CE003
Estatística II

Silvia Shimakura

silvia.shimakura@ufpr.br

Laboratório de Estatística e Geoinformação

Variáveis aleatórias e distribuições de probabilidade

- **Variável aleatória** é um função X que associa a cada evento do espaço amostral um valor numérico $X(\omega) \in R$
- **Distribuição de probabilidade** associa a cada valor de uma v.a. uma probabilidade

Experimento: Lançamento de duas moedas

X=número de caras (C)

$\Omega=\{RR,CR,RC,CC\}$

$X=\{0,1,2\}$

$P(X=0)=1/4$ $P(X=1)=1/2$ $P(X=2)=1/4$

Variáveis aleatórias

Uma variável aleatória pode ser classificada como **discreta** ou **contínua**, dependendo do domínio dos valores de X.

Exemplo:

o número de alunos em uma sala é uma variável aleatória discreta, denotada por X (maiúsculo).

Uma observação dessa variável é denotada pela respectiva letra minúscula, e.g. x=50 alunos

Em geral, denotamos a probabilidade de uma v.a. X assumir determinado valor x como P(x) ou P(X=x)

Distribuições de probabilidade

- Denomina-se de distribuição de probabilidade de alguma variável aleatória, a regra geral que define a

função de probabilidade (fp) (V.A.s discretas)

função densidade de probabilidade (fdp) (V.A.s contínuas)

- Existem muitas distribuições de probabilidade, mas algumas merecem destaque por sua importância prática

Variável aleatória discreta

Uma V.A. é classificada como discreta se assume somente um conjunto enumerável (finito ou infinito) de valores.

Exemplos:

- Número de caras ao lançar 3 moedas
- Número de chamadas telefônicas que chegam à uma central em 1 hora
- Número de votos recebidos
- Aprovação no vestibular
- Grau de queimadura na pele

Variáveis aleatórias discretas

A função que atribui a cada valor da variável aleatória sua probabilidade é denominada função discreta de probabilidade ou **função de probabilidade**, isto é:

$$\mathbf{P(X = x_i) = p(x_i) = p_i,\ i=1, 2, \ldots}$$

com as seguintes propriedades:

i) A probabilidade de cada valor deve estar entre 0 e 1

$$0 \leq p(x_i) \leq 1,\ \forall\ i = 1, 2, \ldots$$

ii) A soma de todas as probabilidades é igual a 1

$$\sum p(x_i) = 1$$

Experimento: Lançamento de duas moedas

Podemos montar uma tabela de distribuição de frequência para a variável aleatória

X = número de caras (C)

Assim podemos associar a cada valor de X sua probabilidade correspondente, como resultado das frequências relativas

P(X = 0) = 1/4

P(X = 1) = 2/4 = 1/2

P(X = 2) = 1/4

X	Freq (f)	Freq Relativa (fr)
0	1	1/4
1	2	2/4
2	1	1/4
Total	4	1

Variáveis aleatórias discretas

Dessa forma, a distribuição de probabilidade da variável aleatória

X = número de caras (C) é:

X	P(X=x)
0	1/4
1	1/2
2	1/4
Total	1

Repare que as propriedades da função de probabilidade estão satisfeitas:

i) As probabilidades estão entre 0 e 1

ii) A soma de todas as probabilidades é 1

Variáveis aleatórias discretas

Dessa forma, a distribuição de probabilidade da variável aleatória

X = número de caras (C) é:

X	P(X=x)
0	1/4
1	1/2
2	1/4
Total	1

Repare que as propriedades da função de probabilidade estão satisfeitas:

i) As probabilidades estão entre 0 e 1

ii) A soma de todas as probabilidades é 1

Qual seria a média desta variável aleatória X?

Esperança

O valor esperado, ou média, ou esperança matemática é uma quantidade utilizada como resumo do comportamento de uma V.A.

A esperança de uma V.A. X é obtida multiplicando-se cada valor de $X = x_i$, por sua probabilidade $P[X = x_i]$, e somando os produtos resultantes

$$\mathbf{E(X) = \sum x_i \cdot P[X = x_i]}$$

A esperança é o valor médio que esperaríamos obter caso o experimento fosse repetido várias vezes.

Esperança

Exemplo: Um pediatra de um hospital público constatou que das crianças internadas durante um ano:

20% não internaram por infecção da faringe,

45% tiveram um internamento por infecção da faringe,

25% tiveram dois internamentos por infecção da faringe,

9% tiveram três internamentos por infecção da faringe e

1% tiveram quatro internamentos por infecção da faringe.

Seja X=número de internamentos por infecção da faringe

A função de probabilidade para X é:

No exemplo temos: 0,20 + 0,45 + 0,25 + 0,09 + 0,01 = 1.

Qual é o número esperado de internamentos por infecção da faringe por ano?

X	$P(X=x_i)$
0	0,20
1	0,45
2	0,25
3	0,09
4	0,01

Esperança

Qual é o número esperado de internamentos por infecção da faringe por ano?

$$E(X)=\Sigma x_i \cdot P(X=x_i)$$

$=0\cdot0,20+1\cdot0,45+2\cdot0,25+3\cdot0,09+4\cdot0,01$

$=1,26$

X	$P(X=x_i)$	$X_i\ P(X=x_i)$
0	0,20	0
1	0,45	0,45
2	0,25	0,5
3	0,09	0,27
4	0,01	0,04
Total	1	1,26

Variância

A variância, como já vimos, mede o grau de dispersão dos valores de uma variável aleatória em torno da sua média ou esperança E(X).

A forma geral para o cálculo é

$$Var(X)=\sum(x_i - E(X))^2 \cdot P[X=x_i]$$

Ou alternativamente

$$Var(X)=E(X^2) - E(X)^2$$

em que

$$E(X^2)=\sum x_i^2 \cdot P[X=x_i]$$

Variância

X	$P(X=x_i)$	$x_i\,P(X=x_i)$	x_i^2	$x_i^2\,P(X=x_i)$
0	0,20	0	0	0
1	0,45	0,45	1	0,45
2	0,25	0,5	4	2,0
3	0,09	0,27	9	2,43
4	0,01	0,04	16	0,64
Total	1	1,26		5,52

$E(X)=1{,}26$

$E(X^2)=5{,}52$

$Var(X)=E(X^2)-E(X)^2$

$=5{,}52-(1{,}26)^2$

$=3{,}93$

$DP(X)=\sqrt{3{,}93}=1{,}98$

Exercício

As probabilidades de um corretor de imóveis vender uma sala comercial com lucro de R$ 3500,00, de R$ 2500,00, R$ 800,00 ou com um prejuízo de R$ 500,00 são 0,33, 0,35, 0,22 e 0,10, respectivamente.

Qual é o lucro esperado do corretor de imóveis?

Exercício

As probabilidades de um corretor de imóveis vender salas comerciais com lucro de R$ 3500,00, de R$ 2500,00, R$ 800,00 ou com prejuízo de R$ 500,00 são 0,33, 0,35, 0,22 e 0,10, respectivamente.

Qual é o lucro esperado do corretor de imóveis?

Lucro (X)	3500	2500	800	-500
P(X=x)	0,33	0,35	0,22	0,10

$$E(X)=3500\times 0{,}33+2500\times 0{,}35+800\times 0{,}22-500\times 0{,}10=\mathbf{2156{,}00}$$

Exercício

As probabilidades de um corretor de imóveis vender salas comerciais com lucro de R$ 3500,00, de R$ 2500,00, R$ 800,00 ou com prejuízo de R$ 500,00 são 0,33, 0,35, 0,22 e 0,10, respectivamente.

Qual é o lucro esperado do corretor de imóveis?

Lucro (X)	3500	2500	800	-500
P(x)	0,33	0,35	0,22	0,10

$E(X)=3500\times0{,}33+2500\times0{,}35+800\times0{,}22-500\times0{,}10=\mathbf{2156{,}00}$

$E(X^2)=3500^2\times0{,}33+2500^2\times0{,}35+800^2\times0{,}22+(-500)^2\times0{,}10=6395800$

$Var(X)=E(X^2)-E(X)^2=1747464$ $DP(X)=\mathbf{1321{,}90}$

Distribuições de Probabilidade

Exemplo: Eficácia de medicamento

- Uma industria farmacêutica afirma que um certo medicamento alivia os sintomas de angina pectoris em 80% dos pacientes.
- Você prescreve este medicamento a 5 dos seus pacientes com angina mas somente 2 (40%) relatam alívio dos sintomas.
- Se a afirmação do fabricante for verdadeira, é possível obter resultados tão ruins ou ainda piores do que os que você observou?

Exemplo: Eficácia de medicamento (cont.)

- Assume-se que:
  - **A**: alívio dos sintomas
  - Afirmação fabricante verdadeira: **$P(A)=0,8$**
  - **X**: nº pacientes que relatam alívio dos sintomas dentre 5 pacientes
- Queremos saber:

$$P(X\leq 2)=P(X=2)+P(X=1)+P(X=0)$$

Exemplo: Eficácia de medicamento (cont.)

X	Sequência	P(Sequência)
2	AANNN	P(AANNN)=P(A)P(A)P(N)P(N)P(N)

Supondo independência entre os pacientes

Exemplo: Eficácia de medicamento (cont.)

X	Sequência	P(Sequência)
2	AANNN	$0{,}8 \times 0{,}8 \times 0{,}2 \times 0{,}2 \times 0{,}2 = 0{,}8^2 \times 0{,}2^3$
2	ANANN	$0{,}8 \times 0{,}2 \times 0{,}8 \times 0{,}2 \times 0{,}2 = 0{,}8^2 \times 0{,}2^3$
2	ANNAN	$0{,}8 \times 0{,}2 \times 0{,}2 \times 0{,}8 \times 0{,}2 = 0{,}8^2 \times 0{,}2^3$
2	ANNNA	$0{,}8 \times 0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}8 = 0{,}8^2 \times 0{,}2^3$
2	NAANN	$0{,}2 \times 0{,}8 \times 0{,}8 \times 0{,}2 \times 0{,}2 = 0{,}8^2 \times 0{,}2^3$
2	NANAN	$0{,}2 \times 0{,}8 \times 0{,}2 \times 0{,}8 \times 0{,}2 = 0{,}8^2 \times 0{,}2^3$
2	NANNA	$0{,}2 \times 0{,}8 \times 0{,}2 \times 0{,}2 \times 0{,}8 = 0{,}8^2 \times 0{,}2^3$
2	NNAAN	$0{,}2 \times 0{,}2 \times 0{,}8 \times 0{,}8 \times 0{,}2 = 0{,}8^2 \times 0{,}2^3$
2	NNANA	$0{,}2 \times 0{,}2 \times 0{,}8 \times 0{,}2 \times 0{,}8 = 0{,}8^2 \times 0{,}2^3$
2	NNNAA	$0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}8 \times 0{,}8 = 0{,}8^2 \times 0{,}2^3$

$$\binom{5}{2} = 10$$

Sequências possíveis

Exemplo: Eficácia de medicamento (cont.)

X	Sequência	P(Sequência)
2	AANNN	$0,8 \times 0,8 \times 0,2 \times 0,2 \times 0,2 = 0,8^2 \times 0,2^3$
2	ANANN	$0,8 \times 0,2 \times 0,8 \times 0,2 \times 0,2 = 0,8^2 \times 0,2^3$
2	ANNAN	$0,8 \times 0,2 \times 0,2 \times 0,8 \times 0,2 = 0,8^2 \times 0,2^3$
2	ANNNA	$0,8 \times 0,2 \times 0,2 \times 0,2 \times 0,8 = 0,8^2 \times 0,2^3$
2	NAANN	$0,2 \times 0,8 \times 0,8 \times 0,2 \times 0,2 = 0,8^2 \times 0,2^3$
2	NANAN	$0,2 \times 0,8 \times 0,2 \times 0,8 \times 0,2 = 0,8^2 \times 0,2^3$
2	NANNA	$0,2 \times 0,8 \times 0,2 \times 0,2 \times 0,8 = 0,8^2 \times 0,2^3$
2	NNAAN	$0,2 \times 0,2 \times 0,8 \times 0,8 \times 0,2 = 0,8^2 \times 0,2^3$
2	NNANA	$0,2 \times 0,2 \times 0,8 \times 0,2 \times 0,8 = 0,8^2 \times 0,2^3$
2	NNNAA	$0,2 \times 0,2 \times 0,2 \times 0,8 \times 0,8 = 0,8^2 \times 0,2^3$
P(X=2)		**$10 \times 0,8^2 \times 0,2^3 = 0,0514$**

$$\binom{5}{2} = 10$$

Sequências possíveis

Exemplo: Eficácia de medicamento (cont.)

