



# **ESTATÍSTICA**

*Introdução ao estudo da Estatística*

- I. Conceitos e objetivos da estatística
- II. Tipos de variáveis
- III. Sistema de gerenciamento de dados e formulários
- IV. Técnicas de amostragem, coleta e crítica dos dados
- V. Exercícios

- No desenvolvimento científico e em nosso próprio dia a dia, estamos sempre fazendo observações de fenômenos, gerando **dados**.
- Os engenheiros estão frequentemente analisando dados de propriedades dos materiais;
- **os profissionais da informática estão avaliando dados de desempenho de novos sistemas computacionais;**

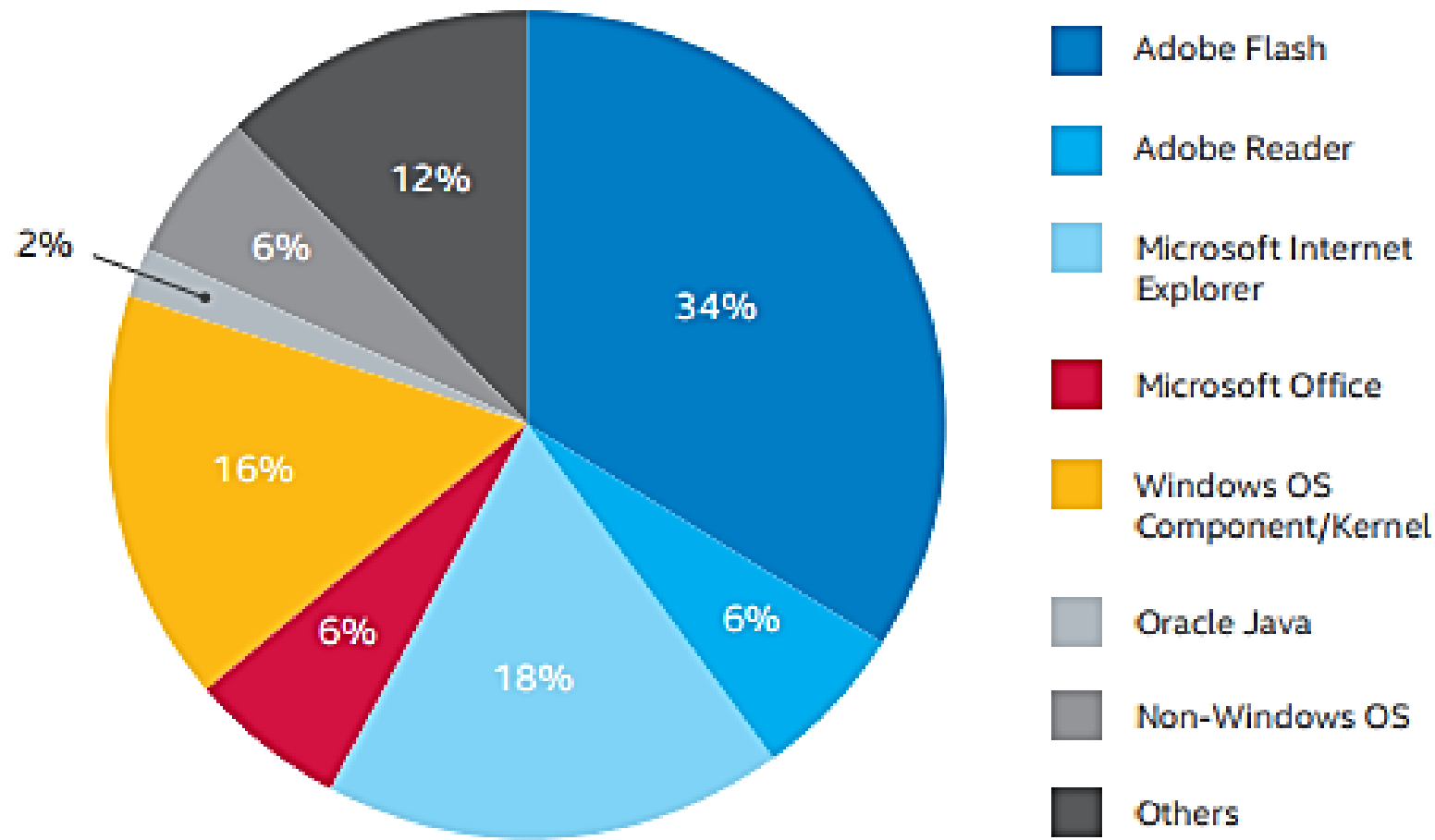
- Todos nós, ao lermos jornais e revistas, estamos vendo resultados estatísticos provenientes do censo demográfico, de **pesquisas eleitorais** etc.
- Eleições 2020...

