



## 16º Congresso Nacional de Iniciação Científica

**TÍTULO:** FERRAMENTAS COMPUTACIONAIS PARA ANÁLISES ESTATÍSTICAS DE GALÁXIAS

**CATEGORIA:** CONCLUÍDO

**ÁREA:** CIÊNCIAS EXATAS E DA TERRA

**SUBÁREA:** COMPUTAÇÃO E INFORMÁTICA

**INSTITUIÇÃO:** CENTRO UNIVERSITÁRIO DAS FACULDADES METROPOLITANAS UNIDAS

**AUTOR(ES):** MARCO PAULO RIBEIRO MEDEIROS

**ORIENTADOR(ES):** AIRTON CARLOS ALMEIDA BORGES

Realização:

**SEMESP**

sindicato das mantenedoras de ensino superior



Apoio:

  
**ENIAC**  
Educação Básica e Superior

**Resumo.** O presente trabalho mostra como ferramentas computacionais podem ser utilizadas em análise de dados científicos. No trabalho, estudos de fenômenos astronômicos são realizados com a ajuda de planilhas eletrônicas e plataforma de tratamento e análise de dados.

O estudo de cores de galáxias para a determinação de informações sobre tipos de populações estelares e características do meio ambiente onde elas se encontram foi feito utilizando dados do “Sloan Digital Sky Survey” (SDSS). A seleção de informações das galáxias foi realizada por meio de programa em linguagem Python em um banco de dados de 88 TB. Uma vez selecionados, esses dados foram analisados estatisticamente, utilizando como ferramenta computacional a linguagem R. A determinação dos parâmetros estatísticos e a elaboração e análise de gráficos e histogramas, representando as propriedades das galáxias, permitiram estudar, por exemplo, diferenças nas populações estelares devido as influências de seu meio ambiente, representadas pela interação gravitacional com outras galáxias.

## **1 – Introdução**

As galáxias são, depois de estrelas e planetas, os corpos celestes mais conhecidos e estudados na Astronomia. Suas características são variadas. Forma, tamanho, quantidade e tipos de estrelas que contém são algumas destas características e propriedades. Elas são importantes para a Astronomia tanto por seu valor intrínseco quanto para o entendimento da criação e evolução do Universo. O primeiro caso é mais objeto da Astrofísica enquanto a Cosmologia trata do segundo. Informações sobre galáxias fornece indicações sobre o que acontece no Universo e vice-versa.

Toda informação que se tem sobre as galáxias vem da luz que produzem e que chegam aos telescópios. Existem diferentes técnicas para observação de galáxias (e de outros objetos), que estão associadas ao modo de capturar a luz e aos instrumentos utilizados. Um exemplo é a fotometria, onde se obtém uma imagem do objeto, permitindo, através de tratamento adequado, que se conheça a quantidade total de luz recebida, denominada luminosidade, representada por sua magnitude (ver abaixo) ou ainda, fazendo com que essa luz passe por um filtro óptico, obtendo assim a magnitude da galáxia em determinada cor. Outra técnica é a espectroscopia, onde a luz do objeto passa por uma rede de difração (análogo a um

prisma) onde ela é decomposta diferentes comprimentos de onda, resultando no seu espectro eletromagnético.

Dependendo da técnica utilizada diferentes características de galáxias são evidenciadas, e podem ser utilizadas para estudos específicos.

Neste trabalho foram utilizadas informações fotométricas, especificamente, as magnitudes aparentes obtidas com diferentes filtros (cores) para o estudo do conteúdo estelar das galáxias – suas populações estelares. O objetivo principal foi o de entender como o meio ambiente influencia as características da população. O meio ambiente aqui é a proximidade física entre duas galáxias, formando um par. No estudo foram comparadas as populações estelares de galáxias em pares com galáxias isoladas. Para isso amostras de objetos, em pares e isolados, foram retiradas do Sloan e comparadas com auxílio de histogramas e medias estatísticas obtidas utilizando a plataforma R.

