

Desempenho do Teste LIS em Imagens Astronômicas

Autor: Thiago Cedran Santos

Bolsa e Orientador: CNPq/V. A. González-López

Introdução

Testes:

Nesse trabalho, foram usados 5 testes de independência para comparar seus resultados ao resultado do teste LIS de independência (<http://arxiv.org/abs/0908.2794>). Os testes utilizados foram: Spearman, Pearson, Kendall, Hoeffding (Hoeffd) e Copula. Para os quais daremos uma breve explicação a seguir.

O **teste de Pearson** testa a correlação de Pearson entre duas variáveis diretamente de seus pontos.

O **teste de Spearman** testa a correlação de Pearson também, porém dos postos das variáveis

E o **teste de Kendall** testa a medida de associação ζ de Kendall.

Esses três testes são testes mais antigos que serão mencionados como testes tradicionais.

Já os **testes de Hoeffding** (Hoeffding W. (1948): A non-parametric test of independence. Ann Math Stat 19:546–57.) e **Copula** (C. Genest and B. Rémillard (2004), Tests of independence and randomness based on the empirical copula process. Test, 13:335–369.) são testes mais atuais e medem a dependência mediante a comparação da distribuição conjunta e das marginais.

A hipótese a testar é :

- H_0 : X e Y são independentes;
- assim cada teste testa alguma característica ou uma medida de dependência que permite inferir se a independência pode ser rejeitada ou não. Cada teste indica a rejeição de H_0 retornando um p-valor com um valor baixo, i.e. menor que determinado nível de significância, previamente escolhido. Já o teste não é capaz de rejeitar H_0 se o seu p-valor for alto.
- No presente estudo baseado na inspeção de imagens bidimensionais, cada ponto (x,y) da amostra corresponde ao pixel dessa imagem. Assim, a rejeição da hipótese de independência por um teste, implica que o mesmo possui mecanismos para detectar alguma relação entre as variáveis X e Y. No caso, significa que o teste detecta a constituição de uma imagem de padrão definido e não aleatório.

Independência vs Dependência

Apresentaremos a seguir simulações de condições parciais para que os resultados deste relatório possam ter um melhor entendimento.

Na imagem 1.1 é apresentado um caso evidente de variáveis independentes e marginais uniformes, onde cada ponto x e y foi criado aleatoriamente.

Nesse caso, todos os testes devem apresentar resultado positivo de independência. Não sendo capazes de rejeitar a independência (hipótese H_0) e fornecendo p-valores altos.

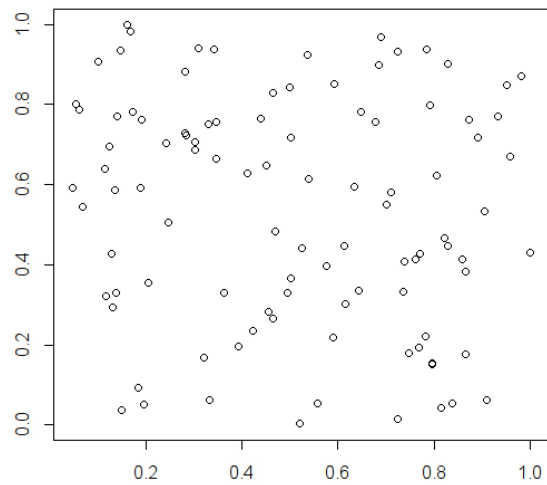


Imagem 1.1 – 100 pontos aleatórios

Porém, casos de independência na vida real podem ser difíceis de identificar como mostra a Imagem 1.2.

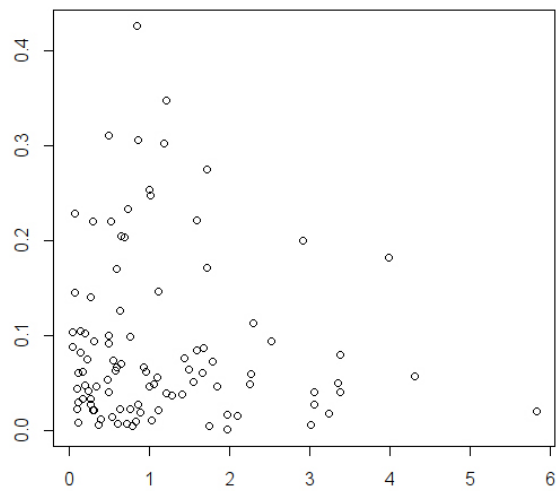


Imagem 1.2 – pontos aleatórios gerados exponencialmente, onde X é $\exp(1)$ e Y é $\exp(10)$

Visualmente podemos julgar este caso como sendo de variáveis dependentes, mas corrigindo X e Y pelos postos, a sua independência fica evidente.

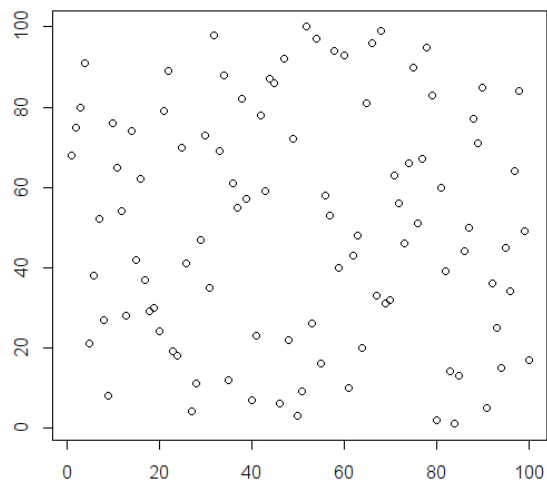
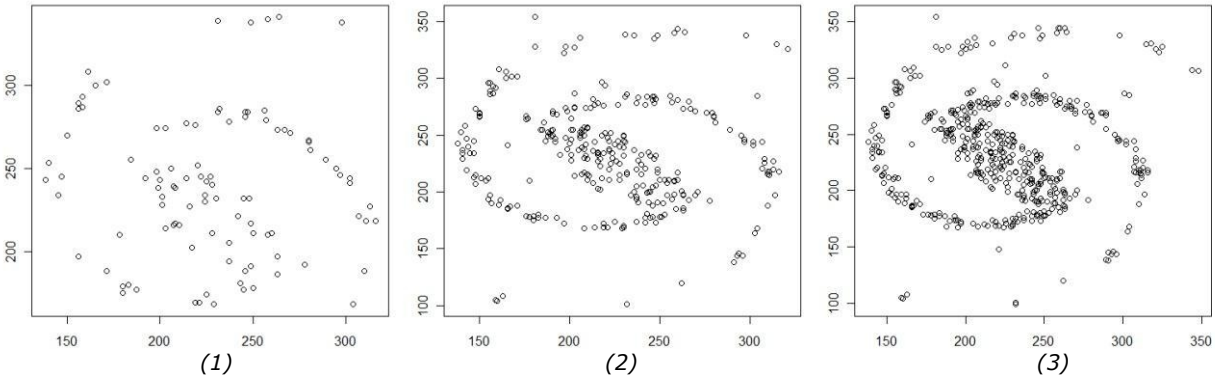


Imagem 1.3 – pontos da Imagem 1.2 corrigidos pelos postos

Resultados

Os resultados foram organizados em 4 cenários principais, tendo em vista algumas semelhanças de desempenho nos testes.

- Objetivo: Analisar a sensibilidade de cada teste para a detecção de padrões da imagem



Como ilustrado nas imagens (1), (2) e (3) desejamos mostrar a eficiência da resposta de cada teste quando o tamanho de pontos é aumentado, isto é, o teste mais eficiente será aquele que retornar um p-valor baixo para um menor tamanho de amostra sendo capaz de detectar a imagem com menos informação (menor tamanho amostral).

Cenário 1

O primeiro cenário é onde, claramente, todos os testes conseguem identificar uma grande correlação para amostras pequenas (menores que 100 pontos).



