PROBABILIDADE E ESTATÍSTICA

Profº Agnaldo Cieslak

Distribuição de Probabilidades

A distribuição de probabilidades mostra o percentual que podemos esperar do resultado de uma V.A., após grande quantidade de observações.

✓ **Em uma distribuição de probabilidades é necessário:**

- $\sum P(x) = 1$, onde x toma todos valores possíveis
- $0 \leq P(x) \leq 1$ para todo o x.

Distribuições de probabilidade → **Distribuições discretas**

Distribuições de probabilidade → **Distribuições contínuas**

Distribuições Contínuas

A distribuição contínua descreve as probabilidades dos possíveis valores de uma variável aleatória contínua. Uma variável aleatória contínua é uma variável aleatória com um conjunto de valores possíveis (conhecidos como a intervalos) que é infinito e incontável. *[support.minitab.com]*

✓Nas distribuições contínuas utilizam-se a probabilidade da ocorrência em um intervalo $P(a < x < b)$;

✓Em uma distribuição contínua, a probabilidade é dada pela área contida no intervalo considerado.

Distribuições Contínuas

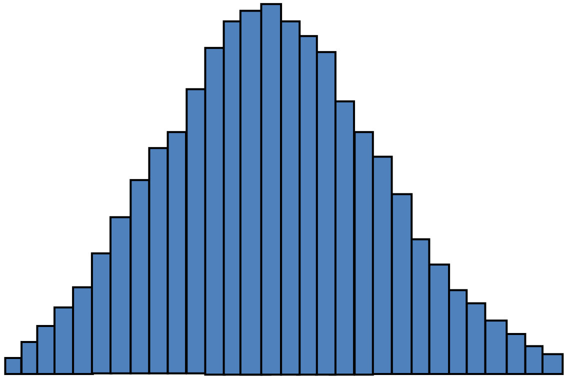
DISTRIBUIÇÕES CONTÍNUAS

(formas)

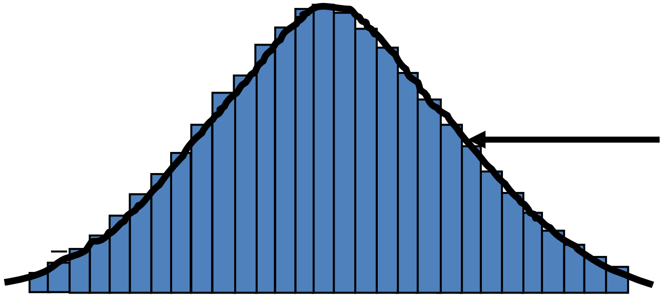
- ✓UNIFORME OU RETANGULAR
- ✓**NORMAL**
- ✓BIVARIADA NORMAL
- ✓EXPONENCIAL
- ✓LOGNORMAL
- ✓WEIBULL
- ✓QUI-QUADRADO χ^2
- ✓**t** DE STUDENT
- ✓**F** DE SNEDECOR
- ✓GAMA
- ✓BETA
- ✓ERLANG

Distribuição Normal

Histórico



No século XVIII, astrônomos e outros cientistas observaram que medidas repetidas de mensurações como a distância à lua variavam como na figura, quando coletadas em grande número.



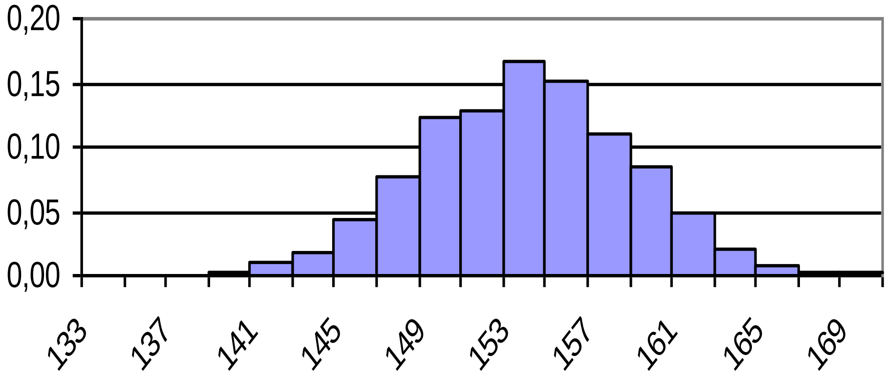
Esta forma gráfica era associada aos erros de mensuração, daí o nome de "Distribuição normal dos erros" e depois "Distribuição normal"

Também é conhecida por "Distribuição Gaussiana", em função do modelo matemático desenvolvido por Karl F. Gauss para este comportamento.

Distribuição Normal - Exemplos

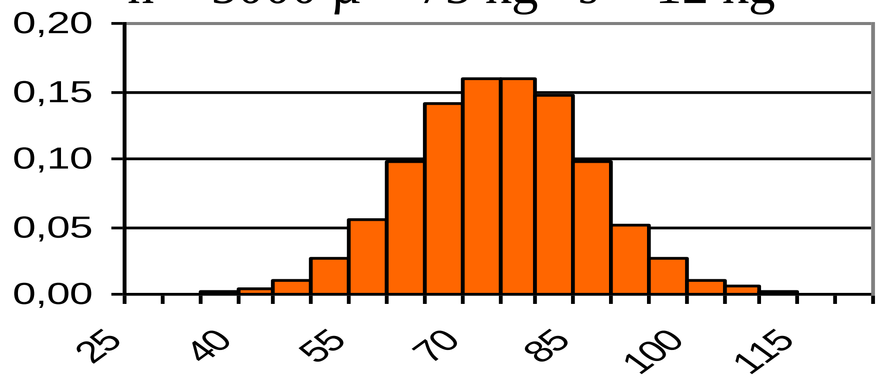
Altura de universitários

n = 3000 μ = 152 cm s = 5 cm



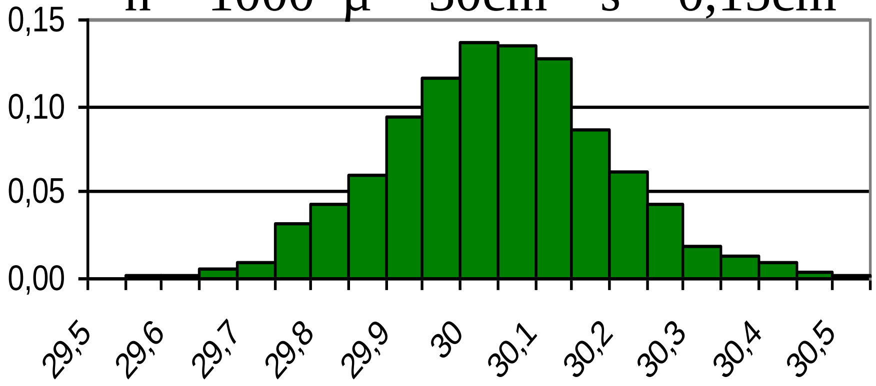
Peso da população adulta

n = 5000 μ = 75 kg s = 12 kg



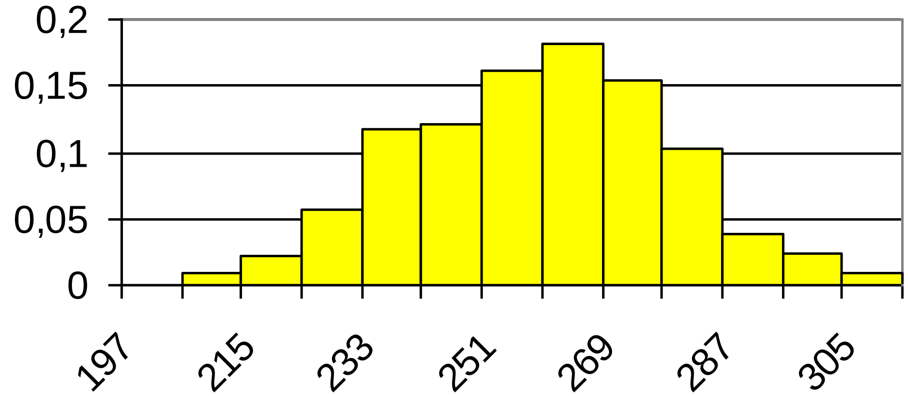
Comprimento de uma régua

n = 1000 μ = 30cm s = 0,15cm



Pessoas num restaurante

μ = 250 por dia s = 20 por dia



IMPORTÂNCIA DA DISTRIBUIÇÃO NORMAL

- ✓Tem boa aproximação para as distribuições de frequência de muitos fenômenos naturais e físicos;

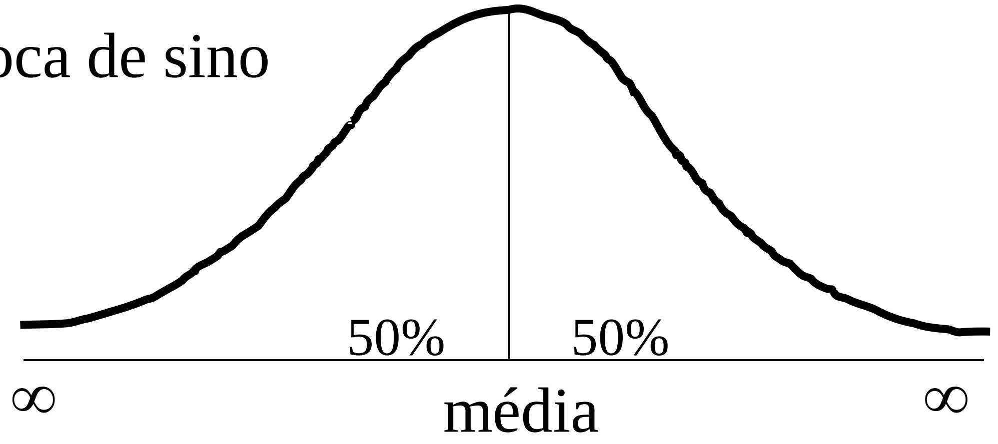
- ✓Representa a distribuição das médias e proporções em grandes amostras, o que tem relevante implicação na amostragem (a mais importante)

