

# Ensino de Estatística

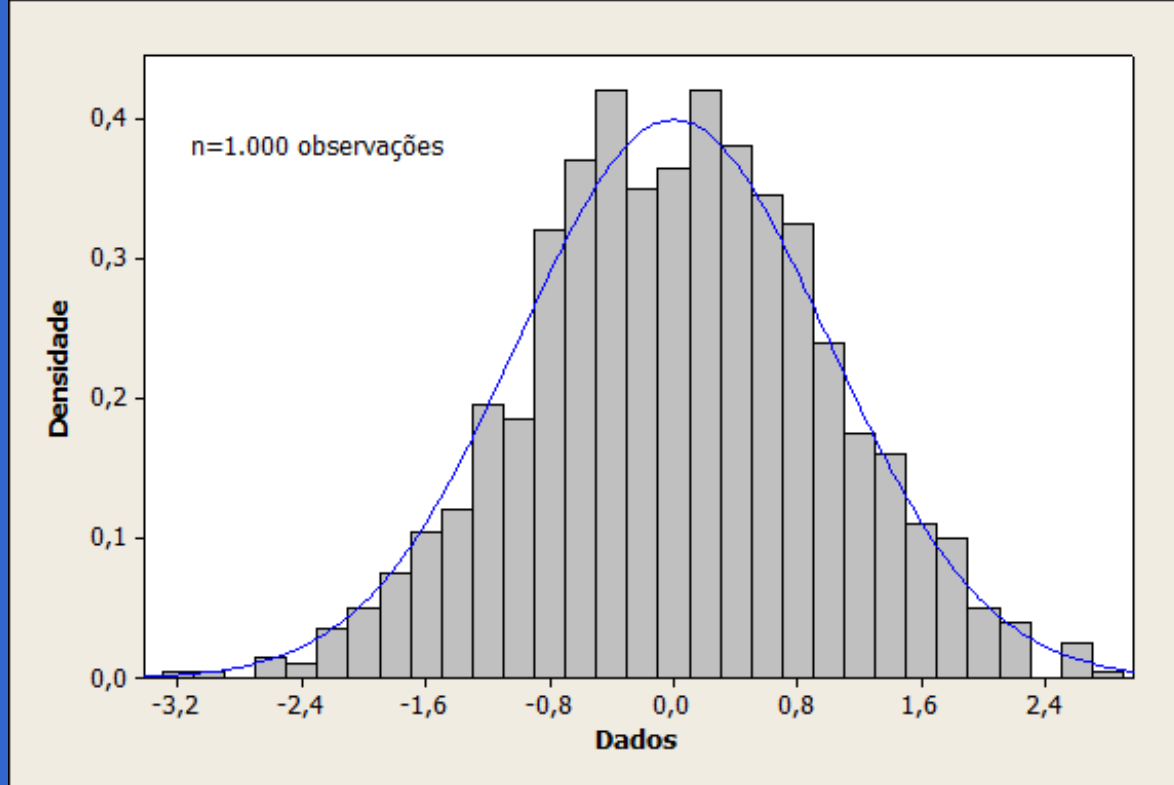
Lupércio França Bessegato  
Ronaldo Rocha Bastos

Departamento de Estatística/UFJF

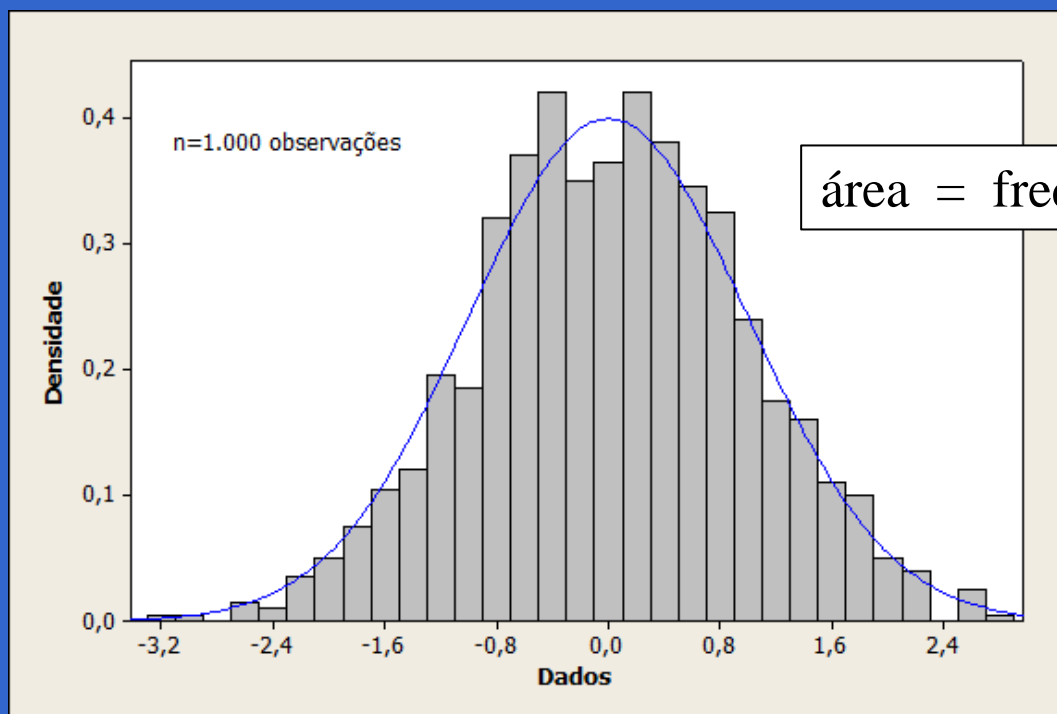
# Distribuição Normal ou Gaussiana

# Exploração de Dados Univariados

- Visualize graficamente seus dados
- Busque padrão global e pontos atípicos
- Resumir numericamente centro e dispersão
- Às vezes, o padrão global de um grande número de observações é tão regular que pode ser descrito por uma **curva suave**



- Distribuição descrita por curva suave
  - √ Mais fácil para trabalhar
- A curva é um modelo matemático
  - √ descrição matemática idealizada

