Estatística aplicada a ensaios clínicos

RAL - 5838

Luís Vicente Garcia

lvgarcia@fmrp.usp.br

Faculdade de Medicina de Ribeirão Preto

Estatística aplicada a ensaios clínicos

aula 6

Variável aleatória

possui resultados ou valores que tendem a variar de uma observação para outra em razão de fatores relacionados com a chance

representa um valor numérico associado a cada um dos resultados de um experimento probabilístico

Quando uma variável tem resultados ou valores que tendem a variar de uma observação para outra em razão de fatores relacionados com a chance ela recebe o nome de

Variável aleatória

Variável aleatória

discreta

contagens

contínua

qualquer valor de determinado intervalo

Distribuição de Probabilidades

é uma distribuição de frequências para os resultados de um espaço amostral

as frequências são relativas ou probabilidades

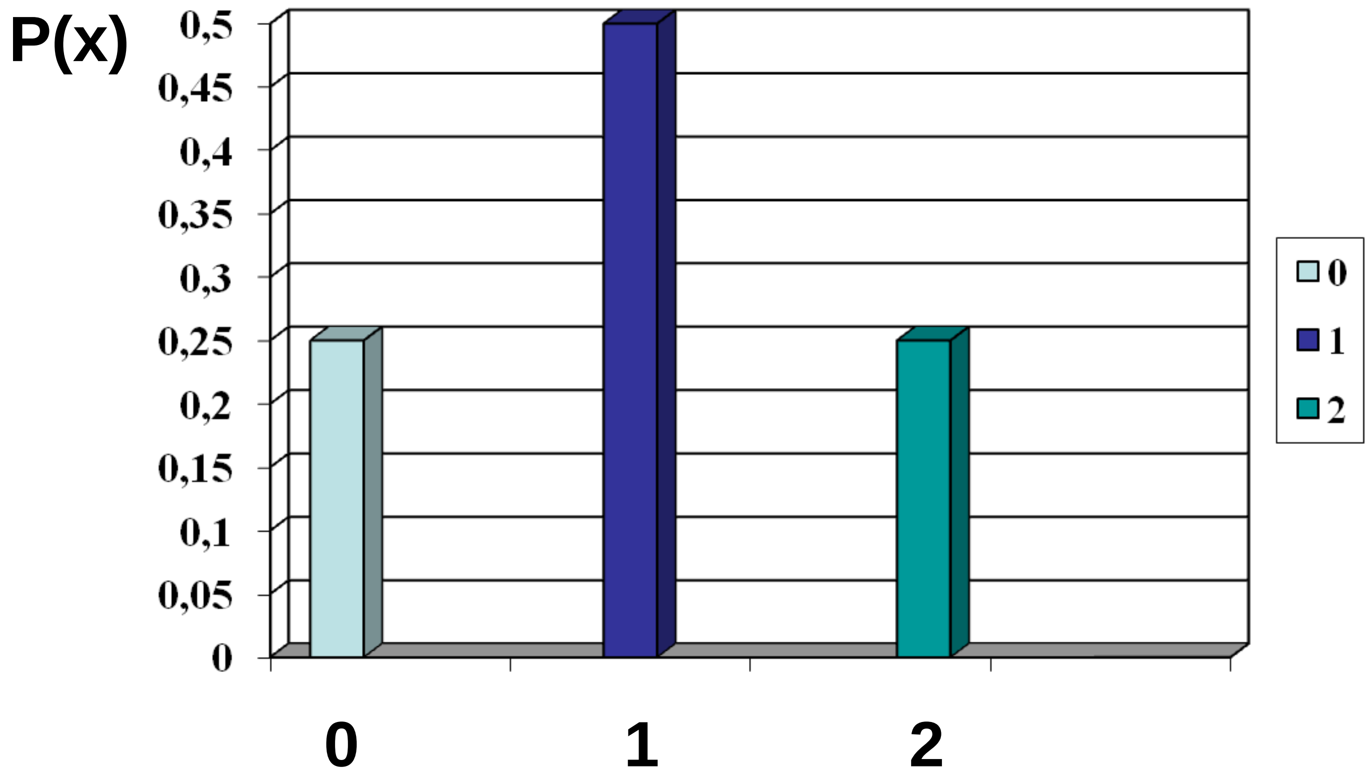
Distribuição de Probabilidades

Mostra a proporção das vezes que a variável aleatória tende a assumir cada um dos diversos valores

Distribuição de Probabilidades

número de caras em duas jogadas de uma moeda

Resultado	VALOR DE V.A.	PROB.	Nº CARAS		P(x)
CC	0	1/4	0		0,25
CK	1	1/4	1	0,75	0,25
KC	1	1/4	1		0,25
KK	2	1/4	2		0,25


P(x)
0,5
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
0
1
2
0
1
2

Distribuição de Probabilidades

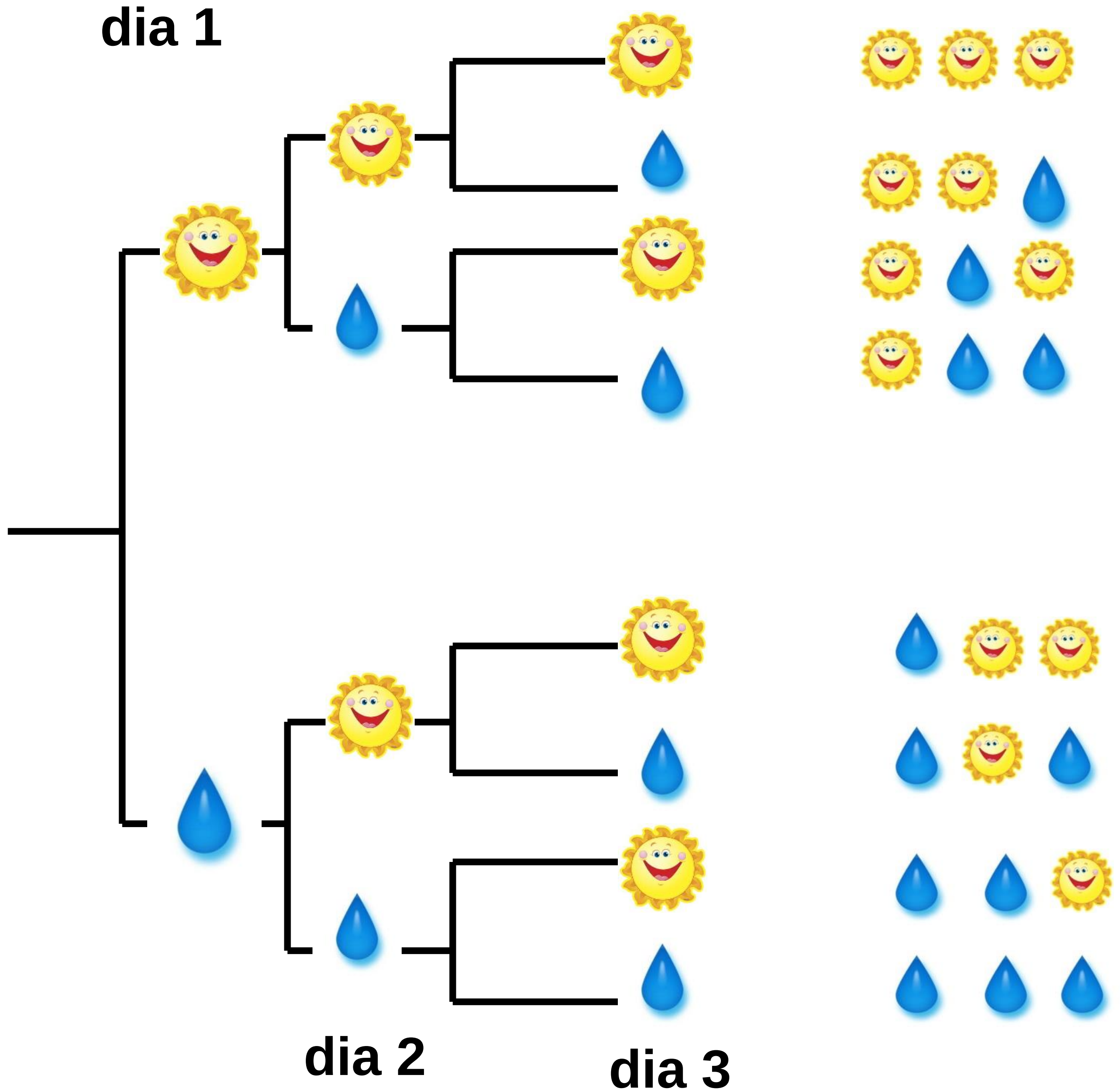
é uma distribuição de frequências
para os resultados de um espaço amostral

