

# USO DA ESTATÍSTICA NO BIG DATE

Lucas Silva Prado Carvalho<sup>1</sup>  
Gleydiston Barreto De Oliveira<sup>2</sup>  
Guilherme Teixeira Dos Santos<sup>3</sup>  
Vitor Giovanny Santos Dias<sup>4</sup>



## RESUMO

Este artigo tem como objetivo apresentar como se integra e como serve o uso da estatística computacional no Big Data (Conceito que está associado a uma grande quantidade de dados), como ela é utilizada nas empresas e como pode auxiliar na tomada de decisões. O artigo vigente visa também expor as áreas que trabalham com essa grande quantidade de informações e que usam a estatística para controlar e ajudar na compreensão das grandes quantidades de dados que são gerados diariamente.

## PALAVRAS-CHAVE

Big data, negócios, estatística, computação

## ABSTRACT

This article aims to present how the use of computational statistics in Big Data is integrated, how it is used in companies and how it can help in decision making. The current article also aims to expose areas that work with this large amount of information and use statistics to control and assist in understanding the large amounts of data that are generated daily.

## KEYWORDS

Big Data. Business. Statistics. Computing

## 1 INTRODUÇÃO

O termo Big Data, bem comum na tecnologia da informação, é utilizado para descrever um grande volume de dados e, seu surgimento está associado a grande geração de informação que vem sendo produzida exponencialmente no mundo. A estatística computacional é uma área da estatística onde se desenvolvem modelos estatísticos em sistemas computacionais para auxiliar e apoiar decisões, como por exemplo negócios, marketing, setor financeiro, medicina, dentre outros. É um trabalho conjunto entre o estatístico e o programador. Os estatísticos são procurados quando existe a necessidade de desenvolver algum tipo de modelagem estatística dentro de um sistema onde não pode ocorrer falhas, pois são informações valiosas e de grande importância. Pode se destacar, como exemplo, o setor financeiro de uma empresa ou um banco.

É de se saber que a estatística é uma ciência nova na área da pesquisa, porém, na antiguidade já fazia operações de contagem populacional e era usada para obtenção de informações sobre os habitantes de uma determinada cidade. Os governantes da Europa Ocidental, depois da idade média, preocupados com as doenças da época e suas mortes por esse motivo, começavam a obter e armazenar informações. Com o passar do tempo e com o crescimento das nações no que se refere à política e comércio, surgiu o interesse em buscar informações sobre esses assuntos.

No item a seguir apresentamos a importância da estatística no desenvolvimento do chamado Big Data.

## 2 DESENVOLVIMENTO

O primeiro relato que se teve sobre o uso da estatística para obter informações de grandes quantidades de dados foi de 1663. John Graunt<sup>1</sup> utilizou uma grande

---

<sup>1</sup> Nascido em 1620, Graunt foi um cientista e demógrafo britânico que contribuiu com seus estudos e observações sobre taxa de mortalidade em Londres. Esta análise, posteriormente, veio a ser compreendida como estatística.

quantidade de informações, das mais diversas fontes, para estudar a epidemia da peste bubônica e a lista de mortalidade na Europa. Graunt, de um jeito bastante simples, conseguiu juntar e analisar fenômenos naturais os associando com números.

A estatística é uma ciência que não se utiliza somente de cálculos, porém colabora ao transformar dados obtidos em informações para diversas situações. A profissão de estatístico vem sendo valorizada em virtude do big data, nesse novo mundo o uso da estatística se tornou indispensável, devido à grande eficiência da utilização de metodologias que podem ser usadas para reverter dados em conhecimento necessários para a realização em tomada de decisão.

O termo "Big Data" foi introduzido pelo chefe da SGI<sup>2</sup>, o cientista John R. Mashey<sup>3</sup>, em 1998. Podemos descrever como uma coleção massiva de dados estruturados e não estruturados, disponibilidade e processamento de grandes volumes de dados de streaming em tempo real, em 2001 o analista Doug Laney<sup>4</sup> definiu esse termo como 3V: volume, velocidade e variedade.

O volume está ligado a grande quantidade de dados, a velocidade se refere ao processamento de dados e a variedade se refere ao aumento da complexidade das análises. Porém a literatura mostra que esse conceito está mais relacionado a 5V e são acrescentadas mais duas características. A veracidade que está diretamente ligada ao quanto uma informação é verdadeira e o valor que está relacionado com o valor obtido desses dados, ou seja, informação útil. O termo Big Data sempre foi um conceito relativo, pois seu tamanho depende de quem está usando os dados.

Com o passar do tempo a tecnologia vem se aprimorando cada vez mais e com isso adquire-se uma maior confiabilidade nas tomadas de decisões o que nas décadas passadas não era possível, uma vez que não possuíam equipamentos e tecnologia suficiente para uma análise confiável de informações.

O uso dos primeiros equipamentos para processar dados são de 1890, durante a realização do Censo dos Estados Unidos, conduzido pelo U.S. Census Bureau<sup>5</sup>. Somente no século XX que começaram a surgir os primeiros sistemas para armazenamento de informações. Em 1927, o engenheiro Fritz Pfleumer criou um método para guardar informações em fitas magnéticas.

O estatístico é o profissional mais indicado para trabalhar com dados e tecnologia. Com o grande aumento de informações fornecidas todos os dias pelos usuários de blogs, site, redes sociais etc..., tornou-se indispensável o uso da estatística para análise desses dados, isso implica relatar que o big data se torna mais complexo para ser analisado por causa da quantidade gigantesca de dados e isso faz com que os estatísticos desenvolvam um trabalho mais analítico e cauteloso. Por conta do grande volume de informações torna-se muito mais difícil de encontrar erros.