- Os dados podem provir de **estudos observacionais ou de experimentos planejados.**
- Ao acompanhar o desempenho de um processo produtivo em sua forma natural, estamos fazendo um **estudo observacional**;
- Ao alterar de forma proposital as variáveis do processo para verificar seus efeitos nos resultados, estamos realizando **um experimento.**

Estatística é a parte da matemática  
que trata dos métodos científicos  
para:



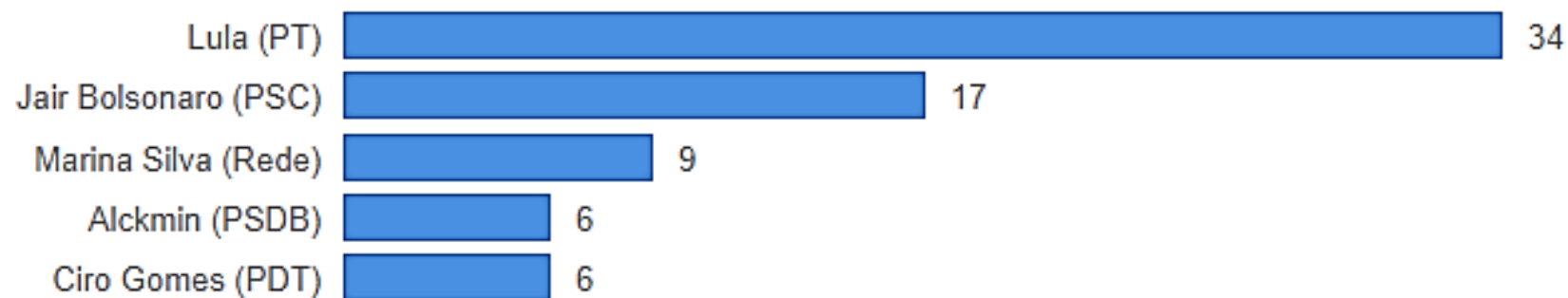
As vulnerabilidades de aplicativos descobertas pelas empresas de segurança em 2014 e 2015, de acordo com relatório da McAfee Labs.



Fonte: McAfee Labs Report 2016 Threats Predictions

## O cenário eleitoral

Pesquisa feita com 2.765 pessoas, entre 29 e 30 de novembro, em 192 cidades. A margem de erro é de dois pontos para mais ou menos. Em %



*Outros: Joaquim Barbosa (Sem Partido) 5%, Alvaro Dias (Podemos) 3%, Manuela D'Ávila (PCdoB) 1%, Michel Temer (PMDB) 1%, Henrique Meirelles (PSD) 1%, □ Paulo Rabello de Castro (PSC) 1%, em branco/nulo/nenhum 12%, não sabe 2%*

Fonte: Datafolha

O GLOBO

(Pesquisa feita entre 29 e 30 de novembro de 2017)

<https://oglobo.globo.com/brasil/lula-segue-na-lideranca-para-2018-bolsonaro-se-consolida-em-2-lugar-aponta-datafolha-22142803>

- Média de 15 pessoas entrevistadas por cidade.
- Não diz em quais regiões do país a pesquisa aconteceu.



