UNIVERSIDADE DO ALGARVE
ESCOLA SUPERIOR DE SAÚDE

Investigação Aplicada I

Aula 5

1º Semestre 2020/21

Licenciatura em Ciências Biomédicas Laboratoriais

Sumário:

Tratamento e análise de dados

Probabilidade

Curva Normal e teste Z

Intervalo de Confiança

TRATAMENTO E ANÁLISE DE DADOS

Distribuição Normal e teste Z

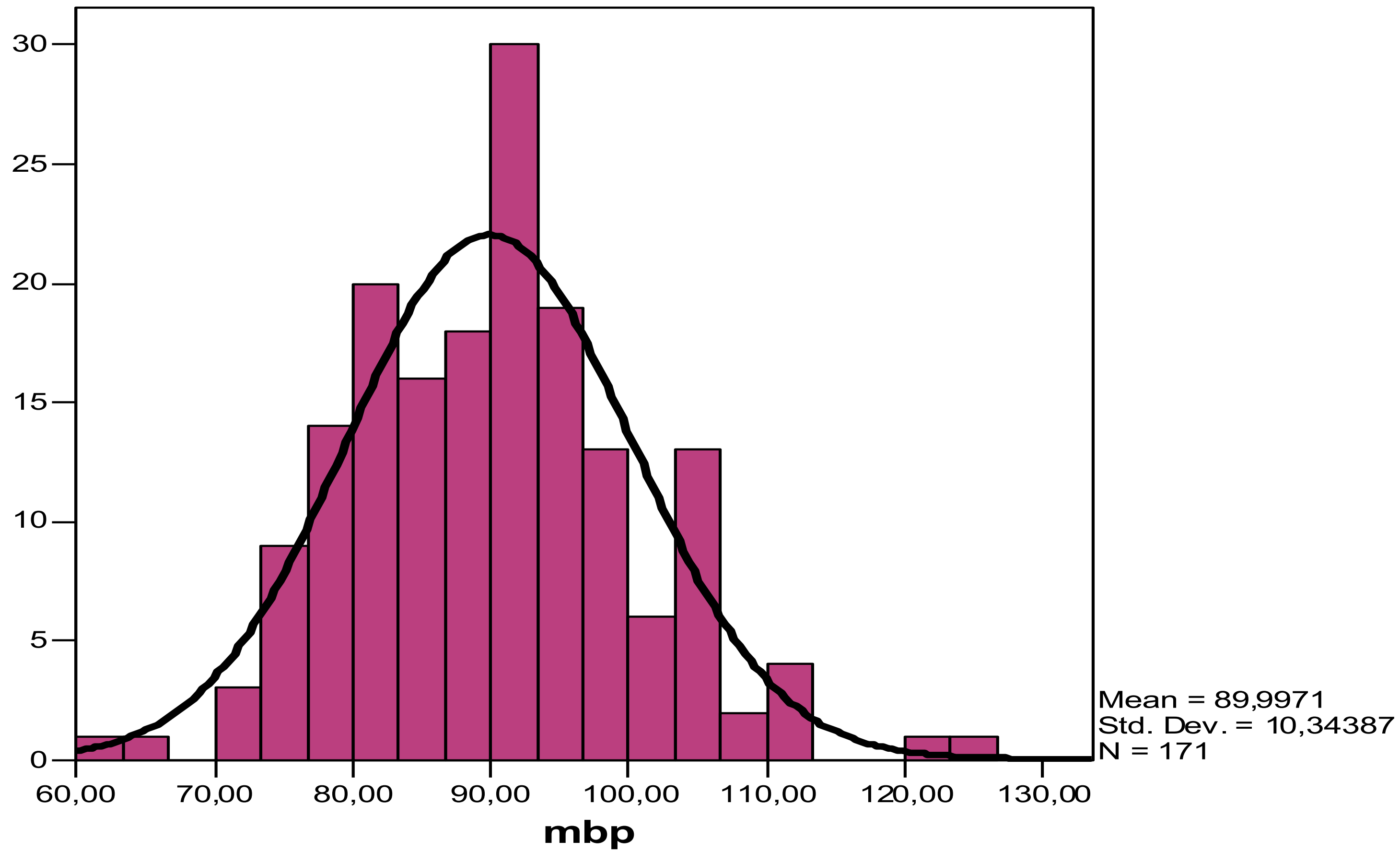
Distribuição Normal

- ✓ A distribuição normal conhecida também como distribuição gaussiana é sem dúvida a mais importante distribuição contínua.

- ✓ A sua importância se deve-se a vários fatores, entre eles podemos citar o **Teorema do limite central**, o qual é um resultado fundamental em aplicações práticas e teóricas.

- ✓ O Teorema do Limite Central garante que mesmo que os dados não sejam distribuídos segundo uma curva normal, a média dos dados converge para uma distribuição normal conforme o número (n) de dados aumenta.

Distribuição Normal



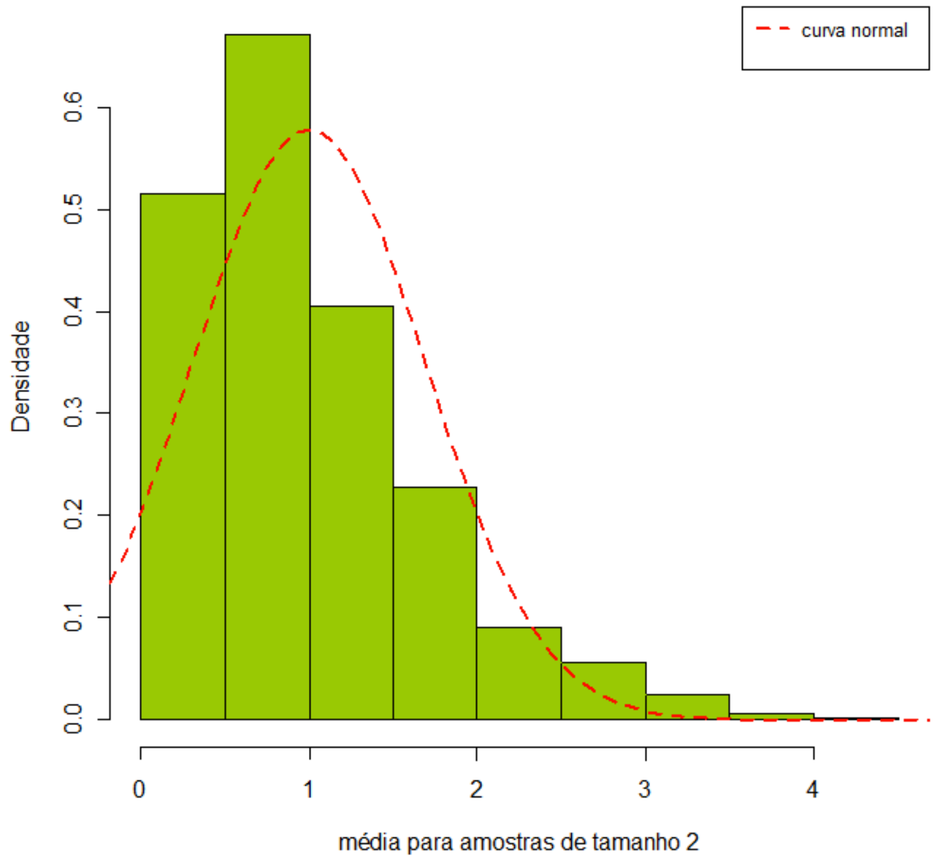
Teorema do Limite Central

- Para a maioria das populações a distribuição de média na amostragem é aproximadamente normal quando o tamanho da amostra é suficientemente grande

