

ET586 - Estatística e Probabilidade para Computação

Gustavo Farani de Farias

Recife, 4 de Maio de 2019

1 Introdução

A estatística pode ser definida como “a parte da metodologia da Ciência que tem por objetivo a coleta, redução, análise e modelagem dos dados”[3]. Ela representa uma ferramenta essencial em qualquer ramo de estudo que envolva o exame de quantidades massivas de dados: geografia, economia, psicologia, química, física, medicina e ciência da computação são alguns exemplos. Mais especificamente, conforme pontua Urdan [4], a estatística permite:

- Centralizar a experiência típica de um conjunto de pessoas, para além das particularidades em nível individual;
- Condensar lotes gigantescos de dados em resumos numéricos;
- Identificar as diferenças mais marcantes existentes entre subconjuntos distintos de um mesmo domínio;
- Investigar como diferentes aspectos (ou variáveis) de um sistema se correlacionam entre si.

A estatística é usualmente subdividida em duas: a descritiva e a inferencial. A frente descritiva estuda como os dados brutos podem ser sumarizados em informações mais facilmente compreensíveis e comunicáveis, seja na forma de medidas numéricas (como a média e a variância) ou de métodos gráficos (Figura 1). Já a inferencial estabelece métodos confiáveis para generalizar em cima de populações inteiras, baseando-se apenas na análise de amostras apropriadamente coletadas.



[1]

Figura 1: A vantagem de usar gráficos é que eles facilitam muito a visualização do conjunto de dados

2 Relevância

Existe uma forte relação mútua entre estatística e ciência da computação. Se, de um lado, o poder computacional trazido por esta fez com que aquela avançasse não apenas do ponto de vista técnico, como também do teórico [2], a primeira, por sua vez, foi responsável por firmar o sustentáculo metodológico que tornaria possíveis prodígios como a aprendizagem de máquina e a mineração de dados, dois dos mais relevantes campos de estudo da computação atual. Isso explica porque alguns autores, na verdade, acabam classificando mineração e aprendizagem como aplicações da estatística: “*Data analysis, machine learning and data mining are various names given to the practice of the statistical inference*” [5].

3 Relação com outras disciplinas

Disciplina relacionada	Explicação
IF670: MATEMÁTICA DISCRETA PARA COMPUTAÇÃO	Cobre conceitos fundamentais ao estudo de probabilidades e das variáveis aleatórias discretas, tais como teoria dos conjuntos, coeficientes binomiais e contagem
MA026: CÁLCULO DIFERENCIAL E INTEGRAL 1	Ferramentas de cálculo, tais como derivadas, integrais, máximos e mínimos são de importância crucial para o trabalho com variáveis aleatórias contínuas e suas distribuições probabilísticas
IF699: APRENDIZAGEM DE MÁQUINA	Grande parte de seus algoritmos se baseiam em conceitos estatísticos, tais como variância, regressão linear e valor-p
IF795: SISTEMAS DE SUPORTE A DECIS. MIN. DADOS	Explora muitos métodos estatísticos avançados, e.g., regressão logística, classificação, <i>bootstrapping</i> , entre outros

Referências

- [1] Image courtesy of jscreationzs at FreeDigitalPhotos.net.
- [2] C. Lauro. Computational statistics or statistical computing, is that the question? *Computational Statistics & Data Analysis*, 23:191-193, 1996.
- [3] Wilton O. Bussab. Pedro A. Morettin. *Estatística Básica*. Editora Saraiva, 2010.
- [4] T.C. Urdan. *Statistics in Plain English, Third Edition*. Taylor & Francis, 2011.
- [5] L. Wasserman and L.A. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004.