

Análise Espacial de Informação Geográfica

Análise Espacial dos Supermercados nos Distritos de Lisboa e Setúbal

Helena Alves

(fc25802@alunos.fc.ul.pt)

RESUMO

“Compreender a distribuição espacial de dados oriundos de fenómenos ocorridos no espaço constitui hoje um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento.” Neste contexto os sistemas de informação geográfica (SIG) assumem um papel de extrema importância no sector económico das empresas, suportando muitas das suas tomadas de decisão. A visualização espacial de um determinado fenómeno facilita, entre outros aspectos, a definição e identificação de novas estratégias de expansão das empresas. O sector do comércio a retalho usufrui bastante das vantagens e benefícios da utilização destes sistemas. Inspirado nestes factos e no crescente aumento nos últimos anos de supermercados, o presente trabalho visou investigar a existência de padrões sistemáticos, identificar os factores que controlam espacialmente esses padrões assim como avaliar o quão acessíveis e próximos estão de determinadas localidades, entre outros. Os supermercados alvo de análise foram: Pingo Doce, LIDL, Jumbo, Intermarché, Feira Nova, E.Leclerc, Continente e Modelo Bonjour e a região de estudo considerada englobou os distritos de Lisboa e Setúbal. Ao longo deste documento será apresentada a metodologia utilizada assim como as conclusões resultantes da utilização de diversas ferramentas de análise espacial e manipulação da informação.

1. Introdução

Nos últimos tempos, tem-se verificado um crescimento considerável no sector de comércio a retalho em supermercados no interior das grandes áreas urbanas. Este crescimento, responsável pela geração de mais empregos, reflecte-se actualmente numa forte presença espacial, influenciando por vezes na própria dinâmica das cidades. Os supermercados expandiram-se de tal forma que quase todos nós temos um, ou mais, supermercados a menos de 10 minutos das nossas casas.

No entanto muitas vezes somos confrontados com questões que nos levam a cogitar porque motivo é que a nossa área de residência não possui um supermercado Feira Nova? Porque que motivo não temos um Pingo Doce ou E.Leclerc perto de nós? O que para nós é por vezes incompreensível, é explicado e sustentado por estratégias de expansão económicas consistentes e bem concertadas, alinhadas com os objectivos de cada cadeia de supermercados.

O presente trabalho, embora tenha como principal objectivo aplicar os conhecimentos teóricos e técnicos adquiridos ao longo do semestre no âmbito da disciplina de Análise Espacial de Informação Geográfica, pretende proceder à “Análise Espacial dos Supermercados nos distritos de Lisboa e Setúbal” de forma a:

- conhecer alguns dos princípios orientadores que norteiam as estratégias de expansão;
- identificar a distribuição espacial de cada cadeia de supermercados;
- identificar possíveis relações espaciais;
- analisar a variabilidade espacial de um fenómeno com a alteração da sua localização;
- entre outros aspectos.

Dados os constrangimentos temporais esta análise focou-se fundamentalmente sobre o conjunto global de supermercados (i.e., conjunto de dados formados por todos os supermercados) existentes nos distritos de Lisboa e Setúbal, e na análise individual dos seguintes supermercados: Pingo Doce, LIDL, Jumbo, Intermarché, Feira Nova, E.Leclerc, Continente e Modelo Bonjour.

Ao longo do presente documento, serão apresentadas as principais conclusões obtidas da análise espacial do padrão pontual (capítulo 3), análise espacial de áreas (capítulo 4), regressão linear (capítulo 5) e análise de redes (capítulo 6). Os resultados apresentados ao longo deste trabalho foram obtidos através dos seguintes softwares: ArcGis, IDRISI, OpenGeoDa e SPRING.

2. Aquisição e Pré-processamento de Dados

A natureza e modo de aquisição dos dados é bastante variada e, obviamente, é determinada pelos objectivos do sistema. Dada a natureza deste trabalho e contrariamente ao que deveria suceder na realidade, não foi efectuado qualquer levantamento topográfico. Os dados foram adquiridos de modo indirecto, tendo-se utilizado para o efeito dados divulgados/partilhados na Internet assim como recorrido a serviços de mapas como o *Google Earth* e outros serviços como o *MapaList* (<http://mapalist.com/>) e *Kml2Shp* (<http://www.zonums.com/online/kml2shp.php>).

O serviço *MapaList* possibilita a criação e gestão personalizada de listas de endereços provenientes do *Google spreadsheets*, possibilitando entre outros aspectos a sua exportação para o *Google Earth*. O *KMLS2Shp* é um serviço que assegura a conversão dos dados *Google Earth* para *Shapefile*.

A tabela seguinte (tabela 1) sintetiza as principais características dos dados recolhidos e/ou resultantes da fase de pré-processamento:

Tabela 1 – Dados recolhidos e/ou resultantes do pré-processamento

Dados	Formato	Tipo	Fonte	Descrição
CAOP 2010	Vectorial	Shapefile de polígonos	Igeo	Carta administrativa oficial de Portugal
Lojas Pingo Doce	Vectorial	Shapefile de pontos	Sites Institucionais Páginas Amarelas Rede Solidária Outras fontes	Localização das lojas Pingo Doce
Lojas LIDL	Vectorial	Shapefile de pontos		Localização das lojas LIDL
Lojas Jumbo	Vectorial	Shapefile de pontos		Localização das lojas Jumbo
Lojas Intermarché	Vectorial	Shapefile de pontos		Localização das lojas Intermarché
Lojas Feira Nova	Vectorial	Shapefile de pontos		Localização das lojas Feira Nova
Lojas E.Leclerc	Vectorial	Shapefile de pontos		Localização das lojas E.Leclerc
Lojas Continente	Vectorial	Shapefile de pontos		Localização das lojas Continente
Lojas Modelo Bonjour	Vectorial	Shapefile de pontos		Localização das lojas Modelo Bonjour
Rede Viária	Vectorial	Shapefile de linhas		CloudMade
Dados INE: - Densidade populacional (N.º km ²) por Local de residência - Poder de compra per capita por Localização geográfica	Excel	Documento .xls	INE	Dados relativos ao ano 2007, da Densidade populacional por Local de residência e Poder de compra per capita por Localização geográfica

Como o universo de análise deste trabalho está restrito aos distritos de Setúbal e Lisboa seleccionou-se meramente a informação relativa a estas áreas administrativas. Para além das restrições de domínio, houve necessidade de se proceder a um conjunto de operações distintas sobre cada uma destas *shapefiles* assim como realizar diversos *joins* entre tabelas de forma a compilar nas tabelas toda a informação necessária para aplicação das técnicas de análise. (figura 1):

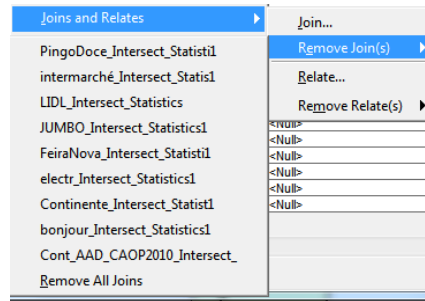


Figura 1 – Join de tabelas

2.1. – Constrangimentos

Contrariamente ao expectável, o levantamento das moradas de todos os supermercados constitui um processo bastante moroso e de alguma forma complexo. Com excepção das Lojas Pingo Doce que indicam explicitamente no site institucional as moradas de todas as suas lojas, constatou-se que as restantes cadeias de supermercados não disponibilizam esta informação de uma forma explícita e, em algumas situações, não facultam de todo esta informação.

Perante este facto houve necessidade de se consultar outras fontes de informação. No entanto, constatou-se que estas, quando confrontadas entre si, apresentavam incoerências evidenciando moradas diferentes para a mesma loja e inclusivamente moradas incorrectas.

Este constrangimento reflectiu-se, por sua vez, na identificação das lojas pelo serviço *MapaList*. Note-se que o rigor da localização deste serviço depende directamente da exactidão das moradas utilizadas. Para minimizar os erros, o processo de localização teve em conta apenas o código postal. No entanto a probabilidade de existência de duplicados e inconsistências é uma realidade.

Dada a natureza deste trabalho, optou-se por ignorar estas limitações, embora tendo sempre presente que era fundamental proceder-se à recolha das moradas de uma forma mais rigorosa e/ou solicitar as mesmas às respectivas instituições.

3. Análise do Padrão de Distribuição de Pontos

3.1. Visualização dos Dados Espaciais

Na análise de padrões pontuais, os eventos de interesse são identificados como pontos localizados no espaço, tendo como objecto de estudo a sua própria localização. No entanto, para além da localização geográfica dos eventos, é fundamental conhecer-se outros parâmetros estatísticos básicos para caracterização dos dados espaciais tais como:

- **Frequência:**

A frequência e, em particular a frequência absoluta, corresponde ao número de vezes que uma determinada variável é observada numa determinada área de estudo. No presente projecto, este parâmetro assume o valor global de **289**. Quando analisarmos individualmente cada uma das cadeias de supermercados (i.e., analisar só Pingo Doce, ou LIDL,..) este valor assume o mínimo de 1 e o máximo pode atingir o valor 55.

- **Central Feature:**

Identifica o evento localizado mais centralmente do conjunto de dados considerado. Esta *feature* tem associada a menor distância total às restantes *features*. Caso seja adicionado um peso na análise como o peso “população”, a *Central Feature* não é apenas o evento mais central, é o evento mais acessível à maior parte das pessoas.

A figura 2 indica o ponto mais central do conjunto global de dados considerado:

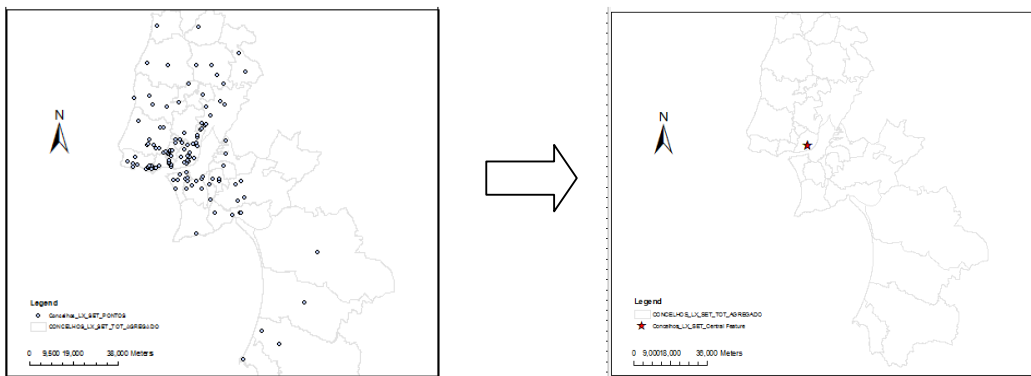


Figura 2 – *Central Feature* (conjunto global de eventos)

Analisando individualmente o conjunto de dados de cada uma das lojas, a *Central Feature* é representado por *features* diferentes, tal como se pode constatar na figura 3:

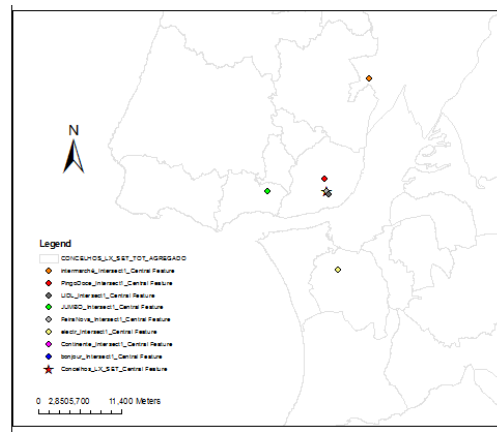


Figura 3 – *Central Feature* (conjunto específico de eventos)

Constata-se que, com excepção das lojas Intermarché, Jumbo e E.Leclerc, as *Central Features* localizam-se no concelho de Lisboa.

- **Centro Geométrico (ou centro médio ou média espacial):**

Representa a *feature* resultante do cálculo da média das coordenadas de localização X e Y. O centro geométrico é útil para comparar a distribuições de diferentes tipos de *features* ou para encontrar a localização com maior concentração de pontos com base no mesmo atributo. A figura seguinte (figura 4) apresenta o centro geométrico do conjunto de dados globais assim como o centro de cada um do conjunto de dados específicos.

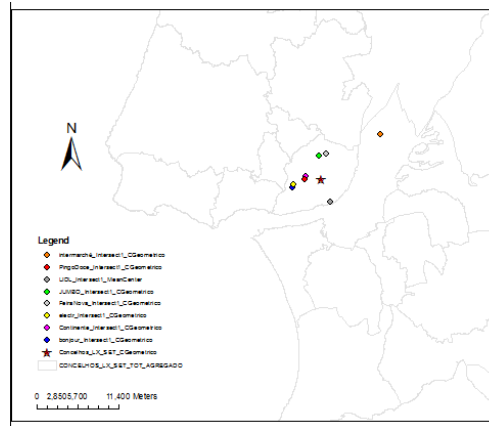


Figura 4 – Centro Geométrico

Com exceção do conjunto de dados relativo às lojas Intermarché, os centros geométricos dos restantes localizam-se todos no concelho de Lisboa.

- **Distância Padrão (ou dispersão espacial):**

Mede a concentração ou dispersão dos eventos em torno do centro geométrico. Pela figura 5, constata-se que o círculo 1 *standard deviation* (círculo azul) delimita, tal como esperado, grande parte dos eventos observados. Conclui-se assim que em círculos com raios menores os eventos são mais localizados enquanto que em círculos com raio maior a ocorrência de um evento é mais comum.

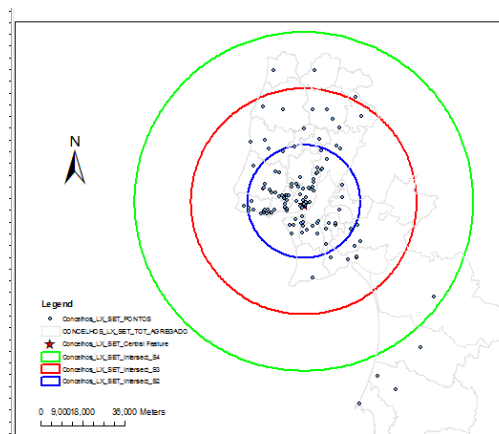


Figura 5 – Distância Padrão (conjunto global de eventos)

- **Distribuição Direccional (*Standard Deviation Ellipse*)**

Permite identificar a tendência direccional dos eventos, com vista a descrever a distribuição espacial e comparar geograficamente as diferentes lojas de supermercados. A figura seguinte apresenta a elipse padrão associada ao conjunto de pontos com a globalidade de lojas Pingo Doce, LIDL, Continente, Jumbo, Feira Nova, E.Leclerc, Modelo Bonjour e Intermarché.

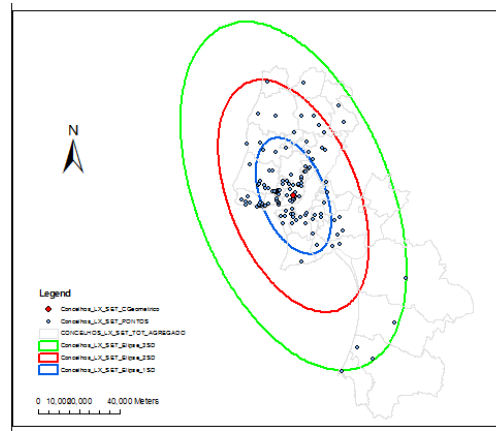


Figura 6 – Distribuição Direccional (conjunto global de eventos)

Uma análise mais atenta da elipse projectada constata que esta evidencia claramente a forma elipsoidal, sendo nítida a diferença entre o semi-eixo menor e maior. Este facto permite concluir que os eventos encontram-se concentrados segundo uma determinada direcção. Observando-se o mapa repara-se que a concentração ocorre essencialmente no concelho de Lisboa.

Outro facto interessante é que a orientação da elipse evidencia que a expansão destas superfícies de retalho ocorre do litoral para o interior, provavelmente acompanhando a tendência de criação de zonas habitacionais afastadas de Lisboa. Mas esta constatação é uma mera suposição, necessitando-se de mais elementos para confirmar esta possível relação de causalidade.