Imagem 1.1.1 – imagem analisada

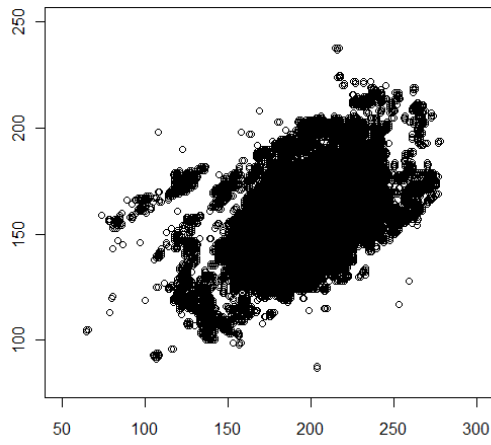


Imagem 1.1.2 – gráfico de todos os pontos coletados da imagem 1.1.1

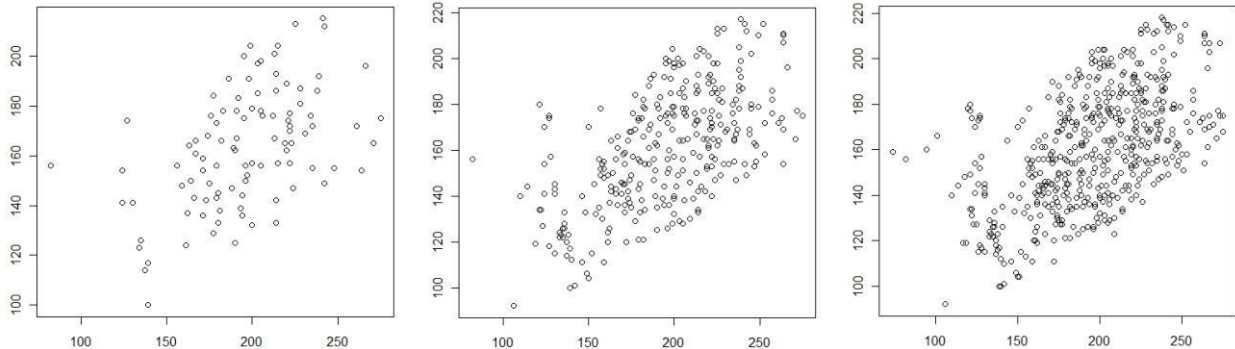


Imagem 1.1.3 – amostras de 100, 300, 500 pontos da imagem 1.1.2

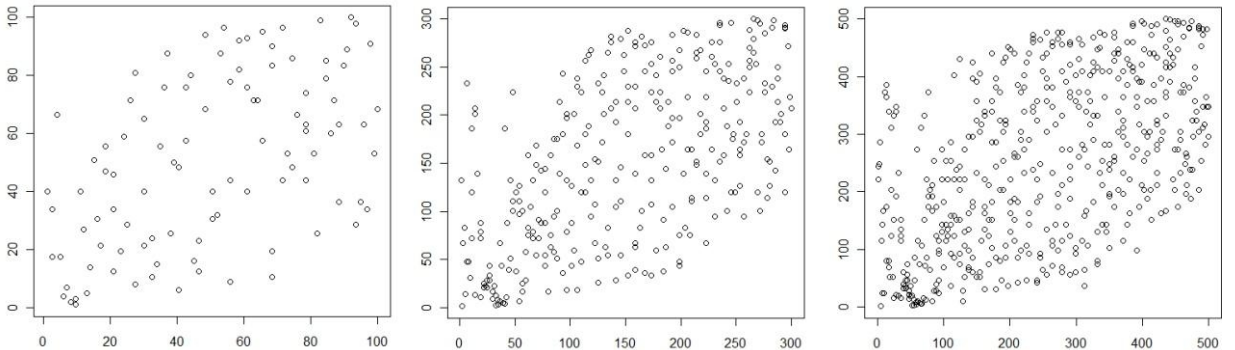


Imagem 1.1.4 – ranks das amostras da imagem 1.1.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.434
200	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.214
300	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.016
400	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.004
500	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.001

Tabela 1.1: testes de correlação aplicado em amostras da imagem 1.1.1



Imagem 1.2.1 - imagem analisada

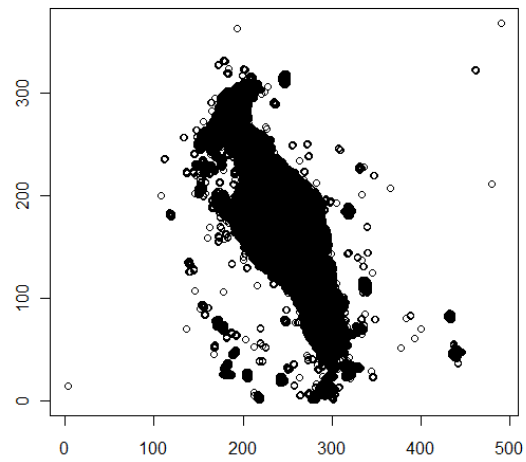


Imagem 1.2.2 - gráfico de todos os pontos coletados da imagem 1.2.1

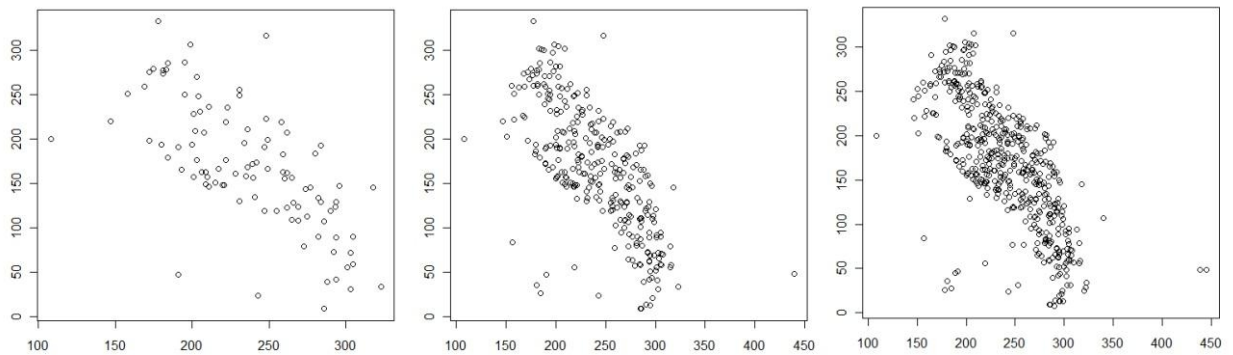


Imagem 1.2.3 - amostras de 100, 300, 500 pontos da imagem 1.2.2

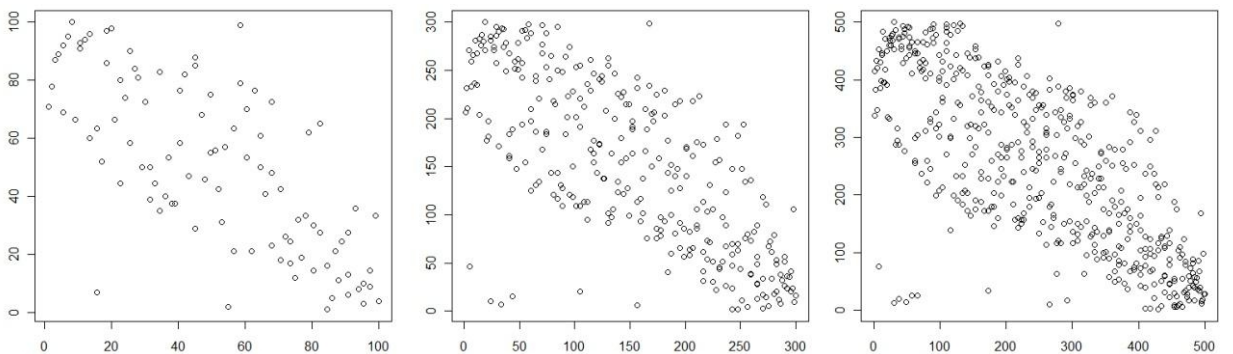


Imagem 1.2.4 - ranks das amostras da imagem 1.2.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
200	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
300	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
400	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
500	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$

Tabela 1.2: testes de correlação aplicado em amostras da imagem 1.2.1

Conclusão cenário 1

Podemos ver, claramente, através dessas duas imagens de amostra que, todos os testes rejeitam a hipótese de independência, identificando com esta rejeição algum padrão de imagem nas figuras.

Os testes tradicionais rejeitam a hipótese de independência pelo formato alongado com um certo ângulo que o padrão da amostra segue. Já os testes que utilizam os postos das amostras rejeitam a independência pela presença dos espaços em branco formados do gráfico dos postos das amostras.

Cenário 2

O segundo cenário encontrado é um cenário no qual o teste de Pearson tem uma maior sensibilidade em relação aos outros testes (logo para 100 amostras), porém o teste LIS tem certa dificuldade para retornar um resultado sensível, inclusive para grandes amostras.