as frequências são relativas ou probabilidades

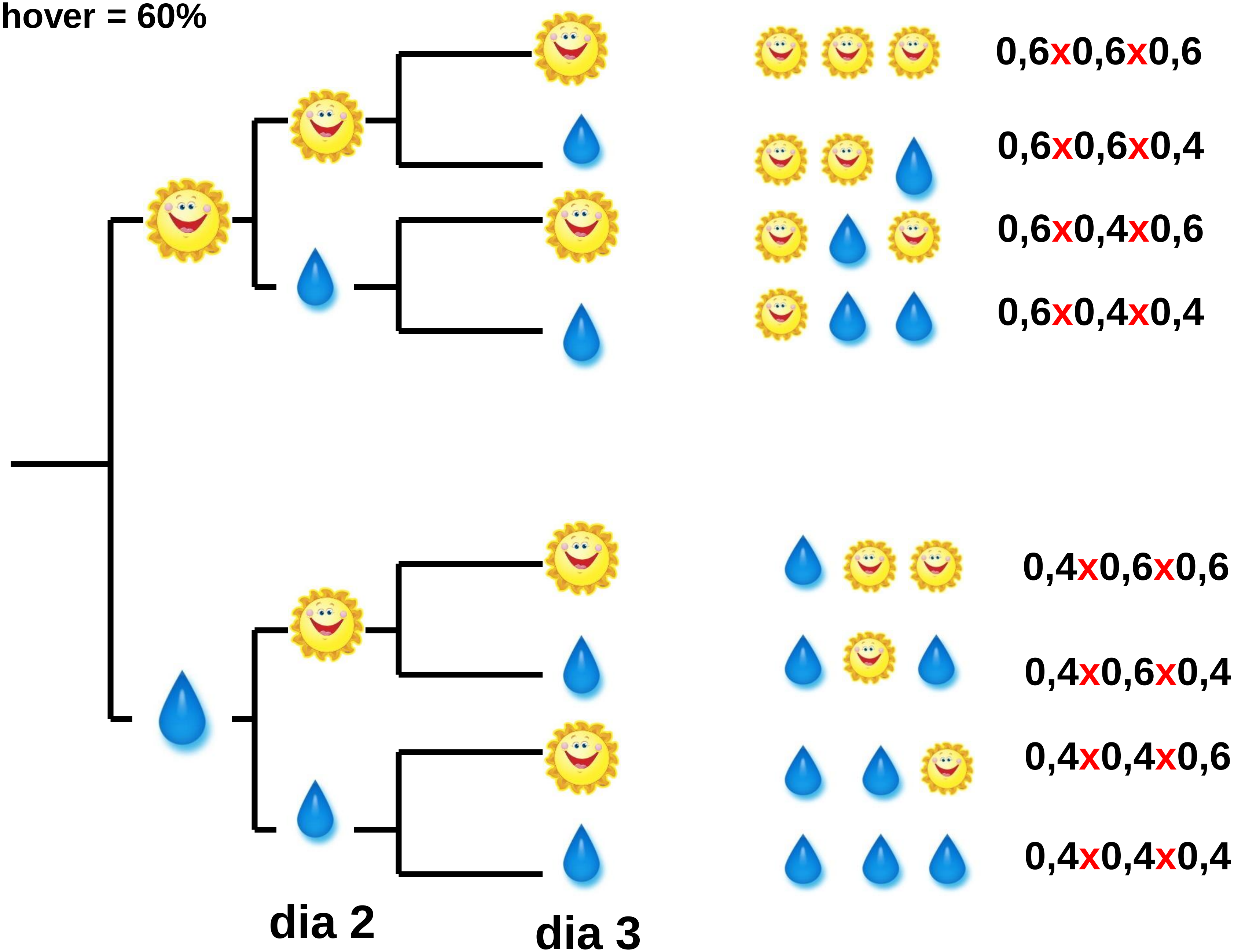
Distribuição de Probabilidades

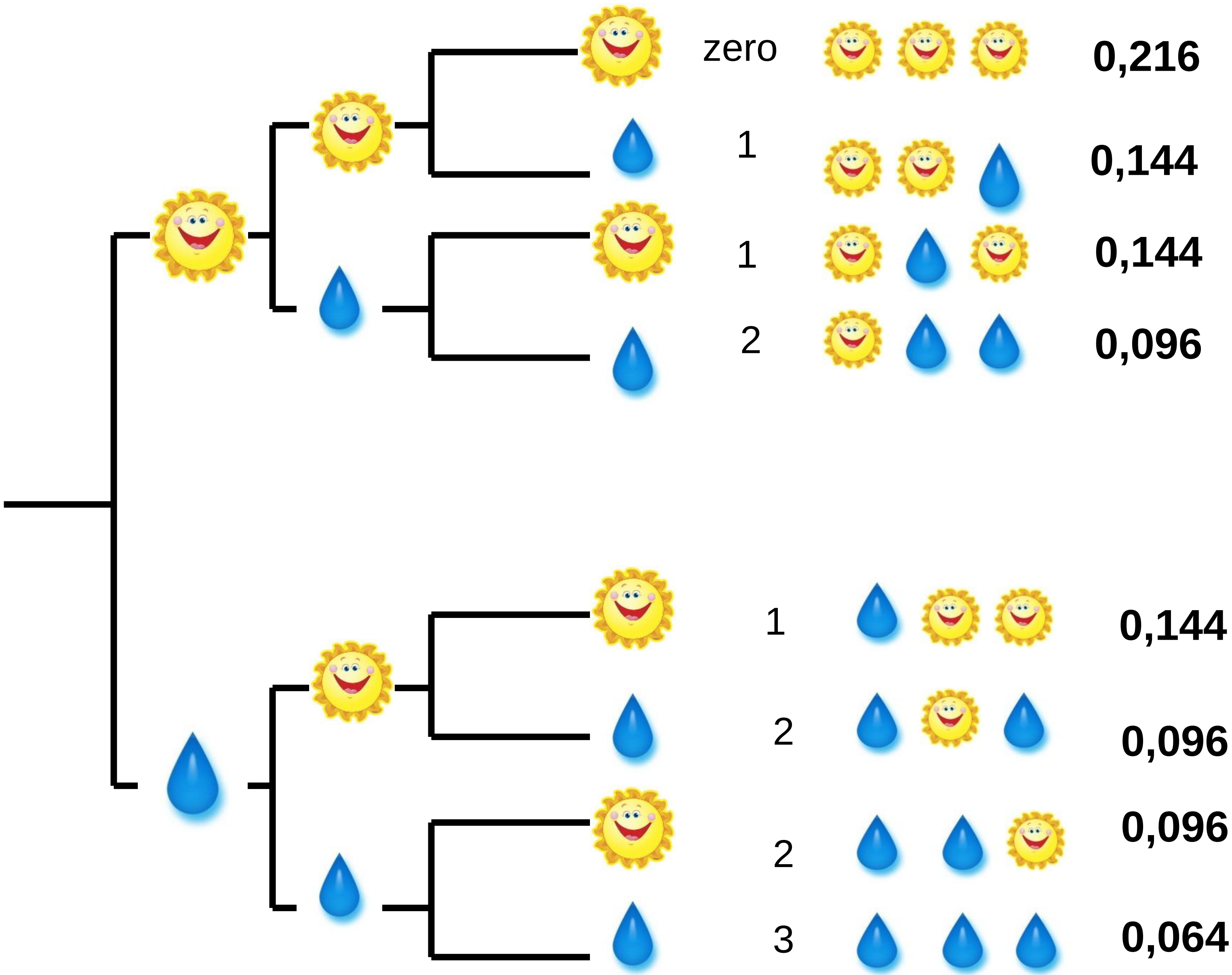
Exemplo: feriadão na praia

P (chover) = 40%
P (não chover) = 60%

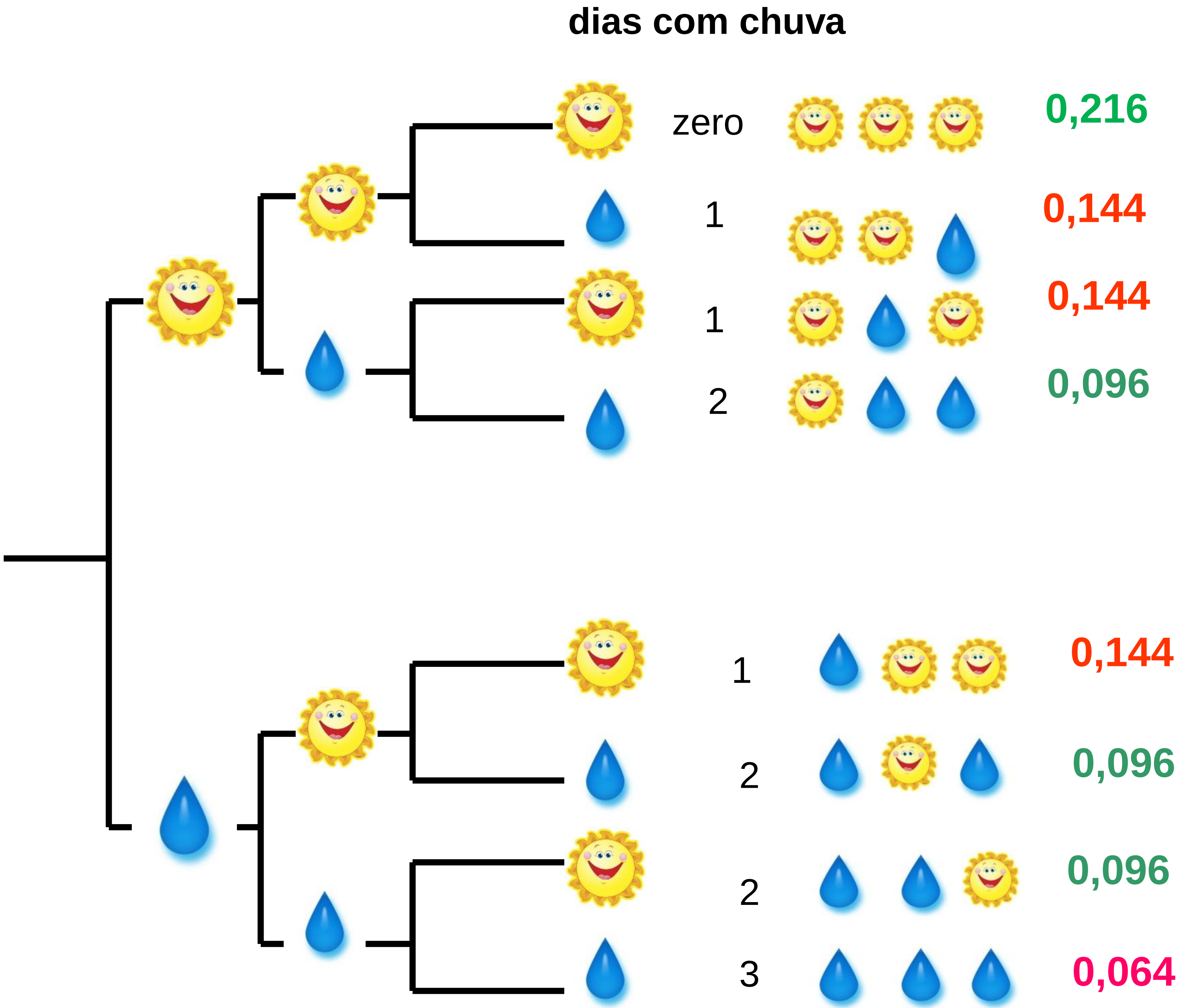

dia 1
dia 2
dia 3

chover = 40%
não chover = 60%



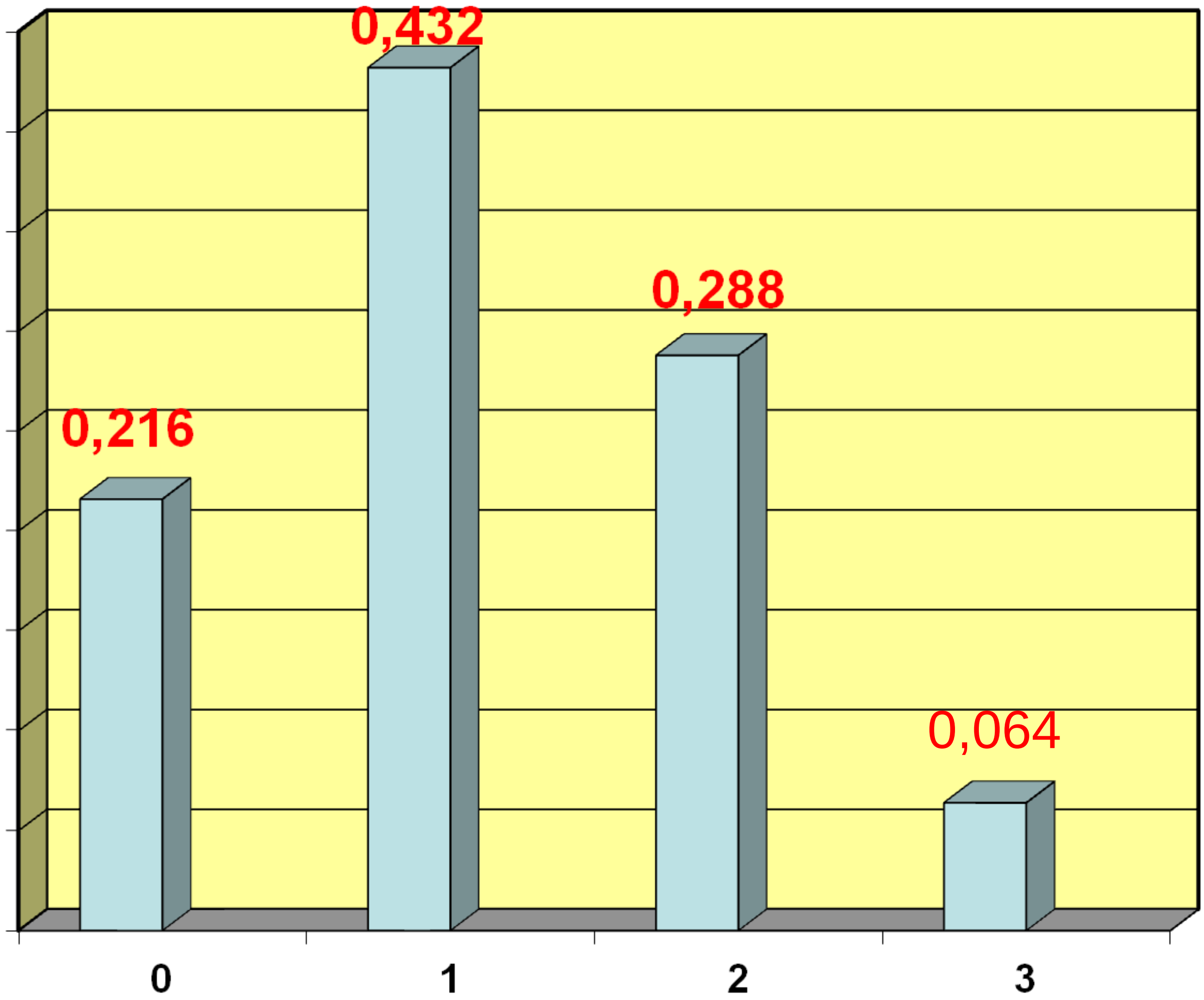


zero
0,216
1
0,144
1
0,144
2
0,096
1
0,144
2
0,096
2
0,096
3
0,064

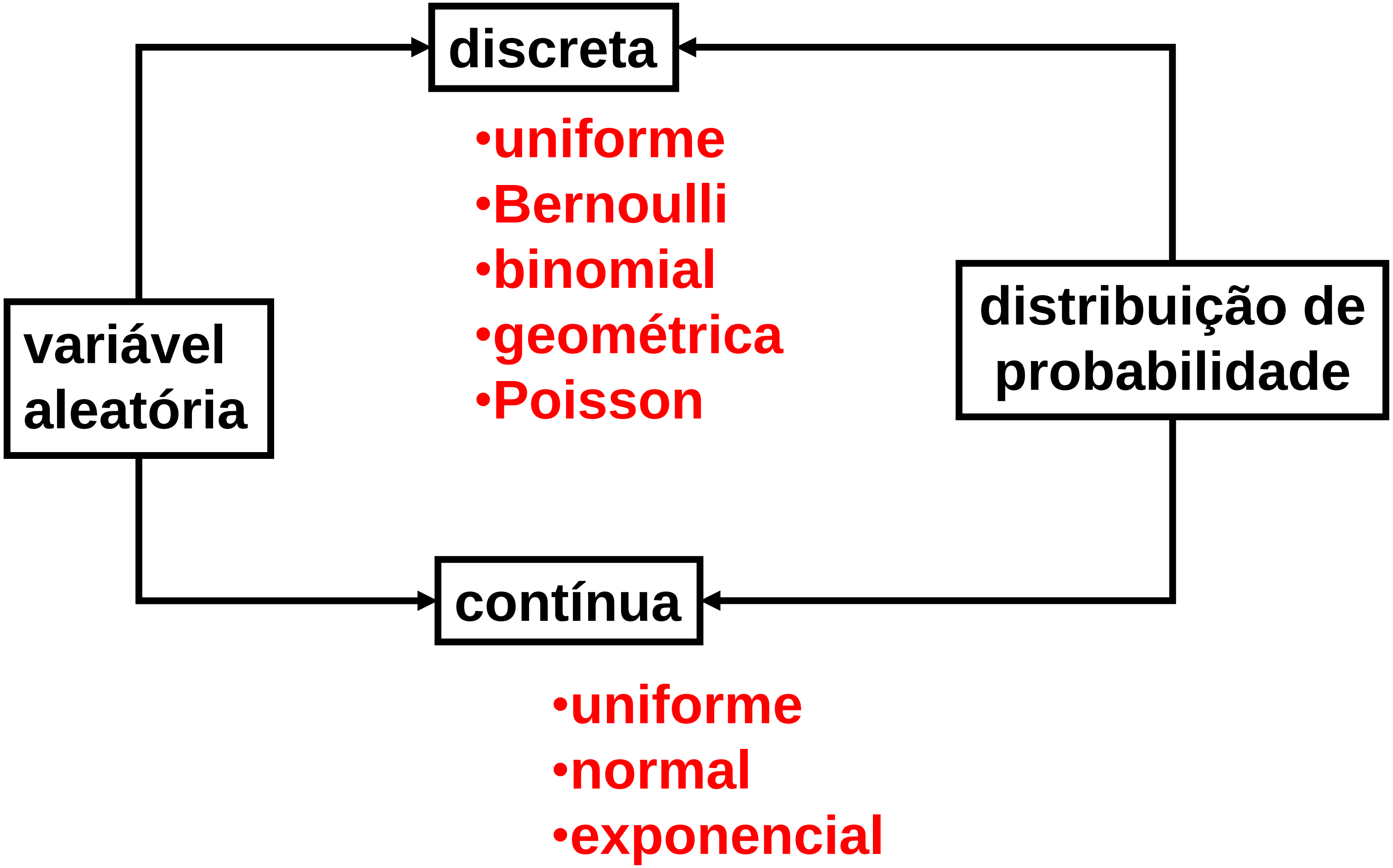

dias com chuva
zero
0,216
1
0,144
1
0,144
2
0,096
1
0,144
2
0,096
2
0,096
3
0,064

distribuição de probabilidade

dias de chuva	contagem	probabilidade
zero	1	0,216
um	3	0,432
dois	3	0,288
três	1	0,064


0,216
0,432
0,288
0,064
0
1
2
3

distribuição de probabilidade



distribuição de probabilidade

➡ Discretas

- uniforme
- Bernoulli
- Binomial
- geométrica
- Poisson

distribuição discretas de probabilidade

x **= número de respostas corretas**

x **= número de chegadas pontuais**

x **= número de pontos feitos num jogo**

distribuição discreta de probabilidade

Uma distribuição discreta de probabilidade enumera cada valor possível da variável aleatória, bem como sua probabilidade.

distribuição discreta de probabilidade

x	$P(x)$
0	0,004
1	0,435
2	0,355
3	0,206

número de filhos em cada família

distribuição discreta de probabilidade

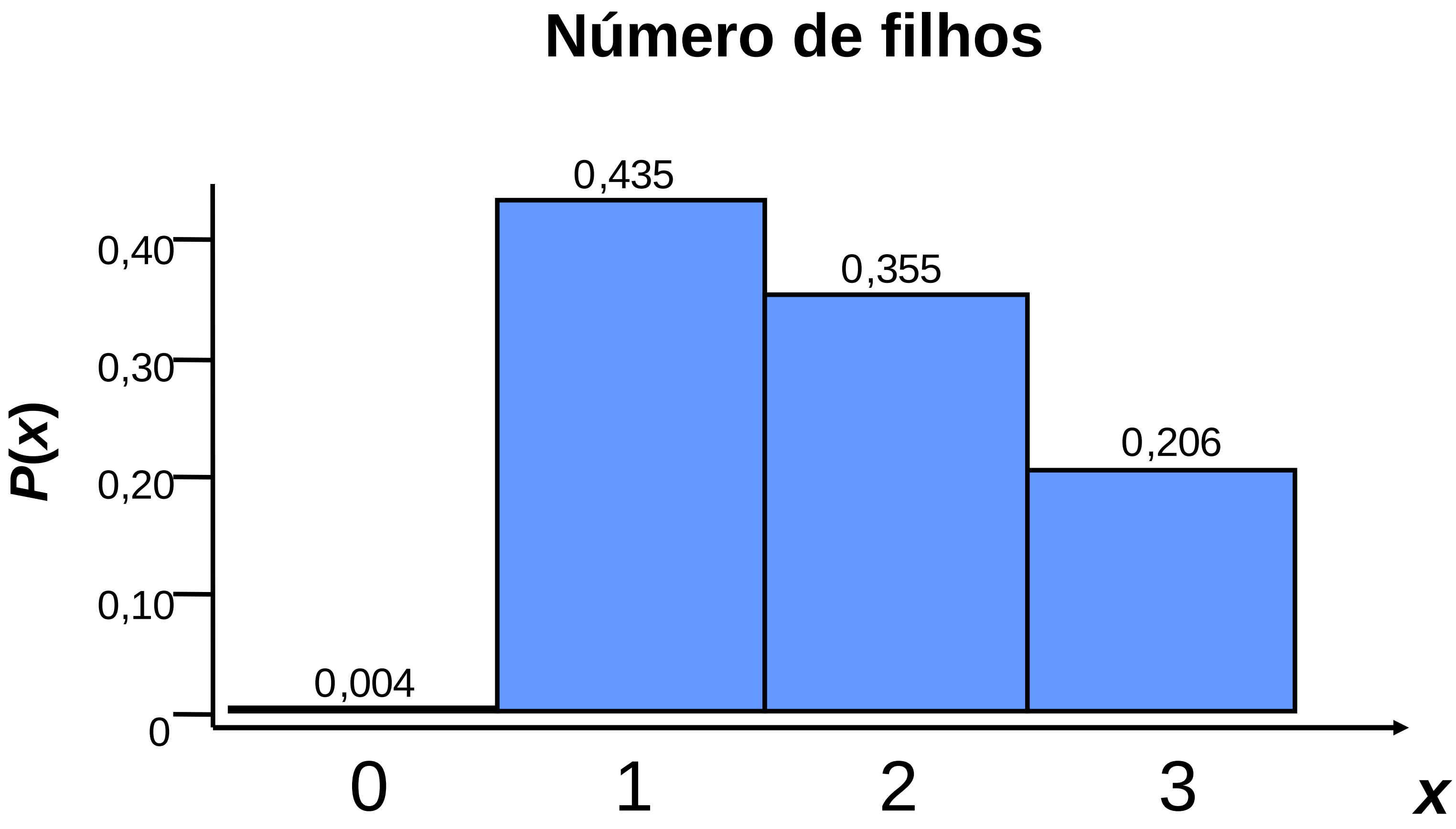
Propriedades

- Cada probabilidade precisa estar entre 0 e 1, inclusive.
- A soma de todas as probabilidades é 1.

$$0 \leq P(x) \leq 1$$

$$\Sigma P(x) = 1$$

Histograma de probabilidade



distribuição discreta

A média

$$\mu = \Sigma x \cdot P(x)$$

A variância

$$\sigma^2 = \Sigma(x - \mu)^2 \cdot P(x)$$

O desvio padrão

$$\sigma = \sqrt{\sigma^2}$$

média

Calcule a média: $$\mu = \sum x \cdot P(x)$$

Multiplique cada valor por sua probabilidade. Some os produtos.

x	$P(x)$	$xP(x)$
0	0,004	0
1	0,435	0,435
2	0,355	0,71
3	0,206	0,618
		1,763

O valor esperado (a média) é de 1,763 FILHOS.

variância e desvio-padrão

A média é de 1,763 FILHOS. $\sigma^2 = \Sigma(x - \mu)^2 \cdot P(x)$

x	$P(x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2 P(x)$
