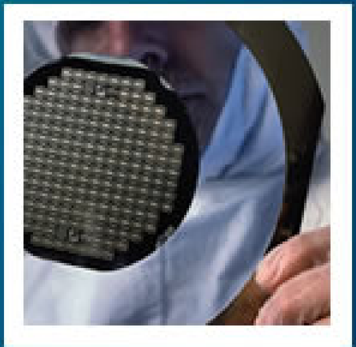
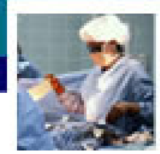
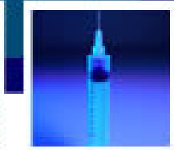
Bioestatística
Aula 5
Distribuções de Probabilidade



Prof. Tiago A. E. Ferreira

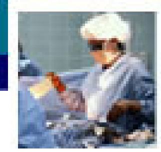
Conceito Função de Probabilidade

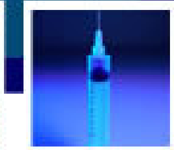




- Dada uma variável aleatória (v.a.) X que pode assumir os valores,
  - $X=\{x_1, x_2, ..., x_n\}$
  - Onde um certo valor x_i tem a probabilidade de ocorrência dada pela função $f(x_i)$
    - *Ou seja,* $x_i \rightarrow f(x_i)$
    - $f(x_i)$ é chamada de **função de probabilidade**

Construção da Função de Probabilidade

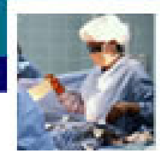


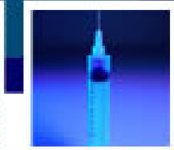


- Suponha que a v.a. X represente o número de óbitos.
  - X={0, 1} (0=não óbito; 1= óbito)
    - Suponha também que um paciente ao dar entrada em uma terapia intensiva tenha probabilidade de 25% de óbito.
    - Assim, se um paciente der entrada na CTI (n=1)

$$X = \{0,1\} \begin{cases} f(0) = 0{,}75 \\ f(1) = 0{,}25 \end{cases}$$

Construção da Função de Probabilidade





- Se dois pacientes derem entrada na CTI (n=2)

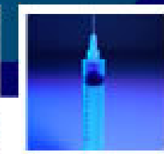
$$X=\{0,1,2\}\begin{cases} f(0)\to p(v_1)\cdot p(v_2)=0{,}75\cdot 0{,}75=0{,}5625 \\ f(1)\to \begin{cases} p(v_1)\cdot p(o_2)=0{,}75\cdot 0{,}25=0{,}1875 \\ p(o_1)\cdot p(v_2)=0{,}25\cdot 0{,}75=0{,}1875 \end{cases} \\ f(2)\to p(o_1)\cdot p(o_2)=0{,}25\cdot 0{,}25=0{,}0625 \end{cases}$$

- Ou,

x	0	1	2	Soma
f(x)	0,5625	0,3750	0,0625	1

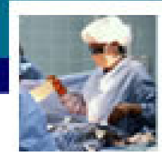
Construção da Função de Probabilidade

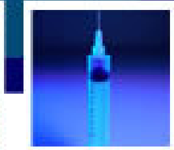




- Observe que tal procedimento para o calculo da função de probabilidade pode ser aplicado para um número qualquer de n, no caso, de pacientes!
  - Como calcular *f(x)* para n=1000
    - É possível ser feito, porém o custo computacional para se obter f(x) por este procedimento é muito alto!
    - Uma forma mais econômica de se obter f(x) é através das distribuições de probabilidades.

Distribuição Binomial



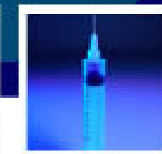


- Seja uma v.a. X binária.
  - Ou seja, a v.a. X tem duas possibilidades de ocorrências!
    - Probabilidade de ocorrer x, dada por *p*
    - Probabilidade de não ocorrer x, dada por *q*
      - Onde, $p + q = 1$ ou $P(X) + P(X') = 1$ ou *100%*
- A probabilidade de k ocorrências da v.a. X em n casos é dada por:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Distribuição Binomial





- O termo $\binom{n}{k}$ é a combinação n, k a k, onde

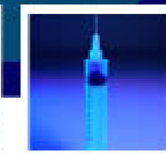
$$\binom{n}{k} = {}_nC_k = \frac{n!}{k!(n-k)!}$$

- Exemplo, dada um string binária de 5 bits
  - Quantas possíveis combinações existem com um bit igual a 1 e 4 bits iguais a 0:
    - 10000, 01000, 00100, 00010 e 00001
      - Ou seja, n=5, k=1,logo

$$ {}_5C_1 = \frac{5!}{1!(5-1)!} = \frac{5\cdot 4\cdot 3\cdot 2\cdot 1}{1(4\cdot 3\cdot 2\cdot 1)} = 5$$

Distribuição Binomial



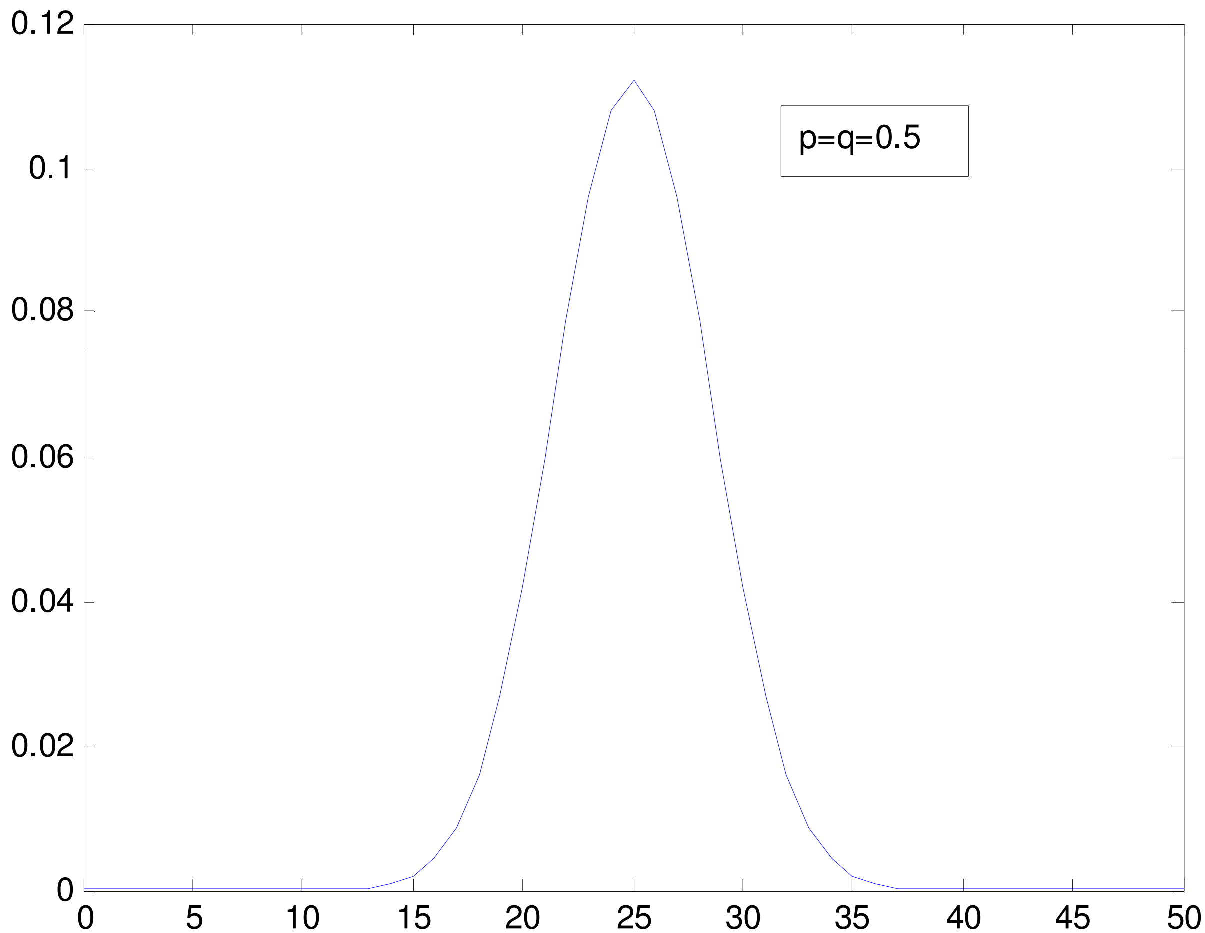


- Então, dada a função de probabilidade binomial,

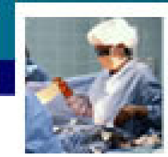
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

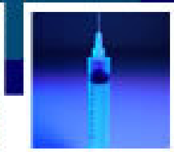
  – Esta irá gerar uma distribuição chamada **distribuição binomial**

