

Proposta de disciplina do PPGFis
FIP00000 - Astronomia E Astrofísica: Astroestatística

- **Semestre:** 2023/2
 - **Carga horária semanal:** 4
 - **Créditos:** 4
 - **Pré-requisitos:**
 - **Professor/Responsável:** José Eduardo da Silveira Costa
-

Súmula

Astroestatística. Ferramentas computacionais para análise e visualização de dados. Conceitos fundamentais de probabilidade e estatística. Distribuições. Testes de hipótese. Correlação. Modelagem de dados. Métodos de Regressão. Análise de Resíduos. Propagação de Erros. Análise bayesiana. Séries temporais.

Objetivos

O objetivo principal da disciplina é apresentar ao estudante uma ampla gama de métodos estatísticos computacionais para uso em análise de dados e na resolução de problemas em Astronomia, Astrofísica e áreas correlatas.

Programa

Introdução: Estatística / astroestatística. Recursos computacionais. Python como linguagem de programação para análise estatística de dados. IDEs e notebooks.

Dados: Dados e metadados. Dados e informação. Leitura de dados. Estruturas de dados. Limpeza de dados.

Conceitos fundamentais de probabilidade e estatística: Variáveis. Tipos de variáveis. População e amostra. Estatística descritiva. Estatística populacional e amostral. Distribuições. Índices de tendência central: médias, mediana e moda. Índices de dispersão: variância,

desvio padrão, quantil, quartil, percentil. Boxplot. Definição de probabilidade. Distribuições de probabilidade. Distribuições de densidade de probabilidade (CDF e ECDF) e de probabilidade acumulada (PDF). Aplicações em Astronomia.

Distribuições de probabilidade: distribuições de probabilidade binomial, poissônica e normal. Quando usar estatística poissônica? Outras distribuições. RMS. Padronização de dados. Razão sinal-ruído (SNR). Coeficiente de variação. Momentos centrais estatísticos. Testes de normalidade. Aplicações.

Representação gráfica de distribuições: histogramas. Critérios de Sturges, Scott e Friedman-Diaconis. Boxplots. Aplicações.

Testes de hipótese. Hipótese nula e hipótese alternativa. Erros de Tipo I e de Tipo II. Testes paramétricos e não-paramétricos. Teste-t de Student. Teste de Wilcoxon. Teste-F, Teste de Brown-Smythe, Teste de Bartlett. Teste de Kruskal-Wallis. Teste de Dunn. Teste de Kolmogorov-Sminnov. Teste de Anderson-Darling. Aplicações.

Comparação de amostras. Testes estatísticos aplicáveis. Métodos gráficos. Aplicações.

Análise de correlação: investigando possíveis correlações entre variáveis. Associações. Correlação e anticorrelação. Medidas paramétricas e não-paramétricas de correlação. Coeficientes de correlação de Pearson, Sperman e Kendall. Aplicações.

Modelagem de dados: inferindo ou ajustando modelos. Regressão. Modelos de regressão. Modelos lineares e não-lineares. Métodos de regressão paramétricos, não-paramétricos e semi-paramétricos. Regressão linear generalizada. Regressão ortogonal. Regressão de Deeming generalizada. Graus de liberdade. Princípio da parcimônia. Maximum likelihood. Método de mínimos quadrados. Aplicações.

Análise de resíduos: Modelos vs. dados: medida da qualidade de um ajuste. Coeficientes de determinação. Resíduos padronizados. Histograma de resíduos. Q-Q plot. A influência dos dados (leverage). Distância de Cook. Avaliação visual de um ajuste. Chi-quadrado. Aplicações.

Pesos: definição de peso de uma medida. Critério de máxima verossimilhança (maximum likelihood). Aplicações.

Propagação de erros: como as incertezas se propagam nos cálculos? Teoria de propagação de incertezas (erros). Estimativa empírica. Combinação de dados. Métodos de re-amostragem: jackknife, bootstrap, métodos de permutação. Monte Carlo. Aplicações.

Estatística Bayesiana. análise bayesiana. Teorema (ou regra) de Bayes. Hipótese. Evidência. Probabilidade a priori. Probabilidade a posteriori. Probabilidade marginal. Aplicações.

Séries temporais: análise de Fourier. Periodogramas. Método (O-C). Aplicações.

Método de Trabalho

O curso será de natureza prática, focando na aplicação de uma ampla gama de métodos estatísticos computacionais, através do uso de módulos da linguagem Python de Programação. Cada estudante deverá reproduzir no computador as demonstrações apresentadas pelo professor e resolver exercícios propostos durante as aulas.