Imagem 2.1.1 - imagem analisada

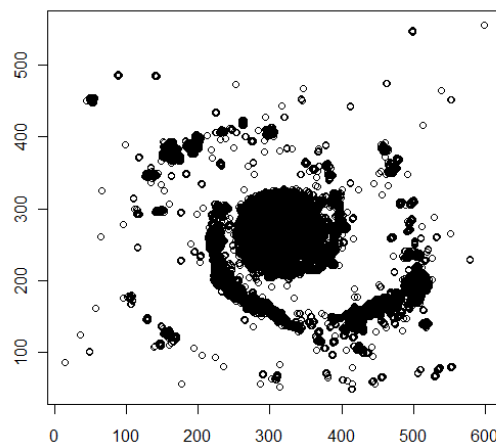


Imagem 2.1.2 - gráfico de todos os pontos coletados da imagem 2.1.1

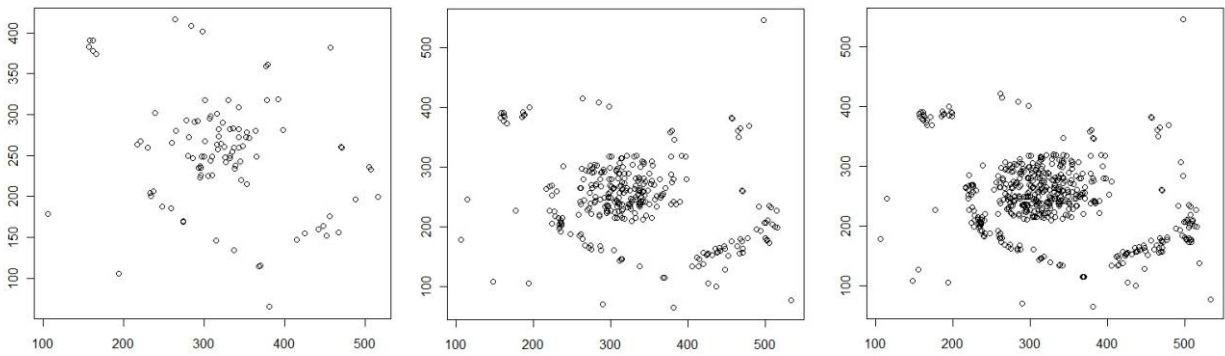


Imagem 2.1.3 – amostras de 100, 300, 500 pontos da imagem 2.1.2

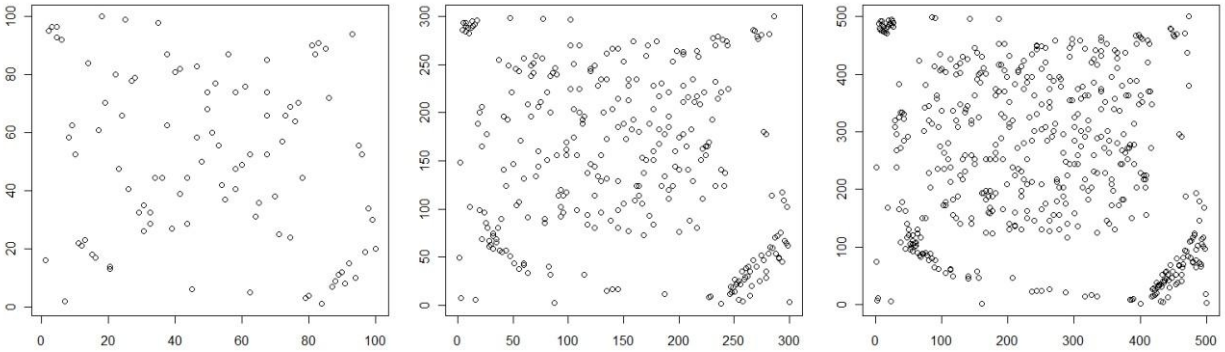


Imagem 2.1.4 – ranks das amostras da imagem 2.1.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.037	0.061	0.003	0.142	0.042	0.669
200	0.001	0.003	$<10^{-3}$	0.004	$<10^{-3}$	0.792
300	$<10^{-3}$	0.002	$<10^{-3}$	0.001	$<10^{-3}$	0.401
400	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.316
500	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.329
600	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.998
700	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.191
800	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.059
900	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.205
1000	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.098

Tabela 2.1: testes de correlação aplicado em amostras da imagem 2.1.1

Conclusão cenário 2

Podemos definir esse desempenho do teste de Pearson pela sua habilidade em detectar dependência na presença de padrões de formato oval, que podem ser relacionados a dados Normalmente distribuídos, ou certas misturas de dados Normalmente distribuídos. Os outros testes conseguiram rejeitar a independência também porém para uma amostragem maior, já o teste LIS não conseguiu detectar essa independência apesar de seguir uma linha para a rejeição no início, ele acabou oscilando não rejeitando a independência, claramente o teste não mostra consistência.

Cenário 3

O terceiro cenário é onde os testes de Spearman, Kendall e Pearson não conseguem identificar uma correlação entre os pontos das amostras. Este cenário pode ser dividido em 3 subcenários.

Cenário 3.1

O primeiro subcenário é onde o teste de Hoeffding e Copula conseguem um resultado bom para valores de amostras pequenos, e o LIS não consegue definir uma resposta concreta (falta consistência.).



Imagem 3.1.1 – imagem analisada

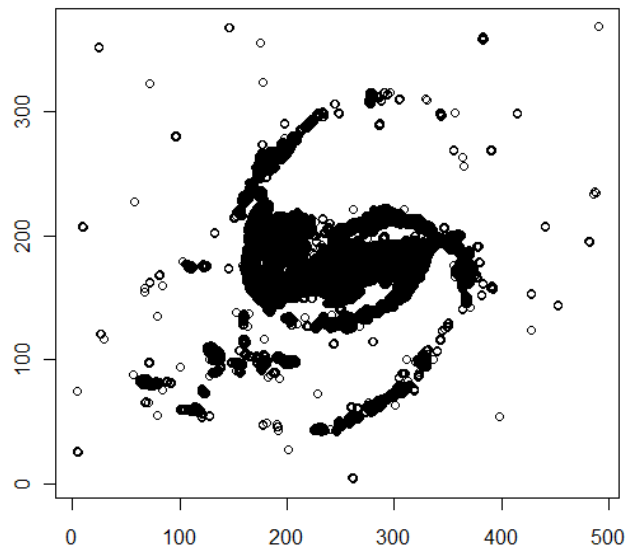


Imagem 3.1.2 – gráfico de todos os pontos coletados da imagem 3.1.1

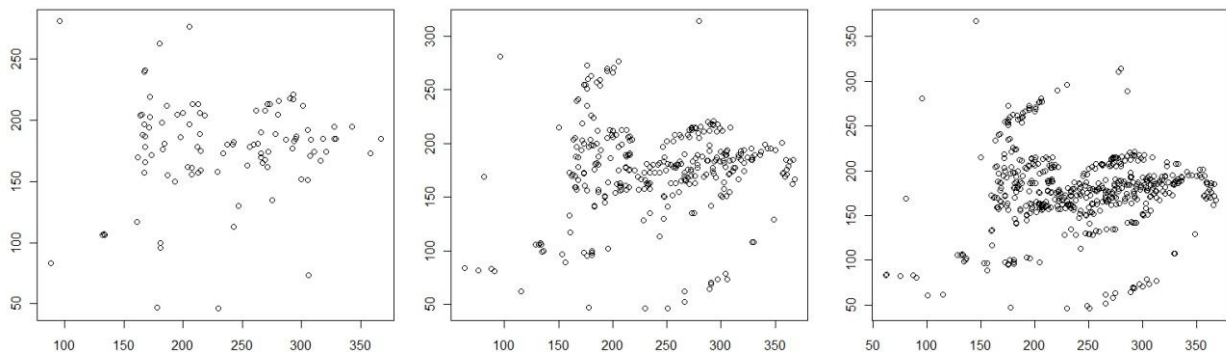


Imagem 3.1.3 – amostras de 100, 300, 500 pontos da imagem 3.1.2

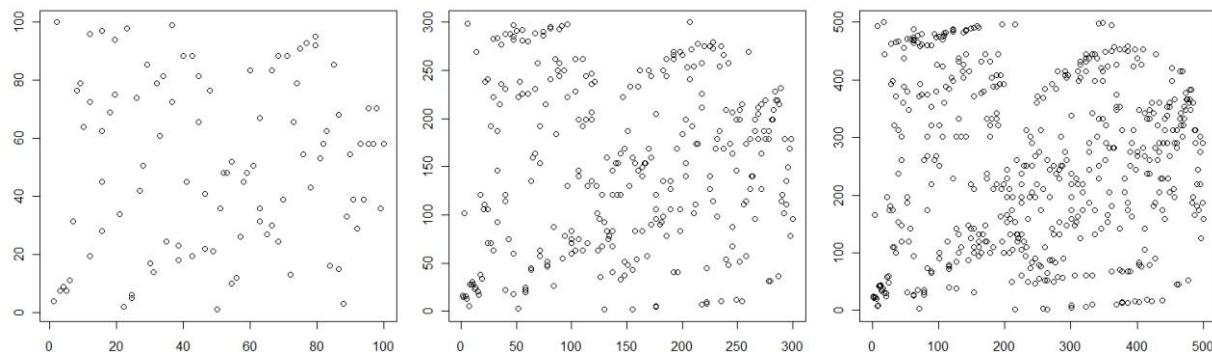


Imagem 3.1.4 – ranks das amostras da imagem 3.1.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.532	0.391	0.356	0.116	0.114	0.696
200	0.352	0.247	0.217	0.004	0.002	0.114
300	0.164	0.093	0.126	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
400	0.181	0.112	0.101	$<10^{-3}$	$<10^{-3}$	0.044
500	0.232	0.123	0.316	$<10^{-3}$	$<10^{-3}$	0.104

Tabela 3.1: testes de correlação aplicado em amostras da imagem 3.1.1

Conclusão cenário 3.1

Neste primeiro subcenário, especulamos que o notório desempenho dos testes de Hoeffding e Copula é decorrente da fundamentação das estatísticas desses testes, todos baseados no conceito de independência, i.e. eles comparam estimativas da distribuição conjunta com o produto das marginais. As imagens dos postos, mostram a clara aparição de espaços em branco, para tamanhos amostrais maiores ou iguais a 300. Já o teste LIS teve um bom desempenho até tamanho 300 de amostras, porém ao passar dessa quantidade de amostras, ele mostrou inconsistência, não sendo capaz de rejeitar a independência a um nível pequeno e fixo (por exemplo 0.01). Nota-se também a falta de consistência verificada nos testes baseados em postos (Spearman e Kendall).

Cenário 3.2

No segundo subcenário, temos os testes Hoeffd e Copula dando bons resultados para valores de amostras médios (maiores que 300).



