Licenciatura em Ciências Biológicas
Universidade Federal de Goiás

Bioestatística

Prof. Thiago Rangel - Dep. Ecologia – ICB
rangel.ufg@gmail.com

Página do curso:
http://www.ecologia.ufrgs.br/~adrimelo/bioestat

Distribuição Normal

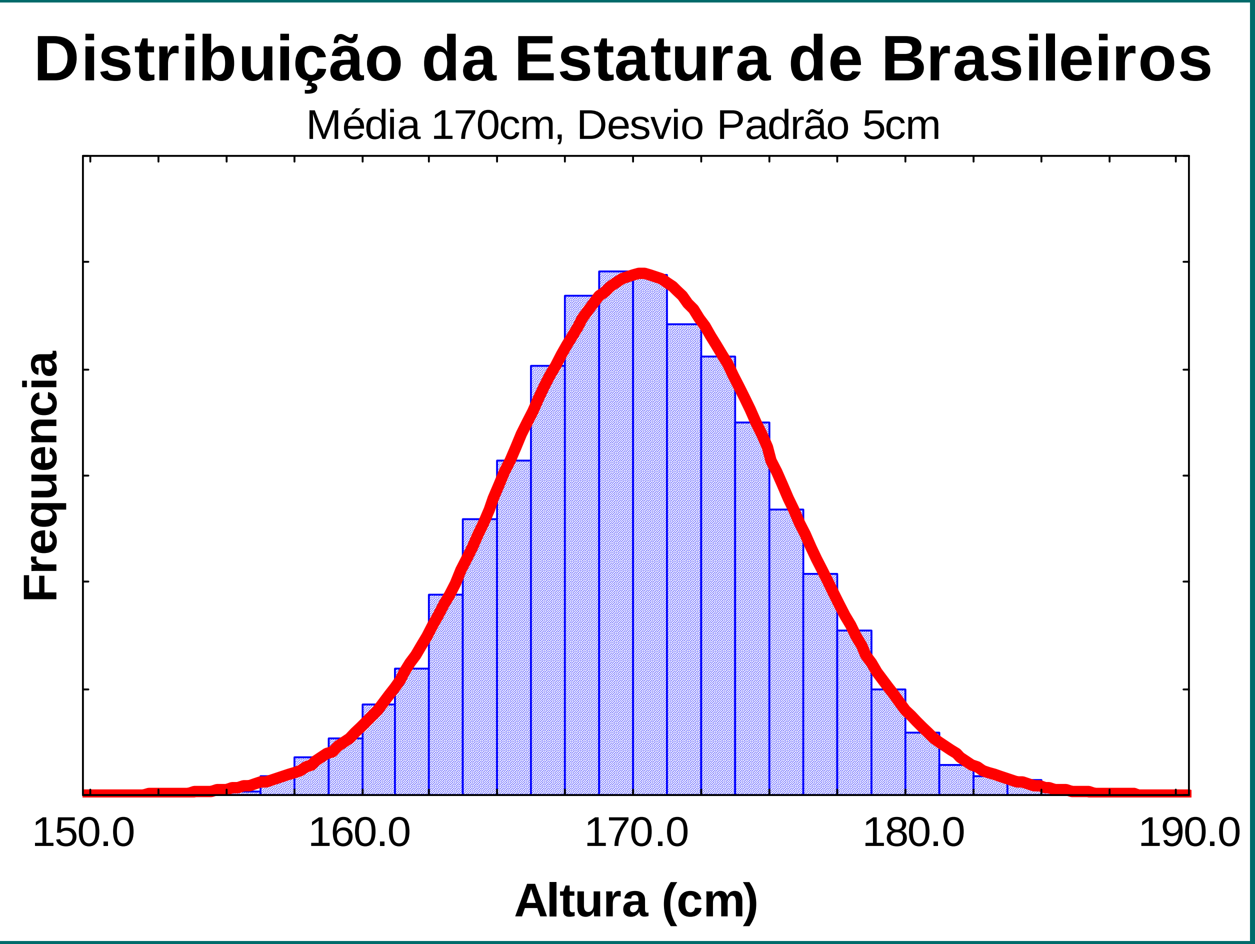
"distribuição de Gauss"
"distribuição Gaussiana"

Em aulas anteriores havíamos visto como gerar uma distribuição de uma dada estatística a partir de nossos próprios dados usando aleatorização.

Entretanto, também é possível utilizar distribuições *teóricas* como referência para o cálculo de probabilidades.

A Distribuição Normal

Descreve fenômenos que são determinados por múltiplas causas, que em geral interagem entre si



Algumas características da Normal

→ A distribuição normal possui a bem característica forma de sino

→ A distribuição normal se estende de -∞ até +∞

→ Portanto, toda população está debaixo da curva. Assim, dizemos que a área sob a curva é 1 ou 100%

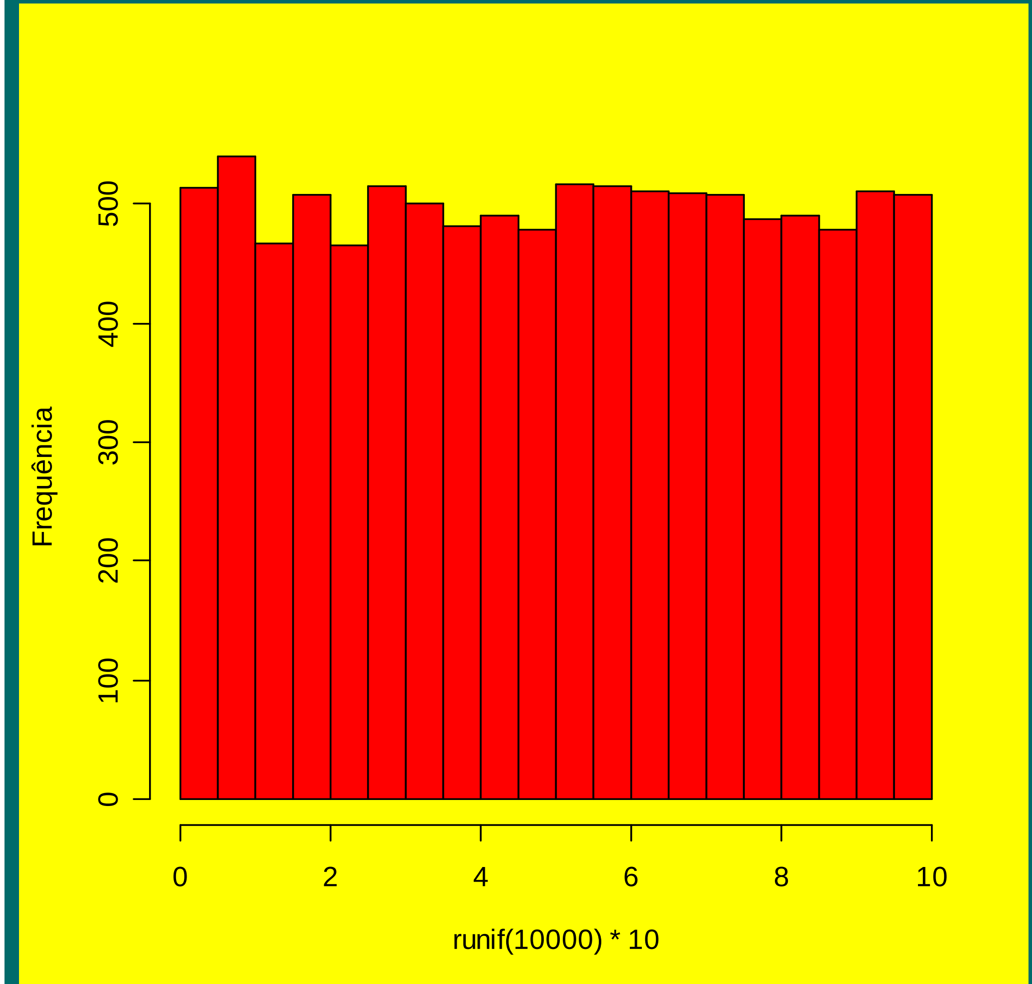
→ A curva normal é definida por dois parâmetros:

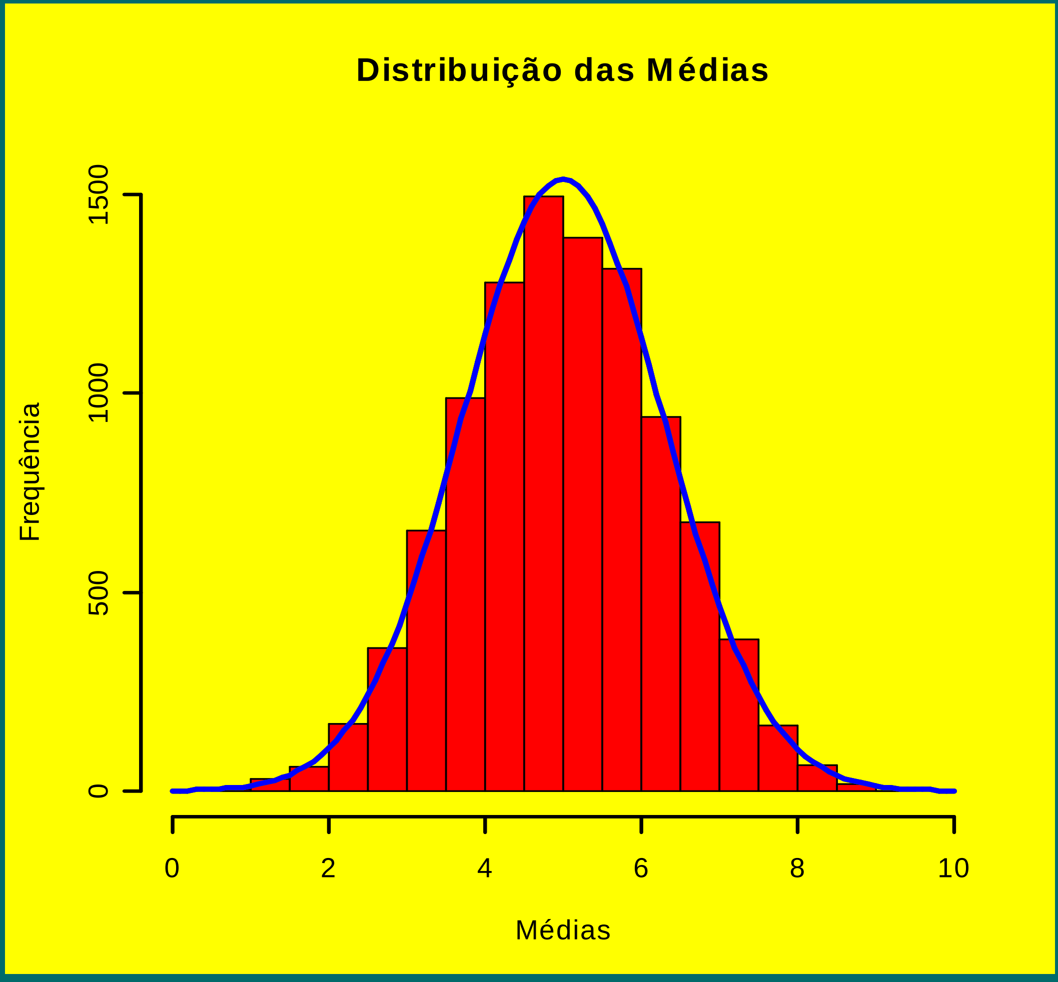
a média μ (lê-se “mi”) e o desvio padrão σ (lê-se “sigma”)