0	0,004	-1,763	3,108	0,012
1	0,435	-0,763	0,582	0,253
2	0,355	0,237	0,056	0,020
3	0,206	1,237	1,530	0,315
				0,601

variância

$\sigma \sqrt{\sigma^2} = \sqrt{0{,}661} = 0{,}775$

O desvio padrão é de 0,775 FILHOS.

distribuição binomial

Características

- **O número de tentativas é fixo (*n*).**
- **As *n* tentativas são independentes e repetidas em condições idênticas.**
- **Para cada tentativa há dois resultados possíveis, *S* = sucesso *ou* *F* = fracasso.**

distribuição binomial

Características

- A probabilidade de sucesso numa tentativa única é p. $P(S) = p$
- A probabilidade de fracasso é q. $P(F) = q$, *onde* $p + q = 1$
- O problema central está em determinar a probabilidade de x sucessos em n tentativas, sendo x = 0 ou 1 ou 2 … n.

A variável aleatória *x* é uma contagem do número de sucessos em *n* tentativas.

Experimentos binomiais

sexta-feira à tarde

convites para balada

prova na segunda

Prova:

3 questões
5 alternativas cada
acerto mínimo = 2

teste

1 – Importante Primeiro Ministro da China – Mao Tsé:

a) Tang
b) Teng
c) Ting
d) Tong
e) Tung

2 – Importante cidade do oriente – Bag_ _:

a) dá
b) dé
c) di
d) dó
e) du

3 – Média é:

a) Leite + café
b) Leite + açúcar
c) valores extremos/2
d) Soma dos valores/número valores
e) puxa-saquismo

Probabilidades binomiais

P de 2 questões em três (5 alternativas)

CCC *CEE*
CCE *ECE*
CEC *EEC*
ECC *EEE*

P (acerto) = 0,2 P (erro) = 0,8

	EEE	CEE ECE EEC	CCE CEC ECC	CCC
certas	0	1	2	3
P(x)	0,512	0,384	0,096	0,008

P(x=3) = P(3 certas) = $0,2^3$ = 0,008

P(x=2) = P(2 certas) = 3 x $0,2^2$ x 0,8 = 0,096

P(x=1) = P(1 certa) = 3 x 0,2 x $0,8^2$ = 0,384

P(x=0) = P(0 certa) = $0,8^3$ = 0,512

Probabilidades binomiais

P de 2 questões em três (5 alternativas)

CCC *CEE*
CCE *ECE*
CEC *EEC*
ECC *EEE*

P (acerto) = 0,2

P (erro) = 0,8

	EEE	CEE ECE EEC	CCE CEC ECC	CCC
certas	0	1	2	3
P(x)	0,512	0,384	0,096	0,008

0,104

Resultados do teste

número de questões corretas = X

Por que esse foi um experimento binomial?

Quais são os valores de n, p e q?

Quais são os valores possíveis de x?