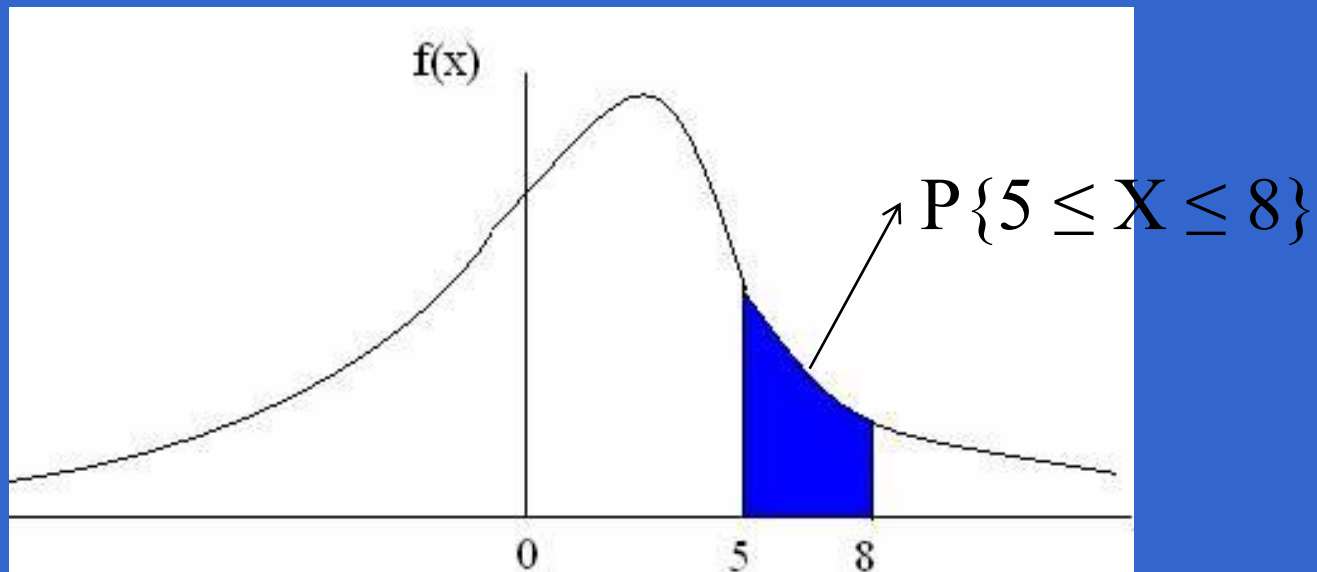


- Áreas das barras em histograma:
  - √ Contagens (ou proporções)
  - √ Área sob a curva é exatamente 1
- Áreas sob a curva
  - √ Representa proporções de observações

# Curva de Densidade

- A curva é denominada curva de densidade
- Propriedades:
  - √ Está sempre sobre ou acima do eixo horizontal
  - √ Tem área exatamente igual a 1 entre ela e o eixo horizontal

- A proporção entre 2 pontos é igual à área sob a curva, entre os dois pontos e o eixo x



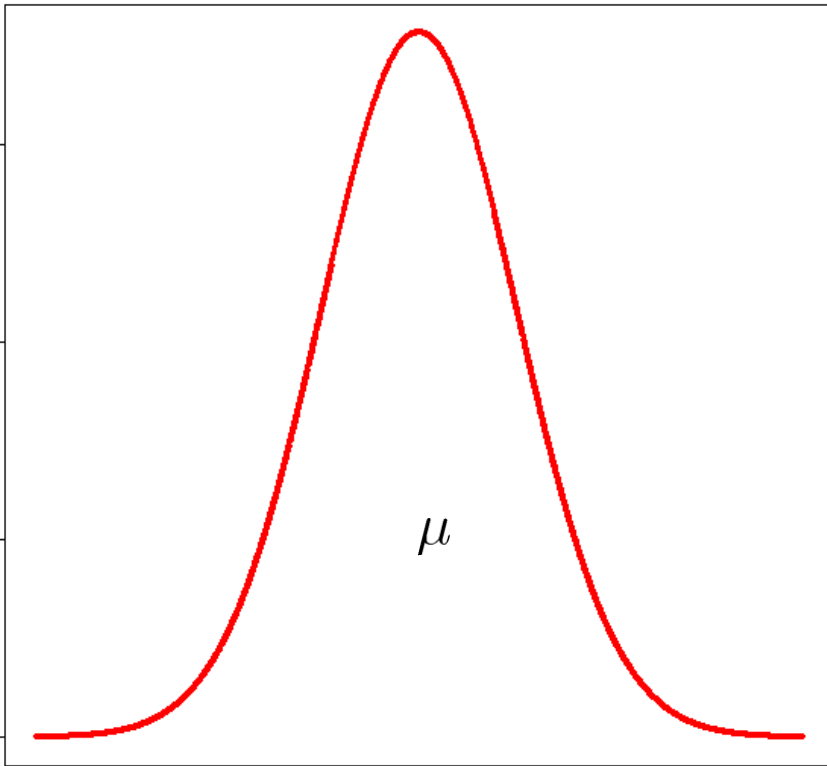
# Curvas Normais

- É uma classe importante de curvas de densidade
- Características:
  - √ São simétricas, unimodais e tem forma de sino
  - √ Descrevem distribuições normais (gaussianas)



# Função de Densidade

Densidade da distribuição Normal

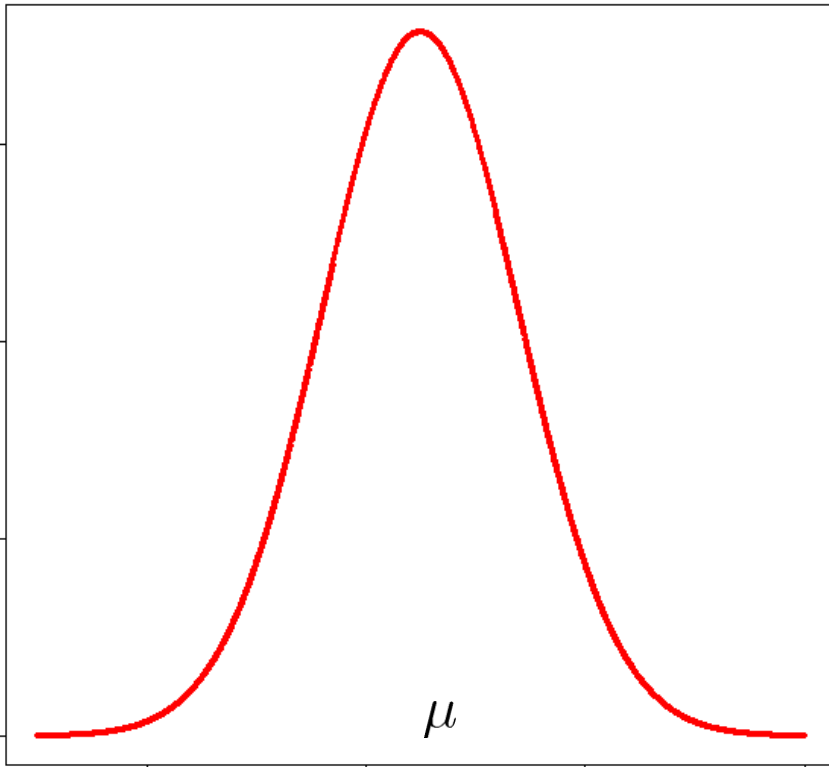


$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- ✓ Gráfico tem formato de sino
- ✓ Parâmetros da distribuição normal:
  - Média ( $\mu$ )
  - Desvio-padrão ( $\sigma$ ) e variância ( $\sigma^2$ )