→ A média, mediana e a moda coincidem no centro da distribuição, pois a distribuição é simétrica

A Distribuição Normal

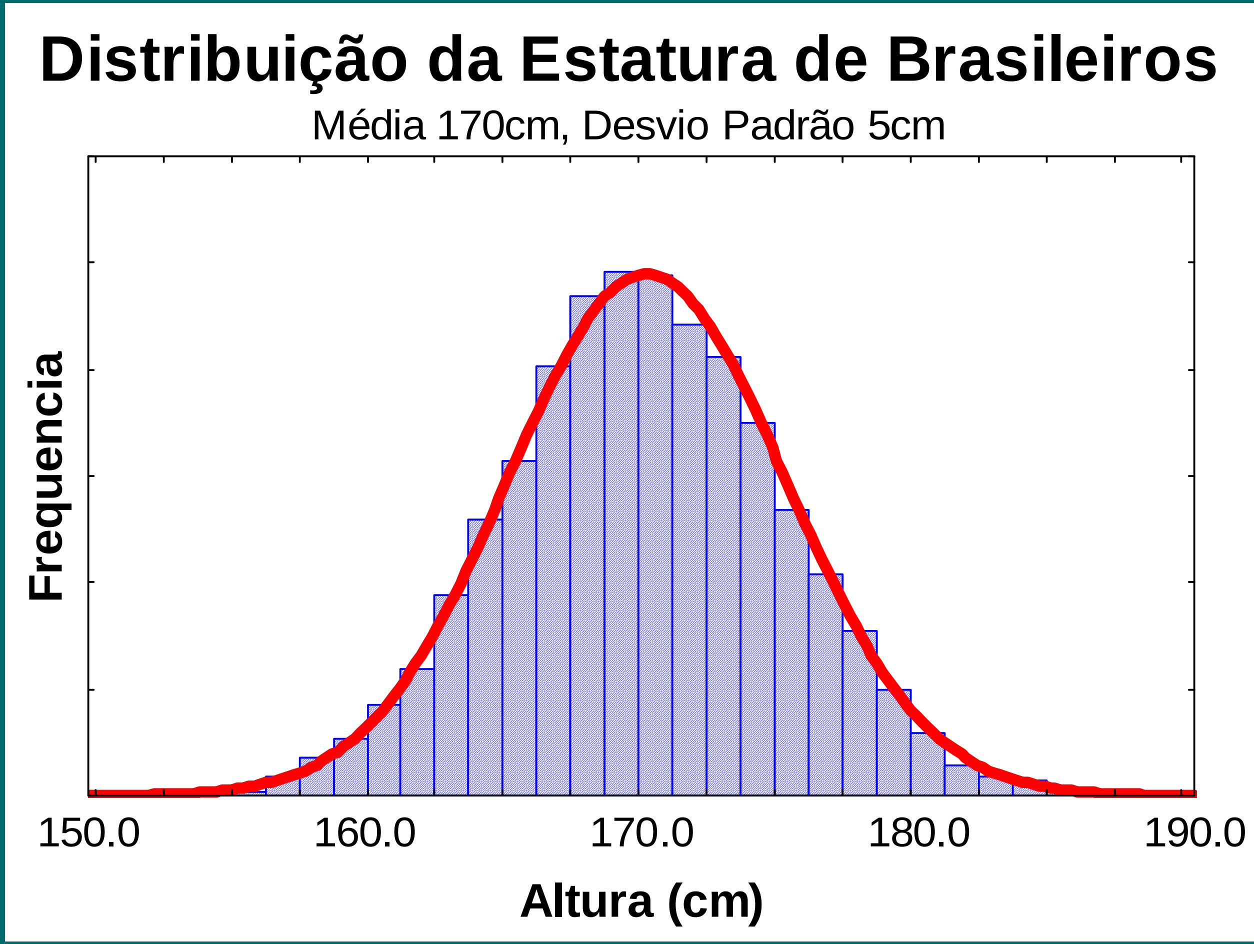
Também podemos gerar uma distribuição normal quando tiramos médias de números sorteados de outras distribuições!





Informações úteis podem ser extraídas do gráfico abaixo:

Se a média da estatura dos brasileiros é 170cm, sabemos que existe 50% de chance de que um indivíduo amostrado aleatoriamente seja mais alto ou mais baixo do que 170cm



Distribuição Normal Padronizada

“Padronizar” significa remover a escala da variável original, em geral transformando-a em um índice.

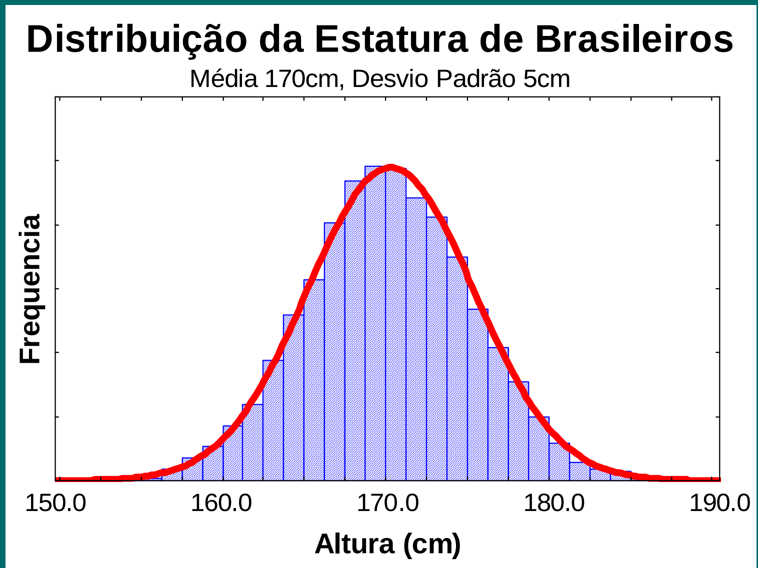
Índices são úteis pois permitem a comparação entre variáveis.

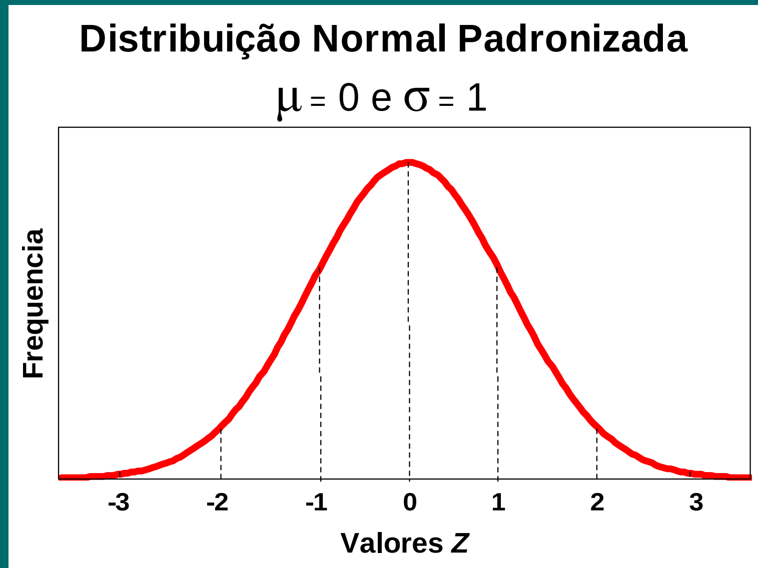
A principal transformação da curva normal é a *Z*. Para transformar sua variável no índice *Z* utiliza-se a seguinte formula:

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

Quando a variável é transformada em *z* ela passa a ter μ = 0 e σ = 1.

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





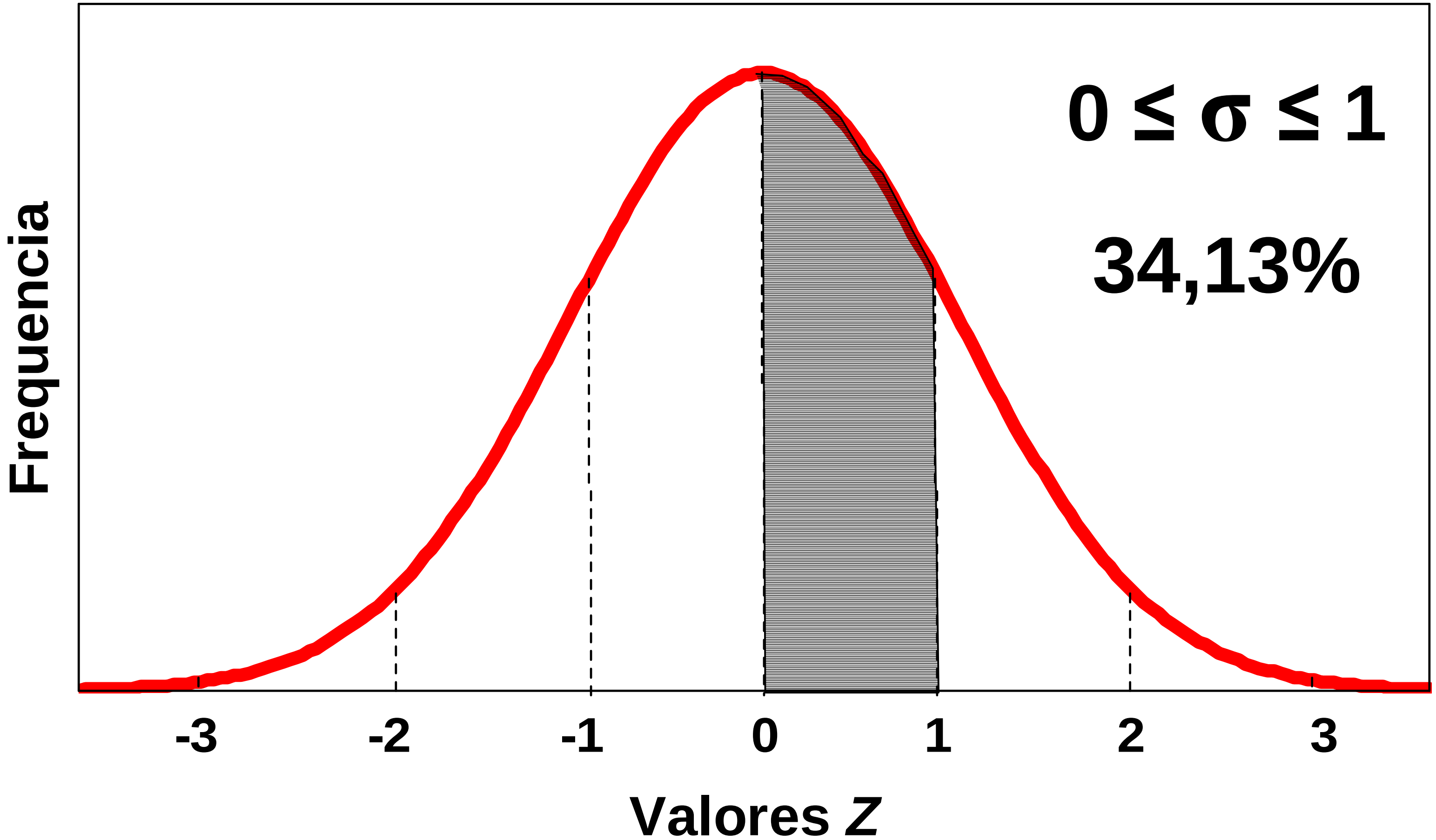
$$x = (z \times \sigma) + \mu$$

Uma amostra de dados reais nunca se ajusta perfeitamente a uma distribuição normal *teórica*, mas frequentemente ela se aproxima bem.