X	Sequência	P(Sequência)
1	**ANNNN**	0,8 x 0,2 x 0,2 x 0,2 x 0,2 = $0{,}8^1$ x $0{,}2^4$
1	**NANNN**	0,2 x 0,8 x 0,2 x 0,2 x 0,2 = $0{,}8^1$ x $0{,}2^4$
1	**NNANN**	0,2 x 0,2 x 0,8 x 0,2 x 0,2 = $0{,}8^1$ x $0{,}2^4$
1	**NNNAN**	0,2 x 0,2 x 0,2 x 0,8 x 0,2 = $0{,}8^1$ x $0{,}2^4$
1	**NNNNA**	0,2 x 0,2 x 0,2 x 0,2 x 0,8 = $0{,}8^1$ x $0{,}2^4$
P(X=1)		**5 x $0{,}8^1$ x $0{,}2^4$ = 0,0064**

$$\binom{5}{1} = 5$$

Sequências possíveis

Exemplo: Eficácia de medicamento (cont.)

X	Sequência	P(Sequência)
1	ANNNN	$0{,}8 \times 0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}2 = 0{,}8^1 \times 0{,}2^4$
1	NANNN	$0{,}2 \times 0{,}8 \times 0{,}2 \times 0{,}2 \times 0{,}2 = 0{,}8^1 \times 0{,}2^4$
1	NNANN	$0{,}2 \times 0{,}2 \times 0{,}8 \times 0{,}2 \times 0{,}2 = 0{,}8^1 \times 0{,}2^4$
1	NNNAN	$0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}8 \times 0{,}2 = 0{,}8^1 \times 0{,}2^4$
1	NNNNA	$0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}8 = 0{,}8^1 \times 0{,}2^4$
P(X=1)		**$5 \times 0{,}8^1 \times 0{,}2^4 = 0{,}0064$**
0	NNNNN	$0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}2 \times 0{,}2 = 0{,}8^0 \times 0{,}2^5$
P(X=0)		**$1 \times 0{,}8^0 \times 0{,}2^5 = 0{,}00032$**

$$\binom{5}{1} = 5$$

Sequências possíveis

$$\binom{5}{0} = 1$$

Exemplo: Eficácia de medicamento (cont.)

X	Sequência	P(Sequência)
1	ANNNN	$0,8 \times 0,2 \times 0,2 \times 0,2 \times 0,2 = 0,8^1 \times 0,2^4$
1	NANNN	$0,2 \times 0,8 \times 0,2 \times 0,2 \times 0,2 = 0,8^1 \times 0,2^4$
1	NNANN	$0,2 \times 0,2 \times 0,8 \times 0,2 \times 0,2 = 0,8^1 \times 0,2^4$
1	NNNAN	$0,2 \times 0,2 \times 0,2 \times 0,8 \times 0,2 = 0,8^1 \times 0,2^4$
1	NNNNA	$0,2 \times 0,2 \times 0,2 \times 0,2 \times 0,8 = 0,8^1 \times 0,2^4$
P(X=1)		**$5 \times 0,8^1 \times 0,2^4 = 0,0064$**
0	NNNNN	$0,2 \times 0,2 \times 0,2 \times 0,2 \times 0,2 = 0,8^0 \times 0,2^5$
P(X=0)		**$1 \times 0,8^0 \times 0,2^5 = 0,00032$**

$$\binom{5}{1} = 5$$

Sequências possíveis

$$\binom{5}{0} = 1$$

$P(X \leq 2) = 0,0514 + 0,0064 + 0,00032 = 0,05812$

Exemplo: Eficácia de medicamento (cont.)

Se a afirmação do fabricante for verdadeira, a chance de se obter resultados tão ruins ou ainda piores do que os observados é de 5,8%.

CONCLUSÃO?

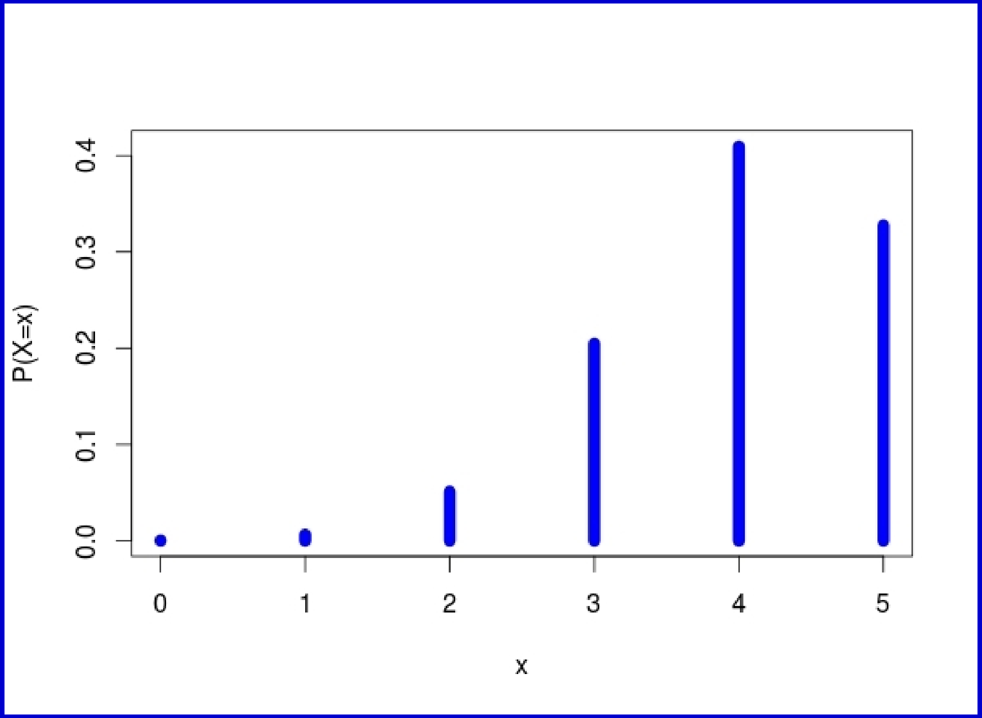
Distribuição Binomial

- n: no. ensaios (independentes)
- X: no. sucessos nos n ensaios
- p: prob. sucesso num ensaio

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

$$P(X=0)+P(X=1)+\ldots+P(X=n)=1$$

Distribuição binomial(5,0.8)



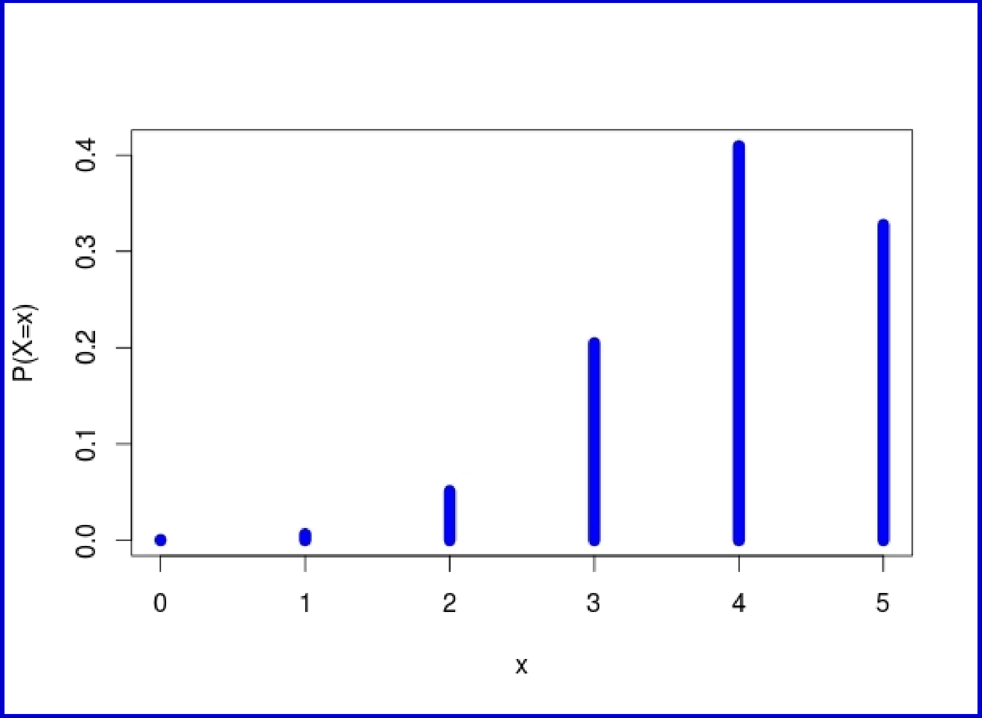
Se n=5 pacientes usarem o medicamento e a prob. alívio dos sintomas for p=0,8

⬇

A cada 5 pacientes espera-se em **média** nxp=4 pacientes com alívio dos sintomas

$$P(X=0)+P(X=1)+\ldots+P(X=5)=1$$

Distribuição binomial(5,0.8)



- Se n=5 pacientes usarem o medicamento e a prob. alívio dos sintomas for p=0,8

⬇

- A cada 5 pacientes espera-se em **média** nxp=4 pacientes com alívio dos sintomas
- A **variância** será

nxpx(1-p)=0,8

$$P(X=0)+P(X=1)+...+P(X=5)=1$$

Calculadora

http://onlinestatbook.com/2/java/binomialProb.html

Exercício

Num munícipio, há uma probabilidade de 0,70 de uma empresa de materiais recicláveis ter seguro contra incêndio; qual a probabilidade de que, dentre cinco empresas:

a) Nenhuma tenha seguro contra incêndio?

b) Exatamente quatro tenham seguro contra incêndio?

Exercício

Num munícipio, há uma probabilidade de 0,70 de uma empresa de materiais recicláveis ter seguro contra incêndio; qual a probabilidade de que, dentre cinco empresas:

X=número de empresas que tem seguro contra incendio

p=Probabilidade de uma empresa ter seguro=0,70

n=5 empresas

a) Nenhuma tenha seguro contra incêndio? P(X=0)

b) Exatamente quatro tenham seguro contra incêndio? P(X=4)

Exercício

Sabe-se que 5% das válvulas fabricadas em uma indústria são defeituosas.

Em lotes de 20 válvulas, são esperadas quantas válvulas defeituosas?

Com qual desvio-padrão?

Variável aleatória contínua

Uma V.A. é classificada como contínua se assume valores em qualquer intervalo dos números reais.

Exemplos:

- Peso de animais
- Tempo de vida de uma pessoa
- Temperatura em uma hora específica do dia
- Salinidade da água do mar
- Retorno financeiro de um investimento

Variável aleatória contínua

- Atribuimos probabilidades à intervalos de valores, por meio de uma função
- As probabilidades são representadas por áreas
- Não podemos atribuir probabilidades à valores específicos: $P(X=x)=0$

Variável aleatória contínua

A função densidade de probabilidade (fdp) atribui probabilidades à intervalos de valores do tipo [a, b], e é definida por

$$P\left(a \leqslant X \leqslant b\right) = \int_a^b f\left(x\right) dx$$

com as seguintes propriedades

i) É uma função não negativa: $f(x) \geq 0$

ii) A área total sob a curva deve ser igual a 1

Variável aleatória contínua

- $P[X = x] = 0$, então

$P[a \leq X \leq b] = P[a < X \leq b] = P[a \leq X < b] = P[a < X < b]$

- Qualquer função que seja não negativa e cuja área total sob a curva seja igual à 1 caracterizará uma VA contínua
- f(x) não representa a probabilidade de ocorrência de um evento
- A área sob a curva entre dois pontos é que fornecerá a probabilidade

Esperança

A **esperança** de uma V.A. contínua tem o mesmo sentido e interpretação da esperança de uma V.A. discreta: **é a média ou valor esperado da V.A.**

A esperança é obtida através da integral do produto de x com a função f(x):

$$E(X) = \int x \cdot f(x)dx$$

Variância

A variância, mede o grau de dispersão dos valores de uma variável aleatória em torno da média ou esperança E(X).