# Características

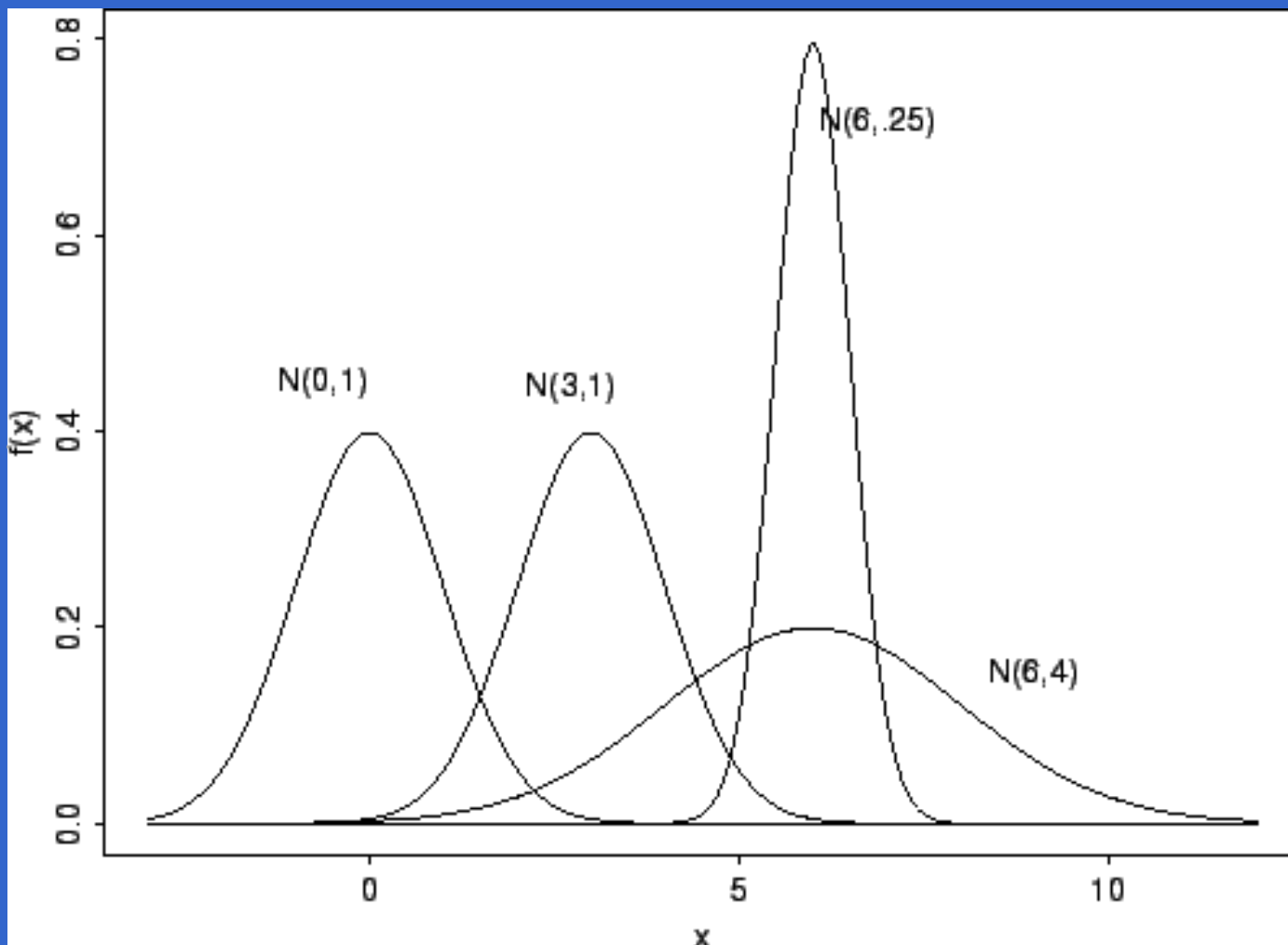
Densidade da distribuição Normal



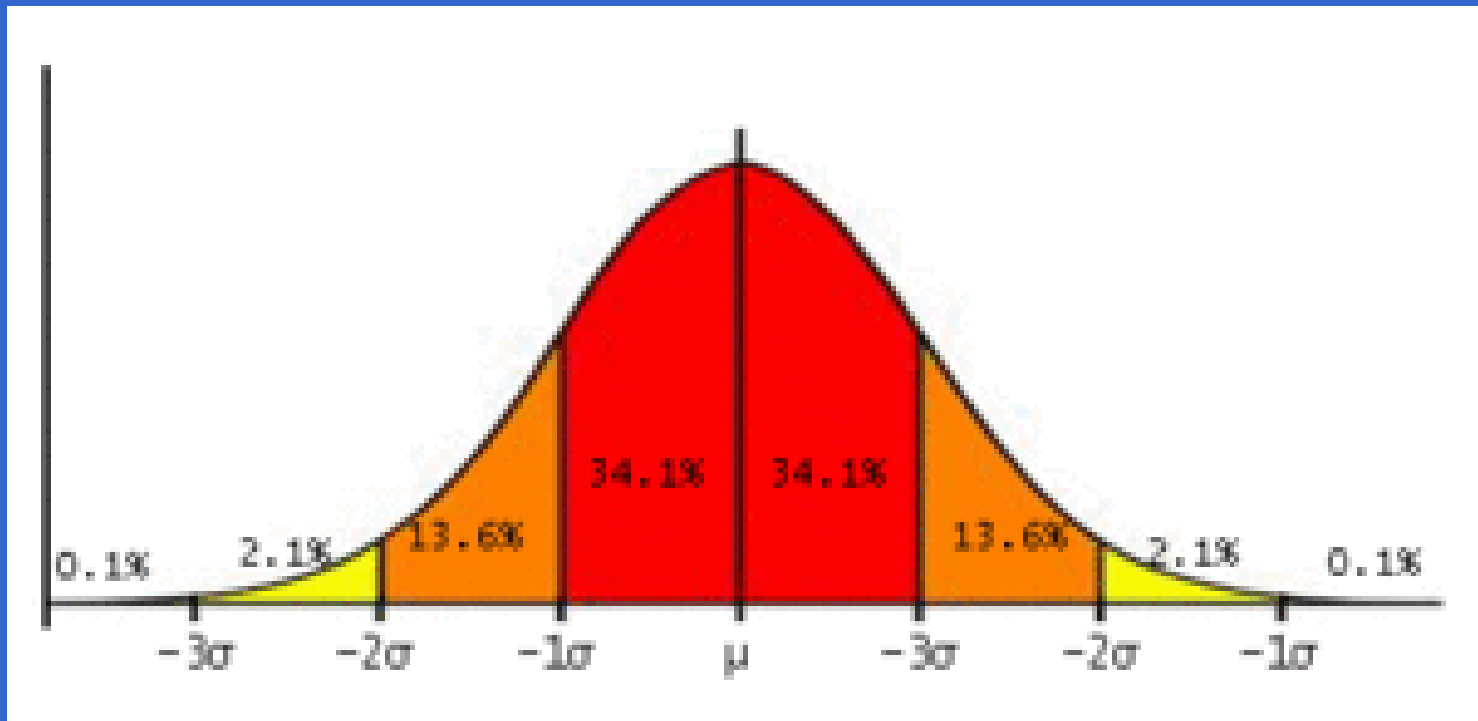
✓ Simétrica em torno da média ( $\mu$ )