Imagem 3.2.1 – imagem analisada

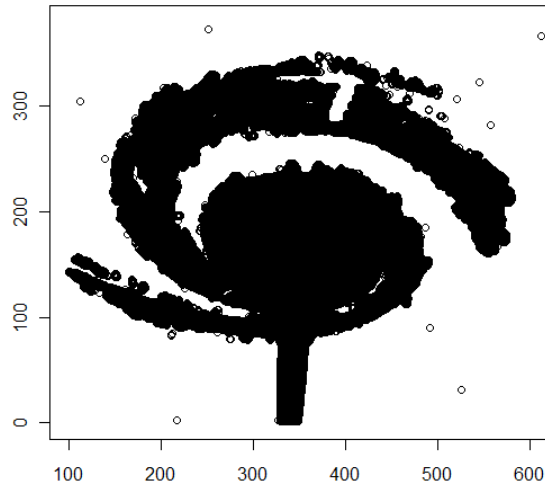


Imagem 3.2.2 – gráfico de todos os pontos coletados da imagem 3.2.1

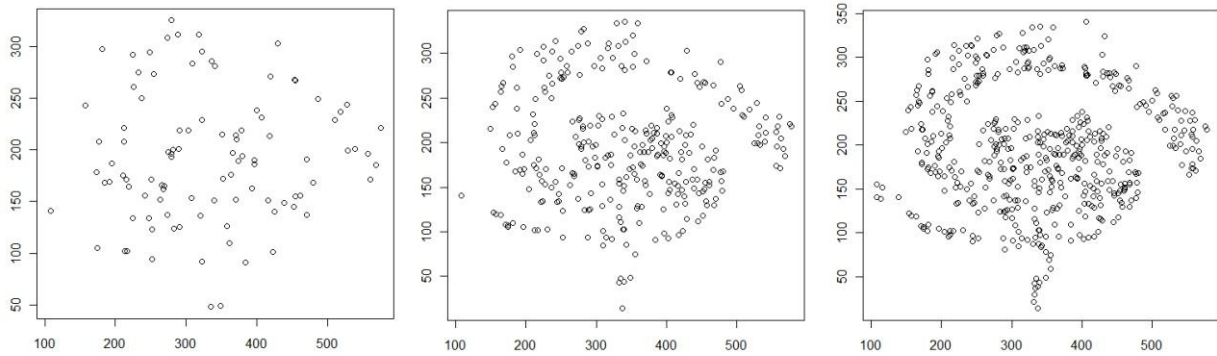


Imagem 3.2.3 – amostras de 100, 300, 500 pontos da imagem 3.2.2

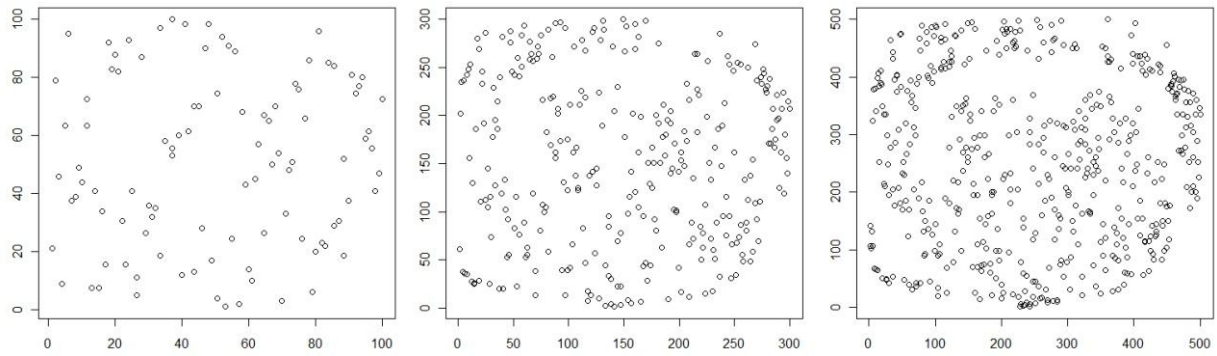


Imagem 3.2.4 – ranks das amostras da imagem 3.2.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.515	0.520	0.662	0.381	0.502	0.187
200	0.707	0.694	0.899	0.220	0.164	0.003
300	0.745	0.864	0.679	0.051	0.032	0.003
400	0.824	0.991	0.927	0.025	0.020	0.074
500	0.926	0.969	0.988	0.009	0.010	0.002

Tabela 3.2: testes de correlação aplicado em amostras da imagem 3.2.1

Conclusão cenário 3.2

Para essa imagem, o teste LIS mostrou-se mais sensível, porém houve uma pequena oscilação no resultado para 400 amostras, decorrente da inconsistência própria do teste.

Cenário 3.3

No terceiro e ultimo subcenário, temos uma resposta boa para os testes de Hoeffd e Copula para valores muito grandes de amostras, enquanto o teste LIS consegue um resultado bom.

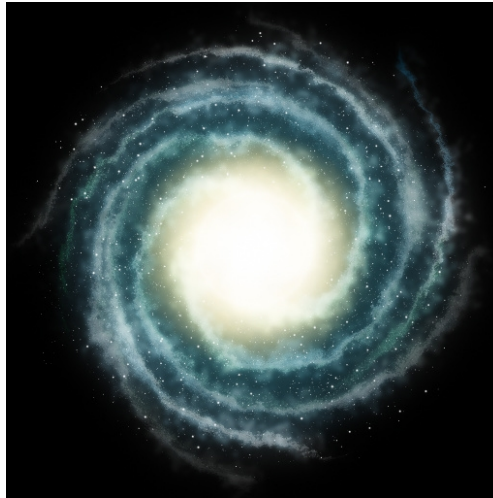


Imagem 3.3.1 - imagem analisada

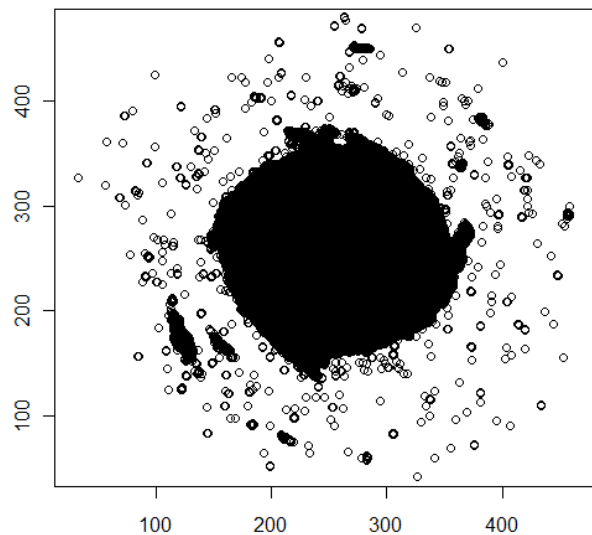


Imagem 3.3.2 - gráfico de todos os pontos coletados da imagem 3.3.1

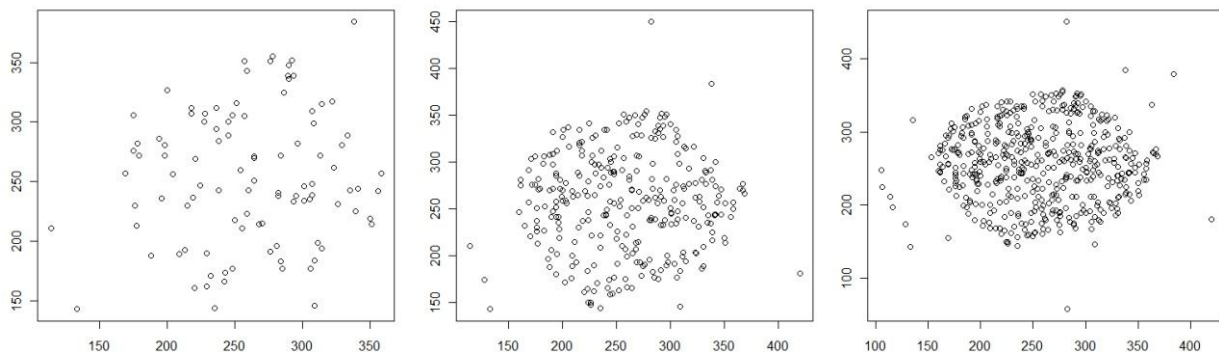


Imagem 3.3.3 - amostras de 100, 300, 500 pontos da imagem 3.3.2

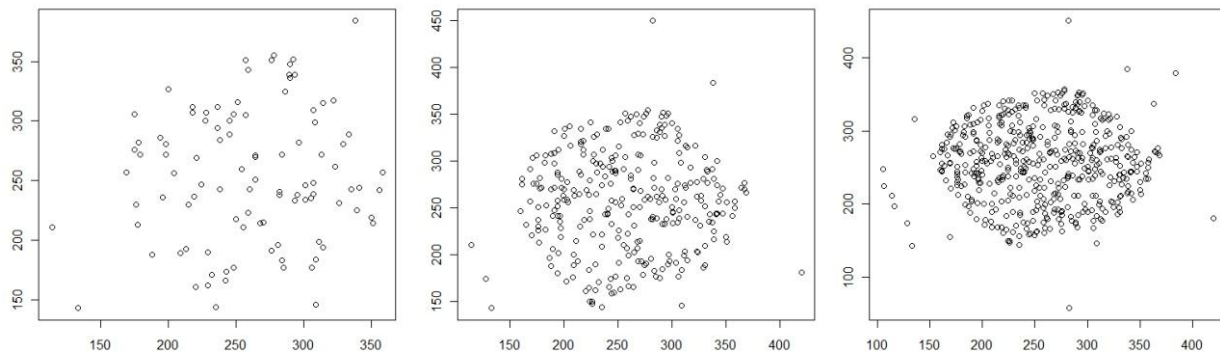


Imagem 3.2.4 – ranks das amostras da imagem 3.2.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.509	0.467	0.307	0.412	0.481	0.869
200	0.674	0.576	0.356	0.203	0.267	0.051
300	0.771	0.670	0.498	0.196	0.264	0.008
400	0.942	0.837	0.602	0.142	0.207	0.005
500	0.984	0.916	0.649	0.112	0.176	0.002
600	0.937	0.858	0.623	0.073	0.138	0.001
700	0.889	0.847	0.649	0.051	0.115	0.001
800	0.938	0.878	0.519	0.044	0.075	0.001
900	0.833	0.897	0.673	0.024	0.040	$<10^{-3}$
1000	0.899	0.957	0.589	0.018	0.032	0.001

Tabela 3.3: testes de correlação aplicado em amostras da imagem 3.3.1

Conclusão cenário 3.3

Para esse tipo de imagem, uma grande concentração circular de pontos com alguns pontos espalhados esparsamente ao redor, o teste LIS é, sem dúvidas, o melhor teste a ser aplicado, encontrando uma dependência já para amostras pequenas, enquanto os testes Copula e Hoeffd não definiram a correlação antes de 500 amostras, e os testes mais tradicionais não conseguem definir para nenhuma quantidade de amostra.