- ✓Exemplos: velocidade de processamento, dureza de um material, resistência de uma peça, durabilidade de um item, pressão sanguínea, avaliação de desempenho

Distribuição Normal

Curva normal típica

Forma de uma boca de sino



Função normal→ X ~N (x; μ, σ)

A variável aleatória X tem distribuição Normal x; μ, σ

Área sob a curva = 1 (0,5 + 0,5)

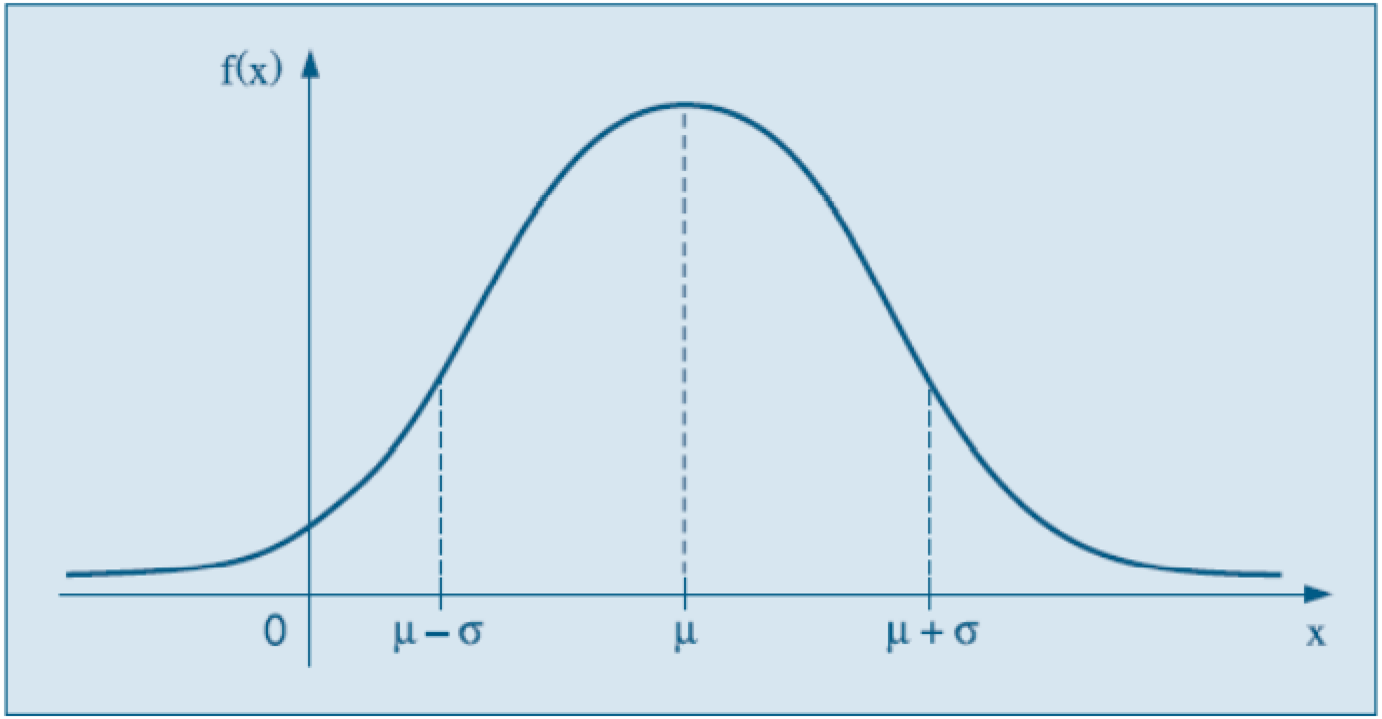
Média = μ

Desvio padrão = σ (S)

Distribuição Normal - Características

1. É simétrica em relação a média
2. Prolonga-se de $-\infty$ a $+\infty$ (apenas em teoria) (assintótica)
3. Especificada por sua média e seu desvio padrão; há uma distribuição normal para cada par (média e desvio padrão)
4. A área total sob a curva é considerada 100% ou igual a 1
5. A área sob a curva entre dois pontos é a probabilidade de uma variável normalmente distribuída tomar um valor entre esses pontos
6. A probabilidade de uma variável aleatória normalmente distribuída tomar exatamente determinado valor (pontual) é zero (característica da distribuição contínua)
7. A área sob a curva entre a média e um ponto arbitrário é função do número de desvios padrões entre a média e aquele ponto

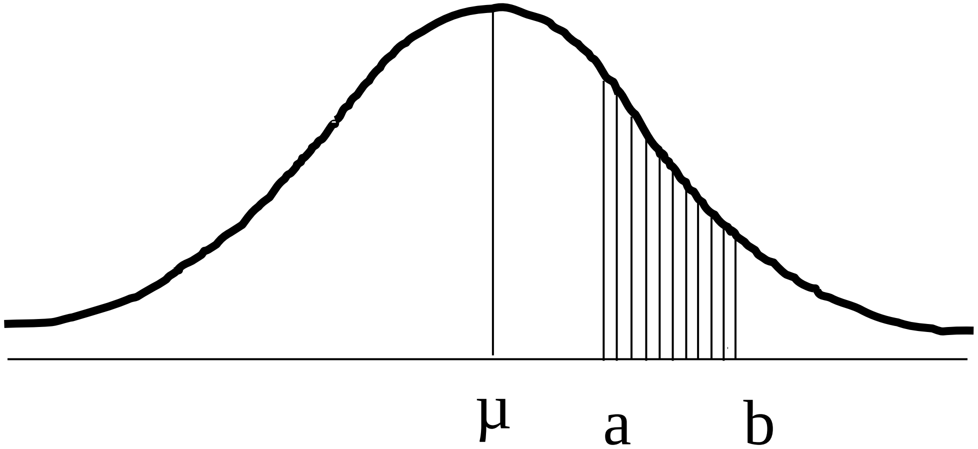
Distribuição Normal



Fonte: Bussab & Morettin (1987).

Distribuição Normal

✓A probabilidade de uma variável aleatória tomar um valor entre dois pontos quaisquer é igual à área sob a curva normal entre aqueles pontos

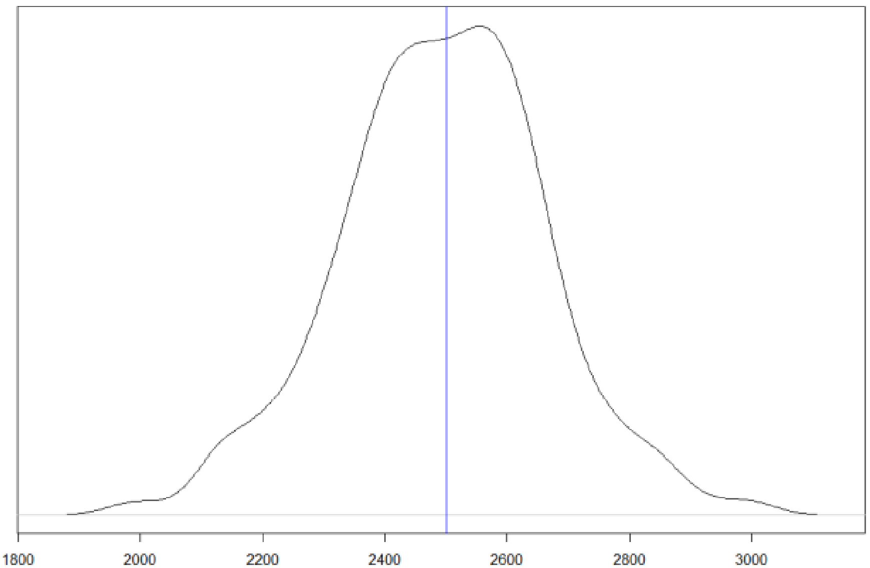


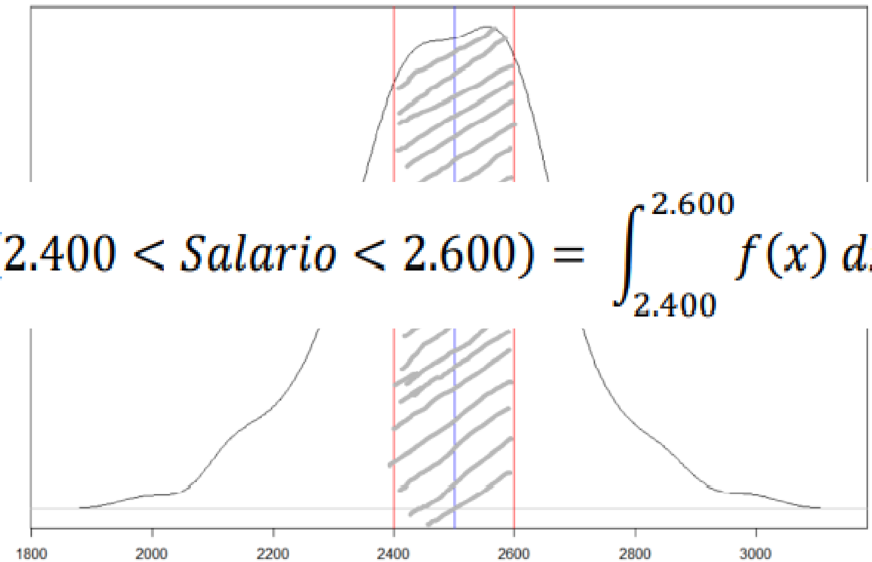
P (a < x < b) = área hachurada sob a curva

