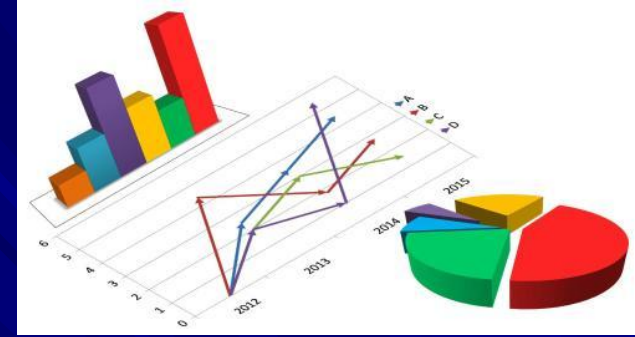




# Análise de Dados

CURSO DE LOGISTICA INTEGRADA DE PRODUÇÃO

*Helio Rosetti Jr.*

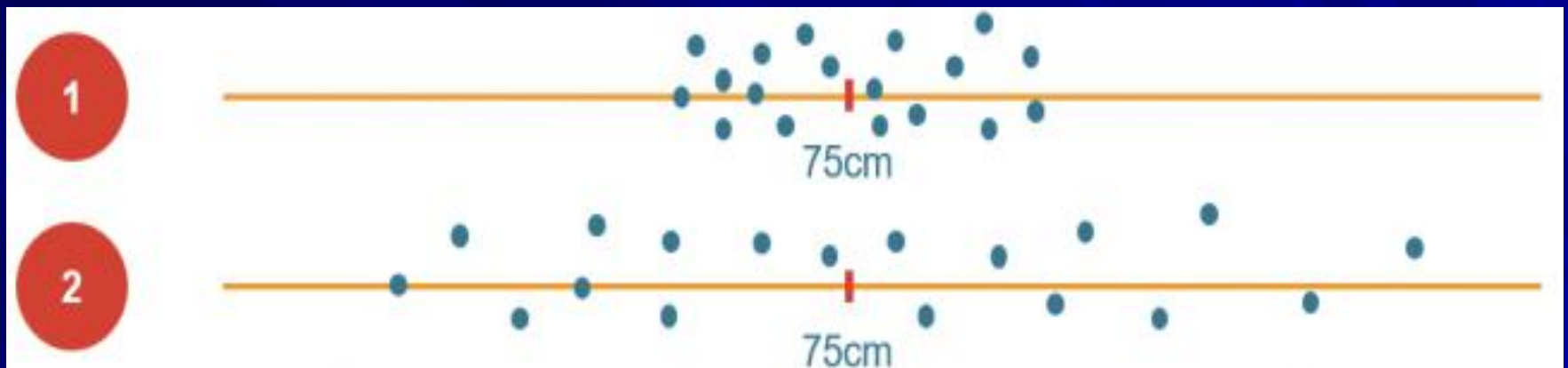
# Medidas de Variabilidade ou Dispersão

Chamamos de Variabilidade a diversificação dos valores de uma variável em torno de um valor de tendência central tomado como ponto de comparação

## **Alguns tipos de Medidas:**

Amplitude Total, Variância, Desvio Padrão e Coeficiente de variação

# Variabilidade



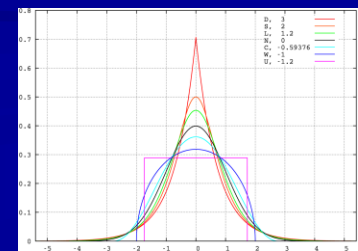
## Um pouco de história

O **desvio padrão** é uma grandeza que remete ao século XIX, no contexto do desenvolvimento da estatística, no Reino Unido.

Enquanto o conceito de *medida de dispersão* foi criado por Abraham de Moivre e usado em seu livro *The Doctrine of Chances* em 1718, o termo **desvio padrão** foi pontualmente usado pela primeira vez por Karl Pearson em 1894, em substituição a termos anteriores como erro médio, utilizado por Carl Friedrich Gauss.

Em 1908, William Gosset (mais conhecido sob o pseudônimo *Student*) definiu o desvio padrão empírico de uma amostra e mostrou que a distinção entre o desvio padrão amostral e o desvio padrão populacional é importante.

Somente em 1918, Ronald Aylmer Fisher definiu a noção da variância no texto *The Correlation between Relatives on the Supposition of Mendelian Inheritance*.



A **Amplitude Total** é a diferença entre o maior e o menor valor observado.

$$AT = x(\text{máx.}) - x(\text{mín.})$$

Onde:

AT = amplitude total

X(máx.) – valor máximo da amostra

X(mín.) – valor mínimo da amostra

### **Variância populacional**

A variância de uma população  $\{x_1, \dots, x_N\}$  de N elementos é a medida de dispersão definida como a média do quadrado dos desvios dos elementos em relação à média populacional  $\mu$ . Ou seja, a variância populacional é dada por:

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N}$$

## Desvio padrão populacional

O desvio padrão populacional de um conjunto de dados é igual à raiz quadrada da variância populacional. Desta forma, o desvio padrão populacional é dado por:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^N \frac{(x_i - \mu)^2}{N}}$$

O **Coeficiente de Variação** é usado para analisar a dispersão em termos relativos a seu valor médio quando duas ou mais séries de valores apresentam unidades de medida diferentes. Dessa forma, podemos dizer que o coeficiente de variação é uma forma de expressar a variabilidade dos dados excluindo a influência da ordem de grandeza da variável.

O cálculo do coeficiente de variação é feito através da fórmula:

$$CV = \frac{s}{\bar{X}} \cdot 100$$

Onde,

$s \rightarrow$  é o desvio padrão

$\bar{X} \rightarrow$  é a média dos dados

$CV \rightarrow$  é o coeficiente de variação

# Análise por meio do Coeficiente de Variação %

Exemplo:

|                  | Média | D. Padrão | CV%   |
|------------------|-------|-----------|-------|
| <b>Estaturas</b> | 175cm | 5 cm      | 2,85% |
| <b>Pesos</b>     | 68 Kg | 2Kg       | 2,94% |

\*Maior grau de dispersão



# Variabilidade e Risco

Um risco faz referência à iminência ou à proximidade de um eventual dano. O conceito está associado à probabilidade de um dano se concretizar.

A percepção de risco de uma classe de ativos é diretamente proporcional à variabilidade de seus retornos.

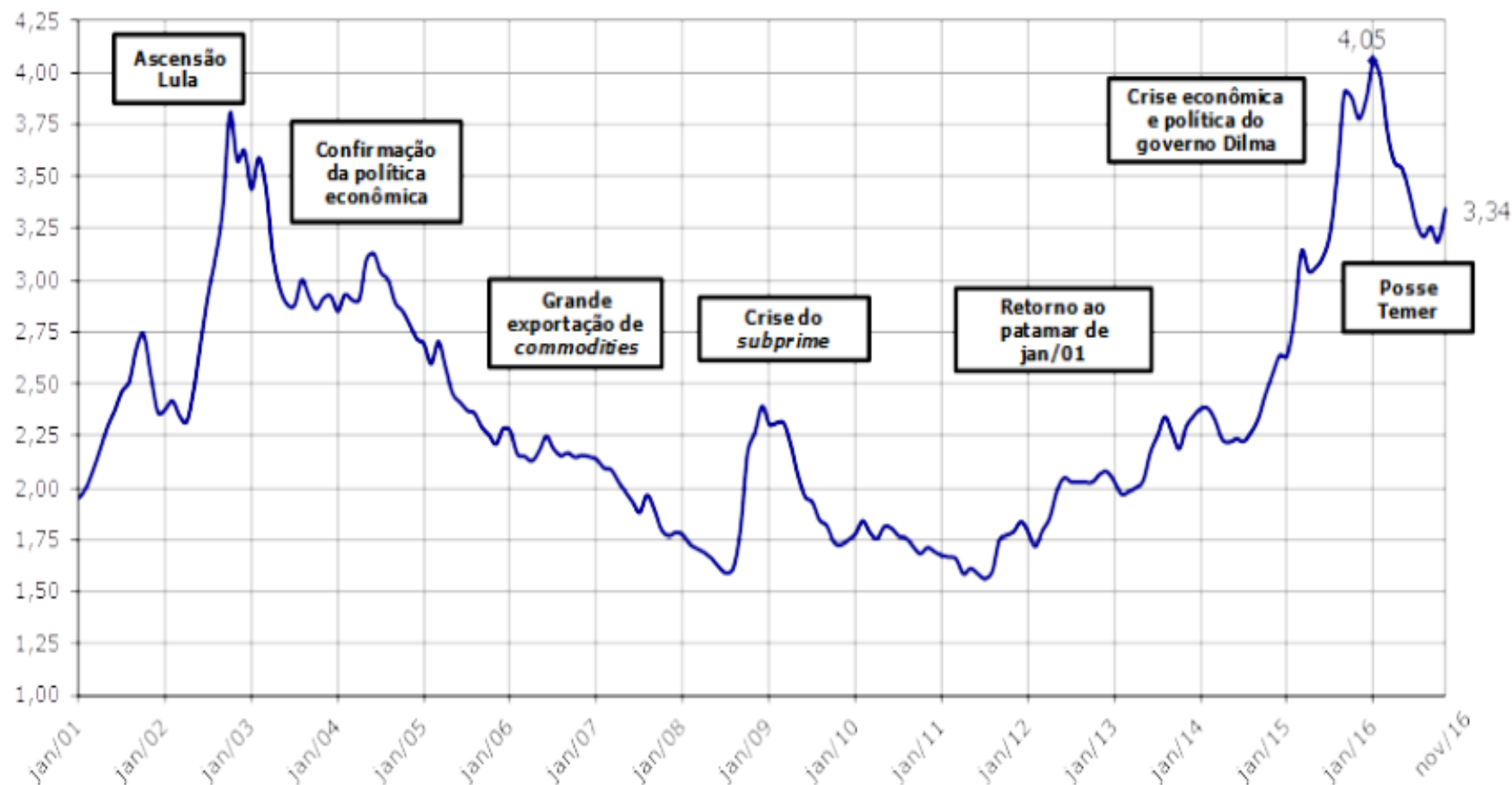
Como resultado, o prêmio de risco que os investidores exigem para investir em ativos, como ações e commodities, é maior do que o prêmio de risco para ativos como contas do Tesouro, que têm uma variabilidade de retorno muito menor.