## Usos e abusos da Estatística

- Usos: empresas, hospitais, governo.
- Abusos: Pequenas amostras, números imprecisos, perguntas tendenciosas, gráficos enganosos, pressão do pesquisador, más amostras.

A Estatística subdivide-se em três áreas:

- Descritiva
- Probabilística
- Inferencial

A **Estatística descritiva** refere-se às seguintes tarefas:

- 1) Encontrar um método apropriado de coletar dados numéricos;
- 2) Determinar um formato eficiente, tal como uma apresentação tabular, para a organização dos dados;

- 3) Apresentar dados numéricos por meio de métodos gráficos;
- 4) Sumarizar ou descrever cada característica através de média, porcentagem ou alguma outra medida apropriada.

## Exemplos:

- Sistema eletrônico do serviço de informação ao cidadão  
<https://esic.cgu.gov.br/sistema/Relatorios/Anual/RelatorioAnualRecursos.aspx>
- DATASUS  
<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/pfet10SC.def>

Em estatística, utilizamos extensamente os termos:

- **População:** é a totalidade dos itens ou objetos considerados. Ex:  $n^0$  de eleitores,  $n^0$  de inscritos no CPF,  $n^0$  de empregados.
- **Amostra:** é a parte da população selecionada para a análise.

A essência de uma análise estatística é tirar conclusões sobre uma **população**, ou universo, **com base em uma amostra de observações**.

- Censo: exame de todas as unidades de observação de uma população. Viável??
- Amostragem: processo pelo qual uma amostra de unidades da população é retirada e observada.

**Deve retratar fielmente a população pesquisada.**



As amostras podem ser classificadas em:

- Amostra probabilística
- Amostra não probabilística

*Qualquer que seja o esquema de amostragem, deve-se sempre garantir que a amostra reflita as características da população da qual foi retirada.*

**Não é o tamanho que garante representatividade da amostra, mas a forma com ela é obtida!!!**

Exemplos:

- Pesquisa sobre o aborto
- Pesquisa de satisfação de uso de aplicativo

(horário, época do mês)

**Variável** é a característica que vai ser observada, medida ou contada nos elementos da população ou da amostra.

Podem ser:

- **VARIÁVEIS QUALITATIVAS** – (ou dados categóricos) característica não numérica. Exemplos: tipo sanguíneo, raça, gênero.

- **VARIÁVEIS QUANTITATIVAS** – obtidas por meio de um processo de medição ou contagem (contínuas e discretas).

Ex: peso, altura, temperatura, comprimento, quantidade de chuva (mm), número de falhas, número de ligações telefônicas, número de mensagens eletrônicas, número de faltas de um aluno numa disciplina, nota final na disciplina, etc.

A variável quantitativa possui o mais alto nível de informação!!!

Para cada variável quantitativa podemos calcular:

- valor médio,
- mediano,
- modal,
- mínimo, máximo,
- desvio padrão, coeficiente de variação, intervalos específicos de variação e outras técnicas analíticas.

Quais são  
as  
variáveis?

Bom Retiro, Joinville - SC		
HOJE	AMANHÃ	10 DIAS
quarta-feira, 28 de fev Tempestades isoladas com trovoadas	80% 	29° 23°
quinta-feira, 1 de mar Chuvas com trovoadas	80% 	28° 23°
sexta-feira, 2 de mar Chuvas com trovoadas	80% 	29° 23°
sábado, 3 de mar Chuvas com trovoadas	80% 	30° 23°
domingo, 4 de mar Pancadas de chuva	60% 	29° 22°
segunda-feira, 5 de mar Pancadas de chuva	60% 	28° 22°

weather.com

- ***Excel***, um dos componentes do pacote Office da Microsoft
- R e RStudio
- BioStat
- SAS
- Formulários do *Google Docs* <https://docs.google.com/forms>.

- COLETA DOS DADOS:

Após a definição do problema a ser estudado o passo seguinte é a coleta dos dados. Como fazer?

- CRÍTICA DOS DADOS

A revisão crítica dos dados procede com a finalidade de **suprimir os valores estranhos ao levantamento**, os quais são capazes de provocar futuros enganos.