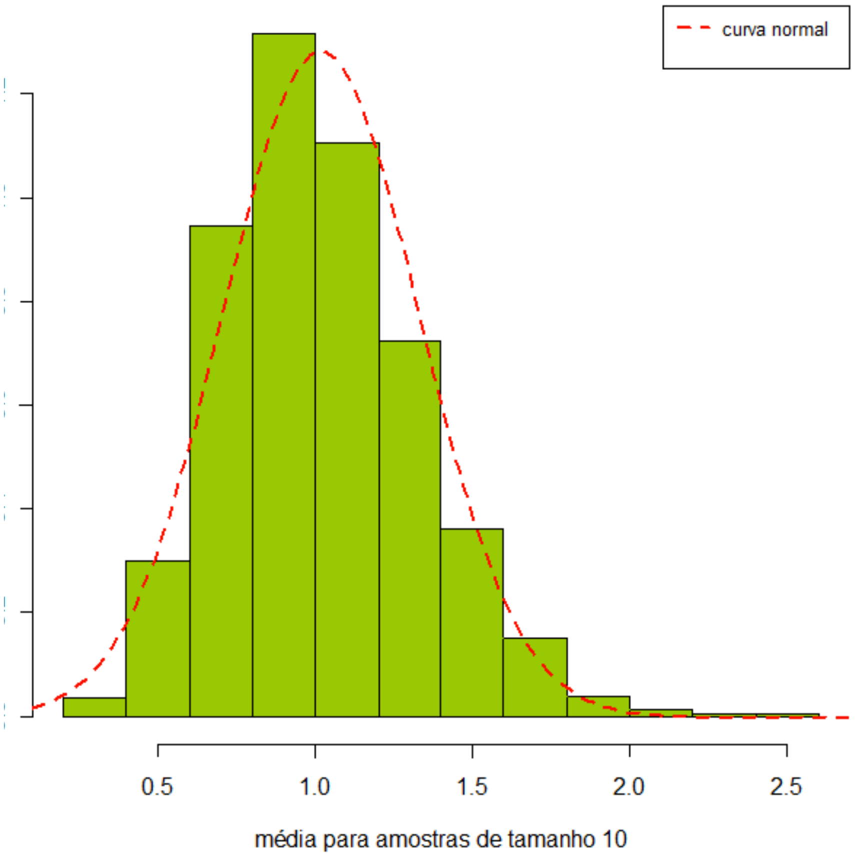
- Se a distribuição da variável X na população é normal:
  - A distribuição de média na amostragem é sempre normal, qualquer que seja o tamanho da amostra.

Convergência de variáveis aleatórias

Histograma de amostras de n=2



Histograma de amostras de n=10



Estimação da média da população

- ✓ Utilizamos os estimadores da população (amostra), sempre para variáveis quantitativas continuas, realizando **testes probabilísticos – Teste Z -** que estimam os valores através de **intervalos de confiança** ;

- ✓ Através destes testes podemos fazer generalizações sobre a população usando uma amostra da mesma (segundo o teorema do limite central)

Probabilidade

Probabilidade define a % de Certeza

Probabilidade objetiva define a a certeza de um acontecimento

A probabilidade exclui os fenómenos aleatórios

(*p*): medida da incerteza da ocorrência de um acontecimento

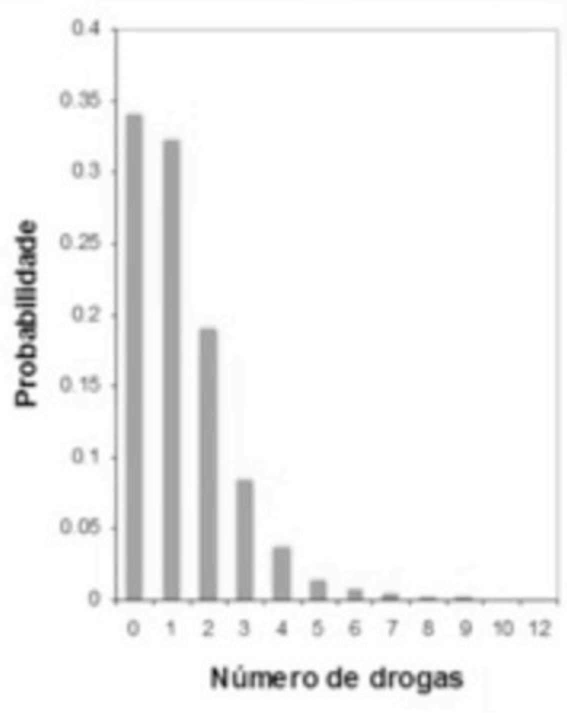
$$0 \leq p \leq 1$$

0	0,5	1
O Acontecimento provavelmente não ocorrerá	Máxima Certeza	O Acontecimento ocorrerá certamente

Probabilidade

Exemplo: Uso de drogas durante a gravidez:

Número de Drogas	Freqüência	P(X=x)
0	1425	0,3405
1	1351	0,3228
2	793	0,1895
3	348	0,0832
4	156	0,0373
5	58	0,0139
6	28	0,0067
7	15	0,0036
8	6	0,0014
9	3	0,0007
10	1	0,0002
12	1	0,0002
total	**4185**	**1,0000**



Probabilidade de uma mulher usar 9 drogas durante a sua gravidez é: 3/4185=0,0007

Através do cálculo da probabilidade podemos traçar uma **linha continua** no gráfico

Probabilidade

Exemplo: Uso de drogas durante a gravidez

A linha continua é uma função que associa uma probabilidade *p*(*X*), a cada valor de uma variável aleatória , *X*.

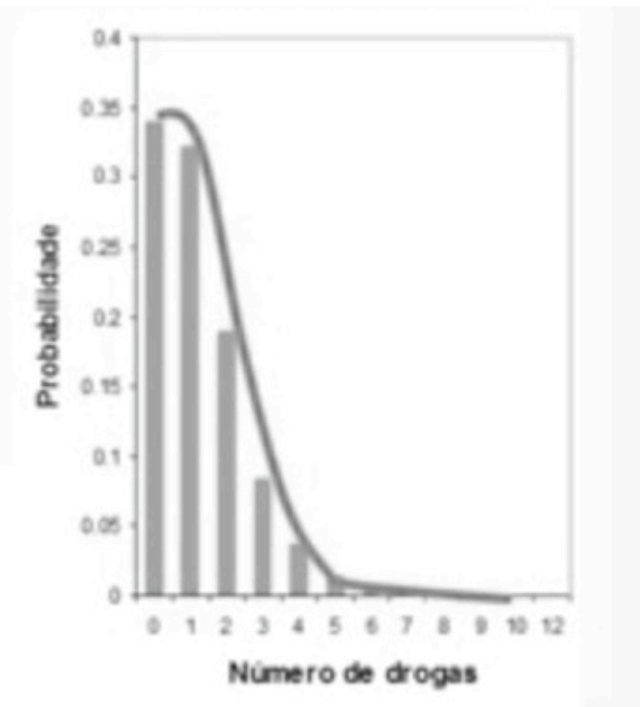
X – Número de drogas

1. $0 \leq p(X) \leq 1$

2. $\sum p(X) = 1 \Rightarrow$ *discreto*

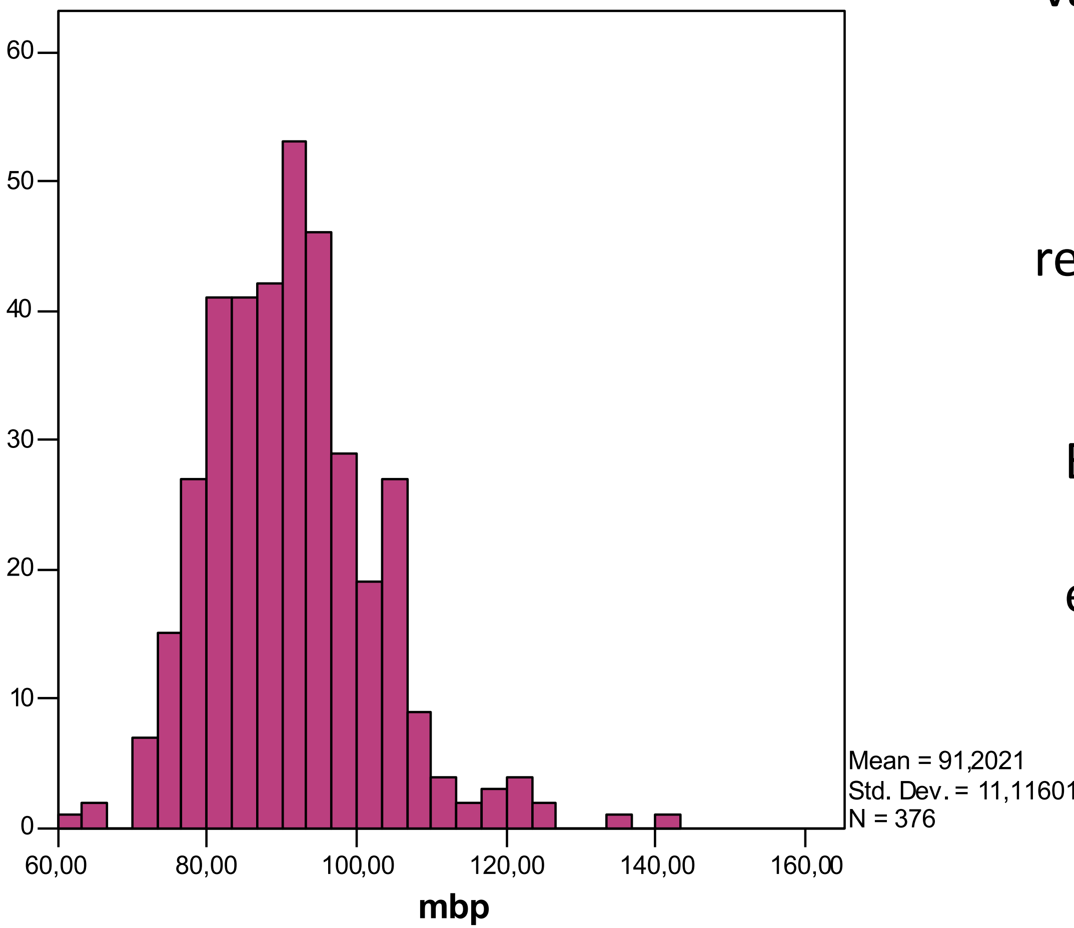
$\int f(X)dX = 1 \Rightarrow$ *continuo*

