	1144,94	1066,12

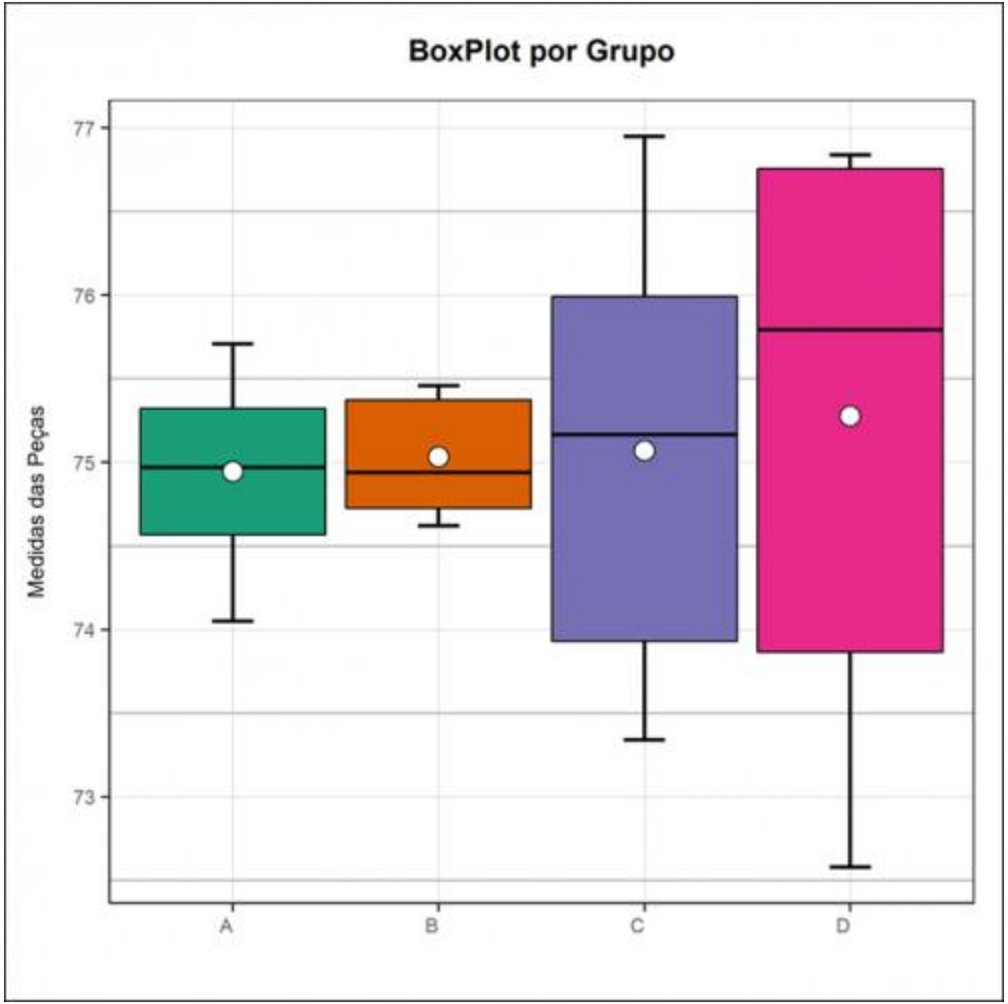
Box-Plot Exemplo 2

- Uma indústria produz uma peça automotiva cujo valor de referência é 75cm. Após verificar lotes com peças fora de especificação, enviaram duas equipes de trabalhadores (A e B) para um treinamento.
- Para verificar a eficiência do treinamento, foram selecionadas 10 peças produzidas pelas equipes A e B e 10 peças produzidas pelas equipes C e D que não participaram do treinamento

A		B		C		D	
75,27	74,93	74,94	74,75	75,93	73,34	75,98	76,75
75,33	74,72	75,25	74,65	76,95	74,04	75,61	76,78
74,58	74,53	75,44	74,94	75,47	75	74,2	74,74
75,01	75,32	74,62	74,92	73,6	76,18	76,44	72,58
75,71	74,05	75,35	75,46	74,85	75,33	76,84	72,86

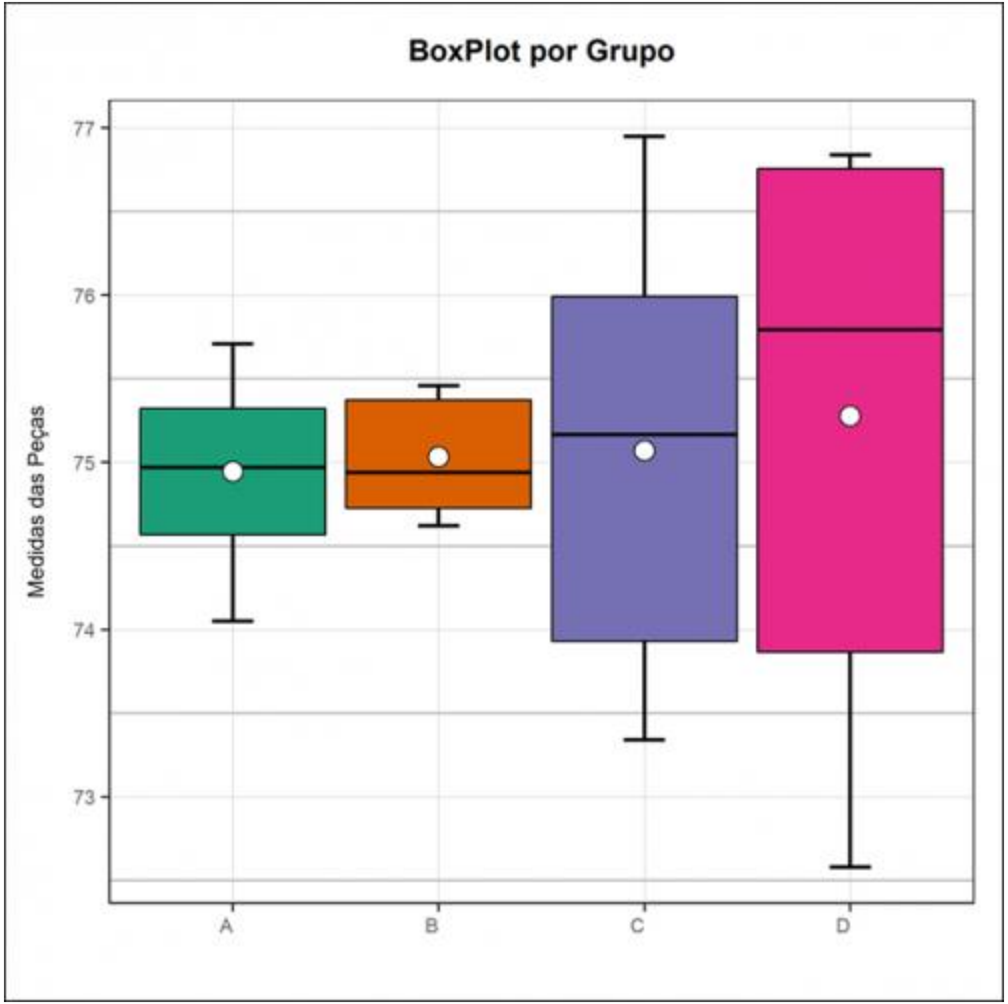
Box-Plot Exemplo 2

- Analisando o gráfico podemos observar que:
  - As equipes A e B produzem peças com menor variabilidade, indicando que o treinamento teve o efeito desejado;
  - A equipe D é a que produz peças com maior variabilidade;
  - A equipe B é a que produz peças com menor variabilidade.



Box-Plot Exemplo 2

- **Considerações:** Como as peças das equipes A e B tem menor variabilidade e com valor médio próximo do valor de referência, vale a pena enviar as demais equipes para o treinamento.



Histograma

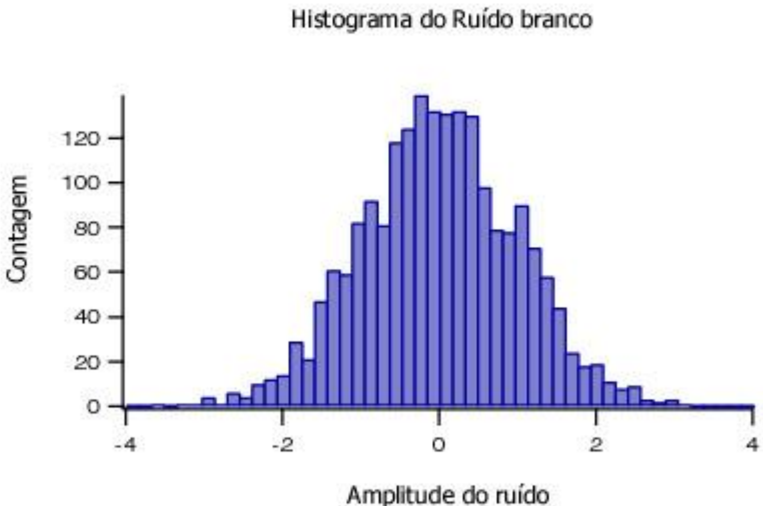
- Organizar os dados coletados em ordem crescente;
- Determinar a amplitude total;
- Dividir a amplitude total em um nº adequado de intervalos de preferência com a mesma amplitude;
- Nº mínimo de intervalos 5, número máximo 20;

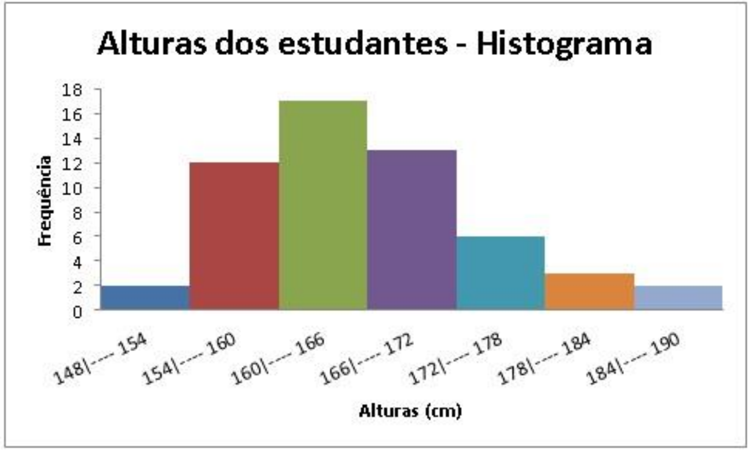
Histograma

- Quando possível os pontos médios dos intervalos devem coincidir com os valores realmente observados
- **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 compará-la com uma curva teórica (Normal).

Gráficos

- **Histograma:**
  - 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 frequências simples ou absolutas.