Um teste de múltipla escolha tem 5 questões, cada qual com cinco alternativas, uma delas correta. Você quer saber qual a probabilidade de 'chutar' certo em exatamente 3 questões. Determine n, p, q e x.

$n = 5$ $p = 1/5$ $q = 4/5$ $x = 3$

Um médico lhe diz que certa cirurgia é bem-sucedida em 80% das vezes. Se a cirurgia for realizada sete vezes, determine a probabilidade de ser bem-sucedida em exatamente seis. Determine n, p, q e x.

$n = 7$ $p = 0,80$ $q = 0,20$ $x = 6$

distribuição binomial

a probabilidade de ocorrerem exatamente *x* sucessos em *n* tentativas é de:

$$\boxed{P(x) = {}_nC_x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}}$$

distribuição binomial

Teste com cinco questões e quatro alternativas

calcular a probabilidade de alguém não acertar nenhuma questão, exatamente uma, duas, três, quatro ou todas as cinco questões do teste.

$$P(0) = {}_5C_0 p^0 q^{5-0} = \frac{5!}{0!(5-0)!} (0,25)^0 (0,75)^5 = 0,237$$

$$P(1) = {}_5C_1 p^1 q^{5-1} = \frac{5!}{1!(5-1)!} (0,25)^1 (0,75)^4 = 0,396$$

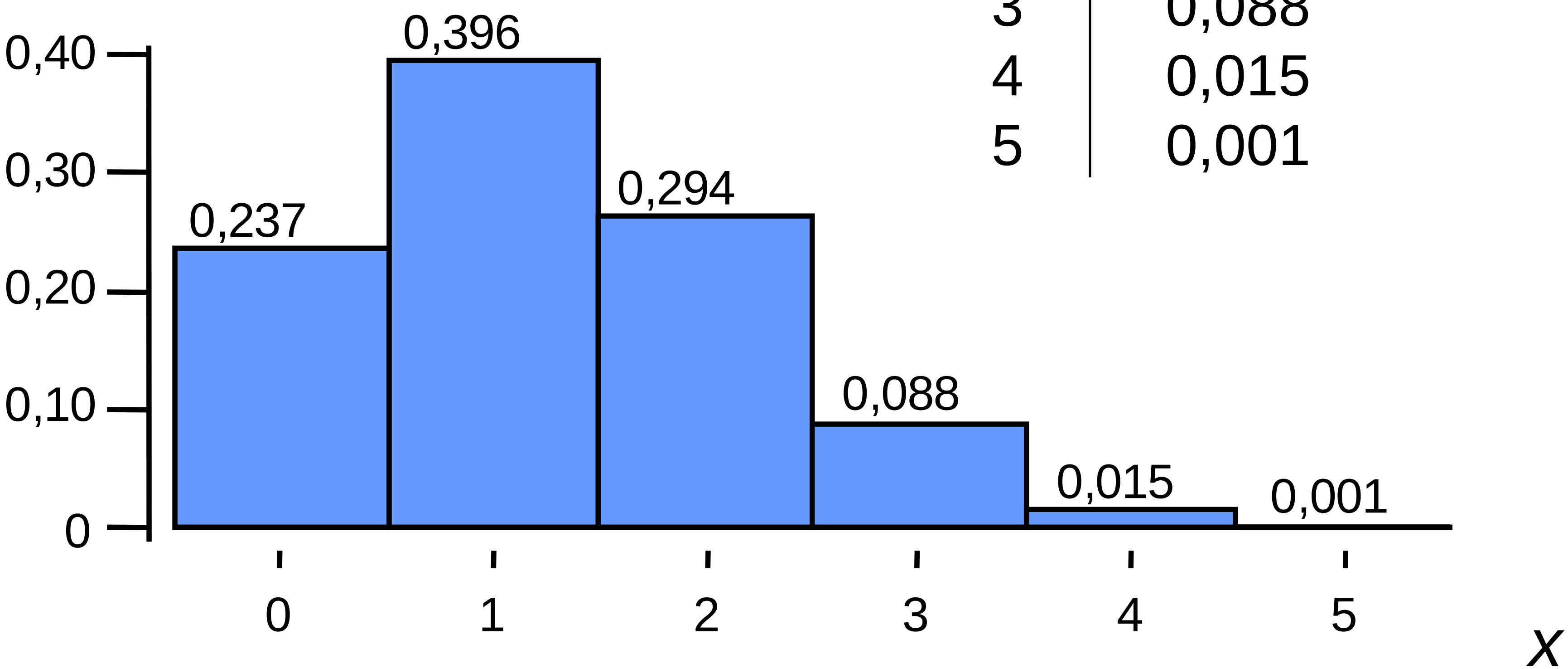
$$P(2) = {}_5C_2 p^2 q^{5-2} = \frac{5!}{2!(5-2)!} (0,25)^2 (0,75)^3 = 0,264$$

$P(3) = 0,088$ $\quad P(4) = 0,015$ $\quad P(5) = 0,001$

distribuição binomial

x	$P(x)$
0	0,237
1	0,396
2	0,264
3	0,088
4	0,015
5	0,001

Histograma binomial



parâmetros para um experimento binomial

Média: $\mu = np$

Variância: $\sigma^2 = npq$

Desvio padrão: $\sigma = \sqrt{npq}$

parâmetros para um experimento binomial

Média: $\mu = np$

Variância: $\sigma^2 = npq$

Desvio padrão: $\sigma = \sqrt{npq}$

$$\mu = np = 5(0{,}25) = 1{,}25$$

$$\sigma^2 = npq = 5(0{,}25)(0{,}75) = 0{,}9375$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{npq} = \sqrt{0{,}9375} = 0{,}968$$

distribuição geométrica

distribuição geométrica é uma distribuição discreta de probabilidade da variável aleatória *x* com as seguintes condições:

- *A tentativa é repetida até que o sucesso ocorra.*
- *As sucessivas tentativas são independentes entre si.*
- *A probabilidade de sucesso é a mesma a cada tentativa*

A probabilidade de que o primeiro sucesso ocorra na tentativa número *x é*: $P(x) = (q)^{x-1}p$, onde $q = 1 - p$.

distribuição geométrica

Segundo uma pesquisa de mercado, a probabilidade de que cada pessoa que entra em determinada loja faça uma compra é de 0,30.

- A probabilidade de que a primeira compra seja feita pela primeira pessoa que entrar na loja é de 0,30. Ou seja: $P(1) = 0{,}30$.

- A probabilidade de que a primeira compra seja feita pela segunda pessoa que entrar na loja é de (0,70) (0,30).
Logo, $P(2) = (0{,}70)\,(0{,}30) = 0{,}21$.

- A probabilidade de que a primeira compra seja feita pela terceira pessoa que entrar na loja é de (0,70)(0,70)(0,30).
Logo, $P(3) = (0{,}70)\,(0{,}70)\,(0{,}30) = 0{,}147$.

A probabilidade de que a primeira compra seja feita pela pessoa número x é de

$P(x) = (0{,}70)^{x-4}(0{,}30)$

distribuição geométrica

Um fabricante de cereais colocou uma peça premiada nas embalagens de seu produto. A probabilidade de ganhar um prêmio é de um para quatro. Determine a probabilidade de que você:

a) *ganhe seu primeiro prêmio na quarta compra*;

$$P(4) = (0{,}75)^3 \cdot (0{,}25) = 0{,}1055$$

b) *ganhe seu primeiro prêmio na segunda ou terceira compra*;

$$P(2) = (0{,}75)^1(0{,}25) = 0{,}1875 \text{ e}$$
$$P(3) = (0{,}75)^2(0{,}25) = 0{,}1406$$
$$\text{Logo, } P(2 \text{ ou } 3) = 0{,}1875 + 0{,}1406 = 0{,}3281$$

c) *não ganhe nenhum prêmio nas quatro primeiras compras*.

$$1 - (P(1) + P(2) + P(3) + P(4))$$
$$1 - (\ 0{,}25 + 0{,}1875 + 0{,}1406 + 0{,}1055)$$
$$= 1 - 0{,}6836 = 0{,}3164$$

distribuição de Poisson

A distribuição de Poisson é uma distribuição discreta de probabilidade de uma variável aleatória *x com as* seguintes condições:

1. O experimento consiste em contar o número de vezes, *x*, que um evento ocorre num intervalo de tempo, área ou espaço.
2. A probabilidade de que o evento ocorra é a mesma em cada intervalo.
3. O número de ocorrências em um intervalo independe do número de ocorrências em outro.

A probabilidade de exatamente *x* ocorrências em um intervalo é

$$P(x) = \frac{m^x e^{-\mu}}{x!}$$

e é um número irracional aproximadamente igual a 2,71828.
μ é o número médio de ocorrências por intervalo.

distribuição de Poisson

Estima-se que, em todo o mundo, os tubarões matem dez pessoas por ano. Determine a probabilidade:

a) de que três pessoas sejam mortas por tubarões este ano

$$P(3) = \frac{10^3\ (2{,}71828)^{-10}}{3!} = 0{,}0076$$

b) de que duas ou três pessoas sejam mortas por tubarões este ano

$$P(2) = \frac{10^2\ (2{,}71828)^{-10}}{2!} = 0{,}0023$$

$P(3) = 0{,}0076$

$P(2 \text{ ou } 3) = 0{,}0023 + 0{,}0076 = 0{,}0099$

distribuição de probabilidade

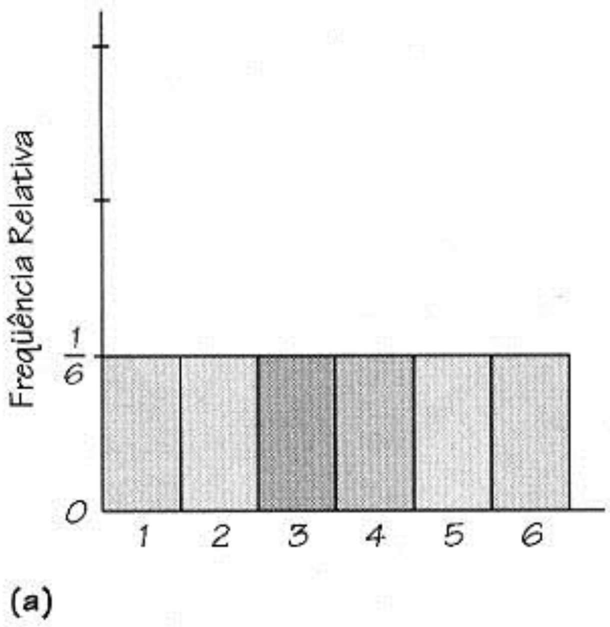
Contínuas

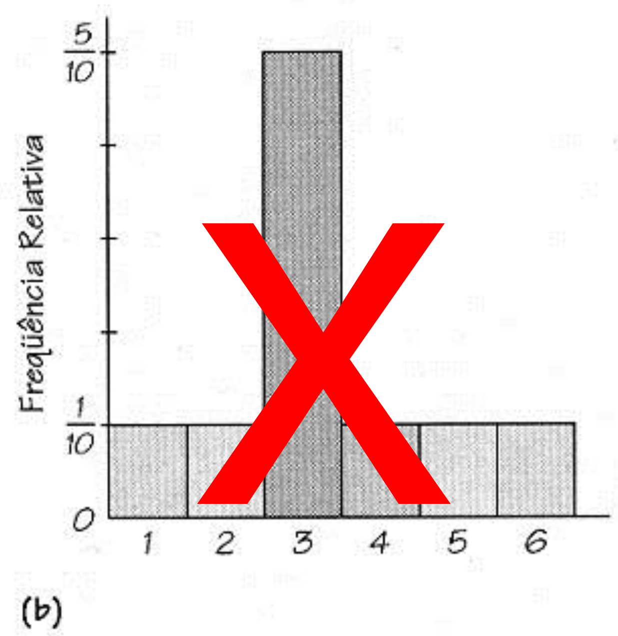
- uniforme
- Normal
- Exponencial

distribuição de probabilidade

◈ CONTÍNUAS

- Uniforme

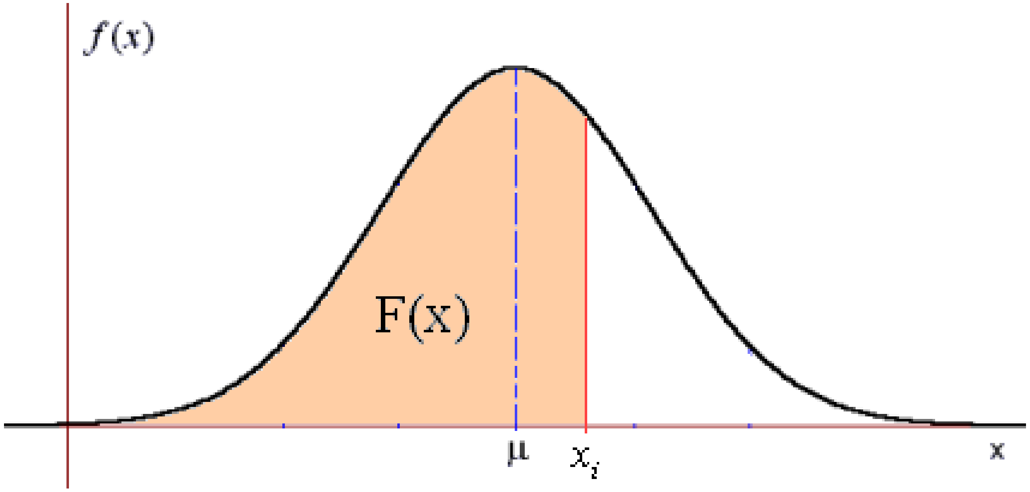




distribuição de probabilidade

◈ CONTÍNUAS

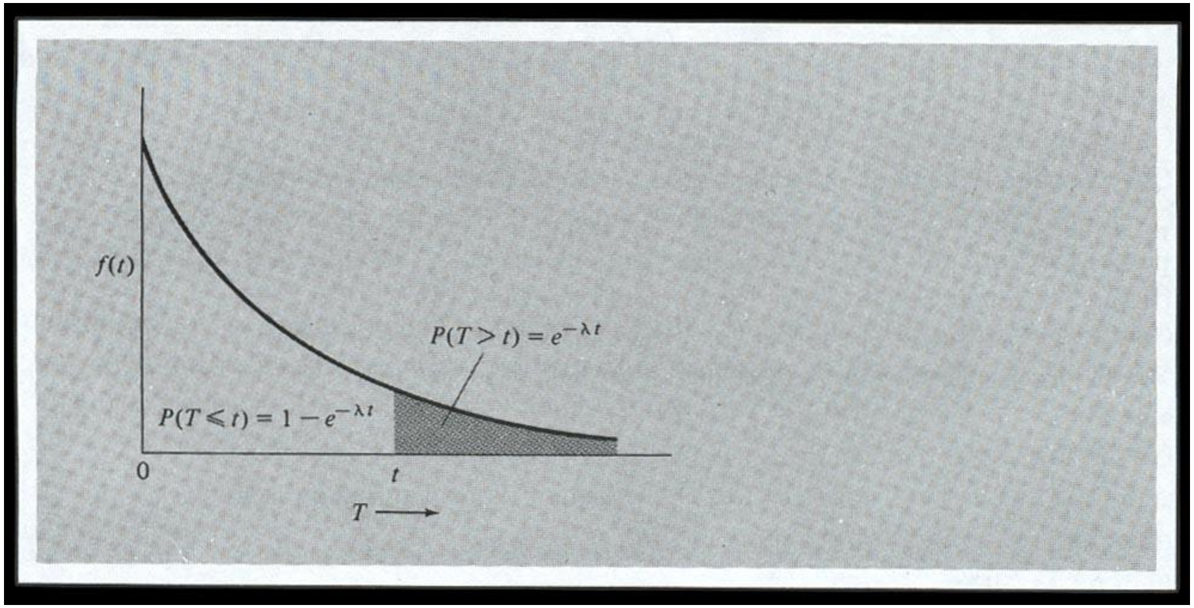
❑ Normal



distribuição de probabilidade

◈ CONTÍNUAS

❑ Exponencial



Teorema de Chebychev

A proporção (ou fração) de qualquer
conjunto de dados a menos
de k desvios-padrão a contar da média
é sempre
ao menos $1 - 1/K^2$

K é um número positivo maior do que 1

Teorema de Chebychev

Consequências para k = 2 e k = 3

Ao menos ¾ (75%) de todos
os valores estão
no intervalo entre –2 e +2 desvios-padrão

Teorema de Chebychev

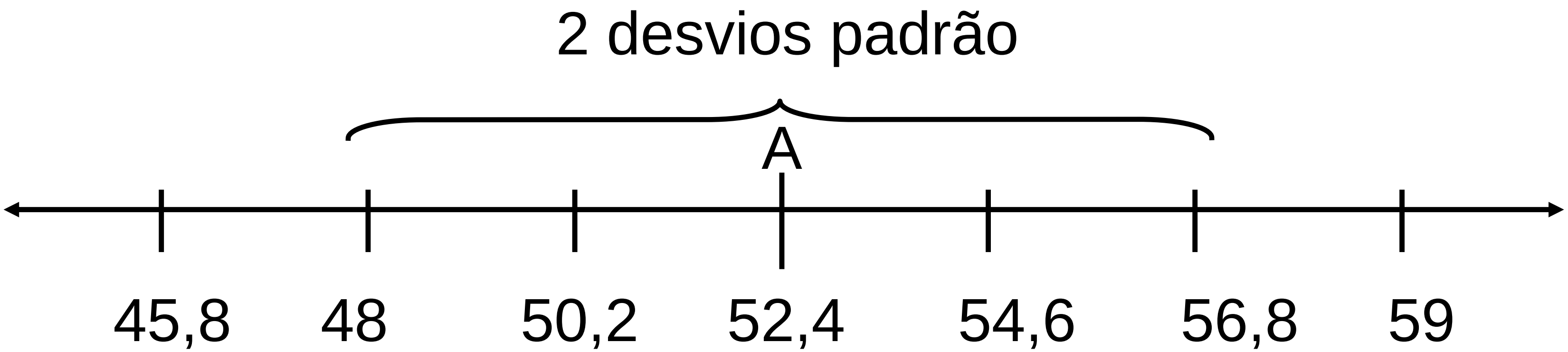
Consequências para k = 2 e k = 3

Ao menos 8/9 (89%) de todos
os valores estão
no intervalo entre –3 e +3 desvios-padrão

Teorema de Chebychev

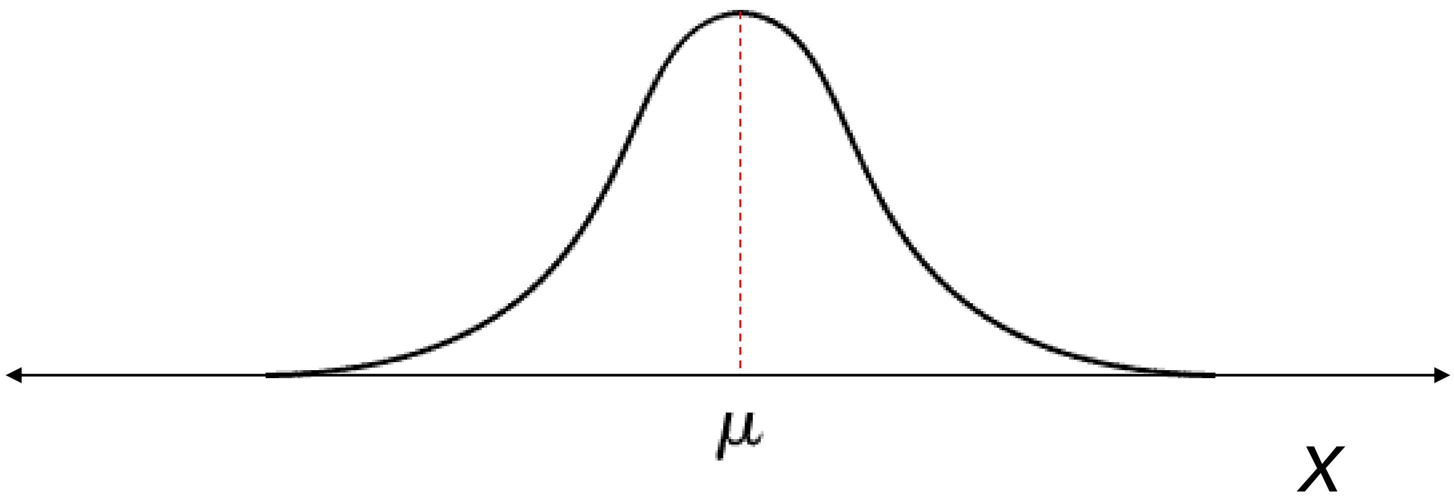
A média feminina nos 400 metros rasos é de 52,4 s.
O desvio padrão é de 2,2 segundos.

Teorema de Chebychev



pelo menos 75% dos tempos femininos nos 400 metros rasos estarão entre 48 e 56,8 segundos.

curva normal

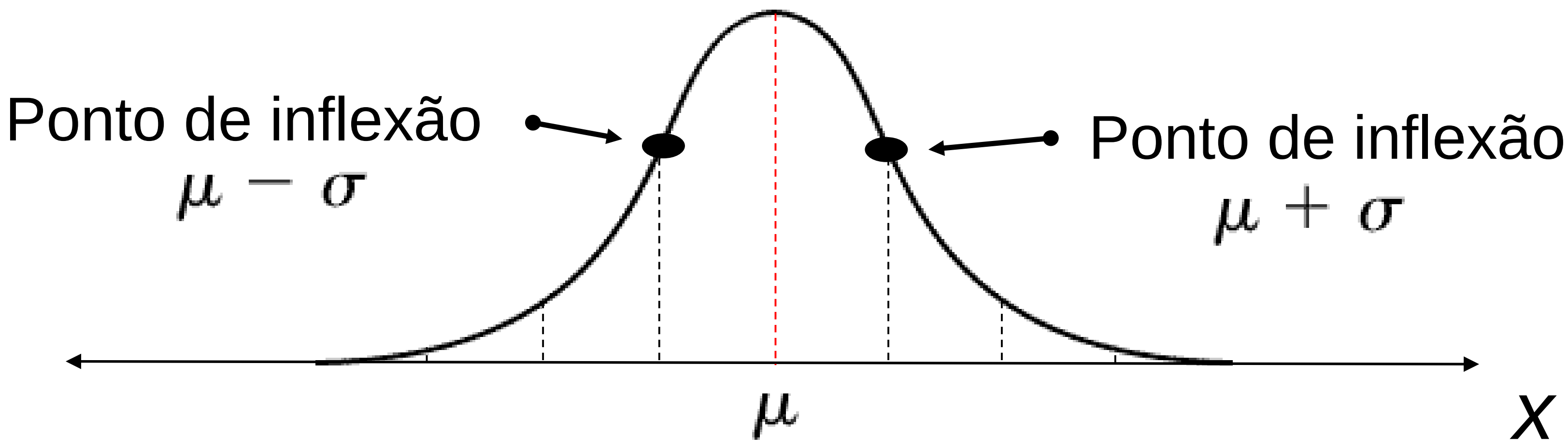


- Suas média, mediana e moda são iguais.

- Tem forma de sino e é simétrica em torno da média.

- A área total sob a curva é de 100%.

curva normal

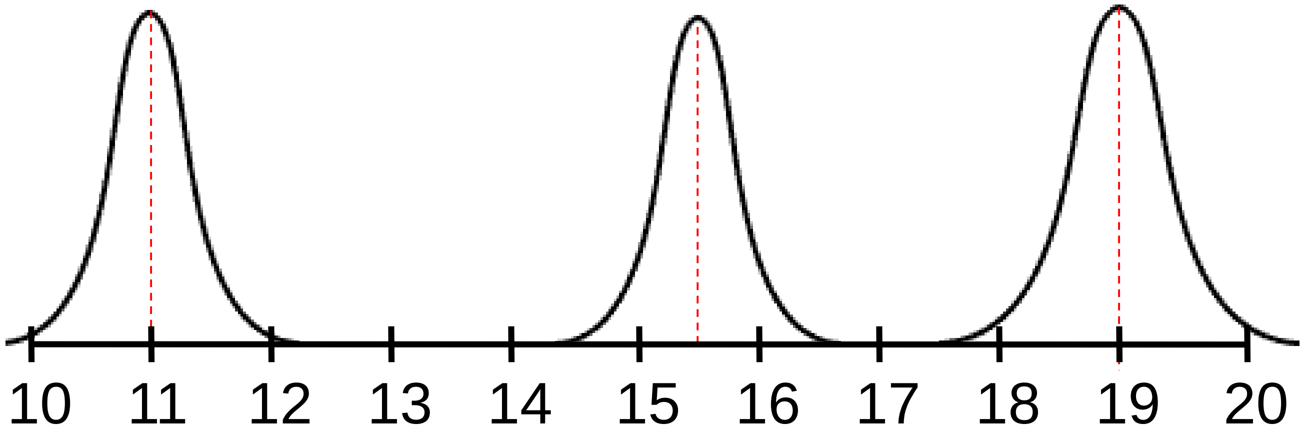


- À medida que a curva se afasta da média, aproxima-se cada vez mais do eixo *x*, mas nunca o toca.