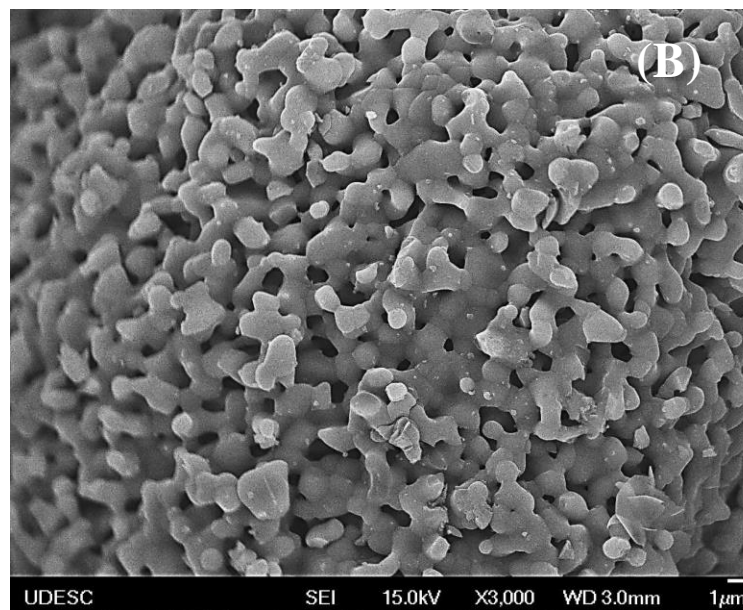
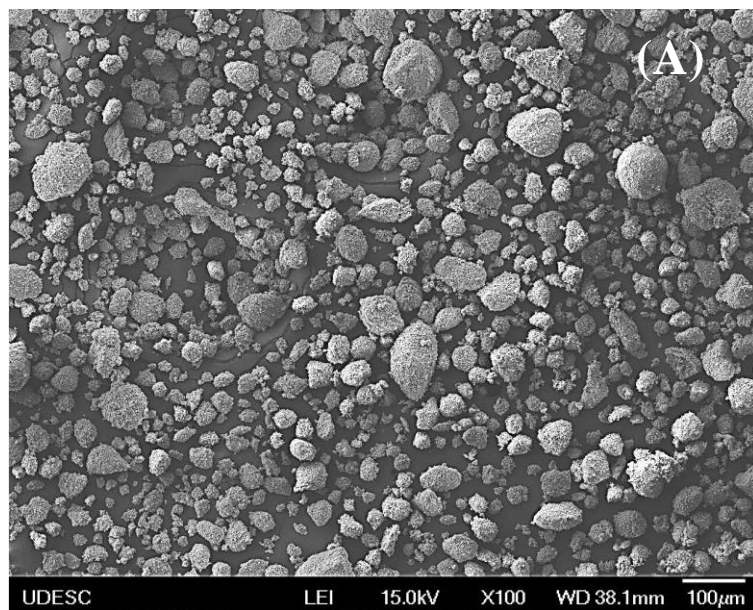
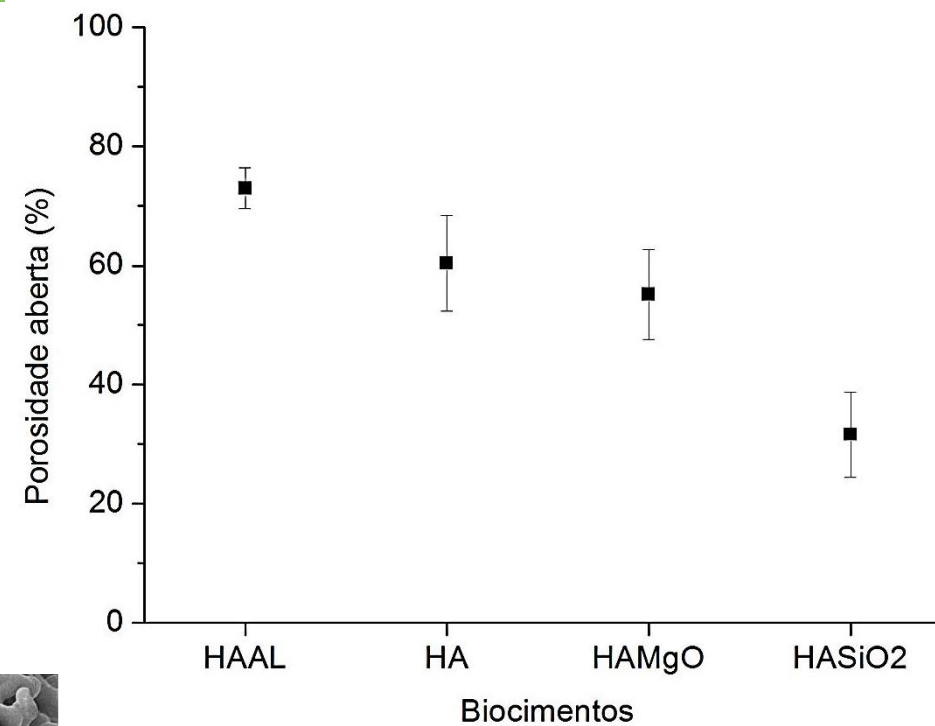