Distribuição Normal

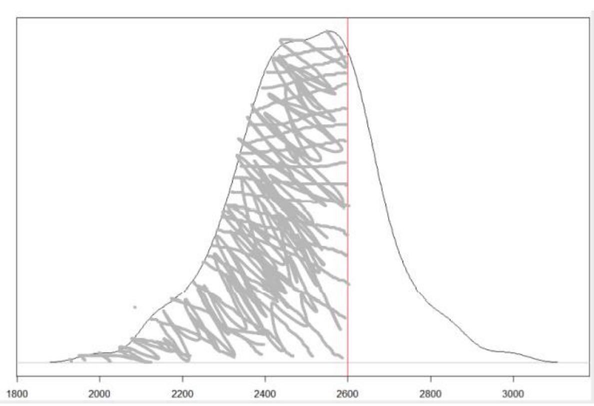
Exemplo: Suponha que a distribuição dos salários dos funcionários de uma empresa siga uma distribuição normal com média μ=2.500 e desvio padrão σ= 170.

Ao selecionar aleatoriamente um indivíduo dessa população, qual a probabilidade de ter salário entre 2.400 e 2.600?

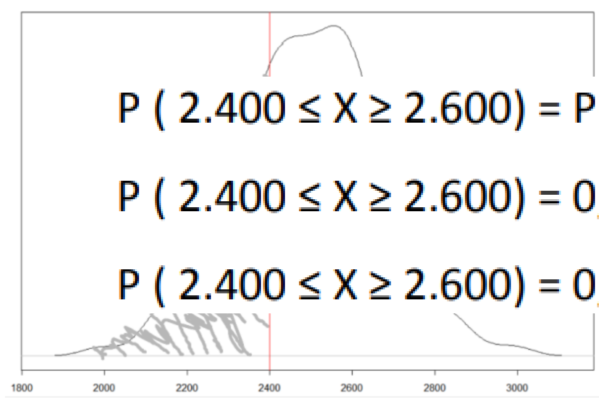


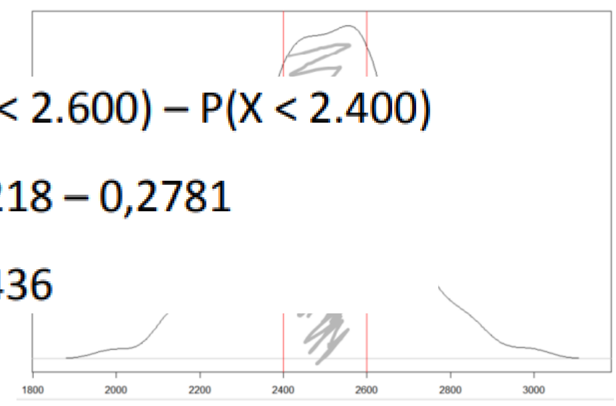


$$P(2.400 < Salario < 2.600) = \int_{2.400}^{2.600} f(x)\,dx,$$



−





P (2.400 ≤ X ≥ 2.600) = P(X < 2.600) – P(X < 2.400)

P (2.400 ≤ X ≥ 2.600) = 0,7218 – 0,2781

P (2.400 ≤ X ≥ 2.600) = 0,4436

Distribuição Normal

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

x – ponto considerado da distrib.

μ - média da distribuição

σ - desvio padrão da distribuição

OBSERVAÇÃO:

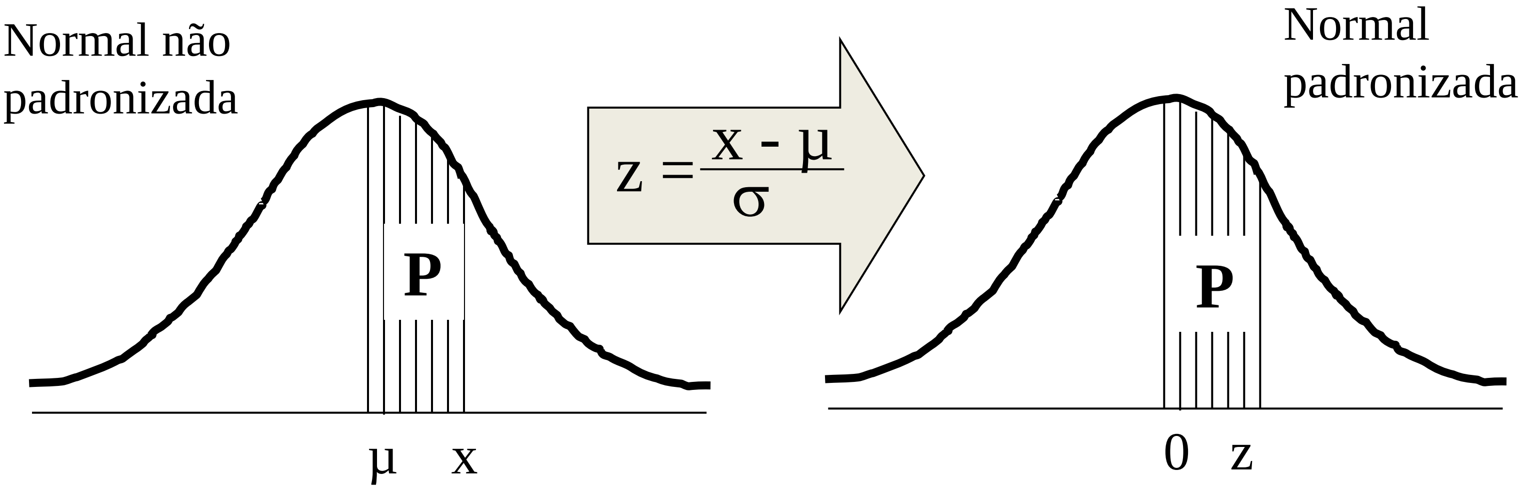
$x - \mu$ = distância do ponto considerado à média

$z = \frac{x-\mu}{\sigma}$ número de desvios padrões a contar da média. Ex.: 2,5

z = valor z ou score z. Pode-se obter valores negativos de z para valores de x inferiores à média

Distribuição Normal

A distância entre a média e um ponto qualquer é dado em número de desvios padrões (z)