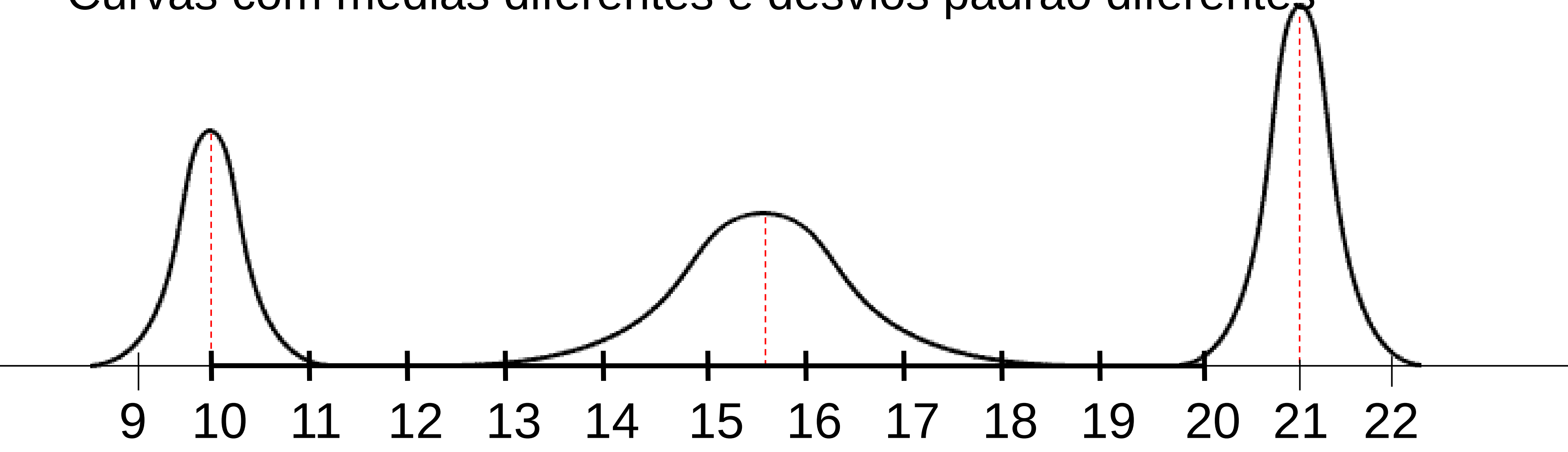
- Os pontos em que a curvatura muda são chamados pontos de inflexão. O gráfico curva-se para baixo entre os pontos de inflexão e, para cima, à esquerda e à direita deles.

Médias e desvios-padrão

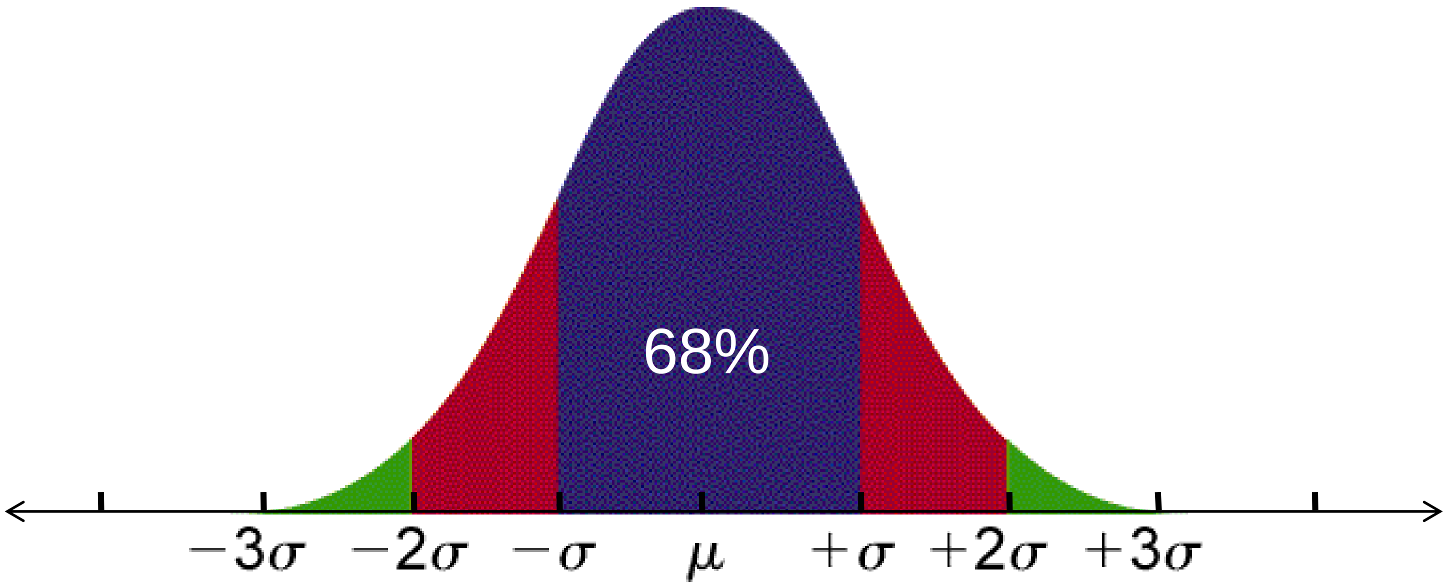
Curvas com médias diferentes e o mesmo desvio padrão



Curvas com médias diferentes e desvios padrão diferentes

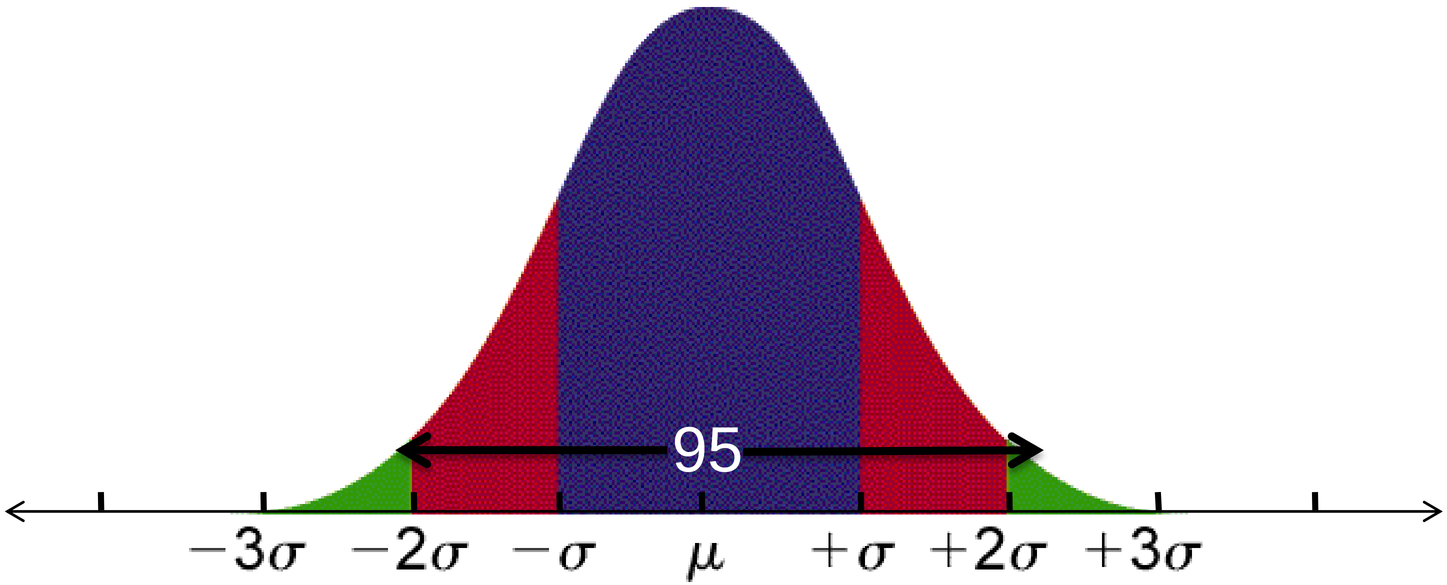


Regra Empírica



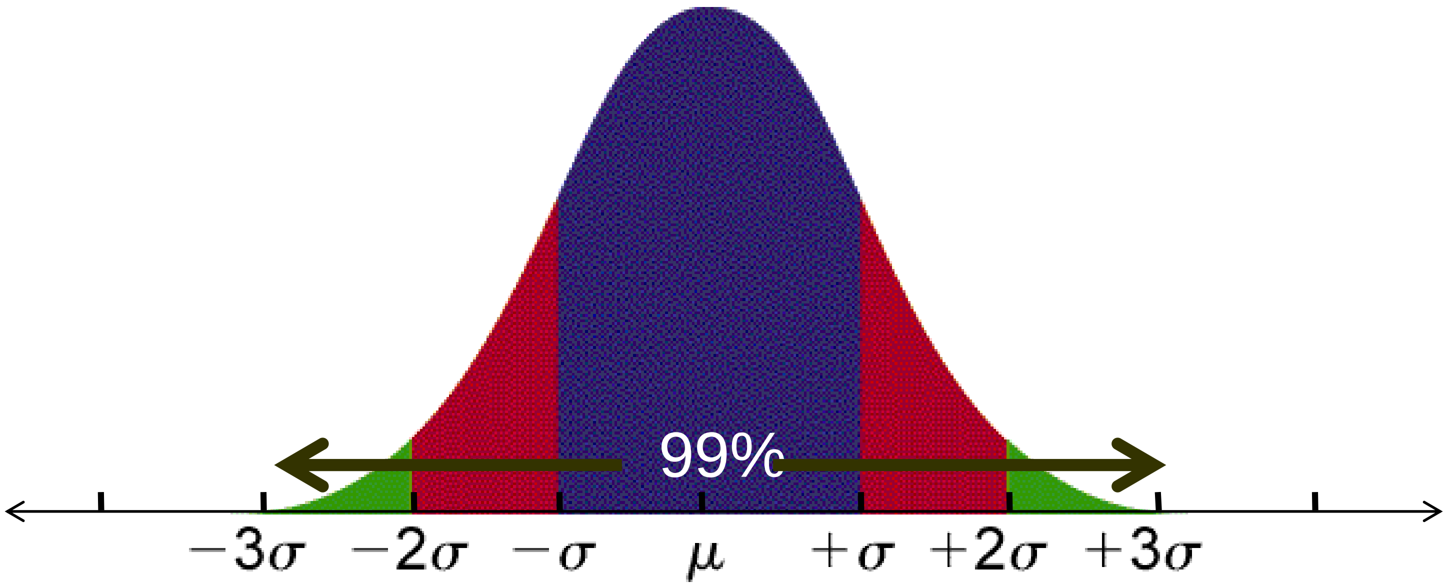
a área entre (– 1dp) e (+1dp) é de cerca de 68%