## Taxa de câmbio R\$/US\$

comercial - venda (média mensal)  
(jan/2001 - nov/2016)



Elaboração: ProFit | Fonte: Banco Central - Sistema Gerenciador de Séries Temporais (SGS)

# Probabilidade



Probabilidade é o estudo das chances de obtenção de cada resultado de um experimento aleatório.

A essas chances são atribuídos os números reais do intervalo entre 0 e 1.

Um **experimento aleatório** pode ser repetido inúmeras vezes e nas mesmas condições e, mesmo assim, apresenta resultados diferentes.

Cada um desses resultados possíveis é chamado de ponto amostral.



## Exemplos de experimentos aleatórios:

### 1) Cara ou coroa

Lançar uma moeda e observar se a face voltada para cima é cara ou coroa é um exemplo de **experimento aleatório**. Se a moeda não for viciada e for lançada sempre nas mesmas condições, poderemos ter como resultado tanto cara quanto coroa.

### 2) Lançamento de um dado

Lançar um dado e observar qual é o número da face superior também é um **experimento aleatório**. Esse número pode ser 1, 2, 3, 4, 5 ou 6 e cada um desses resultados apresenta a mesma **chance** de ocorrer. Em cada lançamento, o resultado pode ser igual ao anterior ou diferente dele.

# Espaço Amostral

O **espaço amostral** ( $\Omega$ ) é o conjunto formado por todos os resultados possíveis de um **experimento aleatório**. Em outras palavras, é o conjunto formado por todos os **pontos amostrais** de um experimento.

Exemplos:

- 1) O **espaço amostral** do experimento “cara ou coroa” é o conjunto  $S = \{\text{Cara}, \text{Coroa}\}$ . Os **pontos amostrais** desse experimento são os mesmos elementos desse conjunto.
- 2) O **espaço amostral** do experimento “lançamento de um dado” é o conjunto  $S = \{1, 2, 3, 4, 5, 6\}$ . Os **pontos amostrais** desse experimento são 1, 2, 3, 4, 5 e 6.

O **espaço amostral** também é chamado de **Universo** e pode ser representado pelas outras notações usadas nos **conjuntos**. Além disso, todas as **operações** entre conjuntos valem também para espaços amostrais.

O **número de elementos** do espaço amostral, número de pontos amostrais do **espaço amostral** ou número de casos possíveis em um espaço amostral é representado da seguinte maneira:  $n(\Omega)$ .

# Evento

Um evento é qualquer subconjunto de um espaço amostral. Ele pode conter nenhum elemento (conjunto vazio) ou todos os elementos de um espaço amostral.

O número de elementos do evento é representado da seguinte maneira:  $n(E)$ , sendo  $E$  o evento em questão.

## Exemplos de eventos:

1) Sair cara em um lançamento de uma moeda

O **evento** é *sair cara* e possui um único elemento. A representação dos eventos também é feita com notações de conjuntos:

$$E = \{\text{cara}\}$$

O seu **número de elementos** é  $n(E) = 1$ .

2) Sair um número par no lançamento de um dado.

O **evento** é *sair um número par*:

$$E = \{2, 4, 6\}$$

O seu **número de elementos** é  $n(E) = 3$ .

# Cálculo da probabilidade

Seja E um evento qualquer no **espaço amostral**  $\Omega$ . A **probabilidade** do evento A ocorrer é a razão entre o número de resultados favoráveis e o número de resultados possíveis.

Em outras palavras, é o número de elementos do **evento** dividido pelo número de elementos do espaço amostral a que ele pertence.

$$P(E) = \frac{n(E)}{n(\Omega)}$$

## Observações:

O número de elementos do **evento** sempre é menor ou igual ao número de elementos do **espaço amostral** e maior ou igual a zero. Por isso, o resultado dessa divisão sempre está no intervalo  $0 \leq P(A) \leq 1$ ;

Quando é necessário usar porcentagem, devemos multiplicar o resultado dessa divisão por 100.







→ **Qual é a probabilidade de, no lançamento de duas moedas, obtermos resultados iguais?**

**Solução:**

Representando cara por C e coroa por K, teremos os seguintes resultados possíveis:

(C, K); (C, C); (K, C); (K, K)

O evento *obter resultados iguais* possui os seguintes casos favoráveis:

(C, C); (K, K)

Há quatro casos possíveis (número de elementos do **espaço amostral**) e dois casos favoráveis (número de elementos do **evento**), logo:

$$P(E) = \frac{n(E)}{n(\Omega)}$$

$$P(E) = \frac{2}{4}$$

$$P(E) = 0,5 = 50\%$$





Um bom exemplo da aplicação do estudo de probabilidade e risco é no mercado de seguros



**Seguro** (do latim "securus") a todo contrato pelo qual uma das partes, segurador, se obriga a indenizar a outra, segurado, em caso da ocorrência de determinado sinistro, em troca do recebimento de um prêmio de seguro.

## Um pouco de história do seguro



**O seguro nasceu da necessidade do homem em controlar o risco. Existem indícios que já na Babilônia, 23 séculos antes de Cristo, caravanas de camaleiros que cruzavam o deserto mutualizavam entre si os prejuízos com morte de animais.**

Na China antiga e no Império Romano também havia seguros rudimentares, através de associações que visavam ressarcir membros que tivessem algum tipo de prejuízo.

Os comerciantes chineses que se aventuravam a transportar as suas mercadorias instalando-as em débeis embarcações que desciam pelas correntezas dos grandes rios continentais e que, para evitar a ruína de alguns deles, distribuíam-nas de modo a que cada barco contivesse uma parte de cada comerciante, estavam aplicando o princípio básico do seguro. Se uma embarcação naufragava, a perda correspondia a uma pequena parte dos bens de cada um.

O mesmo se pode dizer dos comerciantes árabes, que para cruzar os desertos e lugares inóspitos distribuíam os seus bens entre várias caravanas e, dentro da mesma caravana, entre diversos camelos.

Com o Renascimento e a expansão marítima da época Mercantilismo a cobertura aos riscos ganhou nova importância.

Tornaram-se comuns operações chamadas de Contrato de Dinheiro e Risco Marítimo que consistia num empréstimo dado a um navegador, e que previa uma cobrança maior no caso de sucesso da viagem e o perdão da dívida se a embarcação e a carga fossem perdidas.

Foi em virtude dos seguros marítimos que se desenvolveu a gestão de risco na maior parte do mundo.

Essas formas pitorescas foram de extrema importância para garantir a segurança das mercadorias que circulavam por vias terrestres e marítimas.

Nessa época o seguro ainda inspirava dúvidas com relação à integridade das “seguradoras” – que na verdade eram pessoas que assumiam os riscos.





Mas, o seguro foi criando força e conquistando credibilidade, e foi em Gênova, por volta de 1347, que o primeiro contrato de seguros foi escrito.

Nele continha inúmeras cláusulas que garantiam ou isentavam os seguradores de pagarem as indenizações. As primeiras apólices são datadas de 11/07/1385 (Pisa/ Itália) e 10/07/1397 (Florença/ Itália). As apólices tornavam-se comuns no final do século XIV.