Quando esta aproximação for *razoável*, podemos fazer inferência de probabilidades sabendo apenas as estimativas amostrais de μ e σ, pois a área sob a curva é conhecida matematicamente.

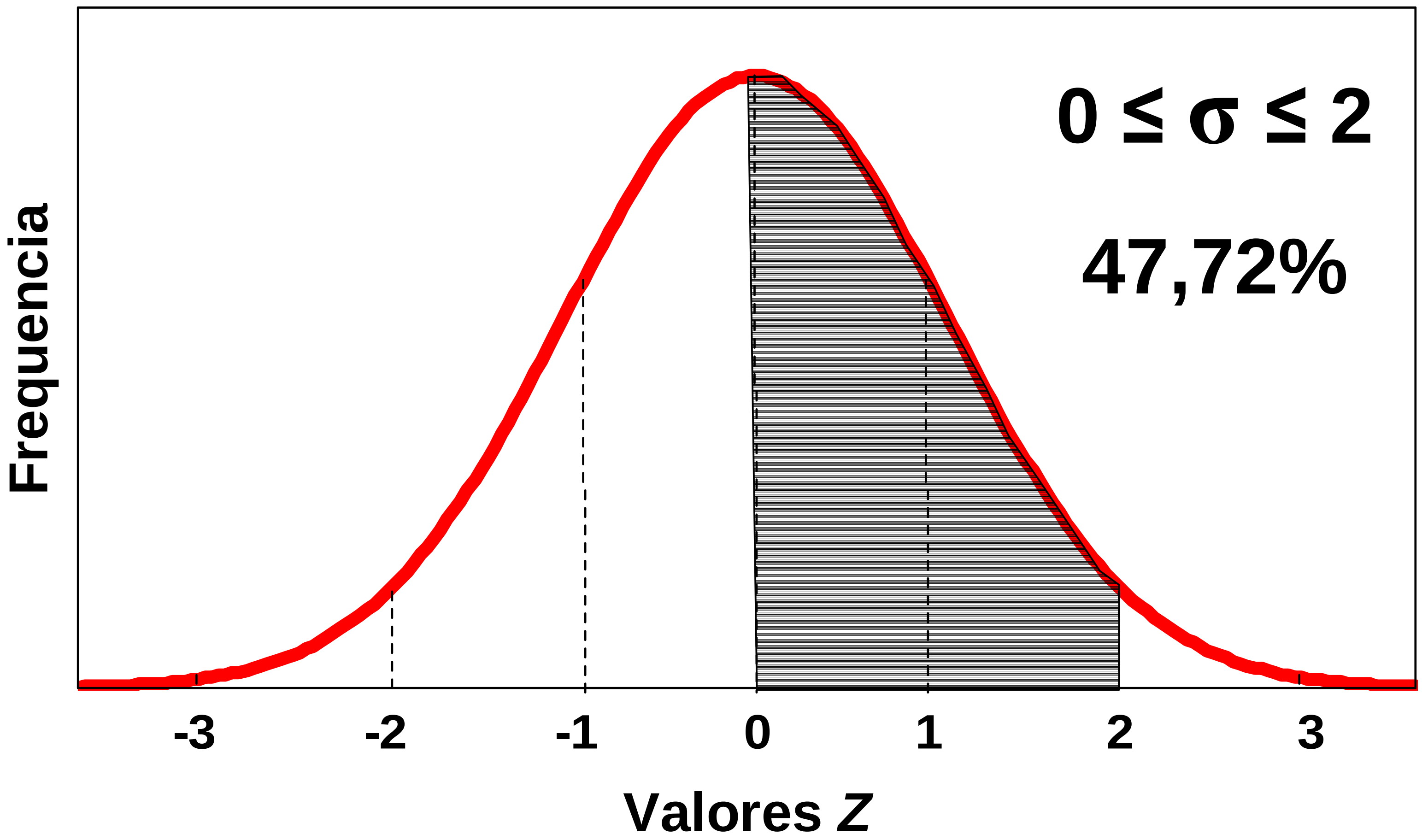
Distribuição Normal Padronizada

$\mu = 0$ e $\sigma = 1$



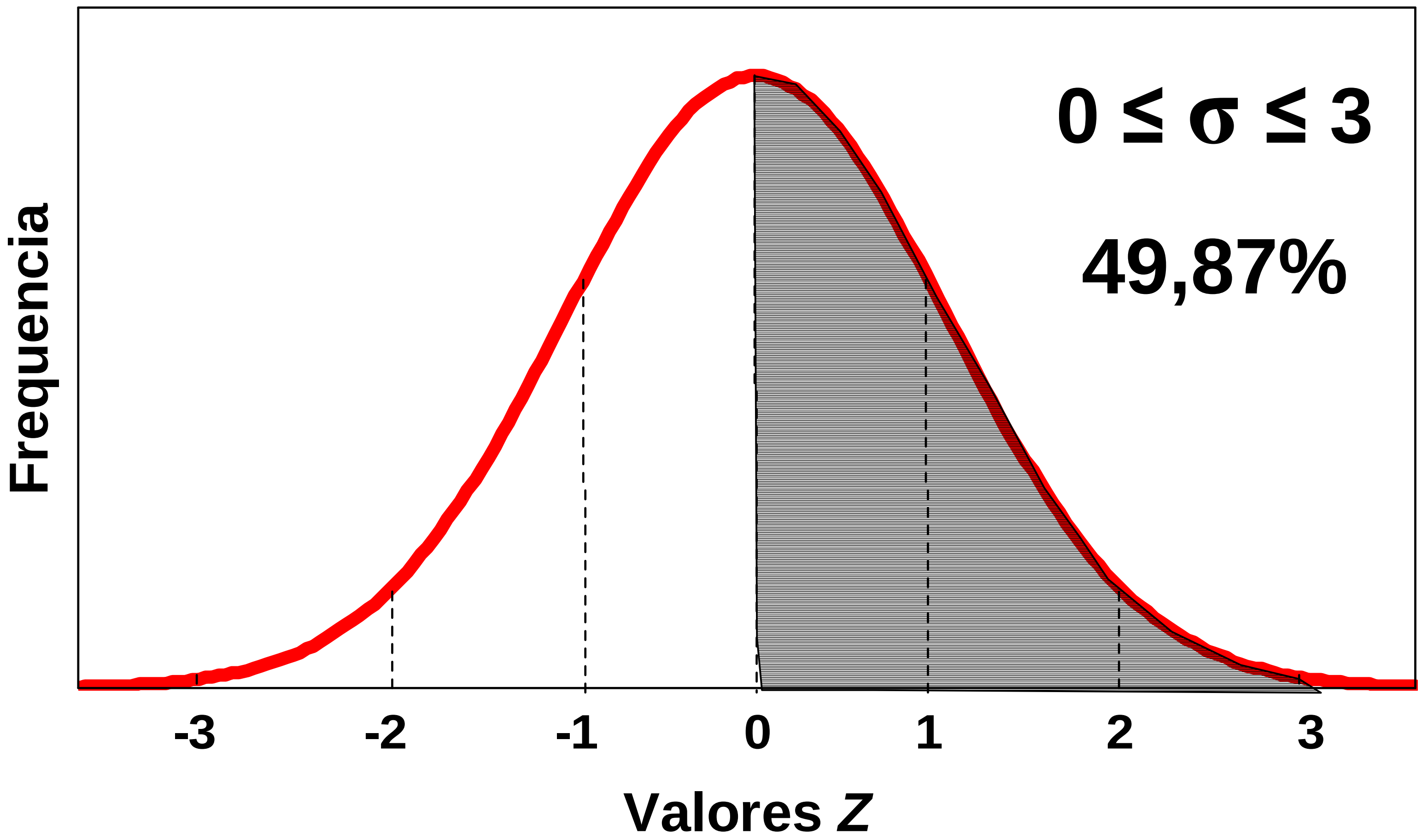
Distribuição Normal Padronizada

$\mu = 0$ e $\sigma = 1$



Distribuição Normal Padronizada

$\mu = 0$ e $\sigma = 1$

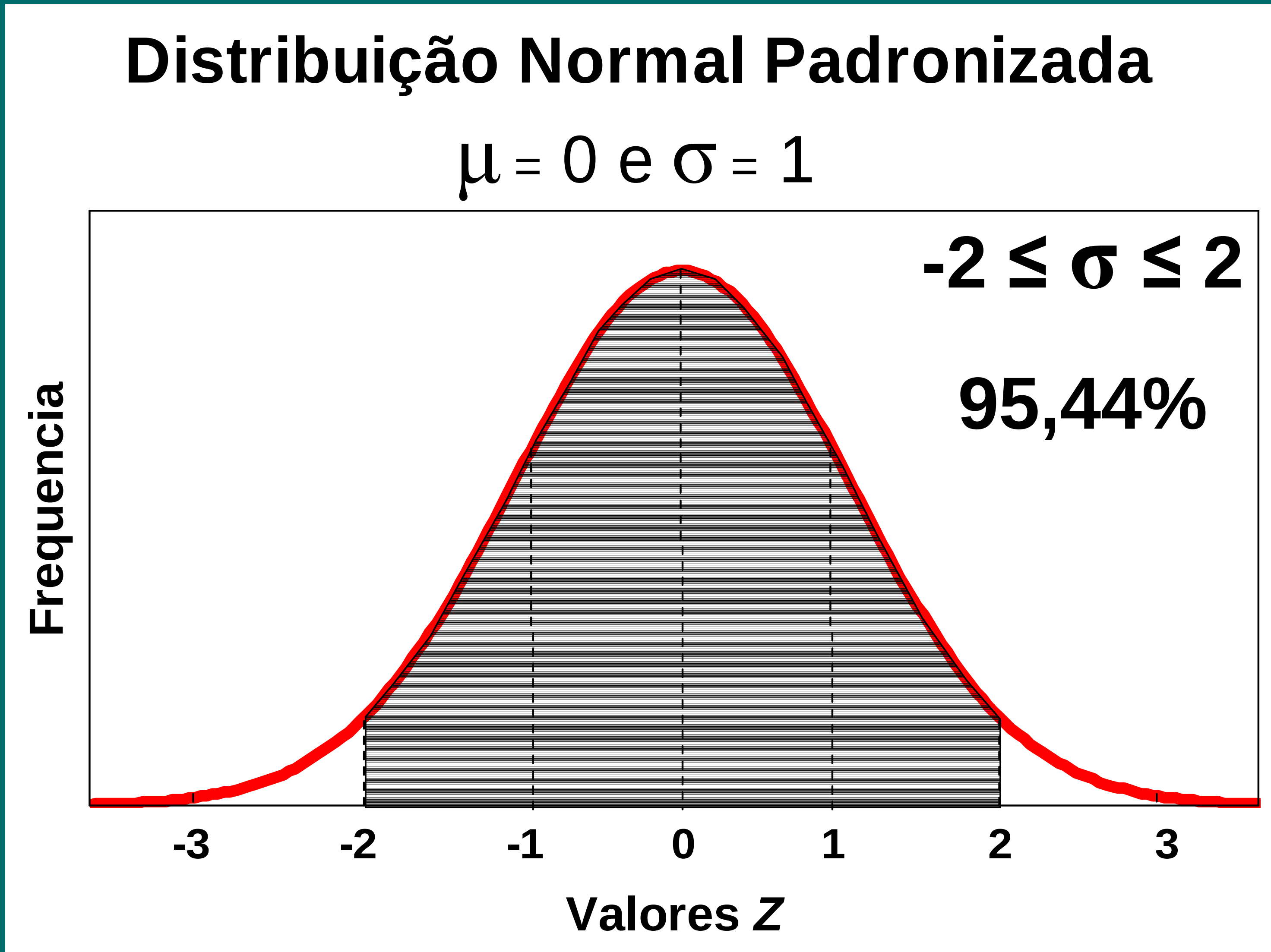


Onde eu encontro esta estimativa da "área sob a curva"?