Função da densidade de probabilidade



Histograma
Curva normal
Probabilidade

Histograma



Representação de variáveis continuas

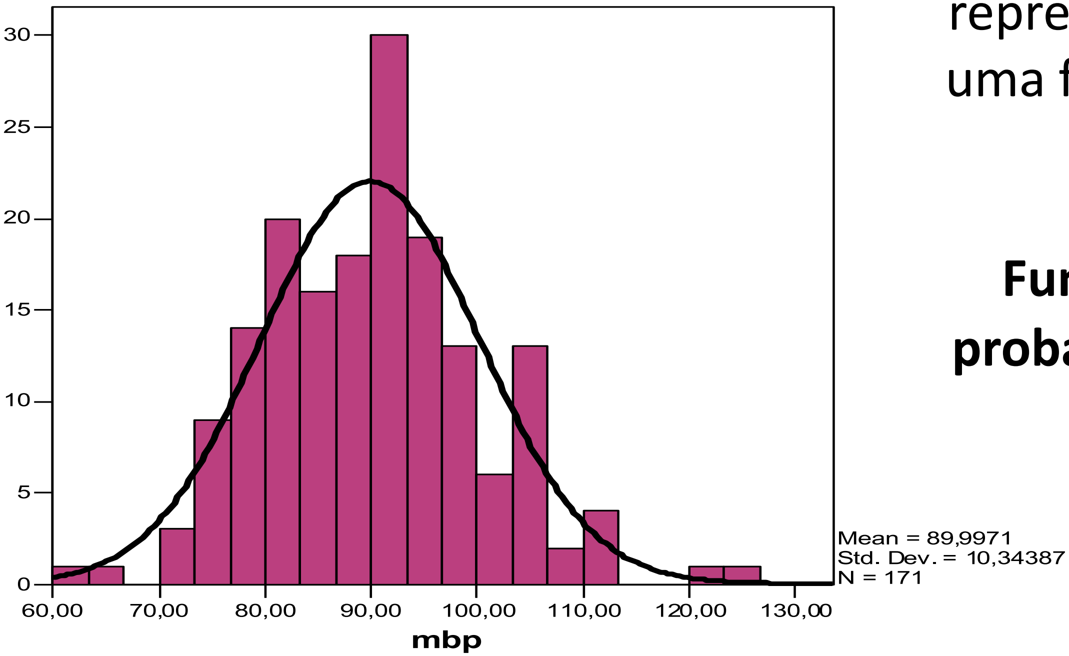
Apropriado para medidas de variáveis de intervalo, razão e ás vezes variáveis ordinais

A largura das barras representa o numero de casos (frequência)

É utilizado para representar percentagens uma vez que estas são mais significativas que os números absolutos

Curva normal
Probabilidade

Histograma



Probabilidade empírica = contagem/ total

Valor esperado de probabilidade: permite representar o todo; carece de uma função que seja aplicável no geral.

Função de Densidade de probabilidade & Distribuição Normal

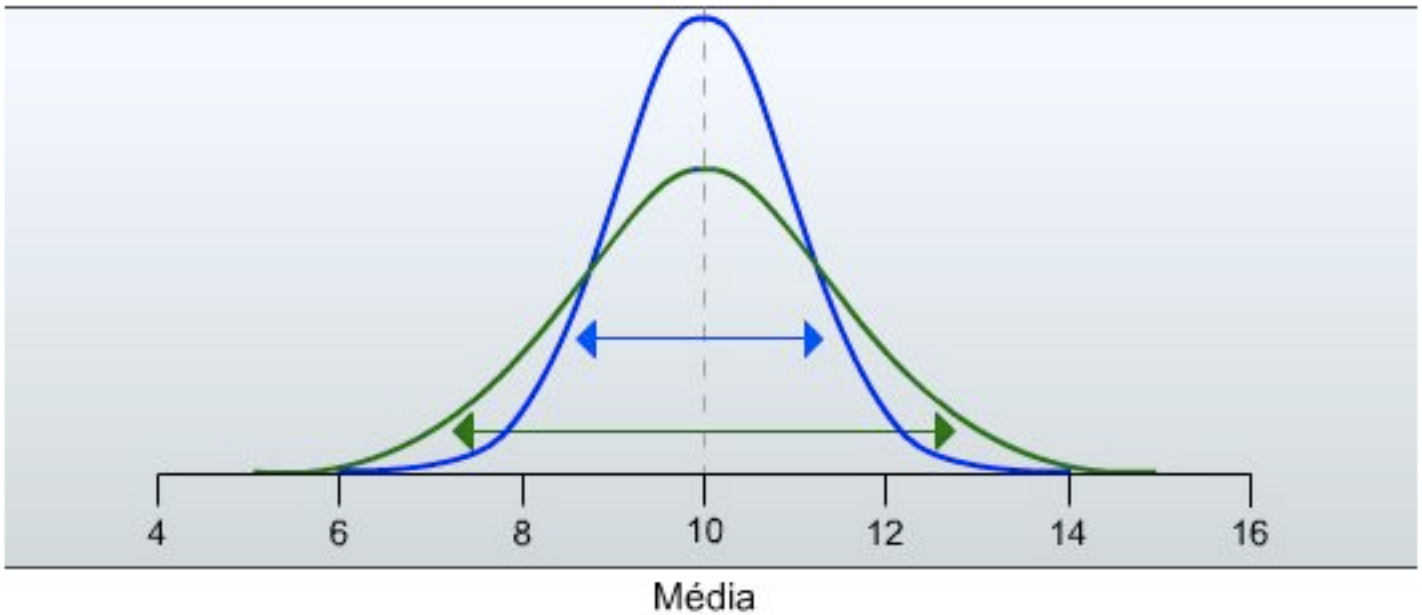
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Probabilidade

Curva Normal

Através de médias e desvio padrão conhecidos (variável quantitativa continua), consegue-se determinar a probabilidade esperada para uma variável através da função de densidade de probabilidade.