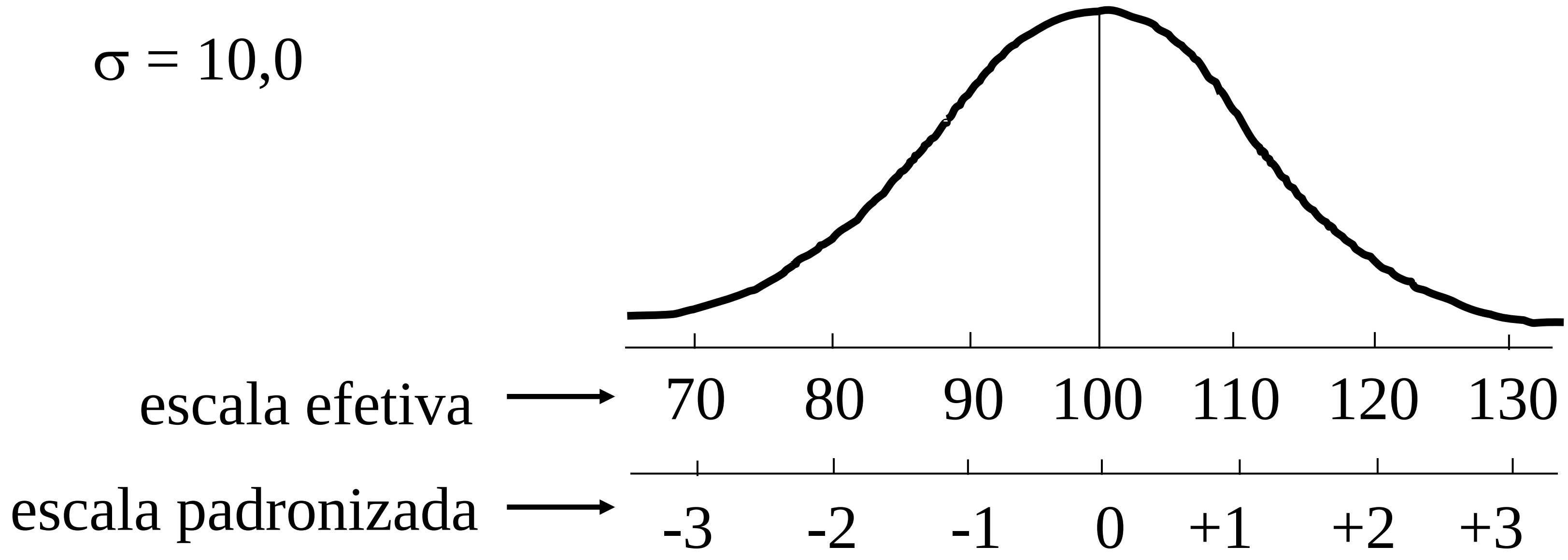


Distribuição Normal

Escala efetiva X Escala padronizada

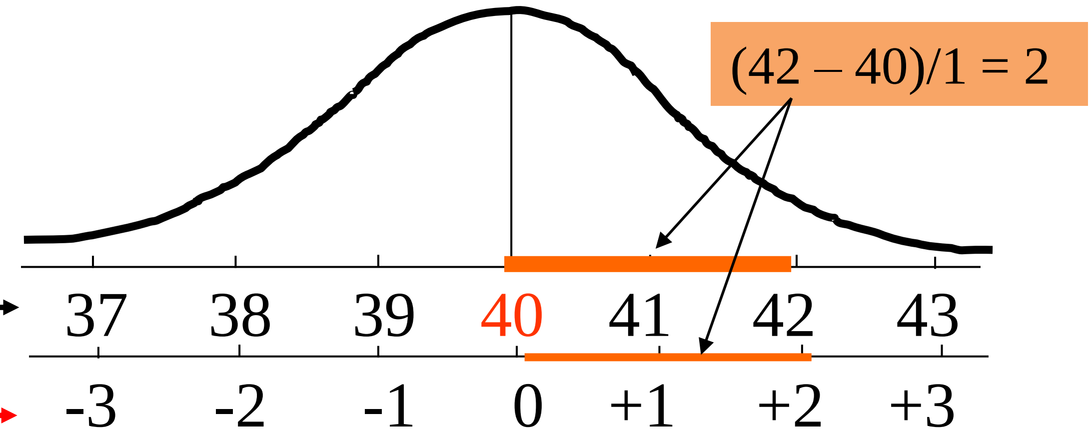
$\mu = 100{,}0$

$\sigma = 10{,}0$



Distribuição Normal

S = 1



escala efetiva

Como calcular Z ?

μ média	σ desvio padrão	x valor considerado	x - μ diferença	(x - μ)/ σ = z diferença relativa
40	1	42	2	2
25	2	23	-2	-1
30	2,5	37,5	7,5	3
18	3	13,5	-4,5	-1,5
22	4	22	0	0

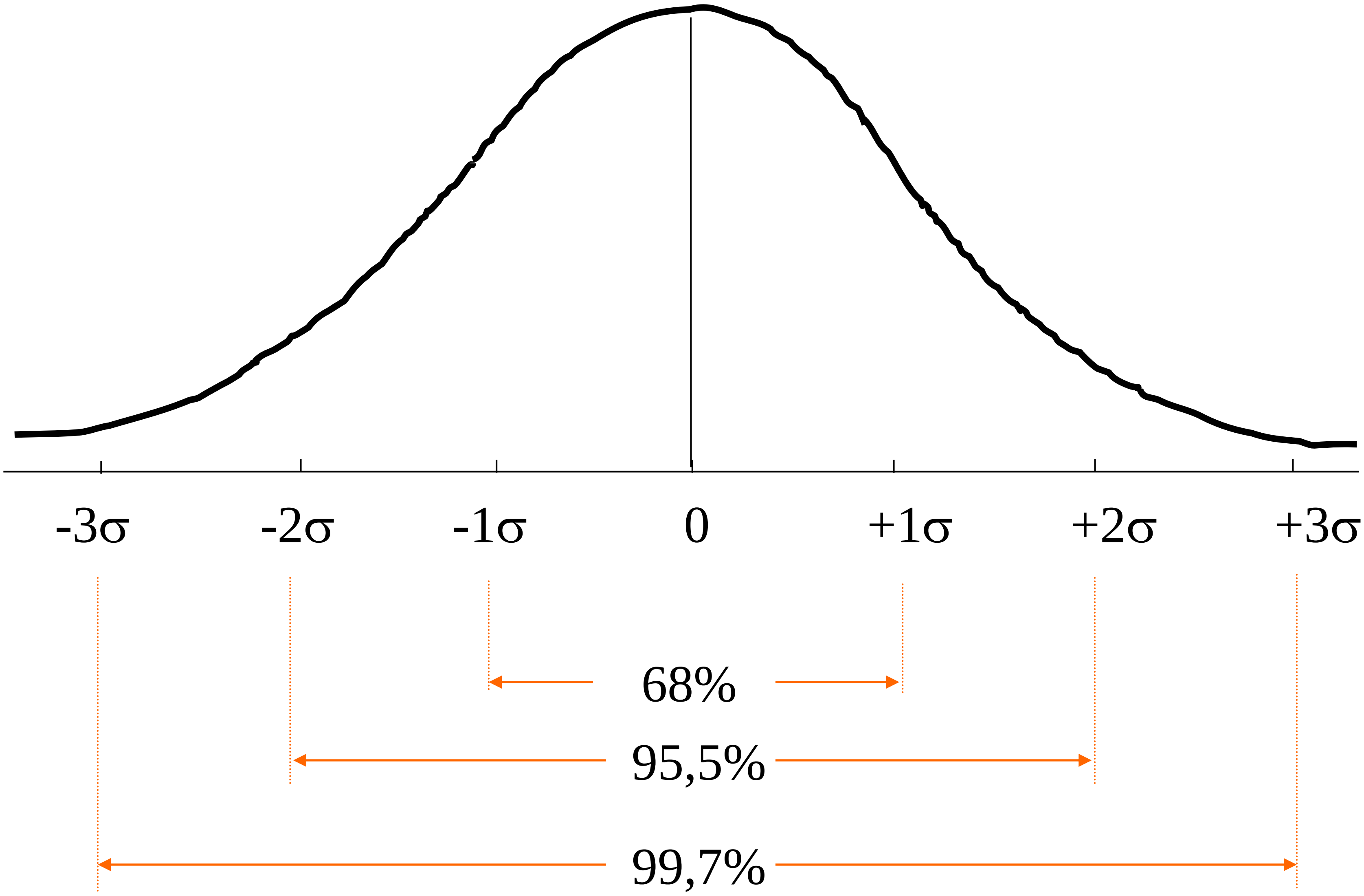
Distribuição Normal

Como calcular o valor efetivo

Passando do valor z para o valor efetivo

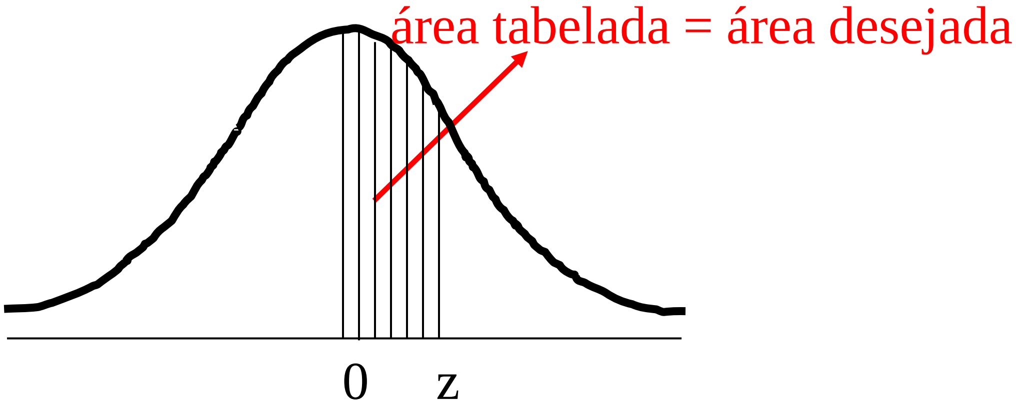
μ média	σ desvio padrão	z valor z	$\mu + z\,\sigma$ cálculo	resultado valor efetivo
20	1	3	20 + 3(1)	23
50	3	-1	50 + 3(-1)	47
60	2	-2	60 + 2(-2)	56
72	5	0,3	72 + 5(0,3)	73,5

Distribuição Normal



Distribuição Normal - Consultando a tabela

Probabilidade de uma variável aleatória normal tomar um valor z entre a média e o ponto situado a z desvios padrões