- área antes de  $\mu$  = área depois de  $\mu$  = 0,5
- média = mediana = moda

✓ Varia de  $-\infty$  a  $+\infty$



- Parâmetro de locação ou localização:  $\mu$
- Parâmetro de escala:  $\sigma$  ( $\sigma^2$ )



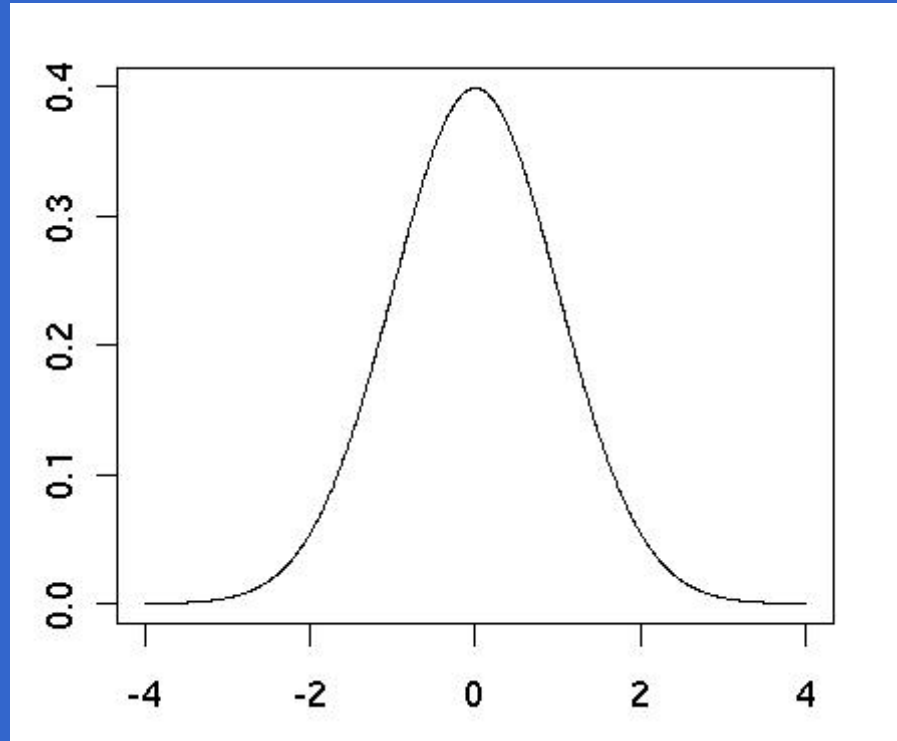
- Áreas de intervalos

$$\sqrt{\mu \pm \sigma} \approx 68\%$$

$$\sqrt{\mu \pm 2\sigma} \approx 95\%$$

$$\sqrt{\mu \pm 3\sigma} \approx 99,7\%$$

# Distribuição Normal Padrão



- $Z \sim N(0, 1)$

√ Média ( $\mu$ ) = 0

√ Desvio-padrão ( $\sigma$ ) = 1

$$Z = (X - \mu) / \sigma$$

- Valores de área tabelados



# Distribuição Normal – Cálculo de Probabilidades

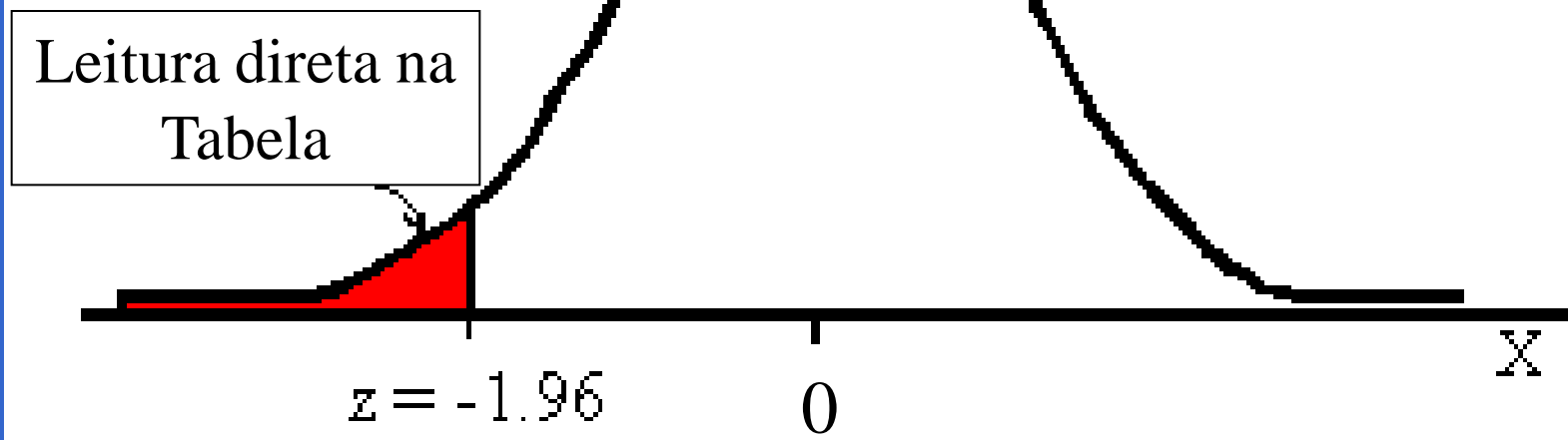
- Seja a variável aleatória  $Z \sim N(0, 1)$ 
  - √ Calcule  $P\{Z < -1,96\}$
  - √ Roteiro:
    - Esboce a curva normal
    - Trace uma linha para  $z = -1,96$
    - Verifique a área que se deseja calcular
    - Determine a área a partir da tabela



