

1 Introdução

Aglomerados de galáxias são as maiores estruturas do Universo observável que podem ter alcançado o estado de equilíbrio dinâmico. Eles são constituídos por algumas dezenas até milhares de galáxias ligadas pela força gravitacional. Para que as galáxias se mantenham próximas umas das outras por longas escalas de tempo é necessário que exista uma considerável força gravitacional que impeça a sua dispersão no espaço. Isto significa que a massa típica desses sistemas é muito grande (o catálogo de Abell, por exemplo, possui sistemas com massa total por volta de 10^{14} a 10^{15} massas solares), o que torna o mapeamento de aglomerados e a correta obtenção de suas propriedades observadas de grande importância tanto para estudos referentes ao processo de formação de estruturas no Universo, como para restringir o modelo cosmológico atual (Velásquez, 2007).

Ao longo do século XX muitos trabalhos foram voltados ao estudo de aglomerados e permitiram a sua crescente caracterização como um sistema físico bastante particular. Dessa maneira, propriedades como: a distribuição de posições e de velocidades em aglomerados, perfis de densidade numérica – entendimento quanto à distribuição das galáxias em relação a sua distância ao centro do aglomerado –, funções de luminosidade – quantificar a distribuição luminosa – e de massa, puderam ser obtidas e utilizadas para descrever aglomerados (Velásquez, 2007). Além de parâmetros dinâmicos e cinemáticos como: velocidade média (adequado a estimativa de distâncias); dispersão de velocidades (compreensão do grau de ligação gravitacional entre galáxias); massas (entender o grau de contribuição de densidade referente a massa total do Universo); e a razão massa/luminosidade (utilizada como indicativo da quantidade matéria escura e para estimar se a distribuição de luz segue a da matéria) (Friaça et al., 2008).

As conclusões alcançadas por esses trabalhos sugerem que os aglomerados iniciaram o seu processo de formação há aproximadamente 10 bilhões de anos, processo que se dá de forma continuada até os dias de hoje. Estudos baseados na distribuição de velocidades de galáxias nesses sistemas indicam que apenas uma fração deles (em torno de 60%) pode ser considerada em estado de equilíbrio dinâmico. Os demais constituem sistemas ainda em formação ou perturbados por interações com outros aglomerados. O grande número de sistemas fora do equilíbrio pode introduzir dificuldades na interpretação das propriedades dinâmicas dos aglomerados (Friaça et al., 2008).

Portanto, o entendimento preciso dos graus de liberdade de aglomerados é de extrema importância para que seja possível realizar inferências dinâmicas sobre esses sistemas. Um dos aspectos menos estudados a respeito de aglomerados é a possibilidade de que eles tenham algum grau de rotação. Não levar em consideração a rotação de

aglomerados pode gerar um erro nas suas estimativas de massa, o que afeta diretamente as restrições cosmológicas fornecidas pela função de massa desses sistemas (Fang et al., 2008).

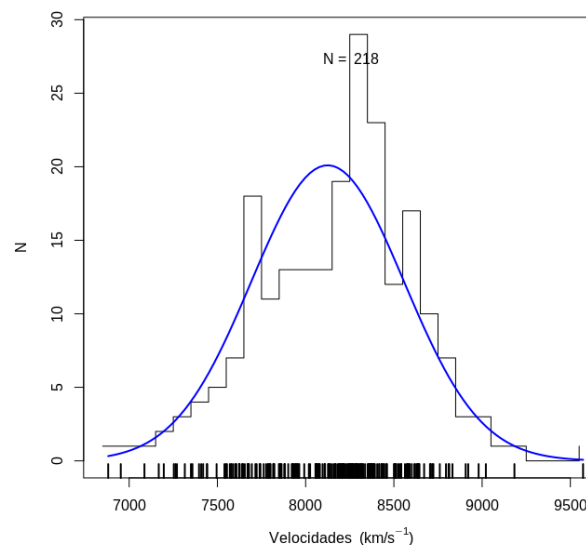
2 Aglomerado de Galáxias

A origem do Universo, de acordo ao modelo cosmológico padrão, se deu há aproximadamente 14 milhões de anos. Desde então o seu processo de expansão ocorre de forma contínua e hierárquica, de modo que unidades menores se fundem formando outras maiores. Aglomerados de galáxias são as maiores estruturas do Universo observável e compõem os objetos de estudo desta dissertação.

Aglomerados de galáxias são definidos basicamente por três componentes: galáxias, meio intra-aglomerado e matéria escura. A maior parte da massa do aglomerado, cerca de 80% do total, é composta de matéria escura (não-bariônica). Do restante, na forma bariônica (feita de prótons e nêutrons), 15% são compreendidos de gás intra-aglomerado (MIA) e apenas 5% da massa de um aglomerado estão na forma de estrelas de galáxias.

A busca por compreender a formação e evolução dos aglomerados de galáxias é uma das questões mais importantes da Astrofísica. No paradigma atual de formação das estruturas, as galáxias e os aglomerados surgem a partir de halos escuros. O resfriamento desses halos ocasiona a formação de estruturas condensadas, onde depois colapsariam os bárions, formando os sistemas astrofísicos conhecidos. Este cenário seria ainda hierárquico, com a formação dos aglomerados ocorrendo após a formação das galáxias, aproximadamente em um desvio para o vermelho $z \approx 2$ (Velásquez, 2007).

Figura 1 – Histograma de velocidades no Aglomerado.



Fonte: Autor

O processo de formação de aglomerados de galáxias ainda não atingiu o seu fim. Enquanto regiões centrais estão em equilíbrio dinâmico, as regiões periféricas (externas) acumulam matéria na forma de galáxias ou grupos de galáxias de modo contínuo. Comumente o entorno dos aglomerados de galáxias é constituído de grupos de galáxias que podem ser absorvidos pelo aglomerado principal ao longo do tempo, ocasionando o aumento de sua massa (Rembold, 2011). Estudos sobre a distribuição de velocidades de galáxias em aglomerados indicam que a mesma possui distribuição Não Rejeita, vide Figura 1, ou muito bem ajustada por uma gaussiana somente na região virializada do sistema (região mais interna do sistema) (Yahil & Vidal 1977), podendo existir sinais de múltiplos modos normais na região mais externa (Ribeiro, Lopes & Trevisan 2011), comprovando a presença de componentes de um sistema em processo de evolução pelo acréscimo de matéria ao seu entorno. Esse acréscimo de matéria, na forma de galáxias ou grupo de galáxias. Isto sugere que a formação de aglomerados de galáxias é um processo contínuo que decorre de fusões sucessivas e encontros gravitacionais de maiores e menores proporções (Nascimento et al., 2016).