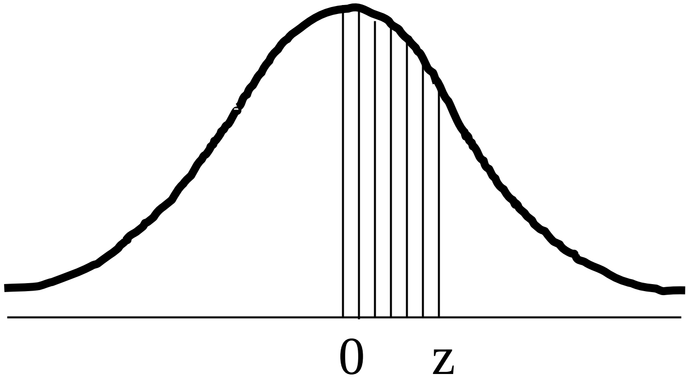


z	área entre a média e z
1,00	0,3413
1,50	0,4332
2,13	0,4834
2,77	0,4972

Distribuição Normal - Consultando a tabela

$z \quad P(0 < x < z)$

$P(x > z) = 0,5 - P(0 < x < z)$



Distribuição Normal - Consultando a tabela

.
.
1,25
1,0 00 01 02 03 04 05 06 ...
.
.
.
1,1
1,2

0,3944

olhando a tabela

Distribuição Normal - Tabela

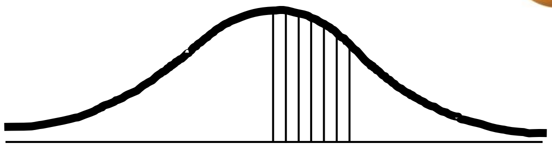
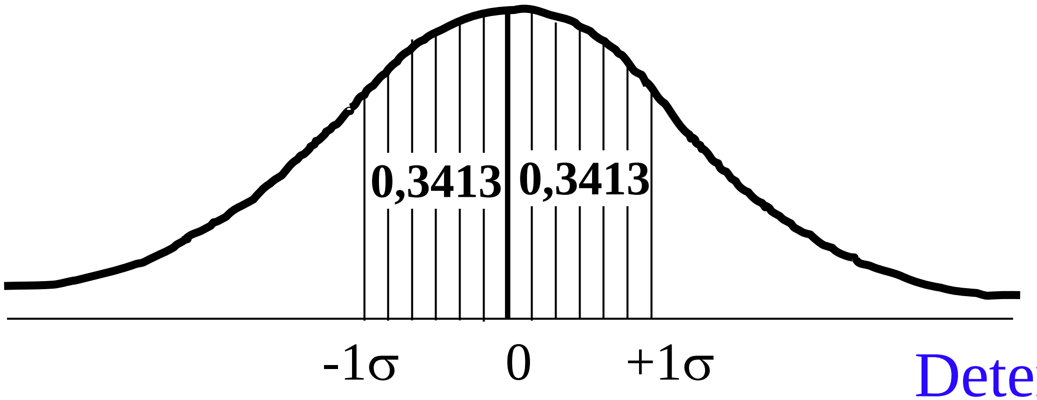


Tabela 5.1 Áreas para a Distribuição Normal Padronizada

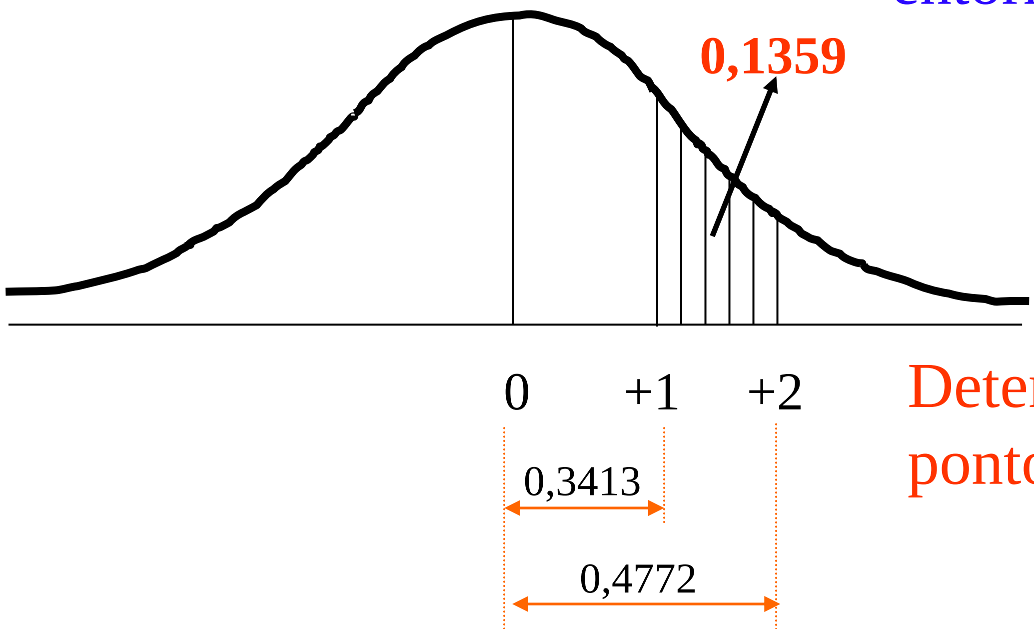
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	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	0,2518	0,2549
0,7	0,2580	0,2612	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
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4986	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
4,0	0,49997									

Distribuição Normal - Cálculo da probabilidade



Exemplos

Determinando a área (probabilidade) sob a curva entre dois pontos entorno da média



Determinando a área entre dois pontos quaisquer

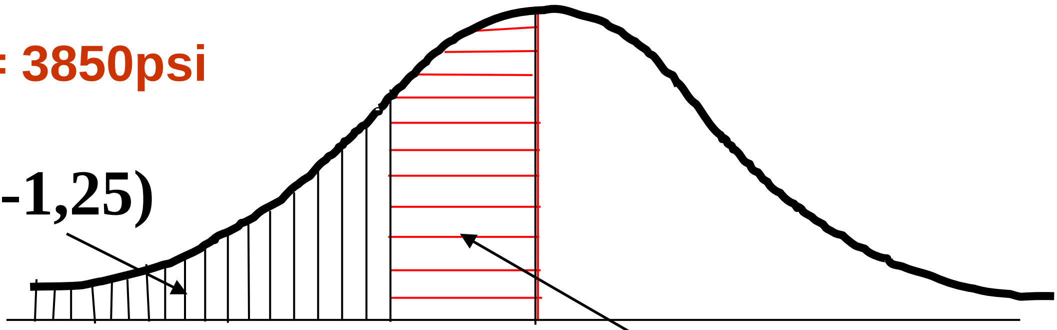
Distribuição Normal - Exemplos

1) Após 28 dias de curagem, o cimento de uma certa marca tem uma resistência compressiva média de 4000psi. Suponha que a resistência tem uma distribuição normal com desvio-padrão de 120psi. Qual a probabilidade de se comprar um pacote de cimento com resistência compressiva de 28 dias menor que 3850psi?

N(μ;σ) = N(4000,120) psi

X = 3850psi

P(z ≤ -1,25)



$$z = \frac{X - \mu}{\sigma} = \frac{3850 - 4000}{120} = -1{,}25$$

(área de -∞ a 4000) = z = -1,25 = 0,3944

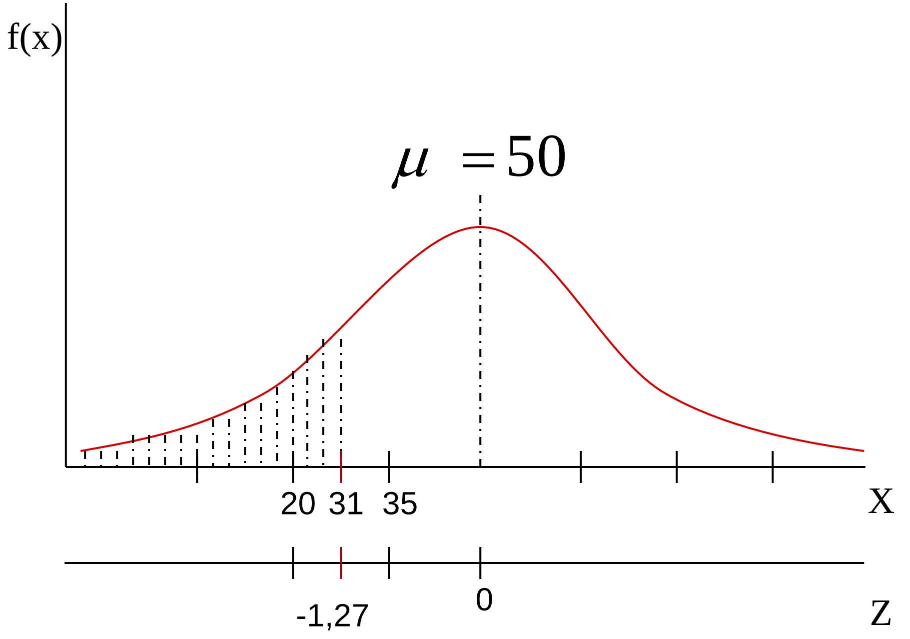
-1,25

Área desejada = 0,50 – 0,3944 = 0,1056 = **10,56%**

$$P(Z \leq -1{,}25) = 0{,}1056 = 10{,}56\%$$

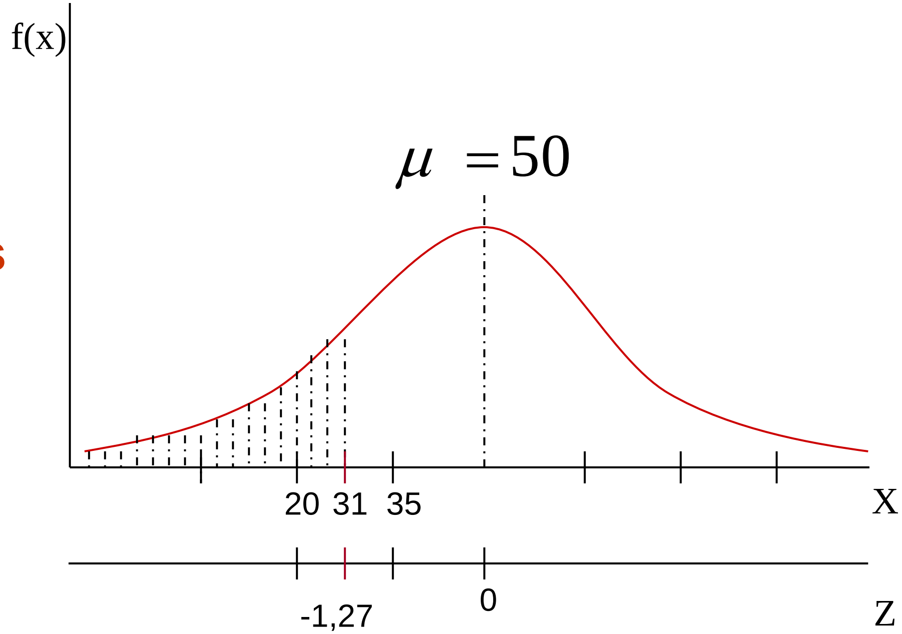
Distribuição Normal - Exercícios

2) Uma grande empresa faz uso de milhares de lâmpadas elétricas que permanecem acessas continuamente. A vida de uma lâmpada pode ser considerada como uma variável aleatória normal com **vida média de 50 dias e desvio-padrão de 15 dias.** Se no dia 1º de agosto foram instaladas 8000 lâmpadas novas, aproximadamente quantas deverão ser substituídas no dia 1º de setembro?