Regra Empírica



a área entre (– 2dp) e (+2dp) é de cerca de 95%

Regra Empírica

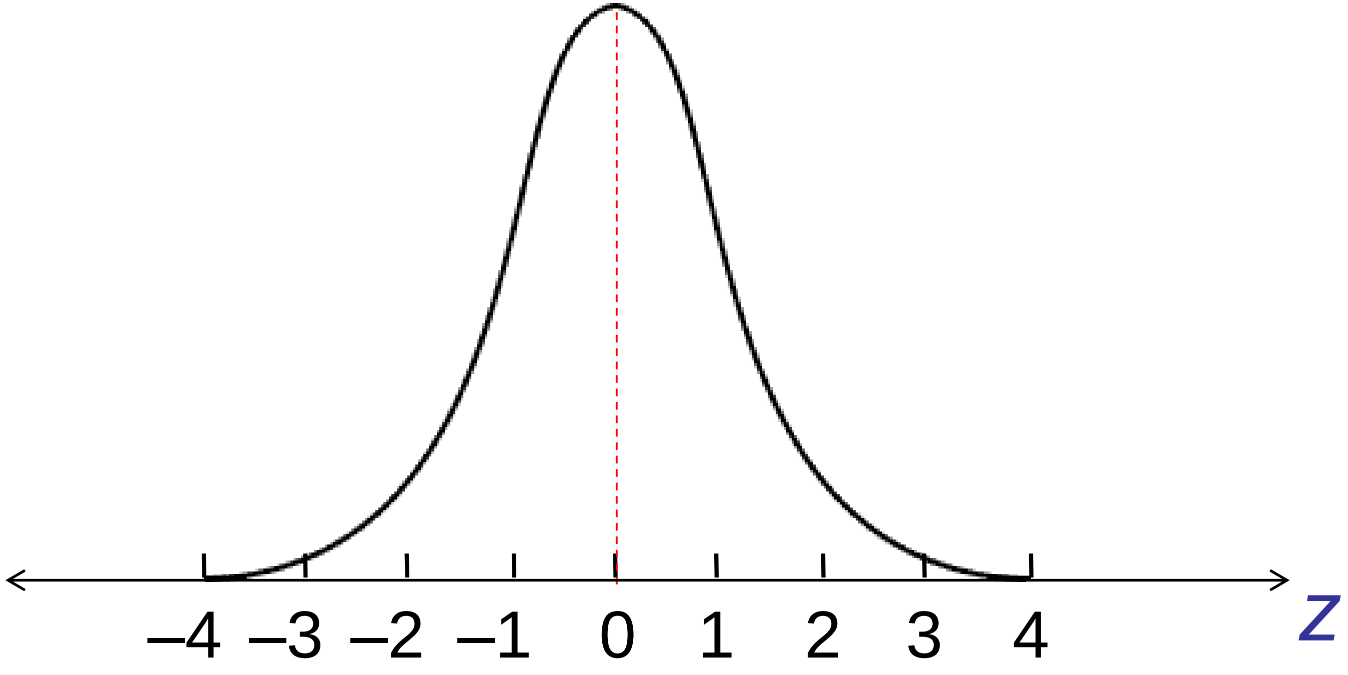


a área entre (– 3dp) e (+3dp) é de cerca de 99,7%

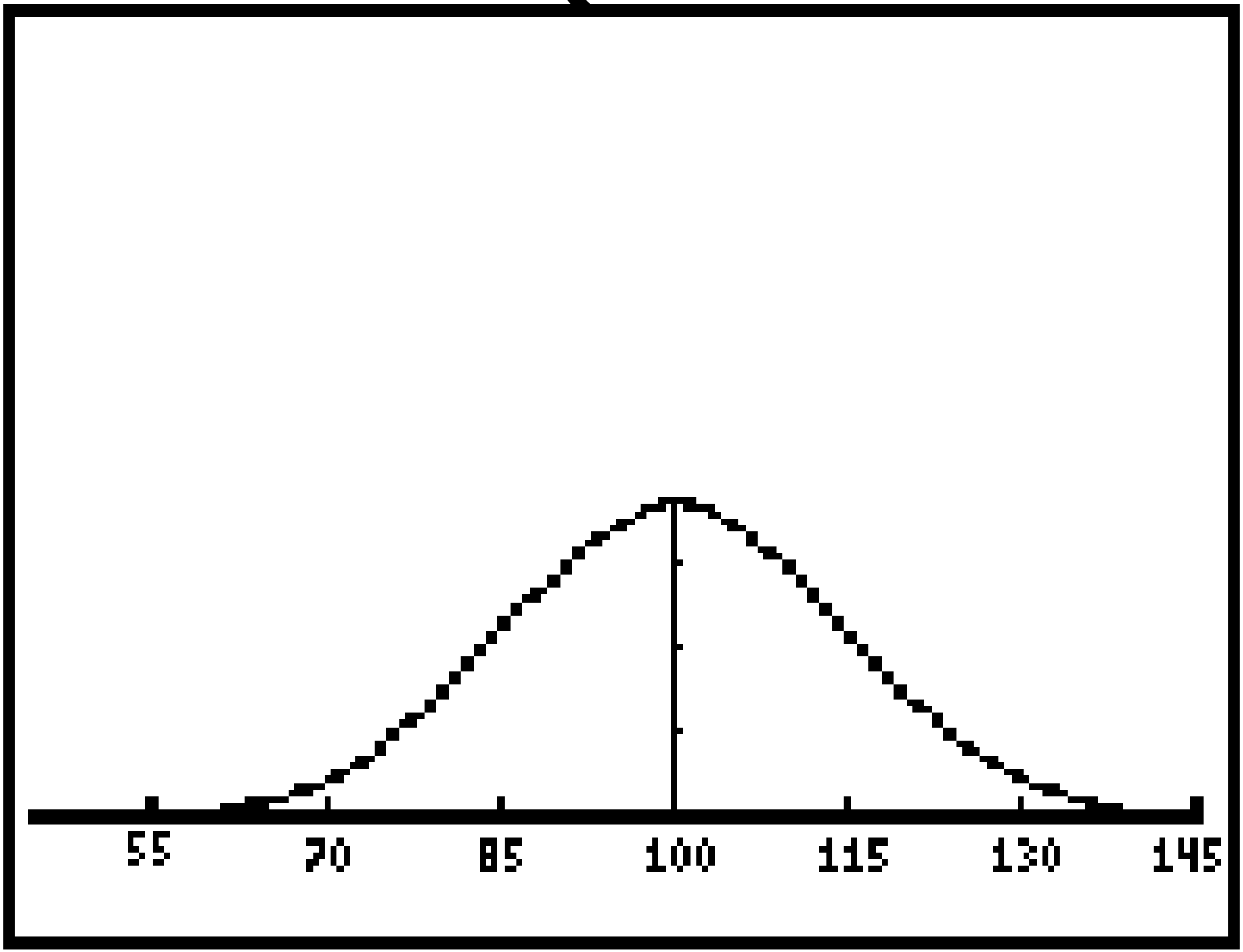
A distribuição normal padrão

A distribuição normal padrão tem média 0 e desvio padrão de 1.

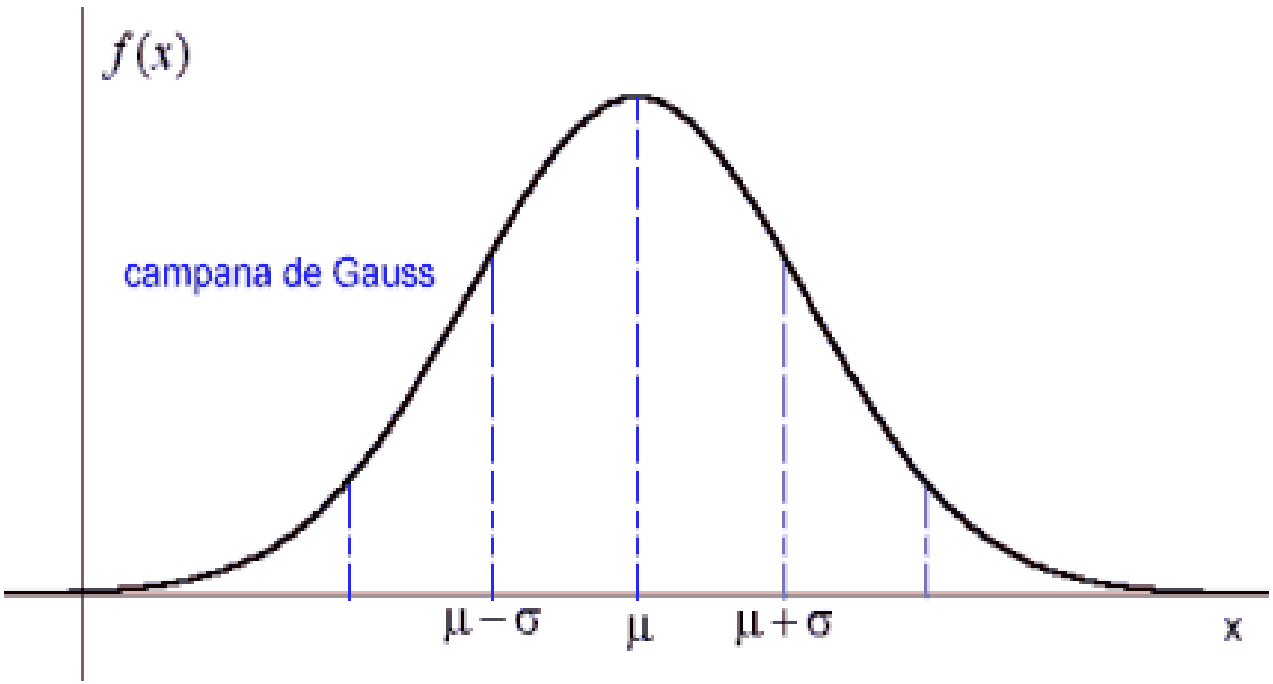
Se usar escores *z*, você pode transformar qualquer distribuição normal numa distribuição normal padrão.