No século XVII, o mercado securitário se expandiu e ganhou novos produtos de cobertura terrestre, especialmente em decorrência do Grande Incêndio de Londres de 1666, que destruiu cerca de 25% da cidade.

Com a Revolução Industrial o seguro acabou se tornando um item praticamente obrigatório em todas as áreas da atividade humana, afinal, os avanços tecnológicos, as atividades de alto risco e os novos meios de transportes podem causar prejuízos de proporções incalculáveis.



# Distribuição de Probabilidade

## Variável Aleatória

Suponhamos um espaço amostral  $S$  e que a cada ponto amostral seja atribuído um número. Fica, então, definida uma função chamada **variável aleatória**, indicada por uma letra maiúscula, sendo seus valores indicados por letras minúsculas.

**Exemplo:**  $X$  é o número de caras no lançamento simultâneo de duas moedas

| PONTO AMOSTRAL | $X$ |
|----------------|-----|
| (Ca, Ca)       | 2   |
| (Ca, Co)       | 1   |
| (Co, Ca)       | 1   |
| (Co, Co)       | 0   |



# Distribuição de Probabilidade

Consideremos a distribuição de frequências relativa ao número de acidentes diários em um estacionamento:

| NÚMERO DE ACIDENTES | FREQÜÊNCIAS   |
|---------------------|---------------|
| 0                   | 22            |
| 1                   | 5             |
| 2                   | 2             |
| 3                   | 1             |
|                     | $\Sigma = 30$ |

# Distribuição de Probabilidade

Em um dia, a probabilidade de:

— não ocorrer acidente é:

$$p = \frac{22}{30} = 0,73$$

— ocorrer um acidente é:

$$p = \frac{5}{30} = 0,17$$

— ocorrerem dois acidentes é:

$$p = \frac{2}{30} = 0,07$$

— ocorrerem três acidentes é:

$$p = \frac{1}{30} = 0,03$$

Podemos então escrever:

| NÚMERO DE ACIDENTES | PROBABILIDADES  |
|---------------------|-----------------|
| 0                   | 0,73            |
| 1                   | 0,17            |
| 2                   | 0,07            |
| 3                   | 0,03            |
|                     | $\Sigma = 1,00$ |

Esta tabela é denominada uma  
**Distribuição de Probabilidade**

Seja  $X$  uma variável aleatória que pode assumir os valores  $x_1, x_2, x_3, \dots, x_n$ . A cada valor  $x_i$  correspondem pontos do espaço amostral. Associamos, então, a cada valor  $x_i$  a probabilidade  $p_i$  de ocorrência de tais pontos no espaço amostral.

Assim, temos:

$$\sum p_i = 1$$

Os valores  $x_1, x_2, \dots, x_n$  e seus correspondentes  $p_1, p_2, \dots, p_n$  definem uma **distribuição de probabilidade**.

## Voltando ao exemplo do lançamento de duas moedas

| PONTO AMOSTRAL | X | P(X)                   |
|----------------|---|------------------------|
| (Ca, Ca)       | 2 | $1/2 \times 1/2 = 1/4$ |
| (Ca, Co)       | 1 | $1/2 \times 1/2 = 1/4$ |
| (Co, Ca)       | 1 | $1/2 \times 1/2 = 1/4$ |
| (Co, Co)       | 0 | $1/2 \times 1/2 = 1/4$ |

$\left. \begin{array}{l} 1/4 \\ 1/4 \\ 1/4 \end{array} \right\} 1/4 + 1/4 = 2/4$

| NÚMERO DE CARAS (X) | P(X)         |
|---------------------|--------------|
| 2                   | 1/4          |
| 1                   | 2/4          |
| 0                   | 1/4          |
|                     | $\Sigma = 1$ |



# Exemplo: lançamento de 3 moedas

| Espaço amostral                                                                     | X= número de caras | Probabilidade                                                   |
|-------------------------------------------------------------------------------------|--------------------|-----------------------------------------------------------------|
|    | → 0                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |
|    | → 1                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |
|    | → 1                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |
|    | → 1                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |
|    | → 2                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |
|    | → 2                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |
|   | → 2                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |
|  | → 3                | $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ |

- ✓ A probabilidade de sucesso em cada tentativa é constante, "p";
- ✓ Cada tentativa é independente (não influencia o resultado) da outra

| X | p(x)          |
|---|---------------|
| 0 | $\frac{1}{8}$ |
| 1 | $\frac{3}{8}$ |
| 2 | $\frac{3}{8}$ |
| 3 | $\frac{1}{8}$ |



Ao definir a distribuição de probabilidade, estabelecemos uma correspondência unívoca entre os valores da variável aleatória **X** e os valores da variável **P**. Esta correspondência define uma **função**; os valores  $x_i$  ( $i = 1, 2, \dots, n$ ) formam o **domínio da função** e os valores  $p_i$  ( $i = 1, 2, 3, \dots, n$ ), o seu **conjunto imagem**.

Essa função, assim definida, é denominada **função probabilidade** e representada por:

$$f(x) = P(X = x_i)$$

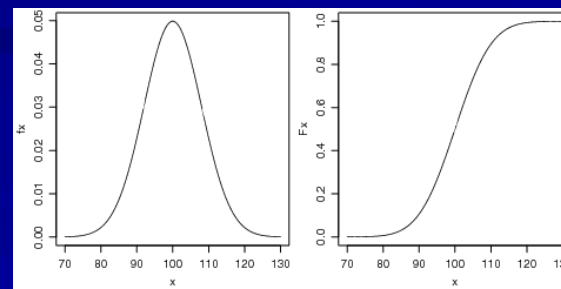
A função  $P(X = x_i)$  determina a **distribuição de probabilidade da variável aleatória X**.

As distribuições de probabilidade podem ser:  
distribuições de probabilidades **contínuas** ou  
distribuições de probabilidade **discretas**

Uma distribuição **discreta** descreve a probabilidade de ocorrência de cada valor de uma variável aleatória discreta. Uma variável aleatória discreta é uma variável aleatória que tem valores contáveis, como uma lista de inteiros não negativos.

| NÚMERO DE CARAS (X)                                                                 | P(X)         |
|-------------------------------------------------------------------------------------|--------------|
| 2                                                                                   | 1/4          |
| 1                                                                                   | 2/4          |
| 0                                                                                   | 1/4          |
|  | $\Sigma = 1$ |

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 intervalos) que é infinito e incontável.



São vários os modelos de distribuição de probabilidade. Seguem alguns:

- Distribuição binomial
- Distribuição beta
- Distribuição de Bernoulli
- Distribuição de Borel
- Distribuição de Cantor
- Distribuição de Poisson
- Distribuição de Weibull
- Distribuição F de Fisher-Snedecor
- Distribuição gama
- Distribuição hipergeométrica
- Distribuição normal
- Distribuição t de Student
- Distribuição Qui-quadrado
- Distribuição triangular

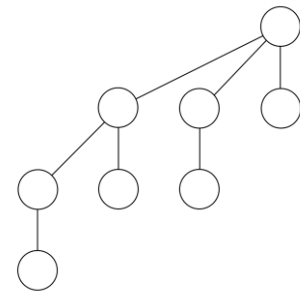
# Distribuição Binomial

## Características

- a. O experimento deve ser repetido, nas mesmas condições, um número finito de vezes ( $n$ ).
- b. As provas repetidas devem ser independentes, isto é, o resultado de uma não deve afetar os resultados das sucessivas.
- c. Em cada prova deve aparecer um dos dois possíveis resultados: sucesso e insucesso.
- d. No decorrer do experimento, a probabilidade  $p$  do sucesso e a probabilidade  $q$  ( $q = 1 - p$ ) do insucesso manter-se-ão constantes.

Resolveremos problemas do tipo: determinar a probabilidade de se obterem  $k$  sucessos em  $n$  tentativas.

Sabemos que, quando da realização de um experimento qualquer em uma única tentativa, se a probabilidade de realização de um evento (sucesso) é  $p$ , a probabilidade de não-realização desse mesmo evento (insucesso) é  $1 - p = q$ .





Suponhamos, agora, que realizemos a mesma prova  $n$  vezes sucessivas e independentes. A probabilidade de que um evento se realize  $k$  vezes nas provas é dada pela função:

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

Obs.: É uma das distribuições mais comuns em Estatística. Uma variável aleatória tem **distribuição binomial** quando o experimento ao qual está relacionada apresenta apenas **2 resultados** (sucesso ou fracasso).

