

## **Título do curso :** Astroestatística com R

Objetivo: Fazer uma introdução aos problemas estatísticos mais encontrados na Astrofísica atual fazendo uso do software estatístico R.

### Contexto:

Cada vez mais o pesquisador em Astrofísica encontra novos e complexos problemas estatísticos, cuja resolução requer o uso de softwares estatísticos computacionalmente potentes que possibilitem o tratamento de grandes conjuntos de dados, e a implementação de códigos de cálculos matemáticos e gráficos. Em particular, na era dos grandes levantamentos do céu feitos com fotometria multibanda, são muito comuns problemas de amostra, análise de dados multivariante, regressão linear ou não linear, probabilidade bayesiana, ou ajuste de modelos, entre outros. Por isso, é importante facilitar aos novos astrônomos um maior entendimento desses problemas estatísticos que irão encontrar no seu trabalho, assim como de um software livre que possam usar para a resolução desses problemas.

Teresa Aparicio Villegas é graduada em Matemática, e a sua tese de doutorado implicava o desenvolvimento de algoritmos estatísticos para o tratamento dos dados de um levantamento do céu com 20 bandas fotométricas. Ademais, durante o período de sua formação, ela participou do "Summer School in Statistics for Astronomers V" no "Center for Astrostatistics" no estado de Pennsylvania, com especialistas do campo, onde especializou sua formação na Astroestatística.

### Conteúdo:

O curso pretende ter um caráter fundamentalmente prático, e o conteúdo vai se ajustar às necessidades gerais dos estudantes e pós-doutorandos do ON. Para isso, as aulas vão ser divididas em uma parte teórica, para dar uma introdução ao problema matemático envolvido, e uma parte prática usando o software R. Se recomendará o uso de laptop para o curso.

O curso seria de 2 créditos, o que equivale a 30 horas de aula.

### A lista de matérias é:

Introduction to R and data exploratory (3h)

Basic Probability (3h)

Inference I: Estimation, Confidence Intervals & Hypothesis Testing (3h)

Inference II: Maximum Likelihood Estimation, the Bayesian Information Criterion (3h)

Correlation & Regression (3h)

Model Fitting and Model Selection (3h)

Methods of Multivariate Analysis (PCA) (3h)

Bootstrap , Markov chain Monte Carlo (3h)

Bayesian Analysis (3h)

Classification and regression analysis with Supervised learning algorithm : Support Vector Machine, Neural Networks (3h)