A figura 7 reflecte as elipses associadas ao conjunto de pontos relativos a cada supermercado:

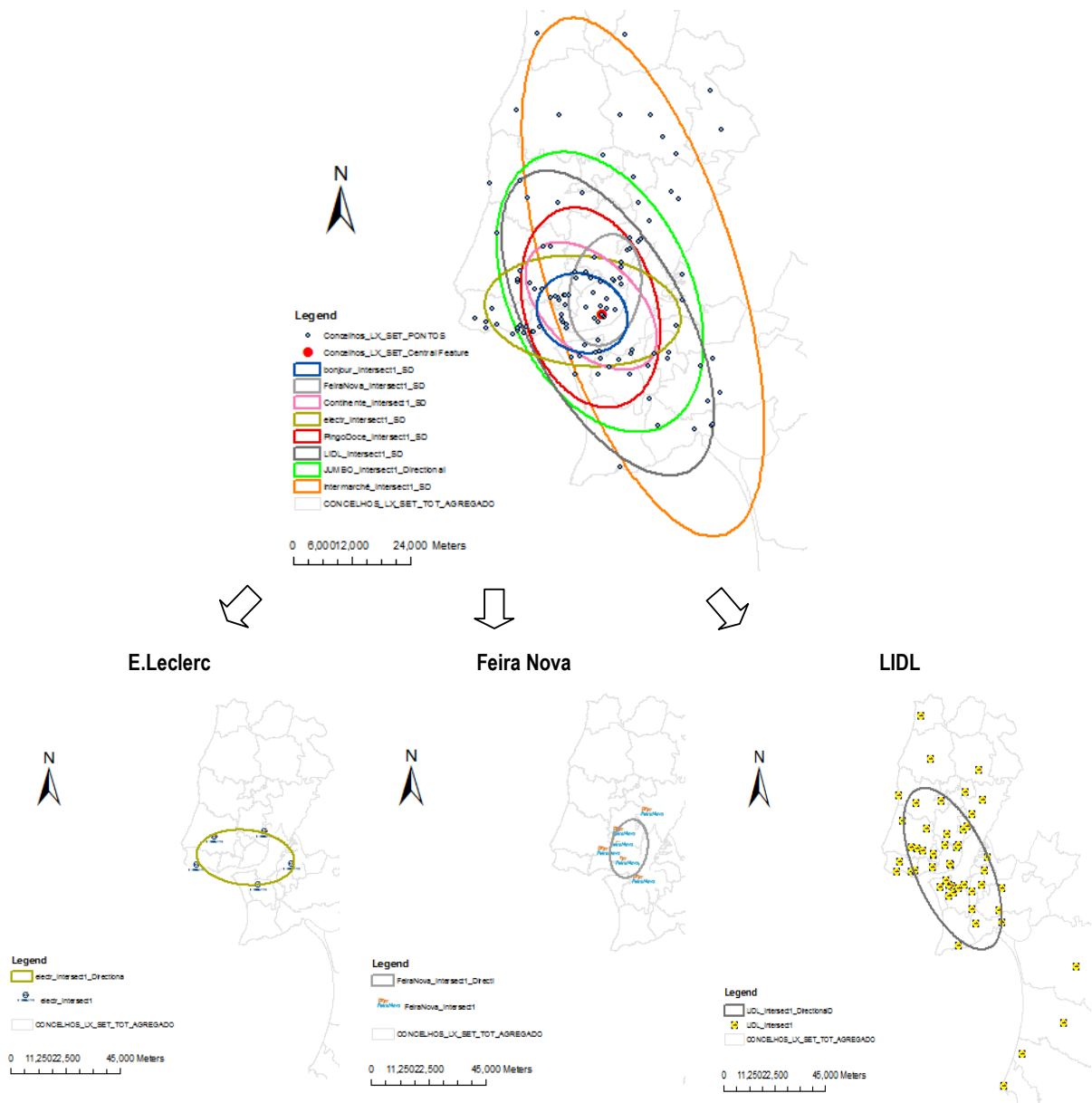


Figura 7 – Distribuição Direccional (conjunto específico de eventos)

Analisando-se a imagem que suporta a projecção das diversas elipses verifica-se que cada conjunto de evento evidencia um comportamento distinto dos restantes. É possível observar através da elipse do desvio padrão que a área da elipse formada pelas lojas Intermarché (elipse cor-de-laranja) é mais extensa em relação aos outros supermercados, abrangendo quase toda área de estudo, e que as lojas do Modelo Bonjour (elipse azul) concentram-se uniformemente na região central, de tal modo que visualmente a elipse se aproxima a um círculo.

Esta figura destaca também as principais orientações das elipses desta amostra.

3.2. Análise Exploratória e Modelação do Processo Pontual

3.2.1. Estimador Kernel

O método Kernel para criar uma superfície contínua de densidade desenha uma região de influência circular ou esférica em redor de cada evento da amostra. Para tal aplica-se uma função matemática sobre os eventos considerados, que efectua a contagem do número de eventos contidos no interior da região de influência, ponderando-os pela distância de cada um à localização de interesse.

Em cálculos de densidade, o tamanho do raio determina o tamanho do Kernel. A figura 8 apresenta as superfícies de densidades resultantes da utilização de diferentes larguras de bandas:

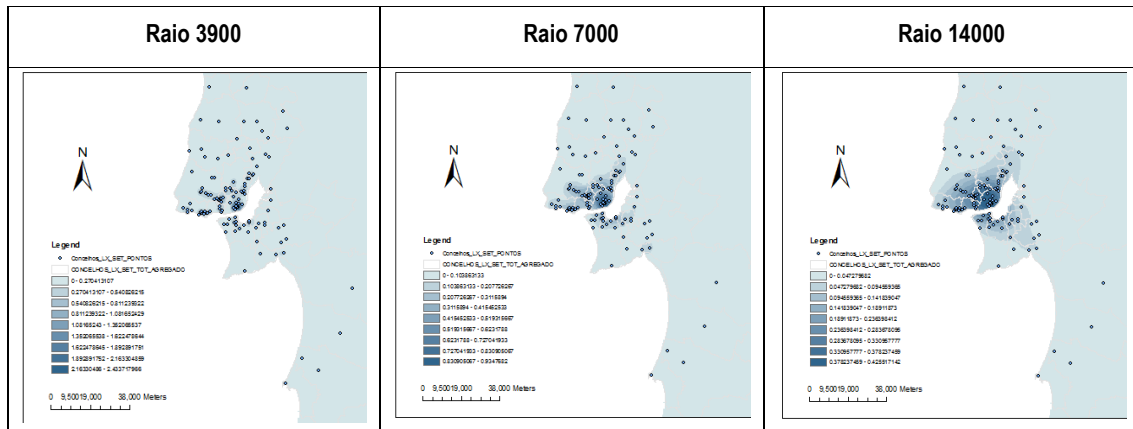


Figura 8 – Cálculo de Densidade Kernel

Analisando estas imagens é possível constatar-se que raios maiores produzem superfícies Kernel mais suaves pois as regiões de influência passam a abranger mais eventos de interesse. Esta suavização não é somente visível nas bordas entre as classes, mas como na intensidade dos próprios eventos.

Estas imagens fornecem uma visão mais detalhada da concentração pontual de lojas nestes dois distritos assim como permitem detectar a localização de *hot-spots*. Concretamente, verifica-se uma maior concentração de eventos no distrito de Lisboa, em particular, nas zonas ribeirinhas (junto do rio). As áreas com maior concentração apresentam uma tonalidade de azul mais intensa. Enquanto que as zonas periféricas evidenciam uma menor concentração e, por sua vez, densidade de eventos.

3.2.2. Método da Contagem por Quadrantes: Census

Para efectuar a análise por quadrantes recorreu-se ao ficheiro template do ArcMap PtPatternAnalysis_v9.mxt, utilizado nas aulas práticas.

Para se perceber melhor o impacto que o tamanho da grelha pode ter nos resultados finais obtidos por este método, optou-se em criar grelhas com diferentes dimensões. A figura 9 apresenta os mapas temáticos associados ao novo atributo 'K', criado automaticamente após execução do método. Este atributo armazena o número de eventos contabilizados por quadrante.

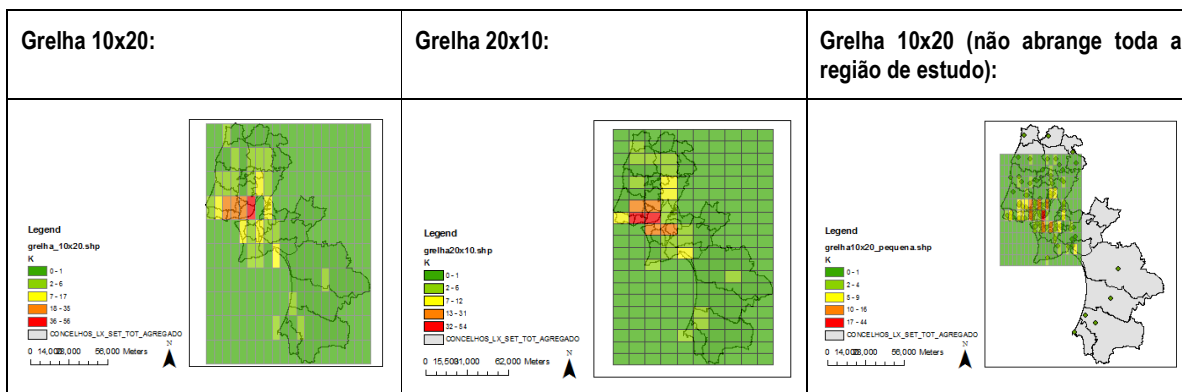


Figura 9 – Mapas Temáticos para o atributo K (grelhas diferentes)

A tabela 2 sintetiza os resultados obtidos em cada execução:

Tabela 2 – Resultados do Método da Contagem por Quadrantes

	Grelha 10x20	Grelha 20x10	Grelha 10x20 (não abrange toda a região de estudo):
Nr. Quadrantes	200	200	200
Pontos	289	289	274
Variância	33.3570	35.4870	16.1631
Variance to Mean Ratio (VMR)	23.0844	24.5585	11.7979

Analisando-se os valores de variância e VMR conclui-se que a utilização de grelhas de diferentes podem conduzir a conclusões distintas, uma vez que afectam a distribuição da frequência pelos diferentes quadrantes. Como tal, a escolha da grelha e dimensão dos respectivos quadrantes é fundamental para a correcta análise dos resultados.

Importa destacar que a execução do método com a grelha 10x20 que não abrangeu toda a região de estudo produziu resultados bastantes inferiores à da execução da grelha 10x20 que abrangeu integralmente a região, apesar do número de pontos não ter reduzido substancialmente.

Não obstante destes factos, o VMR para os três casos é superior a 1, o que sugere a existência de aglomerados – padrão de clusters. Existem células com mais pontos do que o esperado e outras com menos que o esperados.

Para verificar a significância estatística deste método exploratório, verificou-se se o padrão pontual se ajusta ao modelo teórico de referência, através da comparação com a distribuição teórica de referência e realização de teste de hipóteses.

Para tal, calculou-se o teste estatístico do qui-quadrado, dado pela seguinte expressão:

$$\chi^2_{\text{TESTE}} = (m-1) \cdot \text{VMR}$$

Onde:

- m= número de quadrantes
- VMR= *Variance to Mean Ratio*

Esta estatística de teste permitiu avaliar se os padrões são ou não significativamente diferentes do padrão aleatório, isto é, avaliar se a hipótese nula deve ou não ser rejeitada:

$$H_0: \text{VMR} \sim 1 \text{ vs } H_A: \text{VMR} \neq 1$$

O quadro seguinte (tabela 3) sintetiza os valores da estatística de teste do qui-quadrado:

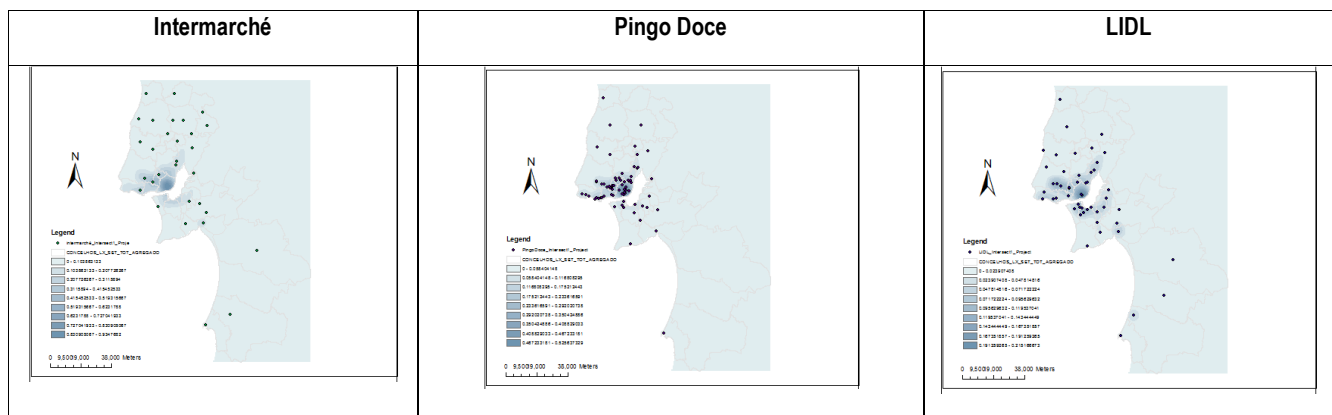
Tabela 3 – Estatística de Teste do Qui-Quadrado

	Grelha 10x20	Grelha 20x10	Grelha 10x20 (não abrange toda a região de estudo):
Nr. Quadrantes	200	200	200
Pontos	289	289	274
Variância	33.3570	35.4870	16.1631
VMR	23.0844	24.5585	11.7979
Teste do Qui-Quadrado $\chi^2_{\text{TESTE}} = (m-1) \cdot \text{VMR}$	4593.7956	4887.1415	2347.7821

Como os valores críticos da distribuição χ^2 para $p=0.99$ e $p=0.01$ ($\chi^2_{\text{CRIT(LOW)}} = 155.5377$ e $\chi^2_{\text{CRIT(HIGH)}} = 248.3382$) são inferiores às diferentes estatísticas de teste, conclui-se que a hipótese nula de aleatoriedade pode ser rejeitada e aceitar-se a hipótese alternativa. Isto é com base nestes resultados pode concluir-se que os padrões de pontos observados são significativamente diferentes do padrão aleatório.

3.2.2.1. Estimador Kernel vs Método da Contagem por Quadrantes – Conjunto de Dados Individual

De forma análoga procedeu-se à determinação das superfícies de densidade Kernel para cada um dos supermercados. A figura 10 apresenta as superfícies Kernel resultantes, considerando-se uma região de influência com um raio de 7000m.



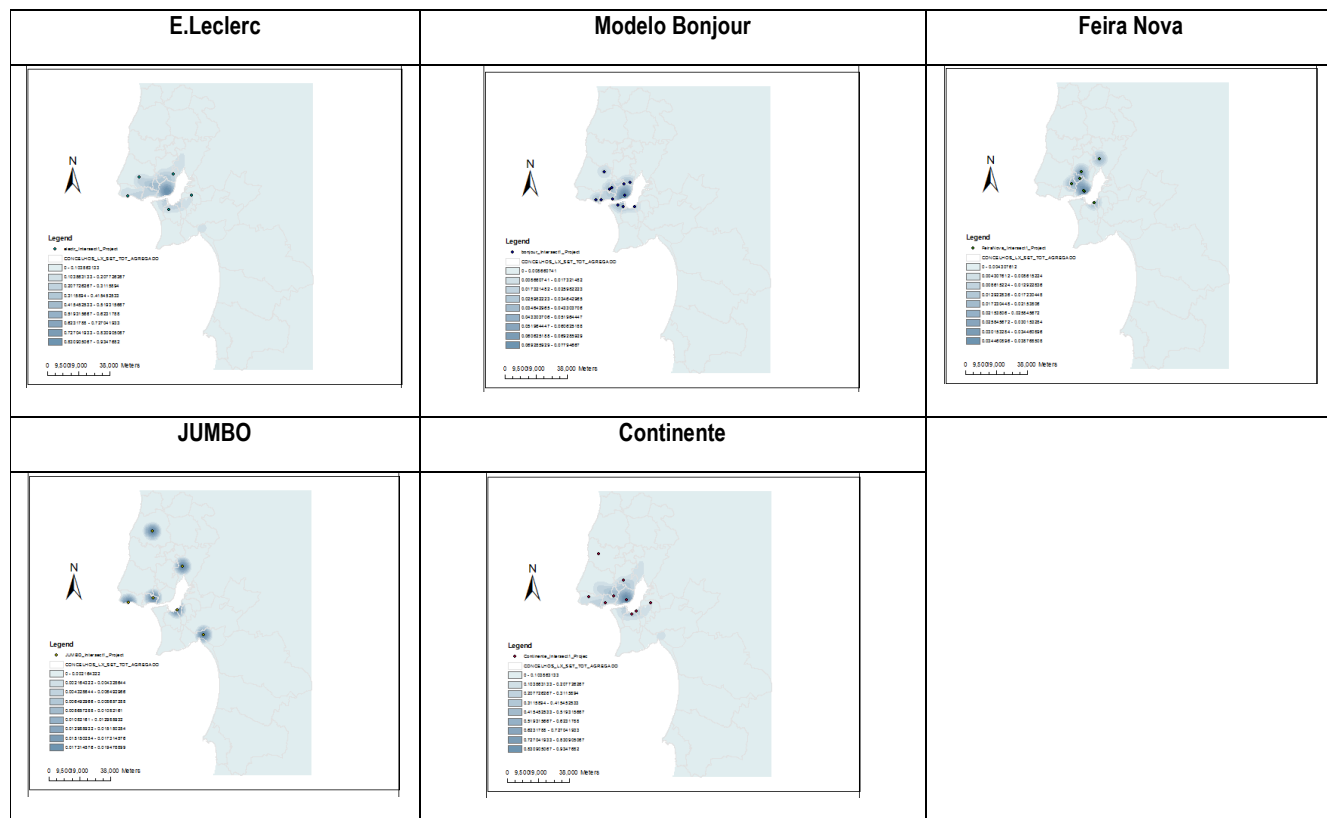


Figura 10 – Superfície de Kernel (conjunto de dados específico)

Tal como na análise global, também estas superfícies de Kernel evidenciam uma maior concentração no centro de Lisboa, contrariamente ao que sucede na periferia. No entanto, constata-se que a superfície de Kernel dos supermercados Jumbo contraria essa evidência, denotando-se alguma dispersão na distribuição das suas lojas.