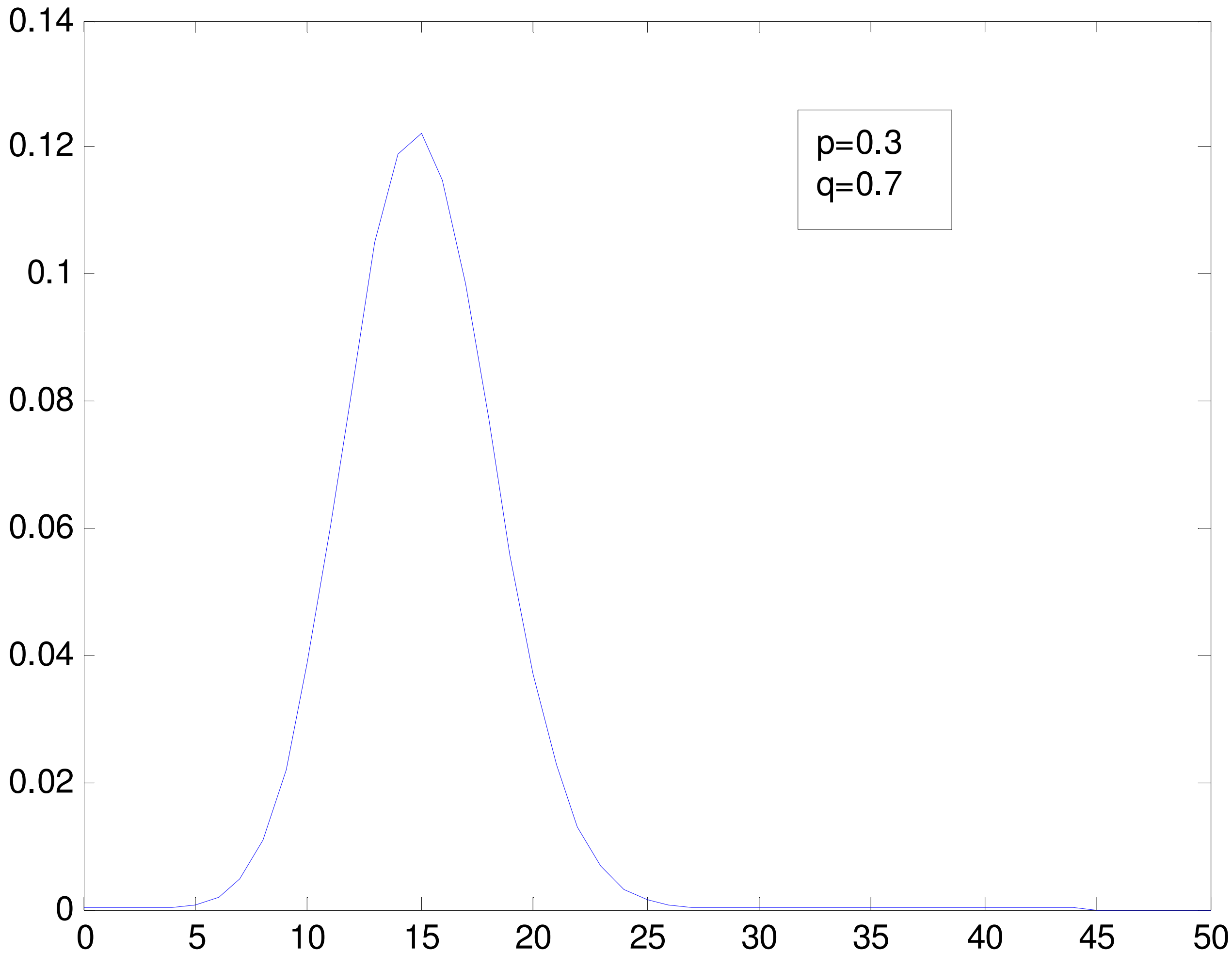
Distribuição Binomial



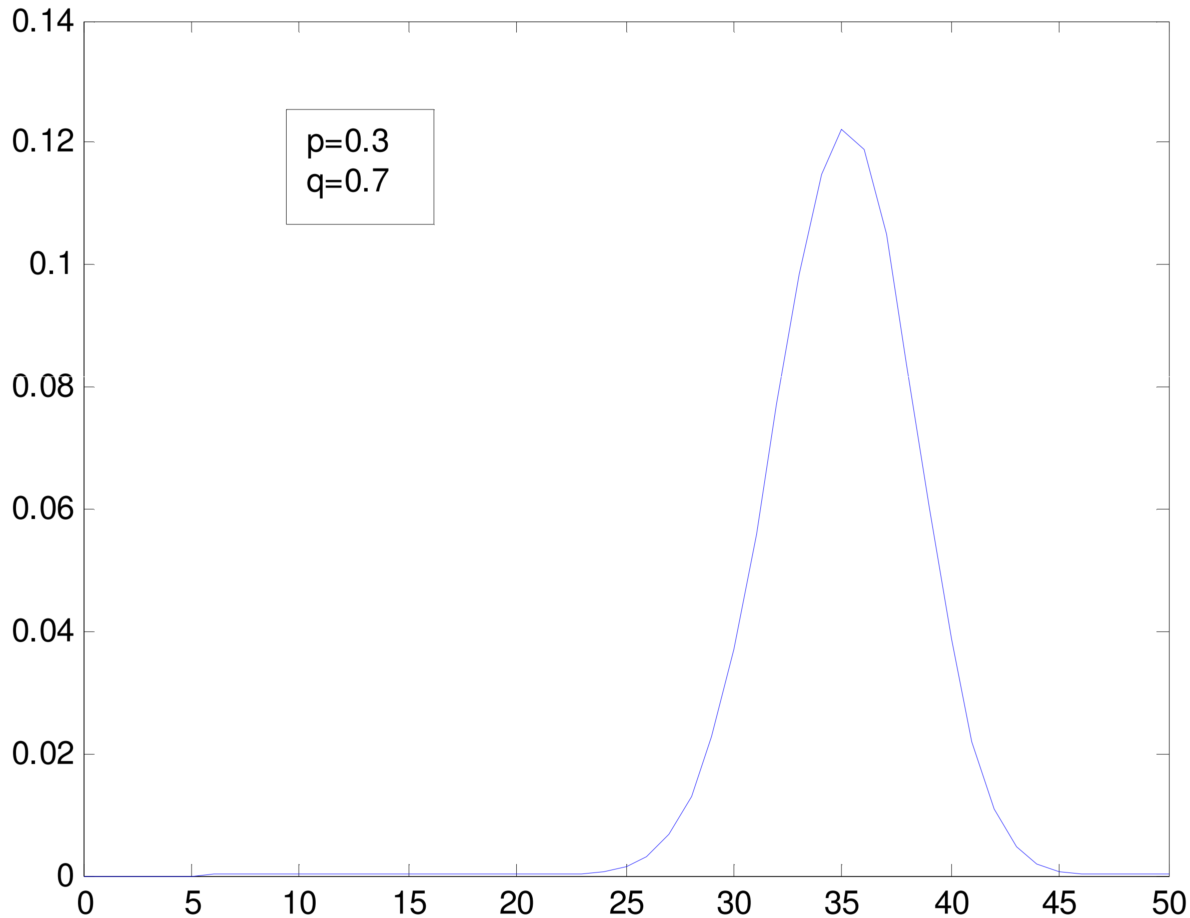
Distribuição Binomial



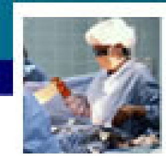


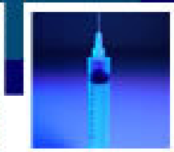


Distribuição Binomial



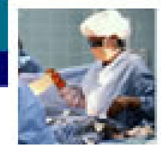
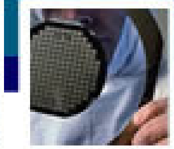
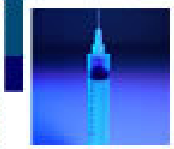
Distribuição Binomial





- Uma distribuição binomial tem os parâmetros:
  - Média = Valor Esperado = $E[X] = \mu = np$
  - Variância = $\sigma^2 = npq = np(1-p)$
  - Desvio Padrão = $\sigma = (npq)^{0,5} = (np(1-p))^{0,5}$

Exemplo

- Dado que a probabilidade de um indivíduo do sexo masculino (M), com mais de 60 anos (+60), com vida sedentária(S) e tabagista (T) de desenvolver uma doença cardiovascular (DCV) é de 40%. Dado também que este estudo foi realizado com 10 indivíduos (n=10), temos
  - $P(DCV|(M) \cap (+60) \cap (S) \cap (T)) = P(X) = p = 0,4$
  - Então a probabilidade de nenhum indivíduo ter DCV (X=0) será: $P(X=0) = {}_{10}C_0 0,4^0 0,6^{10} = 0,0060$ ou 0,60%

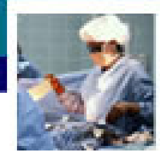
Exemplo

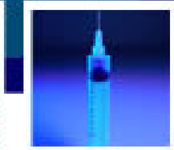
- A probabilidade de se ter menos de três indivíduos com DCV é
  - $P(X<3)=P(X=0,1,2) = P(X=0)+ P(X=1)+ P(X=2)$
    - $P(X=0) = 0.0060$
    - $P(x=1) = {}_{10}C_1 0,4^1 0,6^9 = 0,0403 = 4,03\%$
    - $P(X=2\ ) = {}_{10}C_2 0,4^2 0,6^8 = 0,1209 = 12,09\%$
  - Logo, $P(X<3) = 0,0060+0,0403+0,1209 = 0,1672$ ou 16,72%

Exemplo

- Qual seria a probabilidade de mais de dois indivíduos serem afetados por uma DCV?
  - $P(X>2) = P(X=3, 4, ..., 10) = P(X=3)+...+P(X=10)$
  - Ou, $P(X>2) = 1- P(X = 0, 1, 2)$
    - $P(X>2) = 1 – [P(X=0)+P(X=1)+P(X=2)]$
    - $P(X>2) = 1 – 0{,}1672 = 0{,}8328$ ou 83,28%
- Qual o número esperado de casos de DCV?
  - $\mu = 10 . 0{,}4 = 4$ casos! Com um desvio padrão de $(10.0{,}4.0{,}6)^{0{,}5} \cong 1{,}55$ casos