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	**0,4495**	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	**0,4750**	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817

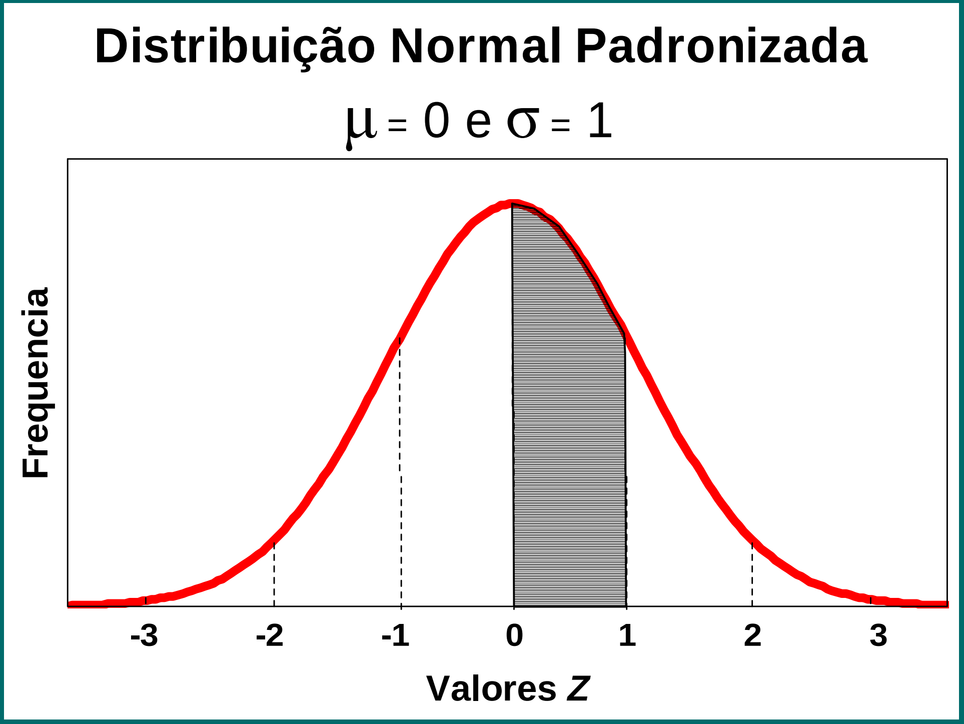
Já que a curva normal é simétrica, podemos incluir no intervalo valores abaixo da média.

Basta multiplicar o valor da área sob a curva por 2:

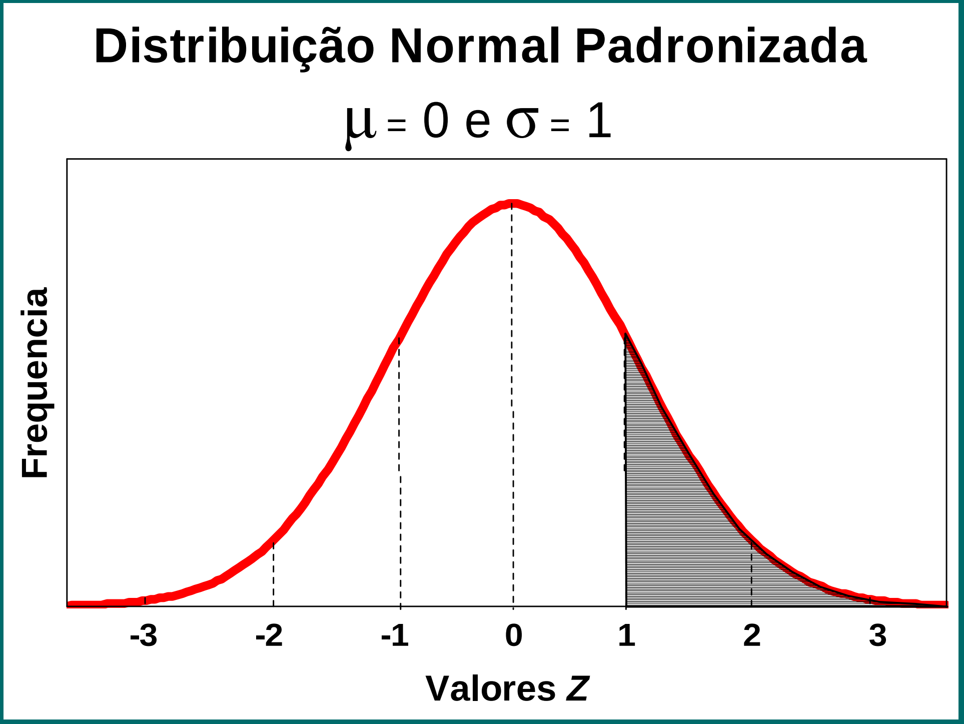


Usando um pouco de lógica e aritmética, podemos também calcular áreas distantes da média:

Se $(0 \leq \sigma \leq 1) = 34{,}13\%$



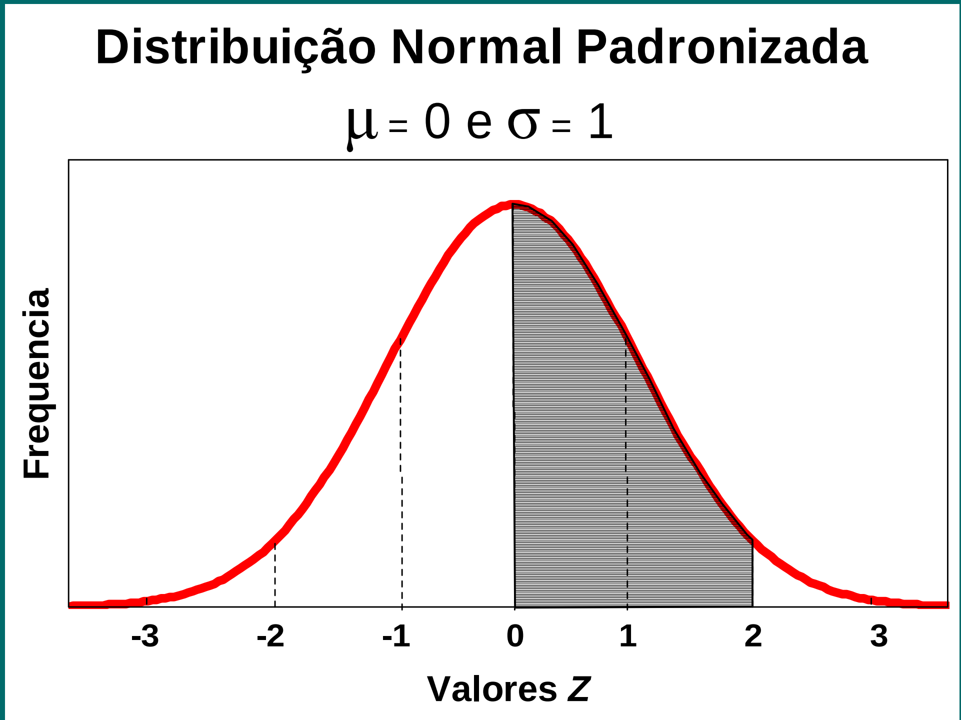
Então $(\sigma > 1) = 15{,}87\%$



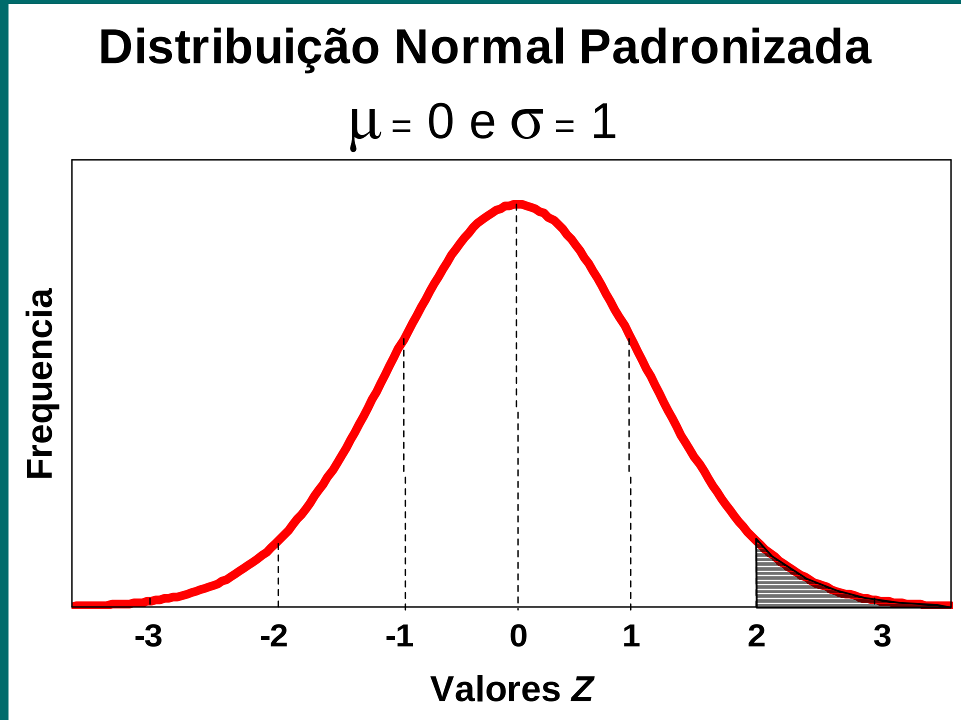
$$50{,}00 - 34{,}13 = 15{,}87$$

Da mesma maneira:

Se $(0 \leq \sigma \leq 2) = 47{,}72\%$



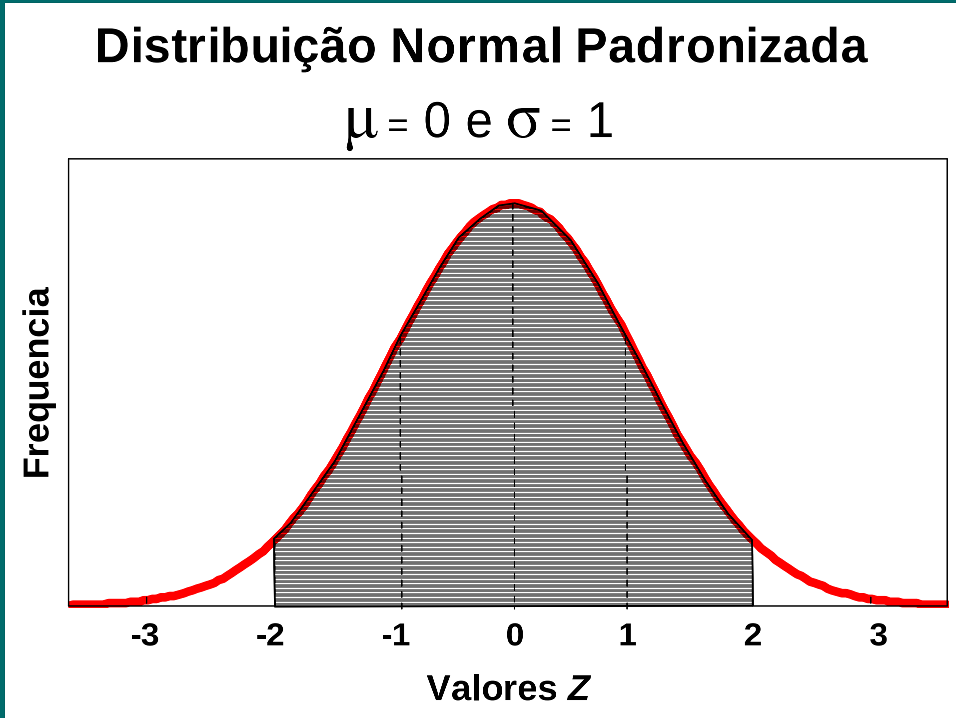
Então $(\sigma > 2) = 2{,}28\%$



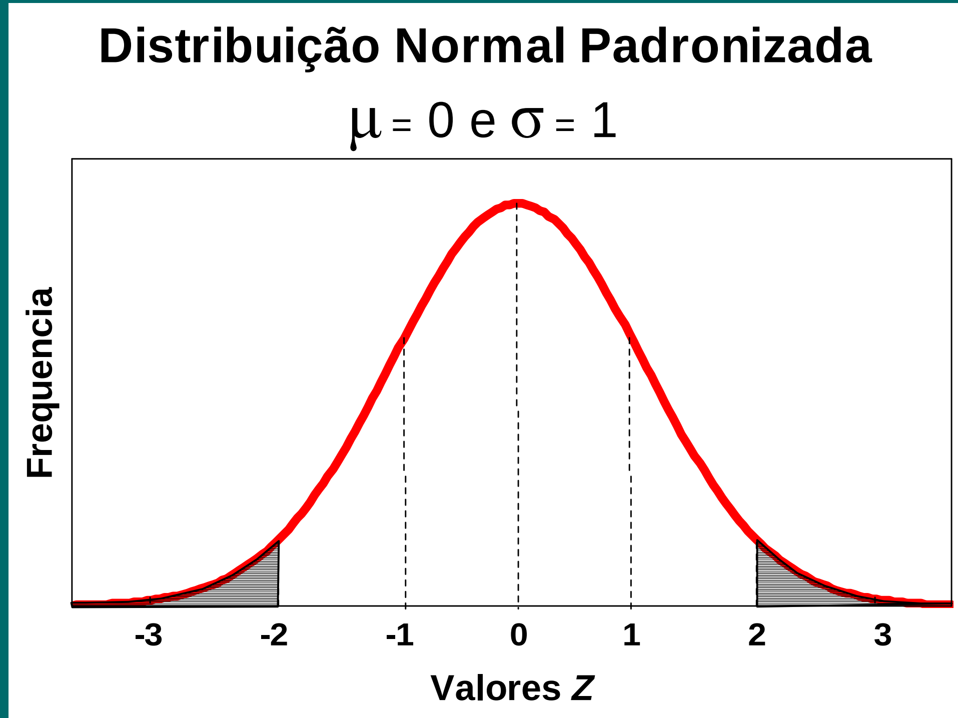
$50{,}00 - 47{,}72 = 2{,}28$

Utilizando a propriedade da simetria:

Se $(-2 \leq \sigma \leq 2) = 95{,}44\%$



Então $(\sigma > \pm 2) = 4{,}56\%$



$$100{,}00 - 95{,}44 = 4{,}56 = [2*(50-47{,}72)]$$

Exemplo 1:

A estatura de brasileiros apresenta distribuição normal, sendo μ = 170cm e σ = 5cm.

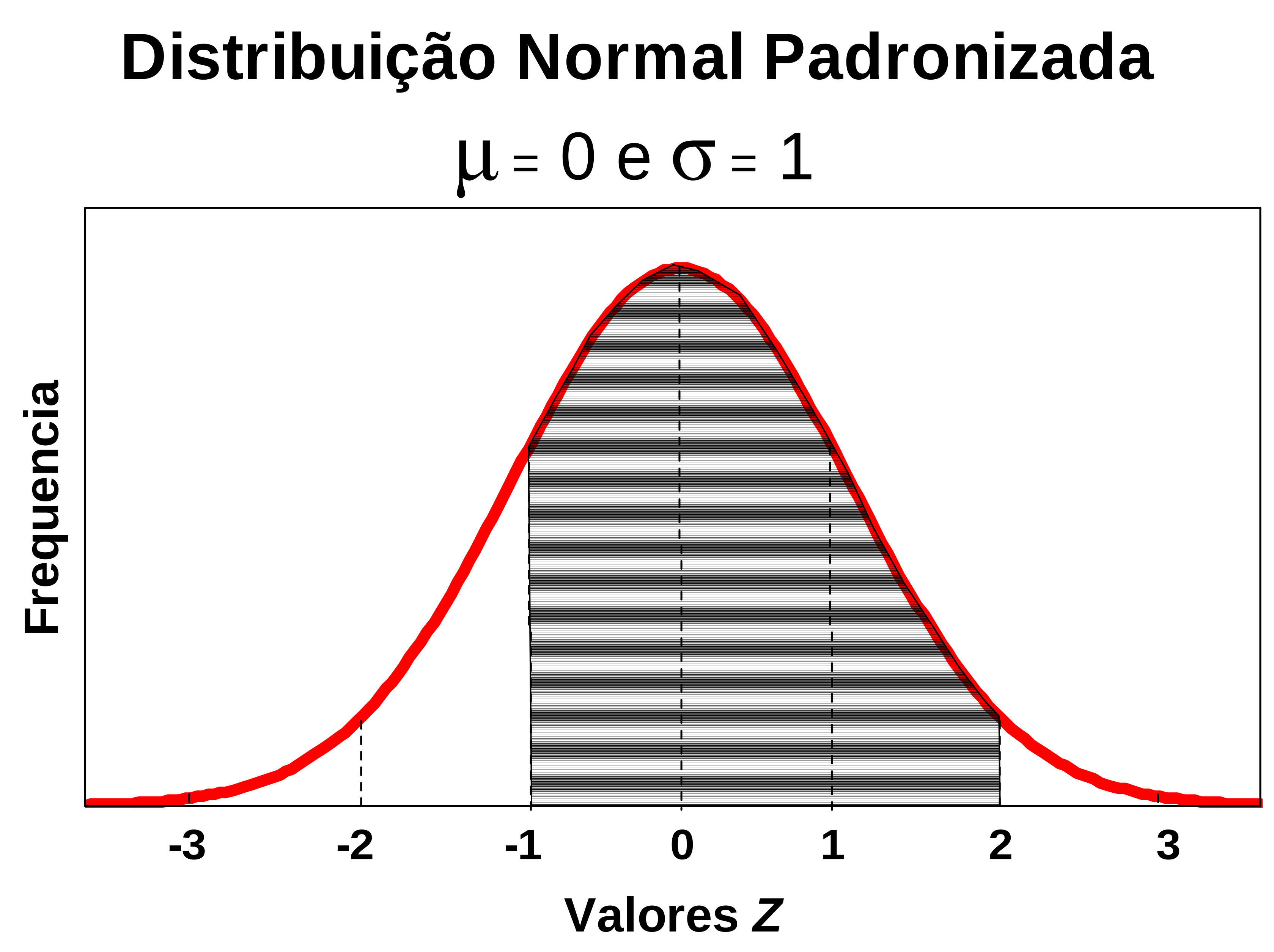
Calcule a probabilidade de um homem apresentar estatura entre 165cm e 180cm.

$$z = \frac{165 - 170}{5} = -1$$

$$z = \frac{180 - 170}{5} = 2$$

Portanto queremos saber a área sob a curva que vai de *z* = -1 até *z* = 2.