A tabela 4 sintetiza os principais resultados obtidos da execução da técnica do método da contagem por quadrantes assim como apresenta as conclusões obtidas quando comparadas as estatísticas de teste com os valores críticos da distribuição χ^2 para $p=0.99$ e $p=0.01$:

Tabela 4 – Resumo dos Resultados Do Método da Contagem por Quadrantes

Método Contagem por Quadrantes:								
	Intermarché	Pingo Doce	LIDL	E.Leclerc	Modelo Bonjour	Feira Nova	Jumbo	Continente
Nr. Pontos	28	124	77	5	15	7	6	14
Quadrantes (m)	200	200	200	200	200	200	200	200
Variância	0.1404	3.9956	1.2768	0.0244	0.1294	0.0438	0.0291	0.1451
VMR	1.0029	6.4445	3.3163	0.9750	1.7250	1.2507	0.9700	2.0729
χ^2	200.57	1288.90	663.26	195.00	345.00	250.14	194.00	414.57

χ^2_{TESTE}	199.5771	1282.4555	659.9437	194.025	343.275	248.8893	193.03	412.5071
χ^2_{CRITICO}	$\chi^2_{\text{CRIT(LOW)}}= 155.5377$ $\chi^2_{\text{CRIT(HIGH)}}= 248.3382$							
Conclusão	Aceita-se H0	Rejeita-se H0	Rejeita-se H0	Aceita-se H0	Rejeita-se H0	Rejeita-se H0	Aceita-se H0	Rejeita-se H0
	Padrão Aleatório	Padrão não Aleatório	Padrão não Aleatório	Padrão Aleatório	Padrão não Aleatório	Padrão não Aleatório	Padrão Aleatório	Padrão não Aleatório

Genericamente, com a aplicação do método da contagem por quadrantes conclui-se que poderão ter um **padrão aleatório** as lojas:

- Intermarché
- E.Leclerc
- Jumbo

E um **padrão não aleatório** as lojas:

- Pingo Doce
- LIDL
- Modelo Bonjour
- Feira Nova
- Continente

3.2.3. Método do Vizinho Mais Próximo

Com vista a explorar e descrever os efeitos de 2ª ordem (locais ou de pequena escala) utilizou-se o método do vizinho mais próximo. Esta técnica de análise espacial consiste em estudar a distribuição espacial entre os eventos pontuais, estimando uma função acumulada através das distâncias h entre os eventos.

Para uma melhor análise e interpretação dos resultados optou-se por executar esta técnica 4 vezes para diferentes distâncias máximas. As figuras abaixo apresentam os gráficos da frequência acumulada da distância de cada evento ao seu vizinho mais próximo. Em todos foi definida uma distância mínima de 0 m e foram considerados 10 intervalos. Note-se que estão a ser apresentados os resultados obtidos sobre o conjunto global de supermercados.

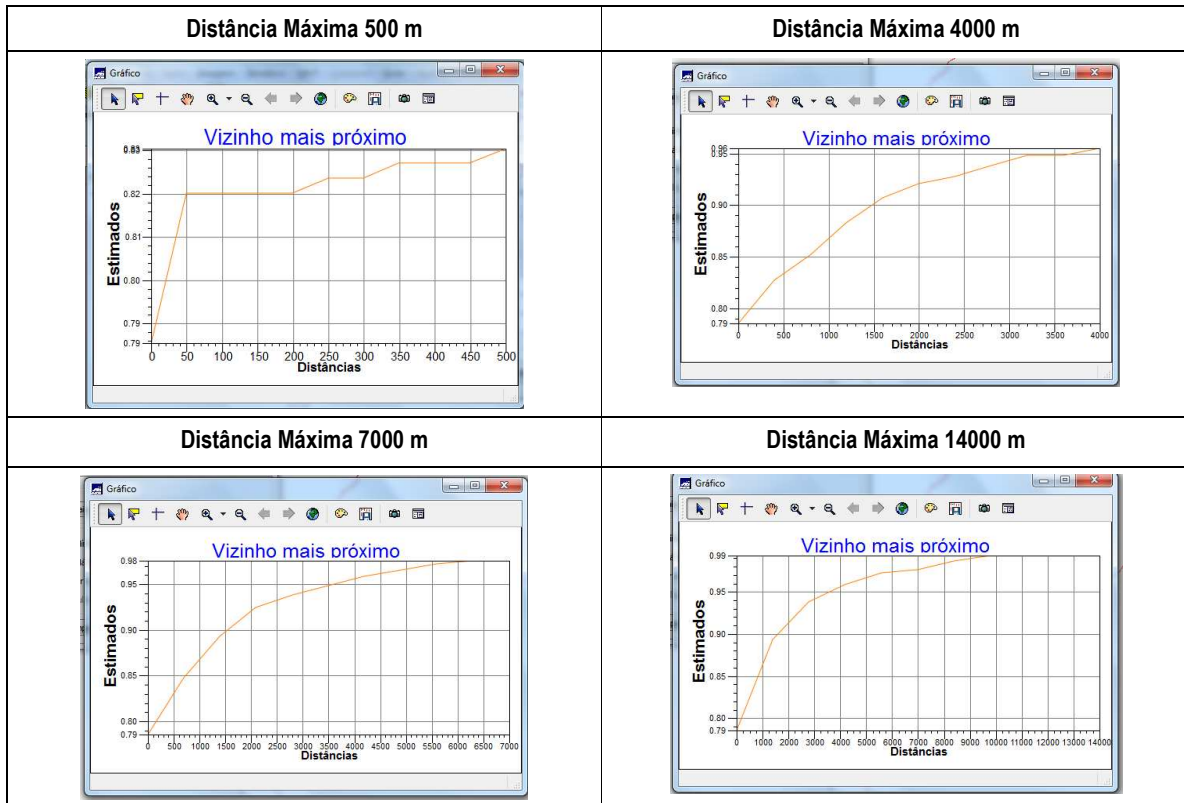


Figura 11 – Método do Vizinho Mais próximo (Gráficos de Frequências Acumuladas)

O gráfico associado à distância máxima de 500 m revela uma curva com um crescimento muito acentuado/brusco até aos 50 m o que é revelador da existência de agrupamento ou iteração espacial entre os eventos nessa faixa de distância. Os restantes gráficos apresentam curvas mais suaves, com uma elevação menos brusca em relação ao primeiro, indicando que as amostras podem estar distribuídas aleatoriamente. À medida que a distância máxima aumenta a inclinação da curva torna-se cada vez mais acentuada.

Para a comparação estatística da distribuição observada dos eventos, com o que seria esperado sob a hipótese de Aleatoriedade Espacial Completa, recorreu-se à funcionalidade disponibilizada pelo software SPRING denominada vizinho mais próximo com simulação.

Embora se tenha executado diversas vezes este método constatou-se que o gráfico resultante pouco diferiu. A figura 12 apresenta o gráfico resultante para uma distância mínima de 0 e máxima de 3000 m, com 10 intervalos e para 10 simulações.

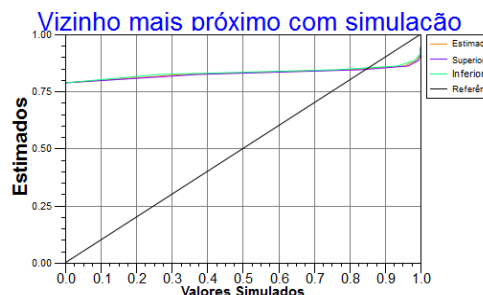


Figura 12 – Gráfico Resultante da Execução do Método do Vizinho Mais próximo Com Simulação

Este gráfico revela a existência de dois comportamentos distintos, um até ao valor simulado próximo de 0.85 outro após este valor. Até 0.85 quer a curva da distribuição quer os envelopes encontram-se acima da recta de 45°, podendo-se considerar uma evidência de não aleatoriedade e, em particular, de agrupamento. Depois dos 0.85 tudo se inverte.

Dado estes resultados, optou-se por avaliar se o padrão de pontos observados é significativamente diferente de um padrão aleatório (hipótese da completa aleatoriedade espacial), através da funcionalidade análoga disponibilizada pelo *software* ArcGis.

Da execução desta técnica sobre o conjunto com todos os supermercados, obtiveram-se os seguintes resultados:

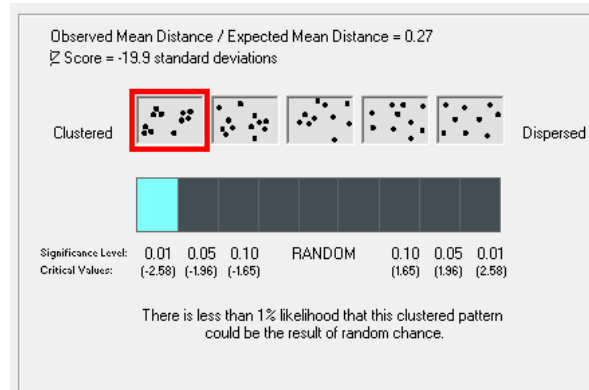


Figura 13 – Representação Gráfica dos Resultados Obtidos pelo Método do Vizinho Mais Próximo (conjunto global de eventos)

Observando-se os resultados obtidos rapidamente se constata que, segundo esta técnica, o padrão observado é significativamente diferente do padrão aleatório, revelando uma tendência global de *clustering*. Com um Z-score de -19.9 existe menos de um 1% de probabilidade deste padrão ser aleatório.

Note-se que o valor indicado na estatística de teste do vizinho mais próximo (0.27) sugeria já a existência de um padrão de pontos aglomerado.

3.2.3.1. Método do Vizinho Mais Próximo – Conjunto de Dados Individual

Aplicando-se esta técnica sobre o padrão de pontos de cada supermercados obtiveram-se os seguintes resultados:

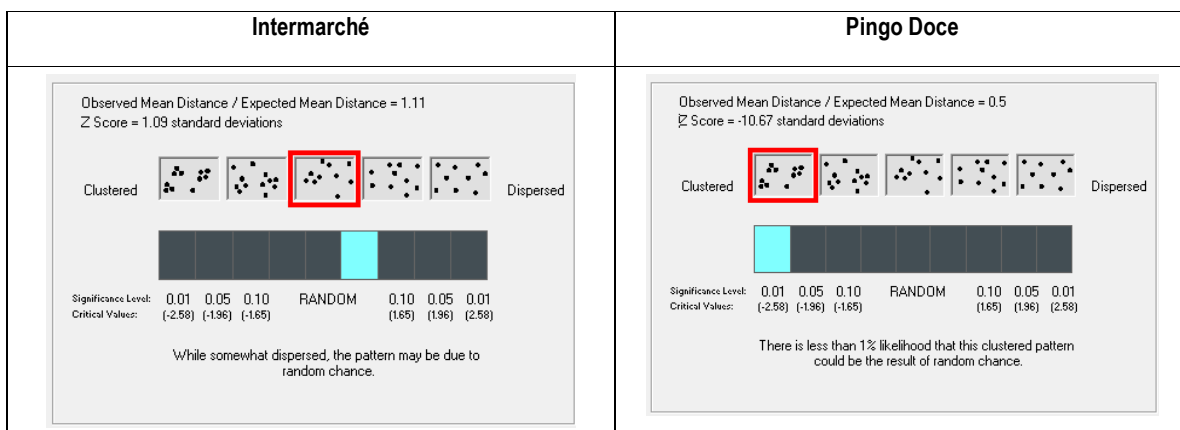




Figura 14 – Representação Gráfica dos Resultados Obtidos pelo Método do Vizinho Mais Próximo (conjunto individual)

Genericamente, com a aplicação do método do vizinho mais próximo conclui-se que poderão ter um **padrão cluster** as lojas:

- Pingo Doce
- LIDL

Um **padrão disperso** as lojas:

- E.Leclerc

- Continente
- Jumbo
- Modelo Bonjour
- Feira Nova

E um **padrão não aleatório** as lojas:

- Intermarché

3.2.4. Função K

Tal como se pode constatar anteriormente, o padrão espacial de eventos está muito associado à escala na qual é feita a análise. Deste contexto destaca-se a função K que permite analisar o padrão espacial em diferentes escalas. A função de K-Ripley é baseada nas distâncias entre todos os eventos da região de estudo.

Com vista a perceber as vantagens desta técnica face às restantes funções (G e F) optou-se por calcular a função K quer para a globalidade de supermercados existentes quer para eventos associados a determinadas cadeias de supermercados. A figura 15 exhibe os resultados obtidos para a totalidade de eventos existente em cada região de estudo:

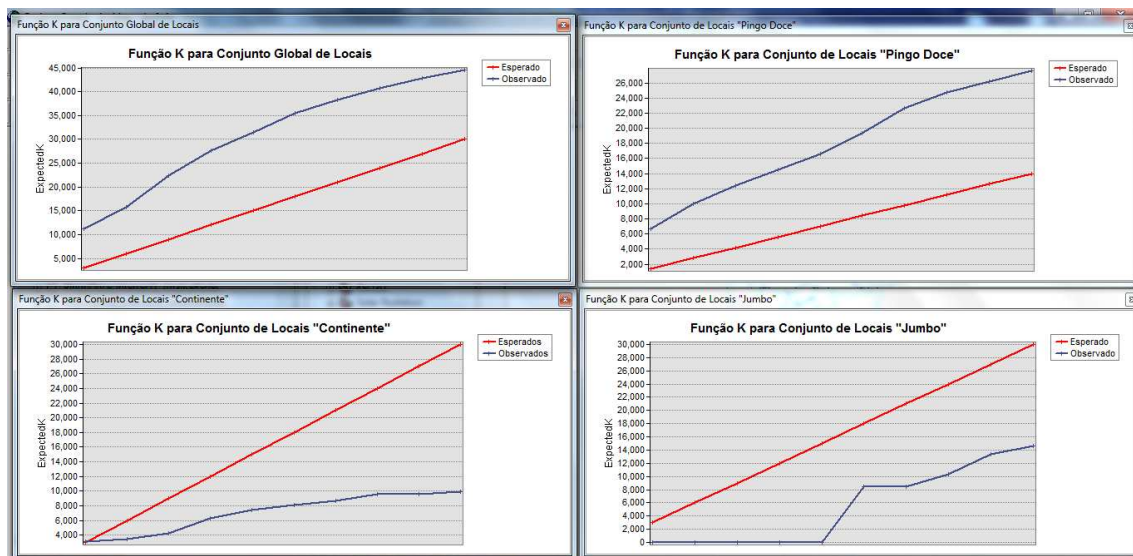


Figura 15 – Função K para a totalidade de eventos existentes na região de estudo

Note-se que nos dois primeiros gráficos os valores observados são significativamente superiores aos valores esperados, e que estes aumentam gradualmente com o aumento da distância. Contrariamente, os gráficos referentes aos supermercados Continente e Jumbo revelam uma tendência inversa, isto é, os valores esperados são superiores aos observados.

Com vista a obter-se significância dos desvios relativamente à aleatoriedade, procedeu-se ao cálculo da função K para um subconjunto dos pontos da área de estudo. Importa referir que no caso do Continente e Jumbo foram considerados todos os pontos dados o seu número reduzido de ocorrências.

