

lambda. It doesn't make sense to say what is the effect of this explanatory variable on a scale with you don't know. This is asking an undefined question. I can see that there maybe some difficulties but I don't think they are the ones that Bickel and Docksun addressed.

**Francisco:** Well, I hope you are enjoying your stay in Brazil.

**David:** Yes, very much indeed. Thank you very much.

## 6.2. ESTATÍSTICA: A TECNOLOGIA DA CIÊNCIA II

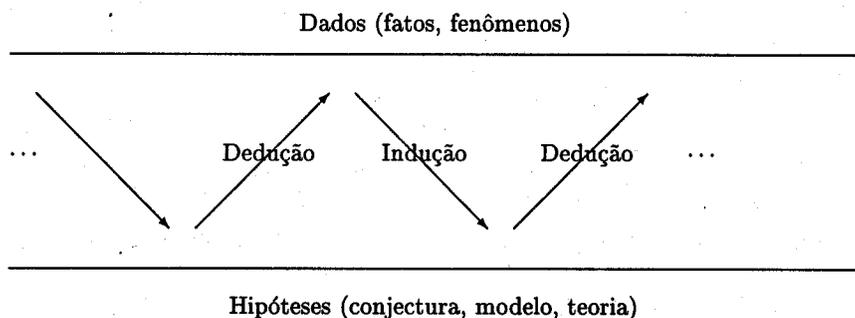
Basilio de Bragança Pereira  
FM, NESC e COPPE/UFRJ  
Basilio@nesc.ufrj.br

### Resumo

Este artigo é uma continuação de Pereira (1977). Apresenta os paradigmas do processo de pesquisa que utilizam a análise de dados. Indica como os avanços da computação influíram no processo, permitindo a análise de sistemas complexos e criando um novo paradigma.

### Estatística e pesquisa científica

No século XX, a concepção de ciência foi de aprendizado através da experimentação e dos dados observados, segundo o qual a procura das causas, das leis, traduz-se em um processo iterativo de observação do real, da realização de experimentos verificatórios e da avaliação quantitativa dos fenômenos em estudo. Uma hipótese inicial leva por um processo de dedução a certas conseqüências que são comparadas com os dados. Quando as conseqüências e os dados falham em concordar, as discrepâncias levam, por um processo de indução, a modificar a hipótese.



Uma parte do conhecimento adquirido é uma mera descrição do que observamos, a parte mais importante é a generalização ou indução que consiste em fazer inferências de experiências passadas para predizer futuras observações.

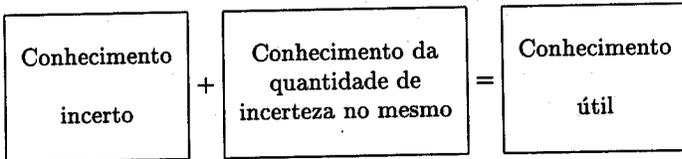
O paradigma da estatística neste contexto é o de constituir o conjunto unificado de métodos e de padrões de trabalho no planejamento e análise dos experimentos.

No final deste século XX, o progresso na computação e o advento de novos sistemas computacionais mudou a natureza dos dados e as possibilidades para sua análise. Novos métodos associados com grande volume de dados, bem como a possibilidade de modelar sistemas complexos criou um novo paradigma. A exploração de grande volume de dados para sugerir uma teoria, isto é, o sentido inverso do outro paradigma. O novo paradigma tem sido denominado: **Mineração de Dados (Data Mining)**. Temos então:

**Paradigma I** “Estatística: Teoria em busca de dados confirmatórios” .

**Paradigma II** “Data Mining: Dados em busca de uma teoria”.

Ambos os paradigmas então devem ser utilizados na busca de conhecimento traduzido na equação de Rao (Rao, 1997).



Em qualquer área o cientista utilizava modelos, que podiam ser **conceituais**, **físicos** ou **matemáticos** e os experimentos poderiam ser **in vitro** ou **in vivo**. Os avanços na computação nos fazem adicionar: **modelos computacionais** e experimentos **in silfco** (simulações, soluções numéricas, etc.).

Portanto, Estatística é uma mistura bem balanceada entre ciência e tecnologia.

Logo, cada vez mais é necessária a interação entre a teoria e aplicação, não só no nível da pesquisa como também profissional e ajudando nas duas tarefas: “do artista que vê a verdade no belo e a do cientista que vê o belo na verdade”.

#### Referências bibliográficas

- Diniz, C. A. e Louzada-Neto, F. (2000). *Data Mining: Uma Introdução*. Texto de Minicurso do 14<sup>o</sup> SINAPE (Simpósio Nacional de Probabilidade e Estatística).
- Pereira, B. de B. (1997). Estatística: a tecnologia da ciência. *Boletim da Associação Brasileira de Estatística*. Ano XIII, n<sup>o</sup> 37, 2<sup>o</sup> Quadrimestre, 27-35.
- Pereira, B. de B. (1995). Estatística em medicina: *p*-variação. *Revista da Sociedade de Cardiologia do Estado do Rio de Janeiro*, vol. VIII, n<sup>o</sup> 3, 73-78.
- Rao, C. R. (1997). *Statistics and Truth: Putting Chance to Work*. World Scientific, 2.ed.