Cenário 4

E como ultimo cenário, obtemos uma resposta ruim de Pearson, uma resposta boa de Kendall e Spearman para amostras maiores que 200~300, e uma sensibilidade maior ainda para os testes de Hoeffd e Copula. No caso, conjecturamos que como os testes que apresentam desempenho bom são baseados em postos esse pode ser o motivo da ineficiência do Pearson, pois as distribuições marginais devem ter um efeito sobre o desempenho dele, assim sendo o que aparentemente suportaria um cenário de independência para o Pearson, é uma situação de dependência clara para testes baseados nos postos das observações.

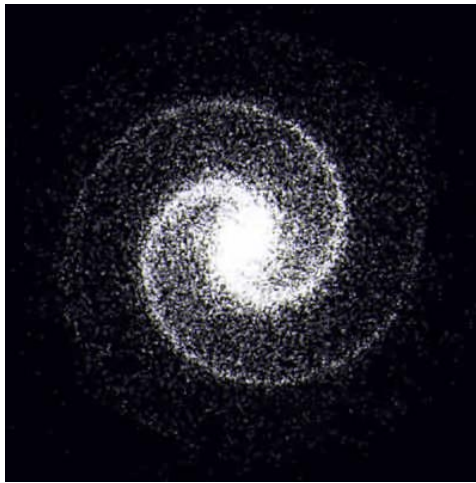


Imagem 4.1.1 - imagem analisada

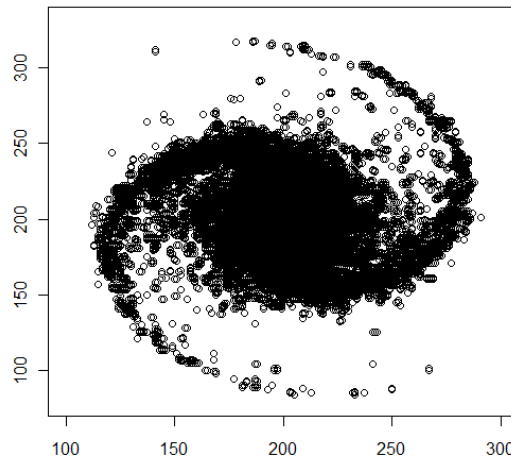


Imagem 4.1.2 - gráfico de todos os pontos coletados da imagem 4.1.1

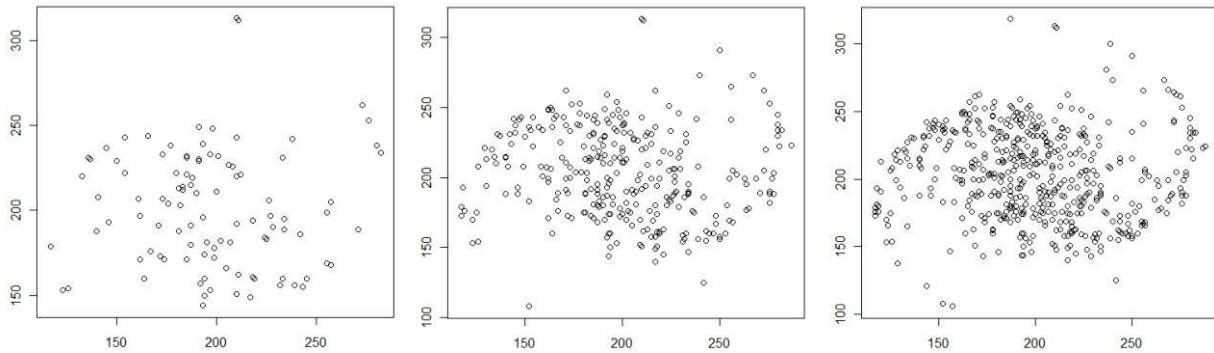


Imagem 4.1.3 - amostras de 100, 300, 500 pontos da imagem 4.1.2

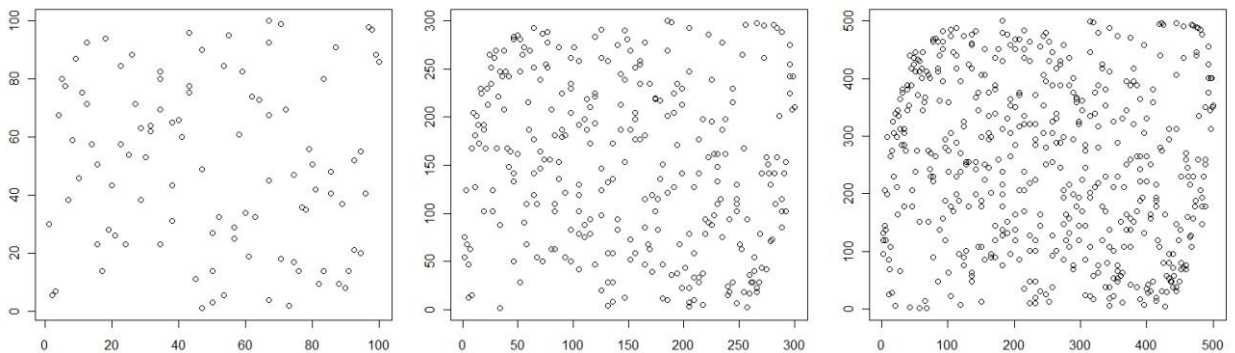


Imagem 4.1.4 - ranks das amostras da imagem 4.1.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.416	0.456	0.990	0.452	0.184	0.021
200	0.515	0.510	0.614	0.208	0.136	0.030
300	0.005	0.005	0.156	$<10^{-3}$	$<10^{-3}$	0.001
400	$<10^{-3}$	$<10^{-3}$	0.075	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
500	0.008	0.009	0.368	$<10^{-3}$	$<10^{-3}$	0.002