No próximo capítulo é apresentado uma breve explicação dos conceitos astronômicos e das grandezas físicas utilizadas no estudo. O capítulo 3 é dedicado à explicação do método de seleção dos objetos que formam o banco de dados. No capítulo 4 são apresentados os dados bem como a análise estatística. As conclusões são apresentadas no capítulo 5. As referências utilizadas são relacionadas no capítulo 6.

## **2 – Magnitudes e Índices de Cor de Galáxias**

De uma forma simples e geral, as galáxias podem ser estudadas em termos do seu conteúdo estelar. Ou seja, os tipos de estrelas de que são constituídas determinam, de certa maneira, suas propriedades gerais tais como luminosidade (brilho) e cores. Por exemplo, galáxias elípticas são formadas por estrelas mais velhas, nos estágios avançados de evolução, por isso, emitem luz em maior quantidade em comprimento de ondas mais longos (mais para o lado vermelho do espectro eletromagnético); já as espirais tem população estelar onde predominam estrelas jovens (estágios iniciais de evolução) responsáveis por uma maior emissão de luz na parte azul e violeta.

As características estelares acima podem ser modificadas caso uma galáxia passe por processos que alterem sua evolução. Um desses processos ocorre quando ela

interage gravitacionalmente com outras galáxias. Se esse é o caso, uma galáxia formada por estrelas mais velhas pode experimentar um processo de criação de novas estrelas, alterando assim sua cor. O presente trabalho estudou as possíveis alterações de cores das galáxias para confirmar se o critério de seleção adotado pode ser o primeiro a ser adotado para a identificação de pares.

Aqui as cores das galáxias são representadas por magnitudes, quantidade de luz coletada em um telescópio, vinda de um objeto e que passa através de um filtro óptico. Por exemplo, a magnitude azul (B de “blue”) de um objeto é a quantidade de luz azul que passa por esse filtro. Existem diferentes sistemas de filtros de luz utilizados em Astronomia. O filtros utilizados nas observações para o obtenção das magnitudes no SDSS são: u (3551Å), g (4686Å), r (6166Å), i (7480Å) e z (8932Å), onde os valores entre parênteses o comprimento de onda central do filtro, medido em angstrom ( $1 \text{ \AA} = 10^{-10} \text{ m}$ ). A figura 1 mostra os filtros do SDSS em termos do intervalo de comprimentos de onda que abrange e de sua sensibilidade (resposta).

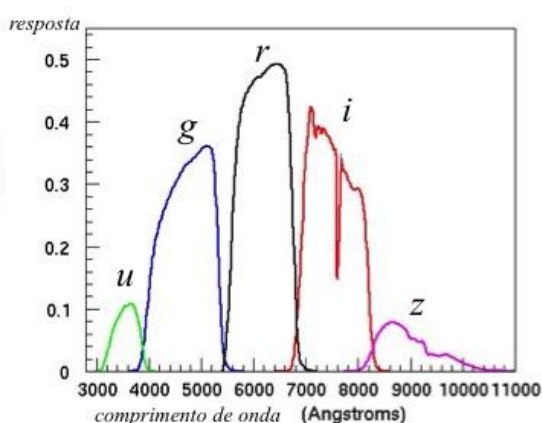


Figura 1 - Sistema de filtros utilizado no SDSS.

Como explicado acima, estrelas mais jovens devem emitir mais luz na parte azul do espectro do que na parte vermelha. Assim, a diferença entre a magnitude no azul e no vermelho é uma indicação do tipo de estrela que predomina na galáxia (que caracteriza sua população estelar). A essa diferença denomina-se “índice de cor”. Quanto maior for índice de cor mais luz azul é recebida da galáxia. Assim, o índice de cor pode ser considerado como um indicador do tipo de estrela que predomina em uma galáxia.

Neste trabalho foram utilizados nas análises os índices de cor (u-r), (u-z) e (g-r).

### 3 – Seleção dos Dados

As amostras das galáxias, tanto em pares quanto isoladas, foram obtidas segundo o critério definido em Rossi, R. et al (2015) a partir do catálogo do Sloan Digital Sky Survey (SDSS) (Ahn, 2014). No presente estudo foram considerados dois tipos de galáxias, a saber, espirais e elípticas.