- APRESENTAÇÃO DOS DADOS

Convém organizarmos o conjunto de dados de maneira:

**Prática e Racional.**

Biocimentos	Porosidade aberta (%)	Desvio Padrão
HAAL	72,99	3,44
HA	60,38	8,01
HAMgO	55,14	7,56
HASiO2	31,62	7,14



- TIPOS DE AMOSTRAGEM

## Amostragem aleatória simples

Também conhecida por amostragem ocasional, acidental, casual, randômica, etc.

Todos os elementos da população têm **igual probabilidade** de serem escolhidos, desde o início até o completo processo de coleta.

## PROCEDIMENTO

1. Devemos enumerar todos os elementos da população
2. Devemos efetuar sucessivos sorteios até completar o tamanho da amostra ( $n$ )

Para realizarmos este sorteio devemos fazer uso das “tábuas de números aleatórios”

Média de idade da turma

**Tabela 6.1** Números Aleatórios

3690	2492	7171	7720	6509	7549	2330	5733	4730
0813	6790	6858	1489	2669	3743	1901	4971	8280
6477	5289	4092	4223	6454	7632	7577	2816	9202
0772	2160	7236	0812	4195	5589	0830	8261	9232
5692	9870	3583	8997	1533	6466	8830	7271	3809
2080	3828	7880	0586	8482	7811	6807	3309	2729
1039	3382	7600	1077	4455	8806	1822	1669	7501
7227	0104	4141	1521	9104	5563	1392	8238	4882
8506	6348	4612	8252	1062	1757	0964	2983	2244
5086	0303	7423	3298	3979	2831	2257	1508	7642
0092	1629	0377	3590	2209	4839	6332	1490	3092
0935	5565	2315	8030	7651	5189	0075	9353	1921
2605	3973	8204	4143	2677	0034	8601	3340	8383
7277	9889	0390	5579	4620	5650	0210	2082	4664
5484	3900	3485	0741	9069	5920	4326	7704	6525
6905	7127	5933	1137	7583	6450	5658	7678	3444
8387	5323	3753	1859	6043	0294	5110	6340	9137
4094	4957	0163	9717	4118	4276	9465	8820	4127
4951	3781	5101	1815	7068	6379	7252	1086	8919
9047	0199	5068	7447	1664	9278	1708	3625	2864
7274	9512	0074	6677	8676	0222	3335	1976	1645
9192	4011	0255	5458	6942	8043	6201	1587	0972
0554	1690	6333	1931	9433	2661	8690	2313	6999
9231	5627	1815	7171	8036	1832	2031	6298	6073
3995	9677	7765	3194	3222	4191	2734	4469	8617

- **Amostragem Sistemática**

Usada quando a população está **naturalmente ordenada**, como fichas em um fichário, listas telefônicas etc.