Distribuição Normal - Exercícios

2) Uma grande empresa faz uso de milhares de lâmpadas elétricas que permanecem acessas continuamente. A vida de uma lâmpada pode ser considerada como uma variável aleatória normal com **vida média de 50 dias e desvio-padrão de 15 dias.** Se no dia 1º de agosto foram instaladas 8000 lâmpadas novas, aproximadamente quantas deverão ser substituídas no dia 1º de setembro?



N(μ,σ) = N(50;15) dias **X = 31 dias**

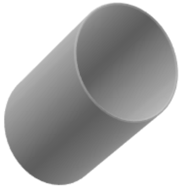
$$z = \frac{X-\mu}{\sigma} = \frac{31-50}{15} = -1{,}27$$

Consultando tabela Z:

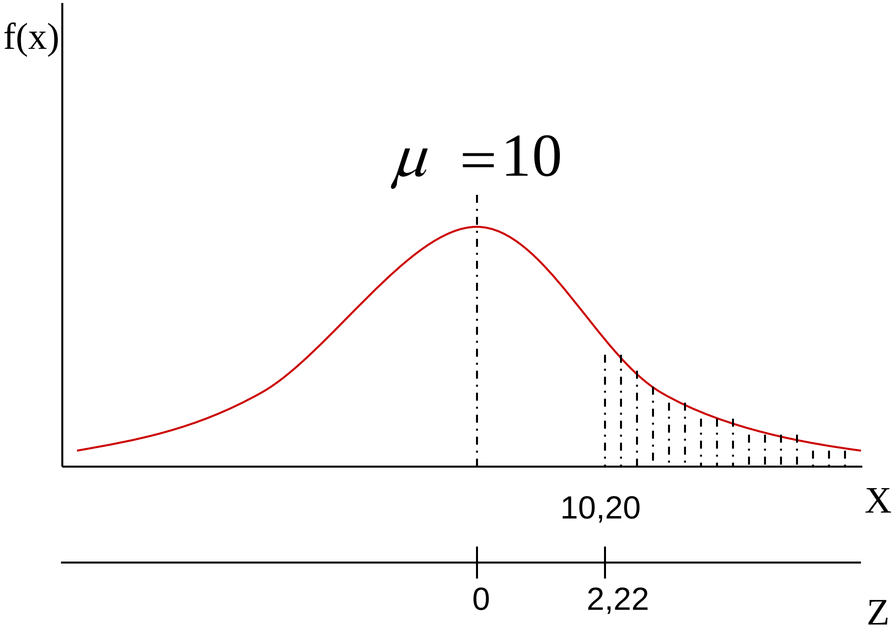
$$P(Z \leq -1{,}27) = 0{,}3980 \quad \text{logo} \quad 0{,}5000 - 0{,}3980 = 0{,}1020 = 10{,}20\%$$

Deverão ser substituídas (0,1020x 8.000) = **816 lâmpadas**

Distribuição Normal - Exercícios



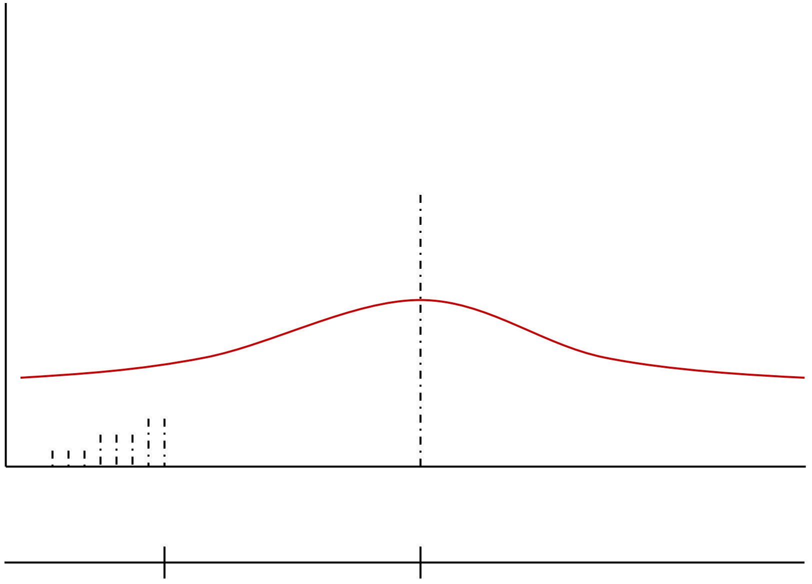
3) Uma indústria siderúrgica produz tubos de aço cujo comprimento pode ser considerado uma variável normalmente distribuída com média **μ=10,00 metros**, e desvio padrão igual a **σ = 0,09 metros**. Quanto refugo a indústria espera produzir se o comprimento dos tubos de aço tiver que ser no máximo, igual a 10,20 m?



Distribuição Normal - Exercícios

Aplicação LOGÍSTICA

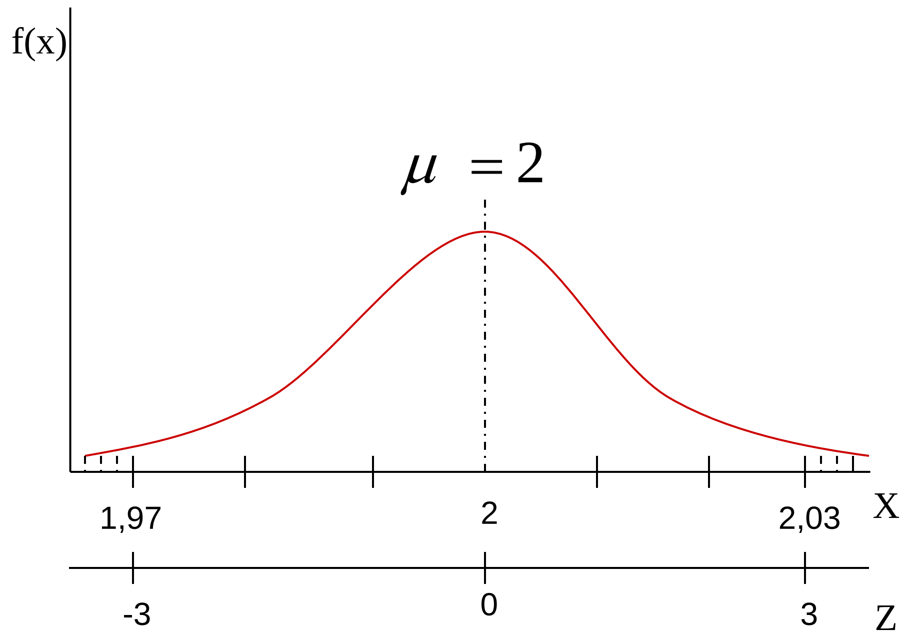
4) O tempo médio que demora para uma viatura de uma determinada empresa atender a uma chamada de emergência é de 8 minutos com desvio-padrão de 3 minutos. Considere o tempo médio como uma variável normalmente distribuída para calcular a probabilidade de uma chamada esperar menos de 4 minutos**.**



Distribuição Normal - Exercícios

Aplicação INDUSTRIAL

5) Um máquina produz peças com o **diâmetro médio de 2,00”** e o **desvio-padrão de 0,01”.** As peças que se afastam da média por mais de **0,03”** são consideradas defeituosas. Qual é a percentagem de peças defeituosa?