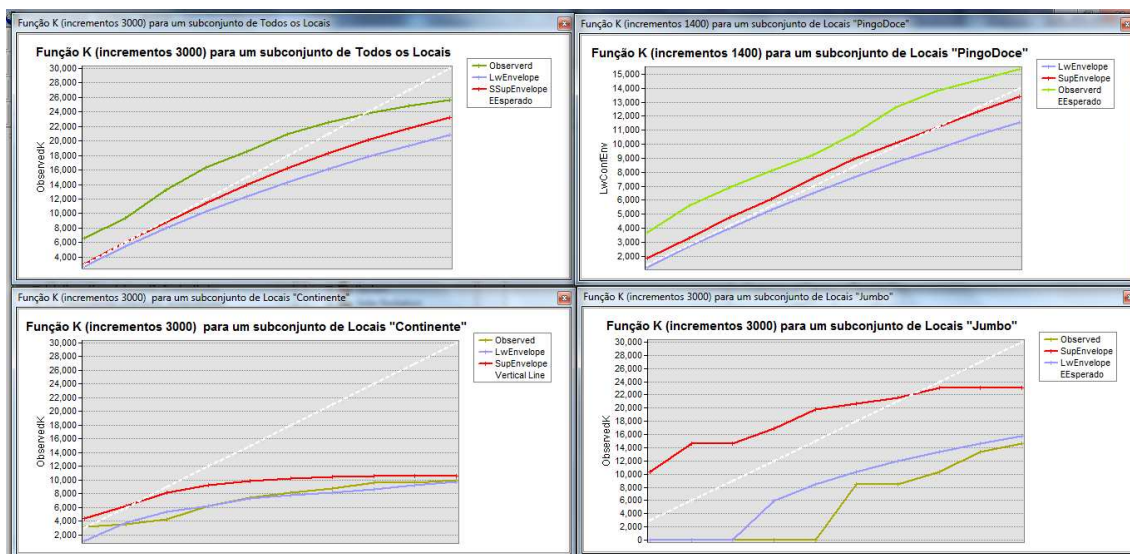


Figura 16 – Função K para a totalidade de eventos existentes na região de estudo

Com exceção do gráfico associado aos supermercados Continente, constata-se que a linha observada nos restantes gráficos encontra-se fora dos limites de confiança, criados pela função K para 99 simulações aleatórias. Nos dois primeiros gráficos a linha está acima do envelope superior, significando a existência de aglomeração, e no último a linha observada está abaixo do envelope inferior, sugerindo dispersão.

Neste três casos pode-se afirmar que, para o subconjunto de pontos considerado, com um nível de confiança de 5% pode-se rejeitar a hipótese nula de aleatoriedade espacial.

O comportamento da linha observada no gráfico dos supermercados Continente não apresenta um comportamento constante, estando por vezes fora outras vezes dentro dos limites de confiança da função K. Como tal nada se pode afirmar com certeza, apesar de se notar uma maior tendência para a linha se encontrar dentro dos limites impostos pela função K e como tal sugerir estar mais próximo de um padrão aleatório.

4. Análise do Padrões de Distribuição de Áreas

Realizada a análise do padrão de distribuição de pontos, procedeu-se à análise de áreas. Com esta pretendeu-se identificar a existência de padrões de distribuição espacial, áreas críticas e de tendências espaciais de crescimento, de forma a auxiliar o entendimento da ocorrência do fenómeno em estudo.

4.1. Visualização de Dados Espaciais

4.1.1.1. Métodos de Corte de Variável

O quadro abaixo reflecte a representação espacial dos dados resultantes dos agrupamentos por **intervalos iguais**, **quantis** e **desvio padrão**. Estes mapas temáticos reflectem a distribuição espacial do número de supermercados nos distritos de Lisboa e Setúbal.

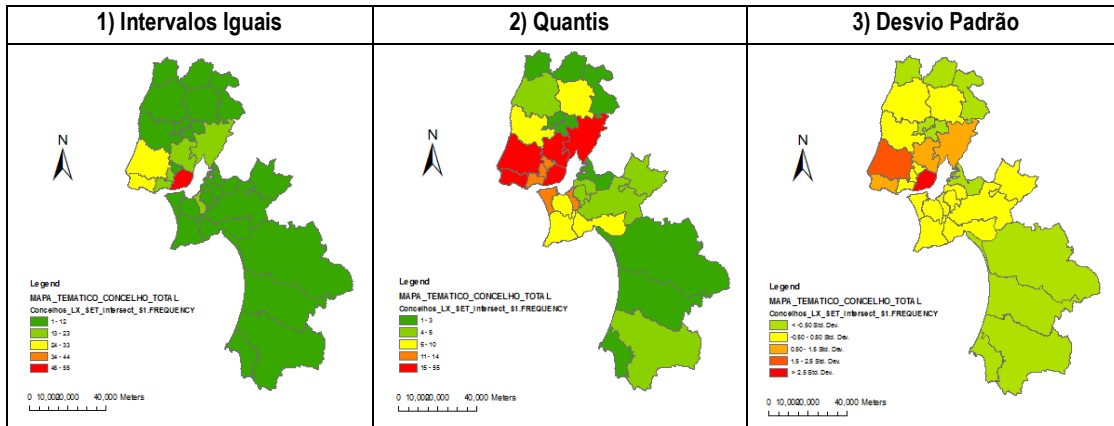


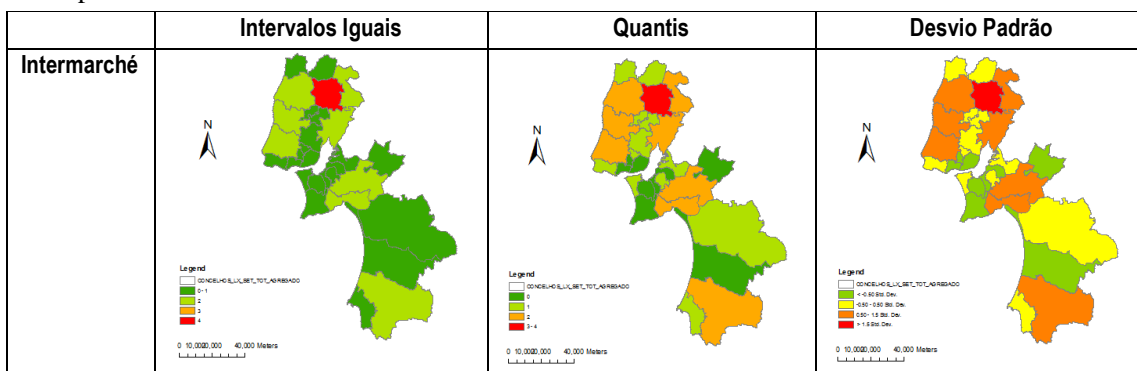
Figura 17 – Mapas resultantes dos Métodos de Corte (conjunto global de eventos)

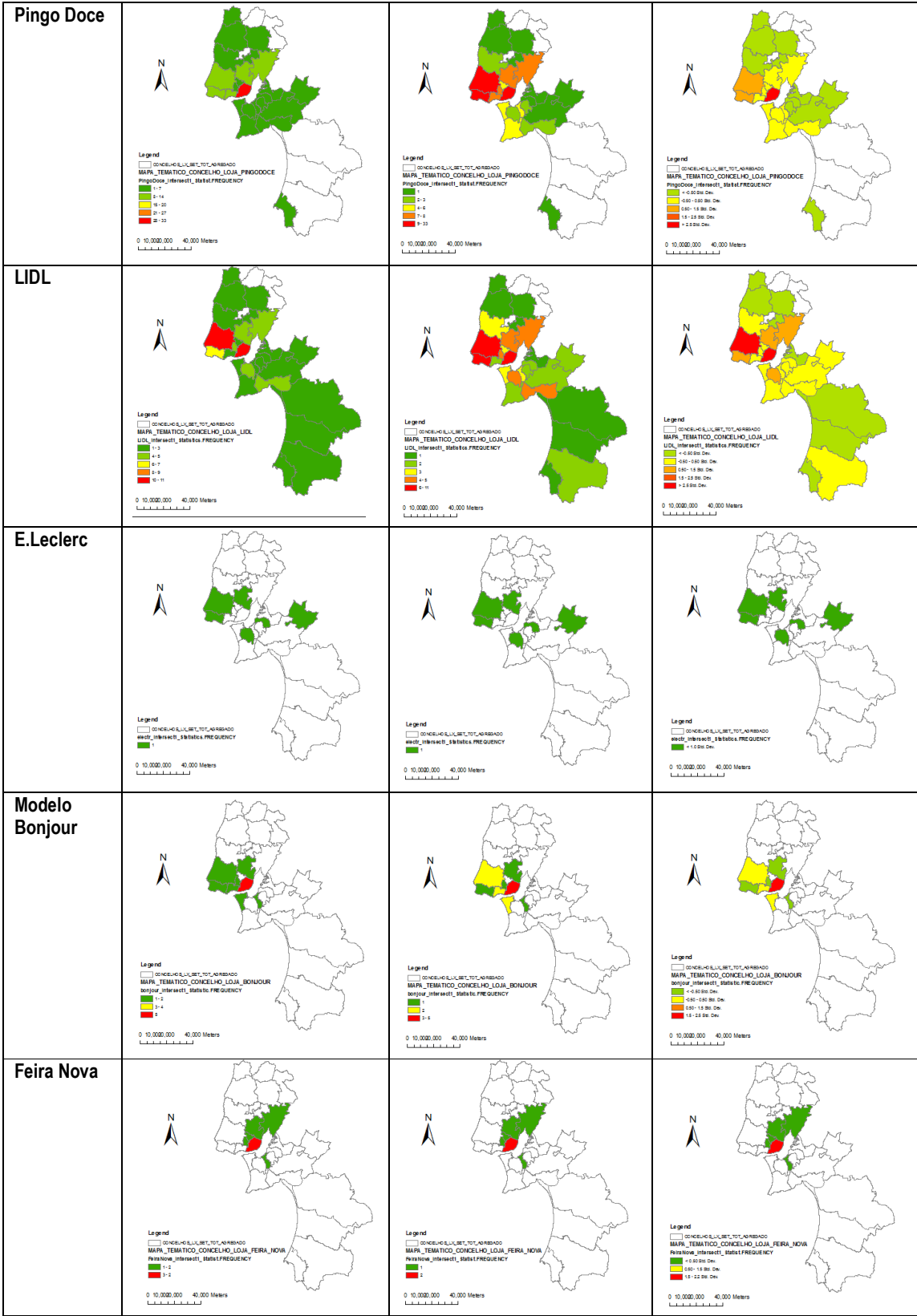
Pela figura 17 constata-se que:

- O gráfico 1 evidencia a dispersão dos supermercados nos 2 distritos. Dada à amplitude dos intervalos considerados este gráfico não revela qualquer distinção no distrito de Setúbal. Note-se que neste método cada classe tem uma amplitude igual.
- O gráfico 2 evidencia a posição relativa de um concelho face aos restantes. Os concelhos próximos do centro de Lisboa apresentam uma maior concentração de supermercados. Destaca-se também o concelho de Santiago de Cacém uma vez que, embora possua uma área aproximadamente igual aos seus concelhos vizinhos, apresenta um maior número de supermercados. Este facto releva que as cadeias de supermercados têm um maior interesse no concelho de Santiago do Cacém. Importa referir que este método é caracterizado por ter o mesmo número de elementos em cada classe e por ordenar o valor do atributo do menor para o maior.
- O gráfico 3 revela os concelhos que mais se afastam mais da média, destacando-se os concelhos de Lisboa e Sintra. Este método caracteriza o comportamento da variável, evidenciando a sua dispersão em torno da média.

Embora da análise a estes três gráficos se possam obter conclusões distintas e por vezes inconsistentes, esta técnica de análise é muito importante na medida em que organiza os dados de diferentes formas. O importante é que se conheçam os dados com que se está a trabalhar assim como o que significam.

Seguidamente são apresentados os mapas resultantes da aplicação deste método sobre cada conjunto de supermercados:





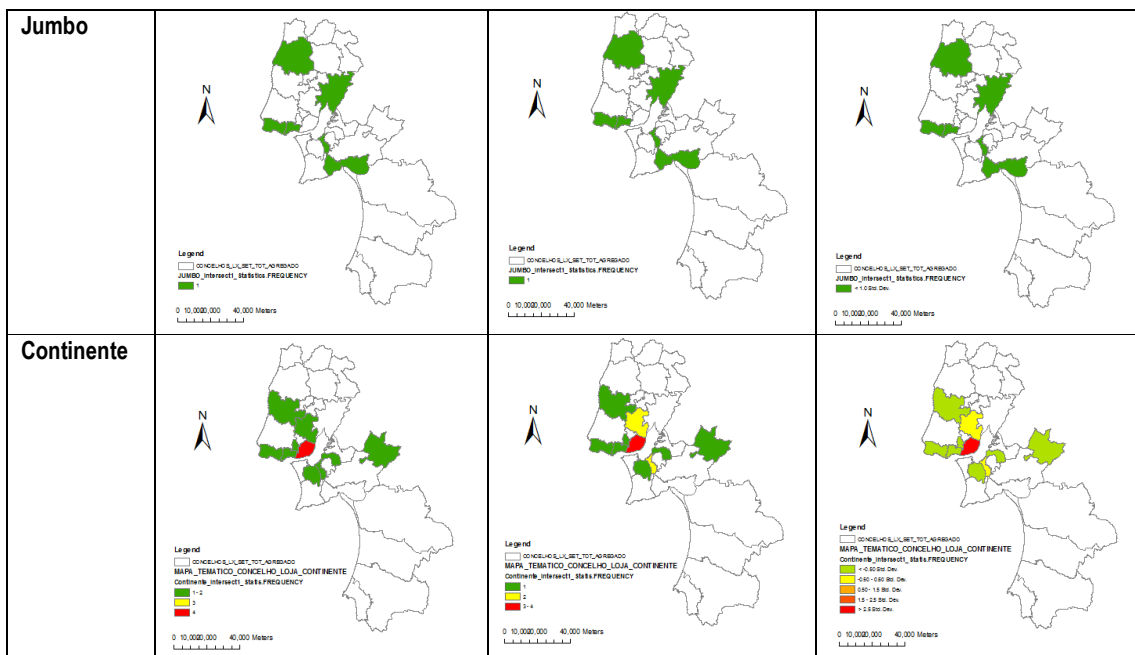


Figura 18 – Mapas resultantes dos Métodos de Corte (conjunto individual de eventos)

4.2. Índices Globais de Autocorrelação Espacial

4.2.1. Índice Global de Moran (I)

Com vista a caracterizar a dependência espacial de todas as observações da região de estudo recorreu-se ao índice global de Moran (I). Com este pretende-se perceber como é que os eventos estão correlacionados no espaço e estimar o quanto o valor observado de um atributo numa região é dependente dos valores desta mesma variável em localizações vizinhas. O índice de Moran fornece uma medida que varia de $[-1,1]$.

Para analisar o índice I utilizou-se numa primeira fase o software OpenGeoDa. A figura 19 apresenta o mapa de dispersão resultante. Note-se que estes valores são referentes à globalidade de supermercados existentes nos distritos de Lisboa e Setúbal.

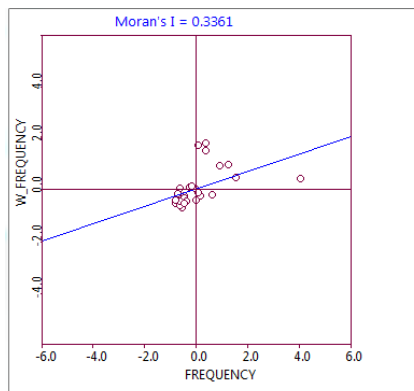


Figura 19 – Mapa de Dispersão (conjunto global de eventos)

O índice I apresenta um valor positivo, afastado do valor 0, evidenciando autocorrelação espacial positiva. O próprio mapa de dispersão comprova essa mesma correlação positiva entre a variável em causa e o valor médio do atributo na sua vizinhança.

Verifica-se ainda uma maior concentração sobre o segundo quadrante (Q2). Seleccionando-se no mapa apenas os pontos contidos no Q2 constata-se que os polígonos correspondentes situam-se mais nas zonas periféricas de Lisboa e Setúbal (figura 20).