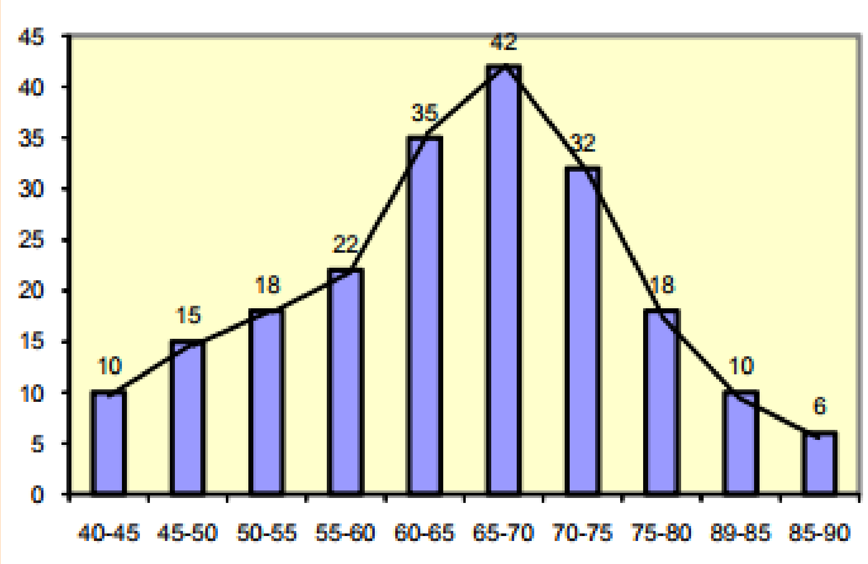




Histograma Simétrico

HISTOGRAMA E POLÍGONO DE FREQUÊNCIA

Pesos (x_1)	Nº alunos (f_1)
40-45	10
45-50	15
50-55	18
55-60	22
60-65	35
65-70	42
70-75	32
75-80	18
89-85	10
85-90	6
Total	208

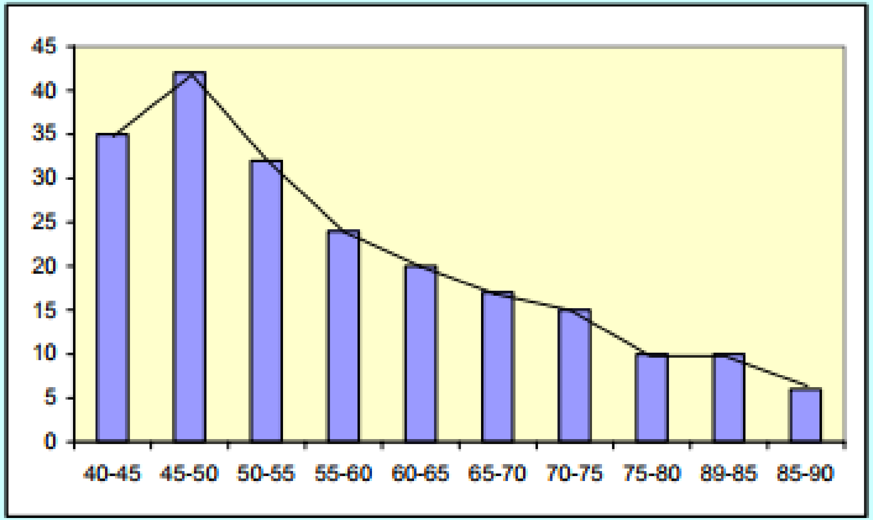


Histograma Assimétrico à Esquerda

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
89-85	10
85-90	6
Total	208

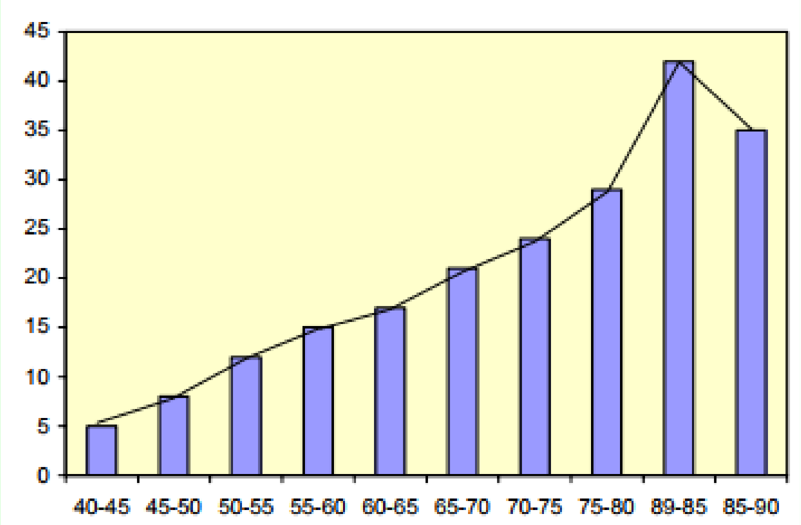


Histograma Assimétrico à Direita

HISTOGRAMA E POLÍGONO DE FREQUÊNCIA

Assimétrico à direita

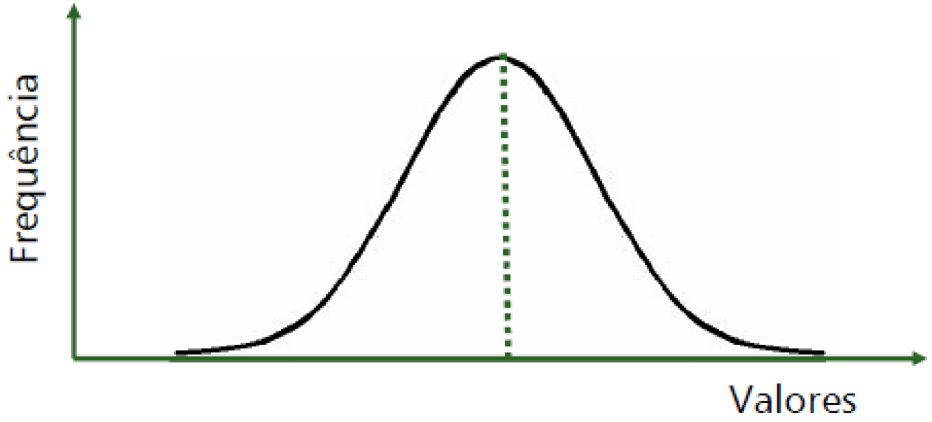
Pesos (x_1)	Nº alunos (f_1)
40-45	5
45-50	8
50-55	12
55-60	15
60-65	17
65-70	21
70-75	24
75-80	29
89-85	42
85-90	35
Total	173



Distribuição Gaussiana

Modelo Gaussiano

Algumas variáveis contínuas exibem um comportamento muito particular quando visualizamos a distribuição de frequências de seus valores.

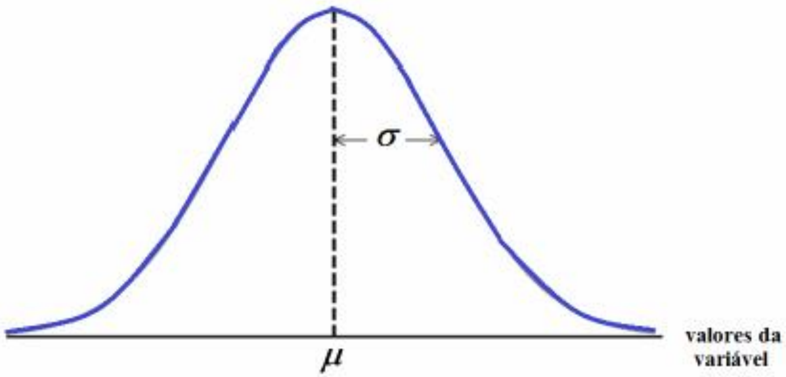


- Concentração de valores em torno de um valor central;
- Simetria em torno do valor central;
- Frequência pequena de valores muito extremos.

Modelo Gaussiano

O matemático alemão **Karl Gauss** popularizou um modelo proposto para a distribuição de probabilidades de variáveis do tipo descrito anteriormente.

A curva descrita por este modelo é conhecida como **Curva de Gauss** (ou também como **Curva Normal**)



Modelo Gaussiano

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

A função de densidade de X só depende de dois valores:

a média μ e o desvio-padrão σ

π e e são constantes conhecidas ($\pi \approx 3{,}14159\ldots$ e $e \approx 2.71828\ldots$)

Modelo Gaussiano

A média μ de uma variável aleatória X que siga o modelo Gaussiano pode assumir qualquer valor na reta real

$$-\infty < \mu < \infty$$

O desvio-padrão σ de qualquer variável aleatória X só pode assumir valores maiores do que zero

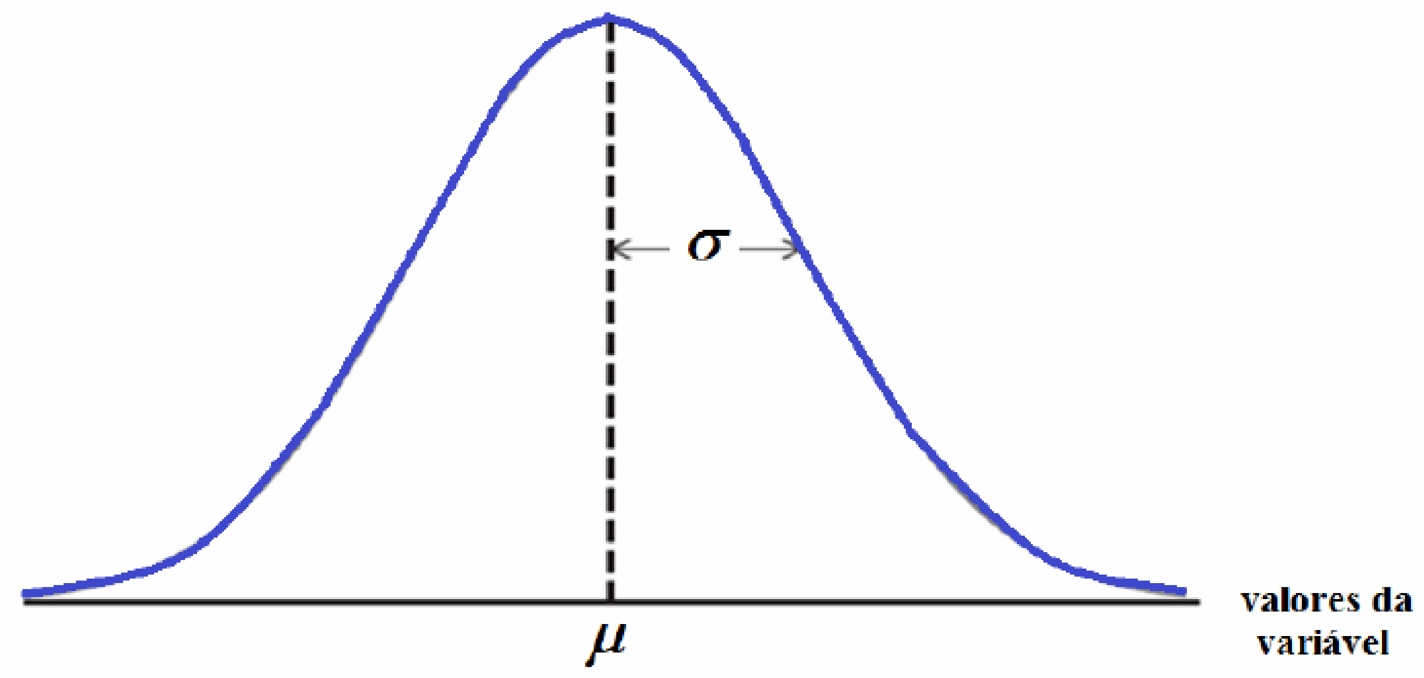
$$\sigma > 0$$

μ e σ são os parâmetros do Modelo Gaussiano

Dizemos que $X \sim$ Normal (μ, σ)