√ Área sob a curva para  $Z < -1,96$ :

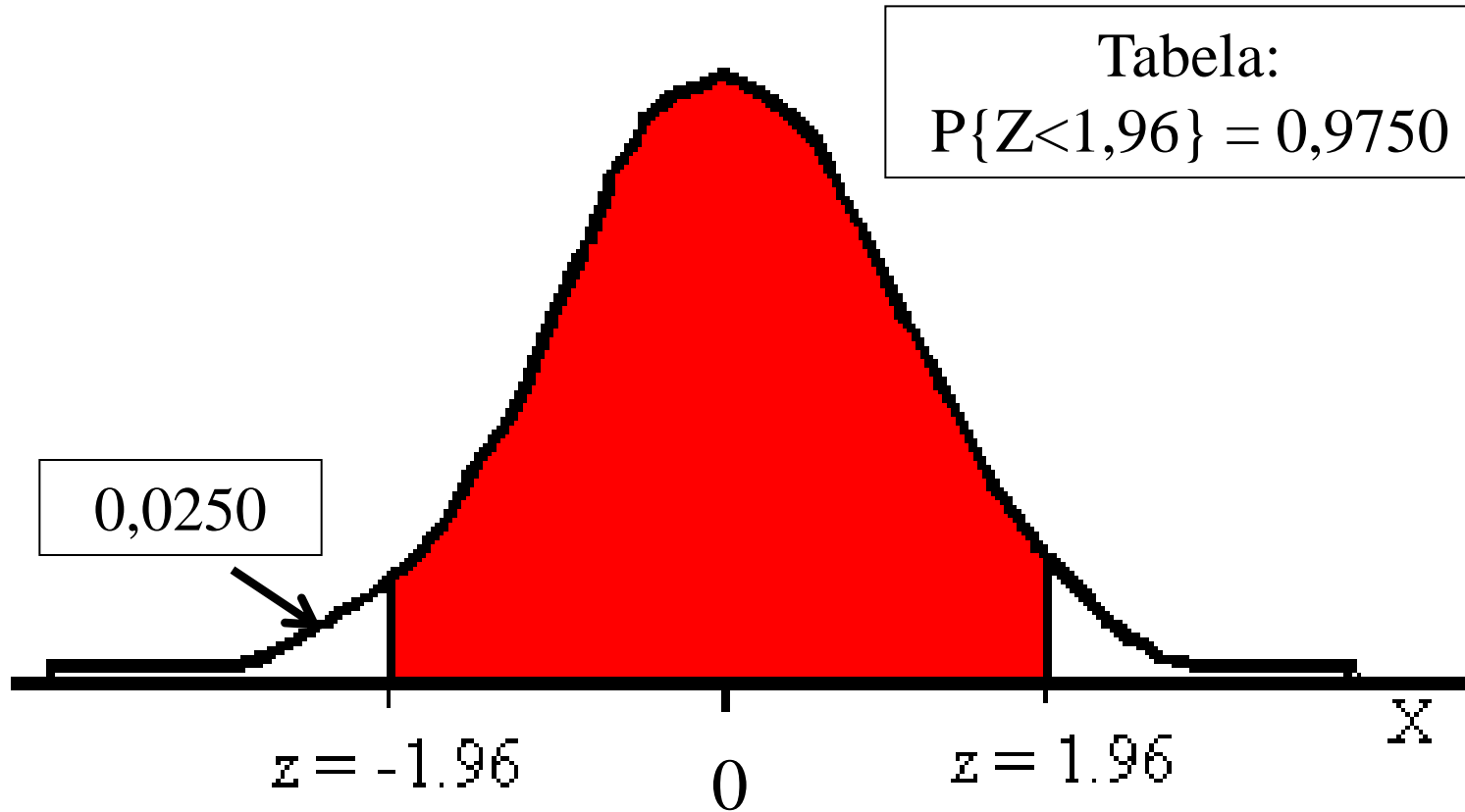
<i>z</i>	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	0.00	<i>z</i>
-3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	-3.80
-3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	-3.70
-3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	-3.60
-3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	-3.50
-3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	-3.40
-3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	-3.30
-3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	-3.20
-3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	-3.10
-3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	-3.00
-2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	-2.90
-2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	-2.80
-2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	-2.70
-2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	-2.60
-2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	-2.50
-2.40	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	-2.40
-2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	-2.30
-2.20	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	-2.20
-2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	-2.10
-2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	-2.00
-1.90	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	-1.90
-1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	-1.80
-1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	-1.70
-1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	-1.60