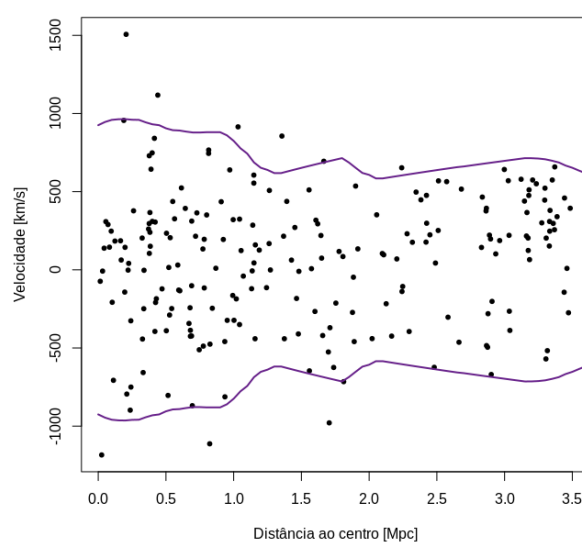
2.1 Distribuição de Velocidades ao longo do Aglomerado

A velocidade de uma galáxia contida em um aglomerado, em uma dada posição, não pode ser maior que a velocidade de escape do sistema, caso isto aconteça a galáxia não pertenceria mais ao aglomerado. A velocidade de escape e a distância ao centro do aglomerado são grandezas inversamente proporcionais, ou seja, a velocidade de escape decresce com o aumento da distância ao centro do aglomerado, portanto é mais fácil o escape de uma galáxia que está na região periférica (de Oliveira e Viegas, 2004).

Para que o aglomerado exista como unidade dinâmica é preciso uma redução na amplitude da distribuição de velocidades das galáxias à medida que haja um afastamento da região central. O grande problema dessa propriedade é o efeito de projeção. As galáxias que estão com distâncias distintas do centro do aglomerado podem parecer ao observador com mesma distância em consequência da observação apenas das posições projetadas no plano do céu (de Oliveira e Viegas, 2004).

Na Figura 2 vemos a distribuição de velocidades do aglomerado em função da distância da galáxia ao centro do aglomerado, onde o estreitamento da distribuição de velocidades define uma espécie de "corneta" que pode ser utilizada para definir os membros de um aglomerado, sendo removidas as galáxias que estejam significativamente acima ou abaixo da "corneta".

Figura 2 – Distribuição de velocidades em função da distância ao centro do Aglomerado.



Fonte: Autor

3 Rotação de Aglomerados

O conhecimento do estado dinâmico de aglomerados de galáxias pode propiciar restrições importantes em cenários cosmológicos, como a determinação da massa total do aglomerado e uma estimativa da quantidade de matéria escura no Universo. A possibilidade da existência de aglomerados em rotação tem sido discutida por muitos autores (por exemplo, veja os estudos de Hwang & Lee 2007; Manaloupoulos & Plionis 2016).

Para detecção de indícios de rotação, Hwang & Lee (2007) utilizaram dados espectroscópicos do *Sloan Digital Sky Survey*¹ (SDSS) e *Two-Degree-Field Galaxy Redshift Survey* (2dF-GRS). A rotação de aglomerados foi modelada como a rotação de galáxias membro e a rotação do gás intra-aglomerado. Eles levantaram inicialmente a hipótese de que a rotação se origina através de fusões de aglomerados. Um aspecto importante do método empregado por Hwang & Lee (2007) é que os aglomerados com rotação devem exibir divisão espacial entre galáxias com velocidades maiores e menores que a velocidade média do aglomerado além de apresentar um pico no mapa de densidade. Nesta pesquisa de Hwang & Lee (2007) foram detectados seis sistemas com rotação, em um total de dozes aglomerados (Abell 0954, Abell 1139, Abell 1399, Abell 2162, Abell 2169, and Abell 2366). Constatou-se ainda que estes aglomerados estão em equilíbrio dinâmico e não sofreram fusão recente, não dando suporte, portanto, à hipótese de interações como causadoras da rotação.

Kalinkov et al. (2008) tentaram obter o gradiente máximo no campo de velocidades de Abell 2107 e determinaram que a direção do coeficiente de correlação linear máximo definiria o eixo maior do aglomerado e o eixo menor seria o de rotação. Foram utilizadas subamostras de galáxias membro, ordenadas de acordo a distância ao centro do aglomerado para definir o grau de rotação do sistema. Esse mesmo aglomerado foi estudado por Oegerle et al. (1992) e foram encontrados indícios de rotação. Materne et al. (1983) apontaram a dificuldade em diferenciar um aglomerado rotativo de dois que se sobrepõem, pelo motivo de estar se fundindo ou se afastando. Porém, o aglomerado Abell 2107 não consiste de dois aglomerados sobrepostos, em consequência do pico estreito representado em seu histograma de velocidades. O indicador mais forte que definiu a rotação foi o ângulo de posição do eixo com o gradiente máximo no campo de velocidades quase coincidindo com o ângulo de posição do eixo com o maior alongamento. Nesse estudo o período de rotação foi calculado em uma volta a cada 2.4×10^9 anos e houve uma correção no valor da massa para $2.8 \times 10^{14} M_{\odot}$ (a massa inicial era

¹Considerado o mais ambicioso mapeamento astronômico que já foi feito. Com este mapeamento, os astrônomos podem observar os padrões de grande escala das galáxias: filamentos e vazios em grandes regiões angulares do Universo.

de $3.2 \times 10^{14} M_{\odot}$, sem levar em conta a rotação).

Baseado no estudo da distribuição de velocidades das galáxias membro, Tovmassian (2015) detectou sinais de rotação em 17 de uma amostra de 65 aglomerados (26%). O método analisa o número de galáxias com velocidades mais baixas e mais altas que a velocidade média do aglomerado em diferentes partes do aglomerado. O método teve mais êxito em aglomerados planos, com $f = a/b > 1.8$ (a e b semieixos - maior e menor - da distribuição de galáxias do aglomerado). Para estes, a taxa de detecção de rotação foi mais alta (7 dos 18 aglomerados planos, 39%). Esse resultado suporta a opinião de que os aglomerados foram originalmente formados a partir das enormes nuvens de gás primordiais e preservaram a rotação das nuvens primordiais, a menos que sofram fusões com outros aglomerados e grupos de galáxias.

Já na tese de Manolopoulou (2014) é realizado um estudo de um novo algoritmo para dedução de rotação usando a velocidade radial projetada². Inicialmente os testes foram realizados em aglomerados gerados em simulações de Monte Carlo para confirmar se o método fornecia indicações robustas de rotação. Em seguida, aplicado em amostras de aglomerados de Abell. Através do teste de Kolmogorov-Smirnov, decidiu-se quanto a sua rotação significativa ou não, seu centro rotacional, orientação do eixo de rotação, amplitude de velocidade rotacional e, finalmente, o sentido de rotação no sentido horário ou anti-horário no plano do céu. Foram encontrados 23 aglomerados possivelmente rotativos dentro de 1.5 Mpc ou a uma distância de 2.5 Mpc do centro do aglomerado, do total de 45 da amostra.