Distribuição de Poisson



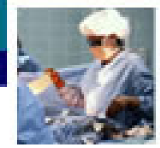


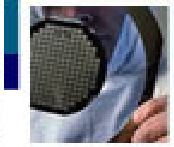
- A distribuição de Poisson é gerada a partir da função de probabilidade

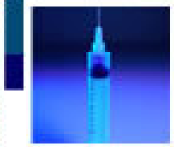
$$P(X = k_i) = \frac{e^{-n \cdot p} \cdot (n \cdot p)^{k_i}}{k_i!}$$

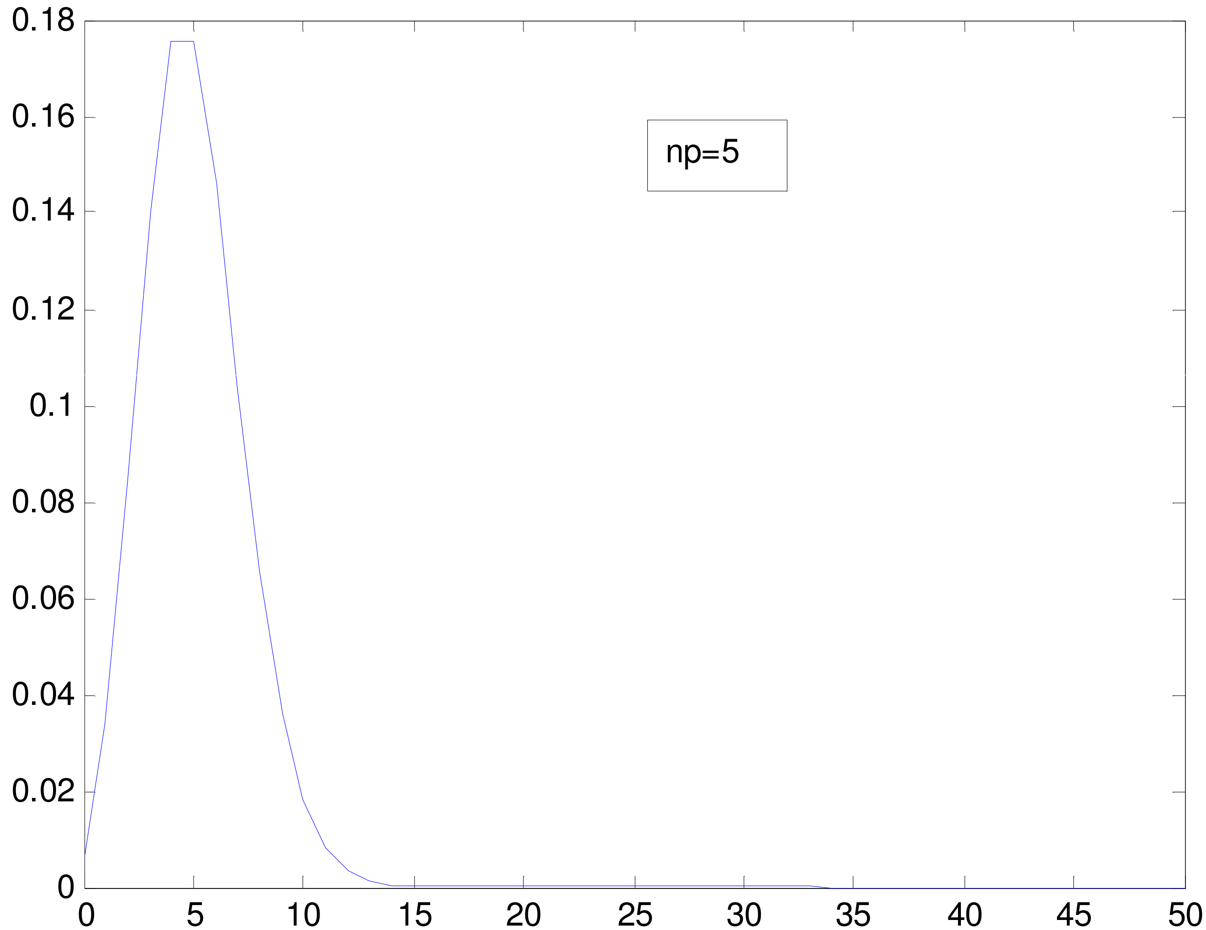
  - Onde $np \leq 7$, ou seja, esta distribuição está relacionada com eventos raros (valor de *p* baixo)