Intervalo de Confiança

✓É uma faixa de possíveis valores em torno da média amostral, e a probabilidade de que esta faixa realmente contenha o valor real da média da população

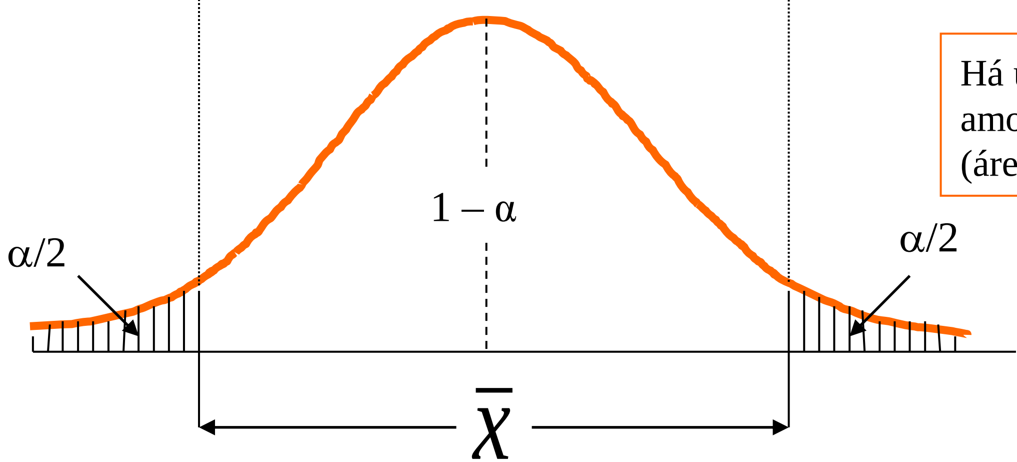
✓O Intervalo de confiança terá uma certa probabilidade chamada de ***nível de confiança*** (simbolizada por $1 - \alpha$) de conter a média da população.

Há uma probabilidade de $1 - \alpha$ da média estar contida no intervalo definido

$1 - \alpha$ = nível de confiança

α = nível de significância (probabilidade de erro)

Há uma probabilidade α de a média amostral estar fora do intervalo definido (área hachurada)

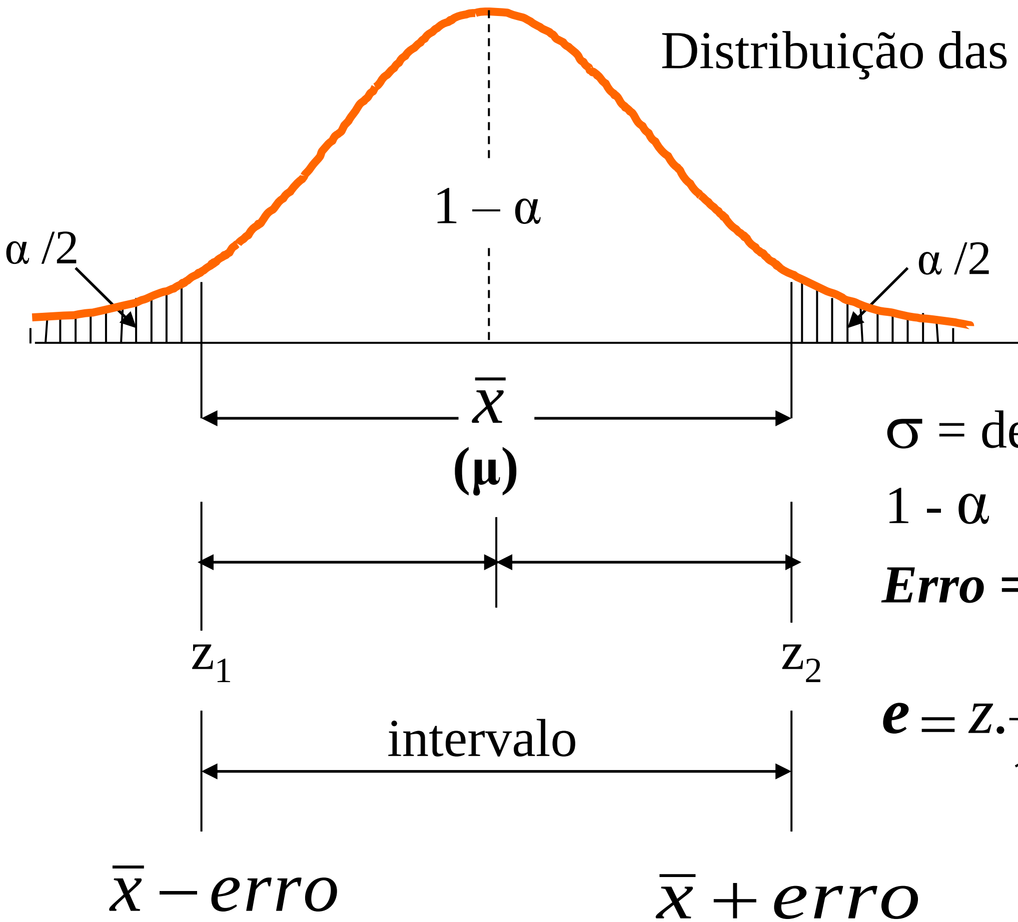


Intervalo de confiança

Se usarmos um desvio padrão em torno da média ($Z = 1$), a chance de erro ao estimar a média será de 31,74%. Mas, se usarmos dois ($Z = 2$), a chance de erro será de 4,56%.

Intervalo de Confiança

Distribuição das médias amostrais



σ = desvio padrão da população

$1 - \alpha$ = nível de confiança

Erro = z . Desvio padrão amostral

$$\boldsymbol{e} = z.\frac{\sigma}{\sqrt{n}}$$

$$P(\bar{x} - e \leq \mu \leq \bar{x} + e) = 1 - \alpha$$

Intervalo de Confiança

Se o desvio padrão da população é conhecido:

$$\mu : \overline{X} \pm z.\sigma_{\overline{X}}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

A estimativa intervalar da média populacional se baseia na hipótese de que a distribuição amostral das médias amostrais é normal. Para grandes amostras se aplica o teorema do limite central.

Para amostras de 30 ou menos observações, é importante saber se a população tem distribuição normal ou aproximada.

Se o desvio padrão da população é desconhecido:

$$\mu = \overline{X} \pm z.\frac{S_X}{\sqrt{n}}$$

μ: média da população

σ: desvio padrão da população

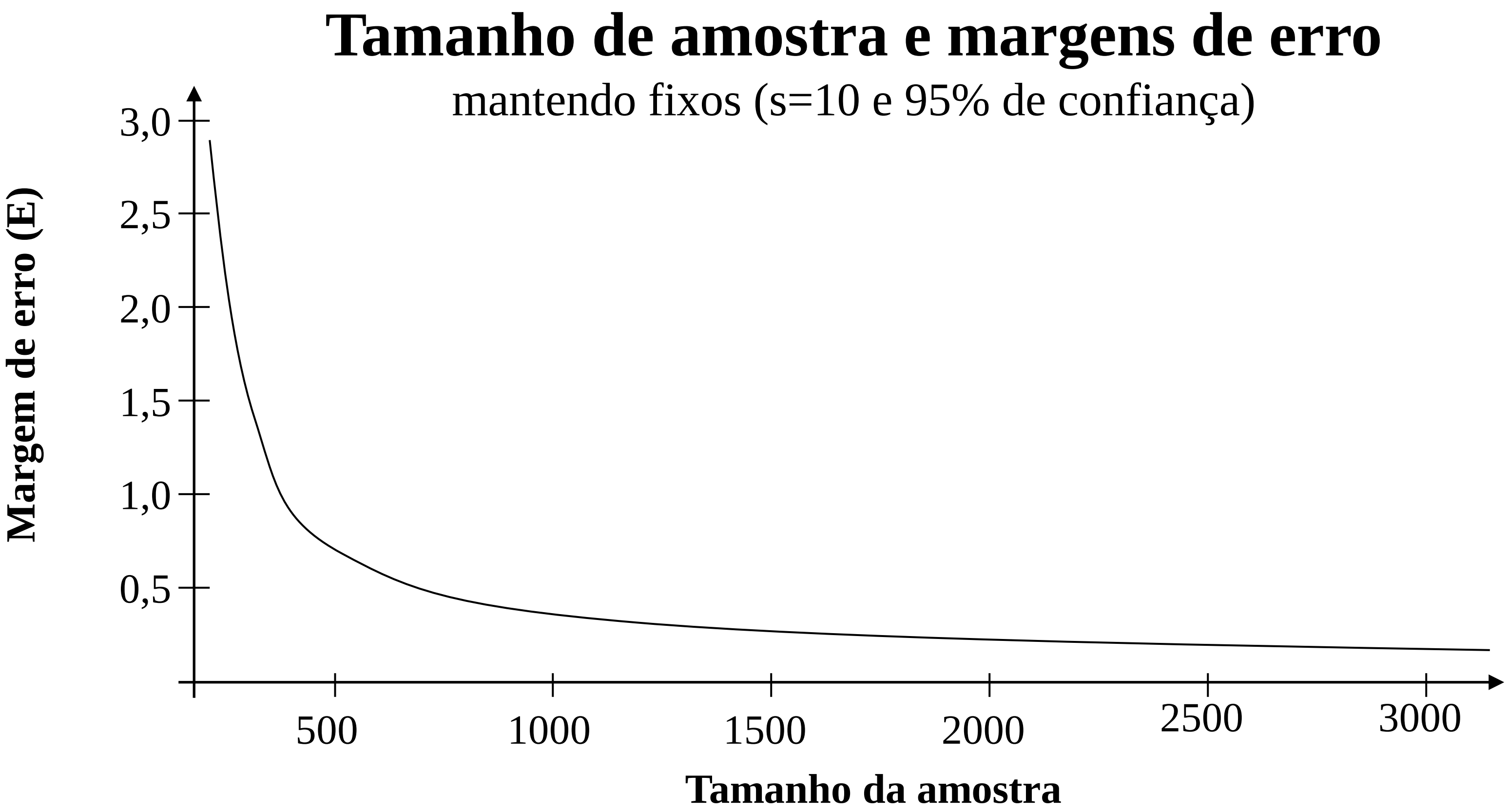
$\overline{X}$: média da amostra

Sx: desvio padrão da amostra

Quando o desvio padrão da população não é conhecido (o que é o caso, geralmente), usa-se o desvio padrão da amostra como estimativa, substituindo-se σ_x por S_x nas equações.