Nascimento et al (2016), a partir de uma amostra de galáxias observadas no Cerro Tololo Interamerican Observatory (CTIO), realizaram um estudo dinâmico em torno do par de aglomerados de Abell (A3407 e A3408). O objetivo era verificar se a amostra correspondia a um simples sistema de galáxias ou a um processo de fusão, melhorando o entendimento desse sistema. Testes estatísticos foram aplicados aos membros mostrando que ambos os sistemas bem como cada aglomerado individual tem uma distribuição de velocidade Gaussiana. Um gradiente de velocidade de $\approx 847 \pm 114 \text{ km s}^{-1}$ foi identificada ao redor do eixo principal da distribuição de galáxias projetada indicando uma possível rotação. O estudo definiu um "gap" na distribuição de velocidades e realizou testes sobre a distribuição espacial de galáxias (em torno do eixo principal do aglomerado) visando identificar diferenças entre objetos com velocidades maiores e menores que a posição do "gap". Esta comparação indicou que havia diferença significativa entre estas subamostras, sugerindo um grau de rotação no sistema A3407+A3408 (vide Figura 3).

²É a velocidade de um objeto na direção da linha de visada, isto é, a velocidade com que o objeto se aproxima ou se afasta do observador.

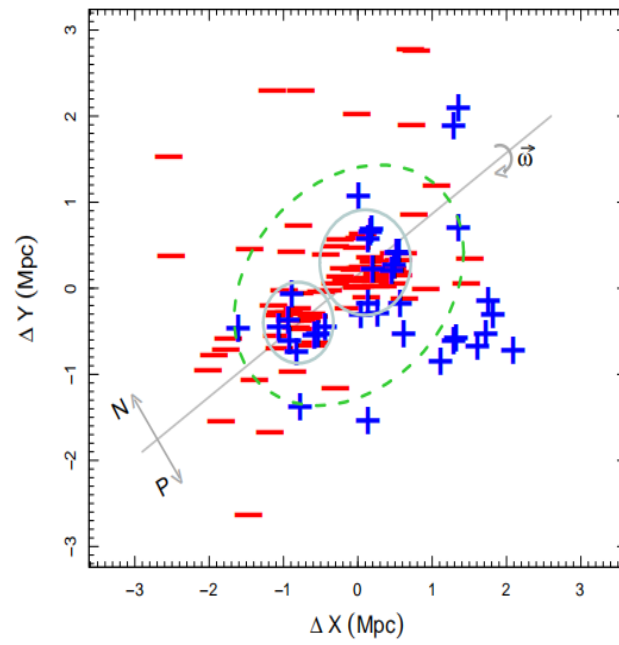


Figura 3 – Distribuição de galáxias no plano do céu do par de aglomerados A3407 e A3408. Fonte: Nascimento et al. 2016.

4 Ferramentas Estatísticas Utilizadas

4.1 Teste de Hipótese

A análise estatística objetiva, especialmente, fazer inferência sobre uma população a partir da observação de uma amostra. Os testes de hipótese representam uma forma de inferência estatística. A hipótese é uma afirmação sobre parâmetros populacionais que devem ser analisadas para verificar sua veracidade. É importante ressaltar que a verdade ou não nunca pode ser determinada, a menos que toda a população seja observada, situação impraticável na maioria das vezes, justificado pelo uso do teste estatístico.

A princípio é necessário estabelecer como verdadeira a **hipótese nula**, denotada por H_0 . Já a **hipótese alternativa** (H_1), contrapõe a hipótese nula, ou seja, H_0 deverá ser rejeitada. É necessário estabelecer um critério auxiliar para decidir a rejeição ou não de H_0 para um teste estatístico. Esse valor, determinado pelo pesquisador antes da análise de dados ou até mesmo na coleta de dados, cenário ideal, é denominado **nível alfa** (α) ou **nível de significância**. Comumente é utilizado como critério de rejeição uma probabilidade de 5%. De acordo com Cramer e Howitt (2004),

O nível em que a hipótese nula é rejeitada é geralmente definido como 5 ou menos vezes fora de 100. Isso significa que tal diferença ou relacionamento é provável que ocorra por acaso 5 ou menos vezes de 100. Este nível é geralmente descrito como proporção 0.05 e às vezes como a porcentagem 5%. O nível de probabilidade de 0.05 foi historicamente uma escolha arbitrária, mas tem sido aceitável como uma escolha razoável na maioria das circunstâncias. Se houver um motivo para variar este nível, é aceitável fazer então. Então, em circunstâncias em que pode haver consequências adversas muito graves se a decisão errada foi feita sobre a hipótese, então o nível de significância poderia ser mais rigoroso em, digamos, 1% (Cramer and Howitt, 2004: 151).

Na realização de testes de hipóteses é possível que erros sejam cometidos, como mostrado no quadro 1. **Erro do tipo I**, denotado por **erro** α , a rejeição de H_0 quando ela é verdadeira. Contrapondo, a não-rejeição de H_0 quando esta é falsa é denominada **erro do tipo II** e representado por β . Esse tipo de teste permite concluir se deve aceitar ou rejeitar a hipótese nula, porém não é possível quantificar o quão provável é o resultado de ocorrer ao acaso. Apoiado por isto, é definido a potência de um teste estatístico $1 - \beta$ como a probabilidade de rejeitar H_0 quando de fato é falsa. Claramente, o teste ideal é aquele em que os valores de α e β são mínimos. Porém, o valor de α é inversamente relacionado com o valor de β , sendo impossível minimizá-los simultaneamente. Geral-

mente, é fixado o nível de significância α e escolhido a região de rejeição que minimiza β , ou seja, que maximize a potência do teste.

Quadro 1 – Tipos de erros em testes de hipótese.

Decisão Estatística	Natureza (estado verdadeiro ou desconhecido)	
	H_0 verdadeira	H_1 falsa
Aceitar H_0	acerto	Erro tipo II (β)
Rejeitar H_0	Erro tipo I (α)	acerto

Fonte:)

O menor nível de significância pode ser definido utilizando o **valor-p** ou “*p-value*”. No teste de hipótese esse valor é comparado ao nível de significância α determinado no início objetivando a tomada de decisão de aceitar ou rejeitar H_0 . Se o valor-p calculado do teste for igual ou maior que α , a H_0 é aceita. Ou seja, a hipótese nula é consistente com os resultados da amostra. Porém, se o valor-p for menor que α , a hipótese nula é rejeitada, a hipótese alternativa, nesse caso, é então aceita como verdadeira.

4.2 Teste de Normalidade

Uma variável aleatória, seja idade de um grupo de pessoas ou ocorrência de um determinado desfecho, pode admitir uma distribuição de frequências da população, contendo diversas formas encontradas na literatura estatística. O intuito desses modelos é caracterizar o comportamento de um determinado evento em função da frequência de sua ocorrência. Se as variáveis forem contínuas, o evento será um intervalo de valores. Portanto, as distribuições de frequências são efetivamente distribuições de probabilidade, em que para um evento teremos associado uma probabilidade de ocorrência (T2).

A inspeção visual pode ser utilizada para avaliação da normalidade. A distribuição de frequência, como exemplo um histograma, relaciona valores observados à sua frequência e pode além de pressupor uma distribuição normal, identifica *insights* sobre lacunas nos dados e outliers. O histograma é composto por barras justapostas em que no eixo horizontal contém a variável de interesse dividida em classes e no eixo vertical a sua correspondente frequência (T2). Para distribuições do tipo normal ou Gaussiana, o histograma constitui formato de sino (Figura 3b).