Distribuição de Poisson

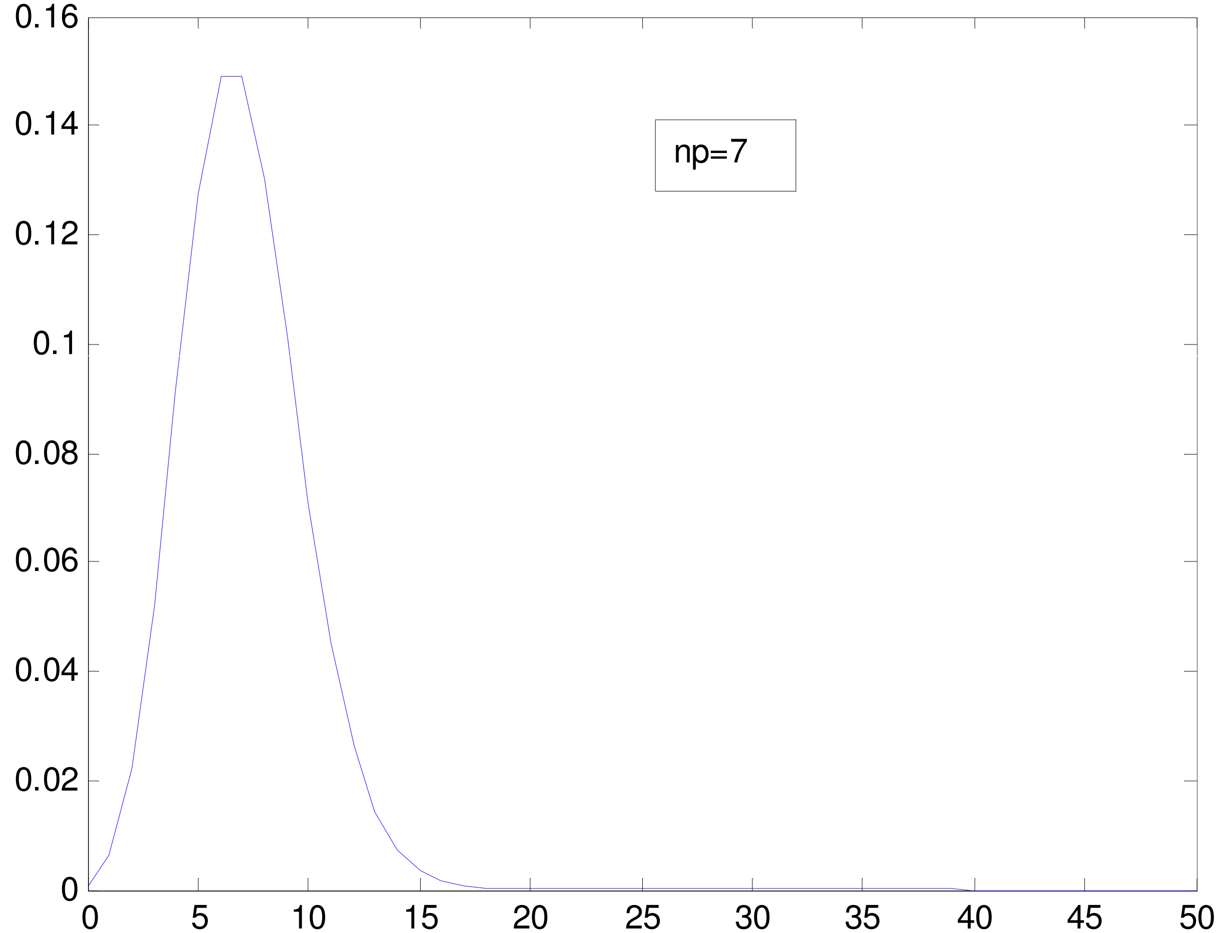




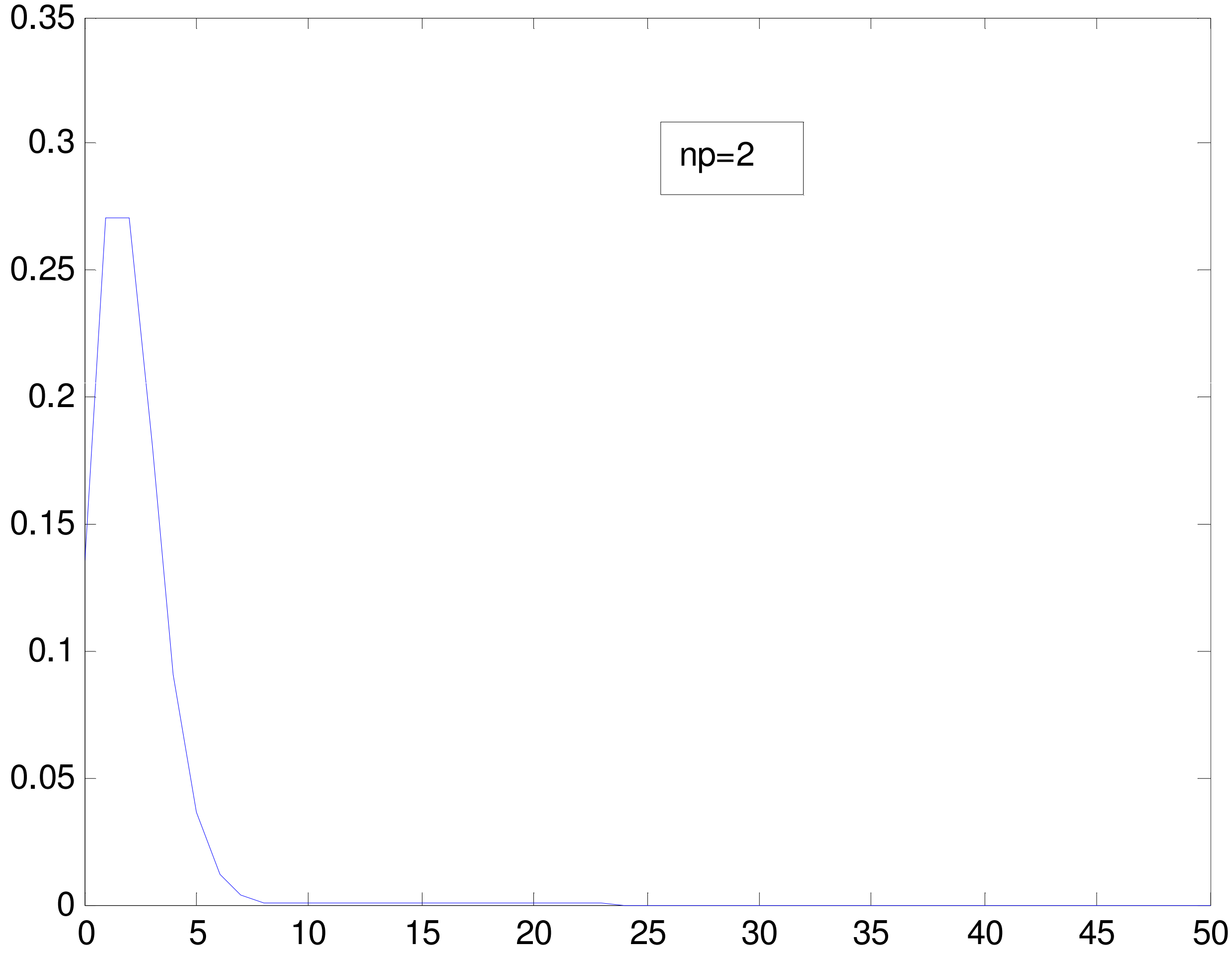




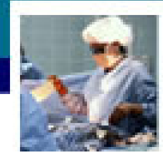
Distribuição de Poisson

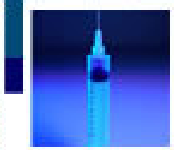


Distribuição de Poisson



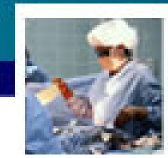
Distribuição de Poisson

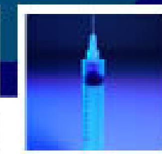




- Parâmetros da função de Probabilidade de Poisson
  - Média = Valor Esperado = $E[X] = \mu = np$
  - Variância = $\sigma^2 = \mu = np$
  - Desvio Padrão = $\sigma = (\mu)^{0,5} = (np)^{0,5}$

Exemplo





- Suponha que a cada mil pessoas que recebem uma certa droga, uma sofre uma reação negativa.
  - Assim, em um total de 500 cirurgias, onde foi empregada tal droga, a probabilidade de uma pessoa sofrer uma reação negativa é,
    - Dado que $\mu = np = 500.0{,}001 = 0{,}5$

$$P(X = 1) = \frac{e^{-0,5}(0,5)^1}{1!} = 0,3033 \; ou \; 30,33\%$$

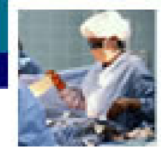
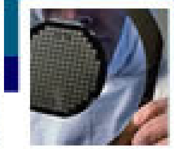
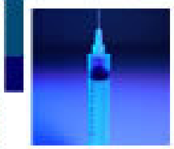
Exemplo

- A Probabilidade de nenhuma reação negativa seria,

$$P(X=1)=\frac{e^{-0,5}(0,5)^{0}}{0!}=0,6065\ ou\ 60,65\%$$

- E a Probabilidade de mais de uma reação
  - P(X>1) = 1 – P(X=0,1) = 1 – [P(X=0)+P(X=1)]
  - P (X>1) = 1 – (0,6065 + 0,3033) = 0,0902 ou 9,02%

Distribuição Normal ou de Gauss

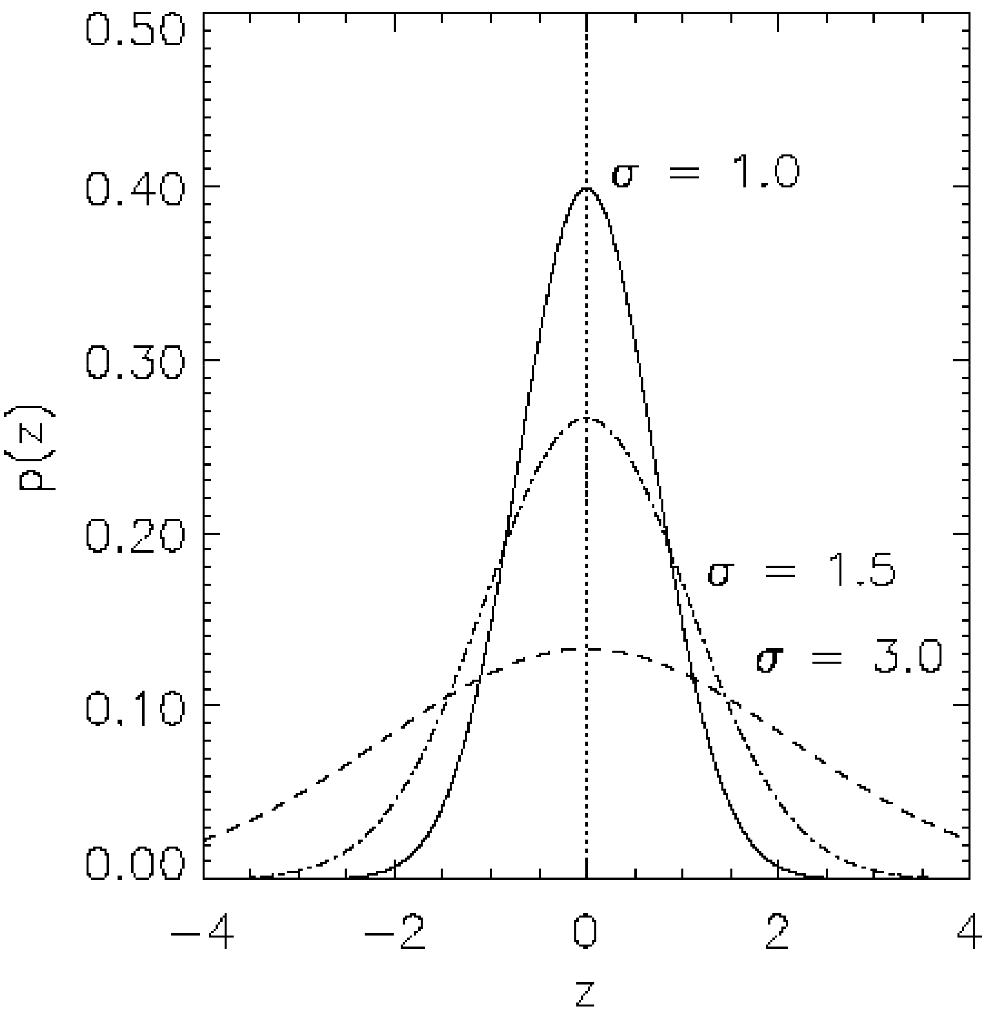
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\sigma \rightarrow$ *desvio padrão*

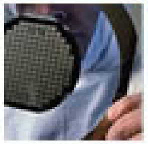
$\mu \rightarrow$ *média da distribuição*

$\frac{1}{\sigma\sqrt{2\pi}} \rightarrow$ *const. de normalização*

Ilustração da Distribuição Normal



Propriedades da Distribuição Normal

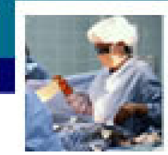



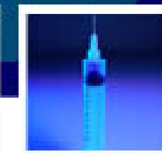

- Para um processo normalizado, temos que o somatório das probabilidades é igual a 1
  - Dada uma distribuição, é equivalente afirmar que a área total abaixo da função de probabilidade deve ser igual a 1 ou 100%.
  - Matematicamente isto pode ser escrito como