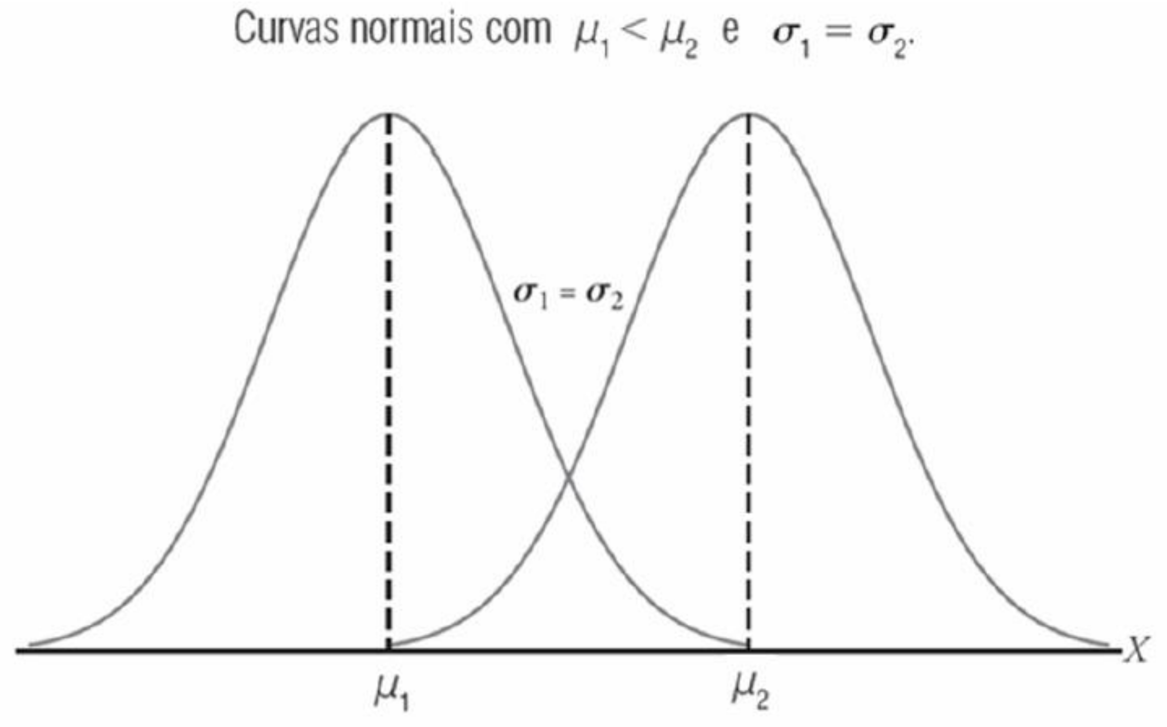
Modelo Gaussiano

A curva gaussiana (ou curva Normal) é definida pela média μ e pelo desvio-padrão σ.



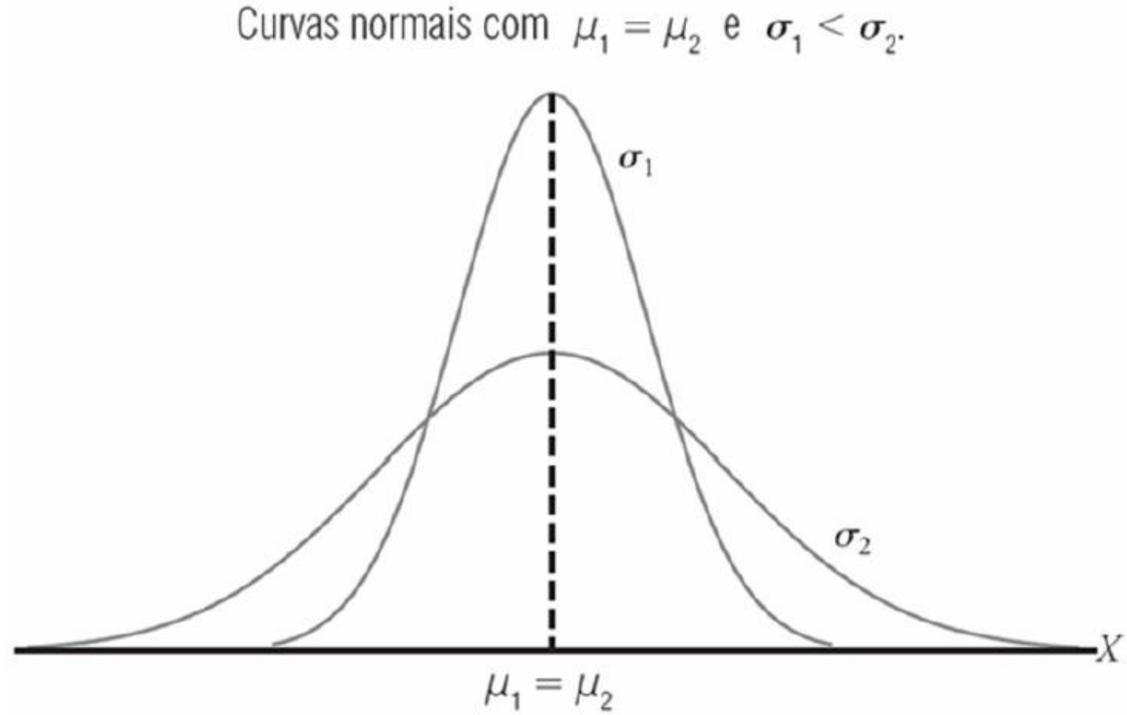
Modelo Gaussiano

- **O parâmetro μ informa onde está centrada a curva gaussiana**



Modelo Gaussiano

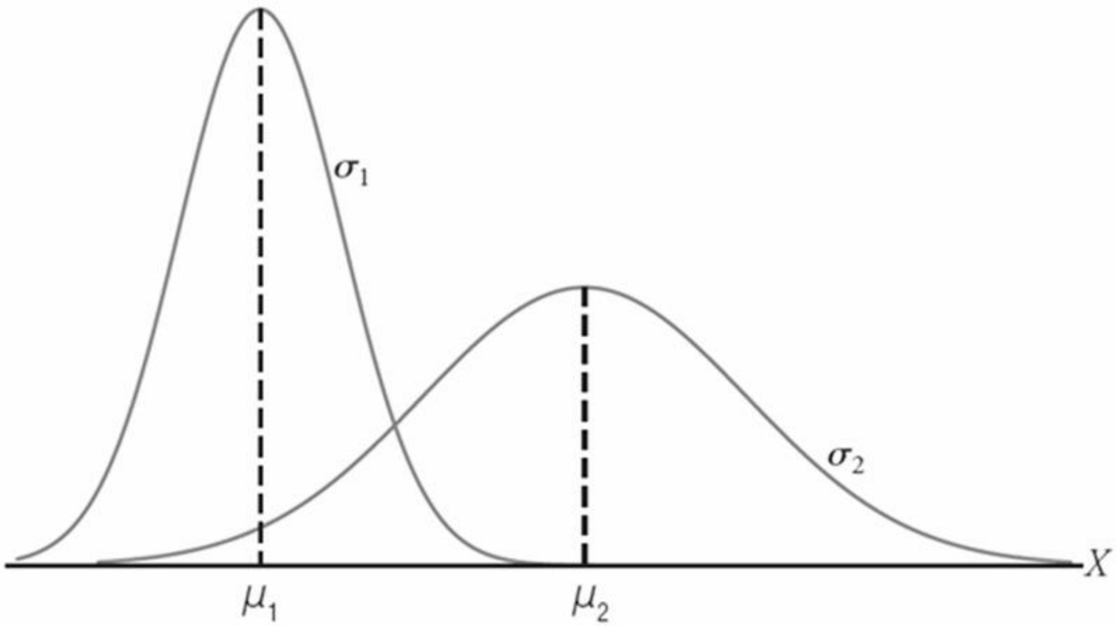
- **A forma do sino (mais "achatado" ou mais "alongado") é dada pelo valor do desvio-padrão σ**



Modelo Gaussiano

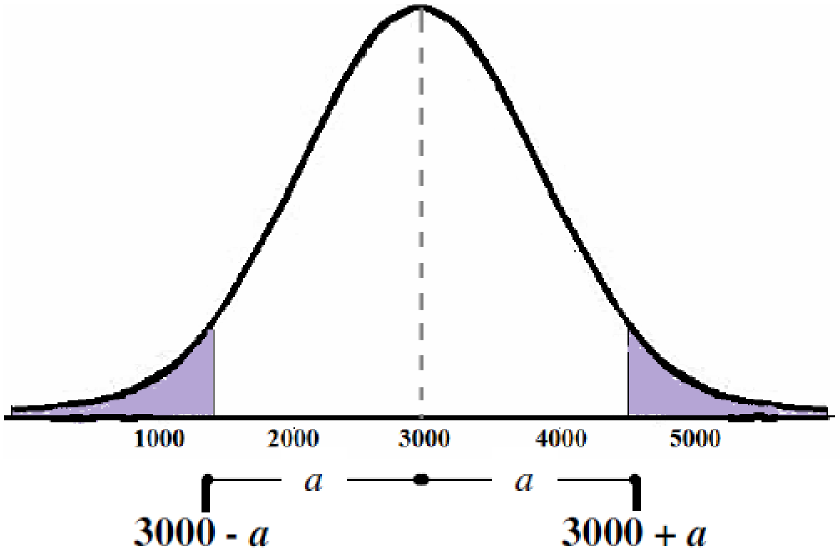
- **Para cada combinação de μ e σ, existe uma curva gaussiana diferente**

Curvas normais com $\mu_1 < \mu_2$ e $\sigma_1 < \sigma_2$.



Modelo Gaussiano

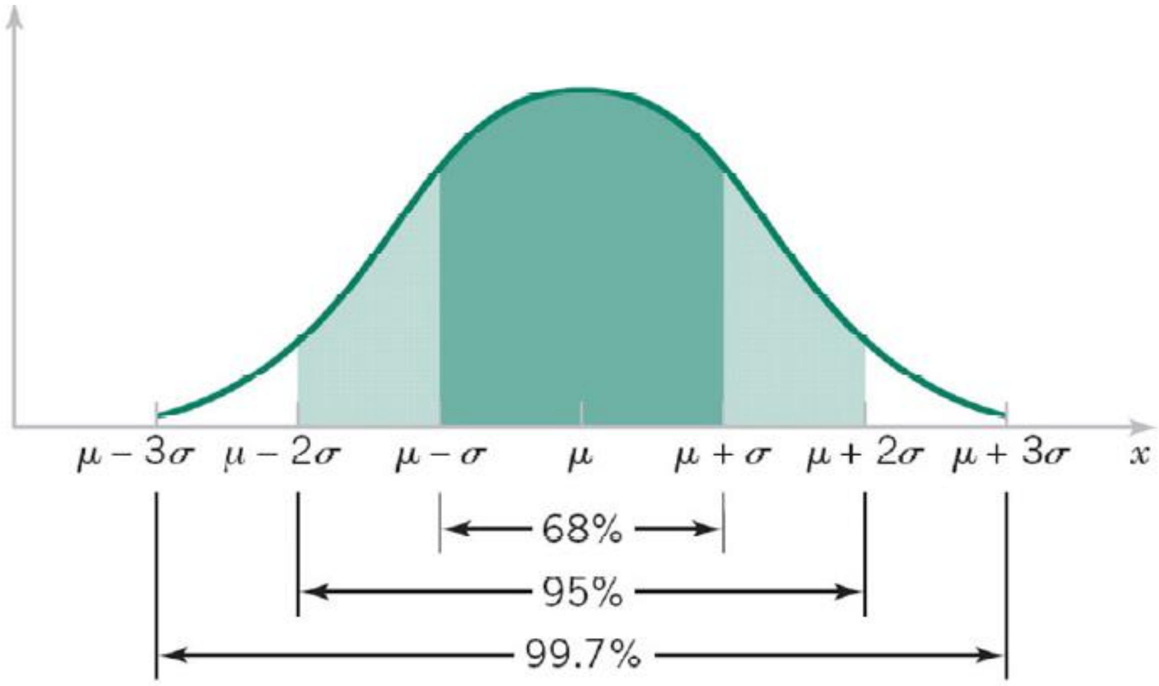
- **A curva gaussiana tem a forma de um sino e é simétrica em torno da média μ;**