QI



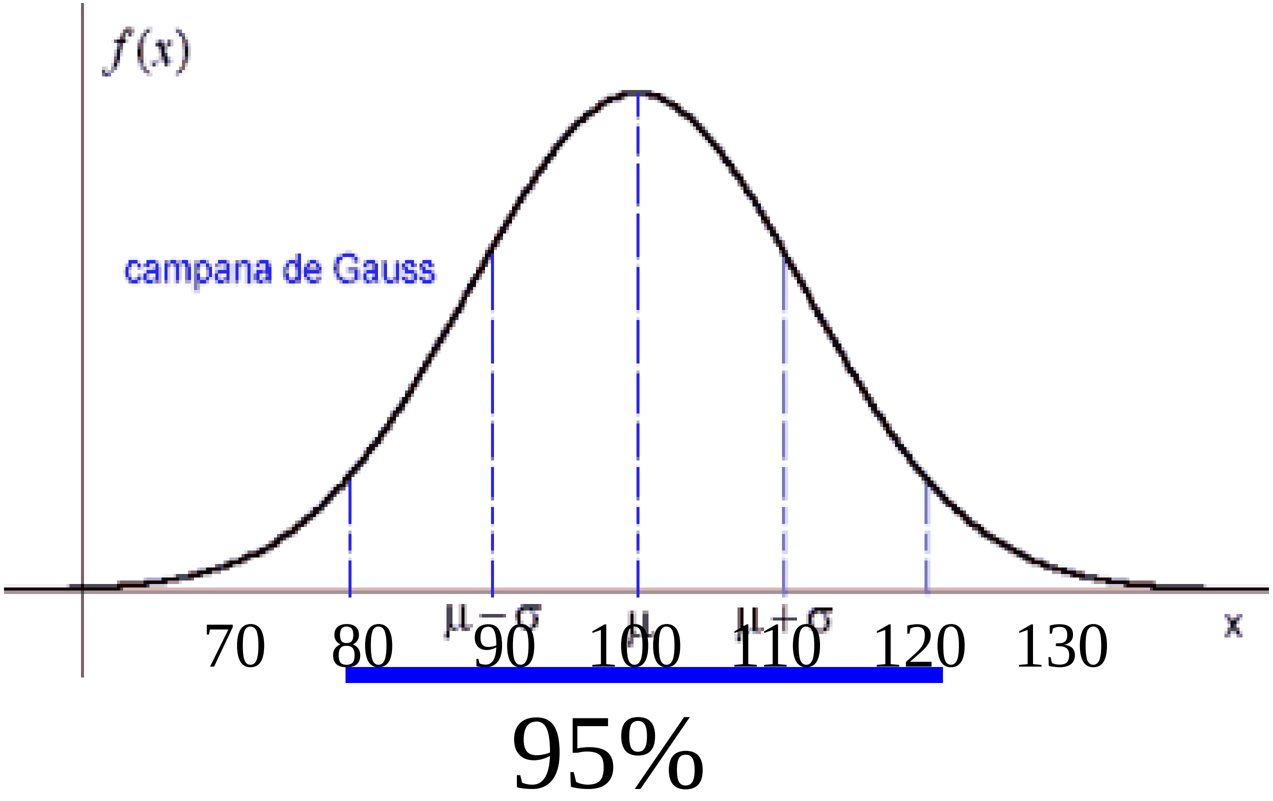
QI masculino:dp = 10



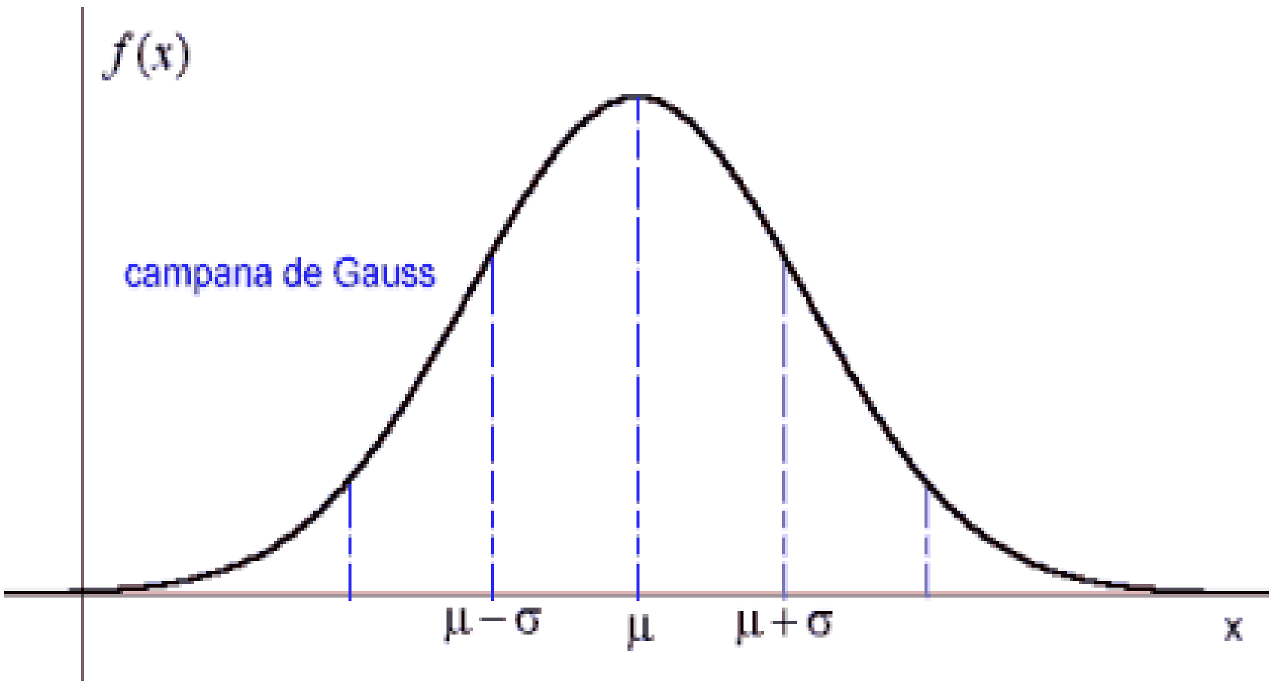
70 80 90 100 110 120 130

68%

QI masculino:dp = 10



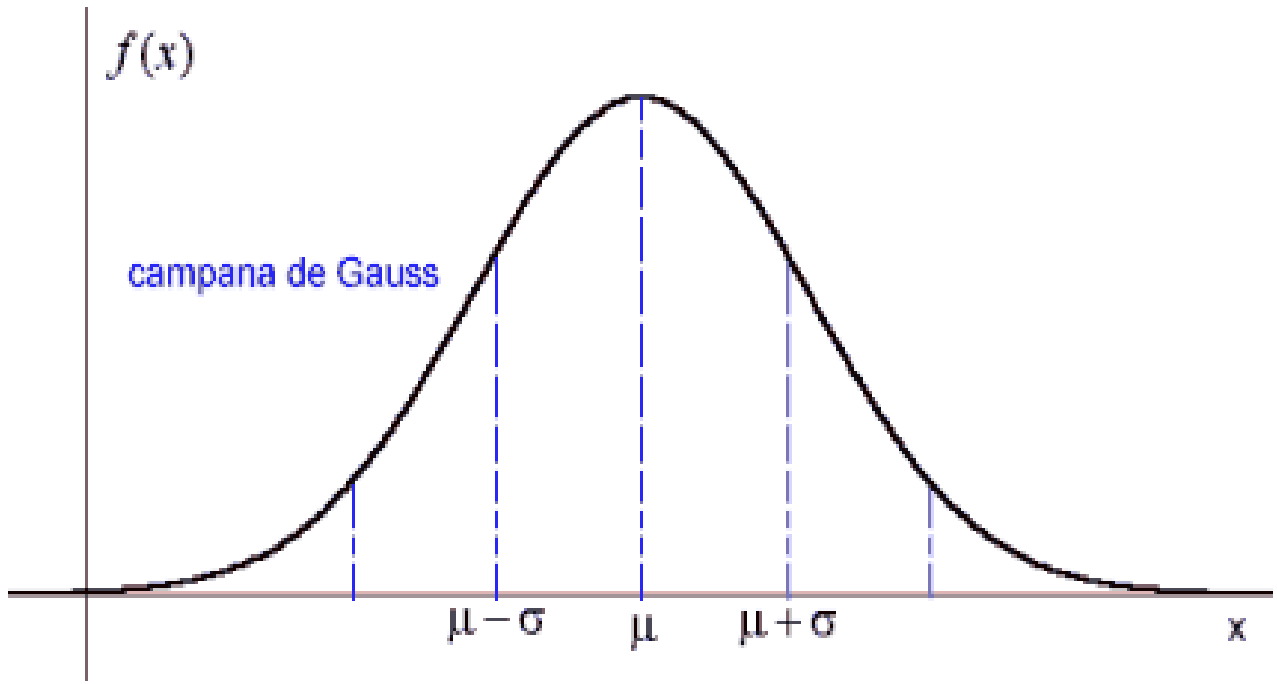
QI masculino:dp = 10



70 80 90 100 110 120 130

99%

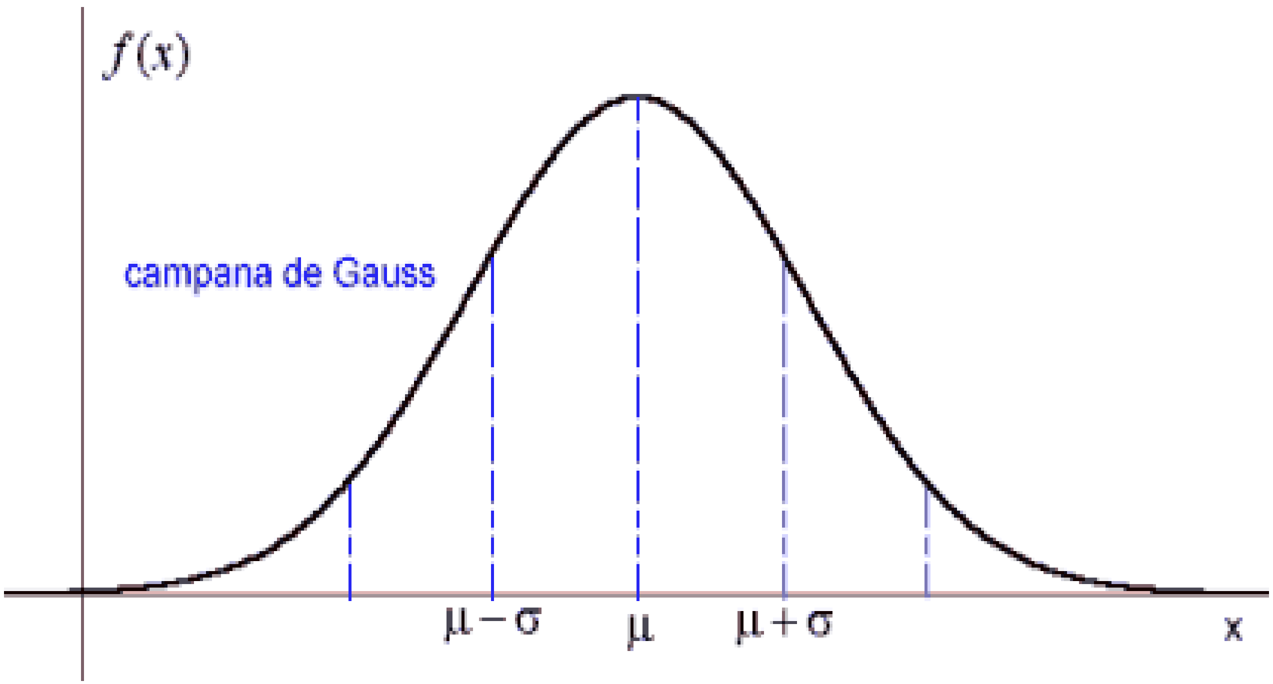
QI FEMININO:dp = 5



75 90 95 100 105 110 115

68%

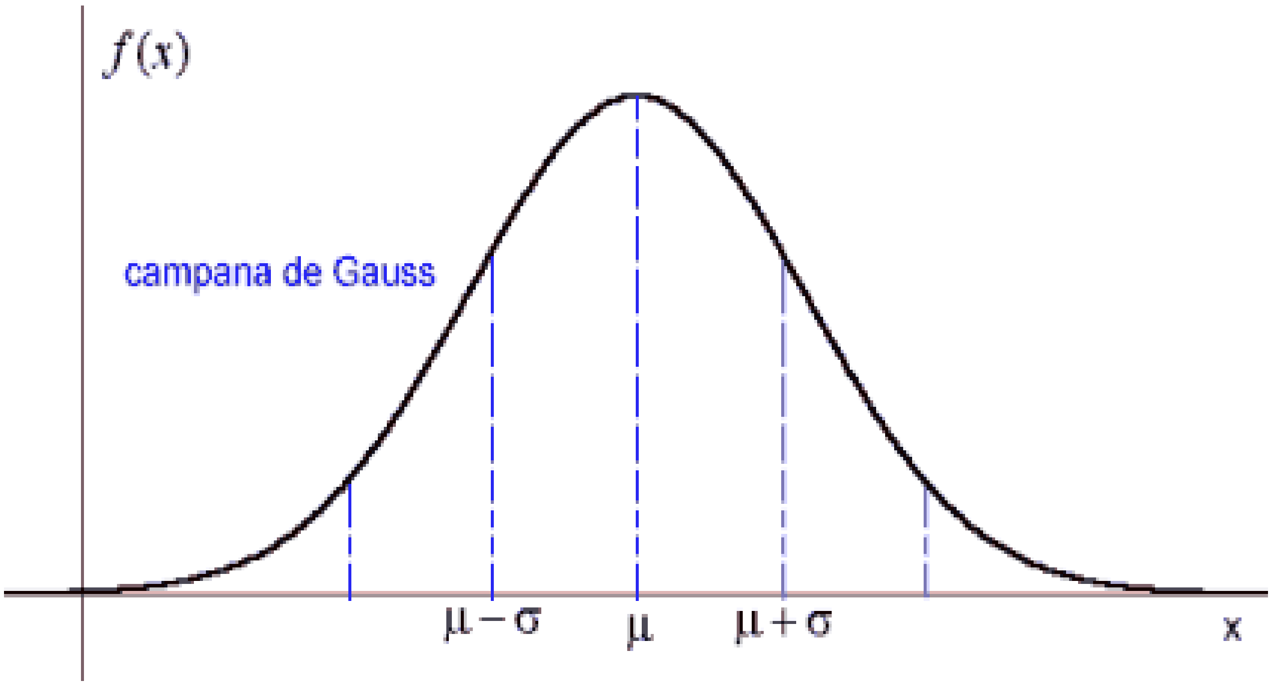
QI FEMININO:dp = 5



75 90 95 100 105 110 115

95%

QI FEMININO:dp = 5



75 90 95 100 105 110 115

99%