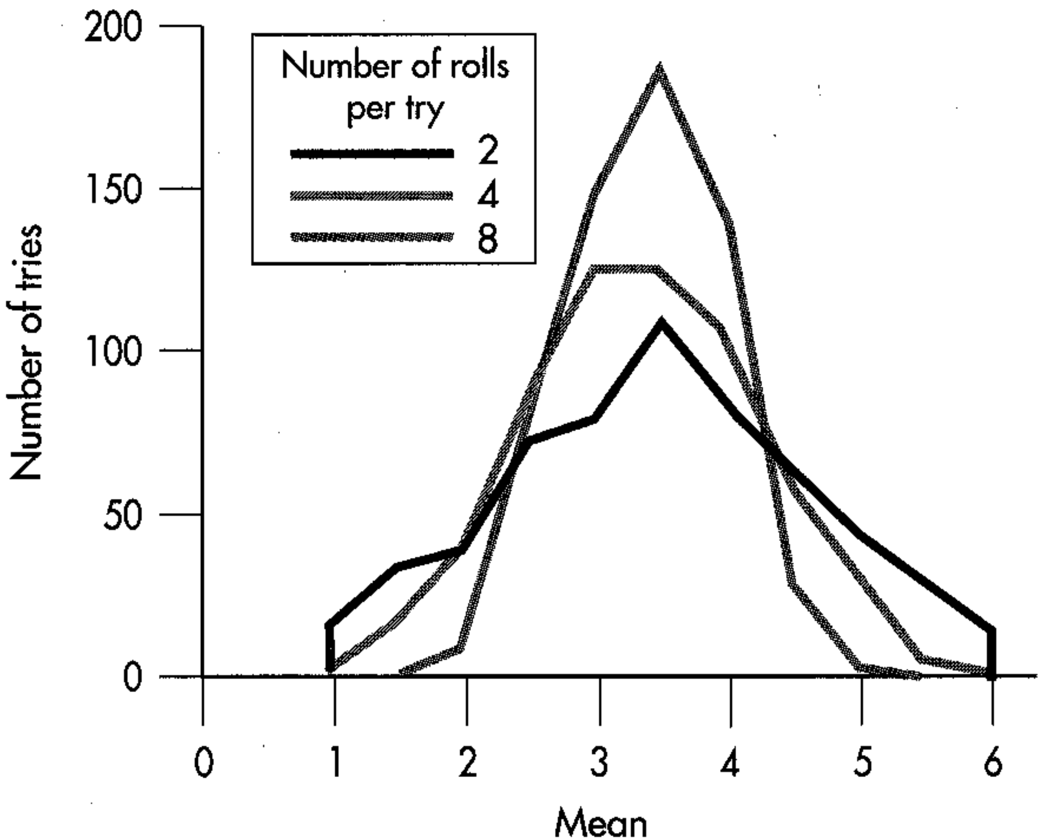
Curva Normal



Curvas Normais com médias iguais e desvios padrão diferentes

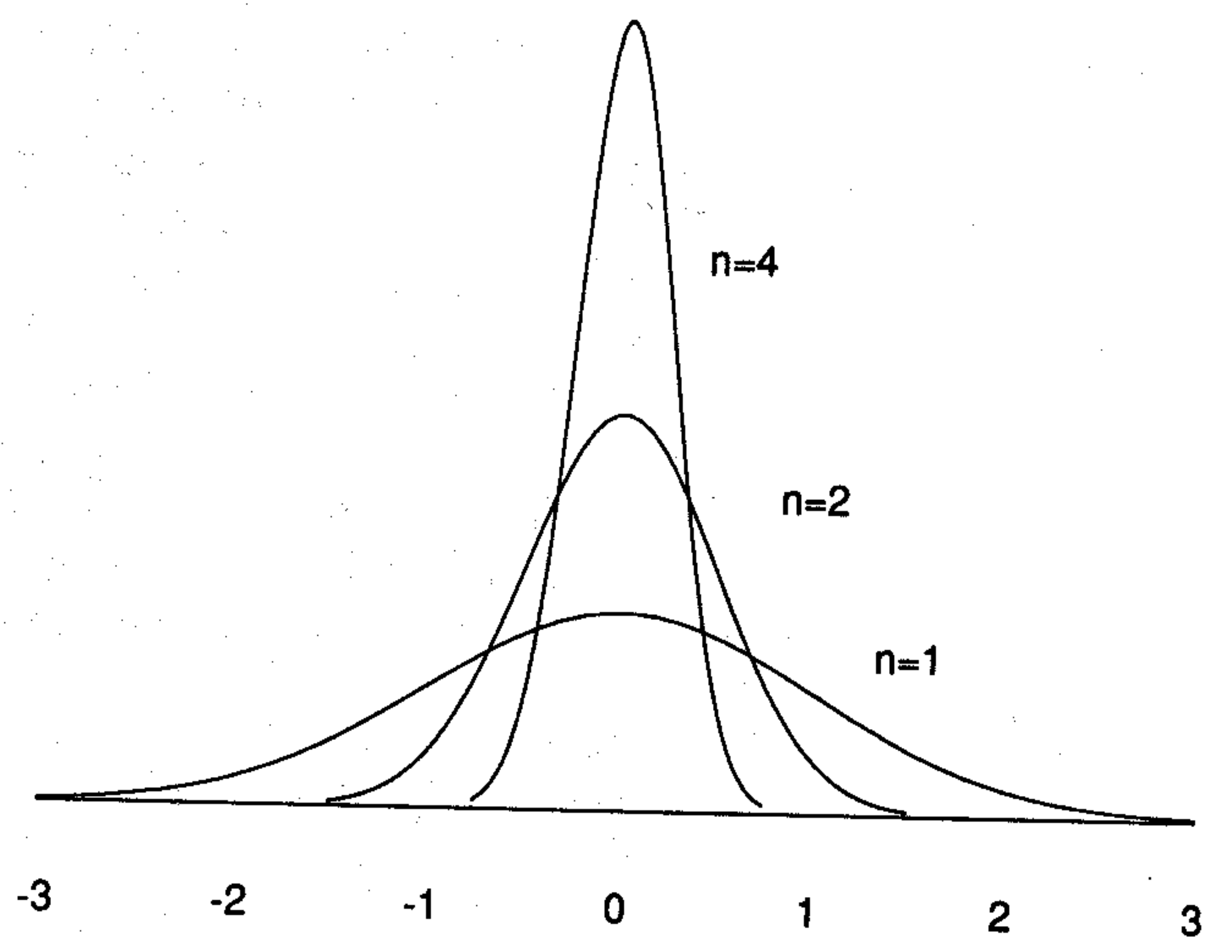
Qual das curvas representa uma amostra com maior *n*?

Curva Normal



Curvas Normais com médias iguais e desvios padrão diferentes

Curva Normal



3 distribuições de amostragem para médias de amostras aleatórias com

n = 1, n=2 e n = 4 de uma população

Estimação da média da população através da amostra

Estimação pontual

– Estimar uma característica da população através de um simples valor

- Estimador: Variável aleatória utilizada para estimar a característica da população
- Estimativa: Valor numérico do estimador

Se a amostra é retirada de uma população infinita, ou com reposição de uma população finita:

Média da população: $E\{\overline{X}\} = \mu$

Variância da população: $E\{s^2\} = \sigma^2$

Distribuição de média na amostragem

- Se a amostra é aleatória, a média da amostra também é aleatória

- Antes de fazer a seleção da amostra a média é uma variável aleatória à qual está associada uma distribuição de probabilidades

- O valor esperado de média na amostragem é : $\overline{X} : E\left\{\overline{X}\right\} = \mu$

- Variância de média na amostragem é: $\overline{X} : Var\left\{\overline{X}\right\} \text{ou } \sigma^2{\overline{x}} = \frac{\sigma^2}{n}$

- O desvio padrão de média na amostragem toma a designação de erro padrão da média e é igual a: $\frac{\sigma}{\sqrt{n}}$

Estimação por intervalos de confiança

- O intervalo de confiança (1 - α)% para um parâmetro, denominado de IC (1 - α)%, é o intervalo de valores (a e b), dentro do qual se encontra o valor do parâmetro que se pretende estimar, com uma probabilidade de (1 - α)%

- O valor de α mais usado é de 5%, o que corresponde ao cálculo de intervalos de confiança a 95%

- Portanto, um tal intervalo tem 95% de probabilidade de conter o parâmetro

Estimação por intervalos de confiança

Se a distribuição de média na amostragem é normal

- Através da normal reduzida é possível calcular a probabilidade da ocorrer entre 2 valores da distribuição

- A estimação dos valores da amostra para os valores da população é feita através de intervalos de confiança (desvios padrões da média)

- A média da amostra representa a perda de 1 grau de Liberdade (-1)

$$\boxed{Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

Então, $\text{Prob}\left\{ -z \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z \right\} = C$ ou $\text{Prob}\left\{ \bar{x} - z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z\frac{\sigma}{\sqrt{n}} \right\} = C$

Ou seja, $\text{Prob}\left\{ \bar{x} - z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}} \right\} = C$

C=Confiança=1-α (com α = risco)