$$P(X < 3000\text{-}a) = P(X > 3000\text{+}a)$$

Simetria

Propriedades da Distribuição Normal

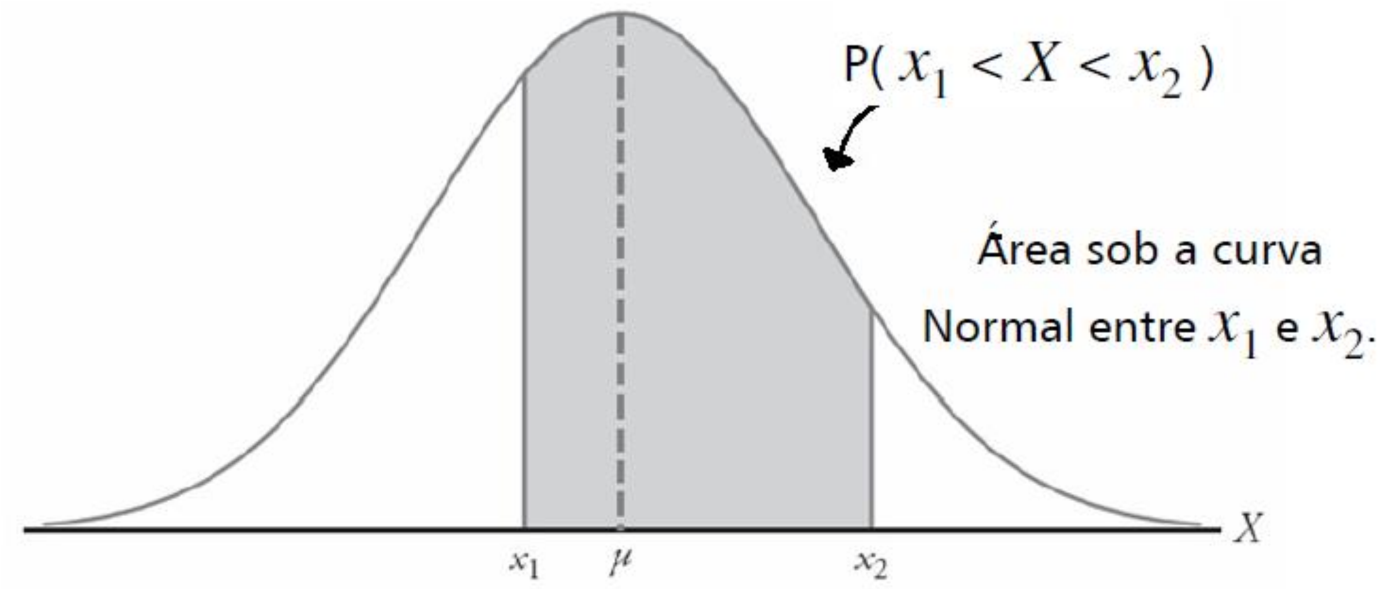


Área fixa entre intervalos simétricos

Cálculo de Probabilidade na Curva Normal

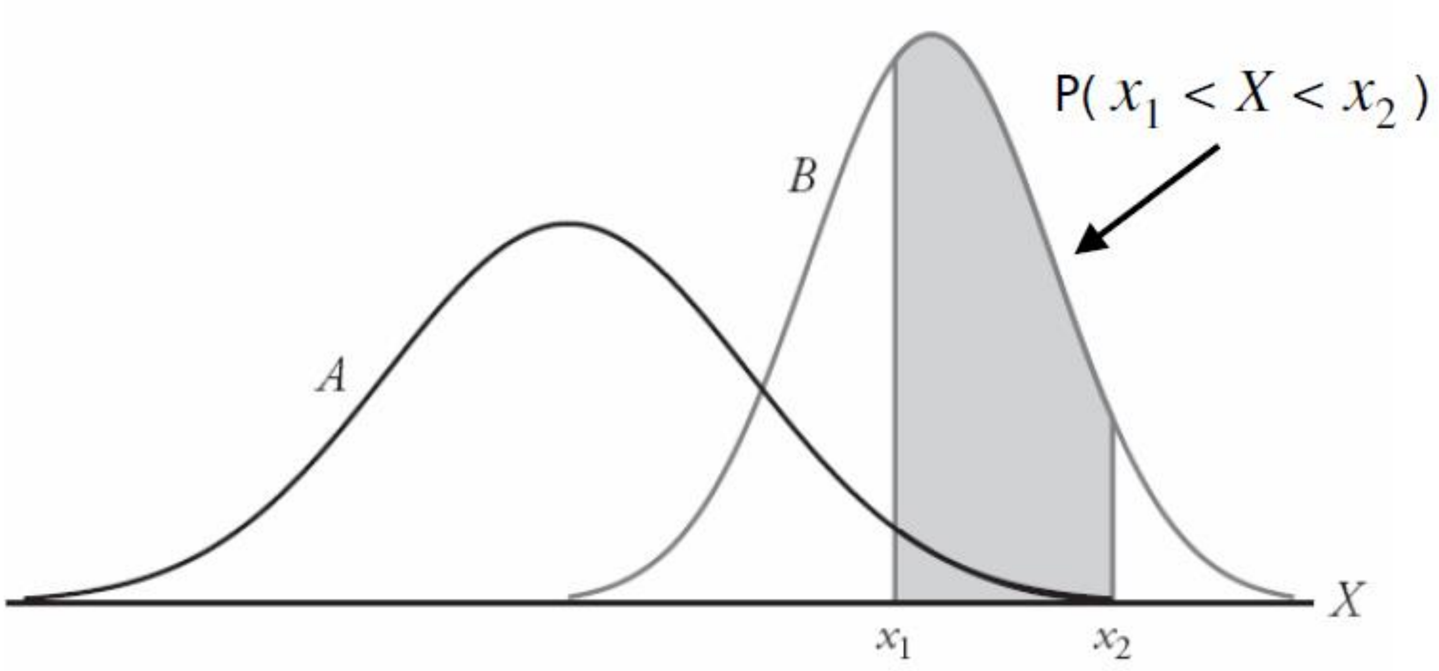
Considere uma variável aleatória X com distribuição Normal (μ, σ). Ou seja, $X \sim \text{Normal}(\mu, \sigma)$

Probabilidade de X estar entre x_1 e x_2: $P(x_1 < X < x_2)$



Cálculo de Probabilidade na Curva Normal

- Curvas Normais diferentes -> áreas diferentes



Distribuição Normal

Distribuição Normal

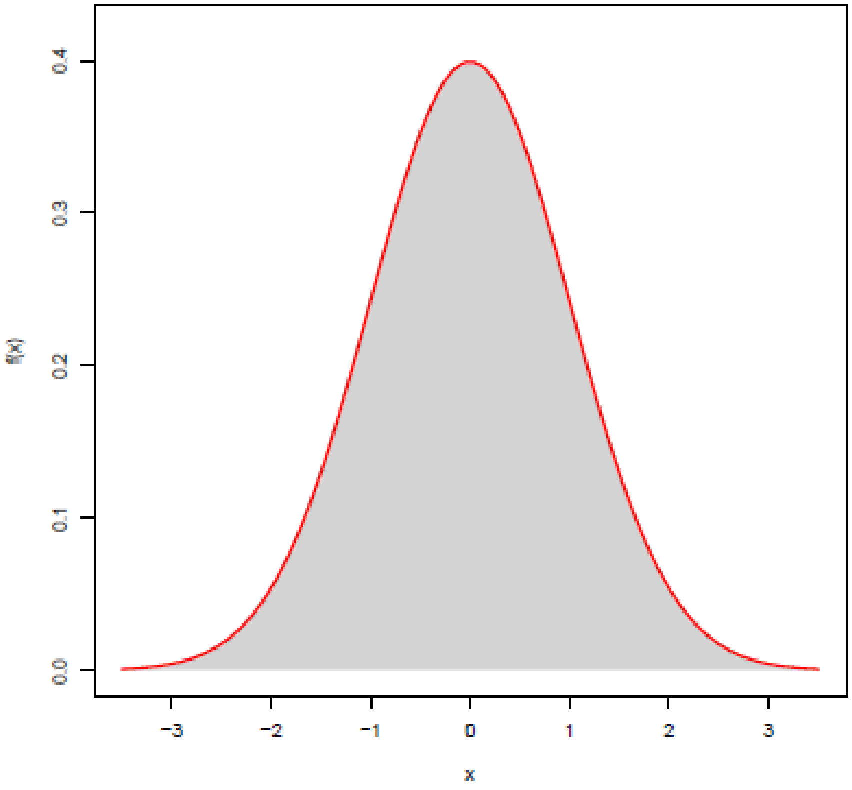
Distribuição Normal

Se X é uma variável aleatória com distribuição normal de média μ e variância σ^2, a função densidade de probabilidade de X é definida por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{\sigma^2}(x-\mu)^2},$$

para $-\infty < x, \mu < +\infty$ e $\sigma > 0$. Notação: $X \sim N(\mu, \sigma^2)$.

Descrição de $f(x)$ de uma N(0,1)



Distribuição Normal

Padronização

Se $X \sim \mathrm{N}(\mu, \sigma^2)$ e $Z \sim \mathrm{N}(0, 1)$ (normal padrão), então

$$P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right),$$

ou seja, todos os cálculos podem ser feitos pela normal padrão.

Distribuição Padrão Normal

Distribuição Padrão Normal

As probabilidades na curva Normal são calculadas com o auxílio de uma tabela.

Como existem infinitas combinações dos valores para μ e σ, seria inviável tabelar as probabilidades de todas as distribuições Normais possíveis.

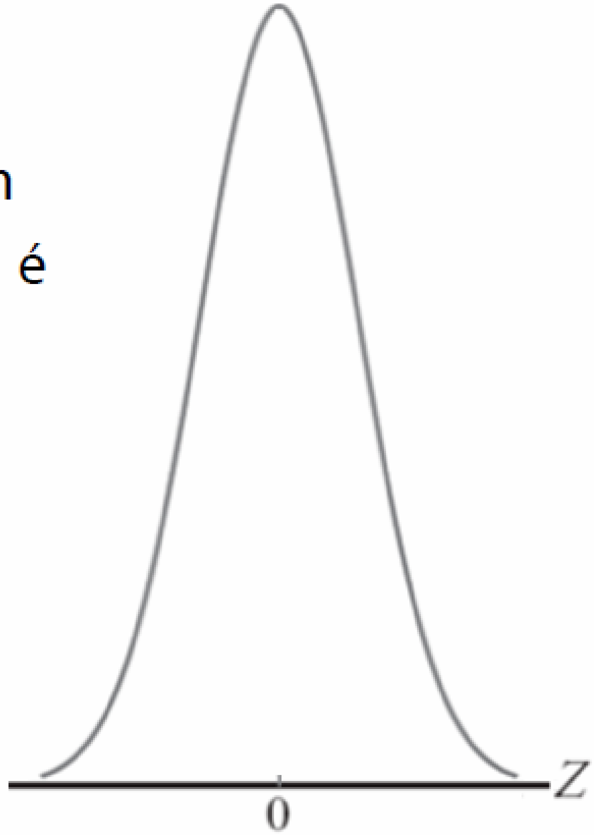
Sendo assim, uma única variável Normal possui suas probabilidades tabeladas: a variável Z com média igual a 0 e desvio-padrão igual a 1.

$$Z \sim \text{Normal}\ (\mu=0\ ;\ \sigma=1)$$

Distribuição Padrão Normal

A variável aleatória Normal com média $\mu=0$ e desvio-padrão $\sigma=1$ é chamada de