Tabela 4.1: testes de correlação aplicado em amostras da imagem 4.1.1

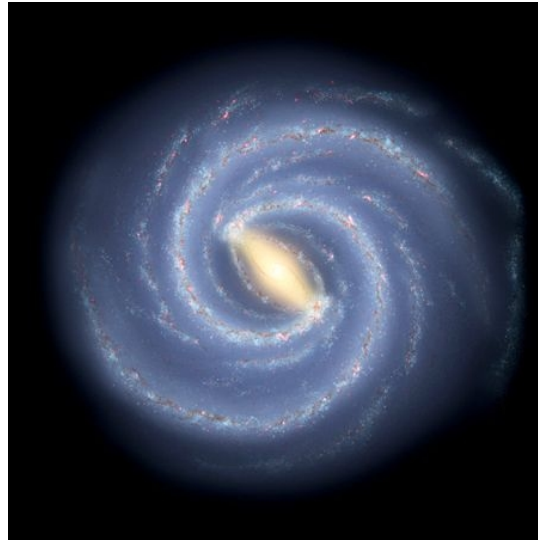


Imagem 4.1.1 – imagem analisada

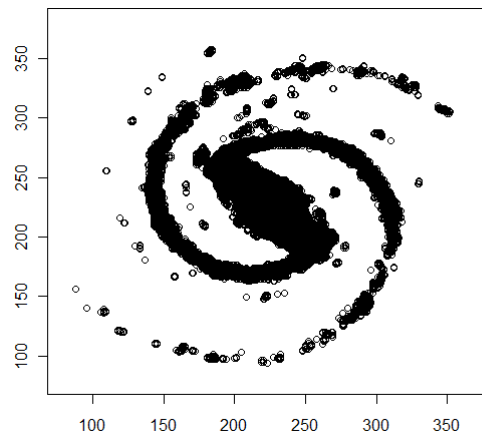


Imagem 4.2.2 – gráfico de todos os pontos coletados da imagem 4.2.1

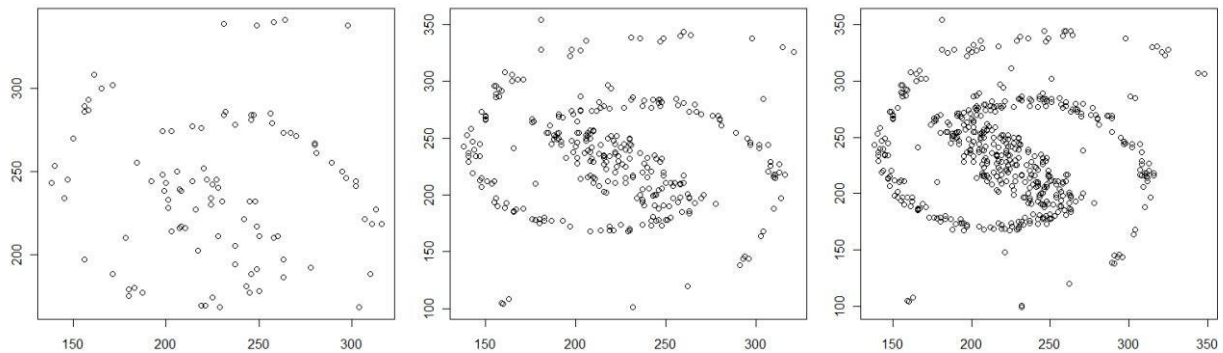


Imagem 4.2.3 – amostras de 100, 300, 500 pontos da imagem 4.2.2

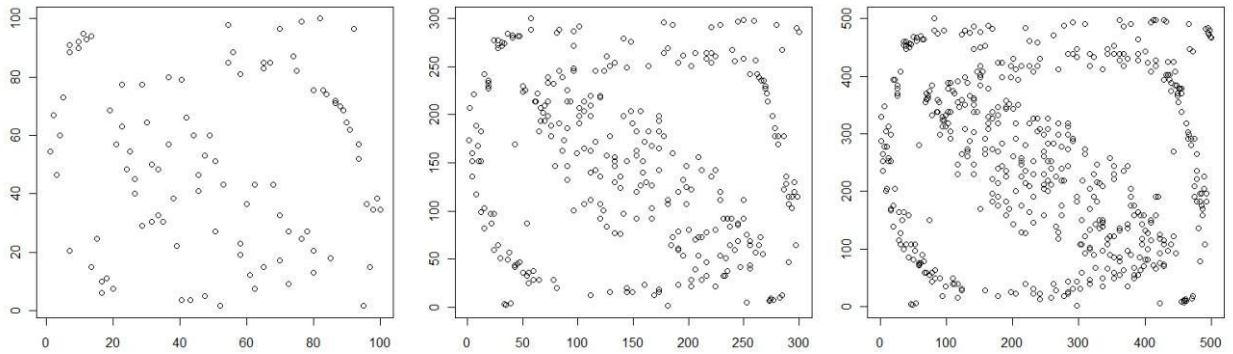


Imagem 4.2.4 – ranks das amostras da imagem 4.2.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.555	0.433	0.591	0.368	0.399	0.789
200	0.021	0.010	0.150	$<10^{-3}$	0.009	0.126
300	0.006	0.001	0.100	$<10^{-3}$	$<10^{-3}$	0.071
400	0.011	0.002	0.197	$<10^{-3}$	$<10^{-3}$	0.011
500	0.018	0.004	0.392	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$

Tabela 4.2: testes de correlação aplicado em amostras da imagem 4.2.1

Conclusão cenário 4

O Teste LIS, para esse tipo de imagem espiralada, acompanha, em média, os testes de Kendall e Spearman, tendo um desempenho razoável para amostras pequenas, e um bom desempenho para amostras maiores. Os testes que se destacam são Hoeffding e Copula.

Dados Extras

Analisando os resultados obtidos, chegamos a alguns padrões de imagens que o teste LIS apresenta algumas respostas singulares e que foram explorados em profundidade.

Circunferências

Para circunferências, o teste LIS tem uma resposta boa porém há uma oscilação dependendo da quantidade de amostras. O teste LIS acompanhou o bom desempenho de Hoeffding e Copula

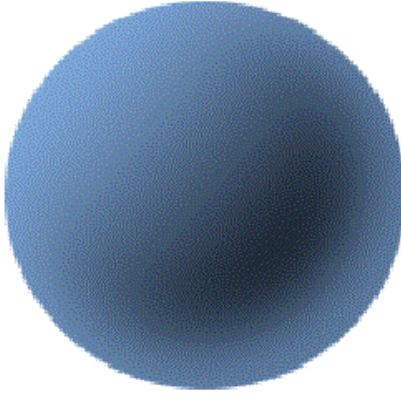


Imagem 5.1.1 - imagem analisada

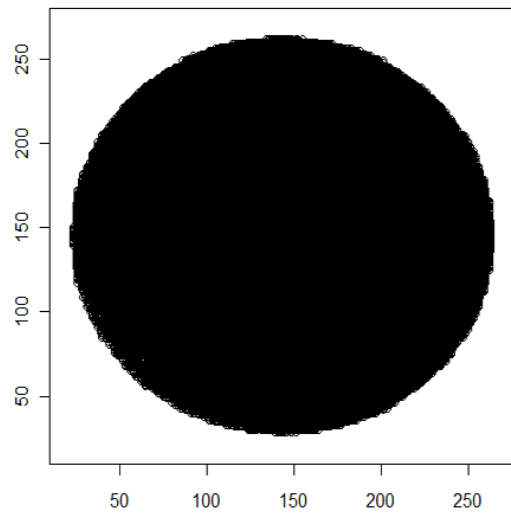


Imagem 5.1.2 - gráfico de todos os pontos coletados da imagem 5.1.1

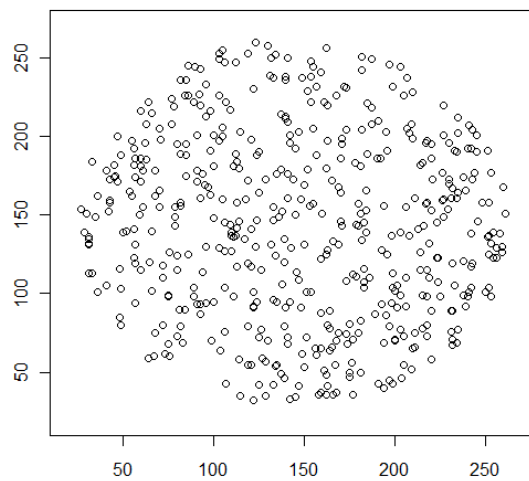


Imagem 5.1.3 - amostra de 500 pontos da imagem 5.1.2

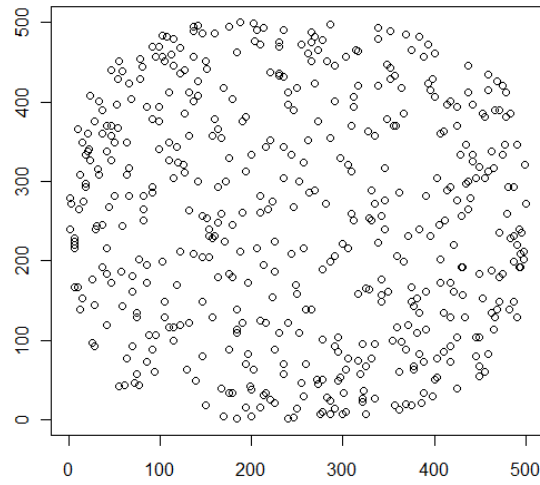


Imagem 5.1.4 – ranks dos pontos da imagem 5.1.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.349	0.388	0.372	0.418	0.545	0.003
200	0.096	0.127	0.115	0.045	0.040	0.001
300	0.091	0.133	0.087	0.025	0.033	0.157
400	0.027	0.051	0.032	0.008	0.006	$<10^{-3}$
500	0.031	0.058	0.037	0.004	0.003	$<10^{-3}$

Tabela 5.1: testes de correlação aplicado em amostras da imagem 5.1.1

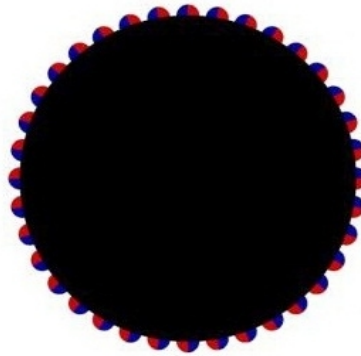


Imagem 5.2.1 – imagem analisada

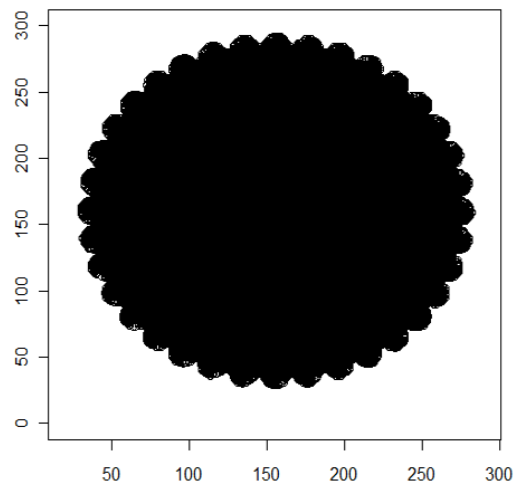


Imagem 5.2.2 – gráfico de todos os pontos coletados da imagem 5.2.1

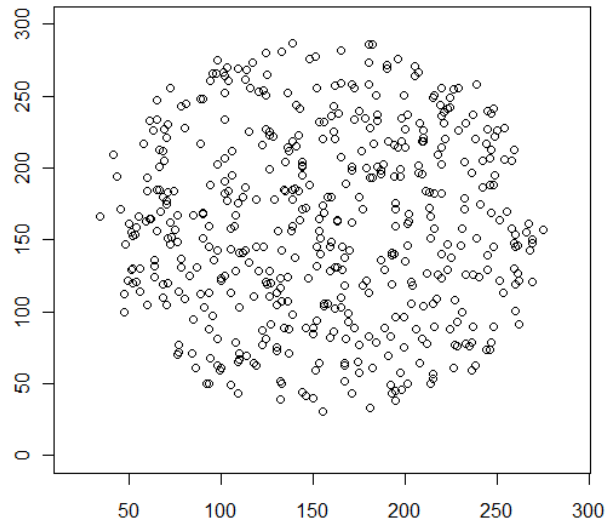


Imagem 5.2.3 – amostra de 500 pontos da imagem 5.2.2

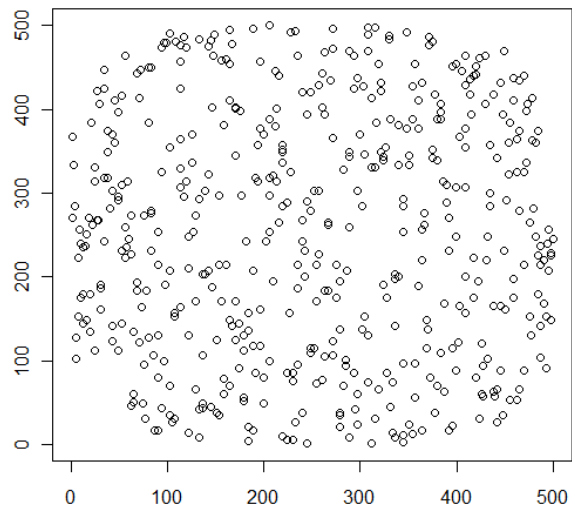


Imagem 5.2.4 – ranks dos pontos da imagem 5.2.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.388	0.378	0.365	0.453	0.730	<10-3
200	0.496	0.510	0.469	0.451	0.607	0.056
300	0.991	0.972	0.908	0.326	0.430	0.169
400	0.680	0.767	0.775	0.083	0.094	0.867
500	0.707	0.758	0.808	0.176	0.193	0.312
600	0.781	0.812	0.864	0.191	0.268	0.234
700	0.761	0.746	0.702	0.152	0.222	0.032
800	0.855	0.845	0.792	0.137	0.184	<10-3
900	0.663	0.670	0.610	0.119	0.158	0.005
1000	0.711	0.720	0.657	0.100	0.142	0.011

Tabela 5.2: testes de correlação aplicado em amostras da imagem 5.2.1

Coroas

Para coroas, o teste LIS tem um ótimo desempenho, superando todos os outros em quase todas as quantias de amostras e sendo comparável com Hoeffding e Copula.