$$\sqrt{P \{Z < -1,96\} = 0,0250}$$

√ Calcule  $P\{-1,96 < Z < 1,96\}$

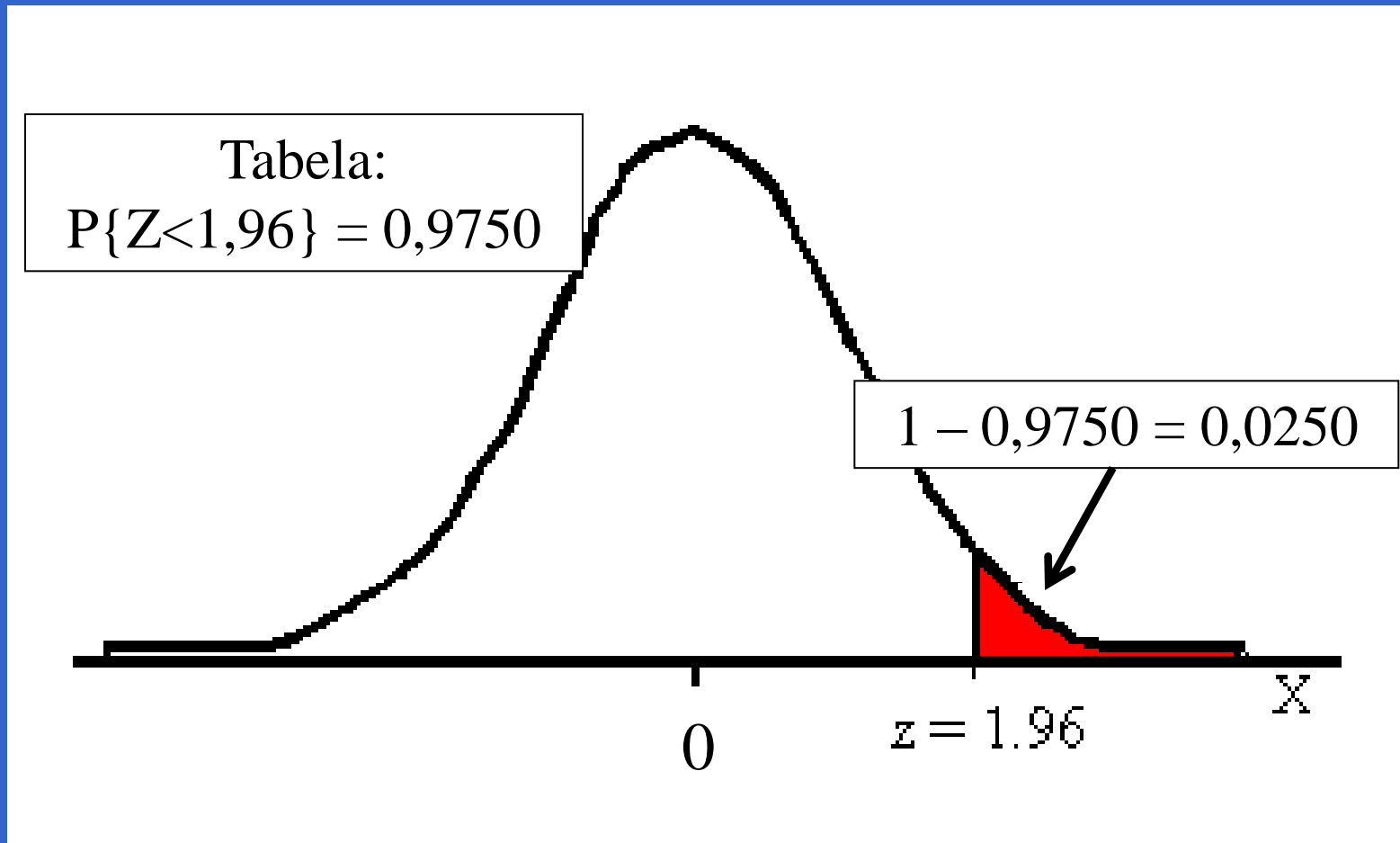


$$P\{-1,96 < Z < 1,96\} = 0,9750 - 0,0250 = 0,9500$$

- Área sob a curva para  $Z < 1,96$

<b>z</b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>	<b>z</b>
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0.90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1.00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40

- Calcule  $P\{Z > 1,96\}$

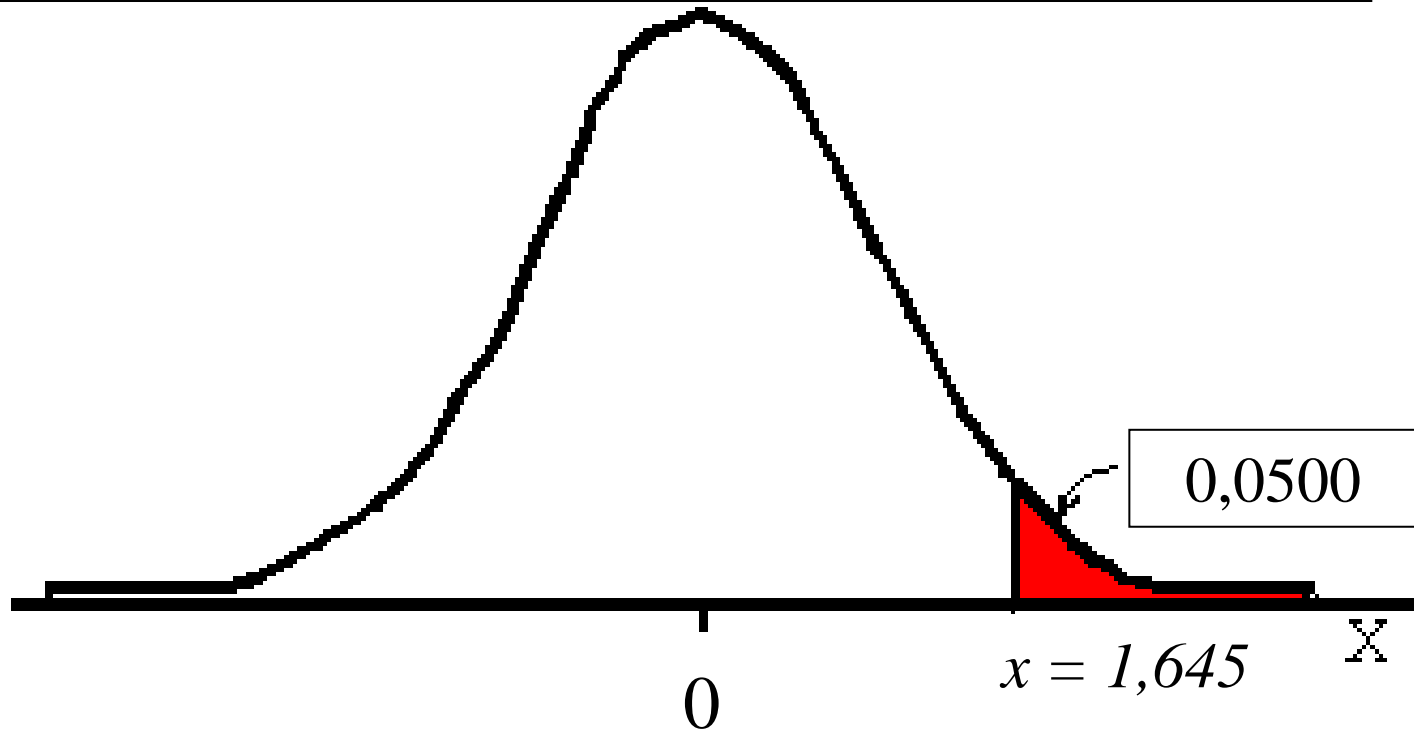


- Probabilidade contida em alguns intervalos

Intervalo	Probabilidade	
$-1 < Z < 1$	$0,8413 - 0,1587 =$	0,6826
$-2 < Z < 2$	$0,9772 - 0,0228 =$	0,9544
$-3 < Z < 3$	$0,9987 - 0,0013 =$	0,9974

- Determinar  $x$ , tal que  $P\{Z > x\} = 0,05$

Tabela: Valor mais próximo de  $P\{Z < x\} = 0,9500$



$$\sqrt{P\{Z < 1,65\}} = 0,9505$$

$$\sqrt{P\{Z < 1,64\}} = 0,9495$$

# Intervalos Simétricos em Torno de Zero

Probabilidade	Intervalo
90%	$-1,645 < Z < 1,645$
95%	$-1,96 < Z < 1,96$
99%	$-2,58 < Z < 2,58$