Variável Normal Padrão



Distribuição Padrão Normal

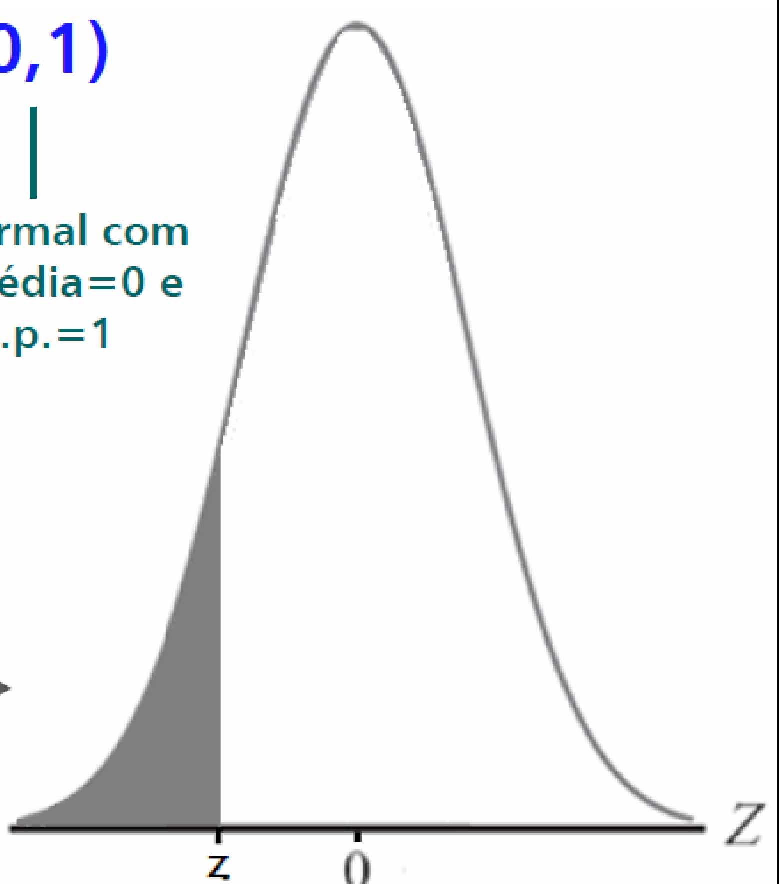
a variável aleatória Z

$Z \sim N(0,1)$

tem distribuição de probabilidade

Normal com média=0 e d.p.=1

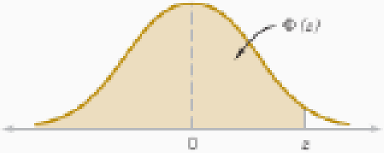
$P(Z < z)$



Função de distribuição acumulada: Tabela *Z*

$$z \geq 0$$

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

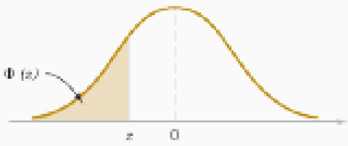


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201

Função de distribuição acumulada: Tabela *Z*

$z \leq 0$

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \, du$$



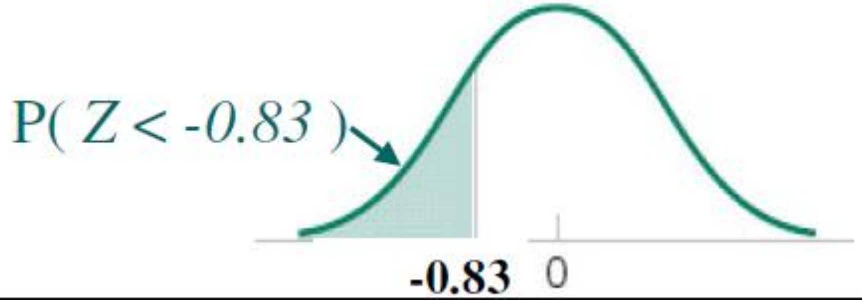
z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.03	−0.01	−0.00
−2.5	0.004799	0.004940	0.005085	0.005234	0.005386	0.005543	0.005703	0.005868	0.006037	0.006210
−2.4	0.006387	0.006569	0.006756	0.006947	0.007143	0.007344	0.007549	0.007760	0.007976	0.008198
−2.3	0.008424	0.008656	0.008894	0.009137	0.009387	0.009642	0.009903	0.010170	0.010444	0.010724
−2.2	0.011011	0.011304	0.011604	0.011911	0.012224	0.012545	0.012874	0.013209	0.013553	0.013903
−2.1	0.014262	0.014629	0.015003	0.015386	0.015778	0.016177	0.016586	0.017003	0.017429	0.017864
−2.0	0.018309	0.018763	0.019226	0.019699	0.020182	0.020675	0.021178	0.021692	0.022216	0.022750
−1.9	0.023295	0.023852	0.024419	0.024998	0.025588	0.026190	0.026803	0.027429	0.028067	0.028717
−1.8	0.029379	0.030054	0.030742	0.031443	0.032157	0.032884	0.033625	0.034379	0.035148	0.035930
−1.7	0.036727	0.037538	0.038364	0.039204	0.040059	0.040929	0.041815	0.042716	0.043633	0.044565
−1.6	0.045514	0.046479	0.047460	0.048457	0.049471	0.050503	0.051551	0.052616	0.053699	0.054799
−1.5	0.055917	0.057053	0.058208	0.059380	0.060571	0.061780	0.063008	0.064256	0.065522	0.066807
−1.4	0.068112	0.069437	0.070781	0.072145	0.073529	0.074934	0.076359	0.077804	0.079270	0.080757
−1.3	0.082264	0.083793	0.085343	0.086915	0.088508	0.090123	0.091759	0.093418	0.095098	0.096801
−1.2	0.098525	0.100273	0.102042	0.103835	0.105650	0.107488	0.109349	0.111233	0.113140	0.115070
−1.1	0.117023	0.119000	0.121001	0.123024	0.125072	0.127143	0.129238	0.131357	0.133500	0.135666
−1.0	0.137857	0.140071	0.142310	0.144572	0.146859	0.149170	0.151505	0.153864	0.156248	0.158655
−0.9	0.161087	0.163543	0.166023	0.168528	0.171056	0.173609	0.176185	0.178786	0.181411	0.184060
−0.8	0.186733	0.189430	0.192150	0.194894	0.197662	0.200454	0.203269	0.206108	0.208970	0.211855
−0.7	0.214764	0.217695	0.220650	0.223627	0.226627	0.229650	0.232695	0.235762	0.238852	0.241964
−0.6	0.245097	0.248252	0.251429	0.254627	0.257846	0.261086	0.264347	0.267629	0.270931	0.274253
−0.5	0.277595	0.280957	0.284339	0.287740	0.291160	0.294599	0.298056	0.301532	0.305026	0.308538
−0.4	0.312067	0.315614	0.319178	0.322758	0.326355	0.329969	0.333598	0.337243	0.340903	0.344578
−0.3	0.348268	0.351973	0.355691	0.359424	0.363169	0.366928	0.370700	0.374484	0.378281	0.382089
−0.2	0.385908	0.389739	0.393580	0.397432	0.401294	0.405165	0.409046	0.412936	0.416834	0.420740
−0.1	0.424655	0.428576	0.432505	0.436441	0.440382	0.444330	0.448283	0.452242	0.456205	0.460172
0.0	0.464144	0.468119	0.472097	0.476078	0.480061	0.484047	0.488033	0.492022	0.496011	0.500000

A Tabela Normal Padrão (Tabela Z) Parte Negativa

Linha: Parte inteira e primeira casa decimal de z

Coluna: Segunda casa decimal de z

z	0.00	0.01	0.02	0.03
-2.9	**0.0019**	**0.0018**	**0.0017**	**0.0017**
⋮		⋮		
-0.8	**0.2119**	**0.2090**	**0.2061**	**0.2033**

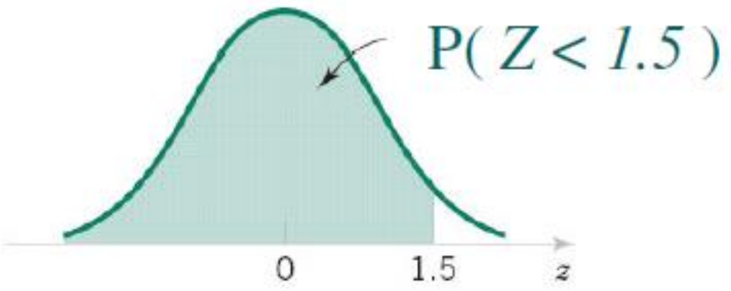


A Tabela Normal Padrão (Tabela Z) Parte Positiva

Linha: Parte inteira e primeira casa decimal de z

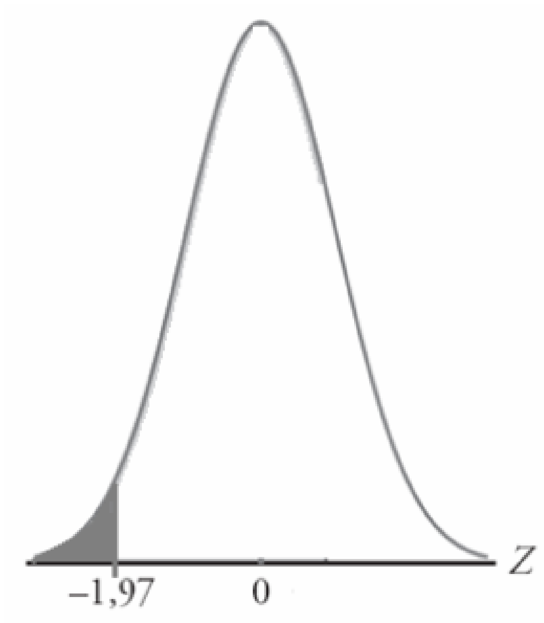
Coluna: Segunda casa decimal de z

z	0.00	0.01	0.02	0.03
0	0.5000	0.5039	0.5039	0.51197
⋮		⋮		
1.5	0.9331	0.9344	0.9357	0.93699



Exemplo: Seja Z uma v.a. normal padronizada. Calcule:

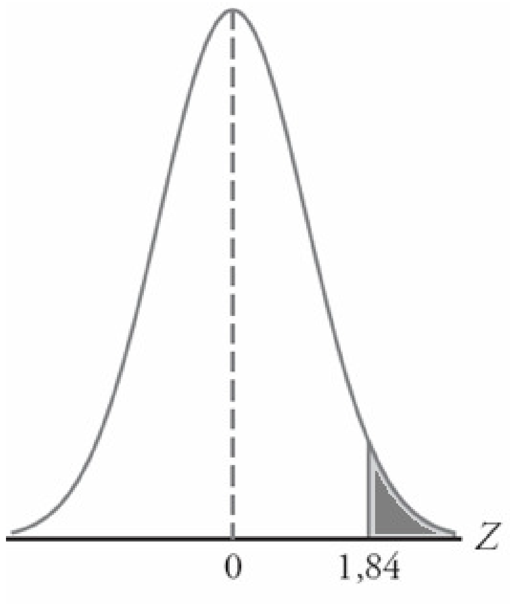
$P(Z < -1.97) = ?$



$P(Z < -1.97) = 0.0244$,

obtida direto da tabela.