Estimação por intervalos de confiança

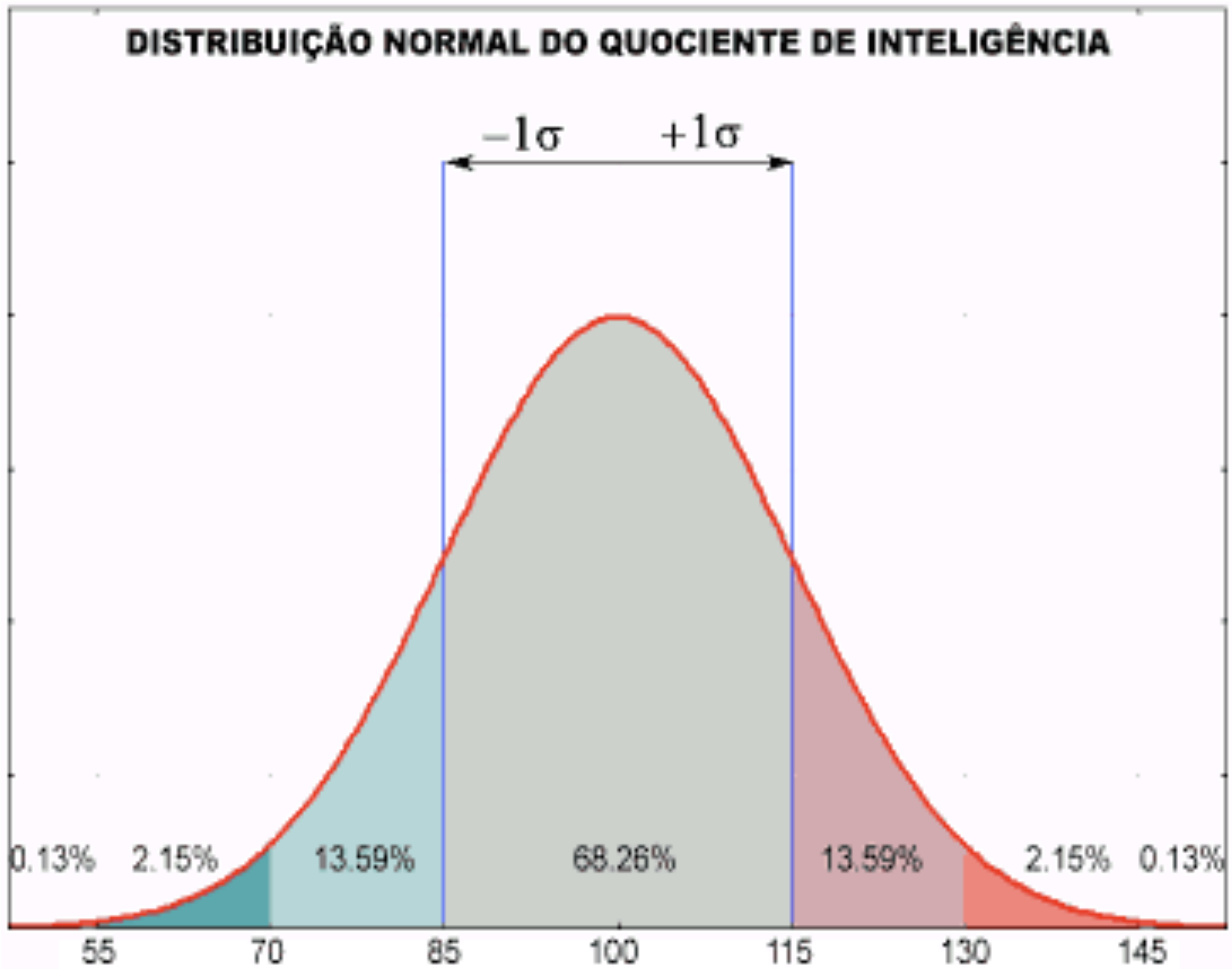
- O intervalo de confiança C (ou risco α) da estimação de μ é:

$$\left[\bar{x} - z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}}\right]$$

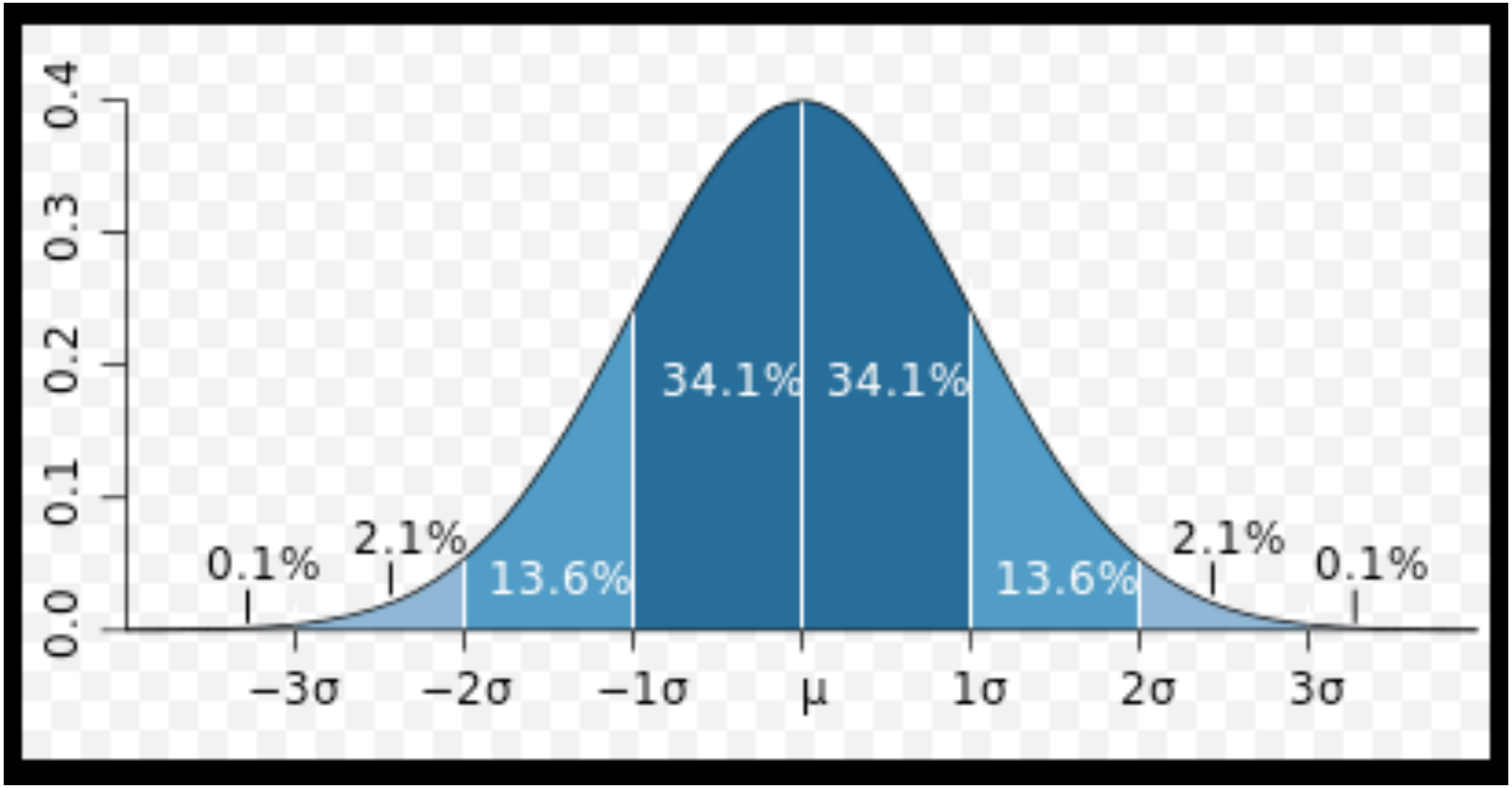
  - Se o desvio padrão da população não é conhecido:

$$\frac{s}{\sqrt{n}}$$ como estimador sem vício de $$\frac{\sigma}{\sqrt{n}}$$

Estimação por intervalos de confiança



Estimação por intervalos de confiança



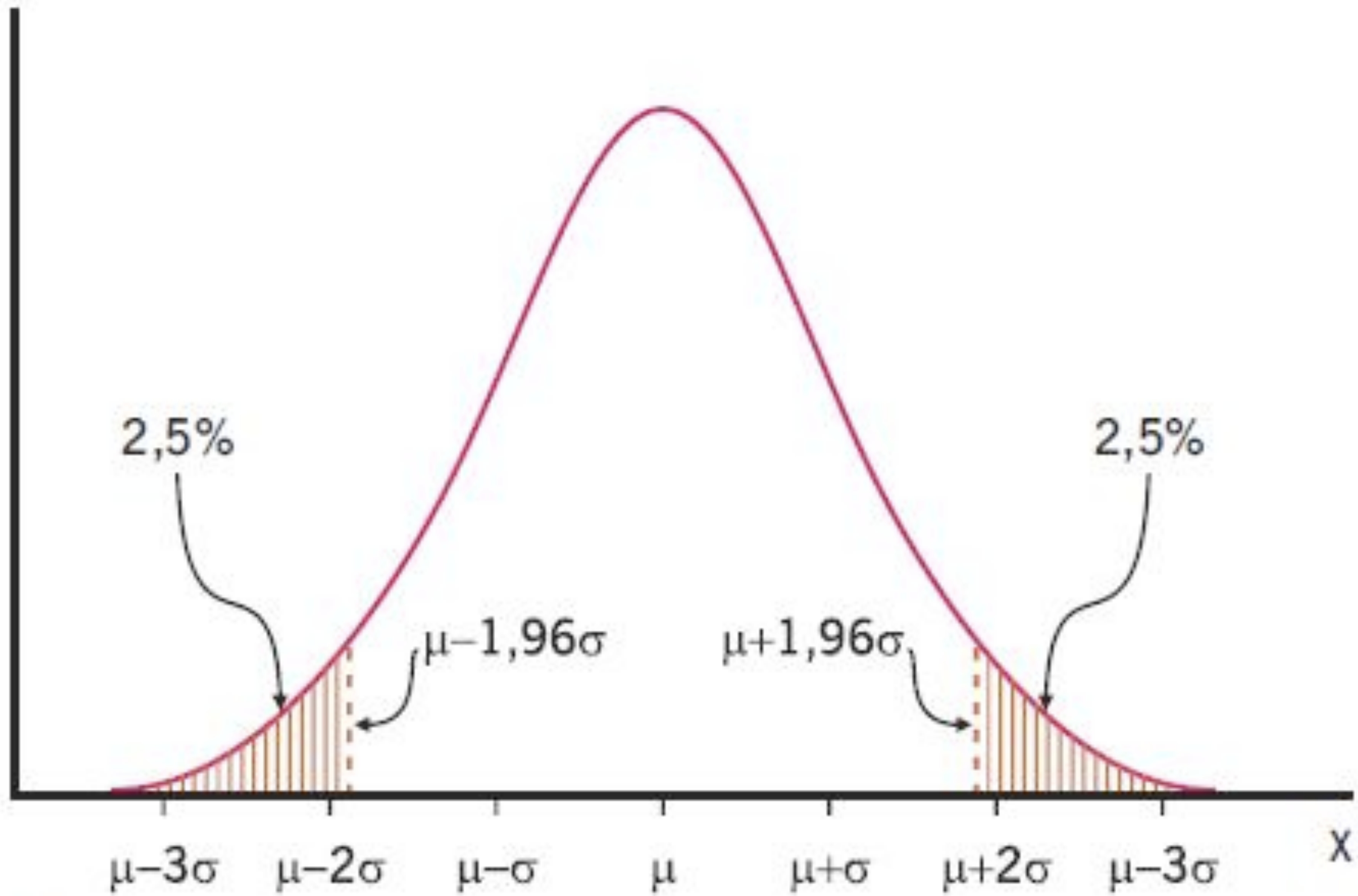
A curva é delimitada por números de desvio padrão iguais que correspondem a Graus de liberdade

A curva é simétrica e tem regiões definidas matematicamente o que torna esta função fácil de trabalhar

34 + 34 = área para 1 desvio padrão

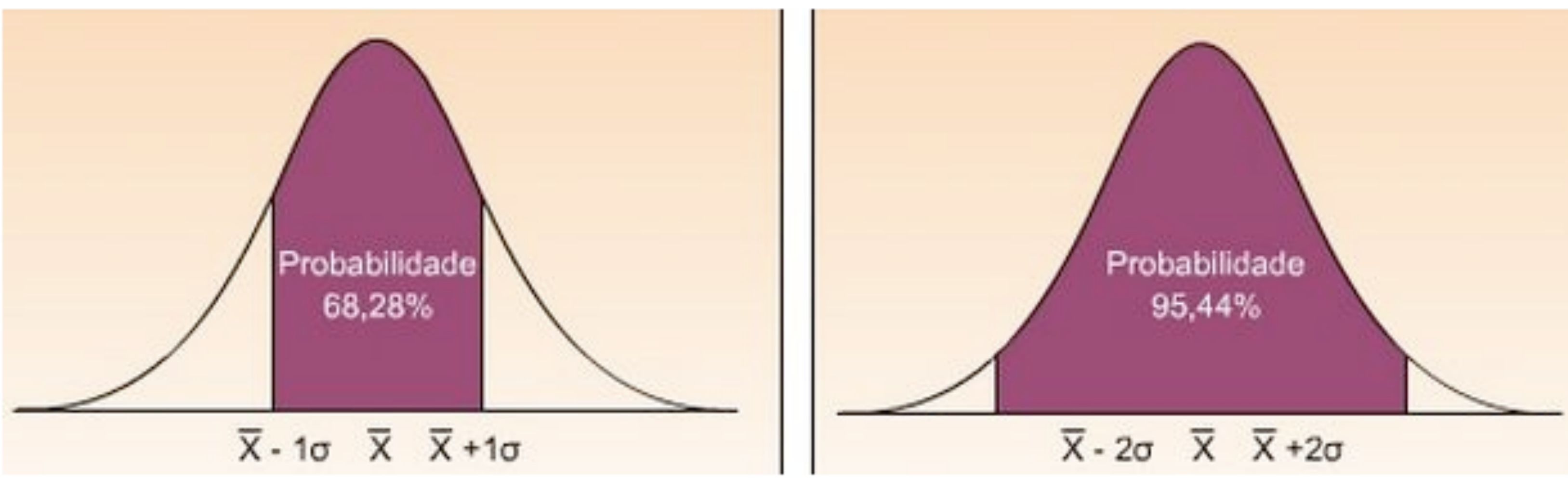
Estimação por intervalos de confiança

σ = SE
U = média



Curva Normal

Probabilidade estimada por intervalos de confiança



Curva Normal

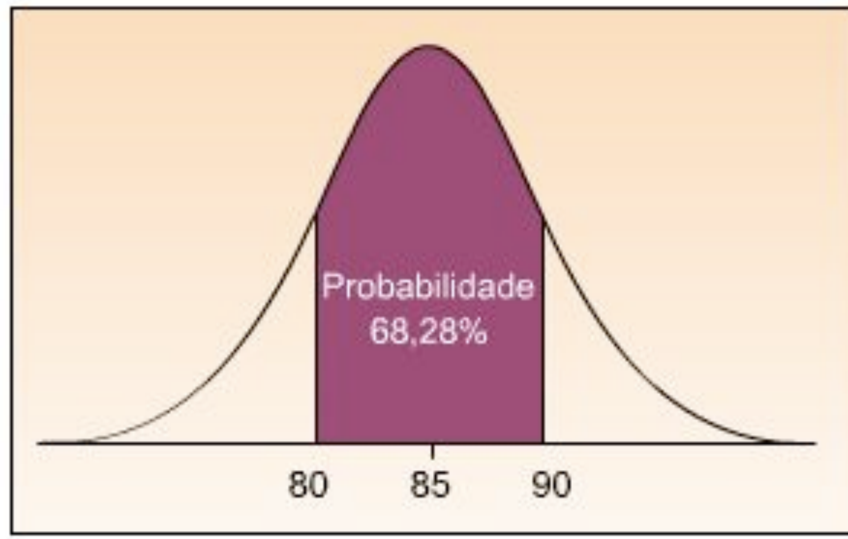
Probabilidade estimada por intervalos de confiança

Sabendo que a média da altura dos pinheiros é 85cm e que o desvio padrão é 5cm:

Neste caso, podemos esperar que **68,28%** dos pinheirinhos tenham alturas contidas no intervalo entre **80cm e 90cm**, uma vez que:

- A média menos 1 desvio-padrão é 80cm (85 - 5); e

- A média mais 1 desvio-padrão é 90cm (85 + 5).



Curva Normal

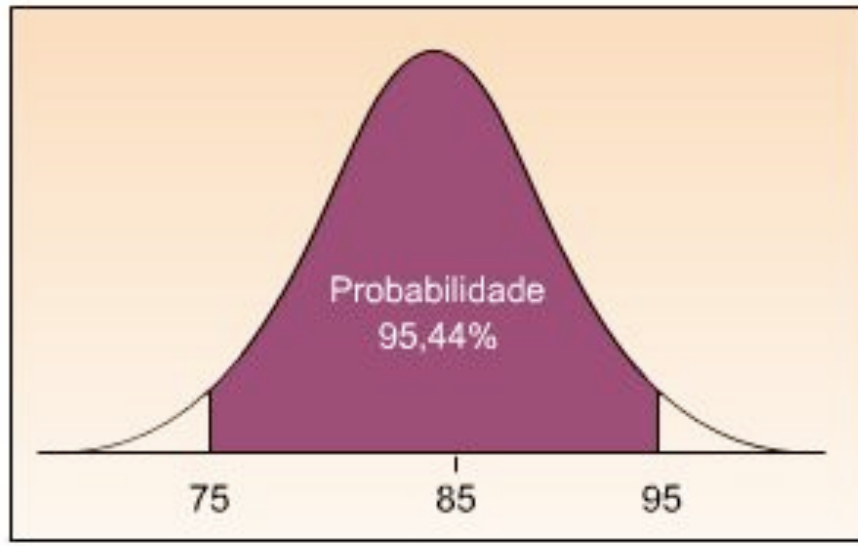
Probabilidade estimada por intervalos de confiança

Sabendo que a média da altura dos pinheiros é 85cm e que o desvio padrão é 5cm:

Também podemos dizer que **95,44%** dos pinheirinhos têm alturas que estão no intervalo de **75cm até 95cm**, uma vez que:

- A média menos 2 desvios-padrão é 75cm (85 - 2x5); e
- A média mais 2 desvios-padrão é 95cm (85 + 2x5).