Requer uma lista dos itens da população, e, assim, padece das mesmas restrições já mencionadas na aleatória ao acaso.

## PROCEDIMENTO

Sejam os seguintes elementos:

$N$ : tamanho da população;

$n$ : tamanho da amostra.

Então, calcula-se o intervalo de amostragem através da razão  $a = \frac{N}{n}$  (onde ***a*** é o número inteiro mais próximo).

Sorteia-se, utilizando a tábua de números aleatórios, um número  $x$  entre **1** e ***a*** formando-se a amostra dos elementos correspondentes ao conjunto de números:

$$\mathbf{x; x+a; x+2a;...; x+(n-1)a}$$



EXEMPLO: Na saída de uma casa noturna com capacidade para 500 pessoas, a polícia resolve fazer uma *Blitz* da Lei seca a 100 m do local. Há 50 canudos descartáveis de bafômetro no local e os policiais decidem fazer uma amostragem sistemática para realizar o teste de embriaguez, já que existe apenas uma saída de veículos e a via é de mão única.

Seja  $N = 500$  (tamanho da população) e  $n = 50$  (tamanho da amostra)

Então:

$$a = \frac{500}{50}$$

$$a = 10$$



$$a = 10$$

Sorteia-se um número entre 1 e 10. Seja 3 ( $x = 3$ ) o número sorteado.

$$\mathbf{x; x+a; x+2a;...; x+(n-1)a}$$

$$3; (3+10); (3+2.10); (3+3.10).....$$

Logo, os carros que corresponderem à ordem 3;13;23;33;... serão os componentes da amostra.

## Amostragem estratificada

No caso de possuir uma população com uma certa característica heterogênea, na qual podemos distinguir subpopulações mais ou menos homogêneas, denominadas de estratos, podemos usar a amostragem estratificada.

Estratificar uma população em  $L$  subpopulações denominada estratos, tais que:

$$n_1 + n_2 + \dots + n_L = n$$

**Exemplo:** Vamos obter uma amostra proporcional estratificada de 10% para a pesquisa de satisfação de 90 alunos de uma universidade que usam os sistemas operacionais Windows e Linux.

54 usam o Windows e 36 usam o Linux.

Queremos saber se os usuários estão satisfeitos com o sistema que estão usando.



a) O primeiro passo é determinar o tamanho da amostra em cada estrato:

Temos aqui dois estratos, Windows e Linux.

Sistema	População	10%	Amostra
Windows	54	$\frac{10 \times 54}{100} = 5,4$	5
Linux	36	$\frac{10 \times 36}{100} = 3,6$	4
<b>Total</b>	<b>90</b>	$\frac{10 \times 90}{100} = 9$	<b>9</b>

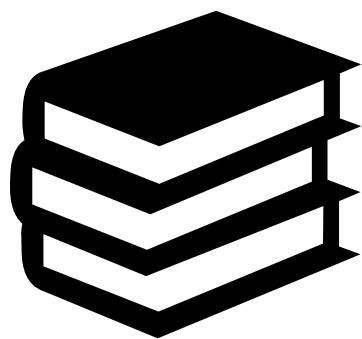
b) Numeramos os alunos de 01 a 90, sendo que de 01 a 54 correspondem ao que usam o Windows e de 55 a 90 os que usam o Linux.

Sorteamos 5 alunos que usam o Windows e 4 alunos que usam o Linux.

c) obtemos **uma amostra aleatória ou sistemática** de cada sistema operacional e reunimos as informações numa só amostra, denominada amostra estratificada.

Podemos então saber desses 90 alunos, qual é a porcentagem que está satisfeita com seu sistema operacional.





**Estudar** as páginas de 1 a 32 do livro **Estatística para cursos de engenharia e informática** (Pedro Alberto Barbetta, Marcelo Menezes Reis, Antonio Cezar Bornia), disponível na biblioteca virtual.



[univille.br](https://univille.br)