$P(Z > 1.84) = ?$

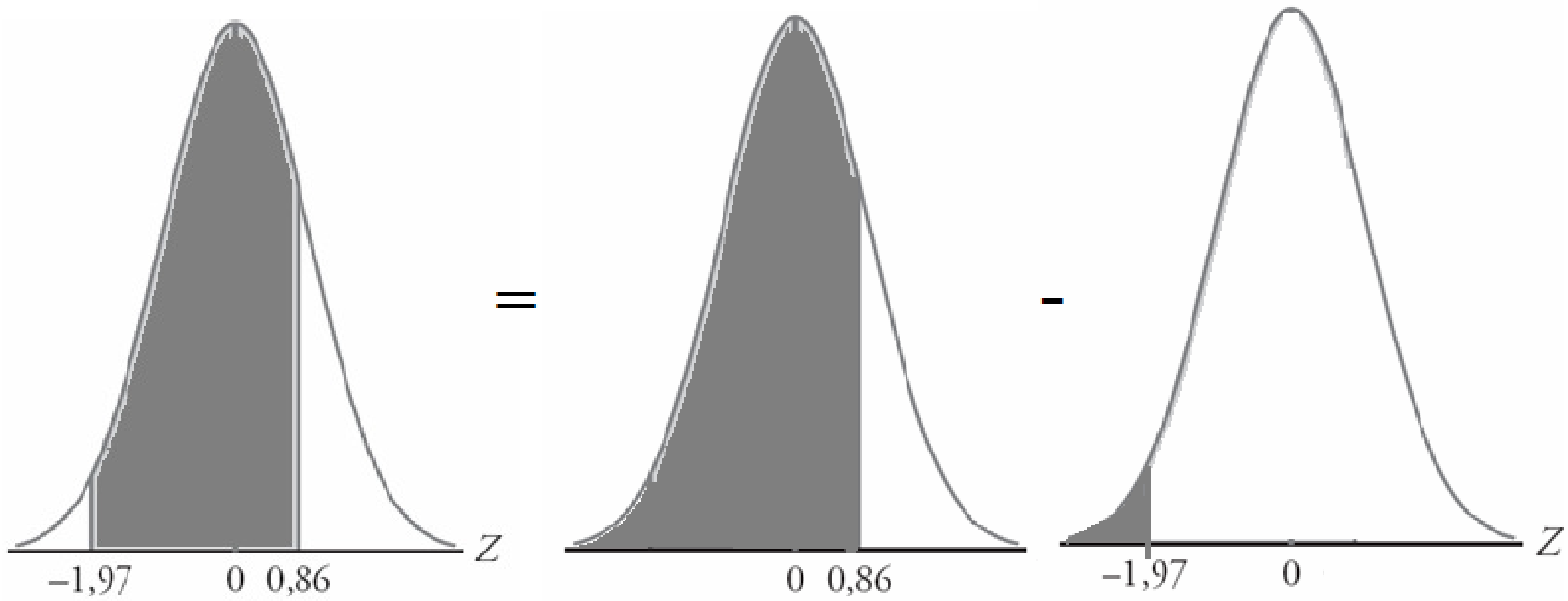


$P(Z > 1.84) = P(Z < -1.84) = 0.0329$,

obtida direto da tabela

e por simetria.

$$P(-1.97 < Z < 0.86) = P(Z < 0.86) - P(Z < -1.97)$$
$$= 0.8051 - 0.0244$$
$$= 0.7807$$

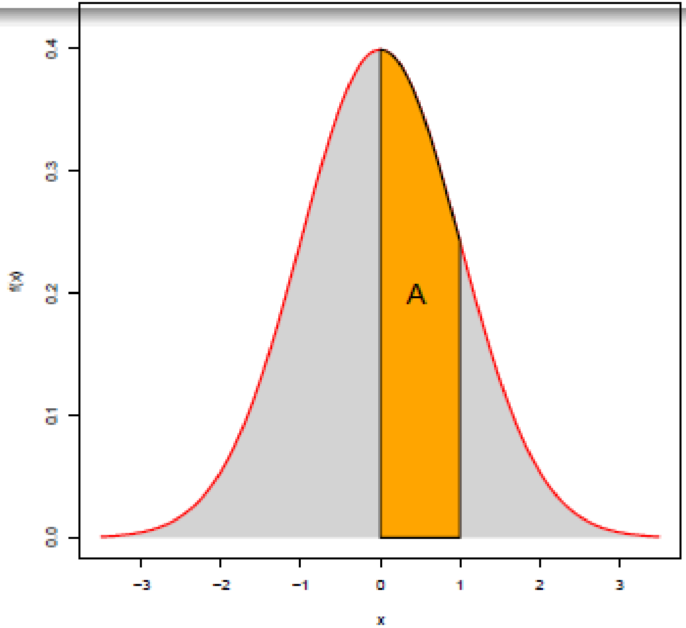


A Tabela Normal Padrão (Tabela Z)

Cálculo de probabilidades

Por exemplo, a probabilidade $A = P(0 \leq X \leq 1)$ pode ser calculada pela diferença

$$P(X \leq 1) - P(X \leq 0) = 0,841 - 0,5 = 0,341.$$

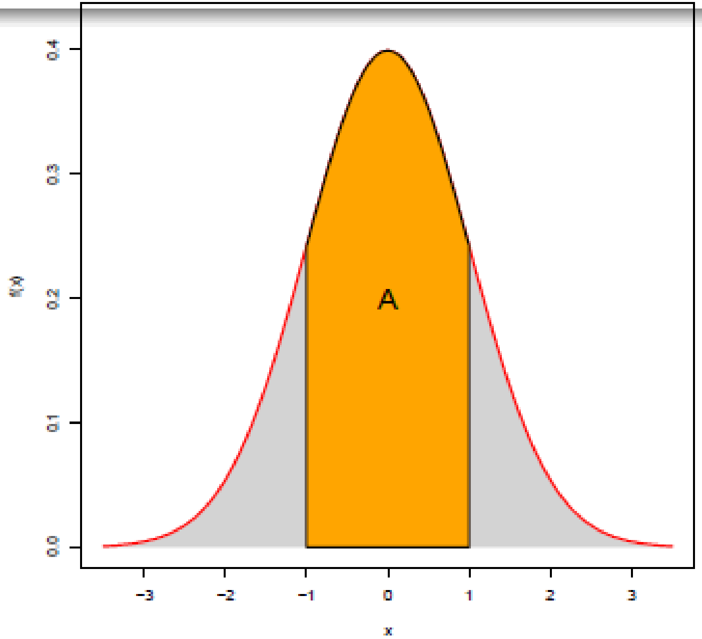


A Tabela Normal Padrão (Tabela Z)

Cálculo de probabilidades

Para calcular a probabilidade $A = P(-1 \leq X \leq 1)$ podemos usar o fato da distribuição ser simétrica na média. Assim,

$$P(-1 \leq X \leq 1) = 2 \times P(0 \leq X \leq 1) = 2 \times 0,341 = 0,682.$$



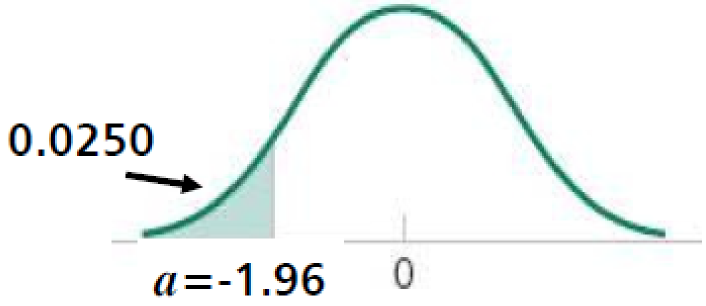
Cálculo de percentis na curva Normal

Percentil de ordem 2.5

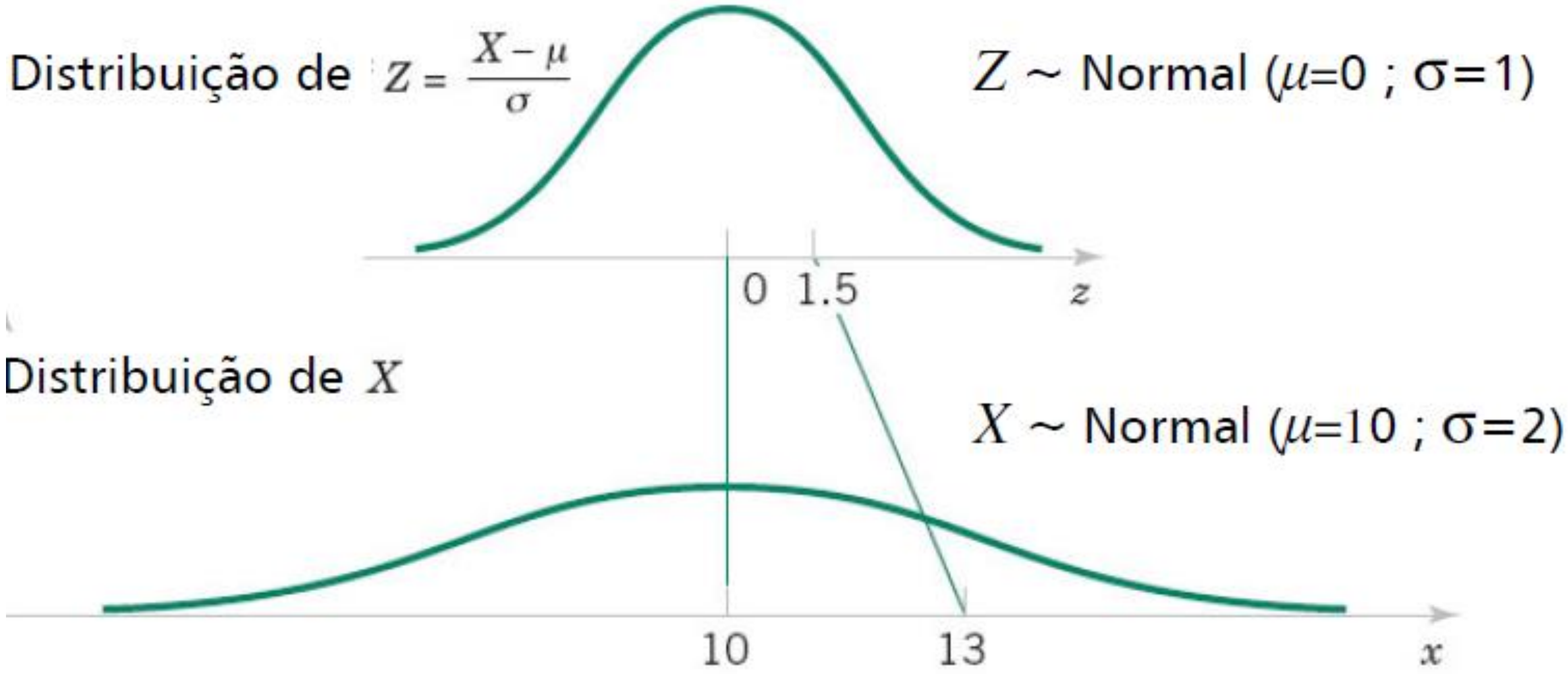
Que valor de Z na tabela Normal Padrão deixa uma área de 0.0250 abaixo dele ?

Ou seja, quem é a tal que P[$Z < a$]=0.0250 ?

a é o percentil 2.5 da curva Normal Padrão



Como usar a tabela Normal Padrão para calcular probabilidades em uma curva Normal qualquer?