Imagem 6.1.1 – imagem analisada

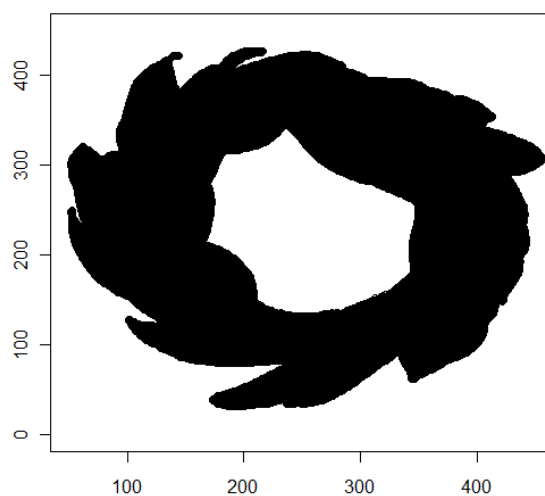


Imagem 6.1.2 – gráfico de todos os pontos coletados da imagem 6.1.1

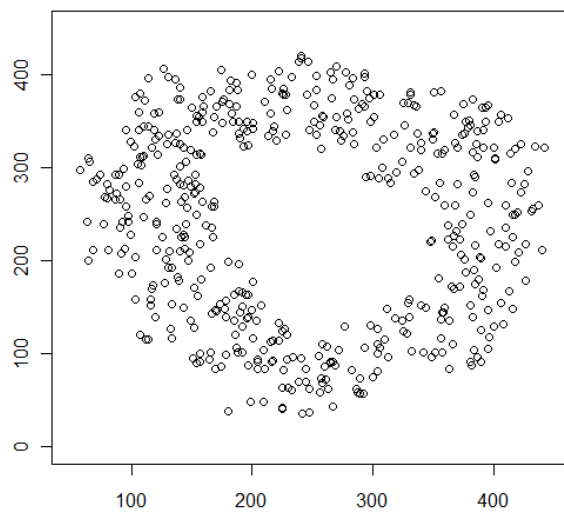


Imagem 6.1.3 – amostra de 500 pontos da imagem 6.1.2

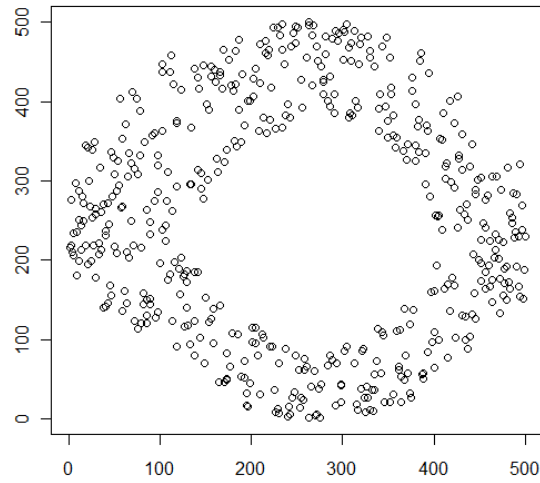


Imagem 6.1.4 – ranks dos pontos da imagem 6.1.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.652	0.933	0.666	0.093	0.091	0.004
200	0.512	0.368	0.488	0.010	0.023	0.002
300	0.596	0.390	0.739	0.001	0.002	0.003
400	0.493	0.396	0.533	$<10^{-3}$	0.004	0.003
500	0.241	0.195	0.268	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$

Tabela 6.1: testes de correlação aplicado em amostras da imagem 6.1.1



Imagem 6.2.1 – imagem analisada

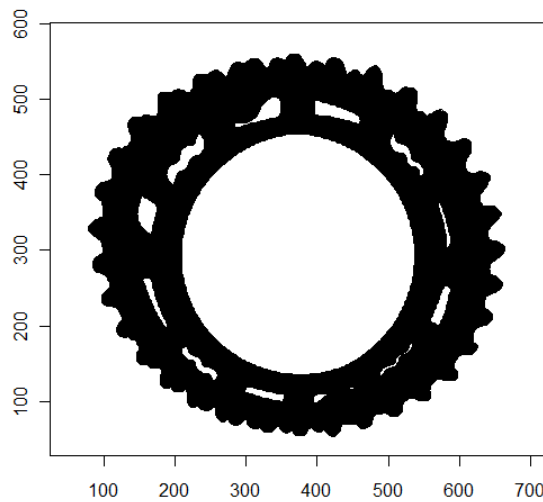


Imagem 6.2.2 – gráfico de todos os pontos coletados da imagem 6.2.1

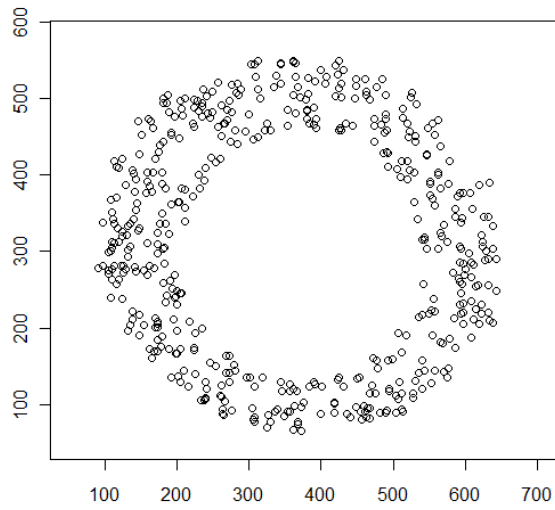


Imagem 6.2.3 – amostra de 500 pontos da imagem 6.2.2

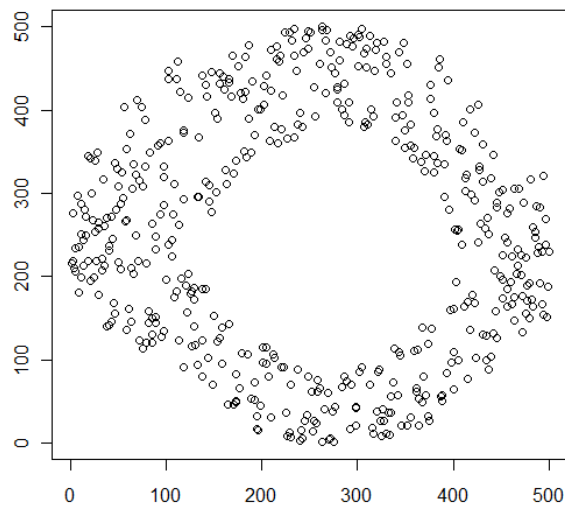


Imagem 6.2.4 – ranks dos pontos da imagem 6.2.3

Número de Amostras	Spearman	Kendall	Pearson	Hoeffd	Copula	LIS
100	0.744	0.611	0.708	0.006	0.032	$<10^{-3}$
200	0.438	0.362	0.367	$<10^{-3}$	0.004	0.021
300	0.334	0.280	0.313	$<10^{-3}$	0.001	$<10^{-3}$
400	0.351	0.280	0.364	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
500	0.158	0.191	0.111	$<10^{-3}$	$<10^{-3}$	0.009

Tabela 6.2: testes de correlação aplicado em amostras da imagem 6.2.1

Apêndice – Relatório Técnico

Primeiramente introduziremos os conceitos de imagens digitais para que o leitor possa acompanhar a natureza dos nossos dados.

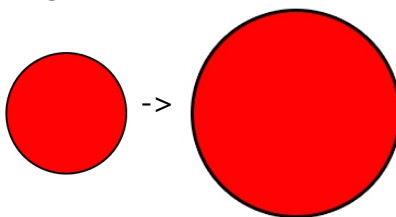
Tipos de imagens digitais

Há, fundamentalmente, dois tipos de imagem digital. Uma delas é do tipo *vetorial* e a outra é do tipo *rastreio*.

Imagens do tipo vetorial são imagens formadas por atributos, sendo esses atributos linhas, curvas, formatos simples e cores, por exemplo. Um círculo feito vetorialmente terá a informação que é um círculo, o seu raio e sua cor. Se quisermos aumentar o tamanho desse círculo mudaríamos apenas o atribuído para seu raio não perdendo a qualidade da imagem.

Frequentemente essa formatação é usada para logotipos de empresas, gráficos de jogos e interfaces de programas.

Ex: imagem de uma circunferência feita vetorialmente, ao aumentarmos a imagem, ela continua com sua qualidade original



Já as imagens do tipo rastreio são constituídas por pixels e podem ser representadas por uma matriz, onde cada elemento dessa matriz contém a informação de cor. A visualização dessa matriz com os elementos sendo representados pelas suas respectivas cores nos formará a imagem desejada. Esse tipo de imagem é muito utilizada para fotos digitalizadas pois não é necessário que a imagem tenha formas simples para ser gerada. O ponto fraco desse tipo de imagem é quando precisamos aumentá-las, como é feita por pixels aumentamos o tamanho de cada pixel, gerando uma imagem quadriculada e de aspecto ruim.