O SDSS é um projeto para mapear objetos celestes de um quarto de todo o céu medindo diferentes propriedades destes. Para esse trabalho foram selecionadas as magnitudes, obtidas com diferentes filtros (ver abaixo) e a posição, representadas pela coordenadas equatoriais do objeto em questão.

O critério para a definição e seleção de um par de galáxias está baseado na distância aparente entre duas galáxias. A partir da posição dos dois objetos (dadas pelas suas coordenadas) pode-se determinar a distância aparente. Duas galáxias são consideradas formando um par se estão separadas por determinado valor máximo. Um algoritmo escrito em Python faz a pesquisa no banco de dados e calcula a distância entre duas galáxias. Se o critério é cumprido o par é selecionado e as propriedades de interesse são gravadas em um arquivo no formato “.csv”. A tabela no Anexo 1 mostra uma parte do catálogo dos pares. O Anexo 2 apresenta o algoritmo de escolha do par. Esse algoritmo foi objeto de estudo do aluno e apresentado como seminário para o orientador.

Foram selecionados mais de 41.000 pares segundo esse critério. Nesses pares 5.114 galáxias são espirais e 1.416 são elípticas. A galáxias restantes nos pares não têm classificação confirmada.

Para comparação das propriedades foram selecionadas todas as galáxias elípticas e espirais (com classificação confirmada) não pertencentes aos pares. Essas galáxias são aqui denominadas de “isoladas”. Dado o grande número desses objetos, para fins de comparação, foram produzidas amostras com 5.000 objetos.

As três amostras, dos pares, das elípticas e das espirais isoladas, constituem três arquivos em formato .xlsx e formam o banco de dados de trabalho.

Cada um dos arquivos apresenta as coordenadas, as 5 magnitudes (cores) u, g, r, i, z (explicadas acima) e os índices que cor (u-r), (u-z) e (g-r). Dadas as informações que estes três índices de cor fornecem sobre as propriedades estelares de galáxias eles foram escolhidos para análise no trabalho.

## 4 – Análise dos Dados

Para a seleção dos dados, como já foi dito acima, foi utilizado a linguagem Python. Python é uma linguagem de programação de alto nível, ou seja, mais próxima a linguagem humana, isso facilita o trabalho da leitura do código escrito. É uma linguagem fortemente “tipada”, isto é que apesar de não ser declarado um tipo de dado para a variável, ela não permite que seja realizado operações com tipos diferentes. Isso evita certos erros, permitindo, quando se trata de seleção de dados, buscar os dados sem que se saiba exatamente seus tipos, mas não permite realizar operações caso sejam de tipos diferentes, evitando falhas.

Como os dados contidos no “*Sloan*”, apesar de serem valores, são diferentes, por exemplo: coordenadas, magnitudes, tipo de galáxia, daí a necessidade da linguagem Python para realizar essa seleção, podendo selecionar esses dados sem um tipo declarado.

Realizada a seleção dos dados a análise estatística dos dados foi feita utilizando-se a linguagem R. Isso devido a grande quantidade de itens analisados, cálculos feitos e gráficos gerados para a realização da comparação entre os valores. A linguagem R é uma linguagem criada especificamente para cálculos estatísticos e gráficos, é uma linguagem bastante simples na criação do *script*, mas muito poderosa. Ela realiza cálculos de forma rápida e exata, é uma linguagem utilizada muito por “*data miners*” (profissionais que trabalham no processo de exploração de uma grande quantidade de dados à procura de padrões – *Big Data*) e estatísticos na análise de dados e criação de softwares estatísticos.

No Anexo 3 pode ser visto um script em R, criado para realizar todos os cálculos e histogramas mostrados abaixo:

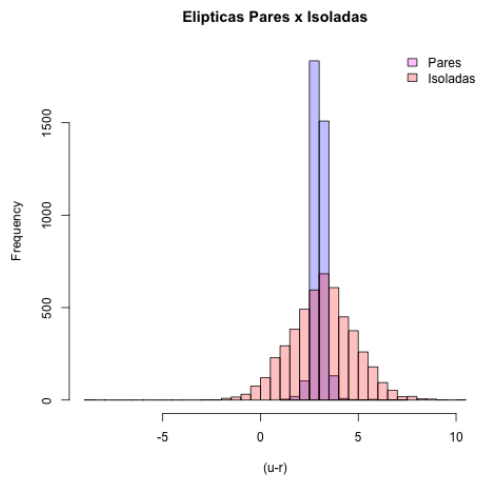


Figura 2a – Índice de cor (u-r) de galáxias elípticas Isoladas x Pares.