Entretanto, a simples constatação por meio de gráficos é subjetiva e não satisfatória, pois depende de uma interpretação visual além de não ser confiável no caso multivariado e especificamente nas situações de muitas variáveis. Desta forma, para inferir sobre a normalidade é necessário utilizar como complemento testes estatísticos

(T1). Como exemplo, podemos citar: o teste de aderência qui-quadrado; Kolmogorov-Smirnov; Lilliefors e Shapiro-Wilk.

Estes testes possuem estatísticas de teste e critérios de decisão diferentes, porém compartilham da hipótese avaliada: a hipótese de nulidade (H_0) especifica que a variável aleatória adere à distribuição normal, sem a necessidade de definir a média ou variância da distribuição. Já a hipótese alternativa (H_1), opõe a hipótese nula (T2).

O resultado que interessa após executar um determinado teste é o seu valor-p ou nível descritivo do teste, referente à probabilidade de que a estatística do teste (como variável aleatória) tenha valor extremo em comparação ao valor observado (estatística) quando a hipótese nula é verdadeira. Sendo o valor-p menor que o nível de significância, logo a hipótese nula é rejeitada. Ou seja, o valor-p representa o menor nível de significância que pode assumir para então rejeitar a hipótese nula. Logo, há significância estatística quando o valor-p é menor que o nível de significância estabelecido (FLÁVIO – ESTUDO DA DISTRIBUIÇÃO).

4.3 Testes de duas amostras

O teste de duas amostras é uma análise estatística projetada para utilizar dados de duas amostras aleatórias. O intuito do teste é determinar se a diferença entre as duas populações é estatisticamente significativa.

Seja X_1, X_2, \dots, X_m e Y_1, Y_2, \dots, Y_N duas amostras aleatórias independentes com funções de distribuição contínua F e G , respectivamente, o teste de duas amostras verifica se

$$H_0 : F = G \qquad H_1 : F \neq G \qquad (1)$$

Isto é, assume como hipótese nula que as médias populacionais das amostras não são significativamente diferentes.

4.4 Testes utilizados

4.4.1 Teste de Cramer-von Mises

Em estatística, o teste de Cramér (conhecido também como phi de Cramer - φ_c) é uma medida de associação entre duas variáveis nominais dado o intervalo de 0 a 1, indicando que um valor mais alto possui forte associação. Fundamentado no teste estatístico do qui-quadrado de Pearson, foi publicado em 1946 por Harald Cramér. A

medida é definida como

$$V = \sqrt{\frac{\chi_{obt}^2}{N.m}} \quad (2)$$

onde χ^2 é o valor obtido do teste estatístico

N é o tamanho da amostra e

m = o menor de $(r - 1)$ ou $(c - 1)$, sendo r o número de linhas e c o número de colunas.

Para entender melhor a utilidade do teste de Cramer é fundamental compreender as formas como os testes estatísticos divergem das medidas de associação para variáveis categóricas. O teste qui-quadrado (χ^2) fornece um teste estatístico de associação entre duas variáveis categóricas (nominais) de uma população única. Ele determina se a associação entre as variáveis é significativa, utilizando como hipótese nula (H_0) que as duas variáveis não são dependentes uma da outra e como hipótese alternativa (H_1) é que existe alguma associação entre duas variáveis.

O teste de Cramer é considerado um dos favoritos entre as medidas baseadas no qui-quadrado. Geralmente, quando o seu cálculo resulta no valor máximo 1 é que exista um forte relacionamento entre duas variáveis. No cálculo de Cramer é levado em consideração as dimensões da tabela, ou seja, diferentes dimensões podem ser comparadas significativamente.

4.4.2 Teste de Hotelling

Um dos mais conhecidos testes de hipóteses multivariados foi proposto por Harold Hotelling em 1947, o teste de T^2 , compara vetores de médias populacionais. Baseado na generalização da estatística *t de Student*, foi o primeiro a levar em consideração a correlação das variáveis na formulação da estatística do teste.

Sendo \mathbf{X} um vetor aleatório com uma dada dimensão, μ o vetor de médias e σ a matriz de covariância. Para \mathbf{X} , sendo uma distribuição Não Rejeita multivariada e com tamanho de amostra aleatória n , a estatística de T^2 é dada por

$$T^2 = n(\bar{\mathbf{X}} - \mu_0) \sum_{p \times p}^{-1} (\bar{\mathbf{X}} - \mu_0) \quad (3)$$

com

$$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0 \quad (4)$$

A equação 3 tem distribuição qui-quadrado com p graus de liberdade. Definindo um nível de significância α , com $0 < \alpha < 1$, para valores de T^2 maiores ou iguais ao valor crítico $\chi_{a,p,c}^2$ dado por $P[\chi_p^2 \geq \chi_{a,p,c}^2]$, a hipótese nula será rejeitada.

Sendo a matriz desconhecida, a estatística de T^2 é dada por

$$T^2 = n(\bar{X} - \mu_0)S^{-1}(\bar{X} - \mu_0) \quad (5)$$

que sob a hipótese nula, tem uma distribuição proporcional a uma distribuição F, ou seja, o valor crítico do teste a um nível de significância α , com $0 < \alpha < 1$, é

$$F_c = \frac{p(n-1)}{n-p} F_{1-\alpha,p,n-p} \quad (6)$$

onde $F_{1-\alpha,p,n-p}$ é a probabilidade acumulada igual a $(1 - \alpha)$ da distribuição de F com p e $n-p$ graus de liberdade.

Sendo S a matriz de covariâncias amostrais ($p \times p$), um estimado não viciado de $\sum_{p \times p}$, dado por

$$\begin{bmatrix} S_1^2 & S_{12} & \dots & S_{1p} \\ & S_2^2 & \dots & S_{2p} \\ & & \ddots & \vdots \\ & & & S_p^2 \end{bmatrix} \quad (7)$$

em que os elementos da diagonal principal de S são as variâncias definidos por

$$S_j^2 = \frac{1}{m-1} \sum_{k=1}^m (x_{jk} - \bar{X}_j), j = 1, 2, \dots, 3 \quad (8)$$

e os elementos fora da diagonal principal são as covariâncias conforme

$$S_{jh} = \frac{1}{m-1} \sum_{k=1}^m (x_{jk} - \bar{X}_j)(x_{hk} - \bar{X}_h) \quad (9)$$

onde x_{jk} e x_{hk} representam os valores amostrais das variáveis X_j e X_h .

5 Linguagem R

R é uma linguagem e ambiente para análise estatística e produção de gráficos, desenvolvida na década de 90 pelos estatísticos Ross Ihaka e Robert Gentleman que utilizavam sistemas pagos em seus projetos. Contém diversos pacotes integrados permitindo que sejam realizadas uma grande variedade de estatísticas (natureza linear, não-linear, séries temporais) e o uso de recursos gráficos avançados, o que a distingue das demais linguagens.