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

  - Assim, a área da distribuição é numericamente igual a sua probabilidade

Propriedades da Distribuição Normal





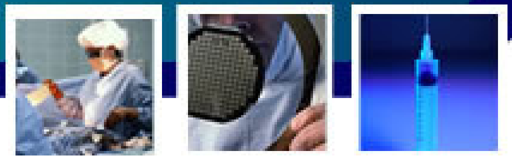
- Tenta a zero quando x tende para ± infinito
- Simétrica em torno da Tendência Central (Média = Mediana = Moda)
- Valores concentrados em torna da Tendência Central
  – As áreas (Probabilidades) para 1, 2 e 3 desvios padrões em torno da média são:

$$\int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 0{,}6826$$

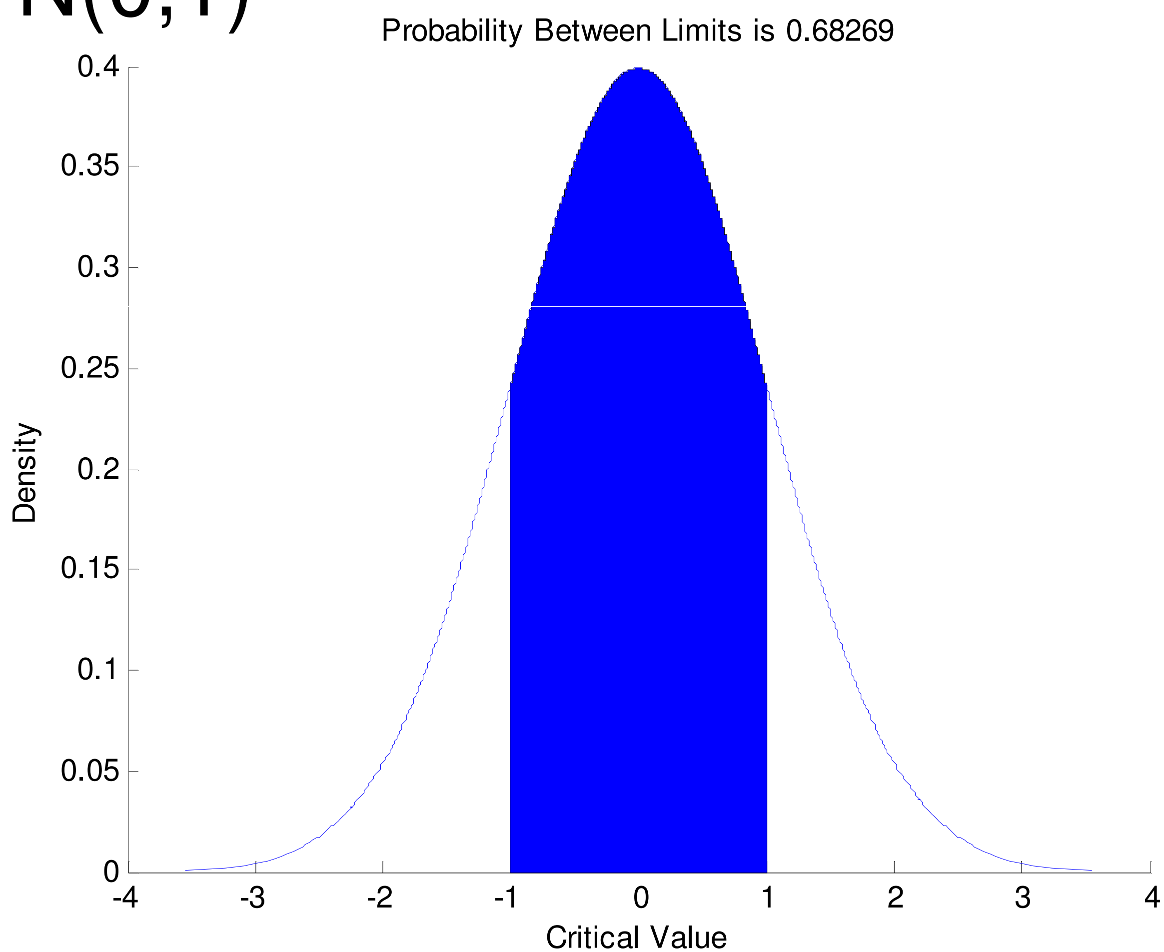
$$\int_{\mu-2\sigma}^{\mu+2\sigma} f(x)dx = 0{,}9546$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx = 0{,}9974$$

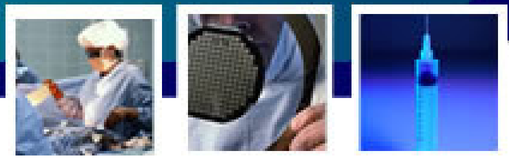
Propriedades da Distribuição Normal



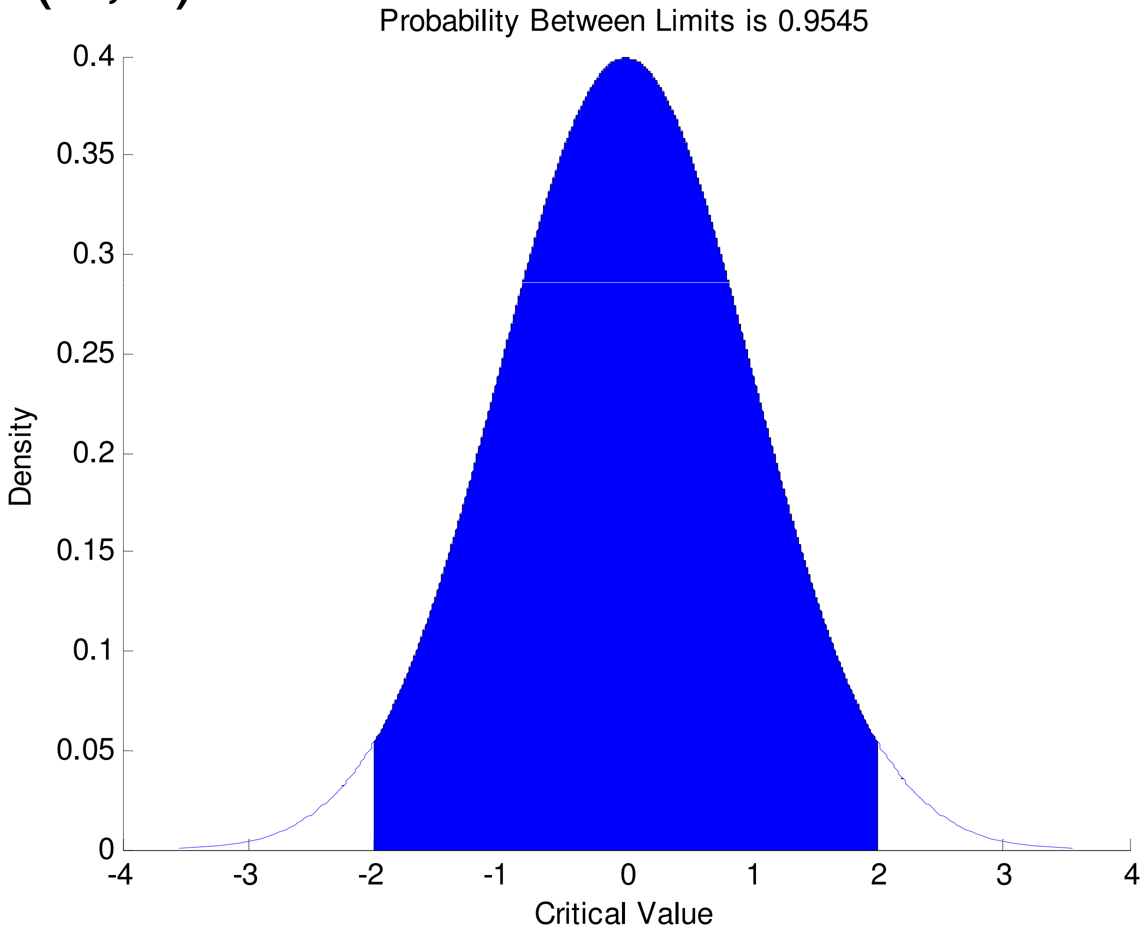
- Para N(0,1)



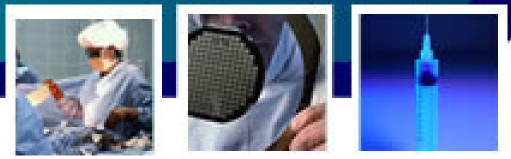
Propriedades da Distribuição Normal



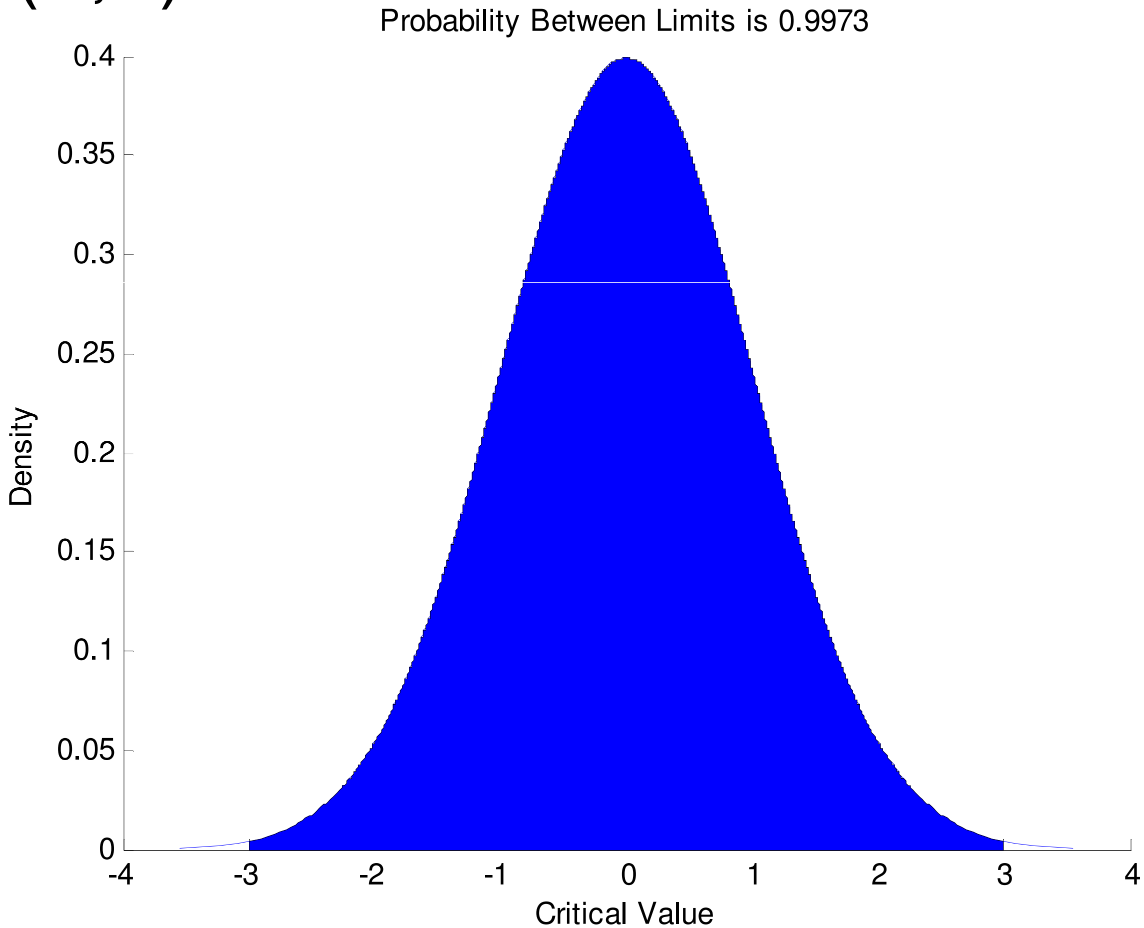
- Para N(0,1)