Cada estudante deverá apresentar dois mini-seminários: o primeiro, na parte inicial do curso, apresentando o problema central de seu projeto de pesquisa e os aspectos estatísticos envolvidos; o segundo, na parte final do curso, mostrando aplicações de métodos estatísticos em seu projeto de pesquisa.

Serão propostos trabalhos e/ou listas de exercícios que deverão ser enviados dentro de prazos pré-definidos.

Ao final da disciplina, será realizada uma prova, envolvendo a aplicação de métodos estatísticos estudados ao longo do curso.

O Moodle será utilizado como plataforma de apoio à disciplina e, eventualmente, algumas aulas poderão ser ministradas no modo EAD, dentro do limite de 20disciplina.

Avaliação

A nota final (NF) do curso será a soma das notas obtidas nos seguintes itens: (a) atividades realizadas em sala de aula: até 6,0 (seis) pontos; (b) listas de exercícios e seminários: até 3,5 pontos; (c) prova final: até 1,5 ponto.

Será atribuído conceito A ao estudante que obter nota final maior ou igual a 9,0 (nove) pontos; conceito B ao estudante que obter nota final entre 7,5 (inclusive) e 9,0 pontos e conceito C ao estudante que obter nota final entre 6,0 (inclusive) e 7,5 pontos.

Ao estudante que ficar com nota final inferior a 6,0 (seis) pontos será possibilitada a realização de um EXAME cobrindo todo o conteúdo da disciplina e valendo até 10 (dez) pontos. Se a nota obtida no exame for igual ou superior a 6,0 (seis) pontos, o estudante será aprovado com conceito C, caso contrário, será reprovado com conceito D.

Bibliografia

BABU, G.J., FEIGELSON, E.D. *Astrostatistics: Interdisciplinary Statistics*. London: Chapman and Hall, 1996. 224p

BEVINGTON, P. R. *Data Reduction and Error Analysis for the Physical Sciences*. 3rd. Edition, McGraw-Hill Education, 2002

DALGAARD, P. *Introductory Statistics with R*. Springer. 2008, 357p.3

FEIGELSON, E.D., BABU, G.J. *Modern Statistical Methods for Astronomers: With R Applications*. Cambridge: Cambridge University Press, 2012, 476p.

HASLWANTER, T. *An Introduction to Statistics with Python*. Springer International Publishing, 2016.

IVEZIC, Z., CONNOLLY, A. J., VANDERPLAS, J. T., GRAY, A. *Statistics, Data Mining, and Machine Learning in Astronomy*. Princeton University Press, 2014, 533p

LIST, L. *Statistical Methods for Data Analysis in Particle Physics*. Springer Lecture Notes in Physics 909, 2016.

NELLI, F. *Python Data Analytics : Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language*. Berkeley, 2015.

PRESS, W. H; TEUKOLSKY, S. A.; VETTERLING, W. T. *Numerical Recipes 3rd edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

SARRO, L. M., EYER, L., OMULLANE, W., DE RIDDER, J. *Astrostatistics and Data Mining*. New York, NY: Springer New York, 2012. 270p. (Spring Series in Astrostatistics)

SIVIA, D. S.; SKILLING, J. *Data Analysis: A Bayesian Tutorial*. 2nd. Edition, Oxford University Press, 2006.

TAYLOR, J.R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, 1997, 327p.

TRUMPLER, R.J., WEAVER, H.F. *Statistical Astronomy*. California: University of California Press, 1953. 646p.

VENABLES, W. N., SMITH, D. M., R Development Core Team, *An Introduction on R*. 2009. 94p.

WALL, J.V., JENKINS, C.R. *Practical Statistics for Astronomers*, Cambridge: Cambridge University Press, 2012. 344p

Referências bibliográficas adicionais poderão ser indicadas durante o curso.

Observações

Em semestres anteriores, introduzi e ministrei a disciplina FIP20129 (Tópicos de Astronomia e Astrofísica: Astroestatística) de 2 créditos, utilizando o ambiente R para estatística computacional como plataforma de trabalho e com o conteúdo programático adaptado à carga horária de 30 horas. Mesmo utilizando os recursos do ambiente R e as metodologias existentes para o ensino de estatística neste ambiente, ficou evidente que uma carga horária de 30 horas é insuficiente para cobrir com a devida profundidade, um conteúdo programático tão diverso e abrangente.

A presente proposta de disciplina difere em dois aspectos importantes: (1) o uso de plataforma de trabalho baseada em Python, linguagem de programação mais amplamente utilizada pelos nossos PGs do que o R, porém com metodologias de ensino adaptadas do R para esta plataforma e (2) carga horária de 60 horas (4 créditos) para melhor distribuição e aprofundamento dos conteúdos estudados.