$P(X = k)$  é a probabilidade de que o evento se realize  $k$  vezes em  $n$  provas;  
 $p$  é a probabilidade de que o evento se realize em uma só prova — **sucesso**;  
 $q$  é a probabilidade de que o evento não se realize no decurso dessa prova — **insucesso**;

$\binom{n}{k}$  é o coeficiente binomial de  $n$  sobre  $k$ , igual a  $\frac{n!}{k!(n-k)!}$ .\*.

Essa função, denominada **lei binomial**, define a **distribuição binomial**.

O nome **binomial** vem do fato de  $\binom{n}{k} p^k q^{n-k}$  ser o termo geral do desenvolvimento do **binômio de Newton**.



Exemplo:

Dois times de futebol, **A** e **B**, jogam entre si 6 vezes. Encontre a probabilidade de o time **A** ganhar 4 jogos.

*Temos:*

$$n = 6, k = 4, p = \frac{1}{3} \text{ e } q = 1 - \frac{1}{3} = \frac{2}{3}$$

$$P(X = 4) = \binom{6}{4} \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^2 = \cancel{15}^5 \times \frac{1}{81} \times \frac{4}{\cancel{9}_3} = \frac{20}{243}$$

$$P(X = 4) = \frac{20}{243} .$$

Seis parafusos são escolhidos ao acaso da produção de certa máquina, que apresenta 10% de peças defeituosas. Qual a probabilidade de serem defeituosos dois deles?

$$N= 6; K= 2; p= 1/10; q= 9/10$$

$$P(X= 2) = 6!/(2! \times 4!)(0,1)^2 \times (0,9)^4 = 0.098415 = \mathbf{9,8415\%}$$

# Tabuleiro de Galton



[http://phet.colorado.edu/sims/plinko-probability/plinko-probability\\_pt\\_BR.html](http://phet.colorado.edu/sims/plinko-probability/plinko-probability_pt_BR.html)

O **Tabuleiro de Galton**, também conhecido como **Quincunx**, é um dispositivo inventado por Francis Galton para demonstrar o teorema do limite central, em particular, que a distribuição normal é aproximada à distribuição binomial. Entre suas aplicações, oferecer ideias sobre regressão para média.

# Distribuição de Poisson

A **distribuição de Poisson** é uma distribuição de probabilidade de variável aleatória discreta que expressa a probabilidade de uma série de eventos ocorrer num certo período de tempo se estes eventos ocorrem independentemente de quando ocorreu o último evento.

A distribuição foi descoberta pelo matemático francês Siméon Denis Poisson (1781–1840).

A **distribuição de Poisson** representa um modelo probabilístico adequado para o estudo de um grande número de fenômenos observáveis. Alguns exemplos:

- *Chamadas telefônicas por unidade de tempo;*
- *Defeitos por unidade de área;*
- *Acidentes por unidade de tempo;*
- *Chegada de clientes a um supermercado por unidade de tempo;*
- *Número de glóbulos sangüíneos visíveis ao microscópio por unidade de área;*
- *Número de partículas emitidas por uma fonte de material radioativo por unidade de tempo.*

A **distribuição de Poisson** foca-se em variáveis aleatórias  $N$  que contam o número de ocorrências discretas de um certo fenômeno durante um intervalo de tempo de determinada duração.

A probabilidade de que existam exatamente  $k$  ocorrências ( $k$  sendo um inteiro não negativo,  $k = 0, 1, 2, 3, \dots$ ) é

$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Onde:

**$e$**  é base do logaritmo natural ( $e = 2.71828\dots$ ),

**$k$**  é o número designado de sucessos.

**$\lambda$**  é o número médio de sucessos em um intervalo específico.



1)Um departamento de polícia recebe em média 5 solicitações por hora.  
Qual a probabilidade de receber 2 solicitações numa hora  
selecionada aleatoriamente?

Solução:

$X$  = número designado de sucessos = 2

$\lambda$  = o número médio de sucessos num intervalo específico (uma hora) = 5

$$P(X=2) = (5^2 \times e^{-5})/2! = 0,08422434 \quad \text{ou} \quad \mathbf{8,42\%}$$

2)A experiência passada indica que um número médio de 6 clientes por hora param para colocar gasolina numa bomba.

Qual é a probabilidade de 3 clientes pararem qualquer hora?

$$P(X=3) = (6^3 \times e^{-6})/3! = 0,08928 \quad \text{ou} \quad \mathbf{8,93\%}$$

# Distribuição Normal

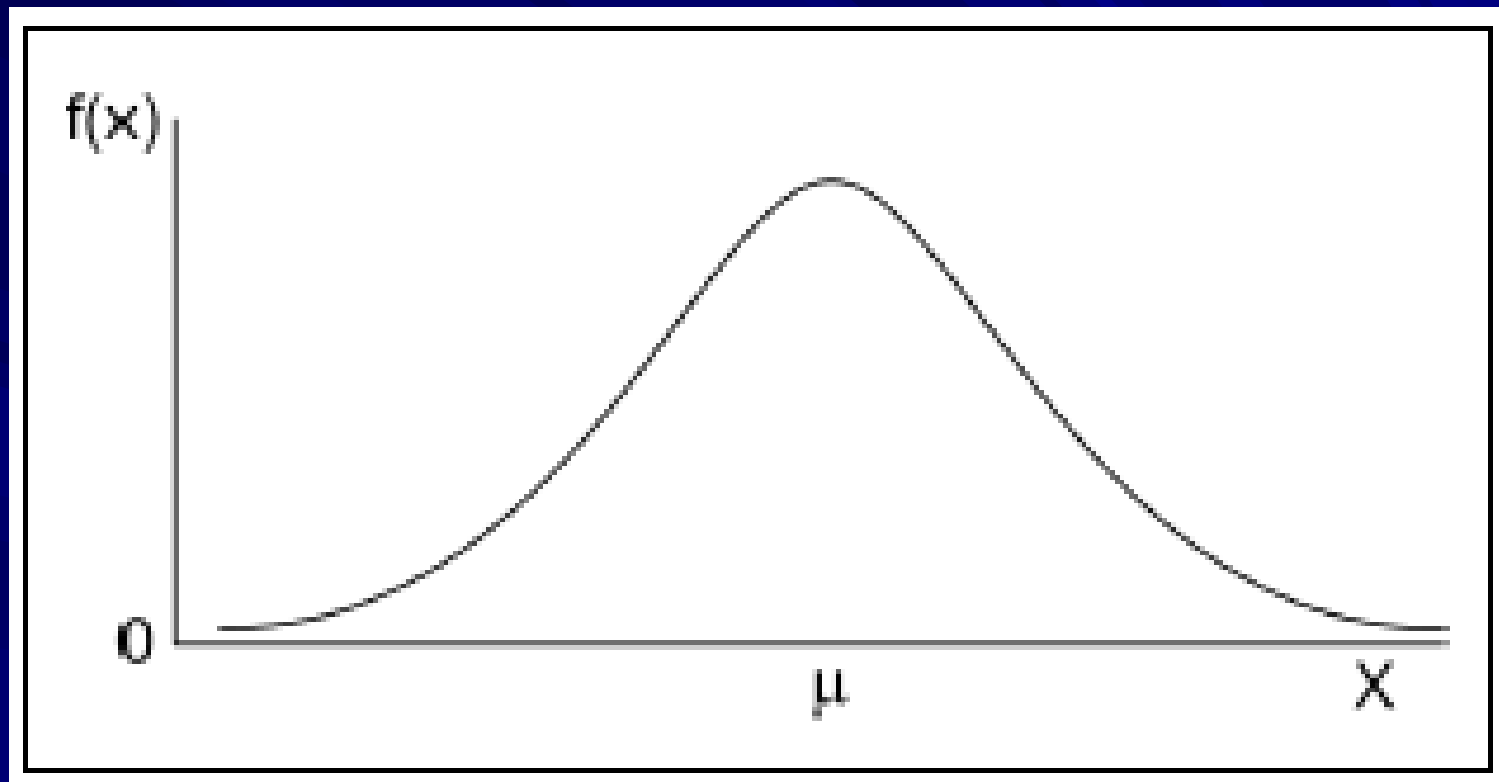
A **distribuição normal** é uma das distribuições de probabilidade mais utilizadas para modelar fenômenos naturais.