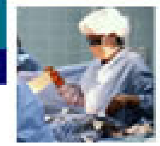
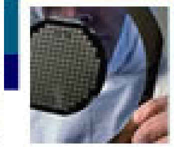
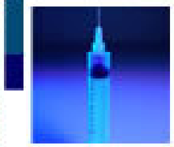
Propriedades da Distribuição Normal



- Para N(0,1)

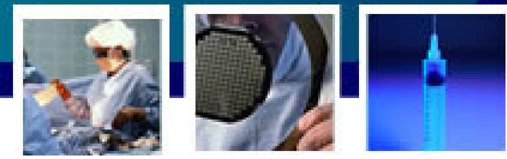


Variável Aleatória Padronizada

- A distribuição Normal costuma ser apresentada em tabelas, facilitando a consulta de sues valores
  - Para tanto deve-se padronizar os valores da média e do desvio padrão para que os dados sejam tabelados
    - Desta forma, padroniza-se para **média zero e desvio padrão unitário**
    - Assim, ao se comparar uma v.a. com uma distribuição normal deve-se garantir que esta v.a. tenha média nula e desvio padrão unitário

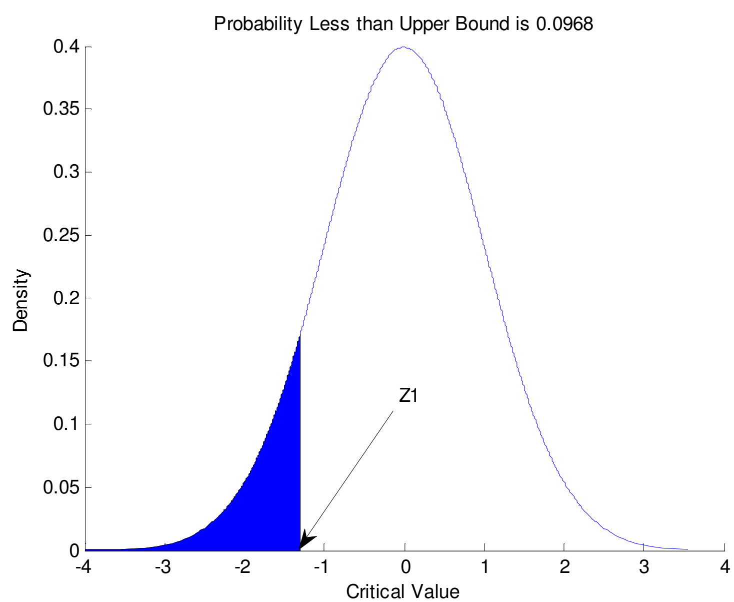
Variável Aleatória Padronizada



- A padronização de uma v.a. é feita por meio da relação

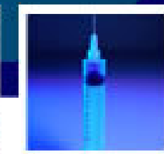
$$z = \frac{x - \mu}{\sigma}$$

- Tabelas:



Exemplo



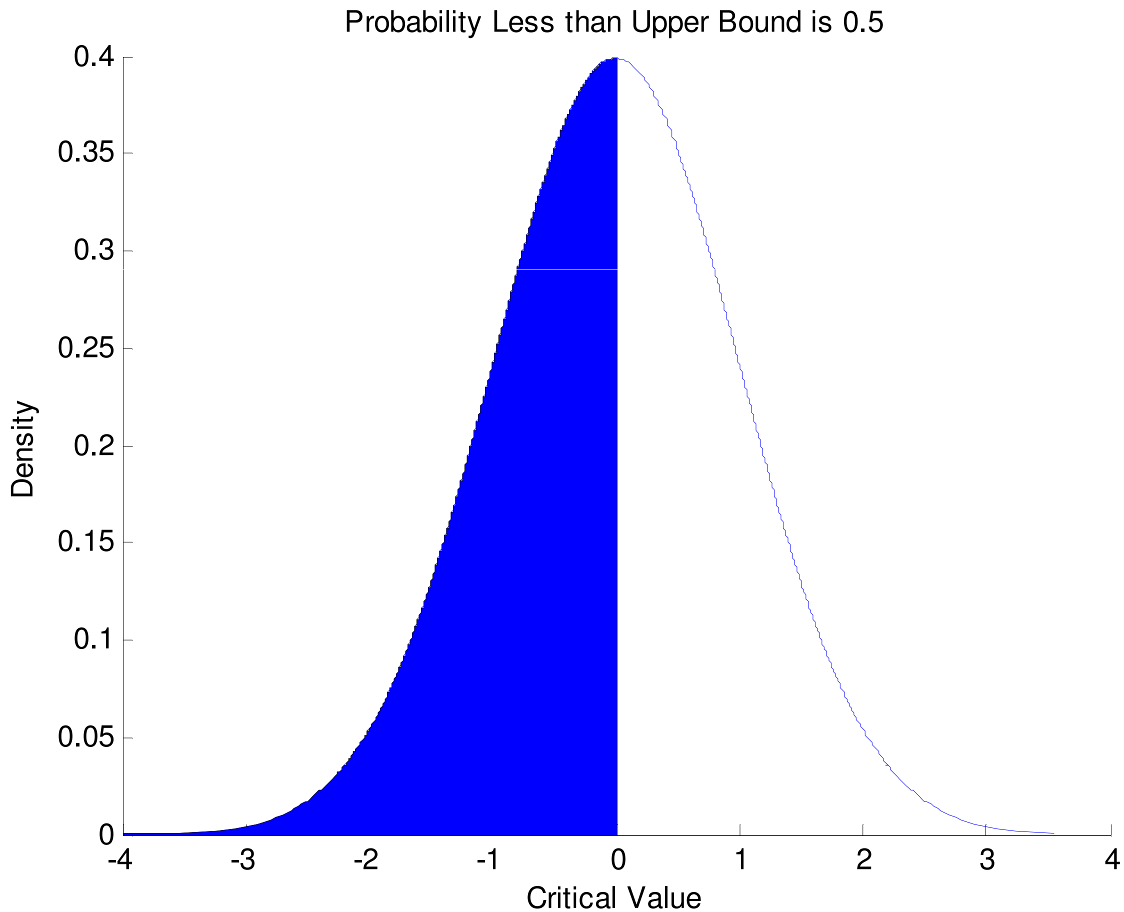


- Suponha que o comprimento de recém-nascidos seja uma v.a. com distribuição N[48,55;2,5]
  - A Probabilidade de um recém-nascido ter comprimento maior do que a média deve ser de 50%
    - Padronizando os dados, isto irá ocorrer em z = 0

$$z = \frac{x - \mu}{\sigma} = \frac{48{,}58 - 48{,}58}{2{,}5} = 0$$

Exemplo

- Pela tabela (ou programa estatístico)