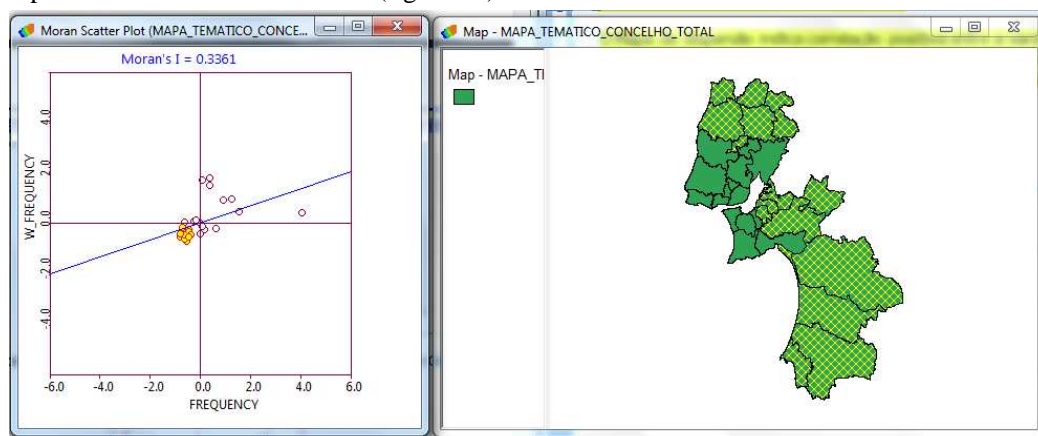


Figura 20 – Mapa de Dispersão com Pontos Seleccionados no Q2 (conjunto global)

Seleccionando-se os eventos presentes no primeiro quadrante (Q1) constata-se que os polígonos que possuem valores e médias positivas são os concelhos de Loures, Lisboa, Sintra, Cascais, Amadora, Odivelas e Oeiras (figura 21).

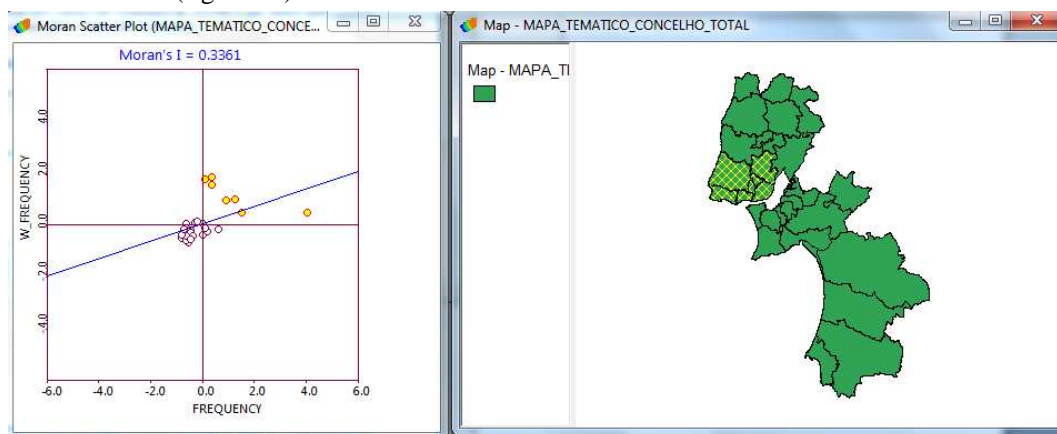


Figura 21 – Mapa de Dispersão com Pontos Seleccionados no Q1 (conjunto global)

Uma vez calculado o índice I, foi fundamental conhecer-se a sua validade estatística. A abordagem utilizada baseou-se na técnica da permutação (teste de pseudo-significância) disponibilizada também por este software.

A figura 22 compila os resultados obtidos da aplicação desta técnica para 99, 999 e 9999 permutações.

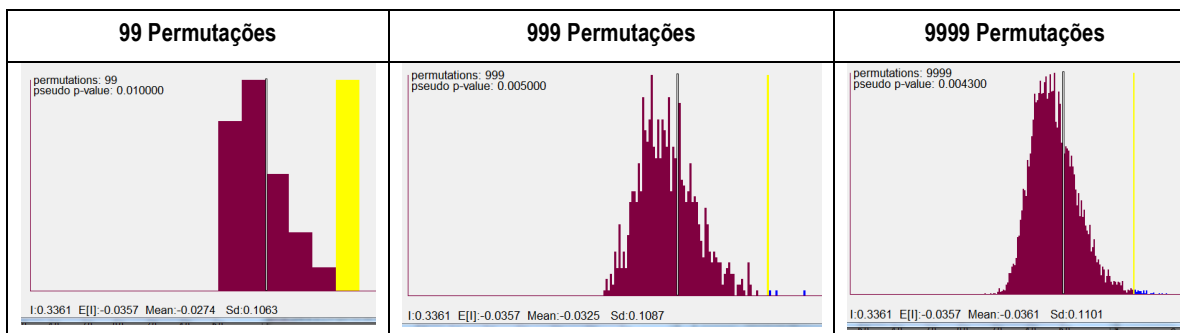


Figura 22 – Teste de Pseudo-Significância Para um Número Diferente de Permutações

Como o valor do índice I (representada pela marca amarela na figura) medido originalmente corresponde a um “extremo” da distribuição simulada, conclui-se que se trata de um valor com significância estatística e, como tal, a hipótese nula de completa aleatoriedade espacial deve ser rejeitada.

O facto da situação observada (marca amarela na figura) se situar em um dos extremos da distribuição simulada, ou seja, das 999 distribuições criadas aleatoriamente, significa que a distribuição observada tem uma situação específica e não provém de uma situação aleatória.

Executada a técnica da permutação, optou-se também estimar a significância de I através de outra abordagem, associando ao índice uma distribuição estatística aproximada (hipótese normal). Para tal, recorreu-se à funcionalidade disponibilizada pelo ArcGis.

Através desta obtiveram-se, para a globalidade dos supermercados, os seguintes resultados (figura 23):

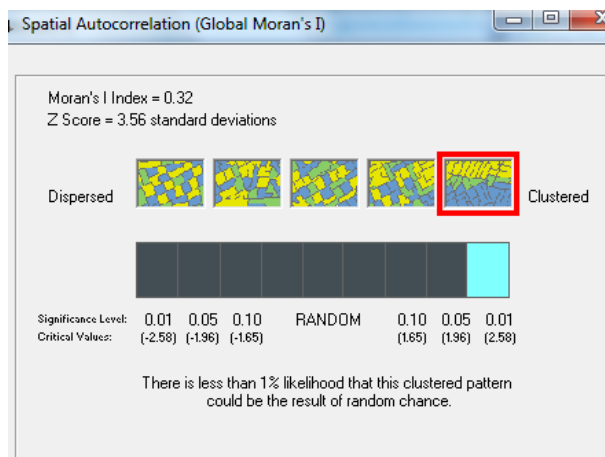


Figura 23 – Índice Global de Moran (conjunto global de eventos)

Esta abordagem permite concluir, com nível de significância de 1%, que a distribuição de supermercados assume um padrão de aglomerados, significativamente diferente do padrão aleatório.

4.2.1.1. Índice Global de Moran (I) – Conjunto Individual de Dados

O quadro seguinte resume os resultados do Índice Global de Moran:

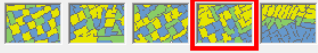

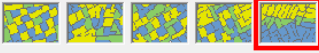









<p>Intermarché</p> <p>Moran's I Index = 0.22 Z Score = 2.13 standard deviations</p> <p>Dispersed  Clustered</p>  <p>Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01 Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)</p> <p>There is less than 5% likelihood that this clustered pattern is the result of random chance.</p>	<p>Pingo Doce</p> <p>Moran's I Index = 0.26 Z Score = 2.93 standard deviations</p> <p>Dispersed  Clustered</p>  <p>Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01 Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)</p> <p>There is less than 1% likelihood that this clustered pattern could be the result of random chance.</p>
<p>LIDL</p> <p>Moran's I Index = 0.18 Z Score = 1.75 standard deviations</p> <p>Dispersed  Clustered</p>  <p>Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01 Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)</p> <p>There is a 5-10% likelihood that this clustered pattern is the result of random chance.</p>	<p>E.Leclerc</p> <p>WARNING 000845: A minimum of 30 input features for analysis is ideal. ERROR 000906: Zero variance: all of the values for your input field are likely the same.</p>
<p>Modelo Bonjour</p> <p>Moran's I Index = -0.39 Z Score = -1.21 standard deviations</p> <p>Dispersed  Clustered</p>  <p>Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01 Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)</p> <p>While somewhat dispersed, the pattern may be due to random chance.</p>	<p>Feira Nova</p> <p>Moran's I Index = -0.4 Z Score = -0.87 standard deviations</p> <p>Dispersed  Clustered</p>  <p>Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01 Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)</p> <p>The pattern is neither clustered nor dispersed.</p>
<p>Jumbo</p> <p>WARNING 000845: A minimum of 30 input features for analysis is ideal. ERROR 000906: Zero variance: all of the values for your input field are likely the same.</p>	<p>Continente</p> <p>Moran's I Index = -0.24 Z Score = -0.52 standard deviations</p> <p>Dispersed  Clustered</p>  <p>Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01 Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)</p> <p>The pattern is neither clustered nor dispersed.</p>

Figura 24 – Índice Global de Moran (conjunto individual)

Genericamente, através dos resultados do Índice Global de Moran pode-se concluir que poderão ter um **padrão cluster** as lojas:

- Pingo Doce
- LIDL
- Intermarché

E um **padrão não aleatório** as lojas:

- Continente
- Modelo Bonjour
- Feira Nova

4.3. Índices Locais de Autocorrelação Espacial

4.3.1. Índice Local de Moran (I)

Com vista a identificar os padrões de associação espacial local, utilizou-se o Índice Local de Moran (I). Este índice permite calcular o produto dos resíduos em cada freguesia (polígono) com os valores dos polígonos adjacentes.

O Índice Local de Moran associado ao conjunto global de supermercados encontra-se patente na figura abaixo (figura 25):

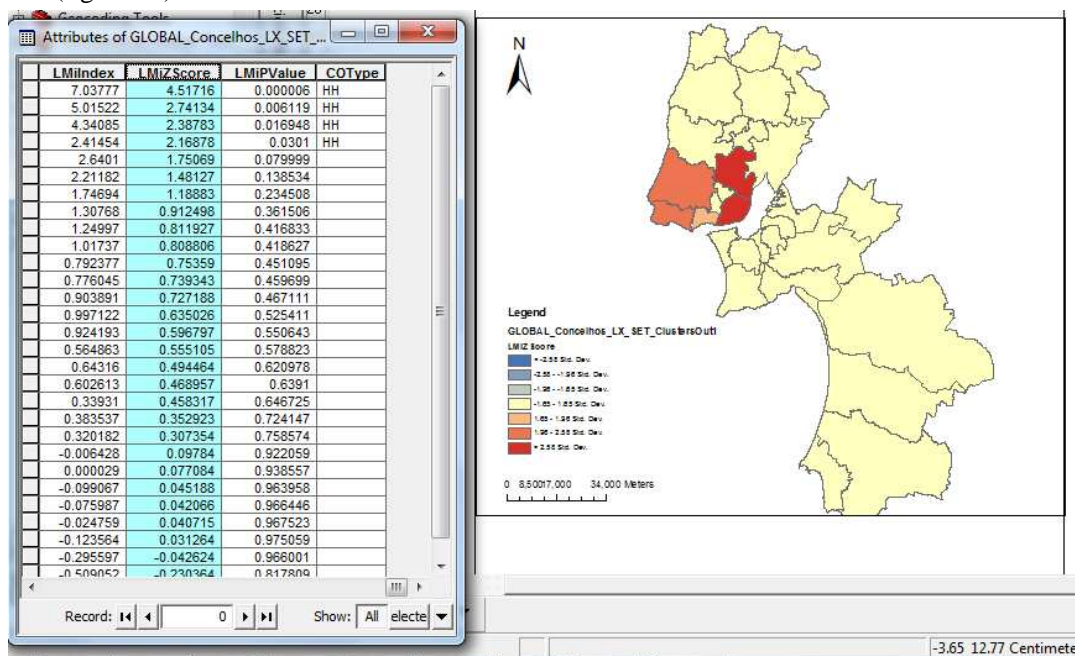


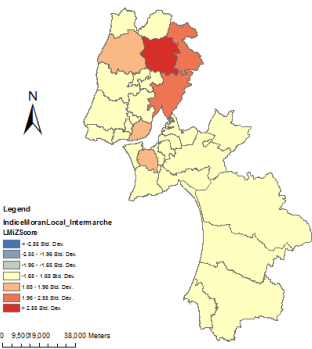
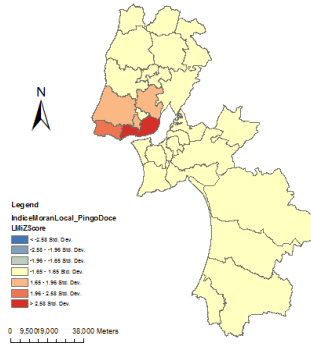
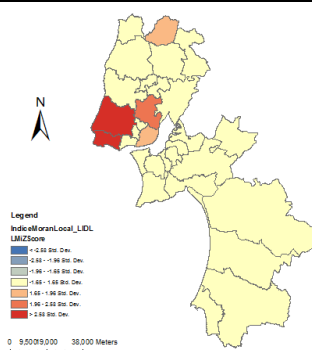
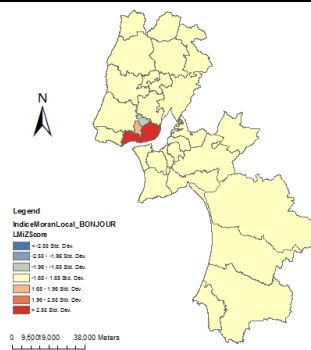
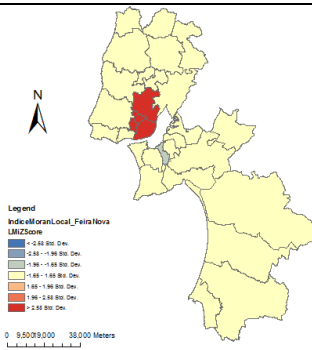
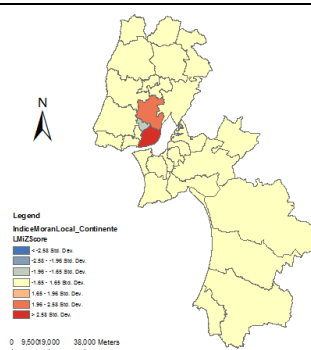
Figura 25 – Mapa do Índice Local de Moran (conjunto global de eventos)

Este mapa reflecte o valor Z-score e, implicitamente, o nível de significância estatística com que cada região é classificada no sentido de contribuir para o valor global de autocorrelação. Assim, conclui-se que o valor do Índice Local de Moran para o concelho de Lisboa (4,51) é estatisticamente significativo ao nível de significância de 1%, pelo que rejeita a hipótese nula de independência espacial.

Este mapa releva ainda, com um nível de confiança de 95%, a presença de duas zonas de aglomeração em Sintra e Cascais. As zonas mais periféricas não apresentam significativa autocorrelação espacial, sugerindo não haver um padrão específico.

4.3.1.1. Índice Local de Moran (I) – Conjunto Individual de Dados

Tabela 5 – Resumo dos Resultados do Índice Local de Moran para Conjunto Individual de Dados

<p style="text-align: center;">Intermarché</p>  <p style="text-align: center;">Alenquer é uma zona de aglomerado, com um nível de 99% de confiança</p>	<p style="text-align: center;">Pingo Doce</p>  <p style="text-align: center;">Oeiras e Lisboa são zonas de aglomerado, com um nível de 99% de confiança</p>
<p style="text-align: center;">LIDL</p>  <p style="text-align: center;">Sintra e Cascais são zonas de aglomerado, com um nível de 99% de confiança</p>	<p style="text-align: center;">Modelo Bonjour</p>  <p style="text-align: center;">Oeiras e Lisboa são zonas de aglomerado, com um nível de 99% de confiança</p>
<p style="text-align: center;">Feira Nova</p>  <p style="text-align: center;">Loures, Odivelas, Amadora e Lisboa são zonas de aglomerado, com um nível de 99% de confiança</p>	<p style="text-align: center;">Continente</p>  <p style="text-align: center;">Lisboa é uma zona de aglomerado, com um nível de 99% de confiança</p>

4.3.2. Estatística G_i e G^* (*hot-spots*)

No âmbito deste trabalho, além do Índice Local de Moran, também foram calculadas as estatísticas G_i e G_i^* . Estas estatísticas são calculadas a partir da soma simples das observações dos atributos vizinhos em relação a uma determinada posição. Importa relembrar que esta abordagem é diferente da lógica de cálculo do índice local de Moran. Enquanto que a estatística G se baseia na soma de valores vizinhos o índice local de Moran baseia-se na covariância. Como tal é expectável que se vá obter resultados diferentes.