A forma geral para o cálculo em V.A.s contínuas é

$$Var(X) = E[(X - E(X))^2] = \int [x - E(X)]^2 \cdot f(x)dx = E(X^2) - E(X)^2$$

em que $E(X^2) = \int x^2 \cdot f(x)dx$

Distribuições contínuas de probabilidade

- Existem diversos modelos contínuos de probabilidade.
- Um dos mais importantes, tanto do ponto de vista teórico quanto prático, é a **distribuição normal**.

Carl Friedr. Gauss
(1777-1855)

Distribuição Normal

- Diversas variáveis contínuas tais como, altura, peso, níveis de colesterol, pressão arterial, medidas de testes psicológicos, podem ser descritas pela distribuição normal

- Formato da curva definido por 2 parâmetros:

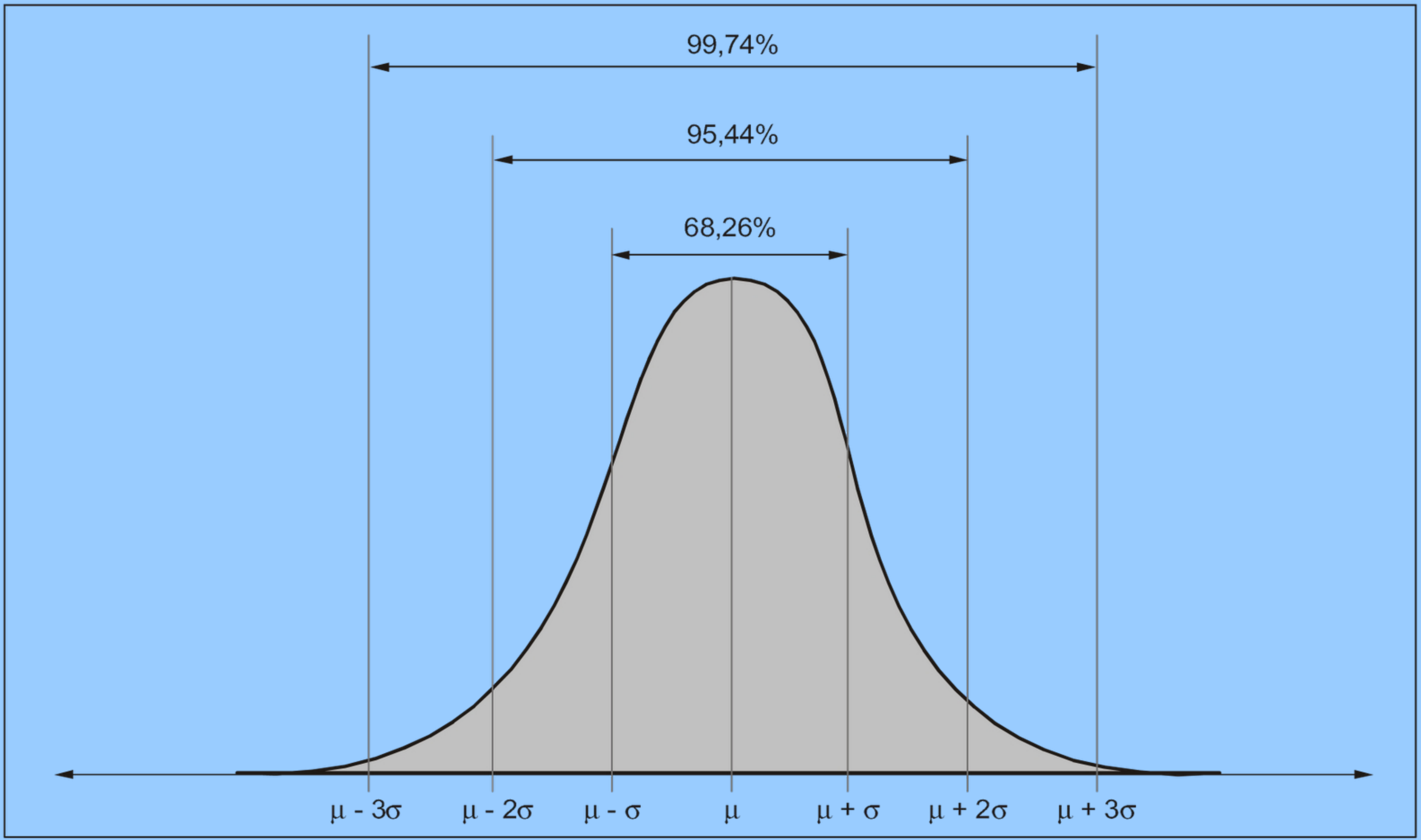
μ σ

Média (Centro)

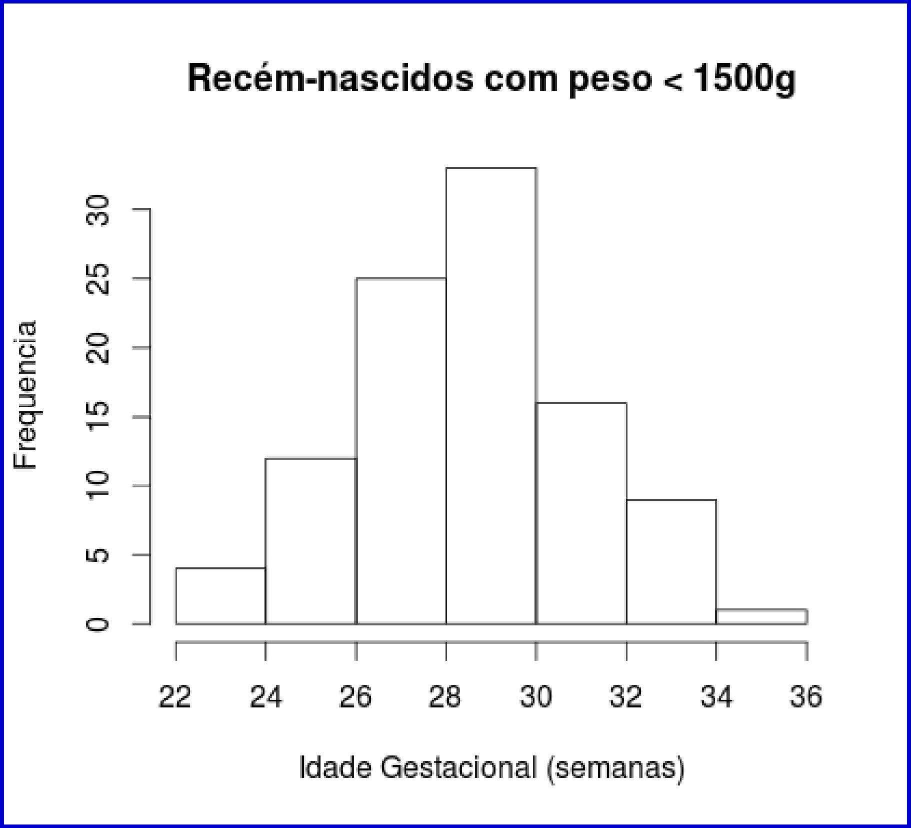
Desvio-padrão (Espalhamento)

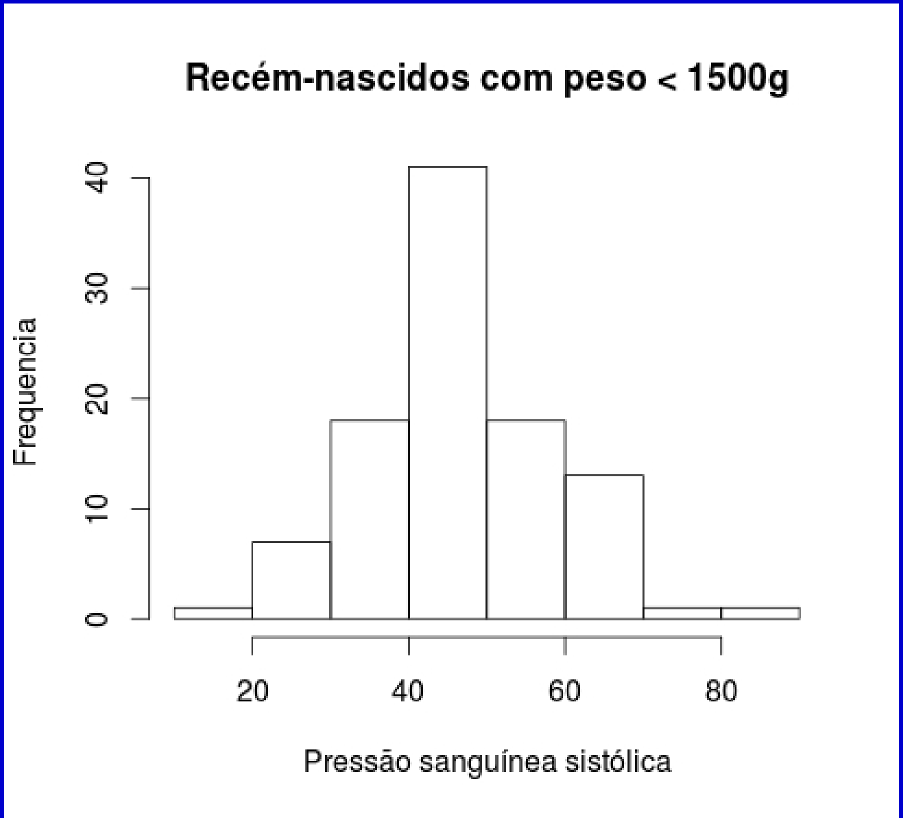
Equação:

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

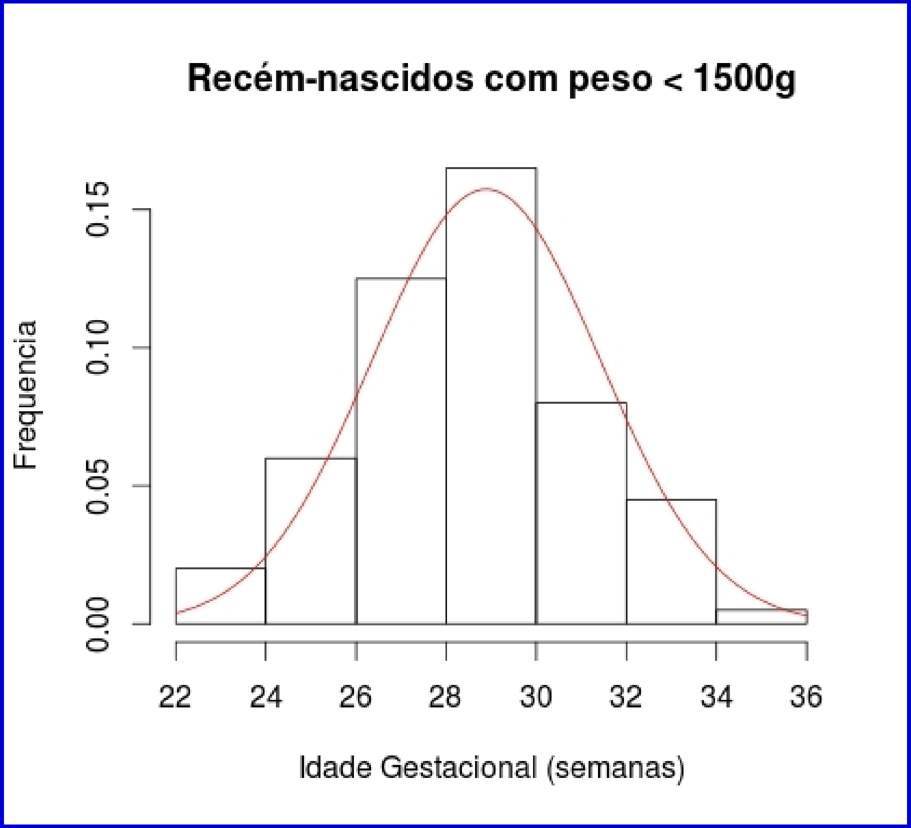


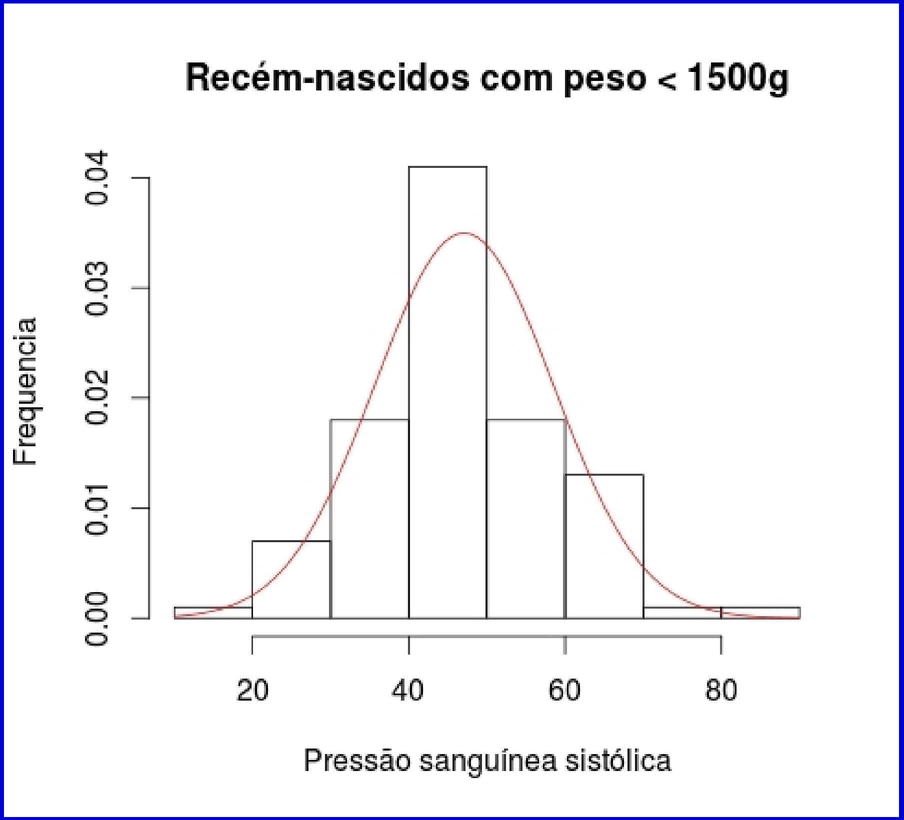
Amostra de 100 recém-nascidos com peso<1500g em Boston, Massachusetts