**2** Silicon Graphics, foi uma empresa americana de computação de alto desempenho onde produzia hardware e software.

**3** John R. Mashey, Ph.D em ciência da computação pela universidade do estado da Pensilvânia.

**4** D. Laney é analista e vice-presidente da equipe de pesquisa de data officer da Gartner. Aborda estratégia de dados, inovação de informações, uso do big data e análise.

**5** Agência governamental responsável pela colheita de dados dos Estados Unidos. Atualmente possui quatro regiões e nove divisões oficiais.

O interesse pela estatística vem aumentando muito que até alguns autores (HAWKINS, 1997; MOORE, 1997) defenderam a estatística como uma disciplina, pois afirmam que a sociedade moderna enfrenta problemas muito polêmicos de ordem social, ética e filosófica, que precisam utilizar dos pensamentos estatísticos para suas resoluções.

Todas as informações são estatísticas, obtidas por meio de médias, que ajudam nas tomadas de decisão. A área financeira está muito intimamente envolvida nestes problemas.

Pode-se definir a Educação Estatística como uma área de pesquisa, cujo objeto de estudo seriam os fatores que interferem no processo ensino-aprendizagem de Estatística; esta área busca o desenvolvimento das habilidades de solução para problemas e análise de dados, possibilitando o desenvolvimento do pensamento estatístico. Já o pensamento estatístico poderia ser definido como a capacidade de utilizar de forma adequada as ferramentas estatísticas na solução de problemas, de entender a essência dos dados e de fazer inferências. Reconhecer e compreender o valor da Estatística e ter a disposição para pensar de maneira probabilística. (CAZORLA, 2002. p. 18-19.)

As oportunidades oferecidas pelo big data são muitas, com isso requer mudança de metodologia e paradigmas. E isto desafia o profissional da estatística, porque atualmente o trabalho está voltado a limpar e organizar grandes bases de informações, ou seja, dados. Por isso a estatística tem ocupado cada vez mais esse espaço, quando se diz respeito a área de análise e coleta de dados.

### 3 CONCLUSÃO

Dado o exposto, podemos concluir que um dos trabalhos mais importantes nesta era da tecnologia é o trabalho do estatístico, pois sua profissão está diretamente ligada com a análise de grandes informações geradas pelo mundo. O cuidado que se deve ter com uma grande quantidade de dados é de fundamental importância, pois essas informações são essenciais nas decisões a serem tomadas pelas empresas e instituições. E, portanto, se faz mais do que necessário incrementar o estudo da estatística como uma disciplina nas instituições.

### REFERÊNCIAS

Big Data. **Wikipédia**, a enciclopédia livre. Disponível em: [https://pt.m.wikipedia.org/wiki/Big\\_data](https://pt.m.wikipedia.org/wiki/Big_data). Acesso em: 22 nov. 2018.

Big Data: O que é e sua importância. **SAS**. Disponível em: [https://www.sas.com/pt\\_br/insights/big-data/what-is-big-data.html](https://www.sas.com/pt_br/insights/big-data/what-is-big-data.html). Acesso em: 23 nov. 2018.

CASATTI, Denise. Como a era do Big Data impacta a carreira dos estatísticos. **Jornal da USP**, 17 de julho de 2017. Disponível em: <https://jornal.usp.br/universidade/como-a-era-do-big-data-impacta-a-carreira-dos-estatisticos/>. Acesso em: 22 nov. 2018.

Christiaan Huygens. **Wikipedia**, a enciclopédia livre. Disponível em: [https://pt.wikipedia.org/wiki/Christiaan\\_Huygens](https://pt.wikipedia.org/wiki/Christiaan_Huygens). Acesso em: 22 nov. 2018.

História da estatística. **Wikipedia**, a enciclopédia livre. Disponível em: [https://pt.wikipedia.org/wiki/Hist%C3%B3ria\\_da\\_estat%C3%ADstica](https://pt.wikipedia.org/wiki/Hist%C3%B3ria_da_estat%C3%ADstica). Acesso em: 22 nov. 2018.

John Graunt, o comerciante que inventou a estatística. **Impa**, 6 de setembro de 2018. Disponível em: <https://impa.br/page-noticias/john-graunt-o-comerciante-que-inventou-a-estatistica/>. Acesso em: 23 nov. 2018.

KAUR, Karambir; KAUR Harinderjit; SURBHI. Convergence of Big Data and Cloud Computing. **International Journal of Information and Technology** (IJIT), v. 2, Issue 3, maio-jun. 2016. Disponível em: <http://www.ijitjournal.org/volume-2/issue-3/IJIT-V2I3P1.pdf>. Acesso em: 22 nov. 2018.

SICA, José Rafael. **Big Data gera valorização da estatística**. IEL – Instituto de Estudos Avançados, Polo Ribeirão Preto, Universidade de São Paulo, 6 de junho de 2018. Disponível em: <http://sites.usp.br/iearp/big-data-gera-valorizacao-da-estatistica/>. Acesso em: 22 nov. 2018.

---

**Data do recebimento:** 21 de julho de 2016

**Data da avaliação:** 9 de novembro de 2016

**Data de aceite:** 12 de dezembro de 2017

---

---

1 Aluno do curso de Ciência da Computação – UNIT. Email: [lucas.prado@souunit.com.br](mailto:lucas.prado@souunit.com.br)

2 Aluno do curso de Ciência da Computação – UNIT. Email: [gleydiston.barreto@souunit.com.br](mailto:gleydiston.barreto@souunit.com.br)

3 Aluno do curso de Ciência da Computação – UNIT. Email: [guilherme.teixeira@souunit.com.br](mailto:guilherme.teixeira@souunit.com.br)

4 Aluno do curso de Ciência da Computação – UNIT. Email: [vitor.giovanny@souunit.com.br](mailto:vitor.giovanny@souunit.com.br)