Exemplo 2:


Distribuição Normal Padronizada
μ = 0 e σ = 1
Frequencia
-3
-2
-1
0
1
2
3
Valores Z

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	**0,4495**	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	**0,4750**	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817

$0 > \sigma > -1 = 34,13\%$ $\quad$ $0 < \sigma < 2 = 47,72\%$

Portanto: 34,13 + 47,72 = 81,85%

Exemplo 2:

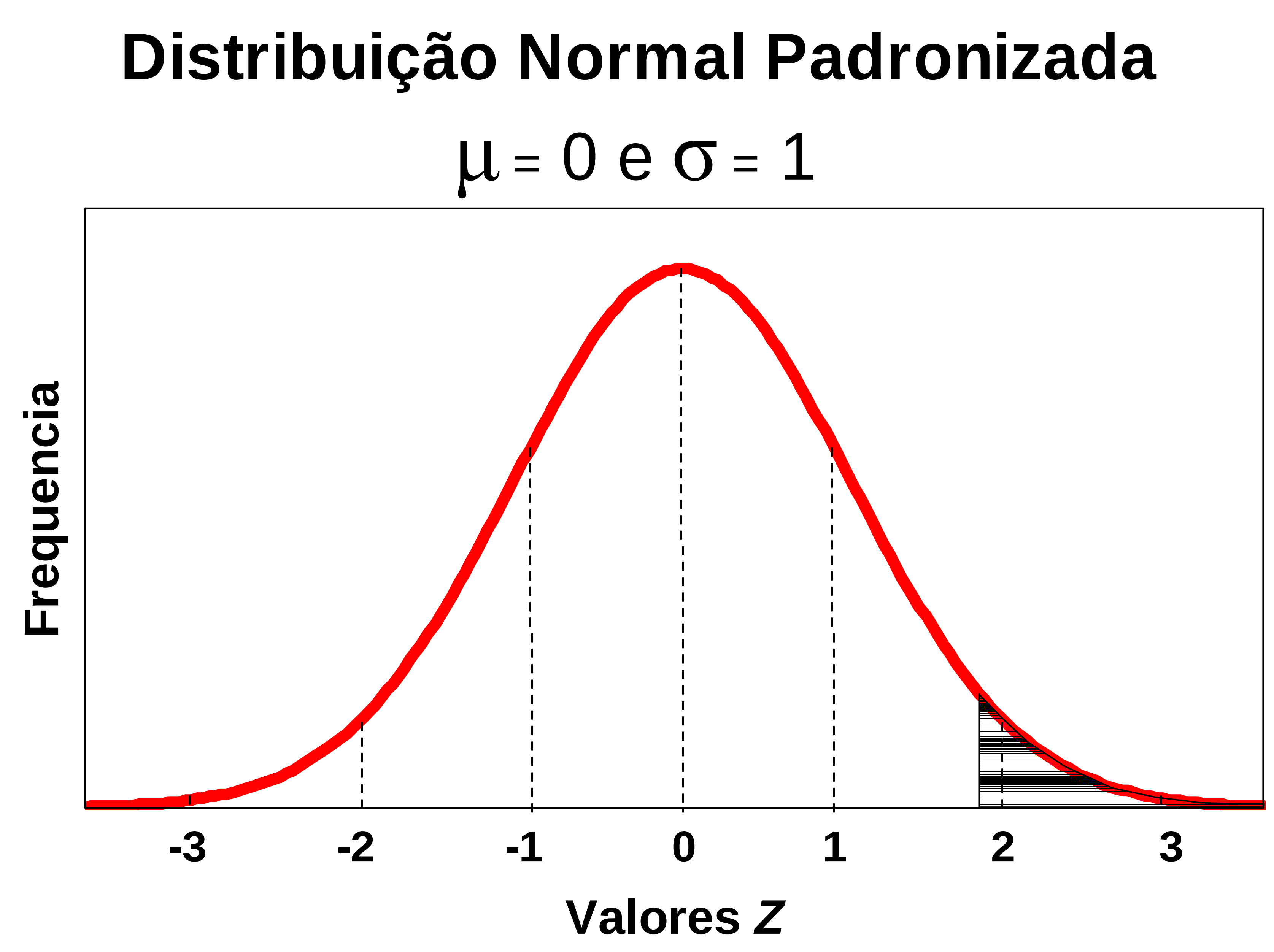
A estatura de brasileiros apresenta distribuição normal, sendo μ = 170cm e σ = 5cm.

Calcule a probabilidade de um homem apresentar estatura maior do que 179,8cm.

$$z = \frac{179{,}8 - 170}{5} = 1{,}96$$

Portanto queremos saber a área sob a curva que vai além de $z > 1{,}96$

Exemplo 2:


Distribuição Normal Padronizada
$\mu = 0$ e $\sigma = 1$
Frequencia
-3
-2
-1
0
1
2
3
Valores *Z*

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	**0,4495**	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	**0,4750**	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817

$0 = \sigma < 1{,}96 = 47{,}5\%$ $\quad$ $50 - 47{,}5 = 2{,}5\%$

Portanto: 2,5%

Exercício 1:

O tempo médio para completar a prova de bioestatística é 58 minutos, e o desvio padrão é 9,5 minutos. Se o professor Adriano quiser que apenas 90% dos alunos terminem a prova, quanto tempo deve da-los? Lembrando...

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

tal que

$$x = (z \times \sigma) + \mu$$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177

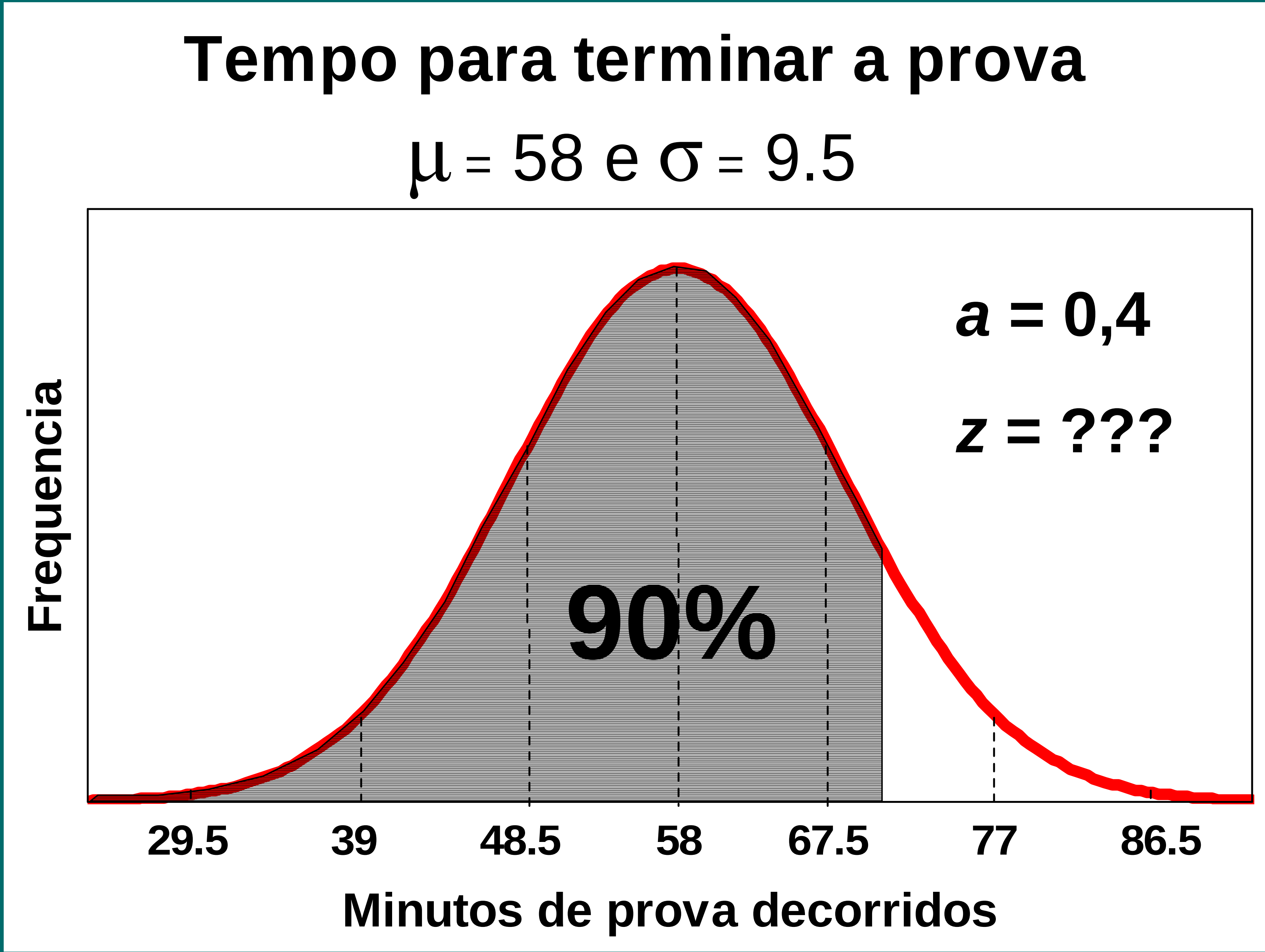
Resolução do exercício 1:

Queremos saber a probabilidade relacionada a 90% dos alunos terminarem.

De início já podemos incluir os 50% de alunos que terminarão até 58 minutos (média).

Então precisamos saber qual o valor de z relativo aos dos demais 40% alunos mais rápidos da turma.

Resolução do exercício 1:



Resolução do exercício 1:

Consultando a tabela descobrimos que 39,97% da curva é preenchida quando z = 1,28

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177

Então, como:

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

$$x = (z \times \sigma) + \mu$$

$$(1{,}28 \times 9{,}5) + 58 = 70{,}16$$

Intervalo de Confiança

O tamanho amostral:

Três pesquisas de opinião a respeito das eleições de 2010 foram publicadas:

- O instituto DataFolha entrevistou 124 eleitores
- O instituto CNT/Sensus entrevistou 584 eleitores
- O instituto Ibope ouviu 3597 eleitores

Assumindo que todos eles utilizaram a mesma metodologia (aleatorização, questionário, etc),

Em qual você confia mais? Por quê?

O tamanho amostral:

Como podemos utilizar nosso conhecimento de probabilidade para "estimar" a

confiança

na minha estimativa?

“Estabilidade” das estimativas:

Imagine que temos tempo e recursos ilimitados para repetir o processo de amostragem infinitamente.

Poderíamos estimar qual é a “estabilidade” das médias estimadas muitas amostras.

Obviamente, quanto maior for as amostras, mais semelhante será a média estimada entre amostras. Assim, quando as amostras forem iguais a população todas terão médias idênticas.

“Estabilidade” das estimativas:

Portanto,

↑ estabilidade   ↑ confiança

Ou seja,

↑ variância da amostra   ↑ incerteza da estimativa

“Estabilidade” das estimativas:

Além disto,

↑ esforço amostral

↑ confiança

Ou seja,

↓ tamanho da amostra

↑ incerteza da estimativa

Erro padrão da média ($s_{\bar{x}}$):

Infelizmente nunca temos recursos ou tempo para fazer muitas amostras independentes.

Entretanto, a variância da média calculada para várias amostras pode ser estimada pela seguinte fórmula:

$$s_{\bar{x}}^2 = \frac{s^2}{n}$$

Assim, o “desvio padrão” da média, conhecido como **erro padrão da média** é dado por:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Erro padrão da média:

Se você queria uma medida de confiança...

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

variância da amostra

esforço amostral

Contextualizando:

A média da altura de 20 alunos da UFG (amostra) é uma estimativa da média de todos alunos da UFG (população).

O erro padrão estima a incerteza que temos em relação a média estimada pela amostra. Ou seja, qual a incerteza que tenho em afirmar que os 20 alunos na minha amostra representam todos os alunos da UFG?