Padronização de uma variável aleatória Normal

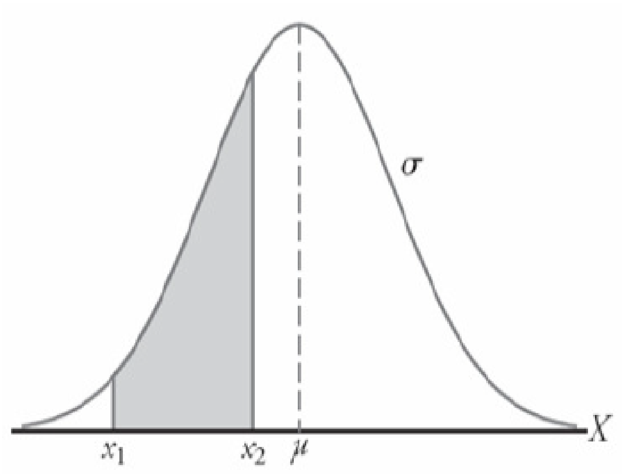
Podemos transformar uma variável aleatória $X \sim \mathit{Normal}\ (\mu, \sigma)$ em uma variável aleatória $Z \sim \mathit{Normal}\ (0, 1)$ usando a expressão:

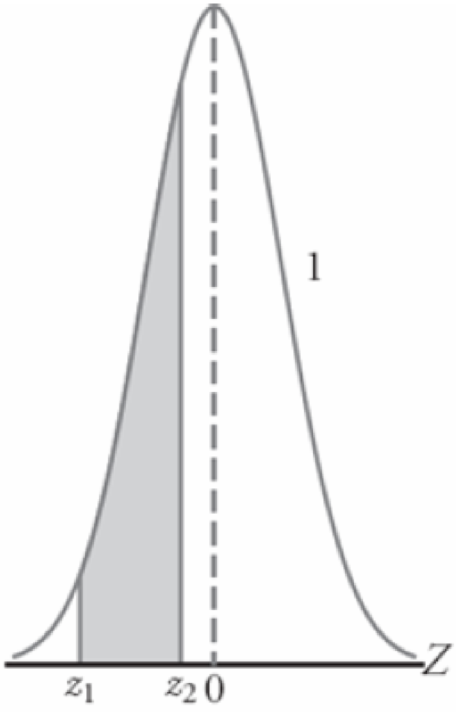
$$Z = \frac{X - \mu}{\sigma}$$

Padronização de uma variável aleatória Normal

$X \sim Normal(\mu, \sigma)$

$Z \sim Normal(0,1)$





$$z_1 = \frac{x_1 - \mu}{\sigma} \qquad z_2 = \frac{x_2 - \mu}{\sigma}$$

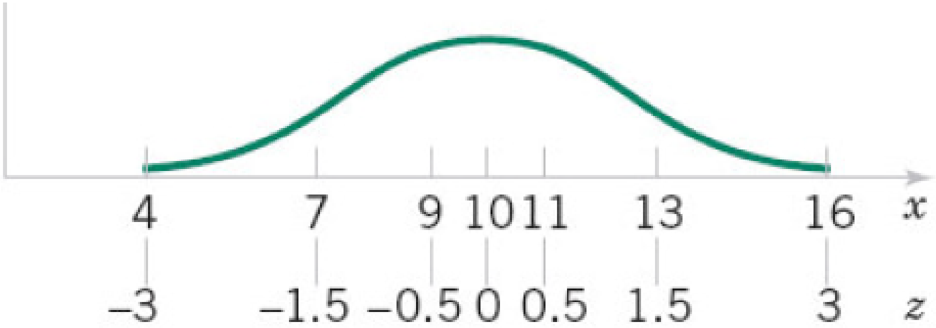
Calculando probabilidades de X utilizando a tabela Z

$$P[X<9]=\quad P\left[\frac{X-\mu}{\sigma}<\frac{9-\mu}{\sigma}\right]=\quad P\left[\frac{X-10}{2}<\frac{9-10}{2}\right]$$

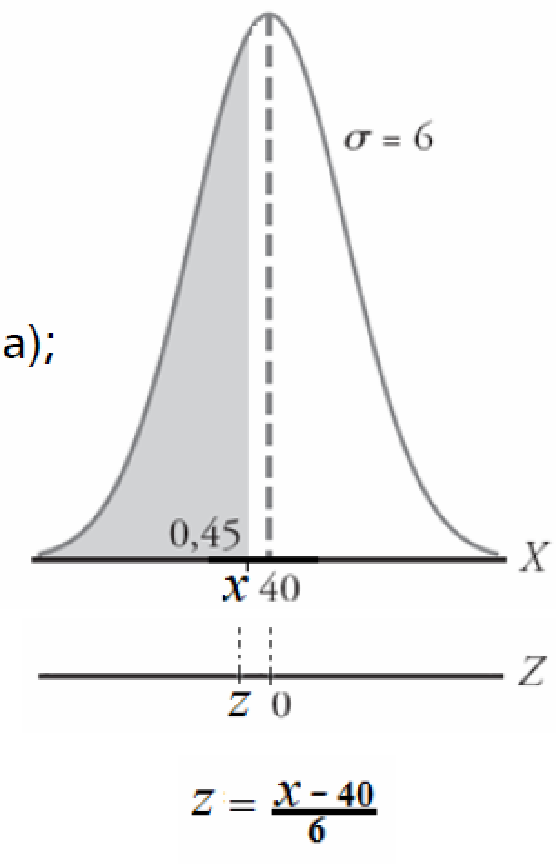
$$=P[Z<-0.5]=0.3085$$

$$P[X>13]=P[\frac{X-10}{2}>\frac{13-10}{2}]=\quad P[Z>1.5]$$

$$=P[Z<-1.5]=0.0668$$



Exemplo 1: Se X tem distribuição Normal com μ = 40 e σ = 6, encontre o valor de x tal que P[$X < x$] =0.45.



Se P[$X < x$] =0.45.

então P($Z < (x$-40)/6) = 0.45.

Mas P($Z <$ -0.13) = 0.45 (da tabela);

Logo (x-40)/6 = -0.13

⇨ x = 40 + (-0.13)6
= 40 - 0.78
= 39.22.

Ou seja, 39.22 é o percentil 45 da distribuição de X.

Exemplo 2: Se X tem distribuição Normal com $\mu = 40$ e $\sigma = 6$, encontre o valor de x tal que $P[X > x] = 0.14$.

Se $P[X < x] = 0.86$

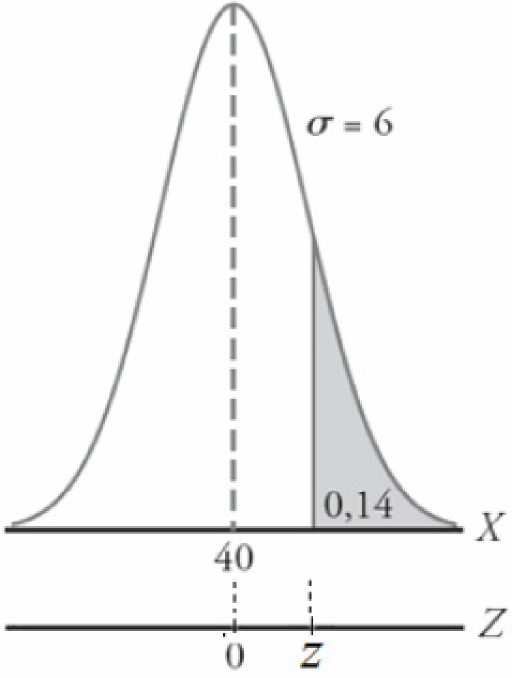
então $P(Z < (x-40)/6) = 0.86$.

Mas $P(Z > 1.08) = P(Z < -1.08)$
$= 0.14$ (da tabela);

Logo $(x-40)/6 = 1.08$

⇨ $x = 40 + (1.08)6$
$= 46.48$

Ou seja, 46.48 é o percentil 86 da distribuição de X.



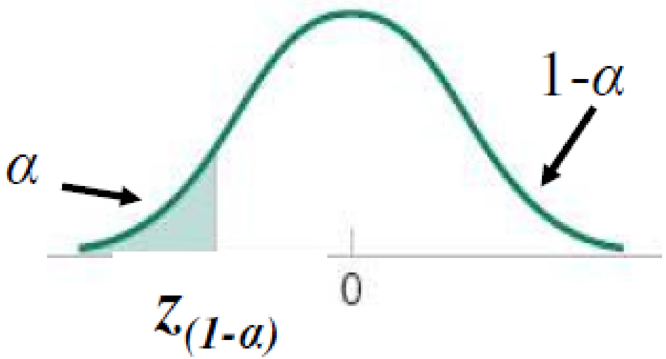
$$z = \frac{x - 40}{6}$$

Cálculo do Percentil de ordem 100α da distribuição Normal

$$P_{100\alpha} = \mu + z_{(1-\alpha)} \cdot \sigma$$

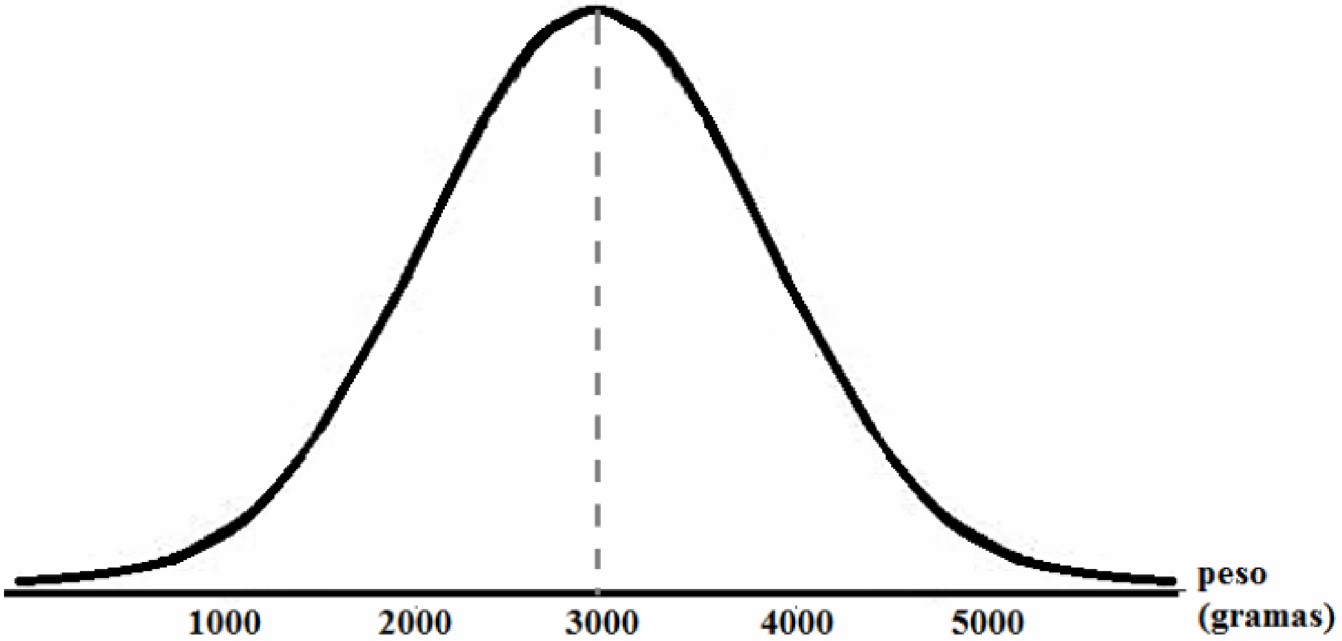
onde α é a ordem do percentil ($0 < \alpha < 1$) e

$z_{(1\text{-}\alpha)}$ é o valor na tabela Z que deixa uma área de (1-α) **acima** dele.

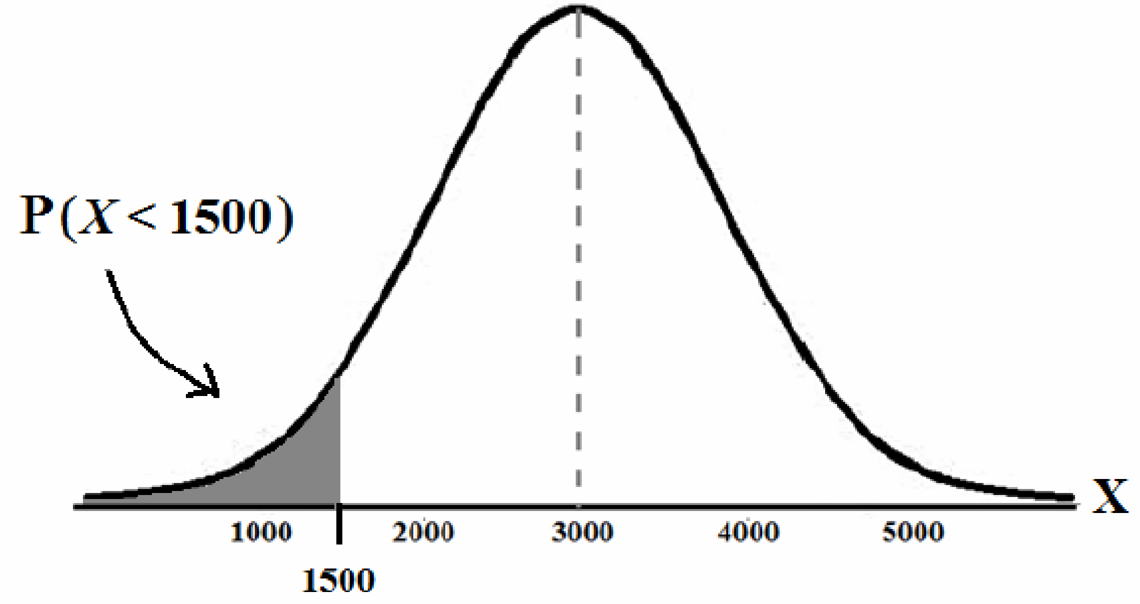


Exemplo Inicial:

Suponha que X é o peso de bebês ao nascer e que, em certa população, X tem distribuição que pode ser aproximada pela Normal com μ = 3000g e σ = 1000g.