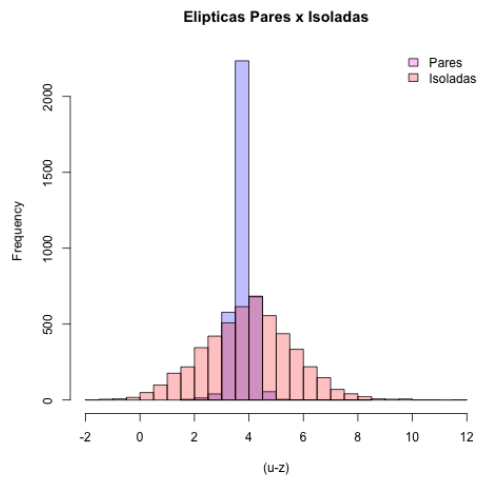


Figura 2b - Índice de cor (u-z) de galáxias elípticas Isoladas x Pares.

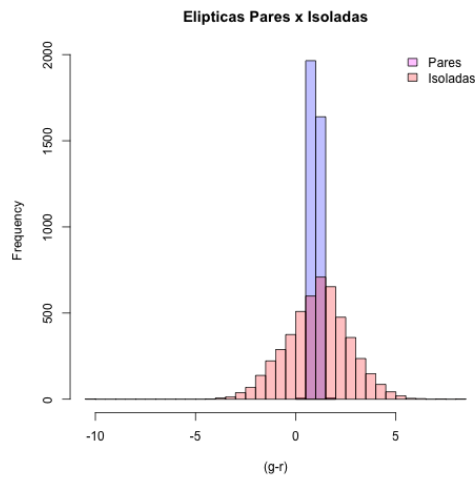


Figura 2c – Índice de cor (g-r) de galáxias elípticas Isoladas x Pares.

Elípticas (U-R)						
	Mediana	1º quartil	3º quartil	Media	Variância	Desvio Padrão
Par	2,98	2,83	3,14	2,99	0,099	0,31
Isolada	3,19	2,04	4,21	3,15	2,801	1,67

Elípticas (U-Z)						
	Mediana	1º quartil	3º quartil	Media	Variância	Desvio Padrão
Par	3,76	3,58	3,96	3,77	0,131	0,36
Isolada	4,00	2,90	5,01	3,96	2,740	1,66

Elípticas (G-R)						
	Mediana	1º quartil	3º quartil	Media	Variância	Desvio Padrão
Par	0,98	0,90	1,09	1,01	0,021	0,15
Isolada	1,17	0,10	2,13	1,11	1,472	1,57

Figura 2d – Dados obtidos com a linguagem R de galáxias elípticas nos índices de cores u-r, u-z e g-r.

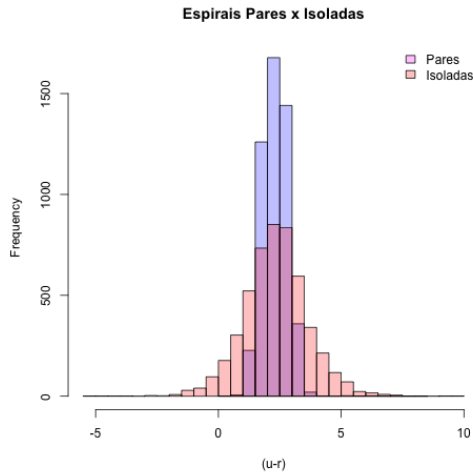


Figura 3a – Índice de cor u-r de galáxias espirais Isoladas x Pares.