Gostaríamos então de saber o **intervalo** que abarcaria *x*% das médias estimadas por amostras independentes de uma mesma população.

Intervalo de confiança para a média:

Nosso já conhecido *z*, da curva normal

$$-z \times s_{\bar{x}} \leq \bar{x} \leq +z \times s_{\bar{x}}$$

Intervalo de confiança para a média:

Qual valor de *z* eu devo escolher?

Depende! Quão confiante você quer estar sobre o intervalo de estimativa da sua média?

Quanto mais confiança quiser, mais certeza terá que o parâmetro populacional estará contido em intervalos de confiança calculados da mesma maneira.

Por outro lado, o custo de ter mais certeza é estimar um intervalo maior.

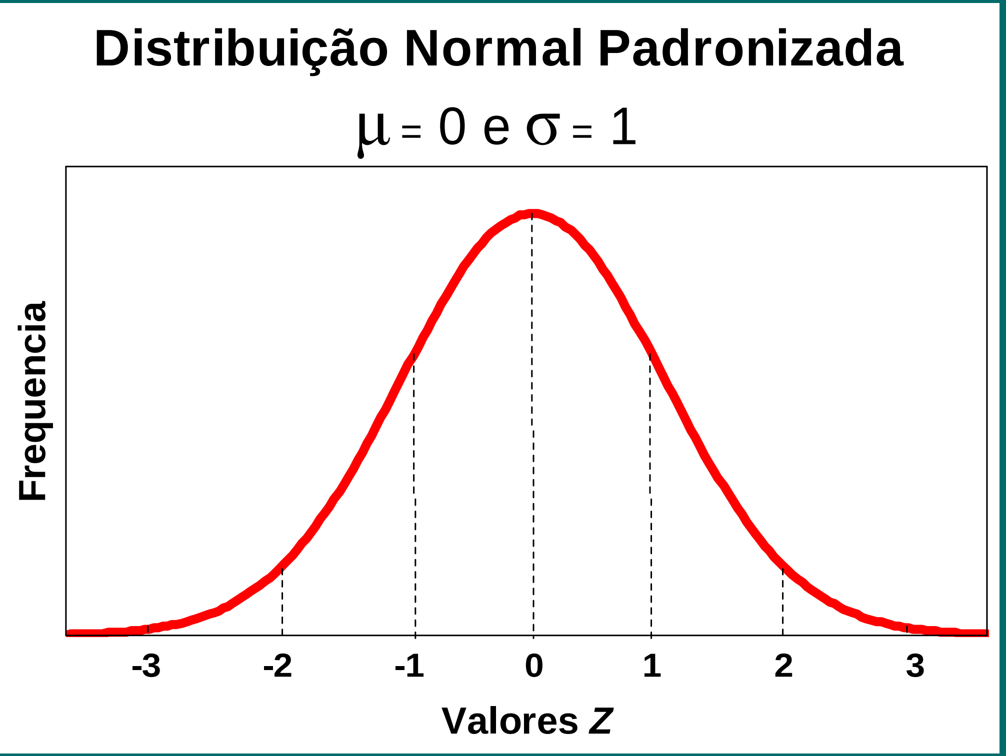
Intervalo de confiança para a média:

Assim,

↑ confiança

↑ valor de *z*

↑ intervalo



Intervalo de confiança para a média:

Infelizmente,

↑ intervalo

↑ confiança

↓informação

Tenho certeza que todos os alunos da UFG estão dentro do intervalo!

$$20cm < 170cm < 320cm$$

Intervalo de confiança para a média:

Mas **confiança** não é baseada no balanço entre variabilidade e esforço amostral?

$$-z \times \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq +z \times \frac{\sigma}{\sqrt{n}}$$

O valor z não aumenta ou diminui a confiança, apenas estima. Para aumentar a confiança você terá que diminuir a variabilidade da amostra ou aumentar o esforço!

Exemplo 1:

Mediu-se a estatura de 40 alunos do curso de Ecologia da UFG. A média e desvio padrão estimados foram: μ = 170 e σ = 5. Qual o intervalo de confiança da média a 95%?

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{40}} = 0{,}79$$

Assim,

$$-z \times s_{\bar{x}} \leq \bar{x} \leq +z \times s_{\bar{x}}$$

$$-1{,}96 \times 0{,}79 \leq 170 \leq 1{,}96 \times 0{,}79$$

Portanto:

168,45 - 171,55

Exemplo 2:

Em uma segunda amostra, mediu-se a estatura de 400 alunos. A média e desvio padrão estimados foram os mesmos: μ = 170 e σ = 5. Calcule o intervalo de 95%

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{400}} = 0,25$$

Assim,

$$-z \times s_{\bar{x}} \leq \bar{x} \leq +z \times s_{\bar{x}}$$

$$-1,96 \times 0,25 \leq 170 \leq 1,96 \times 0,25$$

Portanto:

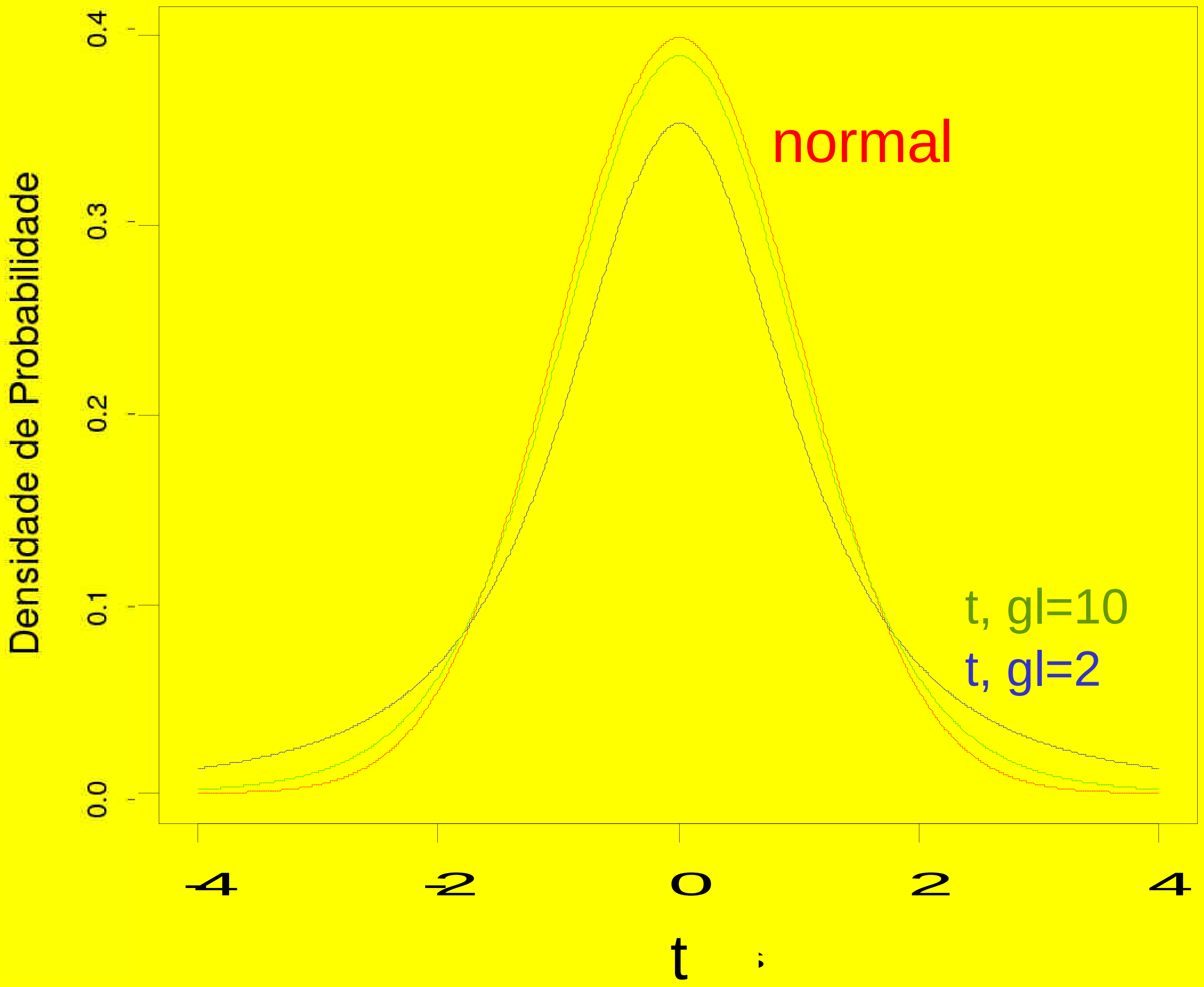
169,51 - 170,49

A questão do tamanho amostral:

A distribuição normal só pode ser utilizada para amostras grandes (>40).

Para amostras pequenas é preciso utilizar a distribuição *t*, que leva em consideração o tamanho amostral.

Para grandes amostras a distribuição *t* converge para a distribuição *z*



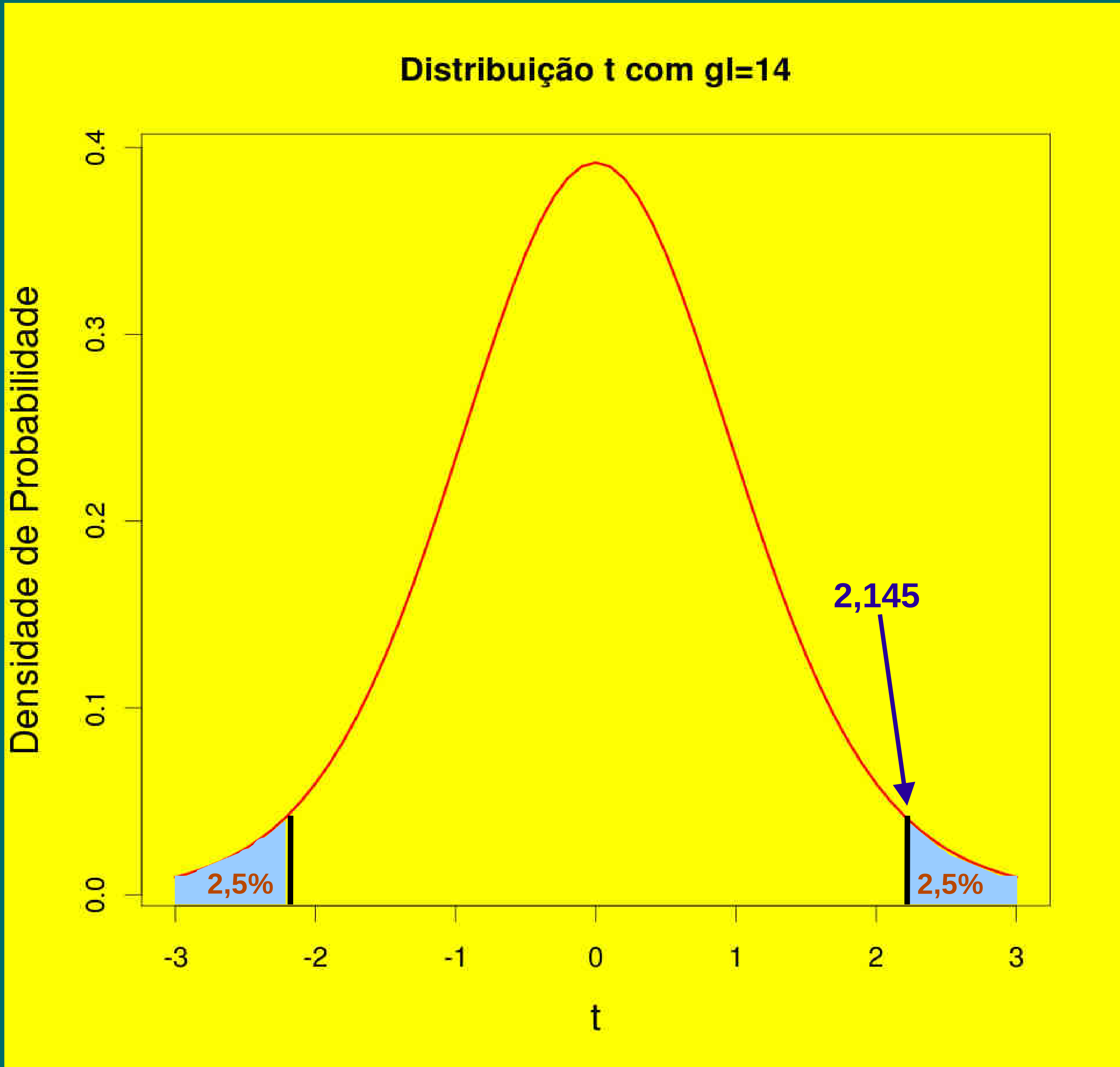
normal
t, gl=10
t, gl=2
Densidade de Probabilidade
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
t