*A distribuição normal também é chamada distribuição gaussiana, distribuição de Gauss ou distribuição de Laplace–Gauss, em referência aos matemáticos, físicos e astrônomos francêss Pierre–Simon Laplace (1749 – 1827) e alemão Carl Friedrich Gauss (1777 – 1855).*

Em termos mais formais, a distribuição normal é uma distribuição de probabilidade absolutamente contínua parametrizada pela sua média ou esperança matemática (número real  $\mu$ ) e desvio padrão (número real positivo  $\sigma$ ). A densidade de probabilidade da distribuição normal é denotada como

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

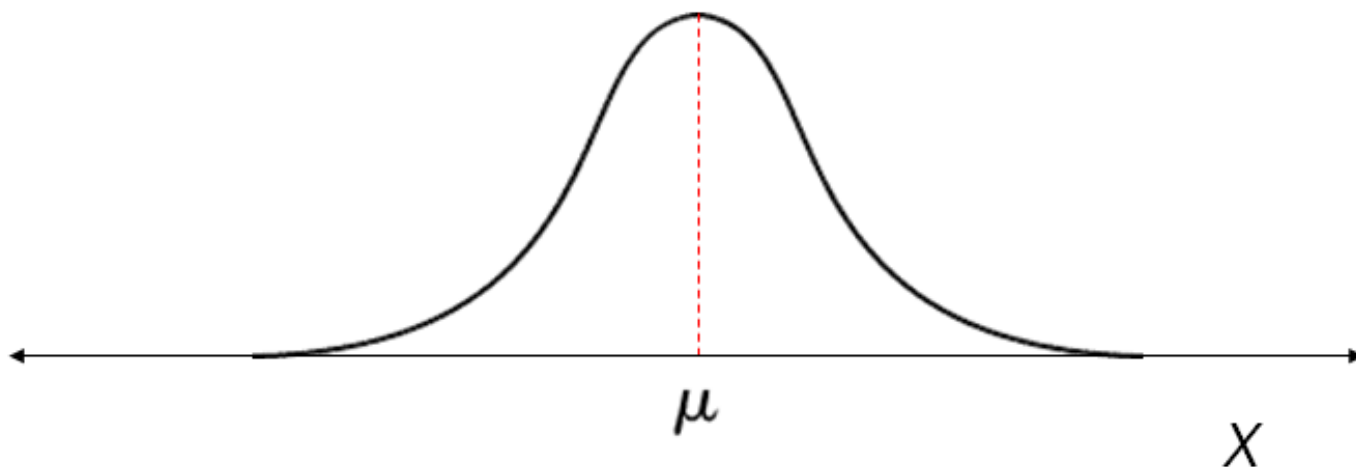
## Gráfico da **curva normal** ou curva de Gauss



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

## Propriedades da curva normal

- 1ª) A variável aleatória  $X$  pode assumir todo e qualquer valor real.
- 2ª) A representação gráfica da distribuição normal é uma curva em forma de sino, simétrica em torno da média ( $\bar{x}$ ), que recebe o nome de **curva normal** ou **de Gauss**.
- 3ª) A área total limitada pela curva e pelo eixo das abscissas é igual a 1, já que essa área corresponde à probabilidade de a variável aleatória  $X$  assumir qualquer valor real.
- 4ª) A curva normal é assintótica em relação ao eixo das abscissas, isto é, aproxima-se indefinidamente do eixo das abscissas sem, contudo, alcançá-lo.
- 5ª) Como a curva é simétrica em torno de  $\bar{x}$ , a probabilidade de ocorrer valor maior do que a média é igual à probabilidade de ocorrer valor menor do que a média, isto é, ambas as probabilidades são iguais a 0,5. Escrevemos:  $P(X > \bar{x}) = P(X < \bar{x}) = 0,5$ .

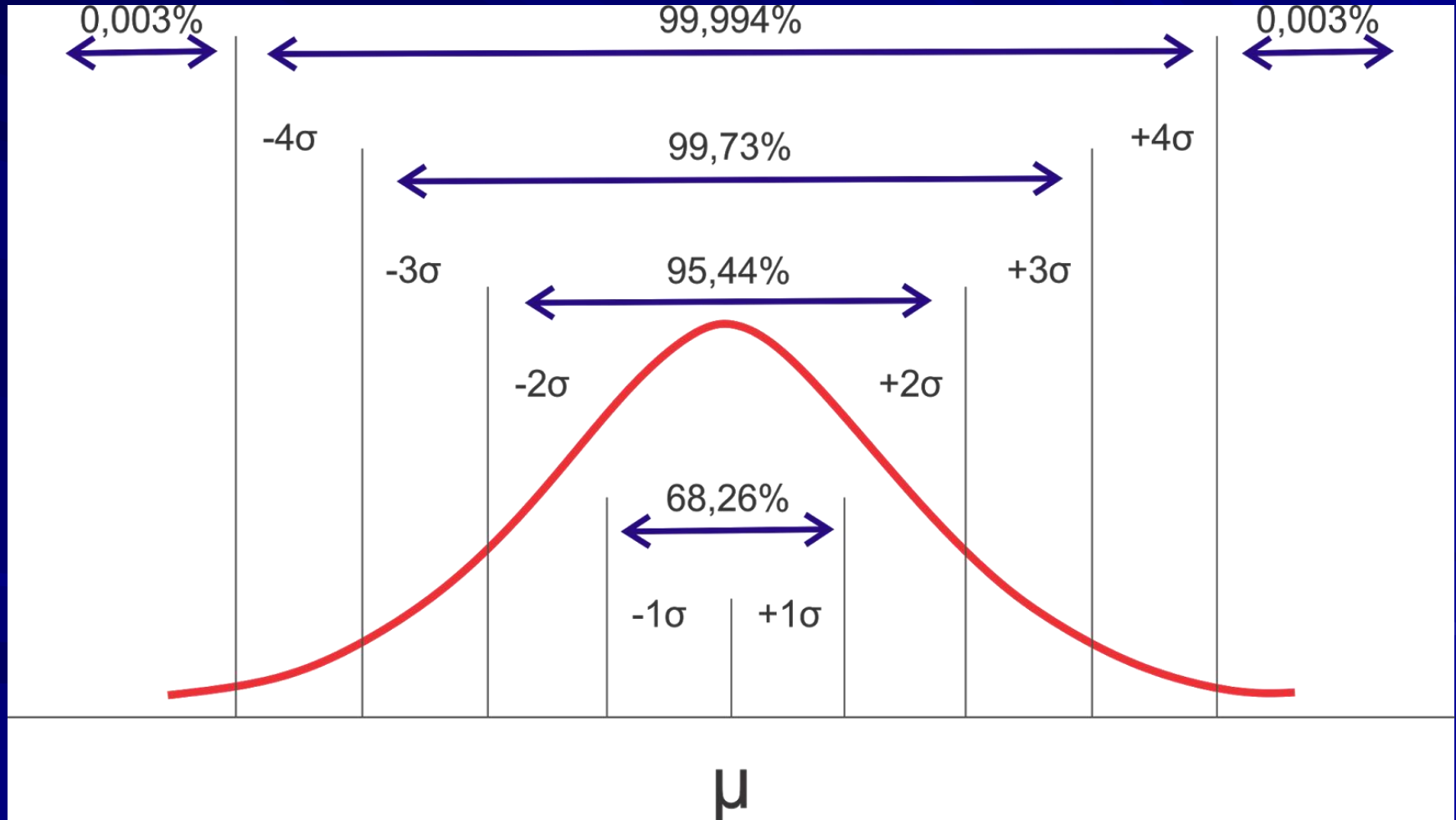


- Suas média, mediana e moda são iguais.
- Tem forma de sino e é simétrica em torno da média.
- A área total sob a curva é de 100%.

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



Na distribuição normal as proporções de observações em determinados intervalos são conhecidas. Exemplo, no intervalo entre média mais ou menos dois desvios-padrão ( $\mu \pm 2\sigma$ ) estão compreendidas cerca de 95% das observações e, no intervalo entre a média mais ou menos um desvio-padrão ( $\mu \pm \sigma$ ) cerca de 68%.