O R é um conjunto integrado de instalações de software para manipulação de dados, cálculos e exibição gráfica. Isto abrange uma instalação eficaz de manipulação e armazenamento de dados, um conjunto de operadores para cálculos em matrizes, uma grande coleção coerente de ferramentas intermediárias para análise de dados e representações gráficas, a inclusão de condicionais (*loops* e funções recursivas) definidas pelo usuário e recursos de entrada e saída.

5.1 Pacotes

5.1.1 Pacote Cramer

Uma rotina em R aplicada para o teste de Cramer de duas amostras. O valor de retorno é um objeto da classe "*cramertest*", contendo, dentre outros, os seguintes componentes:

- *statistic*: valor estatístico do teste de Cramer para observações.
- *conf.level*: nível de significância do teste.
- *p.value*: estimativa do valor-p.

5.1.2 Pacote Hotelling

Uma rotina em R aplicada para o teste de Hotelling de duas amostras. O valor de retorno é uma lista da classe "*hotelling.test*", contendo, dentre outros, o seguinte componente:

- *pval*: estimativa do valor-p.

5.1.3 Pacote Ellipse

Este pacote contém diversas rotinas para desenhar elipses e regiões de confiança do tipo elipse, implementando os gráficos descritos por Murdoch e Chow (1996) e Bates e Watts (1988).

5.1.4 Pacote Astro

O pacote **astro** fornece uma série de funções, ferramentas e rotinas no uso diário da astronomia. Pode-se agrupar essas funções em áreas como a cosmologia ¹, manipulação de arquivos FITS ², funções de tempo e posição.

5.1.5 Pacote cosmoFns

O pacote contém expressões de distância, tempo, luminosidade e outras úteis na cosmologia observacional, compreendendo observações em linhas moleculares. Atualmente codificado apenas para o universo plano.

¹Funções que calculam distâncias, movimentação de volumes, *lookback time* e luminosidade em uma cosmologia plana

²Sistema flexível de transporte de imagens (*Flexible Image Transport System*), um formato de arquivo comum em astronomia

6 Dados e Metodologia

O método proposto por Nascimento et al. (2016) para estudar um par de aglomerados pode ser adaptado para o estudo da rotação de aglomerados individuais. Neste trabalho fazemos esta adaptação e a implementamos em linguagem R. O código resultante é aplicado a um conjunto de três catálogos:

- **Catálogo I:** composto por 20 aglomerados ricos do SDSS, localizados em baixos *redshifts*, com espectroscopia disponível para objetos com $m_r \leq 17.77$.
- **Catálogo II:** NoSOCS (*Northern Sky Optical Cluster Survey*) com baixo redshift ($z \leq 0.10$), compreendendo 183 objetos.
- **Catálogo III:** .

6.0.1 Catálogo I: selec20

A amostra selec20 corresponde a um conjunto de aglomerados ricos (sistemas com mais de 50 galáxias na região virializada) a baixos redshifts ($z < 0.13$), cuja determinação de membros foi feita pelo Dr Paulo Lopes (UFRJ) fazendo uso do programa shiftgapper (Lopes et al. 2009). Esta amostra, por suas características, é extremamente útil para testes e análises exploratória.

6.0.2 Catálogo II: NoSOCS

O NoSOCS (Gal et al. 2000, 2003, 2008) é um catálogo de aglomerados de galáxias elaborado a partir da versão digitalizada do Segundo Observatório de Palomar (POSS-II; DPOSS, Djorgovski et al. 2003). Este catálogo é derivado de campos de alta latitude $|b| > 30^\circ$, cobrindo $\sim 11,000 \text{ deg}^2$ e contendo $\sim 15,500$ aglomerados candidatos. Sua construção é limitada para $r = 19.5$, onde a separação estrela/galáxia é confiável e erros fotométricos são precisos para usar a cor $g-r$ como indicador do *redshift*. A classificação de objetos e a calibração fotométrica são descritas em Odewahn et al. (2004) e Gal et al. (2004), respectivamente.

Os dados foram extraídos do SDSS para cada aglomerado do NoSOCS (Lopes 2003; Lopes et al. 2004) amostrados no DR5. Com o uso de dados fotométricos de alta qualidade, foi possível estimar os novos *redshifts* fotométricos (seguindo Lopes 2007), riqueza e luminosidade óptica (Lopes et al. 2006). Após a remoção de aglomerados duplos, o número de aglomerados do catálogo sofreu uma redução, compreendendo 7414 objetos. Para esses objetos foi aplicado uma técnica de eliminação de interferentes de

galáxias nas distribuições do espaço de fase (*shifting gapper*) e estimado as dispersões e massas de velocidade (M500 e M200).

A amostra final com essas propriedades e em baixo z apresentou 127 agrupamentos. Como o número de amostras do NoSOCS em baixo z não era volumoso, a análise foi estendida para sistemas mais ricos, incluindo os aglomerados CIRS (Rines & Diaferio 2006), com 56 objetos. Portanto, a lista final em baixo *redshift* ($z \leq 0.10$) compreendeu 183 objetos (127 do NoSOCS e 56 do CIRS).

6.0.3 Nosso Método

O método foi baseado no trabalho de Nascimento et al. (2016) que estudou o par de aglomerados A3407 + A3408, no entanto pode ser adaptado para o estudo da rotação de aglomerados individuais.

Foi realizada uma análise na distribuição de velocidades das galáxias membro do aglomerado em busca de "gaps" significativos. O intuito da rotina era identificar a probabilidade de que um "gap", de certo tamanho e em dada localização, possa ser produzido a partir de amostragens aleatórias retiradas de uma gaussiana. As velocidades das galáxias são ordenadas em ordem crescente e o i -ésimo "gap" é definido como $g_i = v_{i+1} - v_i$. O "gap" é ponderado pela sua posição, através de $w_i = i(N - i)$, onde N é o número de galáxias do aglomerado. Os "gaps" ponderados são Não Rejeitados através da divisão por meio da média (MM) da distribuição ordenada do "gap" ponderado dada por:

$$MM = \frac{2}{N} \sum_{i=N/4}^{3N/4} \sqrt{w_i g_i} \quad (10)$$

Investigamos "gaps" com valores maiores que 2.25, uma vez que em retiradas aleatórias de uma gaussiana, "gaps" desse tamanho ocorrem no máximo em 3% dos casos (vide Wainer & Shacht 1978. Beers et al. 1991). Em seguida os dados foram divididos em duas amostras, contendo objetos com velocidades maiores e velocidades menores que o maior "gap" encontrado, referimos aqui de amostras I e II. Para o caso de aglomerados não encontrado "gap" significativo (> 2.25), utilizamos a mediana dos dados como divisor das velocidades do sistema.

A partir de "gaps" identificados na distribuição de velocidade, levantou-se o seguinte questionamento: pode eles indicar um gradiente de velocidade em toda distribuição espacial de galáxias? Para isso, estimamos o eixo principal do aglomerado como o resultante do ajuste de uma elipse aos dados projetados no plano do céu, como ilustrado na figura 4. O ajuste foi feito usando-se o pacote **ellipse** do R.