Portanto, apenas 4,56% (100 - 95,44) restantes estão fora desse intervalo.



Curva Normal
Probabilidade estimada por intervalos de confiança

Exemplo:

- Utilizando os dados das 78 mortes por síndroma de morte súbita:
- X = peso à nascença dos bebés em gramas
- $\bar{X}$= 2993,6 ≈ 2994 g
- Sabe-se que σ = 800 g

- Calcule o intervalo de confiança a 95% para a média dos casos de morte súbita

$$2994 \pm (1{,}96)\left(\frac{800}{\sqrt{78}}\right) = 2994 \pm (1{,}96)(90{,}6) = 2994 \pm 178$$

  – Limite inferior: 2816; limite superior: 3172
  – O intervalo de confiança para 95% é [2816, 3172]
  – Há 95% de confiança de que a média da população do peso à nascença dos bebés com o síndroma da morte súbita se encontra neste intervalo.

Curva Normal

- A curva normal de uma distribuição de dados varia segundo o *N* e a variabilidade da amostra

- Permite interpretar os dados corretamente

- A curva normal permite ver a dispersão dos dados e consequentemente tratamento estatístico dos mesmo.

- Através do cálculo integral a curva normal pode ser padronizada:

- A distribuição normal de média □=0 e □^{2}=1 é denominada distribuição normal reduzida

- A <u>**variável Z**</u> (variável continua) obtém-se da variável X (entre 0 e 1):

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

Curva Normal - Teste Z

A distribuição normal de média $\mu=0$ e $\sigma^2=1$ é denominada distribuição normal reduzida

A variável Z (variável continua) obtém-se da variável X (entre 0 e 1):