# Aplicação

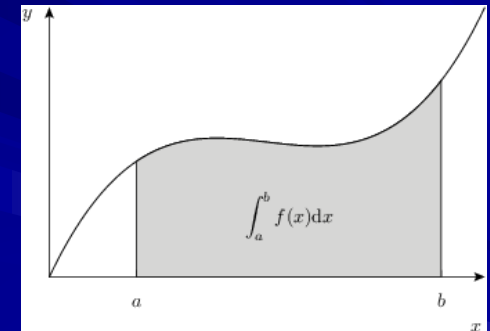
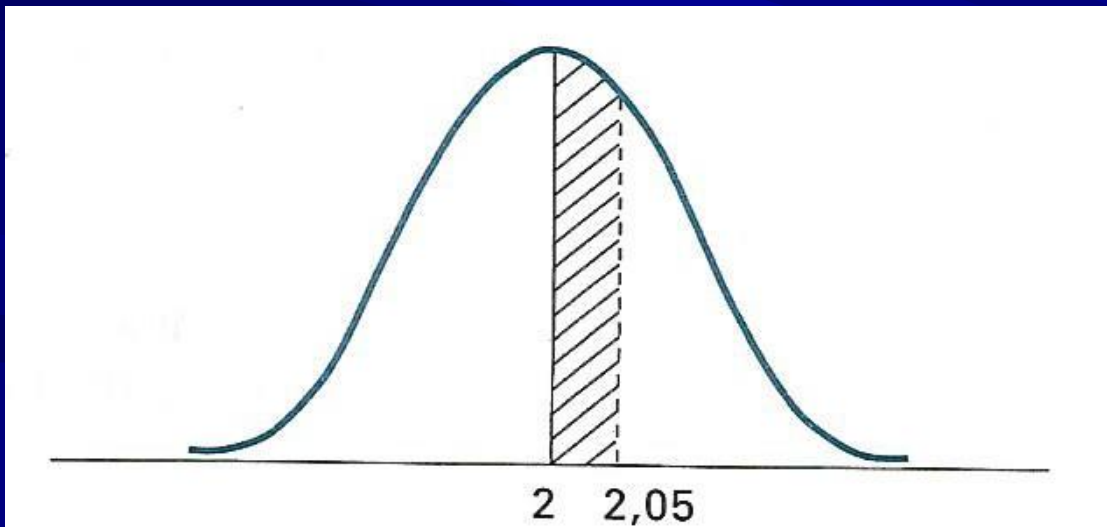
Seja  $X$  a variável aleatória que representa os diâmetros dos parafusos produzidos por certa máquina. Vamos supor que essa variável tenha distribuição normal com média  $\bar{x} = 2$  cm e desvio padrão  $s = 0,04$  cm.

Pode haver interesse em conhecer a probabilidade de um parafuso ter um diâmetro com valor entre 2 e 2,05 cm.

É fácil notar que essa probabilidade, indicada por:

$$P(2 < X < 2,05),$$

Corresponde à área marcada



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

## Normal reduzida ou padrão

### Padronizando a Distribuição Normal

- Utilizando a fórmula de transformação, qualquer variável aleatória normal  $X$  é convertida em uma variável normal padronizada  $Z$

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

onde:

$\sigma$  é o desvio padrão

$\mu$  é a média aritmética

Temos, então, que se  $X$  é uma variável aleatória com distribuição normal de média  $\bar{x}$  e desvio padrão  $s$ , podemos escrever:

$$P(\bar{x} < X < x) = P(0 < Z < z),$$

com  $z = \frac{x - \bar{x}}{s}$ .

Voltemos, então, ao nosso problema.

Queremos calcular  $P(2 < X < 2,05)$ . Para obter essa probabilidade, precisamos, em primeiro lugar, calcular o valor de  $z$  que corresponde a  $x = 2,05$  ( $x = 2 \Rightarrow z = 0$ , pois  $\bar{x} = 2$ ). Temos, então:

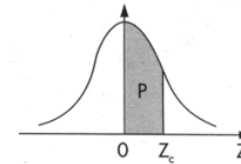
$$z = \frac{x - \bar{x}}{s} = \frac{2,05 - 2}{0,04} = \frac{0,05}{0,04} = 1,25,$$

donde:

$$P(2 < X < 2,05) = P(0 < Z < 1,25)$$

**Indo na tabela**

$Z \sim N(0, 1)$   
Corpo da tabela dá a probabilidade  $p$ , tal que  $p = P(0 < Z < Z_c)$



| parte inteira e primeira decimal de $Z_c$ | Segunda decimal de $Z_c$ |       |       |       |       |       |       |       |       |       | parte inteira e primeira decimal de $Z_c$ |
|-------------------------------------------|--------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------------------------------------------|
|                                           | 0                        | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |                                           |
|                                           | p = 0                    |       |       |       |       |       |       |       |       |       |                                           |
| 0,0                                       | 00000                    | 00399 | 00798 | 01197 | 01595 | 01994 | 02392 | 02790 | 03188 | 03586 | 0,0                                       |
| 0,1                                       | 03983                    | 04380 | 04776 | 05172 | 05567 | 05962 | 06356 | 06749 | 07142 | 07535 | 0,1                                       |
| 0,2                                       | 07926                    | 08317 | 08706 | 09095 | 09483 | 09871 | 10257 | 10642 | 11026 | 11409 | 0,2                                       |
| 0,3                                       | 11791                    | 12172 | 12552 | 12930 | 13307 | 13683 | 14058 | 14431 | 14803 | 15173 | 0,3                                       |
| 0,4                                       | 15542                    | 15910 | 16276 | 16640 | 17003 | 17364 | 17724 | 18082 | 18439 | 18793 | 0,4                                       |
| 0,5                                       | 19146                    | 19497 | 19847 | 20194 | 20540 | 20884 | 21226 | 21566 | 21904 | 22240 | 0,5                                       |
| 0,6                                       | 22575                    | 22907 | 23237 | 23565 | 23891 | 24215 | 24537 | 24857 | 25175 | 25490 | 0,6                                       |
| 0,7                                       | 25804                    | 26115 | 26424 | 26730 | 27035 | 27337 | 27637 | 27935 | 28230 | 28524 | 0,7                                       |
| 0,8                                       | 28814                    | 29103 | 29389 | 29673 | 29955 | 30234 | 30511 | 30785 | 31057 | 31327 | 0,8                                       |
| 0,9                                       | 31594                    | 31859 | 32121 | 32381 | 32639 | 32894 | 33147 | 33398 | 33646 | 33891 | 0,9                                       |
| 1,0                                       | 34134                    | 34375 | 34614 | 34850 | 35083 | 35314 | 35543 | 35769 | 35993 | 36214 | 1,0                                       |
| 1,1                                       | 36433                    | 36650 | 36864 | 37076 | 37286 | 37493 | 37698 | 37900 | 38100 | 38298 | 1,1                                       |
| 1,2                                       | 38493                    | 38686 | 38877 | 39065 | 39251 | 39435 | 39617 | 39796 | 39973 | 40147 | 1,2                                       |
| 1,3                                       | 40320                    | 40490 | 40658 | 40824 | 40988 | 41149 | 41309 | 41466 | 41621 | 41774 | 1,3                                       |
| 1,4                                       | 41924                    | 42073 | 42220 | 42364 | 42507 | 42647 | 42786 | 42922 | 43056 | 43189 | 1,4                                       |
| 1,5                                       | 43319                    | 43448 | 43574 | 43699 | 43822 | 43943 | 44062 | 44179 | 44295 | 44408 | 1,5                                       |
| 1,6                                       | 44520                    | 44630 | 44738 | 44845 | 44950 | 45053 | 45154 | 45254 | 45352 | 45449 | 1,6                                       |
| 1,7                                       | 45543                    | 45637 | 45728 | 45818 | 45907 | 45994 | 46080 | 46164 | 46246 | 46327 | 1,7                                       |
| 1,8                                       | 46407                    | 46485 | 46562 | 46638 | 46712 | 46784 | 46856 | 46926 | 46995 | 47062 | 1,8                                       |
| 1,9                                       | 47128                    | 47193 | 47257 | 47320 | 47381 | 47441 | 47500 | 47558 | 47615 | 47670 | 1,9                                       |
| 2,0                                       | 47725                    | 47778 | 47831 | 47882 | 47932 | 47982 | 48030 | 48077 | 48124 | 48169 | 2,0                                       |
| 2,1                                       | 48214                    | 48257 | 48300 | 48341 | 48382 | 48422 | 48461 | 48500 | 48537 | 48574 | 2,1                                       |
| 2,2                                       | 48610                    | 48645 | 48679 | 48713 | 48745 | 48778 | 48809 | 48840 | 48870 | 48899 | 2,2                                       |
| 2,3                                       | 48928                    | 48956 | 48983 | 49010 | 49036 | 49061 | 49086 | 49111 | 49134 | 49158 | 2,3                                       |
| 2,4                                       | 49180                    | 49202 | 49224 | 49245 | 49266 | 49286 | 49305 | 49324 | 49343 | 49361 | 2,4                                       |
| 2,5                                       | 49379                    | 49396 | 49413 | 49430 | 49446 | 49461 | 49477 | 49492 | 49506 | 49520 | 2,5                                       |
| 2,6                                       | 49534                    | 49547 | 49560 | 49573 | 49585 | 49598 | 49609 | 49621 | 49632 | 49643 | 2,6                                       |
| 2,7                                       | 49653                    | 49664 | 49674 | 49683 | 49693 | 49702 | 49711 | 49720 | 49728 | 49736 | 2,7                                       |
| 2,8                                       | 49744                    | 49752 | 49760 | 49767 | 49774 | 49781 | 49788 | 49795 | 49801 | 49807 | 2,8                                       |
| 2,9                                       | 49813                    | 49819 | 49825 | 49831 | 49836 | 49841 | 49846 | 49851 | 49856 | 49861 | 2,9                                       |
| 3,0                                       | 49865                    | 49869 | 49874 | 49878 | 49882 | 49886 | 49889 | 49893 | 49897 | 49900 | 3,0                                       |
| 3,1                                       | 49903                    | 49906 | 49910 | 49913 | 49916 | 49918 | 49921 | 49924 | 49926 | 49929 | 3,1                                       |
| 3,2                                       | 49931                    | 49934 | 49936 | 49938 | 49940 | 49942 | 49944 | 49946 | 49948 | 49950 | 3,2                                       |
| 3,3                                       | 49952                    | 49953 | 49955 | 49957 | 49958 | 49960 | 49961 | 49962 | 49964 | 49965 | 3,3                                       |
| 3,4                                       | 49966                    | 49968 | 49969 | 49970 | 49971 | 49972 | 49973 | 49974 | 49975 | 49976 | 3,4                                       |
| 3,5                                       | 49977                    | 49978 | 49978 | 49979 | 49980 | 49981 | 49981 | 49982 | 49983 | 49983 | 3,5                                       |
| 3,6                                       | 49984                    | 49985 | 49985 | 49986 | 49986 | 49987 | 49987 | 49988 | 49988 | 49989 | 3,6                                       |
| 3,7                                       | 49989                    | 49990 | 49990 | 49990 | 49991 | 49991 | 49992 | 49992 | 49992 | 49992 | 3,7                                       |
| 3,8                                       | 49993                    | 49993 | 49993 | 49994 | 49994 | 49994 | 49994 | 49995 | 49995 | 49995 | 3,8                                       |
| 3,9                                       | 49995                    | 49995 | 49996 | 49996 | 49996 | 49996 | 49996 | 49996 | 49997 | 49997 | 3,9                                       |
| 4,0                                       | 49997                    | 49997 | 49997 | 49997 | 49997 | 49997 | 49998 | 49998 | 49998 | 49998 | 4,0                                       |
| 4,5                                       | 49999                    | 50000 | 50000 | 50000 | 50000 | 50000 | 50000 | 50000 | 50000 | 50000 | 4,5                                       |