# Outras Distribuições Normais

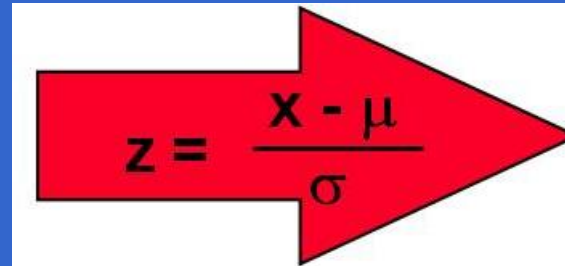
- Caso Geral:
  - √ Média:  $\mu$
  - √ Desvio-padrão:  $\sigma$
- Transformação:

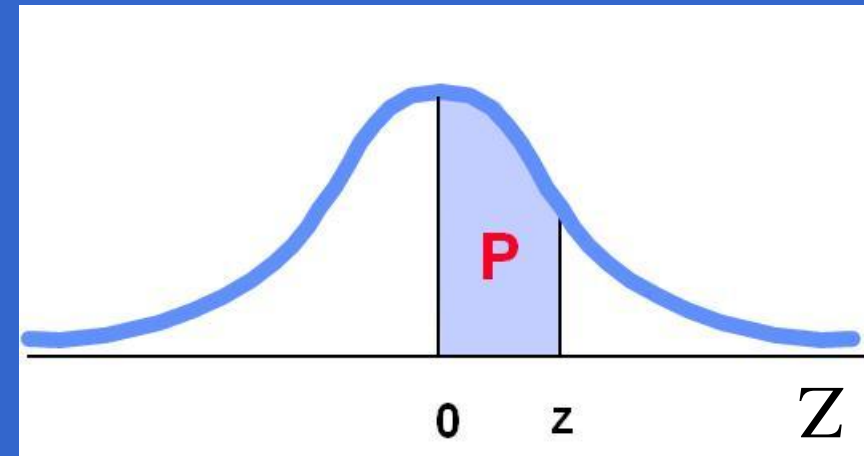
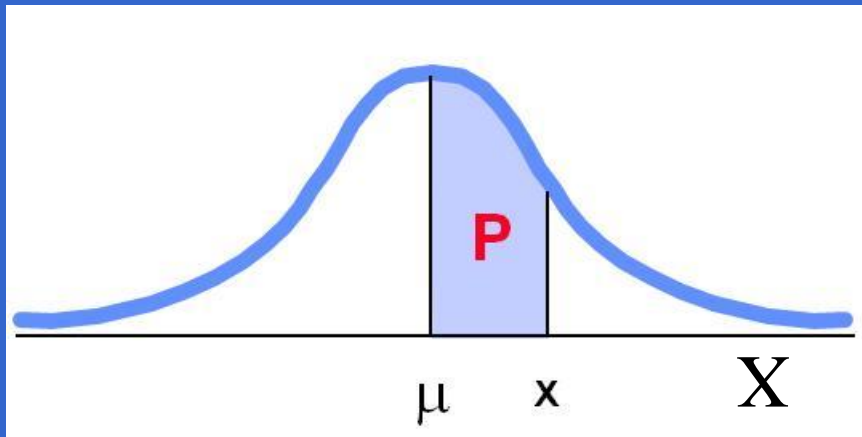
$$Z = \frac{(X - \mu)}{\sigma}$$

- Mesmos procedimentos após transformação (tabela Normal Padrão)



# Conversão na Normal Padrão


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



- $P\{\mu < X < x\} = P\{0 < Z < z\}$

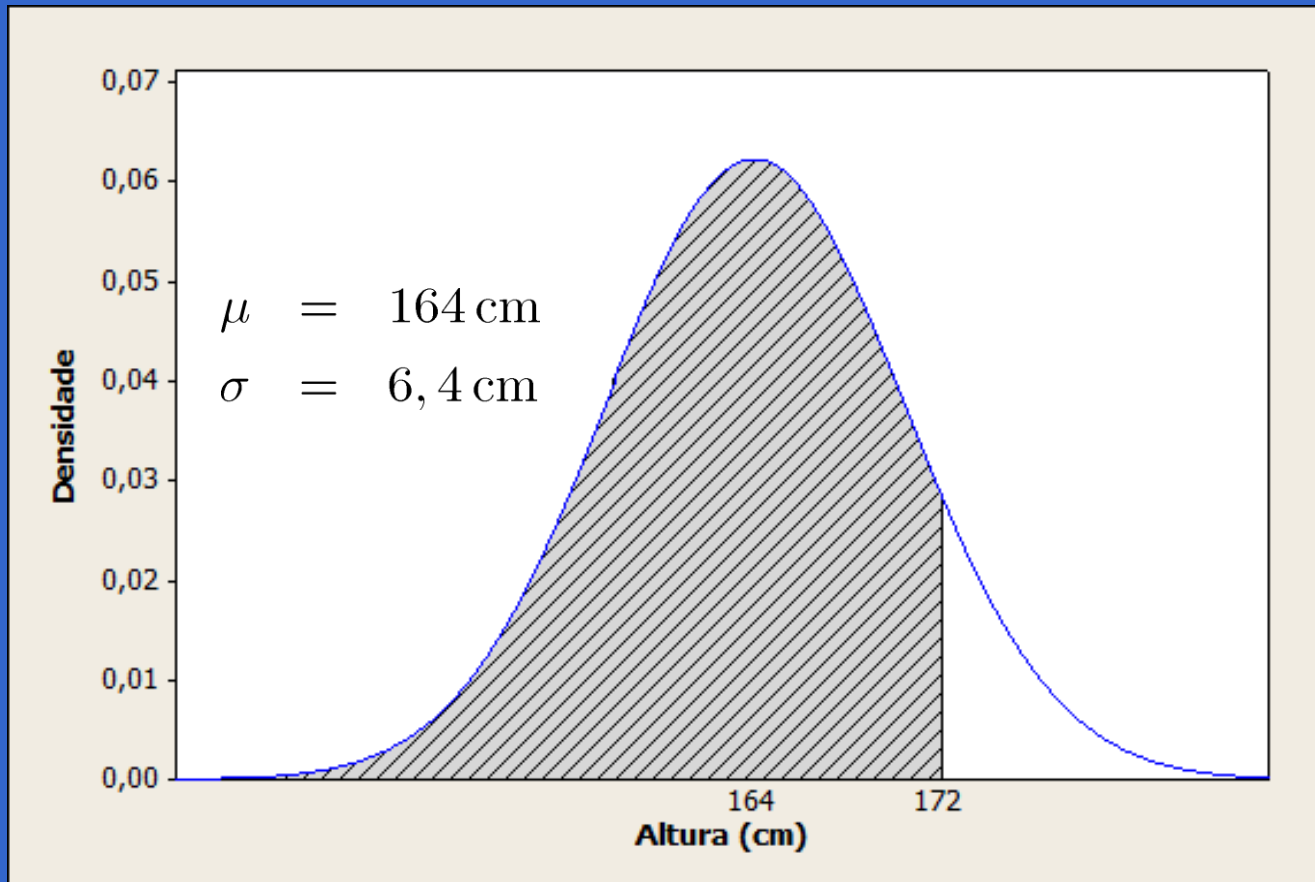
# Exemplo

- As alturas de mulheres com 18 a 24 anos de idade é aproximadamente normal com média 164 cm e desvio-padrão 6,4 cm.