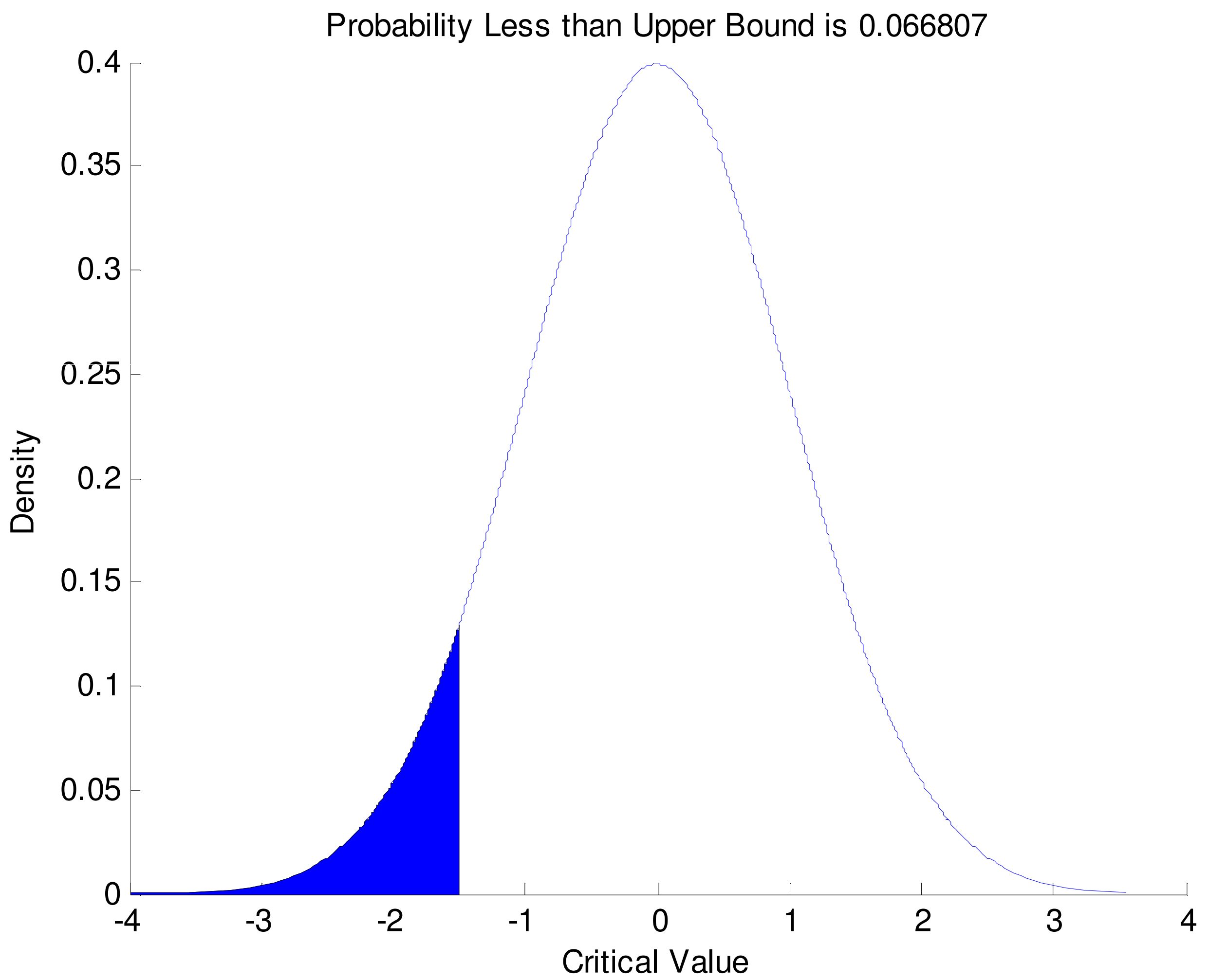
Exemplo

- Probabilidade do comprimento ser inferior a 44,79 cm

$$z = \frac{x-\mu}{\sigma} = \frac{44{,}79-48{,}58}{2{,}5} = -1{,}5$$

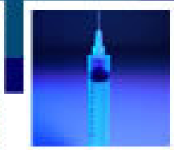
- Logo a probabilidade será a área da curva até z=-1,5
  - P(X≤44,79) = 0,0668 ou 6,68%

Exemplo


Probability Less than Upper Bound is 0.066807
Density
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-4
-3
-2
-1
0
1
2
3
4
Critical Value

Exemplo



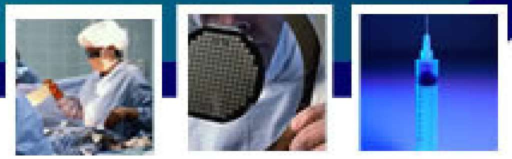


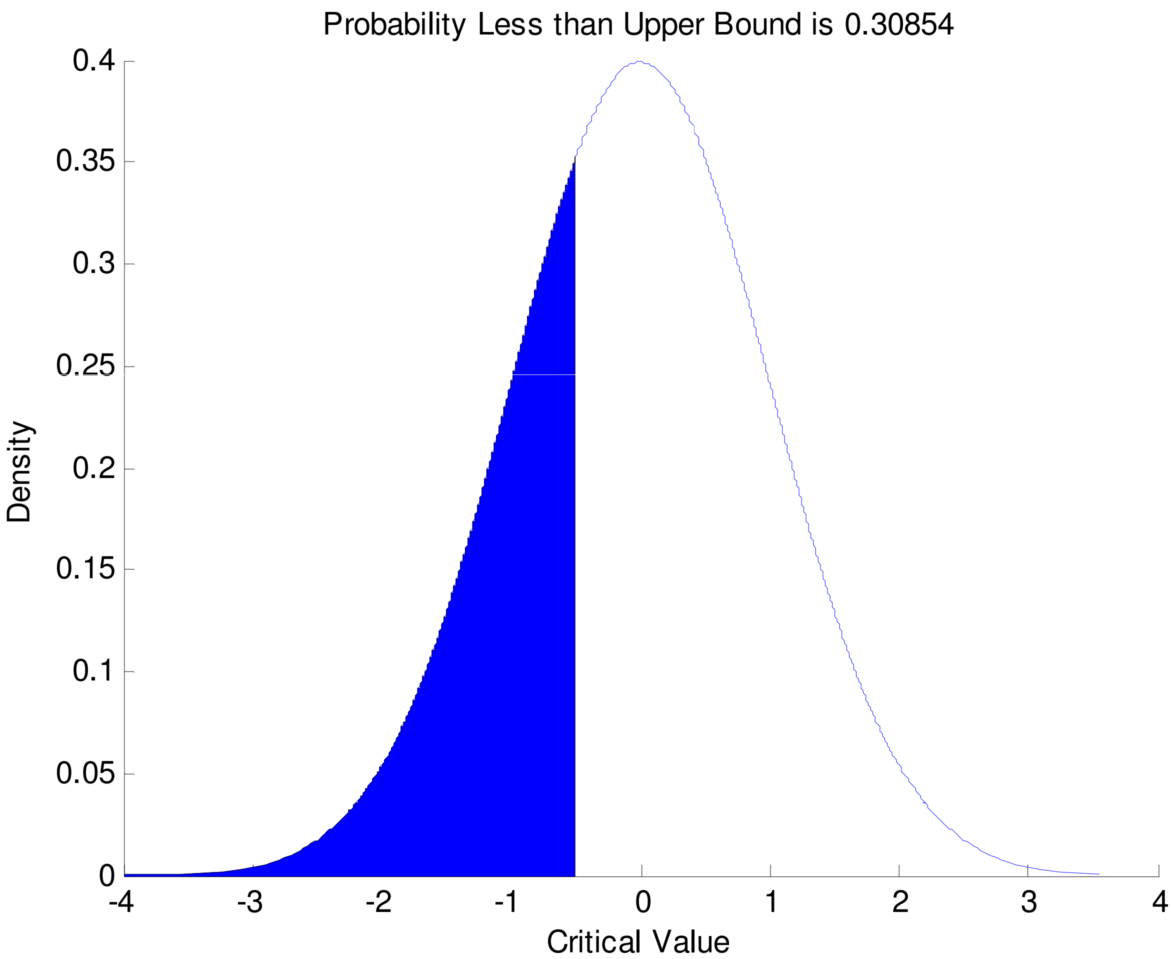
- Probabilidade do comprimento ser inferior a 44,29 cm

$$z = \frac{x - \mu}{\sigma} = \frac{44{,}29 - 48{,}58}{2{,}5} = -0{,}5$$

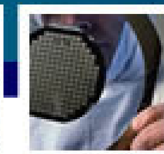
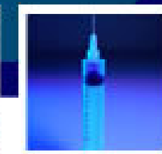
- Logo a probabilidade será a área da curva até z=-0,5
  – $P(X \leq 44{,}29) = 0{,}3085$ ou 30,85%

Exemplo




Probability Less than Upper Bound is 0.30854
Density
Critical Value
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-4
-3
-2
-1
0
1
2
3
4

Exemplo

- Qual a probabilidade de um recém-nascido ter comprimento maior que 44,29 cm?
  - Como a área total é 1
    - Então, $P(X>44{,}29) = 1 - P(X \leq 44{,}29) = 1 - 0{,}3085 = 0{,}6915$ ou 69,15%

Exemplo

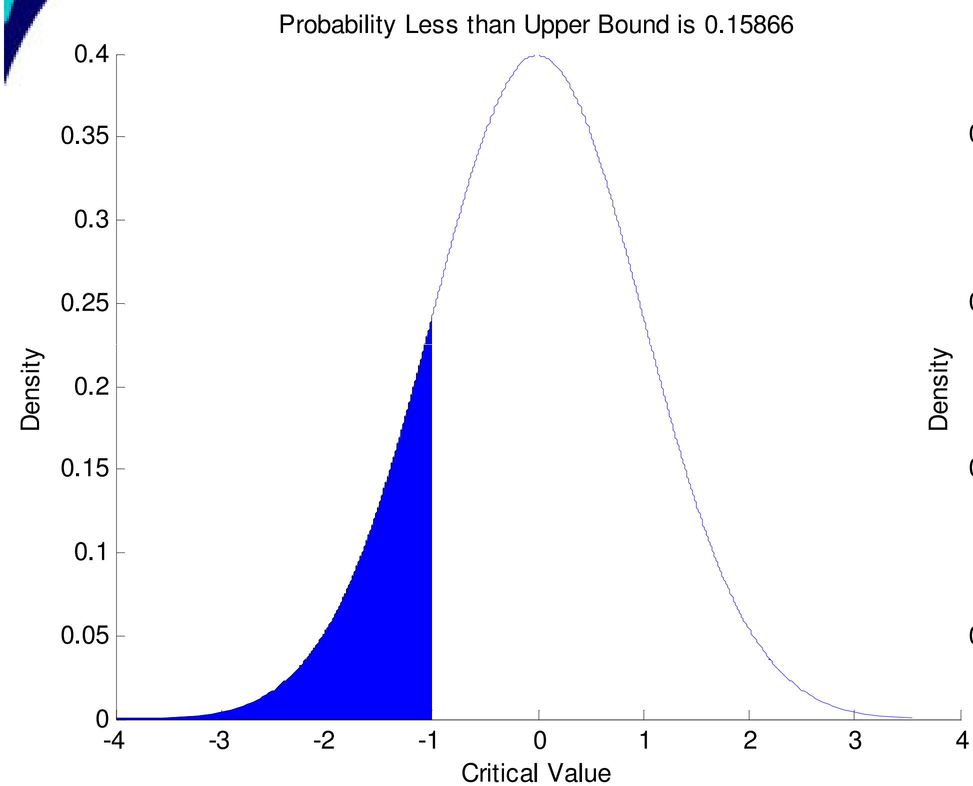
- Qual a probabilidade de um recém-nascido ter comprimento entre 46,01cm e 51,04 cm?