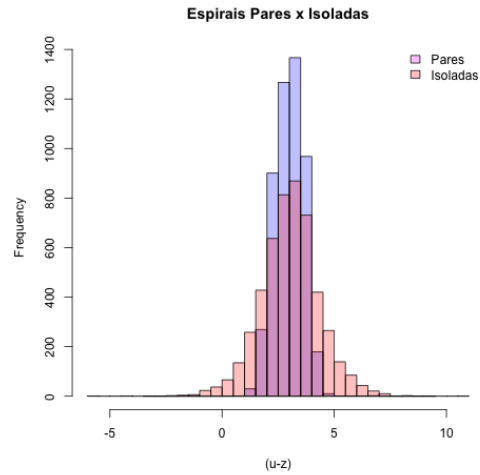


Figura 3b - Índice de cor u-z de galáxias espirais Isoladas x Pares.

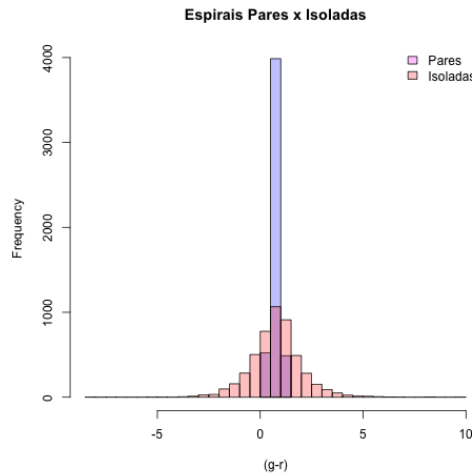


Figura 3c – Índice de cor g-r de galáxias espirais Isoladas x Pares.

Espirais (U-R)						
	Mediana	1º quartil	3º quartil	Media	Variância	Desvio Padrão
Par	2,29	1,93	2,68	2,30	0,262	0,51
Isolada	2,33	1,54	3,09	2,34	1,701	1,30

Espirais (U-Z)						
	Mediana	1º quartil	3º quartil	Media	Variância	Desvio Padrão
Par	3,01	2,52	3,47	2,99	0,412	0,64
Isolada	3,04	2,23	3,80	3,03	1,691	1,30

Espirais (G-R)						
	Mediana	1º quartil	3º quartil	Media	Variância	Desvio Padrão
Par	0,77	0,62	0,90	0,76	0,037	0,19
Isolada	0,78	0,10	1,42	0,78	1,544	1,24

Figura 3d – Dados obtidos com a linguagem R de galáxias espirais nos índices de cores u-r, u-z e g-r.



Nota-se através dos histogramas e dos parâmetros estatísticos, que a variância dos índices de cores das galáxias em pares é bem menor do que das galáxias isoladas, ou seja, existe uma maior concentração de dados em um intervalo menor nas galáxias em pares enquanto existe uma variação maior em um intervalo maior nas galáxias isoladas.

Esta informação nos apresenta o seguinte cenário:

- Galáxias em pares possuem índices de cores positivos, caracterizando populações de estrelas predominantemente azuis, portanto mais jovens.
- Galáxias isoladas possuem índices de cores com variação mais ampla, caracterizando população de estrelas de diferentes cores e idades.

Com estas observações, os dados mostram que o critério escolhido para definir os pares realmente condiz com o que é de conhecimento sobre a interação dos mesmos. Os dados das galáxias em pares indicam que há uma interação física entre os membros do par fazendo surgir novas estrelas.

## **5 – Conclusões**

Pode-se dizer que as ferramentas computacionais são atualmente essenciais na análise de qualquer informação, no nosso caso análise de populações estelares. Através da linguagem R foi possível verificar que o critério de seleção pares de galáxias é válido podendo se utilizado em conjunto com outros.

A metodologia de seleção e análise de dados aqui utilizada se mostrou muito eficiente e simples, permitindo que um trabalho de qualidade possa ser realizado em períodos razoáveis com bom grau de confiabilidade.

## **6 – Referências**

Ahn, Christopher P. et al. The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment.

Astrophysics Journal Supplements Series; V. 211, 17A – 2014

Rossi, R.; Nascimento, A. C.; Barbosa, W.; De Mello, D. F.; Borges, A. C. A.; Goya, M.; Puga, S.

Colliding Galaxies in the Big Data of the Huge Universe (BIDHU) project; American Astronomical Society, AAS Meeting #225, #143.32.