Qual é a porcentagem de bebês que nascem com peso abaixo de 1500g ?

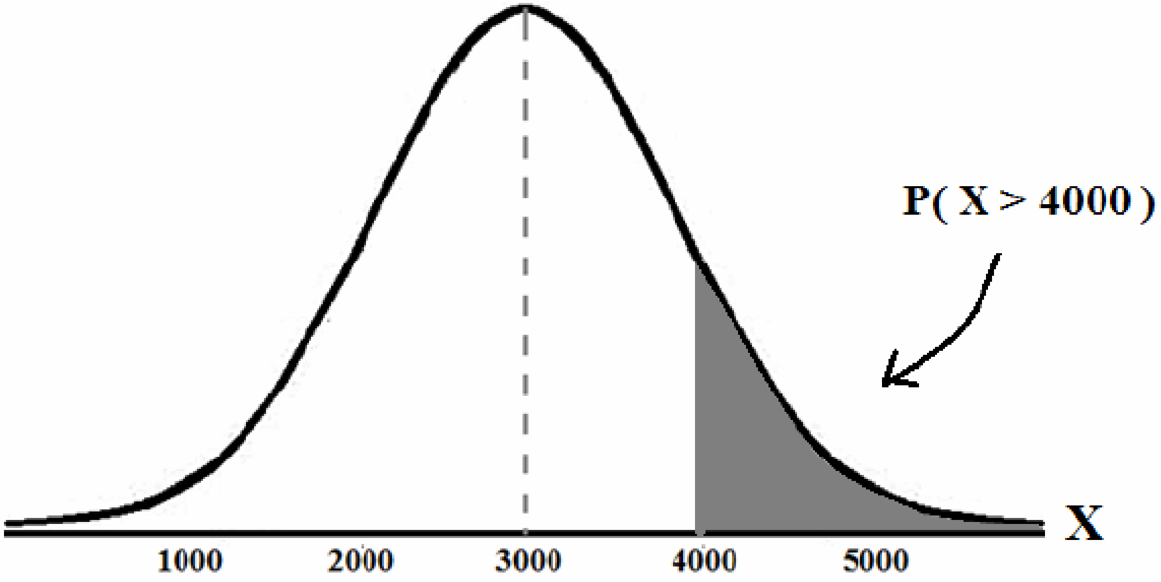


$$P[X < 1500] = P\left[\frac{X-3000}{1000} < \frac{1500-3000}{1000}\right]$$

$$= P[Z < -1.5] = 0.0068$$

0.68% dos bebês têm peso inferior a 1500g.

Qual é a porcentagem de bebês que nascem com peso acima de 4000g ?

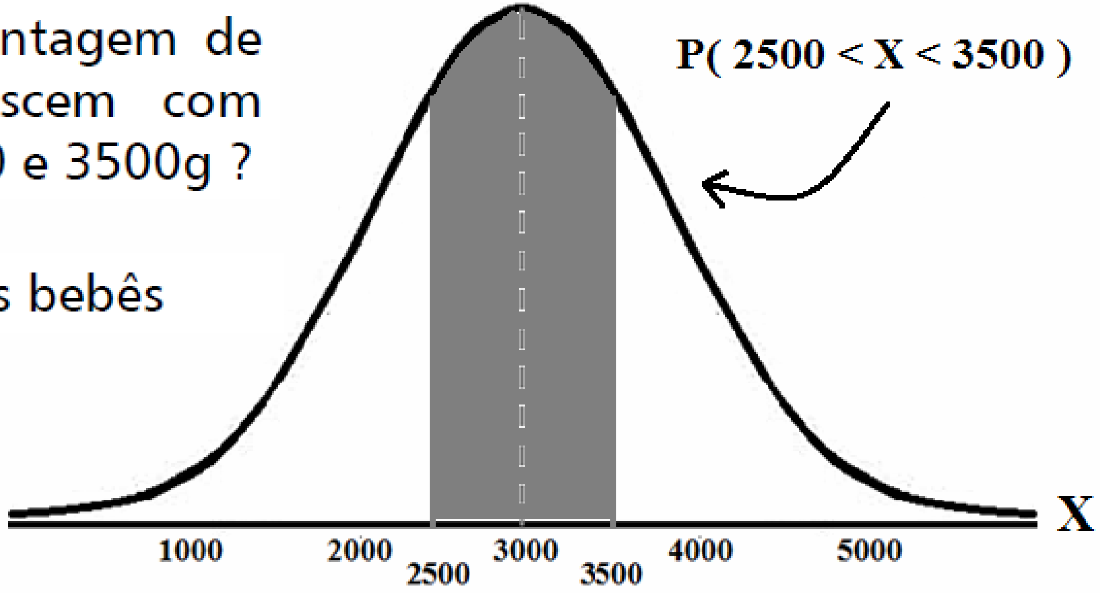


$$P[X > 4000] = P\left[Z > \frac{4000 - 3000}{1000}\right]$$

$$= P[Z > 1.0] = P[Z < -1.0] = 0.1587$$

Qual é a porcentagem de bebês que nascem com peso entre 2500 e 3500g ?

38.30% dos bebês



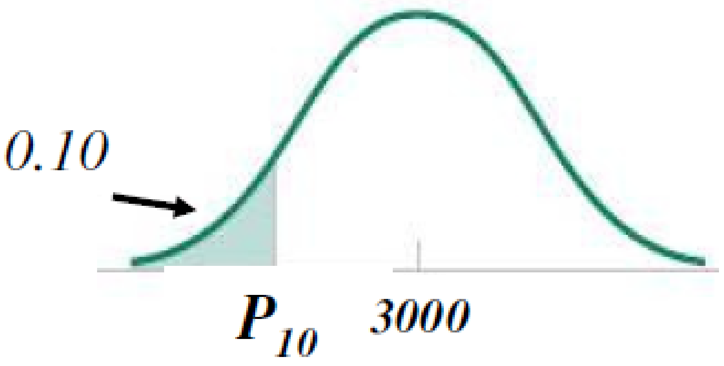
$$P[2500 < X < 3500] = P[X < 3500] - P[X < 2500]$$

$$= P\left[Z < \frac{3500 - 3000}{1000}\right] - P\left[Z < \frac{2500 - 3000}{1000}\right]$$

$$= P[Z < 0.5] - P[Z < -0.5]$$

$$= 0.6915 - 0.3085 = 0.3830$$

Qual valor de peso dos bebês separa os 10% mais leves?



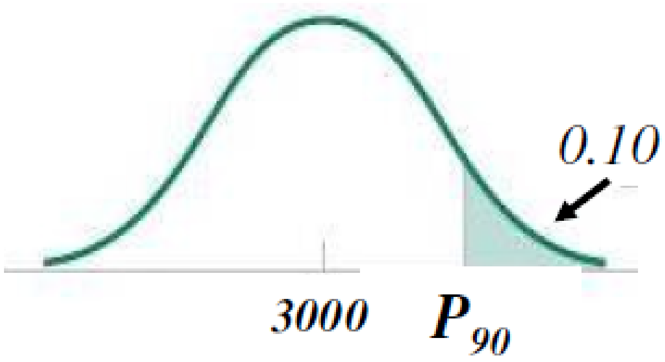
1720 gramas

$$\alpha = 0.10$$

$$\begin{aligned} P_{10} &= 3000 + z_{(1-0.1000)} \times 1000 \\ &= 3000 + z_{(0.9000)} \times 1000 \\ &= 3000 + (-1.28) \times 1000 \\ &= 3000 - 1280 = 1720 \end{aligned}$$

Qual valor de peso dos bebês separa os 10% mais pesados?

4280 gramas



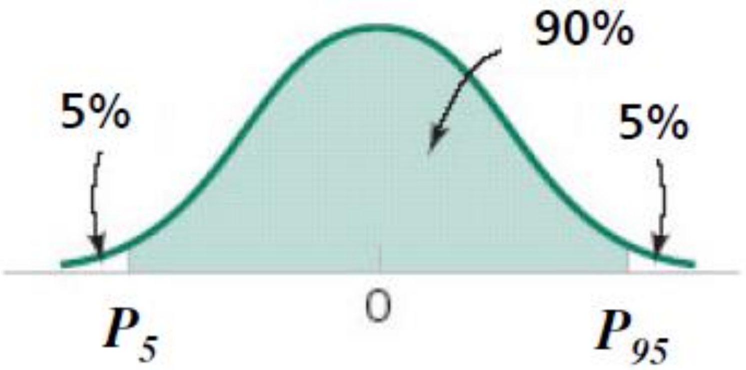
$$\alpha = 0.90$$

$$\begin{aligned} P_{10} &= 3000 + z_{(1-0.9000)} \times 1000 \\ &= 3000 + z_{(0.1000)} \times 1000 \\ &= 3000 + 1.28 \times 1000 \\ &= 3000 + 1280 = 4280 \end{aligned}$$

Aplicações do Modelo Gaussiano: Cálculo de Faixas de Referência

Faixas de Referência são formadas por dois percentis

Exemplo: uma faixa de referência de 90% é formada pelos percentis 5 e 95



Cálculo de Faixas de Referência utilizando o modelo Gaussiano

Seja X a variável aleatória que representa a característica para a qual queremos construir uma faixa de referência.

Exemplo: X é o peso de recém-nascidos

Se $X \sim$ Normal (μ,σ), então uma faixa de referência de $(1\text{-}\alpha)100\%$ é formada pelos percentis

$$P_{(\alpha/2)100} \quad \text{e} \quad P_{(1\text{-}\alpha/2)100}$$

Exemplo: uma faixa de referência de 90% $(\alpha=0.10)$ é formada pelos percentis P_5 e P_{95}

Cálculo de Faixas de Referência utilizando o modelo Gaussiano

No exemplo do peso dos recém-nascidos (X), se pudermos supor que $X \sim$ Normal ($\mu=3000$; $\sigma=1000$),

uma faixa de referência de 80% seria dada pelos percentis $P_{10}=$ 1720 e P_{90} = 4280 (já calculados anteriormente)

Faixa de Referência de 80% para o peso de recém-nascidos

1720 gramas a 4280 gramas

Cálculo de Faixas de Referência utilizando o modelo Gaussiano

Relembrando o cálculo de percentis com o modelo gaussiano e a definição de faixa de referência,

uma faixa de referência de (1-α)100% é dada pela expressão

$$[\mu + z_{(1-\alpha/2)} \cdot \sigma \;\; ; \; \mu + z_{(\alpha/2)} \cdot \sigma]$$

$$z_{(1-\alpha/2)} = -z_{(\alpha/2)} \quad \text{(por simetria)}$$

$$[\mu - z_{(\alpha/2)} \cdot \sigma \;\; ; \; \mu + z_{(\alpha/2)} \cdot \sigma]$$