Na primeira coluna encontramos o valor **1,2**. Em seguida, encontramos, na primeira linha, o valor **5**, que corresponde ao último algarismo do número **1,25**. Na intersecção da linha e coluna correspondentes encontramos o valor **0,3944**, o que nos permite escrever:

$$P(0 < Z < 1,25) = 0,3944$$

Assim, a probabilidade de um parafuso fabricado por essa máquina apresentar um diâmetro entre a média  $\bar{x} = 2$  e o valor  $x = 2,05$  é 0,3944.

Escrevemos, então:

$$P(2 < X < 2,05) = P(0 < Z < 1,25) = 0,3944 \text{ ou } 39,44\%$$

## Exemplo de aplicação

Os salários semanais dos operários industriais são distribuídos normalmente, em torno da média de R\$ 500, com desvio padrão de R\$ 40. Calcule a probabilidade de um operário ter um salário semanal situado entre R\$ 490 e R\$ 520.

*Devemos, inicialmente, determinar os valores da variável de distribuição normal reduzida.*

*Assim:*

$$z_1 = \frac{490 - 500}{40} = -0,25 \text{ e } z_2 = \frac{520 - 500}{40} = 0,5$$

*Logo, a probabilidade procurada é dada por:*

$$P(490 < X < 520) = P(-0,25 < Z < 0,5) = P(-0,25 < Z < 0) + P(0 < Z < 0,5) = 0,0987 + 0,1915 = 0,2902$$

*É, pois, de se esperar que, em média, 29,02% dos operários tenham salários entre R\$ 490 e R\$ 520.*

# Amostragem

Ao conjunto de entes portadores de, pelo menos, uma característica comum denominamos **população estatística** ou **universo estatístico**.

Uma **amostra** é um subconjunto finito de uma população.

**Amostragem** é o processo de obtenção de amostras, que são uma pequena parte de uma população.

As modalidades de amostragem podem ser divididas em dois grandes grupos: **quantitativo e qualitativo**.

# Principais tipos de amostragens

## Amostragem aleatória simples

É aquela em que toda amostra possível de mesmo tamanho tem a mesma chance de ser selecionada a partir da população.

## Amostragem sistemática

Consiste em um elemento aleatório, por exemplo, um nome a cada dez de uma lista, a décima peça produzida em uma linha de produção etc. Sua principal vantagem é sua simplicidade e flexibilidade, sendo mais fácil de instruir os trabalhadores de campo.

## Amostragem estratificada

Consiste em dividir ou estratificar a população em um certo número de subpopulações que não se sobrepõem e então extrair uma amostra de cada estrato. Este tipo de amostragem também é usado quando métodos diferentes de coleta de dados são aplicados em diferentes partes da população.



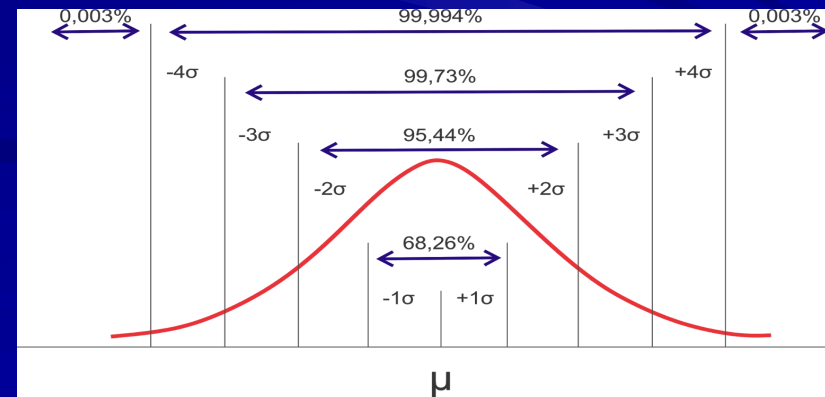
Para ser considerada confiável, a amostra deve ser representativa, isto é, não pode ser tendenciosa. Dois erros podem ser cometidos na definição de uma amostra: os vieses, vale dizer, os desvios da amostra em relação ao universo, e os erros de amostragem, ou seja, as distorções resultantes do fato de se trabalhar com uma parte da população e não com toda ela. Erros de vieses devem ser minimizados; erros de amostragem podem ser calculados.



### Níveis de confiança por unidades de desvios padrões

| Valores de $\sigma$ | Nível de confiança (em %) |
|---------------------|---------------------------|
| 1                   | 68,3                      |
| 2                   | 95,4                      |
| 3                   | 99,7                      |

Quando se diz, por exemplo, que o nível de confiança de tal amostra é de 95% significa que há uma probabilidade de 95% de os resultados obtidos por intermédio da amostra serem válidos para todo o universo. Em outras palavras, se a pesquisa fosse repetida 100 vezes, em 95 casos os resultados estariam dentro de uma faixa previamente estabelecida, porque, em uma amostra, não se esperam resultados precisos. Espera-se que haja alguma variação. É o que se chama de erro máximo de estimativa.



Em termos didáticos, o erro máximo de estimativa significa o quanto o valor encontrado na amostra pode variar, para mais ou para menos, se for aplicado a todo o universo. Por exemplo: se uma pesquisa eleitoral diz que tal candidato tem 52% dos votos com erro de estimativa de 3%, isto quer dizer que a votação pode se situar entre 49% e 55%. Se houver outro candidato com 49% de intenção de votos, diz-se que estão tecnicamente empatados, pois o segundo candidato pode ter de 46% a 52% dos votos.

# Tamanho da amostra

*População infinita:* quando o número de casos de um universo é superior a 100 mil, a fórmula para a determinação do tamanho da amostra é (Cfr. Richardson, 1989, p. 119; Hoel, 1968, p. 127):

$$n = \frac{\sigma^2 pq}{e^2} \quad (1)$$

onde:

$n$  = tamanho da amostra.

$\sigma^2$  = nível de confiança em quantidade de desvios padrões.

$p$  = proporção do universo que possui a propriedade pesquisada.

$q$  = proporção do universo sem a propriedade pesquisada ( $q = 100 - p$ )

$e$  = erro de estimação permitido (em valores percentuais)

Exemplo: Quantos eleitores deverão ser entrevistados em uma pesquisa eleitoral no Brasil para se ter um nível de confiança de 99% e margem máxima de erro de 4%?

Solução: no caso, os valores são:

$\sigma = 3$  (99% de confiança corresponde a  $3\sigma$ , na Tabela 1).

$p = 50$  (usando o valor 50 para  $p$ , maximiza-se o numerador, aumentando a segurança da amostra).