O resultado do cálculo da estatística G_i^* sobre o total conjunto de supermercados encontra-se reflectido na figura 26:

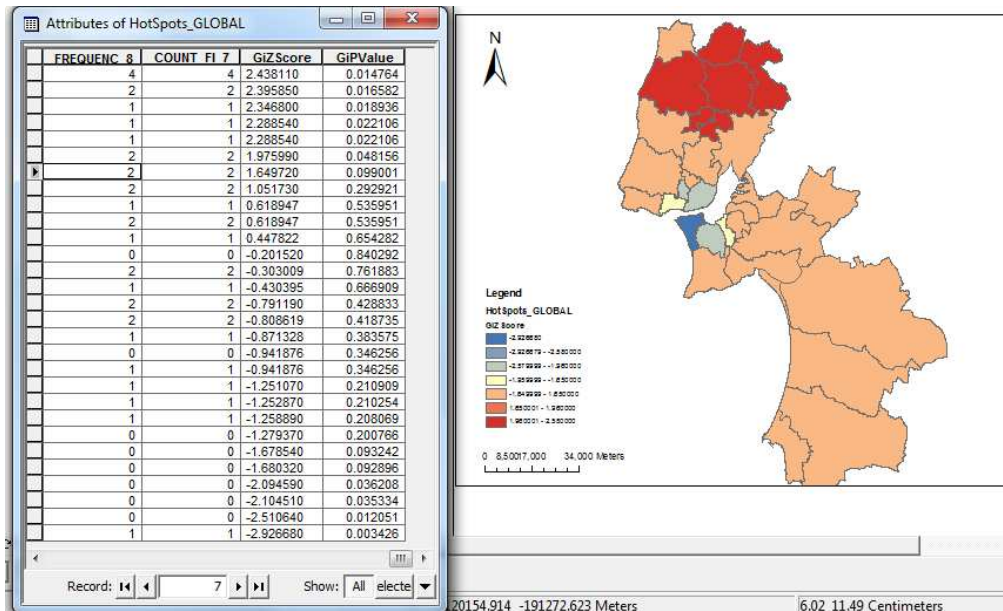


Figura 26 – Estatística G^* (conjunto global de eventos)

Analisando-se o mapa resultante rapidamente se constata a existência de uma mancha a vermelho que sugere a existência de aglomerado significativo ao nível de 5%. Por outro lado, o polígono a azul-escuro indica igualmente a existência de aglomerado ao mesmo nível de significância.

4.3.2.1. Estatística G_i e G^* (*hot-spots*) – Conjunto de Dados Individual

Procedeu-se também ao cálculo da estatística G_i e G_i^* para os conjuntos de pontos relativos a cada supermercado. Tal como referido anteriormente as manchas a vermelho e azul mais forte indicam, com um nível de 99% de confiança, zonas de aglomerado.

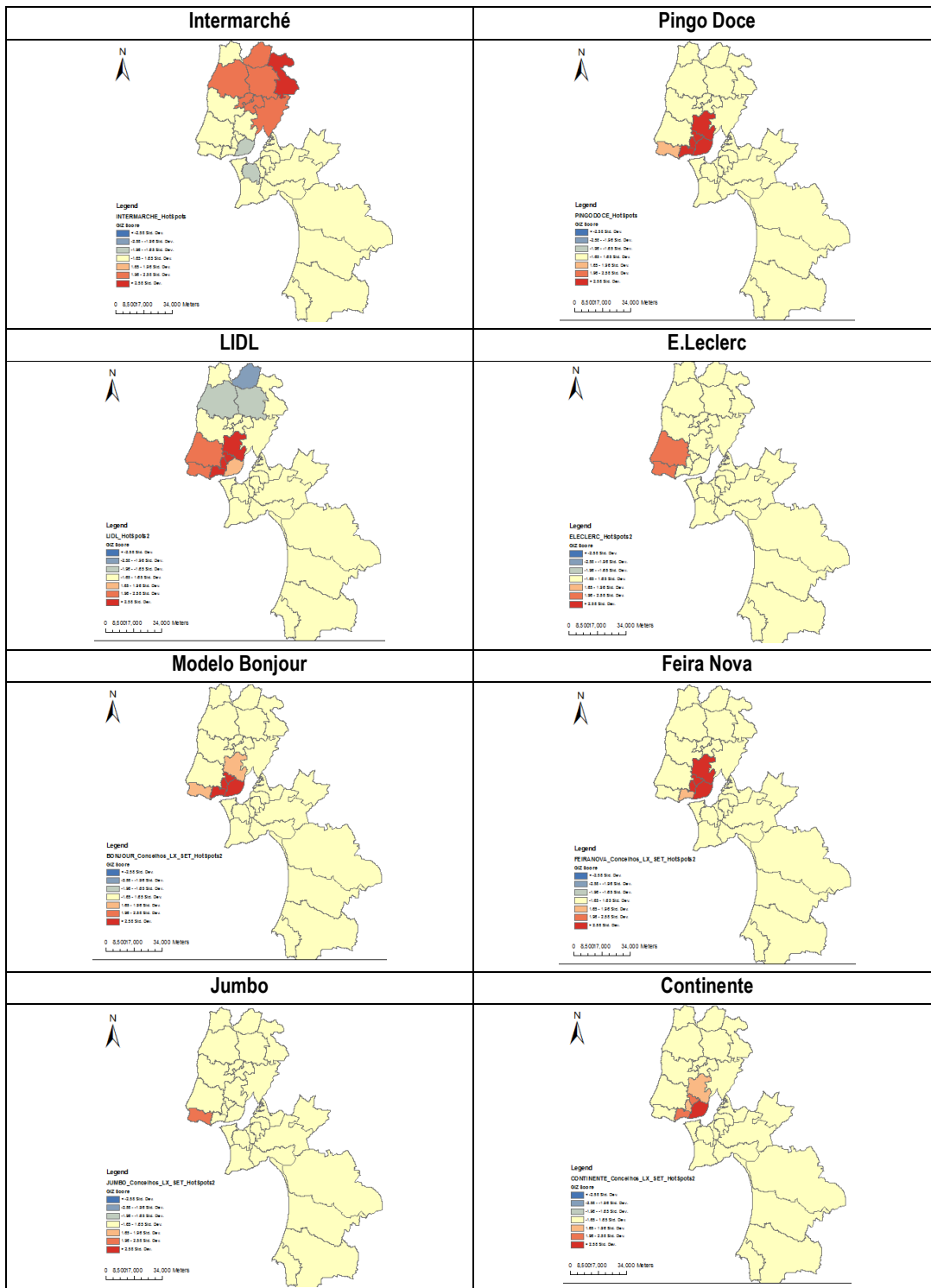


Figura 27 – Estatística G* (conjunto individual de eventos)

5. Regressão Linear

A análise de regressão estuda o relacionamento existente entre duas ou mais variáveis, em particular, entre uma variável chamada a variável dependente e outras variáveis chamadas variáveis independentes. Este relacionamento é representado por um modelo matemático que associa a variável dependente com as variáveis independentes. Este modelo pode ser denominado por **regressão linear simples** ou **regressão linear múltipla** dependendo do número de variáveis independentes incorporadas.

Neste trabalho pretende-se fundamentalmente descobrir quais as variáveis explicativas que contribuem de forma significativa para a relação linear. É óbvio que a análise realizada é muito simplista. No entanto, optou-se por realizar esta “mini”análise com o objectivo de aprender e aplicar as técnicas dadas nas aulas.

Neste sentido a análise de regressão incidiu sobre as seguintes variáveis:

- Densidade Populacional;
- Poder de Compra.

Importa referir que toda esta análise foi efectuada no software IDRISI.

Nos próximos subcapítulos serão apresentados os principais aspectos caracterizadores e resultados obtidos no âmbito desta temática.

5.1. Regressão Linear Aplicada Sobre o Conjunto Global de Eventos

5.1.1. Regressão Linear Simples

Embora o objectivo principal seja explicar o número de supermercados existentes, simultaneamente, pela densidade populacional e poder de compra, numa primeira fase analisou-se individualmente cada uma destas variáveis, i.e.:

- Explicar o número de supermercados existentes pela densidade populacional;
- Explicar o número de supermercados existentes pelo poder de compra.

Para proceder-se à análise de correlação entre cada uma destas variáveis, recorreu-se ao módulo REGRESS do IDRISI. Para tal, criaram-se 3 ficheiros AVL com os valores de cada uma destas variáveis. Da execução deste módulo, obtiveram-se os seguintes diagramas de dispersão e equações de recta de regressão linear (figura 28).

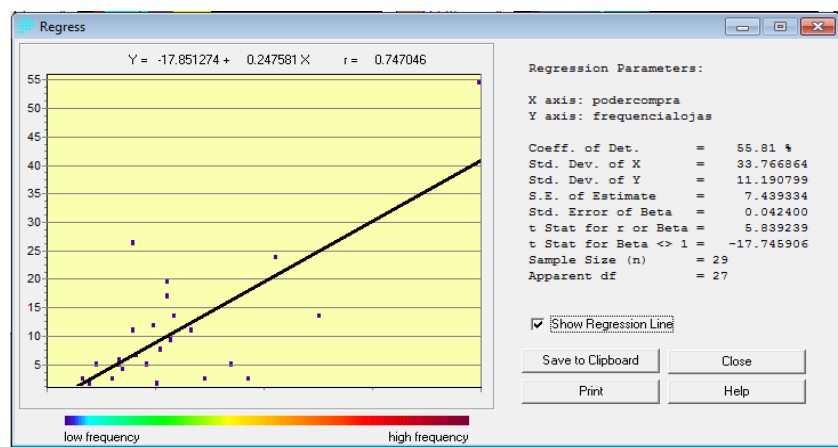


Figura 28 – Diagramas de Dispersão (poder de compra)

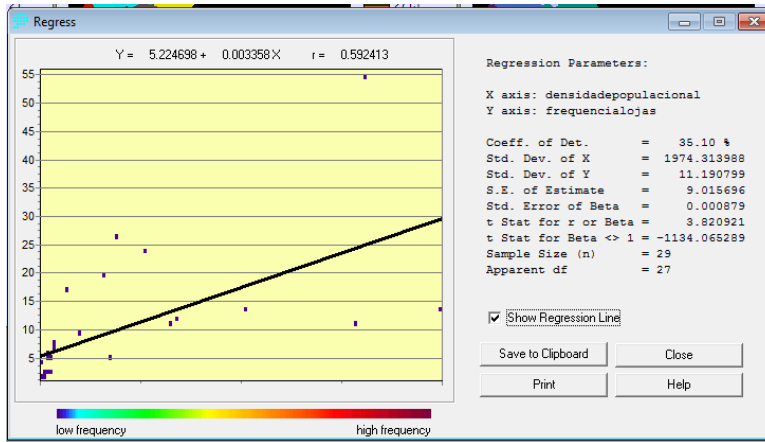


Figura 29 – Diagramas de Dispersão (densidade populacional)

Comparando cada um destes diagramas constata-se que o primeiro evidencia uma relação fortemente positiva entre o número de lojas e o poder de compra. A mesma conclusão é obtida através do coeficiente de correlação. O coeficiente do poder de compra é 0.74 enquanto que o coeficiente da densidade populacional é 0.59.

Obtidas as equações de regressão linear, estimou-se a variável dependente à custa de cada expressão matemática assim como se produziram as respectivas imagens dos resíduos (figura 30 e 31):

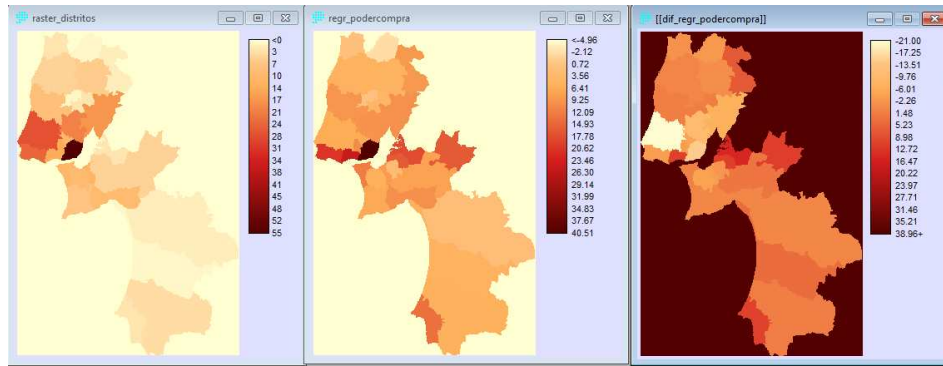


Figura 30 – Imagem Inicial vs Estimada vs Resíduos (poder de compra)

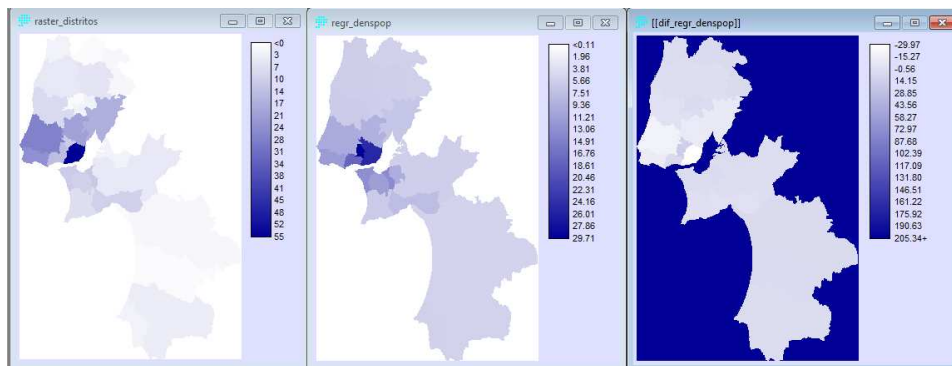


Figura 31 – Imagem Inicial vs Estimada vs Resíduos (densidade populacional)

Note-se que a terceira imagem de cada figura anterior corresponde à imagem dos resíduos. Nelas poder-se-á constatar zonas que evidenciam que ambos os modelos (equações de regressão) não se ajustam correctamente aos valores observados. As zonas mais escuras significam que a diferença entre os valores estimados e os observados é grande e, analogamente, as zonas mais claras indicam que esta diferença é negativa, ou seja, os valores observados são superiores aos estimados.

Para verificar se os resíduos do modelo são normalmente distribuídos procedeu-se à análise dos respectivos histogramas:

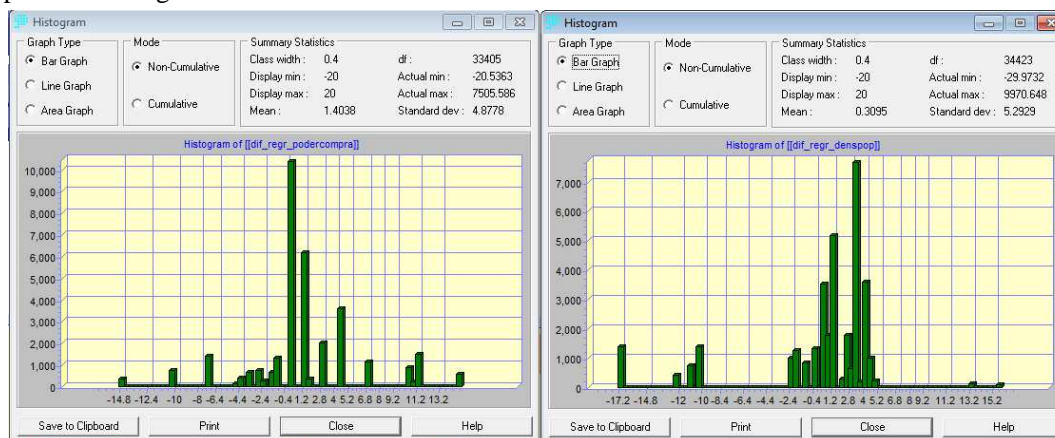


Figura 32 – Histogramas das imagens dos resíduos (poder de compra vs densidade populacional)

Analisando-se cada um dos histogramas constata-se que nenhum é normalmente distribuído. O histograma do poder de compra suscitou algumas dúvidas, mas após uma análise mais atenta verificou-se, por exemplo, a existência de classes com valores inferiores às classes seguintes (por exemplo a classe 2.8). Se seguissem uma distribuição normal as classes seguintes deveriam apresentar valores inferiores e assim sucessivamente. Neste contexto, conclui-se que as observações são não independentes e correlacionam-se.