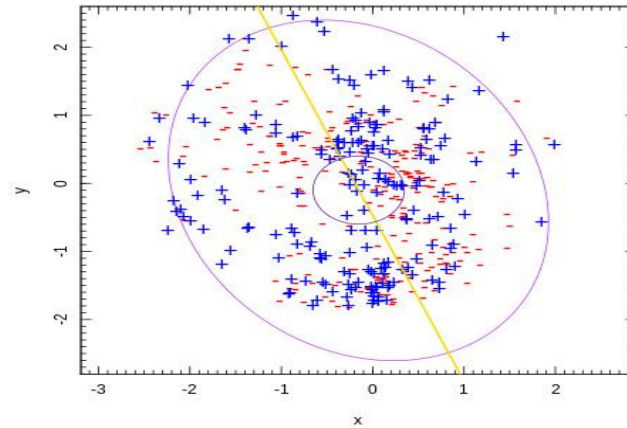


Figura 4 – Ajuste da elipse e eixo principal da distribuição projetada no plano do céu.
Fonte: autor.

As amostras I e II foram então comparadas em relação a sua distribuição de duas maneiras: independente do eixo principal e em cada lado do eixo. Os testes de comparação de duas amostras utilizados foram o teste de Cramer 2D e o de Hotelling, dos pacotes **Cramer** e **Hotelling** do R, respectivamente.

Dado que as distribuições espaciais das amostras I e II sejam distintas com 95% de confiança em relação aos testes acima citados, interpretamos o resultado como sendo uma indicação indireta de rotação nos aglomerados. Para os aglomerados onde isto acontece, traçamos um perfil de velocidade de rotação ao longo da distância ao centro do aglomerado e identificamos a velocidade de rotação através do raio (R200) do aglomerado (intersecção da curva de rotação e o valor de R200), como representado na figura 5.

Uma vez obtido o valor da velocidade de rotação do aglomerado utilizamos a fórmula dada por Lee & Rood (1969) que investigam a dependência da forma e do teorema do virial agindo no movimento orbital. Para que a velocidade angular afete significativamente o cálculo da massa do aglomerado qualquer aceleração centrífuga precisaria ser comparável à aceleração gravitacional. Através da equação 11 encontramos o valor limite de detecção de velocidade de rotação do aglomerado.

$$\omega \cong \omega_{orb} = \frac{1}{r^2} \sqrt{GM_g \Delta (1 - e^2)}, \quad (11)$$

onde ω é a velocidade de rotação, G é constante gravitacional, R e M são raio e massa do aglomerado, respectivamente.

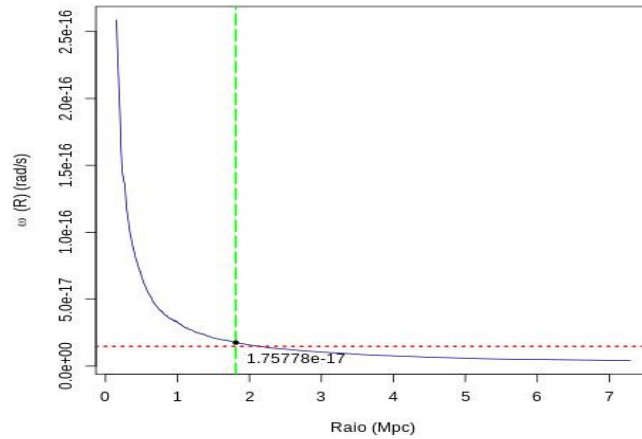


Figura 5 – Perfil de velocidade de rotação. Fonte: autor.

6.0.4 Método de Hwang & Lee - adaptado

O método proposto por Hwang & Lee (2007), também usado para identificar a rotação em aglomerados, foi implementado com o propósito de comparar com o nosso método. Eles utilizam a relação sinusoidal para calcular o eixo de rotação (Θ_o) e a velocidade de rotação (v_{rot}):

$$vp(v_{rot}, \Theta) = v_{sys} + v_{rot} \cdot \sin(\Theta - \Theta_o), \quad (12)$$

onde v_p é a velocidade radial de cada galáxia devido a rotação do aglomerado, v_{sys} é a velocidade peculiar do aglomerado e Θ é o ângulo projetado na posição de cada galáxia no plano céu, partindo do Norte para o Leste. Manolopoulou & Plionis, usaram diferenças de velocidade em relação à velocidade média do aglomerado, o valor de v_{sys} foi igual a 0.

O procedimento de minimização do χ^2 foi utilizado para determinar o melhor ajuste dos valores de v_{rot} e Θ_o , representado na equação 13. Ou seja, o conjunto de valores de v_{rot} e Θ_o , são empregados no cálculo de χ^2 para cada par de parâmetros:

$$\chi^2(v_{rot}, \Theta_o) = \sum_i \frac{(v_{pi} - v_{los,i})^2}{\sigma_i^2}, \quad (13)$$

onde $v_{los,i}$ é a velocidade de linha de visada de cada galáxia e σ_i é a medida em erro.

7 Análise

7.0.1 Catálogo I: selec20

Os resultados da aplicação de nosso método para o catálogo selec20 são apresentados no Apêndice e na tabela 1. Na Figura 6, é exibido a análise de *gaps* para um dos 20 aglomerados da amostra, o aglomerado 08. O gráfico na figura contém o histograma da distribuição de velocidades, o ajuste gaussiano superposto (linha em azul), barras inferiores indicando as velocidades individuais ordenadas em ordem crescente, sendo que em vermelho estão indicados os *gaps* com valores maiores que 2.25, ou seja, os *gaps* significativos. Quando mais de um *gap* é encontrado, escolhemos aquele de maior valor; finalmente, a linha vertical tracejada indica a posição da BCG (*brightest cluster galaxy*) apenas como referência.

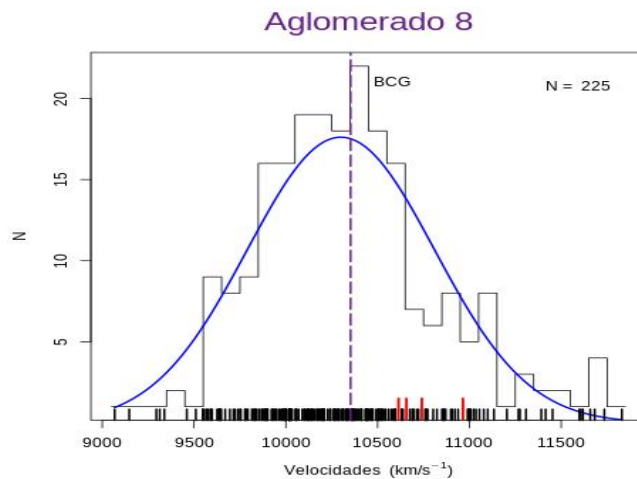


Figura 6 – Histograma Distribuição de Velocidade e Análise de Gaps para o aglomerado 08. Fonte: autor.