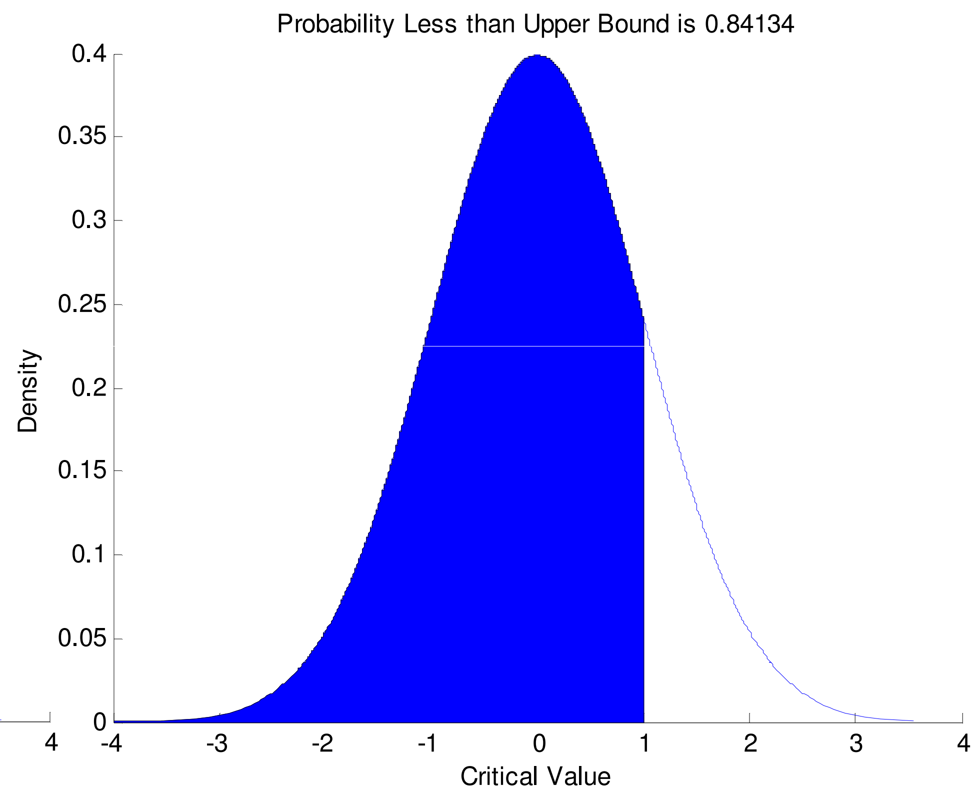
$$z_1 = \frac{x-\mu}{\sigma} = \frac{46{,}01 - 48{,}58}{2{,}5} = -1{,}0$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{51{,}04 - 48{,}58}{2{,}5} = 1{,}0$$

  – A probabilidade será a área entre z_1 e z_2.

Exemplo

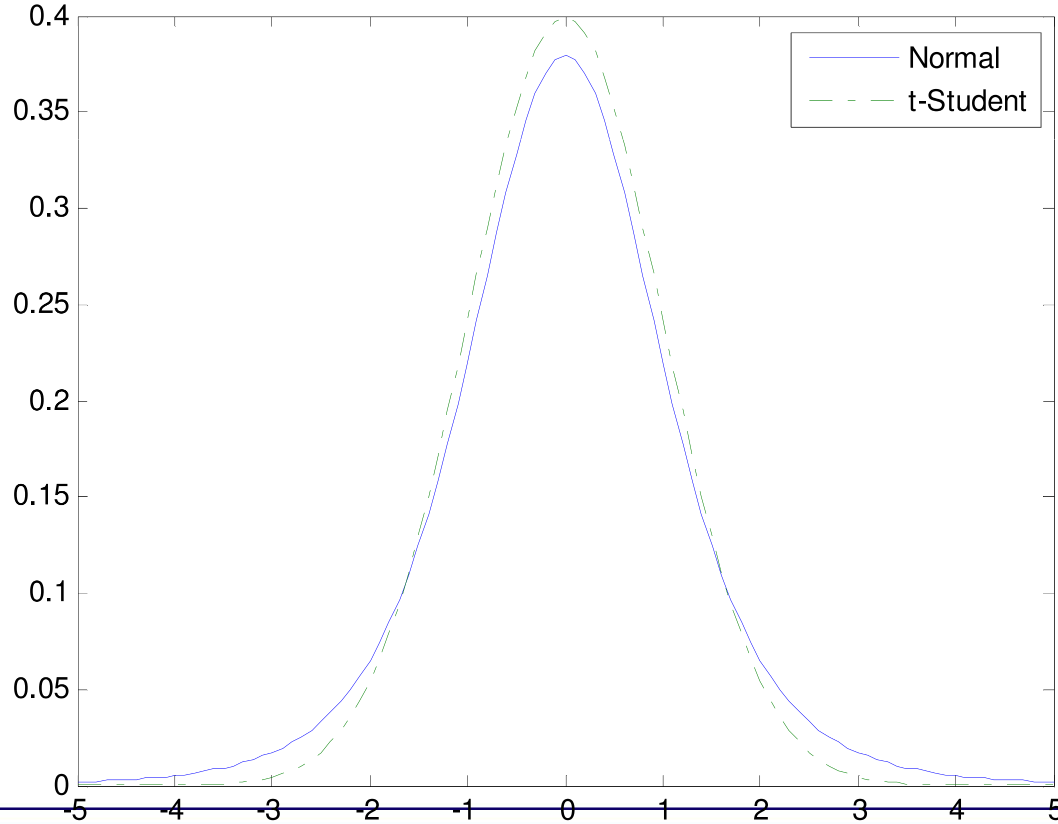




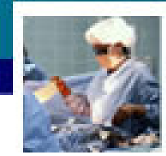
$P(46{,}04 \leq X \leq 51{,}04) = 0{,}8413 - 0{,}1587 = 0{,}6826$ ou 68,26%

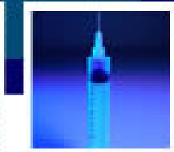
Distribuição t-Student

- A distribuição t-Stuent é bastante semelhante com a distribuição Normal



Distribuição t-Student



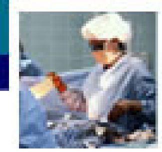


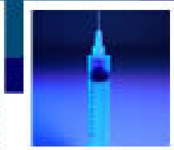
- A distribuição t-Student é bastante utilizada na prática quando temos uma amostragem muito pequena (n<30).
  - Dado que Z é uma N(0,1), então

$$t = \frac{Z}{\sqrt{Y/\nu}}$$

    - Esta é uma distribuição t-Student com ν graus de liberdade.

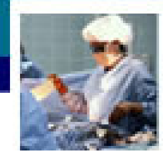
Distribuição de Fisher ou Distribuição F

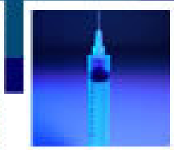




- A distribuição F é geralmente relacionada ao estudo da variância de dados.
  - Está é formada pelo quociente de duas distribuições, cada uma com seus respectivos graus de liberdade, ν_1 e ν_2.
  - Esta distribuição será interessante quando estivermos falando de testes estatísticos, para verificação se dois conjuntos de dados são estatisticamente iguais ou diferentes, por exemplo!

Distribuição Qui-quadrado





- A distribuição Qui-quadrado, ou χ^2, corresponde a distribuição da soma dos quadrados de n v.a. Independentes padronizadas.

$$\chi^2 = \chi_1^2 + \chi_2^2 + \cdots + \chi_n^2$$

$$\chi^2 = f\left(x, \nu\right)$$

  - Onde ν são os graus de liberdade.

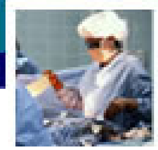
Distribuição t-Student

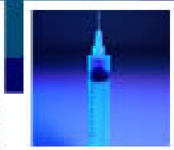
- De forma prática, se x é uma maostra aleatória de uma N(μ,σ), então a estatística

$$t = \frac{\bar{x}_{amostral} - \mu}{s/\sqrt{n}}$$

Terá uma distribuição t-student com n-1 graus de liberdade.

Referências





- Livro Texto: Bioestatística. Teoria e Computacional (Héctor G. Arango). Guanaba
  - Capítulo 6
- Leituras Complementares:
  - Livro: Bioestatística. Princípios e Aplicações (Sidia M. Callegari-Jaqcques)
    - Capítulo 4, 5 e 7.
  - Livro: Estatística Básica (Wilton de O. Bussab e Pedro A. Morettin). Saraiva
    - Capítulo: 6 e 7
  - Livro: Probabilidade. Aplicações à Estatística (Paul L. Meyer). LTC
    - Capítulos 8 e 9