5.1.2. Regressão Linear Múltipla

Garantida a existência de regressão linear entre a variável dependente e cada uma das variáveis independentes, iniciou-se a análise de regressão linear múltipla. Para tal executou-se o módulo MULTIREG, tendo em conta que se pretende explicar o número de supermercados pela densidade populacional e poder de compra.

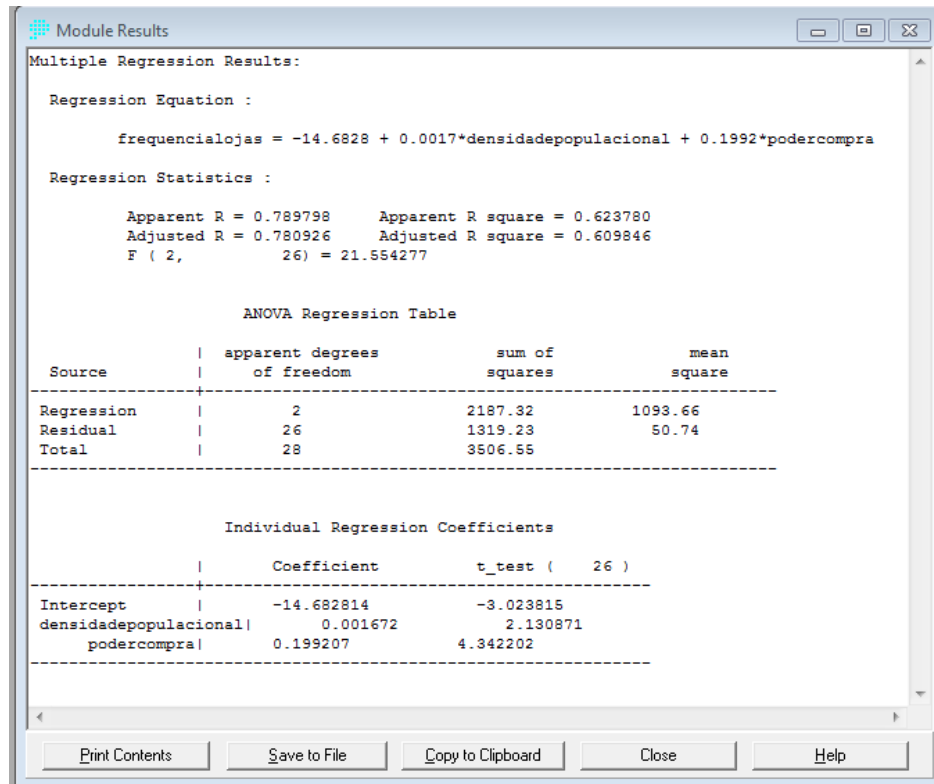


Figura 33 – Resultados da Execução do Módulo de Regressão Múltipla

Com base nos resultados obtidos (figura 33), constata-se que:

- O **coeficiente de correlação múltipla** entre as variáveis independentes e a variável dependente assume um valor próximo de 1 (0.79) que significa que o ajuste ao modelo é bom.
- Cerca de 62% da **variância** é explicada pelas variáveis independentes consideradas.
- O valor de **F(2,26)** é 21.55.

Como este valor é superior ao valor da distribuição de Fisher (2,26) com 99% de confiança (5.526335), conclui-se que a regressão global é significativa. As variáveis independentes em conjunto contribuem significativamente para a predição da variável dependente.

Para avaliar a significância de cada coeficiente foi necessário analisar a tabela da distribuição t-student de forma a obter o valor correspondente a 26 graus de liberdade a um nível de confiança de 99% : t-student $t_{CRITICO}(26) = 3.435$.

Tendo em conta este valor, concluiu-se que como o coeficiente da variável:

- “Poder de compra” possui um valor de t-student t_{TESTE} (4.34) superior ao t-student $t_{CRITICO}$, é altamente significativo (99%).
- “Densidade Populacional” possui um valor de t-student t_{TESTE} (2.13) inferior ao t-student $t_{CRITICO}$, tem uma significância menor.

Neste sentido, a variável “densidade populacional” poderá ser excluída do modelo de predição.

A figura abaixo (figura 34) reflecte o mapa dos resíduos resultantes da regressão linear múltipla:

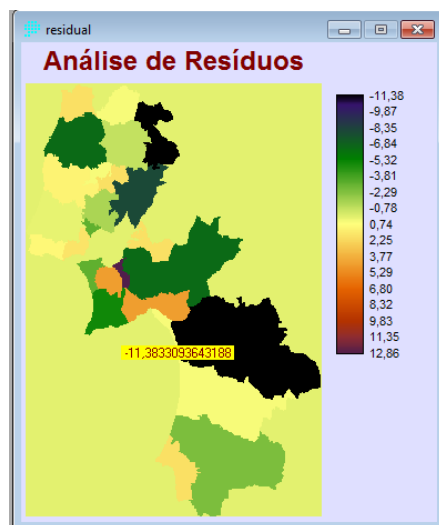


Figura 34 – Mapa dos Resíduos (regressão linear múltipla)

Importa reforçar que as áreas com tonalidade mais escuras sugerem que o modelo não se ajusta adequadamente aos valores observados.

5.2. Regressão Linear Aplicada Sobre o Conjunto de Eventos Individuais

5.2.1. Regressão Linear Simples

A tabela 6 resume os resultados obtidos da análise de regressão linear para cada conjunto individual de supermercados:

Tabela 6 – Resumo dos Resultados da Regressão Linear Simples

	Variáveis Dependentes	Equação da Recta	R	Coefficiente de Determinação
Pingo Doce	Densidade Populacional	$1.531492+0.002106X$	0.633533	40.14%
	Poder de Compra	$-12.433752+0.150711X$	0.77361	60.12%
LIDL	Densidade Populacional	$1.941906+0.000697X$	0.505852	25.59%
	Poder de Compra	$-2.209948+0.045652X$	0.566342	32.07%
Feira Nova	Densidade Populacional	$0.003695+0.000180X$	0.6996059	48.45%
	Poder de Compra	$-0.587780+0.00763X$	0.493199	24.32%
Intermarché	Densidade Populacional	$1.368424-0.000201X$	-0.422243	17.83%
	Poder de Compra	$2.079626-0.008786X$	-0.315954	9.98%
Jumbo	Densidade Populacional	$0.187979+0.000014X$	0.068664	0.47%
	Poder de Compra	$-0.100372+0.002766X$	0.226532	5.13%
E.Leclerc	Densidade Populacional	$0.003695+0.000180X$	0.696059	48.45%
	Poder de Compra	$-0.587780+0.007463X$	0.493199	24.32%

Continente	Densidade Populacional	0.150377+0.000252X	0.545908	29.80%
	Poder de Compra	-1.649533+0.019192X	0.711326	50.60%
Bonjour	Densidade Populacional	0.063890+0.000344X	0.622600	38.76%
	Poder de Compra	-2.221018+0.024648X	0.763815	58.34%

Com exceção do Intermarché e Jumbo, os restantes supermercados evidenciam uma correlação entre o número de lojas e as variáveis independentes, não obstante, de existirem correlações mais fortes do que outras. As correlações fortemente positivas, apresentam um valor de *r* muito próximo de 1.

5.2.2. Regressão Linear Múltipla

Os resultados e conclusões da regressão linear múltipla encontram-se compilados na tabela abaixo (tabela 7). Refere-se que a variável X corresponde à variável independente “densidade populacional” e Y ao “poder de compra”.

Tabela 7 – Resumo dos Resultados da Regressão Linear Múltipla

	Equação da recta	R	R ²	Estatística de Teste F _{TESTE} (2, 26)	Conclusão F _{CRITICO} (2, 26)= 5.526335
Pingo Doce	-10.3493 + 0.0011*X + 0.1189*Y	0.826954	0.683853	28.120049	Regressão é significativa
LIDL	-1.4294 + 0.0004*X + 0.0337*Y	0.623012	0.388144	8.246826	Regressão é significativa
Feira Nova	-0.2942 + 0.0002*X + 0.0030*Y	0.716800	0.513803	13.738128	Regressão não é significativa
Intermarché	1.7622 - 0.0002*X - 0.0039*Y	0.439834	0.193454	3.118112	Regressão não é significativa
Jumbo	-0.1231 - 0.0000*X + 0.0031*Y	0.231981	0.053815	0.739387	Regressão não é significativa
E.Leclerc	-0.2942 + 0.0002*X + 0.0030*Y	0.716800	0.513803	13.738128	Regressão é significativa
Continente	-1.4250 + 0.0001*X + 0.0158*Y	0.745507	0.555781	16.264826	Regressão é significativa
Bonjour	-1.8822 + 0.0002*X + 0.0195*Y	0.814042	0.662664	25.537203	Regressão é significativa

Os resultados associados ao Intermarché e Jumbo já eram previsíveis na medida em que na análise da regressão linear simples observou-se que estes não evidenciaram qualquer correlação entre as variáveis independentes.

A tabela 8 sintetiza as conclusões associadas aos coeficientes individuais de regressão. Refere-se que a variável X e Y correspondem respectivamente às variáveis “densidade populacional” e “poder de compra”.

Tabela 8 – Resumo dos Resultados da Análise Aos Coeficientes Individuais de regressão

	Variável Independente	Estatística de Teste t-student TESTE (26) =	Conclusão t-student CRITICO (26) = 3.435
Pingo Doce	X	2.607421	Pouco Significante
	Y	4.819951	Muito signficante
LIDL	X	1.692364	Pouco Significante
	Y	2.370673	Pouco Significante
Feira Nova	X	3.803739	Muito signficante
	Y	1.251849	Pouco Significante
E.Leclerc	X	3.803739	Muito signficante
	Y	1.251849	Pouco Significante
Continente	X	1.707203	Pouco Significante
	Y	3.884183	Muito signficante
Bonjour	X	2.471477	Pouco Significante
	Y	4.604127	Muito signficante

6. Análise de Redes

Com a análise de redes pretende-se avaliar a estrutura da rede através de operações básicas com vista a resolver os seguintes problemas:

- Qual é o melhor caminho entre o supermercado A e o supermercado B?
- Que supermercado serve um determinado local?
- Qual a acessibilidade de um supermercado a partir de outra localização?

A análise de redes focou-se exclusivamente sobre a rede viária do distrito de Lisboa. A figura 35 indica a região de estudo considerada:

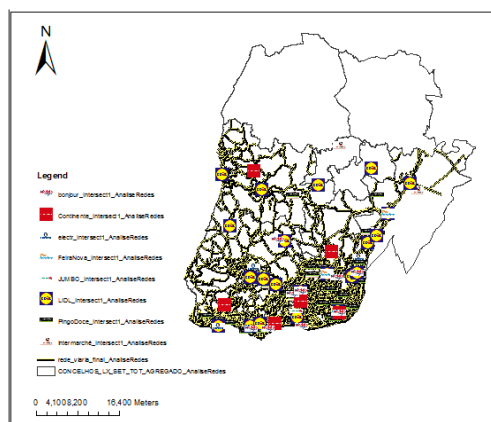


Figura 35 – Mapa dos Resíduos (regressão linear múltipla)

Os subcapítulos subsequentes apresentam os algoritmos de análise de redes utilizados neste trabalho.

6.1. Análise do Caminho Ótimo (Optimal path)

Com este algoritmo pretende-se determinar o caminho óptimo entre 2 ou mais localizações. Neste âmbito, determinou-se o melhor caminho que uma pessoa deverá realizar caso pretenda visitar todos os supermercados Continente e Feira Nova. A figura 36 reflecte o caminho gerado pelo algoritmo:

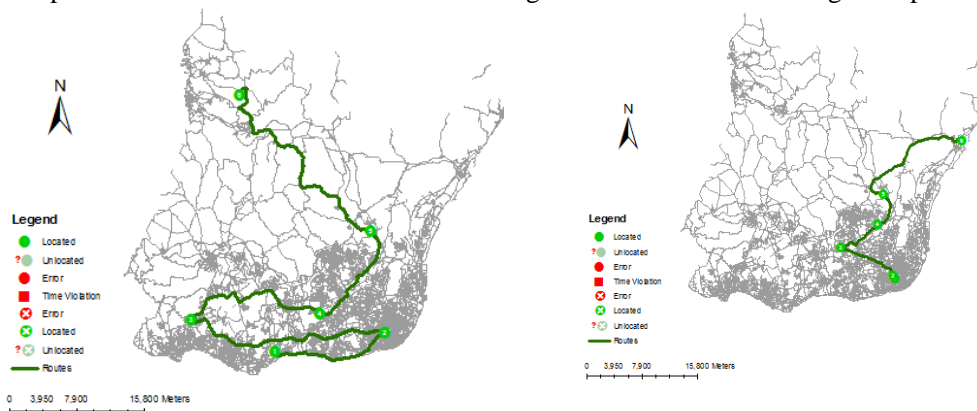


Figura 36 – Caminho Ótimo (Continente)

Em tempos de crise, a poupança no consumo de gasolina e a diminuição das despesas em bens alimentares é algo que todos ambicionamos. Neste contexto a análise do caminho óptimo poderá ser bastante útil nos dias de hoje, na medida em que poderá auxiliar o cidadão na identificação do caminho óptimo a realizar para efectuar as suas compras. A figura abaixo ilustra a trajectória óptima que um cidadão deverá realizar caso pretenda efectuar as suas compras em 5 supermercados diferentes na zona de Sacavém, Loures. Este percurso tem como início e fim uma hipotética residência do cidadão (ponto a verde).

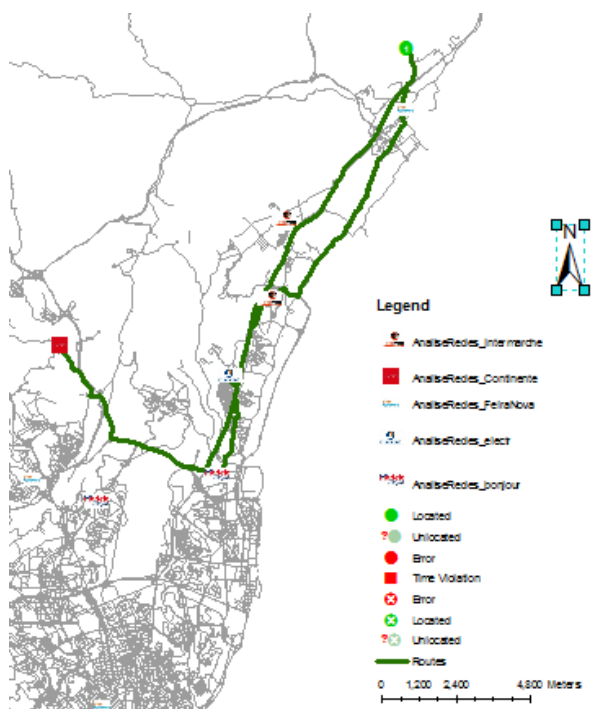


Figura 37 – Caminho Ótimo (trajecto entre diferentes supermercado)

6.2. Localização (Service Area)

Com a aplicação deste algoritmo pretendeu-se identificar as zonas em Lisboa que podem (ou não) ser alcançadas por um supermercado a menos de uma determinada distância. Nesta análise consideraram-se as distâncias: 1500, 3000, 4500 e 10000 m. A figura abaixo (figura 38) expressa os resultados alcançados:

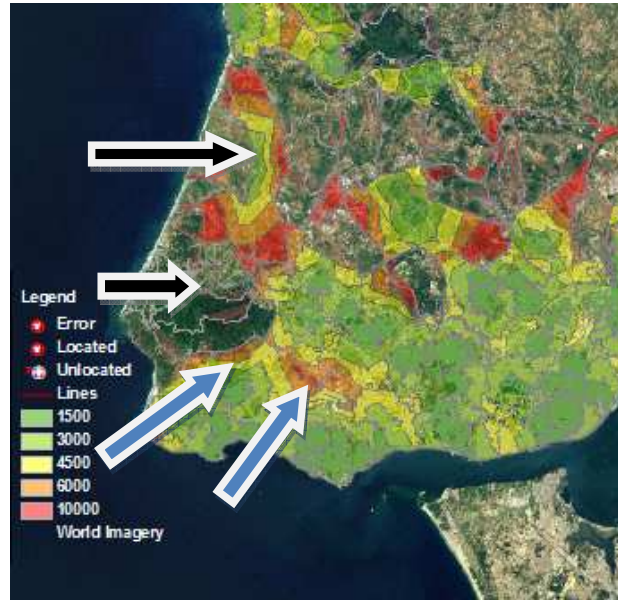


Figura 38 – Service Area (globalidade de eventos na região de Lisboa)

Analisando-se a imagem constata-se que, com excepção das áreas com espaços verdes, toda a área do distrito de Lisboa tem ao seu dispor um supermercado a menos 10.000m. Uma análise mais cuidada sugere que supermercados a distâncias superiores a 6000m são mais frequentes nas zonas de periferia. Detectaram-se nos concelhos de Cascais, Sintra e Oeiras a existência de zonas com supermercados a mais de 6000m.