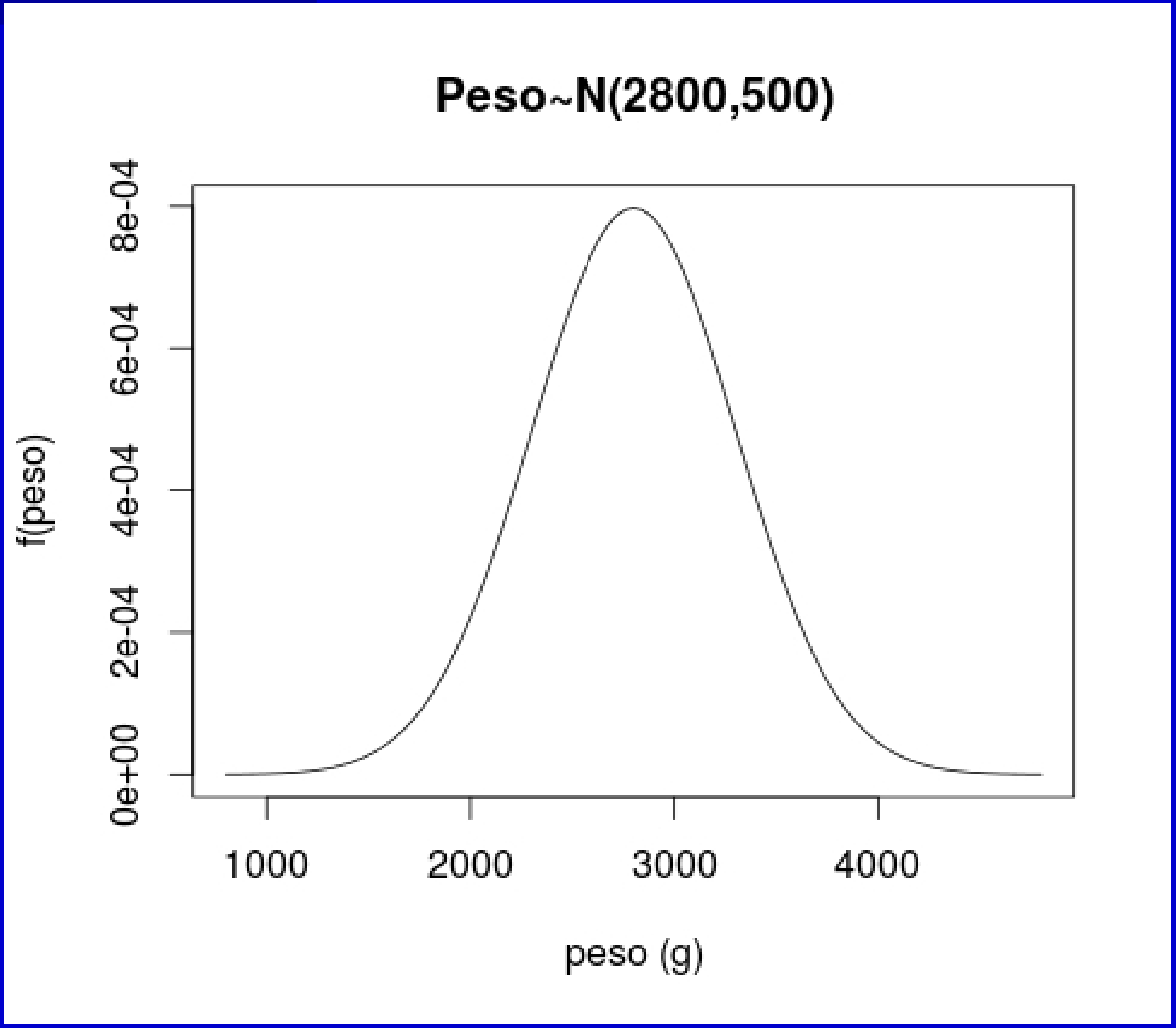


Amostra de 100 recém-nascidos com peso<1500g em Boston, Massachusetts

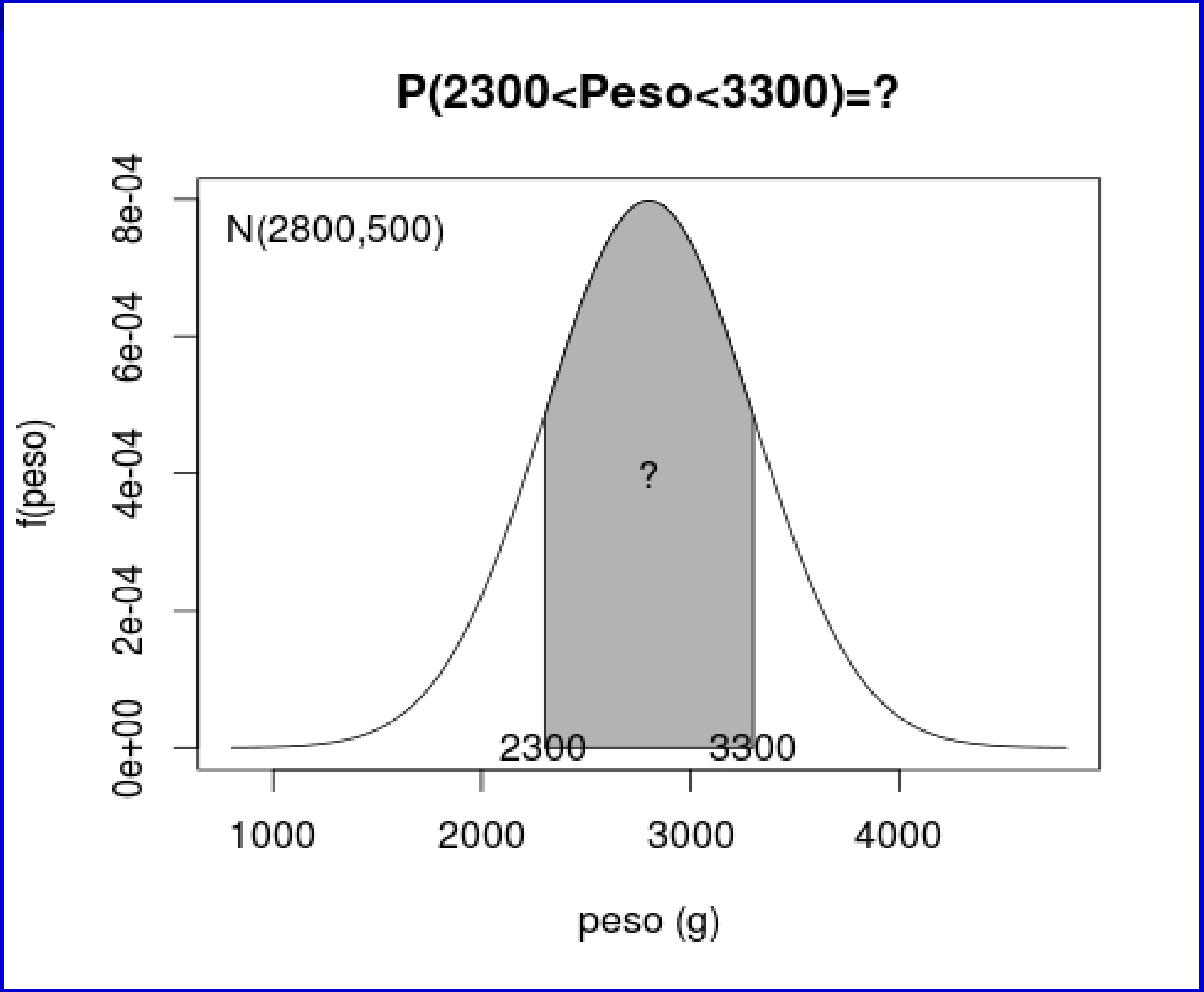




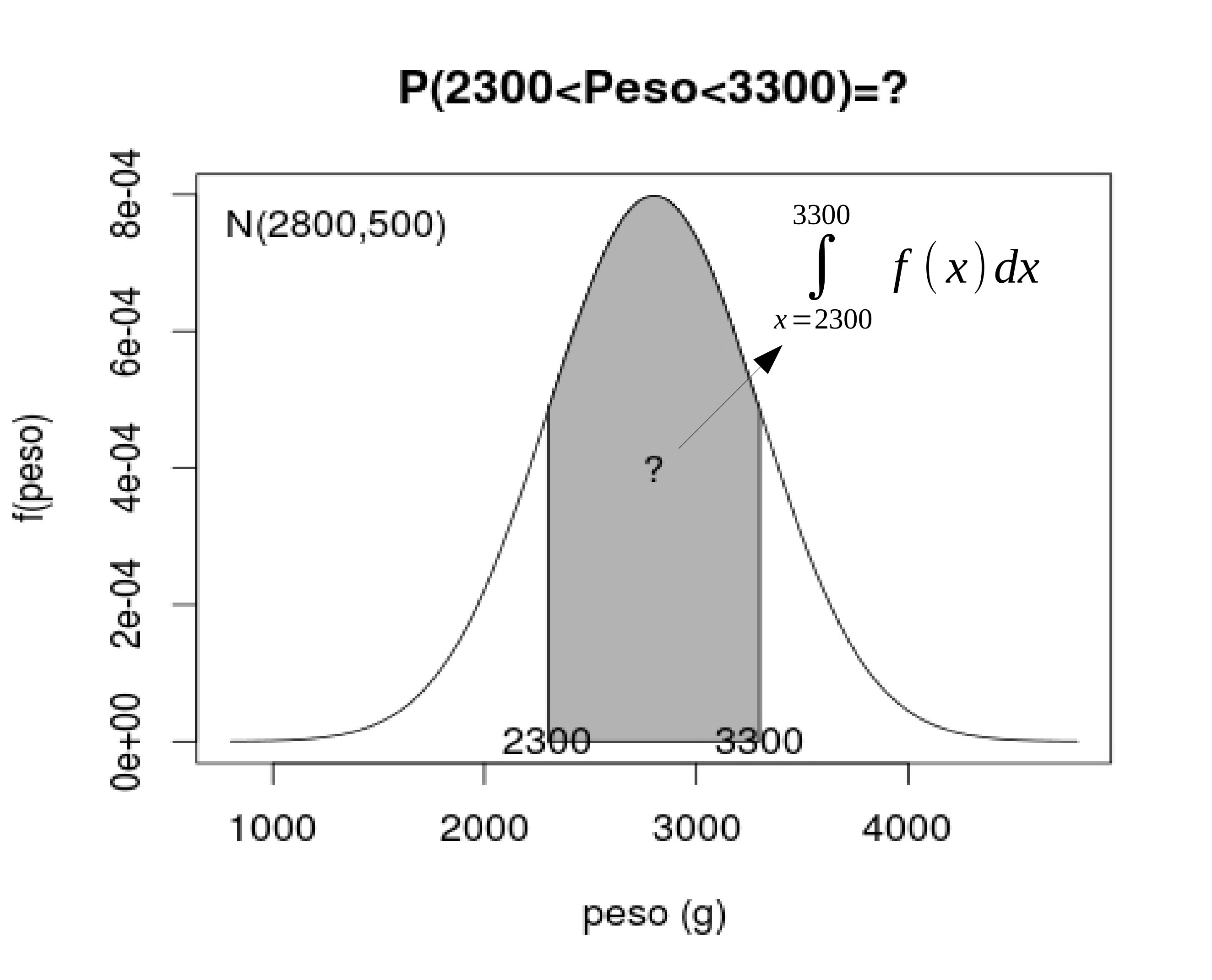
Exemplo: peso de recém-nascidos



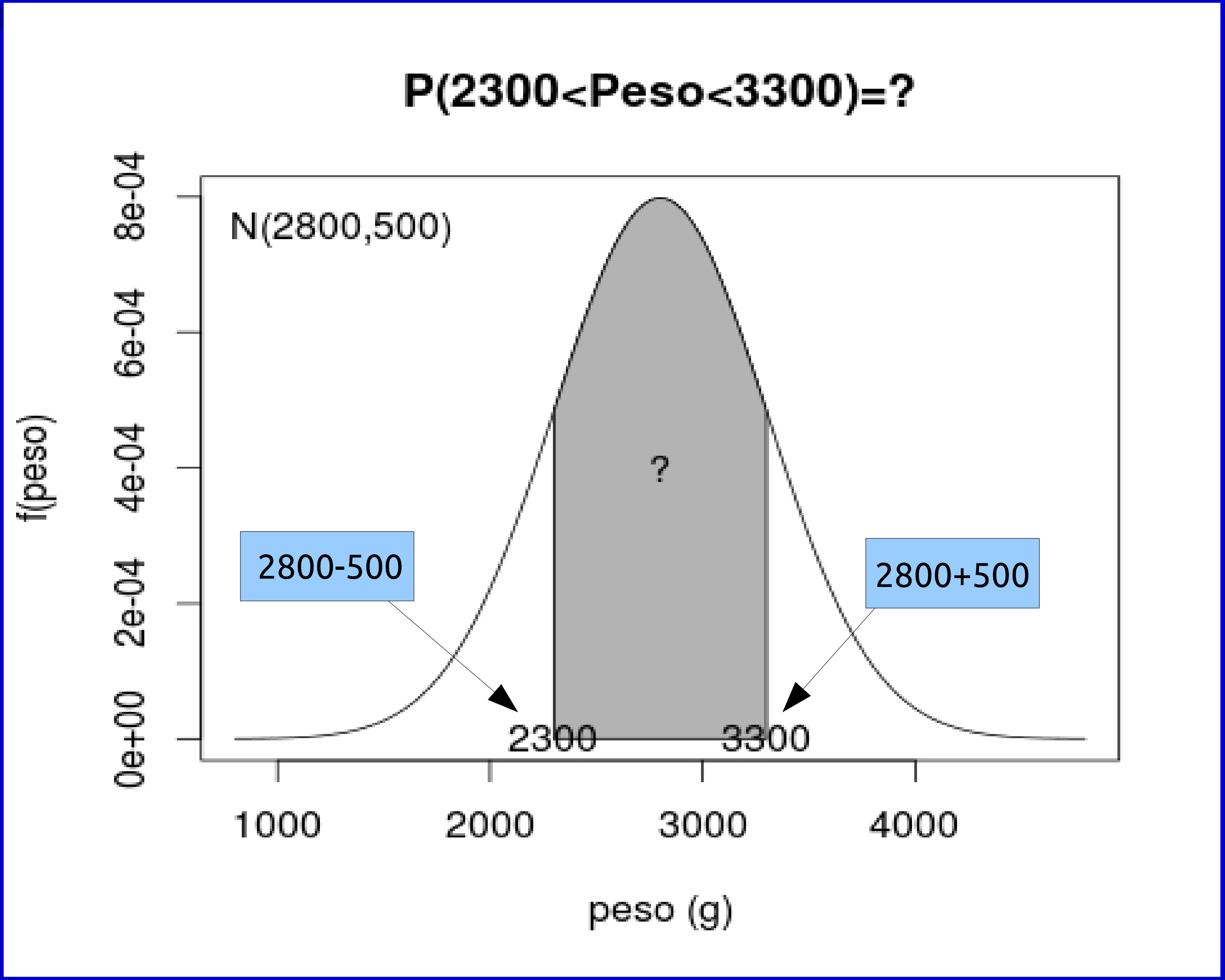
Exemplo: peso de recém-nascidos



Exemplo: peso de recém-nascidos

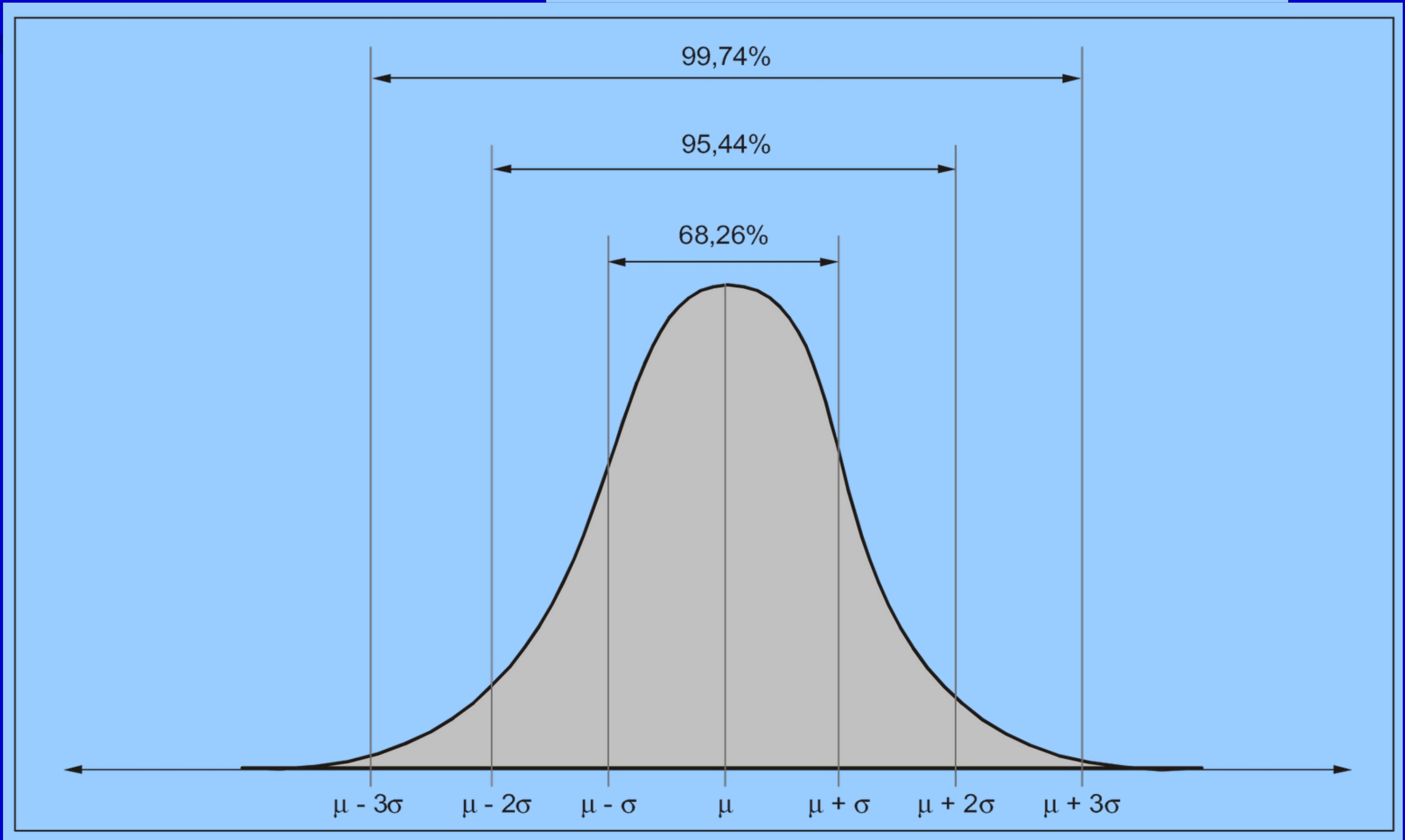


Exemplo: peso de recém-nascidos

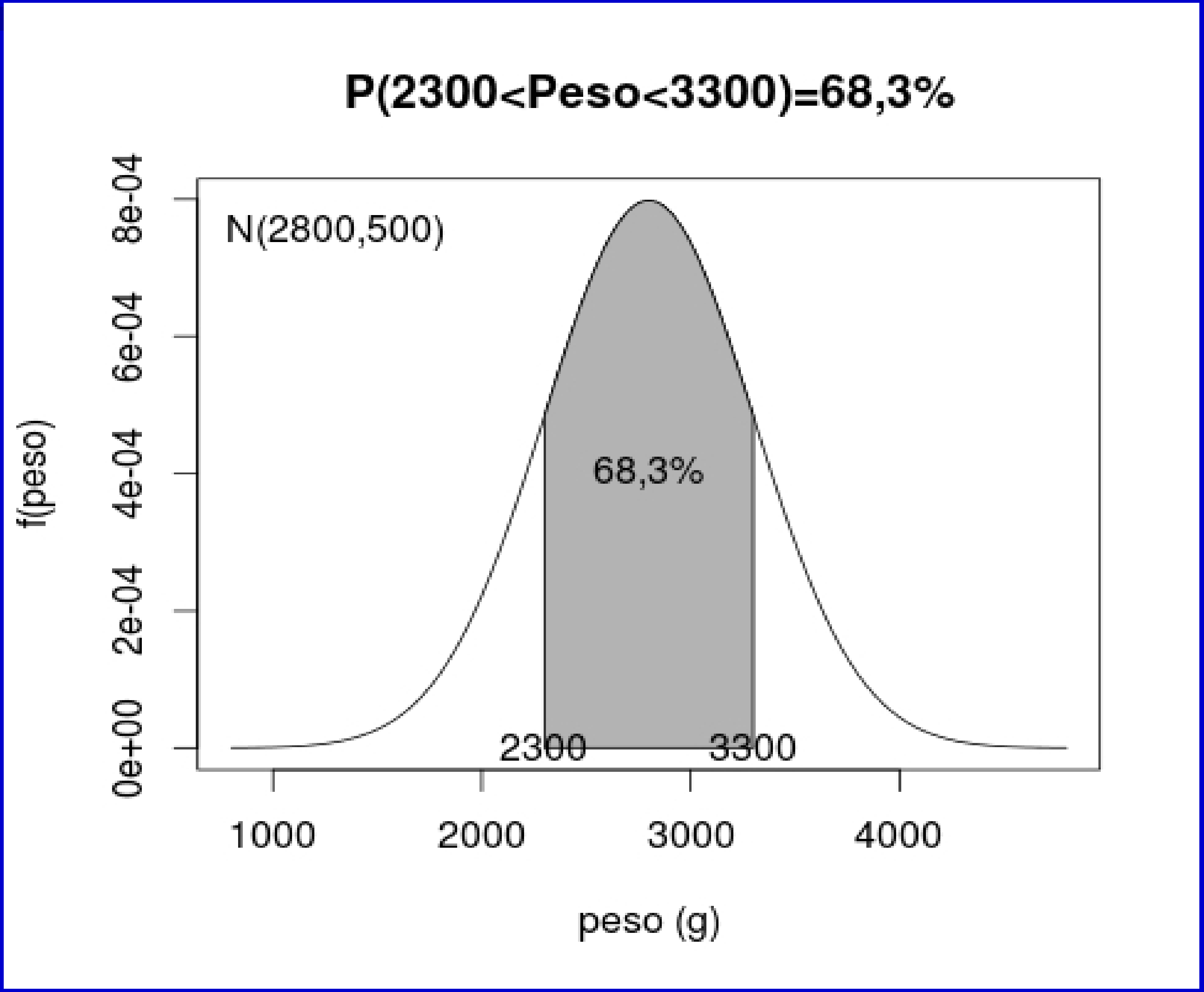


Equação:

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

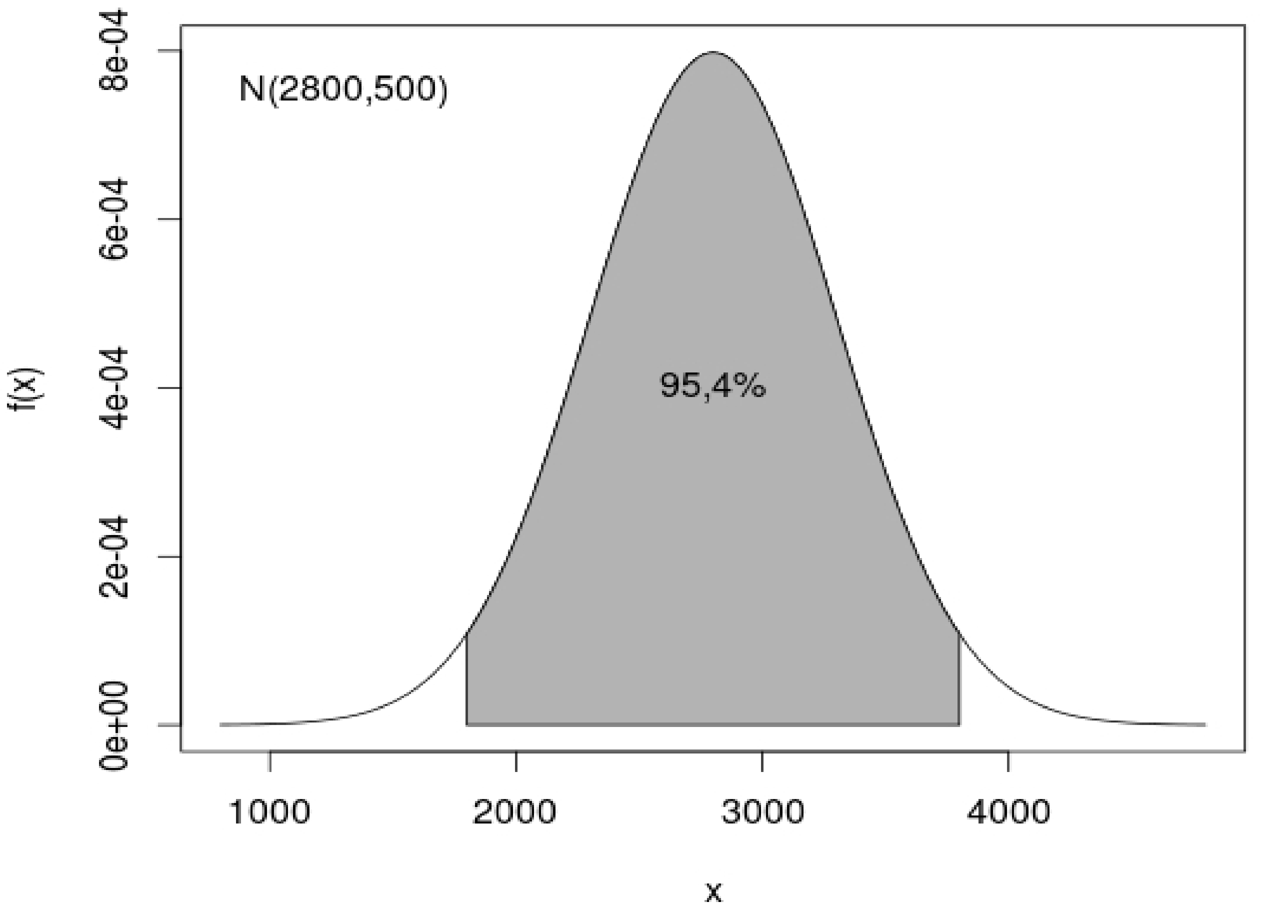


Exemplo: peso de recém-nascidos



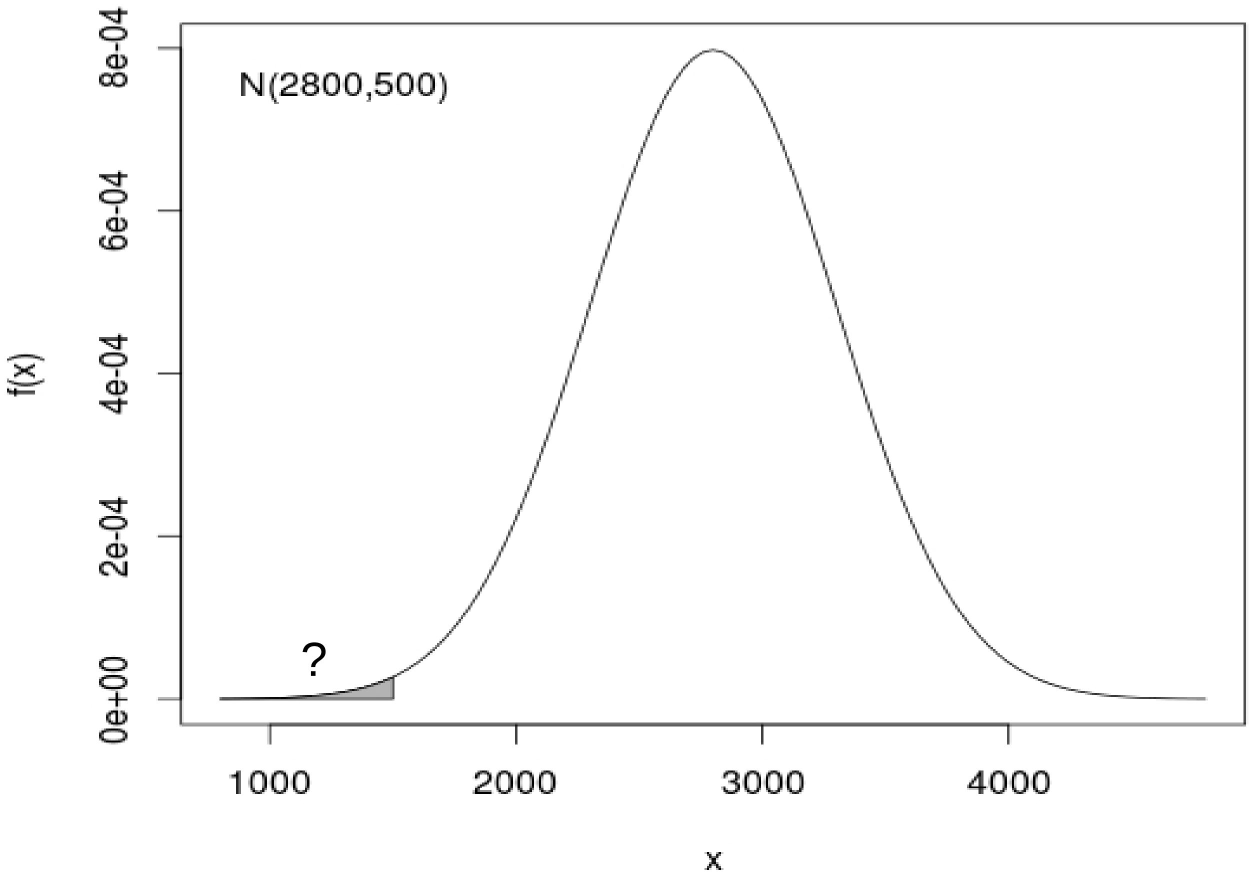
Exemplo: peso de recém-nascidos

P(1800<Peso<3800)=95,4%

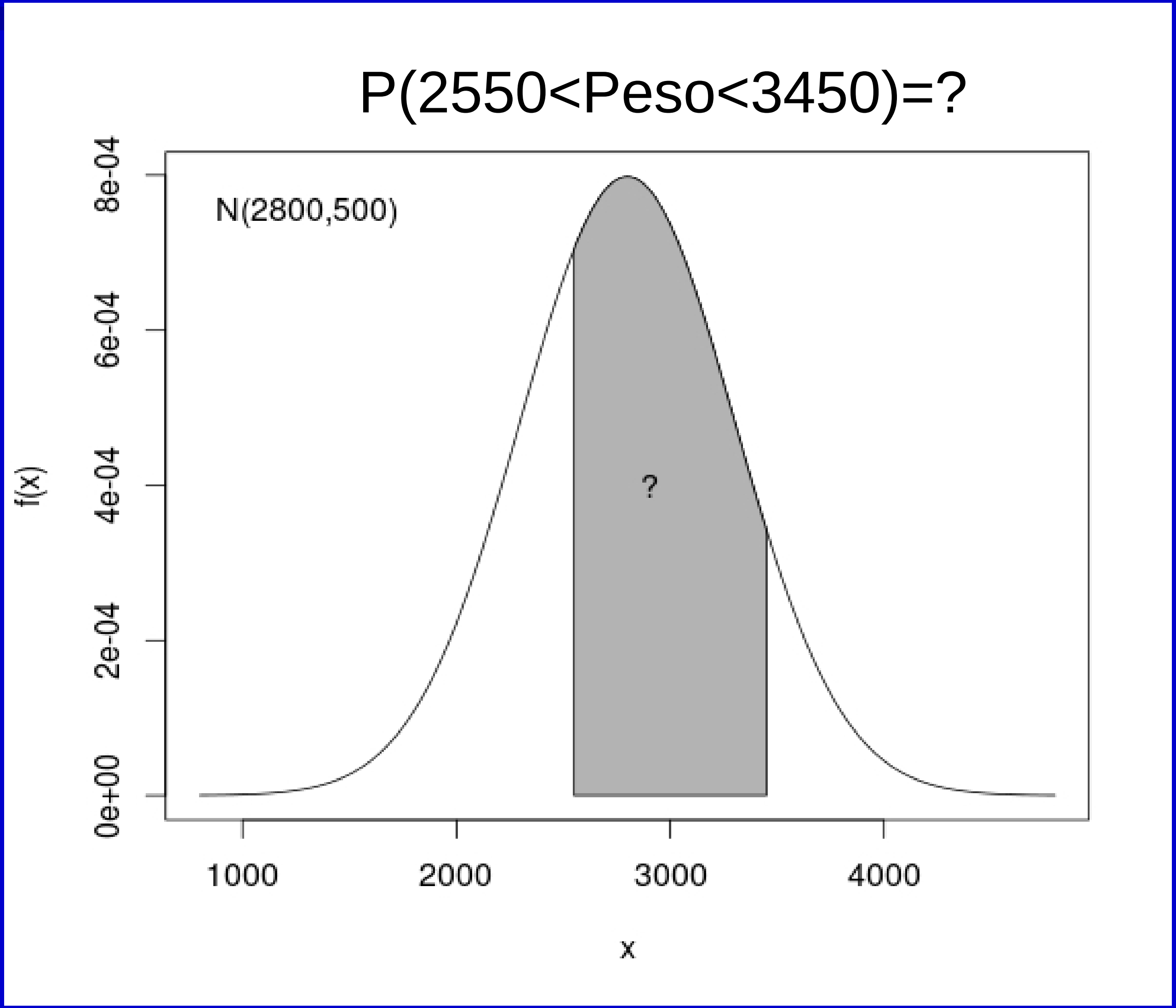


Exemplo: peso de recém-nascidos

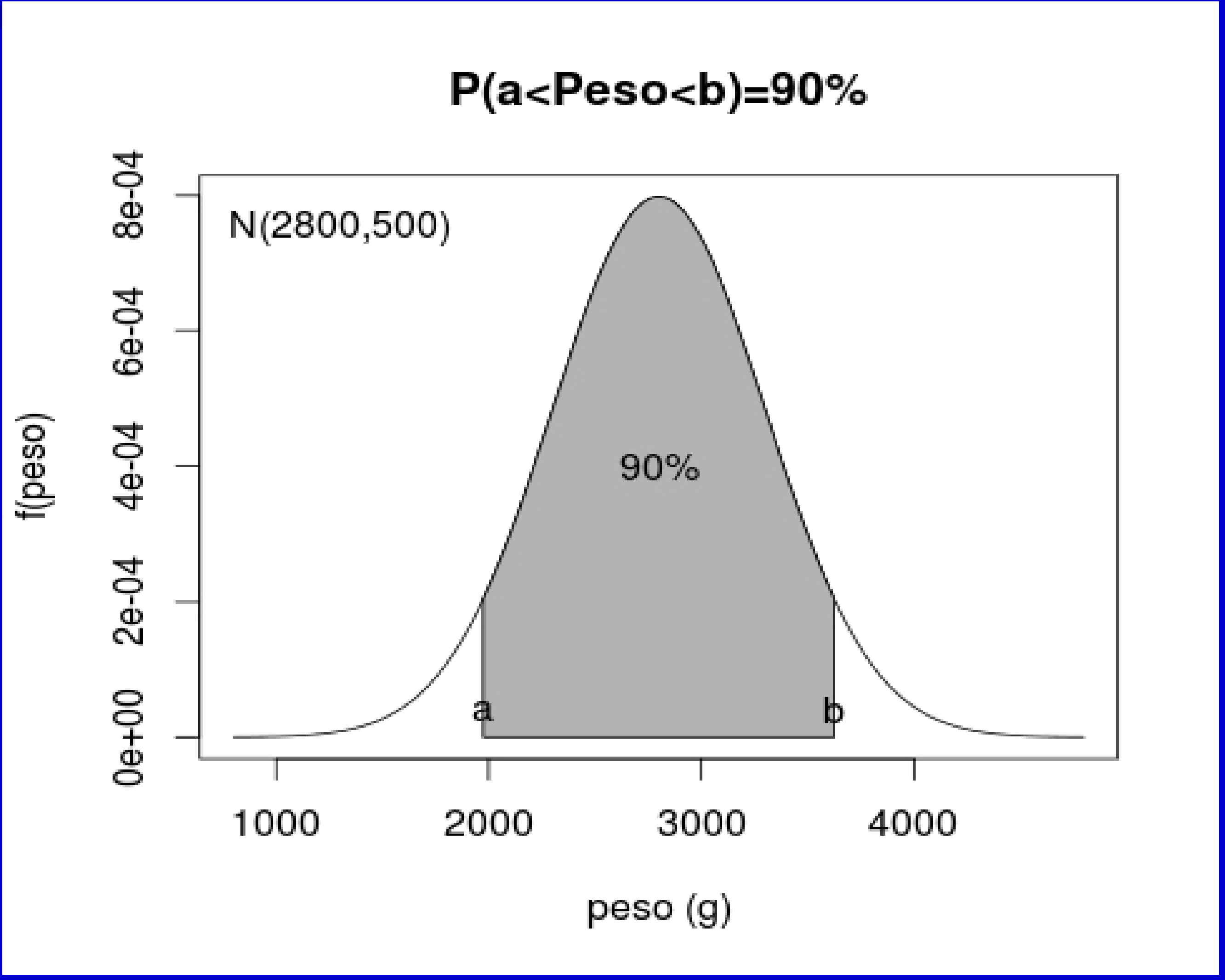
P(peso<1500)=?