Na Figura 7, vemos o ajuste da elipse na distribuição (X,Y) projetada no plano do céu, com os pontos acima (+) e abaixo (–) em relação à posição do *gap* principal, indicados em azul e vermelho, respectivamente. A distribuição de galáxias em torno do *gap* de velocidades pode ser utilizada como um indicador indireto da presença ou não de rotação. Estudamos o quanto diferem espacialmente as galáxias de acordo com a sua posição em relação ao eixo principal. A hipótese nula dos testes é a de que os pontos + e – foram retirados da mesma população. Aplicamos dois testes estatísticos, Cramer 2D e Hotelling, em três cenários diferentes: todos os pontos do gráfico, acima e abaixo do eixo principal (Tabela 1).

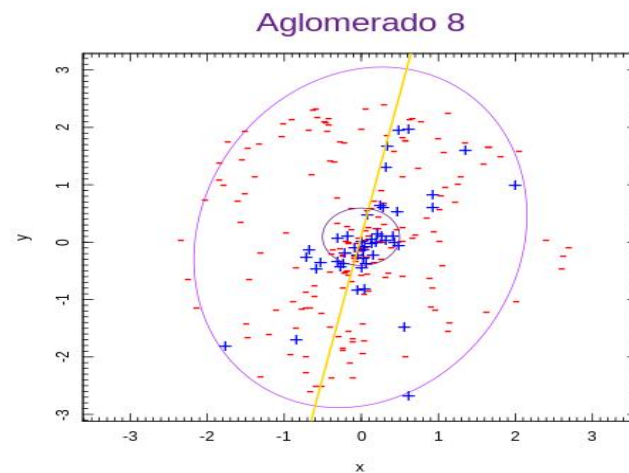


Figura 7 – Ajuste da elipse e eixo principal da distribuição projetada no plano do céu para o aglomerado 08. Fonte: autor.

Tabela 1 – Teste Cramer e Hotelling aplicado no catálogo selec20.

Cluster	Cenário 1		Cenário 2		Cenário 3		Nº galáxias
	Cramer	Hotelling	Cramer	Hotelling	Cramer	Hotelling	
01	0.496503	0.616007	0.768232	0.871778	0.364635	0.746935	223
02	0.003996	0.007875	0.076923	0.006919	0.021978	0.091680	244
03	0.298701	0.561228	0.648352	0.202482	0.060939	0.129923	177
04	0	9.769963e-15	0	2.334383e-07	0.006993007	0.01569305	435
05	0.097902	0.048305	0.044955	0.072319	0.082917	0.396323	951
06	0.281718	0.265025	0.472528	0.804144	0.257742	0.414097	165
07	0.041958	0.588022	0.556444	0.634652	0.038961	0.884234	313
08	0.009990	0.444646	0.011988	0.080753	0.071928	0.055154	225
09	0.028971	0.230194	0.035964	0.037384	0.095904	0.036469	215
10	0.000999	0.006792	0.003996	0.025815	0.502498	0.803647	308
11	0.000999	0.003167	0.335664	0.900630	0.000999	0.069658	651
12	0	1.130531e-10	0	3.967377e-05	0	1.773186e-07	173
13	0.106893	0.698218	0.133866	0.151622	0.265734	0.332495	218
14	0.038961	0.005413	0.475525	0.211045	0.760240	0.501961	127
15	0.117882	0.574480	0.341658	0.224218	0.036963	0.060953	114
16	0.044955	0.518194	0.002997	0.019457	0.408591	0.477522	144
17	0	2.202682e-13	0	3.47107e-05	0.2367632	0.3371383	773
18	0	0.04275908	0.01498501	0.007541272	0.1298701	0.753498	233
19	0.087912	0.406300	0.299700	0.295311	0.137862	0.239693	113
20	0.118881	0.903751	0.072927	0.317330	0.128871	0.229752	210

O teste de Cramer para duas amostras pode ser usado para dados univariados e multivariados, como neste trabalho. Para o cálculo do valor crítico uma rotina de *bootstrap* é utilizada e métodos de permutação são usados para obter o valor-p do teste. Neste trabalho usamos 1000 réplicas dos dados a cada caso. O teste de Hotelling multivariado compara médias em duas amostras. A rejeição ou não da hipótese nula é feita em todos os casos para um nível de 95% de confiança.

Consideramos evidência significativa de rotação se houve rejeição da hipótese nula em pelo menos um dos cenários testados. Isto nos leva a 14 aglomerados com evidência de algum grau de rotação. São eles os aglomerados: 02, 04, 05, 07, 08, 09, 10,

11, 12, 14, 15, 16, 17 e 18. Em uma análise mais criteriosa verificamos que:

- 28.57% dos aglomerados houve rejeição da hipótese nula (indicando rotação) em pelo menos um dos testes (Cramer e Hotelling) nos três cenários (todos os pontos, acima e abaixo do eixo principal).
- 42.86% dos aglomerados houve rejeição da hipótese nula em pelo menos um dos testes em dois cenários.
- 7.14% dos aglomerados houve rejeição da hipótese nula nos dois testes em apenas um cenário.
- 7.14% dos aglomerados houve rejeição da hipótese nula em pelo menos um dos testes em apenas um cenário.

Finalmente, calculamos o perfil de velocidade de rotação para estes quatorze aglomerados, como ilustrado na Figura 8. A velocidade de rotação foi calculada de maneira cumulativa contra o raio projetado das galáxias de acordo com

$$\omega = \Delta V/R \quad (14)$$

onde ΔV é a diferença de velocidade entre os pontos + e – internos a R .

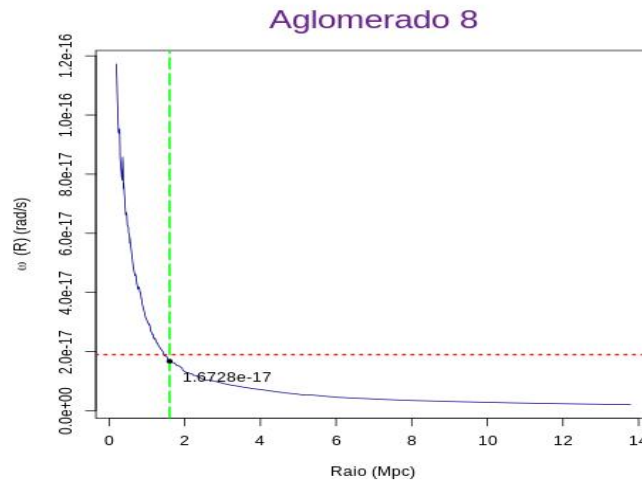


Figura 8 – Perfil de velocidade de rotação para o aglomerado 08. Fonte: autor.

No gráfico o valor exibido é a velocidade de rotação do aglomerado determinado a partir do R200 e a linha na vertical, o valor em y, é a velocidade limite para detecção de rotação no aglomerado, calculado pela fórmula da equação 11.