A questão do tamanho amostral:

Para encontrarmos a área sob a curva *t* precisamos saber o esforço amostral e o *t*-valor.

Alternativamente podemos encontrar o *t*-valor quando tivermos a área sob a curva e o esforço amostral.

O esforço amostral é estimado pelo tamanho amostral menos o número de parâmetros a ser estimado. É o famoso **<u>número de graus de liberdade</u>!**


Distribuição t com gl=14
Densidade de Probabilidade
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
t
2,145
2,5%
2,5%

	0,200 $t_{0.800}$	0,150 $t_{0.850}$	0,100 $t_{0.900}$	0,050 $t_{0.950}$	0,025 $t_{0.975}$	0,020 $t_{0.980}$	0,015 $t_{0.985}$	0,010 $t_{0.990}$	0,005 $t_{0.995}$
*d.f.*1	1,37638	1,96261	3,07768	6,31375	12,70615	15,89447	21,20505	31,82096	63,65590
2	1,06066	1,38621	1,88562	2,91999	4,30266	4,84873	5,64280	6,96455	9,92499
3	0,97847	1,24978	1,63775	2,35336	3,18245	3,48191	3,89606	4,54071	5,84085
4	0,94096	1,18957	1,53321	2,13185	2,77645	2,99853	3,29763	3,74694	4,60408
5	0,91954	1,15577	1,47588	2,01505	2,57058	2,75651	3,00288	3,36493	4,03212
6	0,90570	1,13416	1,43976	1,94318	2,44691	2,61224	2,82893	3,14267	3,70743
7	0,89603	1,11916	1,41492	1,89458	2,36462	2,51675	2,71457	2,99795	3,49948
8	0,88889	1,10815	1,39682	1,85955	2,30601	2,44899	2,63381	2,89647	3,35538
9	0,88340	1,09972	1,38303	1,83311	2,26216	2,39844	2,57381	2,82143	3,24984
10	0,87|906	1,09306	1,37218	1,81246	2,22814	2,35931	2,52749	2,76377	3,16926
11	0,87553	1,08767	1,36343	1,79588	2,20099	2,32814	2,49067	2,71808	3,10582
12	0,87261	1,08321	1,35622	1,78229	2,17881	2,30272	2,46070	2,68099	3,05454
13	0,87015	1,07947	1,35017	1,77093	2,16037	2,28160	2,43585	2,65030	3,01228
14	0,86805	1,07628	1,34503	1,76131	***2,14479***	2,26378	2,41490	2,62449	2,97685
15	0,86624	1,07353	1,34061	1,75305	2,13145	2,24854	2,39701	2,60248	2,94673
16	0,86467	1,07114	1,33676	1,74588	2,11990	2,23536	2,38155	2,58349	2,92079
18	0,86205	1,06717	1,33039	1,73406	2,10092	2,21370	2,35618	2,55238	2,87844
20	0,85996	1,06402	1,32534	1,72472	2,08596	2,19666	2,33625	2,52798	2,84534
22	0,85827	1,06145	1,32124	1,71714	2,07388	2,18289	2,32016	2,50832	2,81876
24	0,85686	1,05932	1,31784	1,71088	2,06390	2,17155	2,30692	2,49216	2,79695
26	0,85567	1,05752	1,31497	1,70562	2,05553	2,16203	2,29581	2,47863	2,77872
30	0,85377	1,05466	1,31042	1,69726	2,04227	2,14697	2,27827	2,45726	2,74998
40	0,85070	1,05005	1,30308	1,68385	2,02107	2,12291	2,25027	2,42326	2,70446
80	0,84614	1,04319	1,29222	1,66413	1,99007	2,08778	2,20949	2,37387	2,63870

Exercício 1:

A estatura de 500 goianos escolhidos ao acaso é $\mu = 164$ e $\sigma = 3$. A estatura de 21 alunos da UFG, também escolhidos ao acaso, é $\mu = 166$ e $\sigma = 4$. Os alunos da UFG representam uma amostra aleatória dos goianos?

Os goianos:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{500}} = 0{,}13$$

Assim,

$$-t_{499{,}5\%} \times s_{\bar{x}} \leq \bar{x} \leq +t_{499{,}5\%} \times s_{\bar{x}}$$

$$-1{,}9749 \times 0{,}13 \leq 164 \leq +1{,}9749 \times 0{,}13$$

$$-0{,}25 \leq 164 \leq +0{,}25$$

163,75 - 164,25

Os alunos da UFG:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{21}} = 0{,}56$$

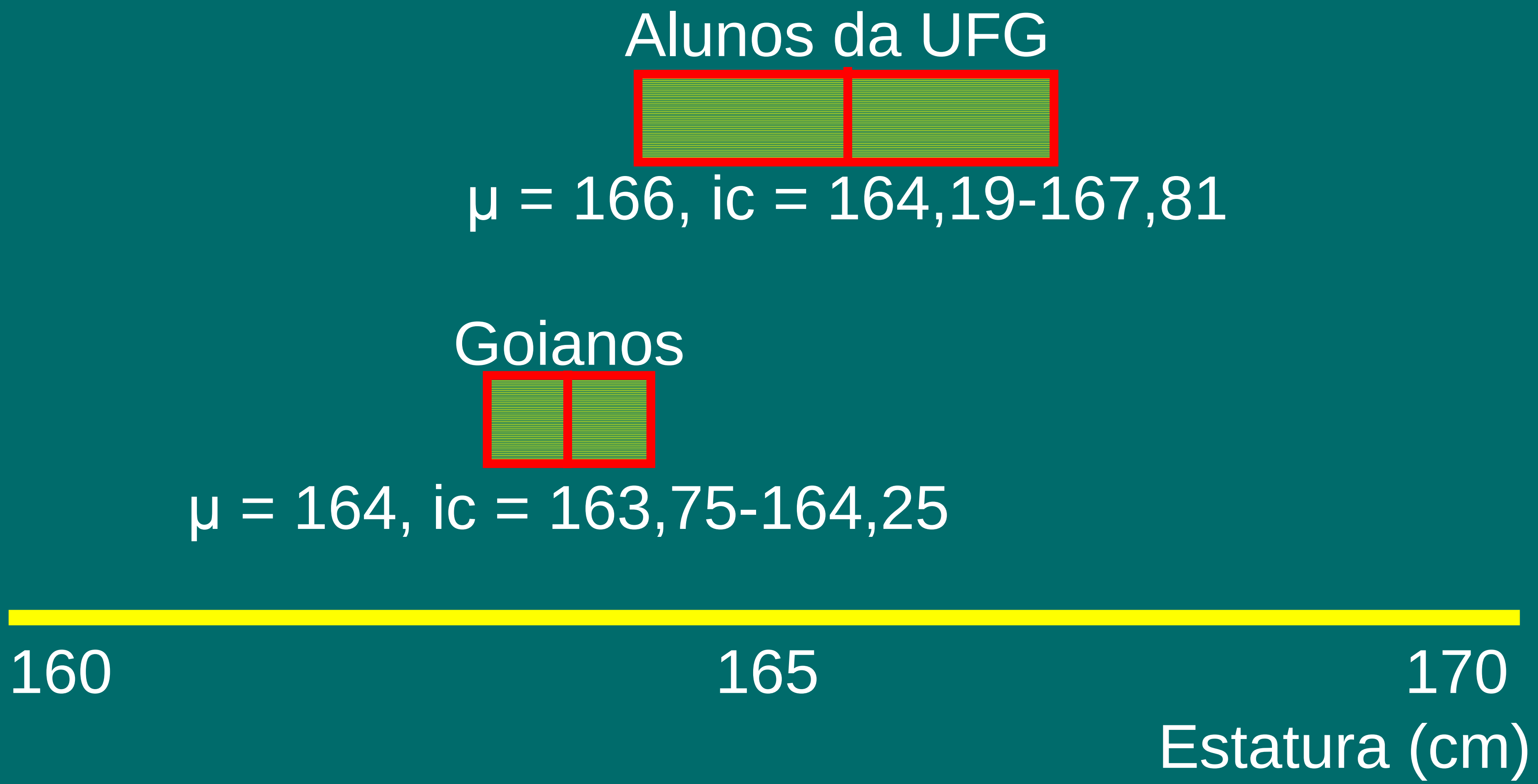
Assim,

$$-t_{20{,}5\%} \times s_{\bar{x}} \leq \bar{x} \leq +t_{20{,}5\%} \times s_{\bar{x}}$$

$$-2{,}085 \times 0{,}87 \leq 166 \leq +2{,}085 \times 0{,}87$$

$$-1{,}81 \leq 166 \leq +1{,}81$$

$$164{,}19 \ \ - \ \ 167{,}81$$



Apesar dos alunos da UFG serem em média mais altos do que a média dos goianos, os intervalos de confiança se sobrepõe. Portanto não possível afirmar que os alunos são significativamente diferentes dos goianos

Exercícios para casa:

Todos os exercícios propostos dos capítulos 10 e 11 de: Sônia Vieira. 2008. Introdução à Bioestatística. 4a ed. Elsevier.