Exemplo: peso de recém-nascidos



Exemplo: peso de recém-nascidos



Padronização

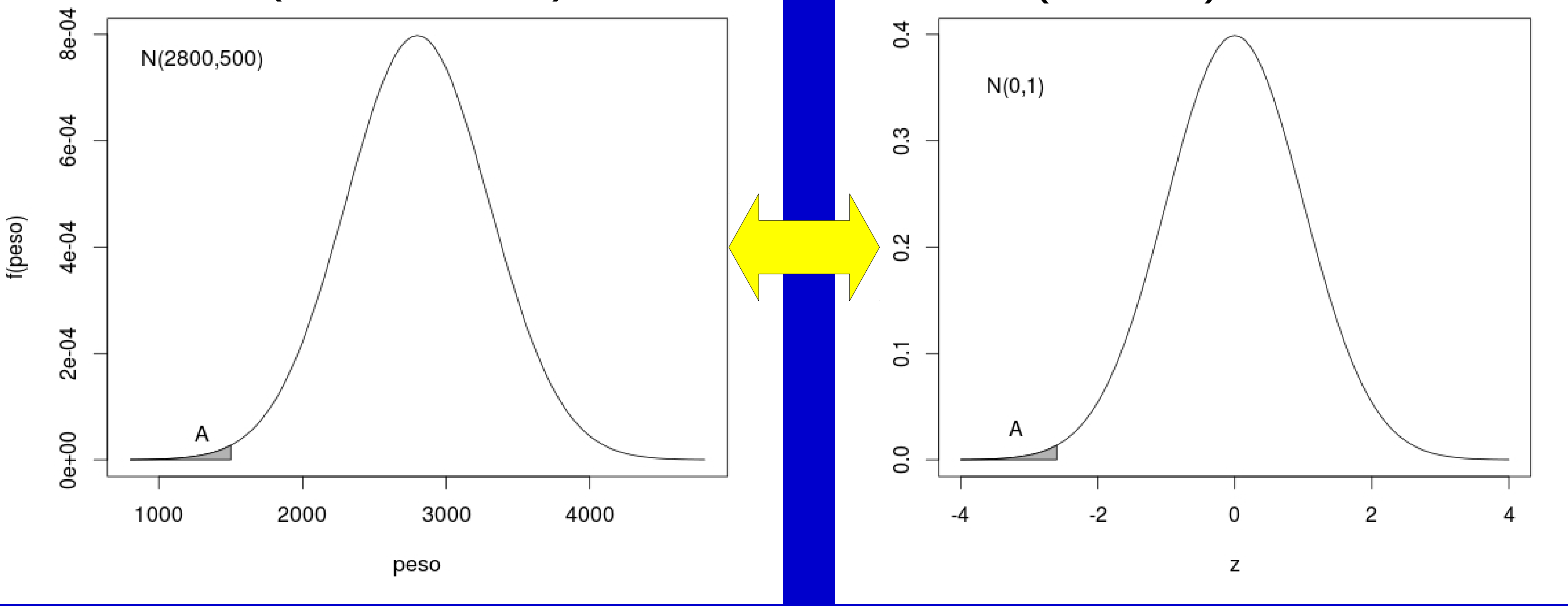
X~N(μ,σ) é transformada numa forma padronizada Z~N(0,1)

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

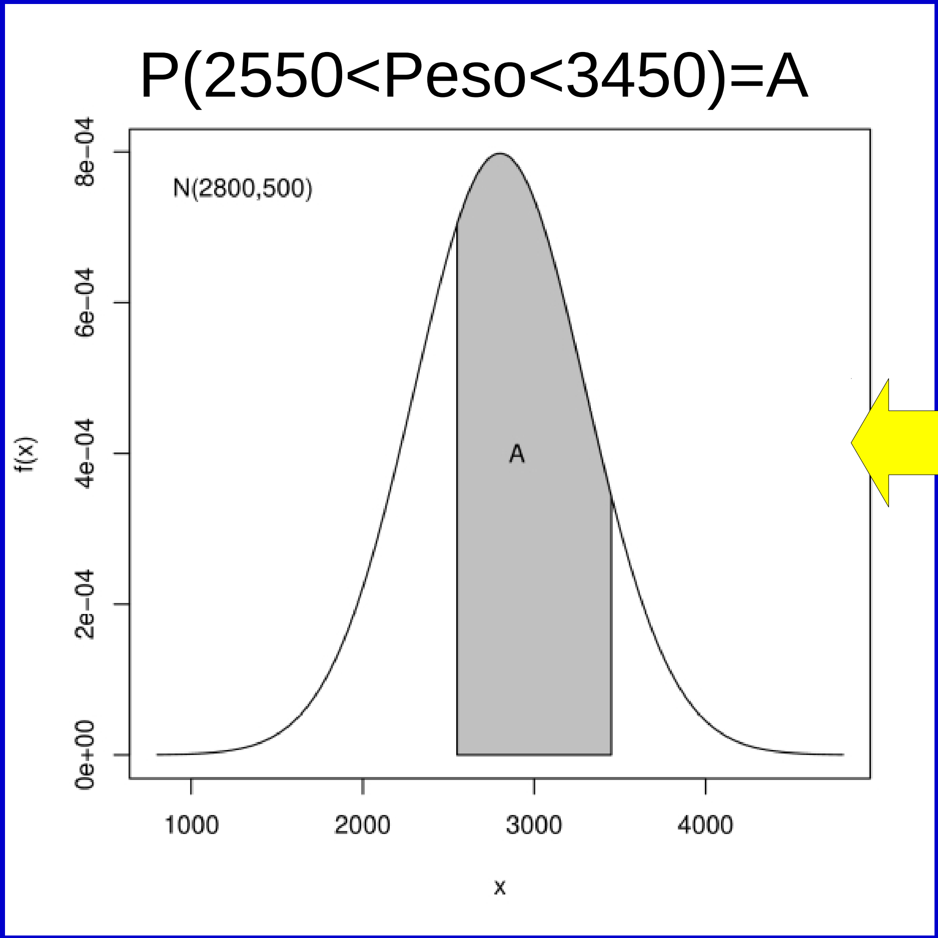
Padronização

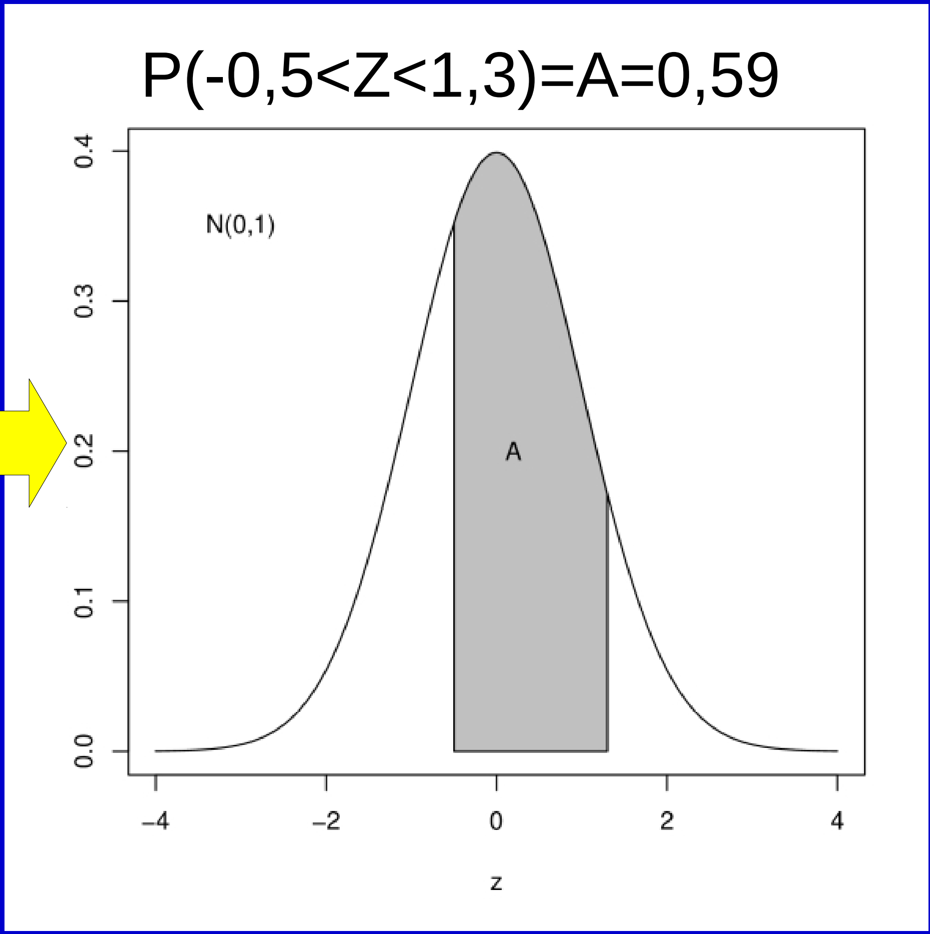
Peso~N(2800,500) é transformado em Z~N(0,1)