A título de curiosidade, procedeu-se à análise deste algoritmo a conjuntos específicos de supermercados. As figuras seguintes para além de evidenciarem uma imagem com as *service areas* associadas aos supermercados em causa (Continente, Feira Nova e E.Leclerc), apresenta mais 2 imagens que resultaram da sobreposição das *service areas* com os mapas temáticos da densidade populacional e poder de compra.

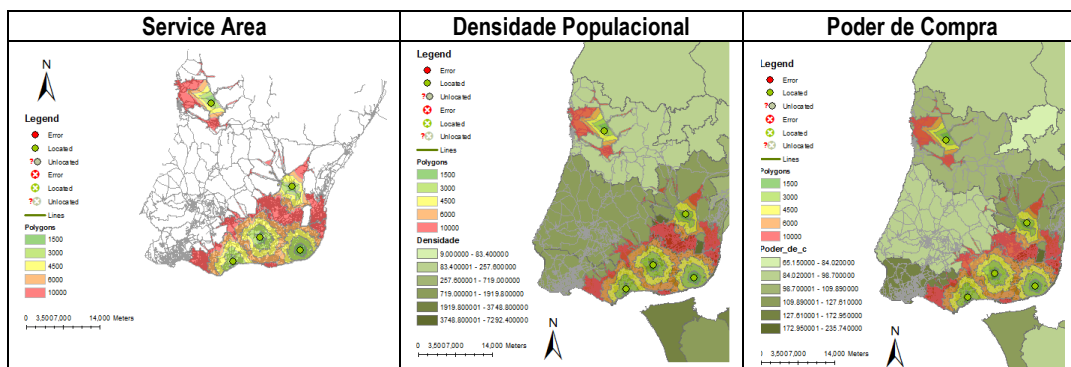


Figura 39 – Service Area (Continente)

Através da figura 39, pode-se constatar que:

- Estes supermercados estão localizados em zonas com forte densidade populacional.
- Estão presentes em zonas com forte poder de compra.
- O supermercado no concelho de Mafra sugere que a cadeia de supermercados Continente privilegia o poder de compra à densidade populacional.

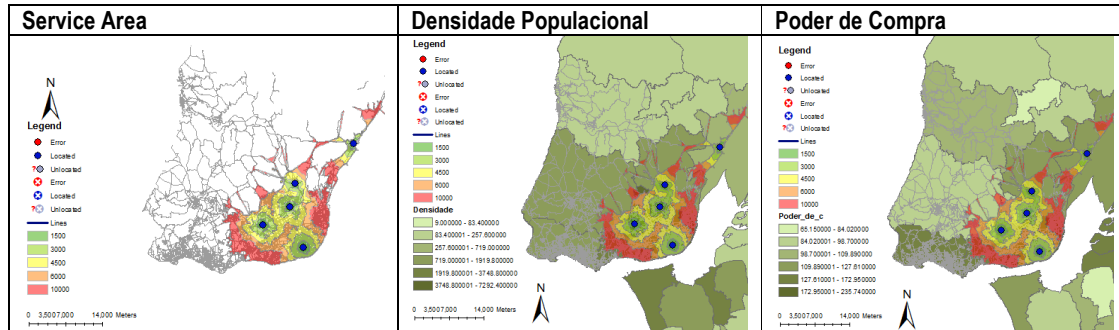


Figura 40 – Service Area (Feira Nova)

Através da figura XX, pode-se constatar que:

- Estes supermercados estão localizados em zonas com forte densidade populacional.
- Privilegiam zonas com forte poder de compra.

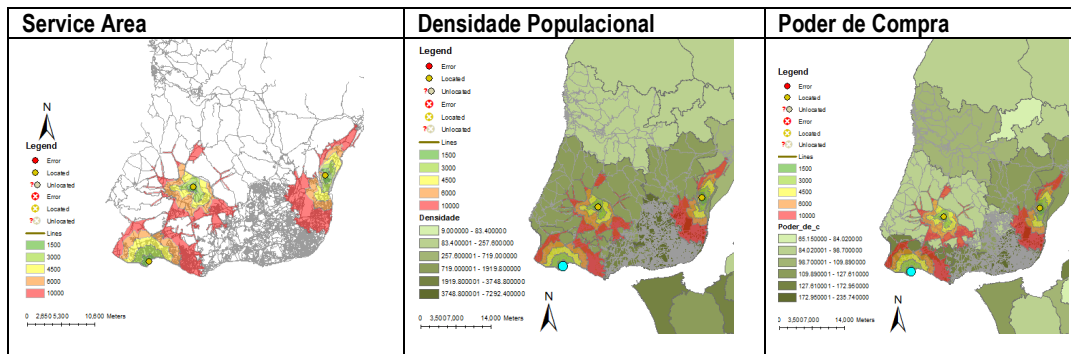


Figura 41 – Service Area (E.Leclerc)

Através da figura XX, pode-se constatar que:

- Estes supermercados estão localizados em zonas com forte densidade populacional, embora não se encontrem presentes no concelho de Lisboa.
- Não privilegiam zonas com forte poder de compra.

6.3. Análise da Acessibilidade (Closest facility)

A análise da acessibilidade pretende avaliar o quão acessível um lugar é de outros lugares. Nesta análise consideraram-se os mesmos supermercados que o algoritmo service área considerou: Continente, Feira Nova e E.Leclerc. Como “incidentes” consideraram-se pontos localizados em zonas que não possuem nas suas proximidades um supermercado da mesma marca e em zonas com forte poder de compra.

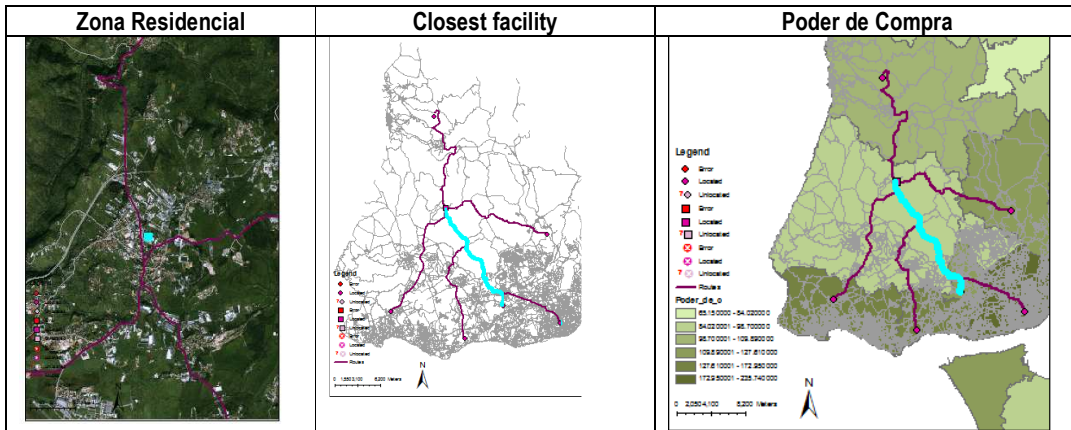


Figura 42 – Closest Facility (Continente)

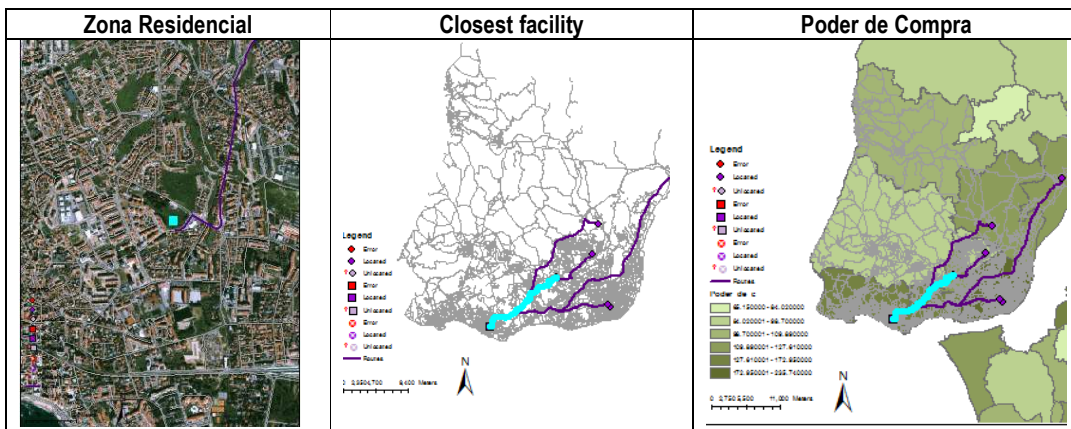


Figura 43 – Closest Facility (Feira Nova)

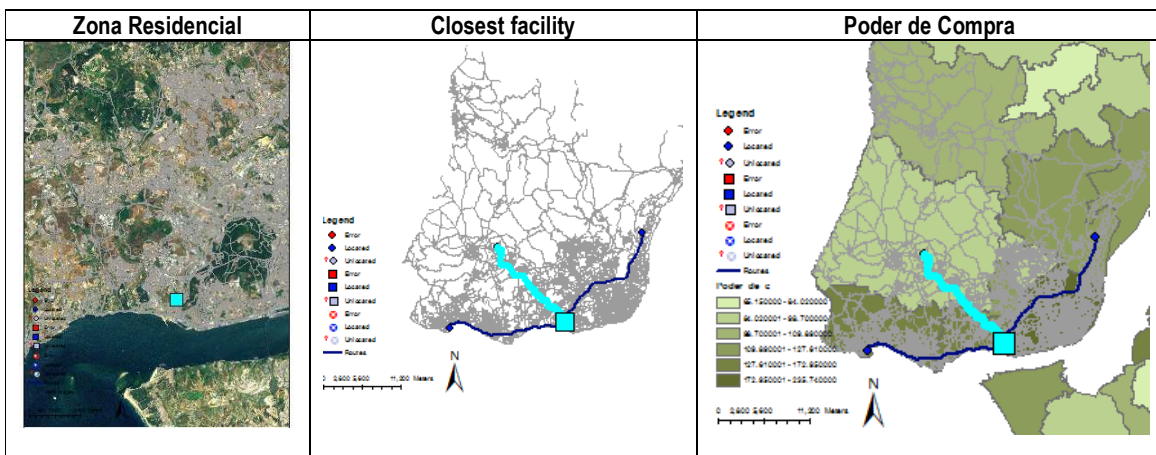


Figura 44 – Closest Facility (E.Leclerc)

7. Conclusões

Os objectivos propostos para este projecto foram atingidos com sucesso pois através da utilização das diferentes ferramentas de análise espacial foi possível perceber qual a sua orientação de expansão assim como compreender o padrão de distribuição dos supermercados. A tabela 9 sintetiza algumas das conclusões obtidas ao longo deste trabalho:

	Global	Intermarché	Pingo Doce	LIDL	E.Leclerc	Modelo Bonjour	Feira Nova	Jumbo	Continente
Padrão Pontual:									
Método Contagem por Quadrantes	Cluster	Aleatório	Cluster	Cluster	Aleatório	Cluster	Cluster	Aleatório	Cluster
Método do Vizinho Mais Próximo	Cluster	Aleatório	Cluster	Cluster	Disperso	Disperso	Disperso	Disperso	Disperso
Áreas:									
Índice Global de Moran	Cluster	Cluster	Cluster	Cluster	x	Aleatório	Aleatório	x	Aleatório

Tabela 9 – Resumo dos Resultados

Por outro lado, embora tenham sido analisadas um número reduzido variáveis independentes explicativas do modelo de expansão de supermercados, conclui-se também que o número de supermercados pode ser explicado pela densidade populacional e poder de compra. No entanto, importa referir que nem todos os supermercados conseguem ser explicados por estas variáveis.

Na análise de redes verificou-se através da aplicação do algoritmo *service area* que as zonas na periferia de Lisboa possuem um supermercado tipicamente a mais de 10 minutos. Esta tendência é invertida quando nos aproximamos do centro da cidade. Este algoritmo permitiu identificar novas localizações para os supermercados Continente, Feira Nova e E.Leclerc. Neste âmbito, percebeu-se claramente da potencialidade que o algoritmo do caminho óptimo pode ter nos dias de hoje.

Não obstante da análise efectuada ter incidido sobre dados inconsistentes e, por vezes, em pressupostos irrealistas, são óbvias as vantagens e benefícios da utilização das diferentes técnicas de análise espacial. No entanto, constatou-se que, embora a análise espacial favoreça a compreensão de um determinado fenómeno, exige algum cuidado no seu manuseamento e interpretação uma vez que podem produzir resultados e conclusões contraditórias entre si. Para minimizar erros de interpretação é fundamental conhecer-se muito bem os dados que estão a ser analisado assim como ter sempre presente os desígnios do projecto.

Referências Bibliográficas

- Análise de Regressão e Correlação, <http://www.estv.ipv.pt/PaginasPessoais/malva/MetodosElectro/Regress%C3%A3o.pdf> (consultado dia 08/07/2011)
- Apostila de ArcGIS, Laboratório de Topografia e Cartografia Universidade Federal do Espírito Santo <http://www.ltc.ufes.br/geomaticsee/Apostila%20de%20ArcGIS%20%28Geodatabase%29.pdf>
- Druck, S.; Carvalho, M.S.; Câmara, G.; Monteiro, A.V.M. “Análise Espacial de Dados Geográficos”, <http://www.dpi.inpe.br/gilberto/livro/analise/> (consultado 08/07/2011).
- Eastman J.R., “ Manual do usuário”, http://www.ecologia.ufrgs.br/labgeo/arquivos/downloads/Tutorial_Idrisi_for_Windows_2.pdf (consultado dia 08/07/2011)
- Godoy, C. G. M., (2010) “Análise Espacial das Aglomerações de Comercio e Prestação de Serviços a Varejo No Sistem Urbano”, <http://www.lume.ufrgs.br/handle/10183/29129> (consultado dia 08/07/2011)
- Gomes, R. J., (2008) “Análise Espacial dos Acidentes de Transito do Município de Vitória Utilizando Sistema de Informações Geográficas”, (consultado dia 08/07/2011)
- Índice do poder de compra, INE http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=56951125&DESTAQUESmodo=2 (consultado dia 20/06/2011).
- Jacomio, C.A., Achibana, V.M., Imai, N.N, Flores, E.F., “Aplicação De Técnicas De Estatística Espacial Na Caracterização Dos Casos De Dengue No Município De Presidente Prudente – SP No Período De 1999 A 2007”, Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da geoinformação, http://www.ufpe.br/cgtg/SIMGEOIII/IIISIMGEO_CD/artigos/CartografiaeSIG/SIG/A_192.pdf (Consultado dia 08/07/2011)
- Queiroz, M.P., (2003) “Análise Espacial Dos Acidentes De Trânsito Do Município De Fortaleza”, http://www.google.pt/url?sa=t&source=web&cd=3&ved=0CCoQFjAC&url=http%3A%2F%2Fwww.det.ufc.br%2Findex.php%3Foption%3Dcom_docman%26task%3Ddoc_download%26gid%3D115%26Itemid%3D137&rc=t-j&q=Marcelo%20Pereira%20Queiroz%20AN%C3%81LISE%20ESPACIAL%20DOS%20ACIDENTES%20DE%20TR%C3%82NSITO&ei=mh4YTpeBIJGn8QOopu0d&usq=AFQjCNEJWNnGhoRnOwtje16OPdDkj4LlXw , (consultado dia 08/07/2011)
- Mota, M., “Concepção de Curricula em Análise Espacial para o Terceiro Ciclo do Ensino Básico”, <http://www.isegi.unl.pt/servicos/documentos/TSIG/TSIG0007.pdf> (consultado dia 08/07/2011)
- Ortiz, J.L., Vettorazzi, C.A., Couto, H.T.Z, Gonçalves, J.L.M, Relações espaciais entre o potencial produtivo de um povoamento de eucalipto e atributos do solo e do relevo(consultado dia 08/07/2011)
- Ribeiro, B.M.G, “SER-301: ANÁLISE ESPACIAL DE DADOS GEOGRÁFICOS”, 2008
- Slides da aulas teorias de AEIG (2010/2010)
- Tutoriais das aulas práticas de AEIG (2010/2010)