



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

**EXPLORACIÓN Y ANÁLISIS ESPACIAL
EN ELECCIONES LOCALES.
LA MEDICIÓN Y DETECCIÓN DE LOS USOS
ELECTORALES DEL GASTO SOCIAL EN LA ALCALDÍA DE
COYOACÁN, CIUDAD DE MÉXICO**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

**LICENCIADO EN MATEMÁTICAS
APLICADAS Y COMPUTACIÓN**

P R E S E N T A:

DIEGO ALEXIS GALVÁN SANDOVAL



**DIRECTOR DE TESIS:
DR. EMANUEL RODRIGUEZ DOMINGUEZ**

**Santa Cruz Acatlán, Naucalpan, Estado de México ,
Junio 2021**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dedicatoria

Para mi padre y madre por su apoyo invaluable e incondicional que me han dado durante toda mi vida y a los que les debo más de lo que podría retribuirles. A mi asesor y profesores por su apoyo y conocimiento que de ellos he adquirido.

Investigación realizada en el marco del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM: IA302720 – Antropología de los gobiernos locales en México: los sentidos de la gestión pública y las innovaciones autonómicas.

Índice

Introducción.....	7
Capítulo 1 Planteamiento y justificación del problema	9
1.1. Antecedentes.....	9
1.2. Geografía electoral.....	11
1.3. Estadística espacial.....	13
1.4. Econometría espacial.....	14
1.5. La problemática	15
1.6. Metodología e información.....	17
1.7. Objetivos	18
Capítulo 2 Métodos estadísticos en la investigación social.....	19
2.1. Análisis exploratorio de Datos y Análisis Exploratorio de Datos Espaciales ..	20
2.2. Unidad espacial de análisis.....	21
2.3. Indicadores de competitividad electoral.....	22
2.4. Método de ponderación de área simple	24
2.5. Método de agrupación K-means.....	26
2.6. Matriz de contigüidad y vecindarios	27
2.7. Retardo espacial.....	30
2.8. Coeficientes de correlación	31
2.9. Autocorrelación espacial o dependencia espacial.....	33
2.10. Indicadores de asociación espaciales globales y locales.....	35
2.11. Análisis de regresión, modelos espaciales y medidas de comparación.....	44
Capítulo 3 Los usos electorales de la pobreza: desarrollo social, programas de atención a la pobreza y disputas electorales.....	55
3.1. Clientelismo político y prácticas clientelares en México.....	55
3.2. Clientelismo político en Coyoacán 2018	57
3.3. Espacios vulnerables, marginados y votantes	62
Capítulo 4 Geografías de poder; comportamientos electorales en Coyoacán 2009 –2018.....	65
4.1. Comportamientos electorales 2009-2012. Los partidos tradicionales.....	67
4.2. Comportamientos electorales 2015-2018. Nuevas opciones políticas.....	71
Capítulo 5 Análisis estadístico y representación de datos: los usos electorales del gasto social en la Alcaldía de Coyoacán 2018	80

5.1. Datos y variables.....	80
5.2. Análisis exploratorio de datos espaciales.....	81
5.3. Modelos de regresión.....	85
Conclusiones	88
Bibliografía.....	91
Anexos	97
Anexo 1. Glosario	97
Anexo 2. Software libre utilizado	100
Anexo 3. Simulador y códigos utilizados.....	100

Introducción

Las Matemáticas Aplicadas a las Ciencias Sociales han significado una manera precisa de análisis y estudio de fenómenos sociales de manera cuantitativa. En la actualidad y gracias a un mayor acceso a grandes cantidades de información los métodos matemáticos (principalmente estadísticos) para el análisis de datos han ido penetrando cada vez más en las investigaciones dentro de áreas político-sociales. El matemático Daniel Peña (Peña, 2019) refiere tres puntos importantes para justificar el uso de las Matemáticas como herramientas en la adquisición de conocimiento en las Ciencias Sociales. Primero, porque su aplicación obliga a definir y plantear variables e hipótesis de manera clara. Segundo, porque permite la importación de modelos matemáticos, aplicados con éxito en otras áreas, para la explicación de fenómenos sociales, y finalmente, por la gran cantidad de datos y el poder de cómputo disponible en nuestros días que apoyan y facilitan el desarrollo de análisis cuantitativos. Lo que obliga a los investigadores de las Ciencias Sociales a incorporar herramientas matemáticas a su bagaje, además de ofrecer un campo fértil para la aplicación de los conocimientos de los académicos del área matemática y computacional interesados en temas relacionados con la sociedad.