P(Peso<1500)=A   P(Z<-2,6)=A=0,005

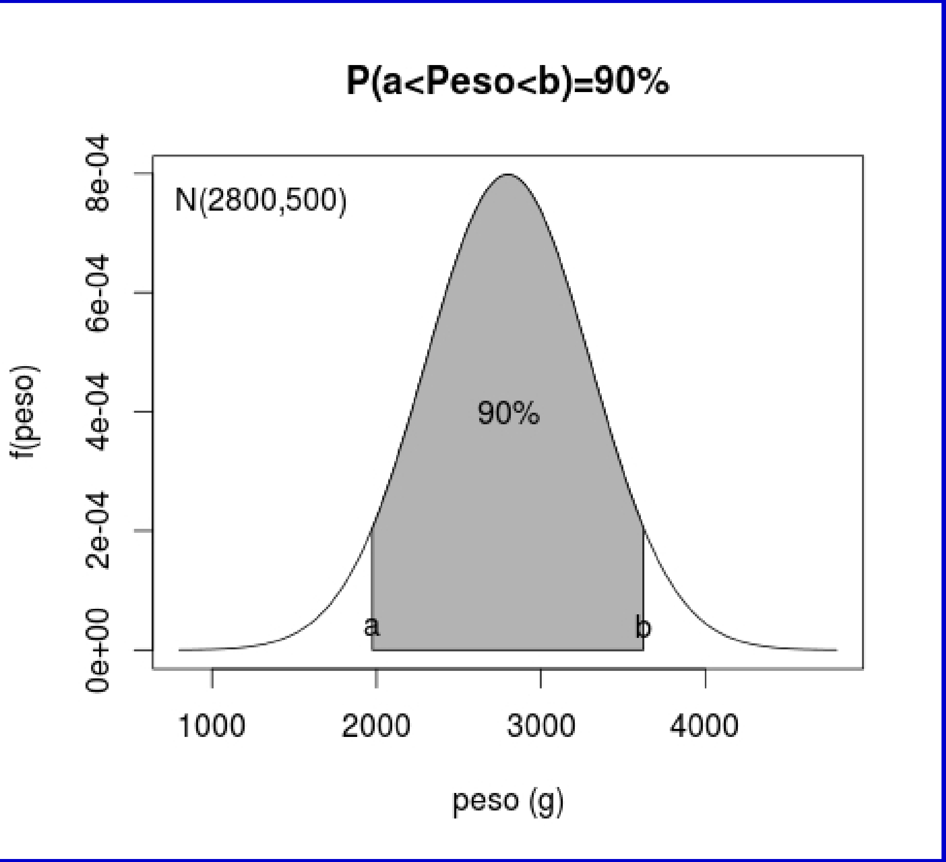


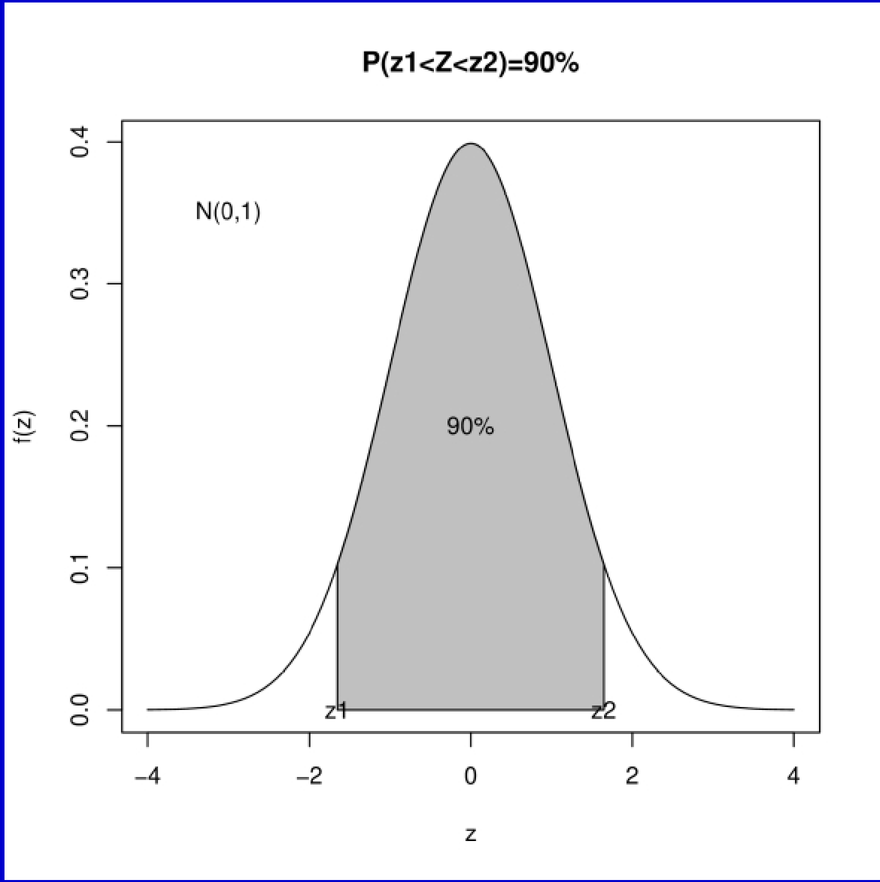
Exemplo: peso de recém-nascidos





Exemplo: peso de recém-nascidos





$$z1 = \frac{(a - 2800)}{500} = -1{,}65$$

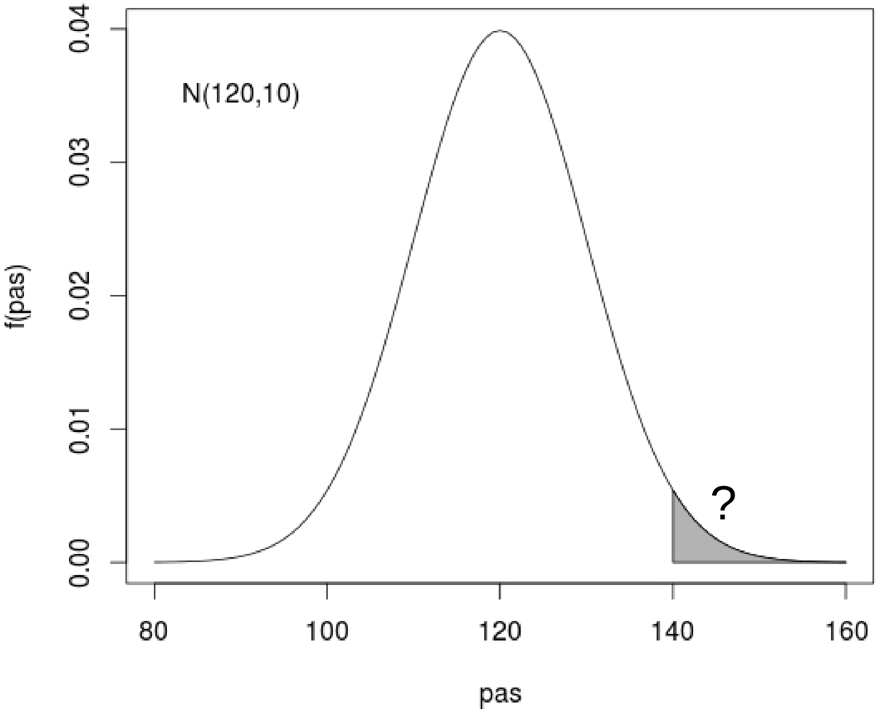
$$z2 = \frac{(b - 2800)}{500} = 1{,}65$$

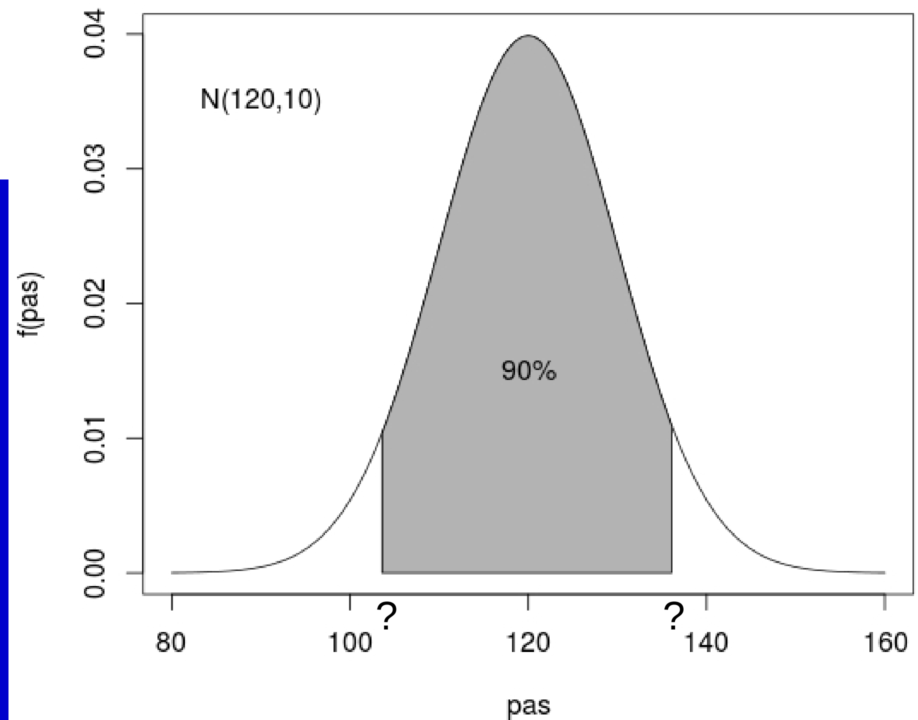
Exemplo: PAS

Suponha que a pressão arterial sistólica de pessoas jovens saudáveis seja N(120,10)

Qual é o percentual dessas pessoas com pressão sistólica acima de 140mmHg?

Qual é o intervalo simétrico em torno da média que engloba 90% dos valores das pressões sistólicas de pessoas jovens e saudáveis?


N(120,10)
?
f(pas)
pas
0.00
0.01
0.02
0.03
0.04
80
100
120
140
160


N(120,10)
90%
?
?
f(pas)
pas
0.00
0.01
0.02
0.03
0.04
80
100
120
140
160

Calculadora

http://onlinestatbook.com/2/calculators/normal.html

Exercício

Supondo que o coeficiente de inteligência (QI) de pessoas saudáveis segue uma distribuição Normal e que o QI médio seja 100, com desvio-padrão 15.

- Qual é a probabilidade de uma pessoa dessa população ter QI acima de 130?
- Qual é a probabilidade de uma pessoa dessa população ter QI entre 90 e 110?
- Se quisermos definir um ponto de corte para o QI que deixa somente 1% das pessoas dessa população acima desse valor, qual seria?
- Qual seria a faixa de referência de 95% para QI nessa população?