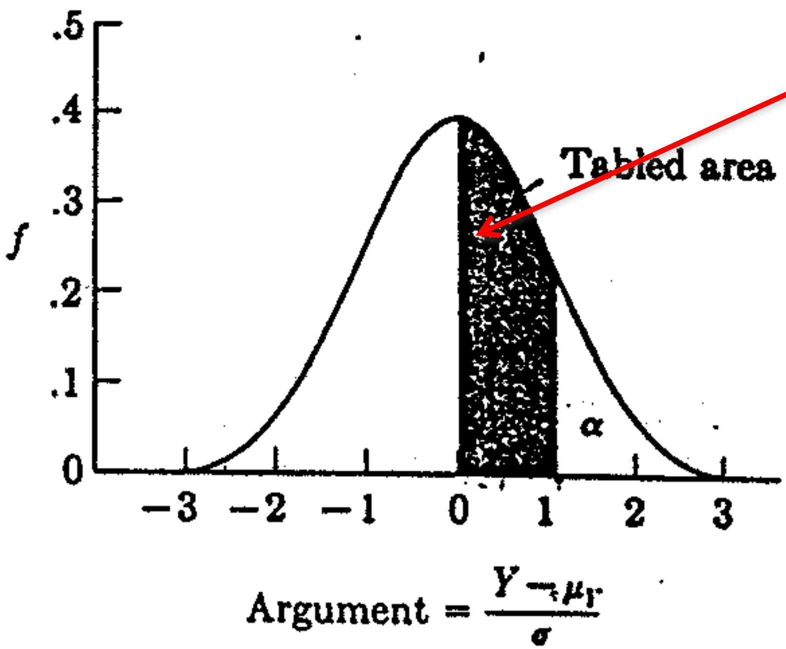
$$Z = \frac{X - \mu}{\sigma}$$

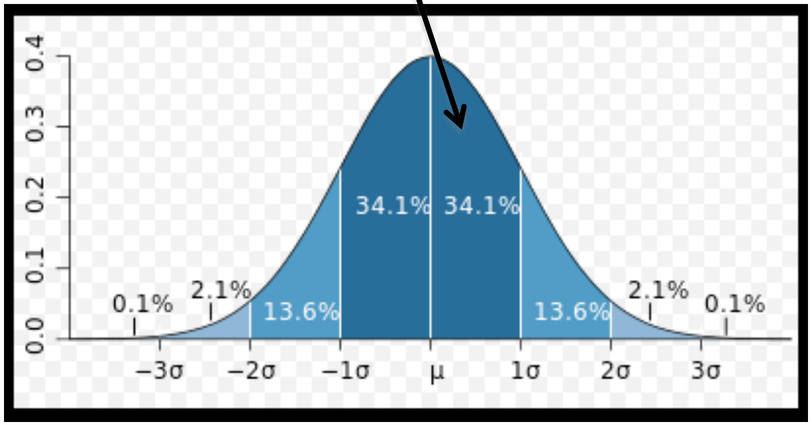
$Z > 0 \rightarrow X > u$

$Z < 0 \rightarrow X < u$

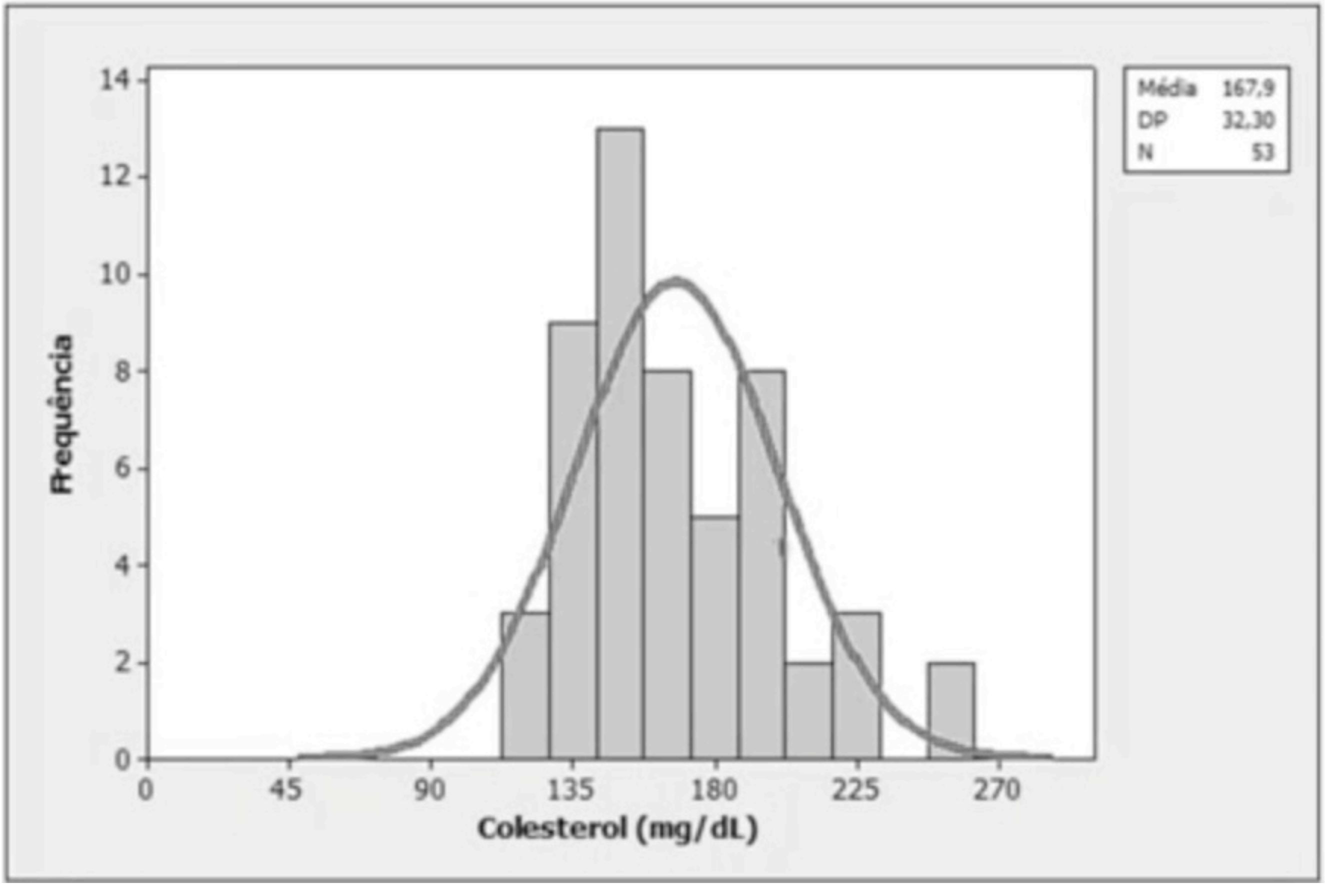
Curva Normal - Teste Z

P (0 $\leq$ Z $\leq$ 1) = 0,34





Curva Normal - Teste Z



Média: 167,9
Desvio padrão 32,30
N= 53

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

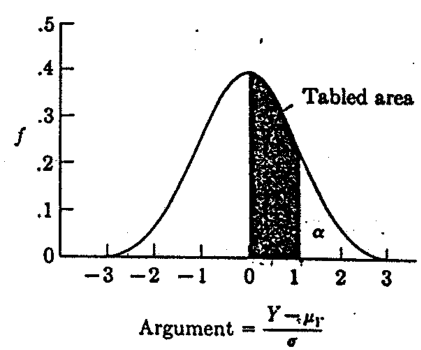
- Qual a probabilidade de escolher uma pessoa ao acaso e ela ter valores de colesterol entre 90 e 167,9 (média)?

Z = (90-167,9)/32,3 = -2,41

- 90 é distante da média -2,41 valores

→ Com este numero podemos ver na tabela a probabilidade

Standard deviation units	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.499767									
3.6	.499869									
3.7	.499892									
3.8	.499928									
3.9	.499952									
4.0	.499968									
4.1	.499979									
4.2	.499987									
4.3	.499991									
4.4	.499995									
4.5	.499997									
4.6	.499998									
4.7	.499999									
4.8	.499999									
4.9	.500000									



Média: 167,9
Desvio padrão 32,30
N= 53

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

Z = (90-167,9)/32,3 = -2,41

Curva normal Probabilidade

Standard deviation units	0	1	2	3
0.0	.0000	.0040	.0080	.0120
0.1	.0398	.0438	.0478	.0517
0.2	.0793	.0832	.0871	.0910
0.3	.1179	.1217	.1255	.1293
0.4	.1554	.1591	.1628	.1664
0.5	.1915	.1950	.1985	.2019
0.6	.2257	.2291	.2324	.2357
0.7	.2580	.2611	.2642	.2673
0.8	.2881	.2910	.2939	.2967
0.9	.3159	.3186	.3212	.3238
1.0	.3413	.3438	.3461	.3485
1.1	.3643	.3665	.3686	.3708
1.2	.3849	.3869	.3888	.3907
1.3	.4032	.4049	.4066	.4082
1.4	.4192	.4207	.4222	.4236
1.5	.4332	.4345	.4357	.4370
1.6	.4452	.4463	.4474	.4484
1.7	.4554	.4564	.4573	.4582
1.8	.4641	.4649	.4656	.4664
1.9	.4713	.4719	.4726	.4732
2.0	.4772	.4778	.4783	.4788
2.1	.4821	.4826	.4830	.4834
2.2	.4861	.4864	.4868	.4871
2.3	.4893	.4896	.4898	.4901
2.4	.4918	.4920	.4922	.4925
2.5	.4938	.4940	.4941	.4943
2.6	.4953	.4955	.4956	.4957

Média: 167,9
Desvio padrão 32,30
N= 53

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

Z = (90-167,9)/32,3 = -2,41

- A probabilidade de escolher um paciente ao acaso com valores entre 90 e a média é de 49%

Curva normal Probabilidade

Trabalho Prático I : Estimação por intervalos de confiança

Exercício:

Quais os valores do intervalo de confiança de 95% estimados para uma média de 100 e desvio padrão de 27, numa amostra com *n*=81 ?

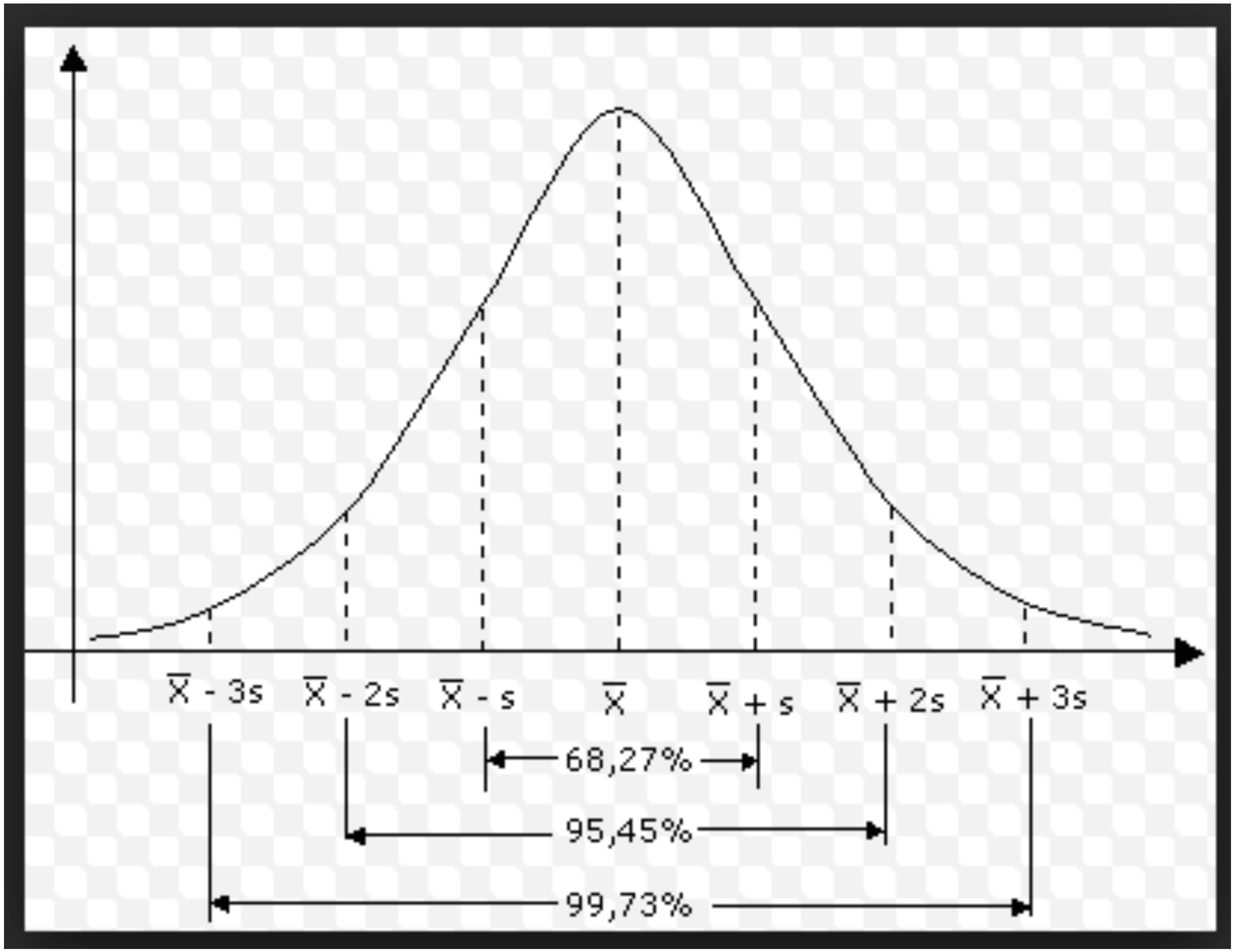
Cálculo do erro padrão:

$$\mathbf{SE = SD/\sqrt{\mathit{n}}}$$

Cálculo do Intervalo de confiança de 95% (95% CI)

$$\mathbf{95\% = \bar{X} \pm 1.96\ (SE)}$$

Trabalho Prático I : Estimação por intervalos de confiança



95% = $\overline{X} \pm 1.96$ (SE)

99% = $\overline{X} \pm 2.58$ (SE)

SE = SD/√*n*

SE = Standard Error = Erro Padrão

SD = Standard Desviation = Desvio Padrão

Trabalho Prático II : Estimação por intervalos de confiança

Exercícios:

Os valores de uma determinada investigação formam uma distribuição normal.
Média = 70; SD= 15

1.68% dos valores estão entre _______ e ________

2.96% dos valores estão entre _______ e ________

3.Numa distribuição assimétrica positiva os valores estendem-se à _______ ou estão representados valores mais ______ da distribuição

(Direita/esquerda & altos/baixos)