7.0.2 Catálogo II: NoSOCS

No catálogo de NoSOCS, diferente do selec20, 56.33% dos aglomerados (um total de 89) não apresentaram *gaps* significativos, nestes casos os dados foram divididos pelo cálculo da mediana. Além disso, 13.66%, cerca de 25 aglomerados, continham um total inferior a 20 objetos, o que tornava o cálculo de detecção de rotação inviável. Por esse motivo, reduzimos a amostra de 183 para 158 aglomerados.

Aplicamos os mesmos cenários e testes usados no catálogo selec20 e verificamos que:

- 34.43% dos aglomerados apresentaram indicação direta de rotação, levando em conta um dos testes (Cramer e Hotelling) em pelo menos um cenário. Considerando-se que 47.62% apenas as amostras sem *gap*, ou seja, em casos onde os dados foram divididos pela mediana.
- Desse total, tendo em conta a rejeição da hipótese em ambos os testes para os três cenários um total de 6.35%, dois cenários 15.87% e um cenário 9.52%.
- Já a rejeição da hipótese em pelo menos um dos testes para os três cenários um total de 11.11%, dois cenários 15.87% e um cenário 41.27%.

Os resultados são apresentados nas tabelas 2 3 e no Apêndice.

Tabela 2 – Teste Cramer e Hotelling aplicado no catálogo NoSOCS com rotação detectada utilizando gap.

Cluster	Cenário 1		Cenário 2		Cenário 3		Nº galáxias
	Cramer	Hotelling	Cramer	Hotelling	Cramer	Hotelling	
01238	0.003996	0.005606	0.021978	0.025209	0.056943	0.150563	44
01836	0	0.000395	0.006993	0.010808	0.022977	0.001210	86
02440	0.017982	0.090600	0.069930	0.005757	0.150849	0.137841	41
02447	0.511488	0.646515	0.031968	0.053283	0.743256	0.548081	46
04681	0.014985	0.017471	0.024975	0.047805	0.204795	0.212804	162
04703	0.001998	0.000551	0.027972	0.018654	0.109890	0.149901	91
06256	0.035964	0.016620	0.026973	0.013306	0.283716	0.368284	37
06547	0	0.005209	0.562437	0.849069	0.001998	0.005002	112
07217	0.077922	0.038481	0.382617	0.433854	0.152847	0.456518	24
07703	0.043956	0.079676	0.005994	0.015114	0.586413	0.980806	48
09148	0.012987	0.004465	0.048951	0.112433	0.022977	0.089178	31
10008	0.001998	0.293476	0.029970	0.387069	0.025974	0.143509	123
10013	0.008991	0.031328	0.287712	0.832952	0.377622	0.582334	86
10015	0.003996	0.057669	0.092907	0.279563	0.029970	0.057429	139
10020	0.000999	0.008616	0.003996	0.403677	0.004995	0.016639	114
10021	0.004995	0.030595	0.012987	0.015253	-	-	89
10023	0.040959	0.084825	0.078921	0.277558	0.110889	0.277176	107
10024	0.037962	0.074983	0.090909	0.255343	0.231768	0.413339	113
10027	0.163836	0.413906	0.268731	0.393231	0.035964	0.059391	86
10029	0.025974	0.056490	0.048951	0.165669	0.057942	0.058164	135
10030	0	3.06E+01	0.013986	0.015258	0	5.74E+01	87
10037	0.000999	0.008946	0.032967	0.285959	0.070929	0.157548	145
10043	0.020979	0.059925	0.053946	0.061660	0.188811	0.271167	218
10044	0	1.10E+01	0.003996	0.008143	0.001998	0.001075	177
10045	0.103896	0.281427	0.037962	0.093578	0.481518	0.961108	55
10051	0.045954	0.240987	0	0.003778	0.214785	0.067199	189
10053	0.065934	0.559379	0	0.000141	0.001998	0.001089	447
10054	0	0.006898	0.201798	0.576037	0	0.010700	537
10056	0.025974	0.020923	0.160839	0.047179	0.318681	0.373086	100
10059	0.041958	0.424066	0.111888	0.290470	0.309690	0.404923	150
10060	0.201798	0.917590	0.005994	0.048580	0.831168	0.647436	133
10063	0.099900	0.695755	0.049950	0.189107	0.033966	0.102787	219
10064	0.096903	0.042881	0.279720	0.139438	0.628371	0.789767	133

Tabela 3 – Teste Cramer e Hotelling aplicado no catálogo NoSOCS com rotação detectada utilizando mediana.

Cluster	Cenário 1		Cenário 2		Cenário 3		Nº galáxias
	Cramer	Hotelling	Cramer	Hotelling	Cramer	Hotelling	
01831	0	0.002671	0.001998	0.031291	0.009990	0.004074	72
02137	0.036963	0.347093	0.248751	0.329532	0.161838	0.732693	34
02433	0.118881	0.301255	0.846153	NA	0.032967	0.216530	31
03176	0.040959	0.437995	0.155844	0.161899	0.111888	0.723103	26
03691	0.060939	0.031896	0.464535	0.534411	0.435564	0.115439	43
03907	0.114885	0.168692	0.015984	0.065279	0.606393	0.998517	44
04048	0.139860	0.023276	0.252747	0.213522	0.817182	0.59885	30
04409	0.000999	0.001251	0.101898	0.178182	0.006993	0.01041	23
04470	0.047952	0.110327	0.511488	NA	0.056943	0.136180	35
04479	0.090909	0.173748	0.830169	0.069908	0.013986	0.020507	25
04672	0.354645	0.600566	0.540459	0.433777	0.168831	0.024215	48
05359	0.096903	0.116698	0.010989	0.025032	0.483516	0.702417	62
05447	0	0.001094	0.023976	0.046308	0.007992	0.015626	57
06207	0.626373	0.378981	0.268231	0.007777	0.709290	0.595039	21
06286	0.129870	0.054084	0.019980	0.013322	0.130869	0.159239	70
07395	0.328671	0.777906	0.409590	0.673840	0.136863	0.000785	29
08022	0.266733	0.857210	0.040959	0.259307	0.147852	0.150297	51
08721	0.032967	0.129166	0.225774	0.515822	0.084915	0.275898	47
09061	0.302697	0.037116	0.805194	0.896726	0.517482	0.067197	46
10006	0.040959	0.084762	0.195804	0.602401	0.018981	0.005725	74
10016	0.000999	0.000896	0.034965	0.072791	0.000999	0.005635	92
10018	0.398601	0.331314	0.047952	0.807257	0.330669	0.141546	65
10026	0.033966	0.036372	0.531468	0.680673	0	0.001887	130
10028	0.018981	0.018787	0.231768	0.274238	0.053946	0.052530	80
10031	0	6.43E-03	0.002997	0.001092	0	1.80E-04	141
10035	0.122877	0.156720	0.505494	0.996701	0.029970	0.074546	58
10036	0.410589	0.917501	0.056943	0.045570	0.670329	0.544576	119
10040	0.000999	0.001944	0.036963	0.059351	0.007992	0.009938	99
10050	0.020979	0.595124	0.068931	0.279501	0.077922	0.050991	77
10055	0	0.362105	0.260739	0.180886	0	0.000790	104