Frequentemente essa formatação é usada para fotografias e imagens cujas limitações vetoriais não permitem seu processamento.

Ex: imagem de uma circunferência feita do tipo rastreio, ao aumentarmos podemos ver seus pixels



Nesse trabalho utilizamos imagens do tipo rastreio.

Preparação do programa

Para a aquisição e processamento dos dados, foutilizado o programa R, que pode ser obtido através do website oficial (<http://www.r-projeci t.org>)

Para todo o processo, é necessário ter as seguintes bibliotecas instaladas:

- PNG

- Hmisc
- Copula

Para instalar as bibliotecas, você deve seguir o seguinte procedimento:

- Abra o programa R
- No menu superior, vá na opção "Packages->Install package(s)"
- Selecione um mirror (o utilizado para o processamento foi "Brasil (SP 1)")
- Selecione o pacote a ser instalado e confirme a instalação
- Faça este processo para todas as bibliotecas necessárias

Antes de iniciar o processamento, é necessário definir um diretório de trabalho. Para definir um diretório de trabalho, siga os seguintes passos:

- Na janela do programa vá na opção "File->Change dir..."
- Selecione a pasta que deseja para seu diretório de trabalho
- Confirme clicando em "Ok"

Detalhe: a imagem a ser analisada deve, necessariamente, estar na pasta definida como diretório de trabalho e, necessariamente, estar em uma extensão ".png" (formato de processamento do pacote PNG).

Processamento da imagem

Na aquisição dos pontos da figura foi usado a seguinte rotina:
(Nessa rotina, o processo pega todos os pontos mais "escuros" e forma um gráfico com os pontos coletados)

```
#comando para carregar o pacote png
library(png)

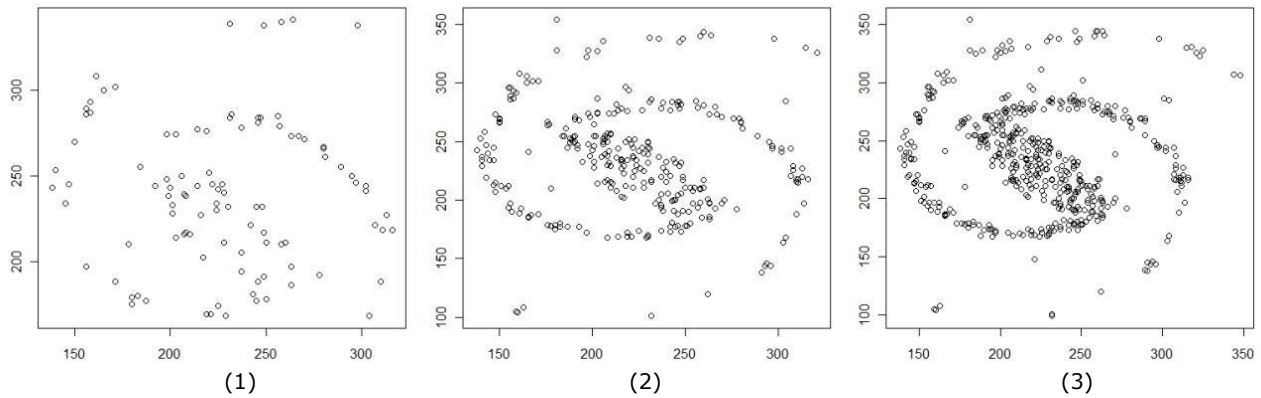
#comando para ler a imagem a ser analisada no formato png
img<-readPNG("imagem.png",native=FALSE)
#esse comando monta uma matriz com os dados da cor de cada pixel da imagem

#ler posições dos pontos da imagem
L<-numeric(length=0)
C<-numeric(length=0)
l=0
#preencher com o tamanho da largura da imagem
tam2=
#preencher com o tamanho da altura da imagem
tam1=
#para obter esses tamanhos, clique com o botão direito em cima do arquivo da imagem
e vá na opção "propriedades" e aparecerá os dados "largura x altura"
for(i in 1:tam1){
  for(j in 1:tam2){
    L[l]=i
    C[l]=j
    #pode-se trocar esse valor para ajustar a intensidade do pixel a ser considerado
    if(img[i,j]>0.6)
      l=l+1
  }
}
```

#essa rotina passará elemento por elemento da matriz, guardando as posições dos elementos definidos na condição "if"

Objetivo

Para melhor entendimento da necessidade da seguinte rotina, uma breve explicação do objetivo do trabalho:



Como ilustrado nas imagens (1), (2) e (3) desejamos mostrar a eficiência da resposta de cada teste quando o tamanho de pontos é aumentado, isto é, o teste mais eficiente será aquele que retornar um p-valor baixo para um menor tamanho de amostra sendo capaz de detectar a imagem com menos informação (menor tamanho amostral).

Para a amostragem, os seguintes comandos foram usados (foi feita uma amostragem de 1000 pontos para caso seja necessária um teste para grandes valores de amostragem):

```
#passar uma amostra de 1000 conjuntos de pontos aleatórios da imagem para uma matriz
a<-sample(1,1000)
matriz<-matrix(nrow=1000,ncol=2)
#cria uma matriz cuja primeira coluna são valores "x" e a segunda coluna são valores "y"
matriz[,1]=C[a]
matriz[,2]=L[a]
```

Montar uma tabela com os testes para 100, 200, 300, 400 e 500 amostras dos pontos

```
#carrega as bibliotecas necessárias para os testes de independência
library(Hmisc)
library(copula)

#monta uma matriz para guardar os resultados dos testes
testes<-matrix(nrow=5,ncol=5)

#teste de spearman
testes[1,1]<-cor.test(matriz[1:100,1],matriz[1:100,2],method="spearman")$p.value
testes[2,1]<-cor.test(matriz[1:200,1],matriz[1:200,2],method="spearman")$p.value
testes[3,1]<-cor.test(matriz[1:300,1],matriz[1:300,2],method="spearman")$p.value
testes[4,1]<-cor.test(matriz[1:400,1],matriz[1:400,2],method="spearman")$p.value
testes[5,1]<-cor.test(matriz[1:500,1],matriz[1:500,2],method="spearman")$p.value

#teste de kendall
testes[1,2]<-cor.test(matriz[1:100,1],matriz[1:100,2],method="kendall")$p.value
testes[2,2]<-cor.test(matriz[1:200,1],matriz[1:200,2],method="kendall")$p.value
testes[3,2]<-cor.test(matriz[1:300,1],matriz[1:300,2],method="kendall")$p.value
testes[4,2]<-cor.test(matriz[1:400,1],matriz[1:400,2],method="kendall")$p.value
testes[5,2]<-cor.test(matriz[1:500,1],matriz[1:500,2],method="kendall")$p.value

#teste de Pearson
testes[1,3]<-cor.test(matriz[1:100,1],matriz[1:100,2],method="pearson")$p.value
testes[2,3]<-cor.test(matriz[1:200,1],matriz[1:200,2],method="pearson")$p.value
testes[3,3]<-cor.test(matriz[1:300,1],matriz[1:300,2],method="pearson")$p.value
testes[4,3]<-cor.test(matriz[1:400,1],matriz[1:400,2],method="pearson")$p.value
testes[5,3]<-cor.test(matriz[1:500,1],matriz[1:500,2],method="pearson")$p.value
```

```
#teste Hoeffd
testes[1,4]<-hoeffd(matriz[1:100,1],matriz[1:100,2])$P[1,2]
testes[2,4]<-hoeffd(matriz[1:200,1],matriz[1:200,2])$P[1,2]
testes[3,4]<-hoeffd(matriz[1:300,1],matriz[1:300,2])$P[1,2]
testes[4,4]<-hoeffd(matriz[1:400,1],matriz[1:400,2])$P[1,2]
testes[5,4]<-hoeffd(matriz[1:500,1],matriz[1:500,2])$P[1,2]
```

```
#teste Copula
d <- indepTestSim(100,2)
testes[1,5]<-indepTest(matriz[1:100,1:2],d)$pvalue
d <- indepTestSim(200,2)
testes[2,5]<-indepTest(matriz[1:200,1:2],d)$pvalue
d <- indepTestSim(300,2)
testes[3,5]<-indepTest(matriz[1:300,1:2],d)$pvalue
d <- indepTestSim(400,2)
testes[4,5]<-indepTest(matriz[1:400,1:2],d)$pvalue
d <- indepTestSim(500,2)
testes[5,5]<-indepTest(matriz[1:500,1:2],d)$pvalue
```

#o teste LIS foi rodado externamente ao R por problemas de processamento, porém é possível obter sua biblioteca para rodá-lo no próprio programa R

Salvar os dados em um arquivo "dados.RData" no diretório de trabalho

```
save(img, C, L, l, a, matriz, x, testes, file = "dados.RData")
```

Para acessar a matriz com os resultados o seguinte comando deve ser digitado:

```
testes
#na matriz mostrada cada linha corresponde a certa quantidade de amostras, para esse programa a linha 1 são 100 amostras, a linha 2 são 200 amostras, a linha 3 são 300 amostras, a linha 4 são 400 amostras e a linha 5 são 500 amostras. E cada coluna corresponde a um teste, a primeira coluna é o teste de Spearman, a segunda coluna é o teste de Kendall, a terceira coluna é o teste de Pearson, a quarta coluna é o teste Hoeffd e a quinta coluna é o teste Copula
```