

SEMINÁRIOS DA COMPUTAÇÃO

EDIÇÃO 2016

1

**PH.D. MARCEL DE TOLEDO
VIEIRA**

*Professor Associado II do
Dept. de Estatística/UFJF*

marcel.vieira@ice.ufjf.br

"Sampling Methods in the Big Data Era: Challenges and Implications"

Statistical inference using registers and administrative data sources is a very recent research topic that is attracting the attention of survey sampling methodologists. Big Data is not a new concept for survey statisticians, and particularly for those who work on official censuses and surveys. The role of surveys as traditional data gathering exercise in many parts of the world is changing, and we are now seeing a greater interest to consider surveys as one among many ways of collecting and analysing data for providing information to the society. In a world of complementary data sources, the role of survey and census methodologists needs to change to accommodate the new design and inference problems that are arising. Another challenge is the integration of surveys and censuses with external data sources. Survey methodologists are invited to develop design and inference methods for a future where official statistics will be produced from a variety of data sources, with surveys not necessarily constituting one of these sources. The following issues will be discussed in this presentation: as you have the entire big data set does it mean you should always use it? is it worth the effort, time and computing costs for analysing big data over parallel computers? would our probability of being closer to the truth when analysing the entire population versus analysing the sample differ substantially? or are the current computer capability enough for us to consider sampling less important?

2

FELIPE LEITE FAGUNDES

Mestrando do PGCC/DCC

felipe.fagundes@ufjf.edu.br

"Aprendizado de Métrica Utilizando uma Função de Distância Parametrizada e o Algoritmo K-means"

O Aprendizado de Métrica tem sido de grande utilidade no desenvolvimento de algoritmos de Aprendizado de Máquinas, sobretudo para a solução de problemas relacionados à análise de *cluster*. Este problema de aprendizado foi originalmente formulado como um problema de otimização, com o objetivo de minimização de um conjunto parametrizado de distâncias de Mahalanobis, sujeito às restrições de não negatividade e de desigualdade triangular. Nesta apresentação será descrita a proposta de um novo método de aprendizado de métrica, que minimiza um conjunto parametrizado de distâncias Euclidianas e utiliza o algoritmo K-means para a geração de potenciais candidatos.

Data: 10 de Novembro de 2016.

Horário: 10H00.

Local: Auditório de DCC/Estatística (prédio do DCC/Estatística - UFJF).