Para amostras menores que 30, a aproximação normal não é adequada. Devemos então usar a distribuição t. A forma da distribuição t é bem parecida com a normal.

Intervalo de Confiança



• Os ganhos em precisão conseguidos com aumentos fixos dos tamanhos das amostras não são constantes;

• Tamanho de amostra 5.000 podem ser um perda de tempo e dinheiro porque elas fornecem <u>pouca precisão adicional</u>;

Intervalo de Confiança

$$\bar{x} - e \leq \mu \leq \bar{x} + e \quad ou \quad \mu = \bar{x} \pm e$$

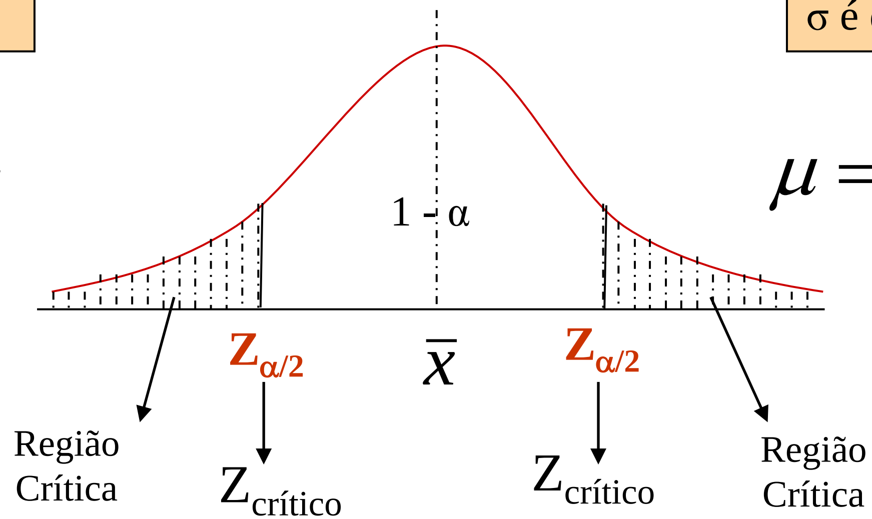
Quando tem n > 30 e

σ é conhecido

$$\mu : \overline{X} \pm z.\sigma_{\overline{X}}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$e = z.\frac{\sigma}{\sqrt{n}}$$



Quando tem n > 30 e

σ é desconhecido

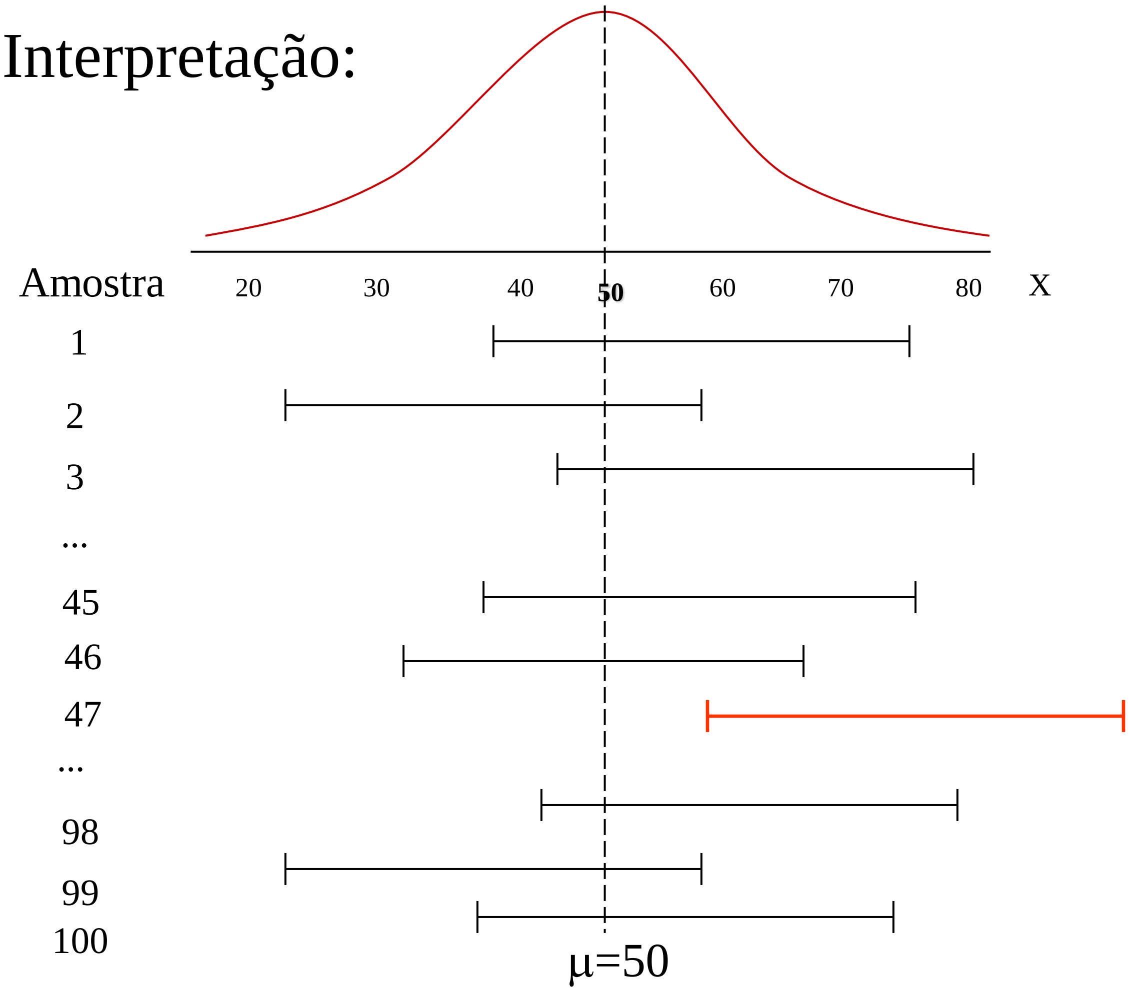
$$\mu = \overline{X} \pm z.\frac{S_X}{\sqrt{n}}$$

$$e = z.\frac{S_X}{\sqrt{n}}$$

Substitui o desvio padrão da população σ pelo desvio padrão da amostra s

Intervalo de Confiança

Interpretação:



Se forem retiradas várias amostras aleatórias de tamanho n da população e para **cada amostra**, seja construído um intervalo de (1-α) de confiança para a variável desejada.

Os intervalos obtidos serão diferentes, mas (1-α)% destes intervalos conterão entre os seus intervalos o valor real do parâmetro.

✓Ao nível de 95% de confiança espera-se que em 100 intervalos para as amostras, 95 deles contenham a média μ

Imagine que tivéssemos uma amostra de tamanho tão grande que tendesse ao infinito. O que ocorreria?

O erro seria próximo de zero ou desconsiderável e a média da amostra seria igual a média da população, sem a necessidade de estimar um intervalo.

Determine o valor crítico $Z_{\alpha/2}$ que corresponde ao grau de confiança indicado:

a) 99%

b) 94%

c) 92%

d) 90%

Intervalo de Confiança – Desafio 1

Uma das linhas de produção de uma siderúrgica fabrica folhas de flandres. Havia uma preocupação com a possibilidade de haver uma quantidade fora da faixa de especificação de dureza (LIE = 58,0 HR e LSE = 64,0 HR). A área de qualidade da empresa decidiu estimar a dureza média das folhas de flandres (μ) coletando uma amostra aleatória de 49 folhas.

Medidas de dureza (HR) das folhas-de-flandres fabricadas pela siderúrgica						
61,0	60,2	60,3	60,3	60,0	61,0	60,3
60,0	60,0	60,9	61,0	61,2	59,2	60,9
60,0	60,5	59,8	59,3	61,0	59,6	59,8
59,6	60,1	58,0	59,8	58,9	57,6	58,0
60,5	60,1	61,6	61,1	59,7	58,3	61,6
59,5	59,0	60,3	58,7	59,6	54,2	60,3
61,0	59,7	59,9	59,9	60,0	58,6	59,9

$\overline{X} =$

$s =$

Para um grau de confiança de 95%, determine a margem de erro (E) e o intervalo de confiança para média populacional (μ).

Intervalo de Confiança – Desafio 2 e 3

2- Uma máquina automática de suco industrial é regulada de modo que a quantidade suprida de cada vez, tenha distribuição aproximadamente normal com desvio-padrão de 35ml. Determine um intervalo de 96% de confiança para a quantidade média de toda produção, sabendo que uma amostra de 30 embalagens teve um conteúdo médio de 290 ml.

3- Uma amostra aleatória de 40 contas não-comerciais na filial de um banco acusou saldo médio de R$140,00 com desvio-padrão de R$30,00.

a) Construa um intervalo de 95% confiança para a verdadeira média.

b) Construa um intervalo de 99% confiança para a verdadeira média.

c) A que conclusão podemos chegar com os resultados das letras anteriores?