$q = 50$  ( $q = 100 - p$ ).

$e = 4$  (margem de erro de 4%).

Logo, substituindo os valores na fórmula (1), obtém-se:

$$n = \frac{3^2 \times 50 \times 50}{4^2} = \frac{9 \times 50 \times 50}{16} = 1.406,25$$



*População finita:* ao se tratar de população com menos de 100 mil indivíduos, ela é considerada finita. Há mais um parâmetro, o tamanho da população (N), e a fórmula recebe alguns ajustes para aumentar a precisão:

$$n = \frac{\sigma^2 pq.N}{e^2 (N-1) + \sigma^2 pq} \quad (2)$$

Considere-se uma empresa com 10 mil empregados, distribuídos por nível hierárquico e por gênero conforme a Tabela 2, a seguir.

*Tabela 2*

**Funcionários da empresa  
segundo os níveis hierárquicos e o gênero**

| Nível<br>hierárquico | Homens |    | Mulheres |    | Total  |     |
|----------------------|--------|----|----------|----|--------|-----|
|                      | %      | f  | %        | f  | %      | f   |
| Estratégico          | 300    | 3  | 200      | 2  | 500    | 5   |
| Tático               | 800    | 8  | 700      | 7  | 1.500  | 15  |
| Operacional          | 5.400  | 54 | 2.600    | 26 | 8.000  | 80  |
| Total                | 6.500  | 65 | 3.500    | 35 | 10.000 | 100 |

Fonte: Dados fictícios.

Qual será o tamanho da amostra, com nível de confiança de 95%, e erro de estimação permitido de: *a.* 3%; e *b.* 4%?

Para ambas as questões, a resposta está na aplicação da fórmula (2):

$N = 10.000$

$p = 50$

$q = 50$ .

$\sigma = 2$  (nível de confiança de 95%)

- a. Para a questão *a*,  $e = 3$ . Logo, aplicando a fórmula, obtém-se:

$$n = \frac{2^2 \times 50 \times 50 \times 10.000}{3^2 (10.000 - 1) + 2^2 \times 50 \times 50}$$

$$n = \frac{4 \times 2.500 \times 10.000}{9 \times 9.999 + 4 \times 2.500} = \frac{100.000.000}{99.991} = 1.000,09$$

- b. Para a questão *b*,  $e = 4$ , aplicando a fórmula, obtém-se:

$$n = \frac{2^2 \times 50 \times 50 \times 10.000}{4^2 (10.000 - 1) + 2^2 \times 50 \times 50}$$

$$n = \frac{4 \times 2.500 \times 10.000}{16 \times 9.999 + 4 \times 2.500} = 588,99$$

Esses valores, para ambas as situações, são válidos para a empresa como um todo. Admita-se, porém, que o gênero pudesse influenciar o tipo de comportamento estudado. Nesse caso, seria necessário estratificar a amostra, ou seja, deveria contar com representantes de cada gênero na mesma proporção do universo. Dessa forma, a amostra nas duas situações devia ter 65% de homens e 35% de mulheres, vale dizer:

- a. Situação *a*:

$$1.000 \times 0,65 = 650 \text{ homens}$$

$$1.000 \times 0,35 = 350 \text{ mulheres.}$$

- b. Situação *b*:

$$588 \times 0,65 = 382 \text{ homens}$$

$$588 \times 0,35 = 206 \text{ mulheres.}$$

Admita-se, ainda, que o comportamento dos funcionários pudesse ser influenciado pelo nível hierárquico. Nesse caso, com raciocínio análogo, chegar-se-ia à seguinte estratificação das amostras:



a. Situação *a*:

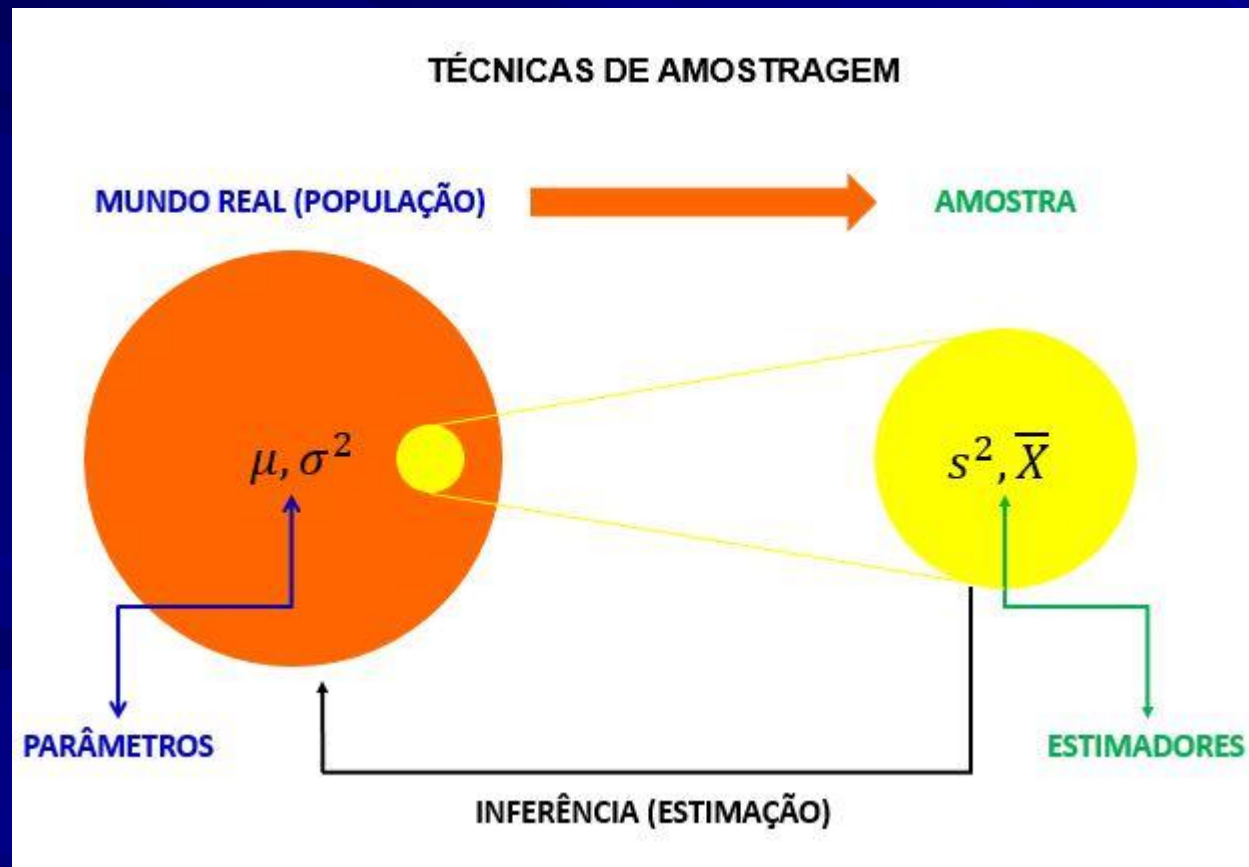
50 funcionários do nível estratégico (5% da amostra);  
150 funcionários do nível tático (15% da amostra); e  
800 funcionários do nível operacional (80% da amostra).

b. Na situação *b*, pelo mesmo raciocínio, a amostra teria 30, 88 e 470 funcionários dos níveis estratégico, tático e operacional, respectivamente.

# Estimativa

A **estimativa** é um valor (ou valores) que atribuímos a um **parâmetro** de uma **população** baseado em um valor da **estatística** correspondente da **amostra**.

O **estimador** é a estatística da amostra utilizada para estimar um parâmetro da população.



Exemplo:

Imagine que você quer saber o tempo médio de duração de um casamento no Brasil, onde sua população seriam todos os divorciados.

No entanto, como não é viável analisar toda a população, calculamos uma estimativa, a partir de uma amostra representativa da população.

Nesse caso, você terá que extrair uma amostra dentre as pessoas divorciadas, e registrar a duração do casamento. Utilizando essas informações poderemos calcular uma média aritmética da amostra.

E então com base nesse valor, atribuiremos valores à média populacional  $\mu$ .

Para materializar, suponhamos que você extraia uma amostra representativa de tamanho  $n$ , e descubra que o **tempo médio de duração de um casamento no Brasil** é de 15 anos.

Nesse caso, consideramos o valor como uma estimativa de  $\mu$ . Como a estatística da amostra, utilizada para estimar um parâmetro da população, é chamada de **estimador**, a média aritmética da amostra, representa um estimador para a média aritmética da população  $\mu$ .

Então a minha estimativa é 15 anos, e o meu estimador é a média (poderia ser variância, desvio-padrão, mediana, moda).