$\sqrt{X}$ : altura de mulheres entre 18 e 24 anos (cm)

$$X \sim N(164, 6,4)$$

- ✓ Encontre a proporção de mulheres com altura inferior a 172 cm



## √ Padronização

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{172 - 164}{6,4} = 1,25$$

## √ Pela tabela

$$P \{ Z < 1,25 \} = 0,8944$$

$$P \{ X < 100 \} = 0,8944 = 89,44\%$$

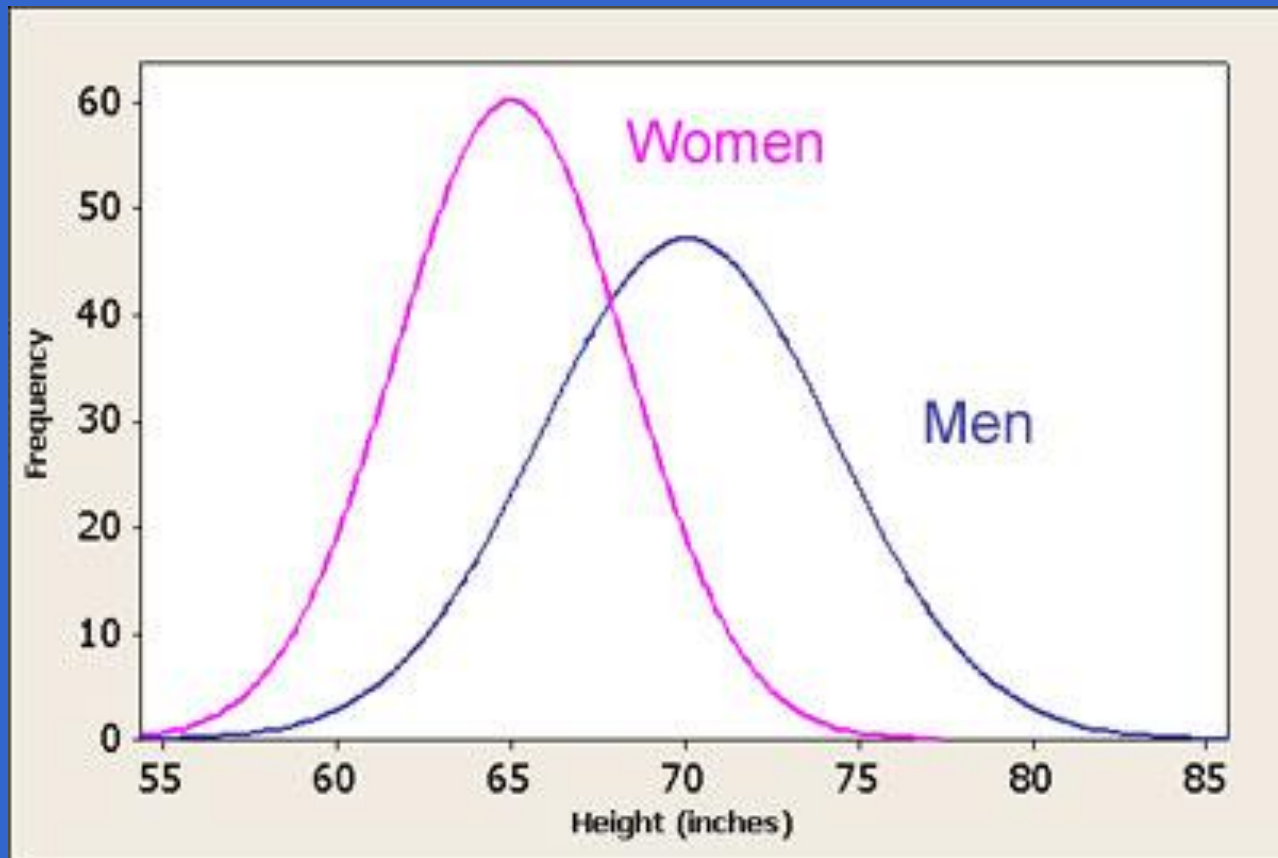
- Qual o valor de altura que delimita 5% das mulheres mais altas?

$$X = \mu + Z\sigma = 164 + 1,645(6,4) = 174,5\text{cm}$$

# Aplicações da Distribuição Normal

- Usada como um modelo para estudar uma grande variedade de variáveis
  - √ Objetivo: responder questões sobre probabilidades relacionadas com essas variáveis
- Exemplos:
  - √ Altura humana
  - √ Inteligência
  - √ Em várias situações quando a resposta é influenciada por diversos fatores

# Exemplo para altura



# Atividade nº 9



# Alguns pontos sobre a distribuição Normal

- A distribuição normal é um **modelo**, ou seja, a aproximação de um processo aleatório, e não uma regra;
- Nem todas as distribuições empíricas podem ser aproximadas por uma Normal; (por ex. preços de imóveis, tempos entre chegadas, precipitação diária, gastos em planos de saúde, tempos de estadia em um hotel)

# Alguns pontos sobre a distribuição Normal (cont.)

- Área sob a curva: metáfora do grão de areia ou do pixel em uma foto ou tela. (Ler o artigo da Atividade 11, cuja referência está no final e responder as perguntas formuladas)

# Atividade nº 10

# Atividade nº 11

# Referências

- AGRESTI, A.; FINLAY, B. *Métodos estatísticos para as ciências sociais*. Penso, 2012.
- LEE.,J.T. e LEE, H.S. Visual representations of empirical probability distributions when using the granular density metaphor. In K. Makar, B. de Sousa e R. Gould (eds), *Sustainability in Statistics Education*. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona, USA. The Netherlands: ISI/IASE, July 2014.
- MAGALHÃES, M.N.; LIMA, A.C.P.L. *Noções de Probabilidade e Estatística*. Edusp, 2011.
- MOORE, D. S.; MCCABE, G. P. *Introdução à prática da estatística*. LTC, 2002.
- WILD, C.J. E SEBER, G.A.F. *Encontros com o acaso: um primeiro curso de análise de dados e inferência*. LTC, 2000.