A lo largo de esta tesis se hace uso de herramientas de la geografía electoral, matemáticas y computacionales en el estudio del caso de las elecciones para alcalde en Coyoacán, Ciudad de México del año 2018. Dichos métodos permiten la inclusión y tratamiento de la variable geográfica, es decir, la localización donde ocurrieron los hechos y su interrelación con las áreas aledañas agregando una dimensión más al análisis. Estas elecciones, enmarcadas en un contexto de acusaciones mutuas entre los candidatos de incurrir en prácticas de compra y coacción del voto, y en particular del uso de programas sociales en beneficio del partido en el poder (Partido de la Revolución democrática, PRD), llegó hasta los tribunales. Sin embargo, debido a que se consideró que no existían pruebas suficientes de estos actos, dichas acusaciones fueron desestimadas. Por lo que un análisis cuantitativo y cualitativo de los datos de esta elección es efectuado en el presente trabajo en orden de explorar lo ocurrido.

La tesis se encuentra dividida en dos grandes partes: la primera, consta de una revisión de la literatura especializada en actos de compra y coacción del voto y en la descripción de los métodos estadísticos y las técnicas aplicables a la geografía electoral. La segunda consta de la aplicación de estas técnicas en el caso particular de Coyoacán a nivel de sección electoral. En este trabajo se analizó el padrón de

beneficiarios del programa social “A tu lado” enfocado a apoyar a la población en pobreza de Coyoacán por medio de transferencias de dinero. En su conjunto ambas partes permiten relacionar la cantidad de beneficiarios del programa social y otras variables con la preferencia electoral hacia el PRD de manera positiva, esto es, un aumento en el número de beneficiarios generaría un aumento en la preferencia del voto hacia el PRD. Además, se pudieron definir e identificar espacios de alta competitividad electoral, gran cantidad de votantes y un grado de desarrollo social bajo que pueden ser relevantes en análisis futuros a un nivel individual.

Capítulo 1

Planteamiento y justificación del problema

1.1. Antecedentes

El análisis electoral a nivel geoespacial (en adelante geografía electoral) vincula de manera interdisciplinaria elementos de la geografía política, la estadística aplicada y la ciencia política para definir un campo de estudio que comprende el análisis del voto y los resultados electorales desde una perspectiva espacial y temporal, con el objetivo de explicar la relación del espacio, los actores que en él se desenvuelven, las territorialidades que en él se expresan y los resultados electorales. Es decir, este subcampo de la geografía busca explicar las preferencias de la población votante tomando en cuenta el territorio donde habitan, las relaciones de poder y las características socio-económicas que éstos poseen (Monzón, 2009). La geografía electoral se caracteriza por su remarcado énfasis en los métodos cuantitativos y cualitativos; los estudios dentro de este campo se basan en una metodología estadística para explorar y explicar los resultados de los comportamientos políticos; no obstante, desde el análisis político contemporáneo a las metodologías estadísticas se suman los análisis cualitativos para obtener mejores resultados. La geografía electoral ofrece un campo fértil para el análisis estadístico de las elecciones proveyendo de herramientas para el estudio de comportamientos electorales, que tienen un potencial en la toma de decisiones y permiten determinar lugares en donde se reproducen prácticas que afectan el desarrollo democrático de los comicios.

Desde sus inicios la geografía electoral echó mano de representaciones gráficas, pues el insumo básico de esta disciplina son los mapas que muestran la realidad de manera espacial. Ejemplo de ello es el trabajo de Paulin, quien en 1932 presentó una serie de mapas de las elecciones de los Estados Unidos, en los que se encontraron relaciones entre la preferencia partidista y las características sociodemográficas de la población (porcentaje de personas de color, servicios educativos, etcétera) vale decir que en estos mapas también se reflejaban datos que aludían a las variables económicas de la región (precisando las actividades productivas de los territorios analizados) (Prescott, 1959). Se puede considerar que los estudios de E. Krehbeil

(Krehbiel, 1916) son pioneros en este campo, pues analizaron las elecciones parlamentarias en Inglaterra entre los años 1880 y 1910, encontrando que los votantes de las regiones industrializadas en general tenían preferencia por los representantes liberales, mientras que las zonas agrícolas y menos industrializadas se decantaban por representantes conservadores. En el trabajo de Krehbeil era evidente la influencia de factores geográficos en la construcción de la preferencia política (Prescott, 1959).

Hoy día los estudios en la geografía electoral se encuentran prácticamente en todos los países democráticos. Además, los avances tecnológicos han facilitado el acceso a la información, el diseño y la publicación de mapas electorales. En México, la geografía electoral cuenta con más de 40 años de existencia. Los primeros trabajos en esta área fueron publicados en la década de los setenta (Zavala, 2012), aunque por desgracia esta disciplina no ha tenido un amplio desarrollo en nuestro país (Vilalta, 2008). Algunos autores atribuyen esto al desprecio hacia la geografía electoral por parte de los estudiosos de las ciencias sociales, a lo que se suma la dificultad y complejidad en las cuestiones técnicas vinculadas a la estadística aplicada y al análisis espacial (Zavala, 2012). Sin embargo, gracias al desarrollo tecnológico es posible el acceso a gran cantidad de información georreferenciada y al uso de programas computacionales que sirven para el análisis de datos espaciales. Esto permite el aprovechamiento de las características descriptivas y predictivas de la geografía electoral, las cuales pueden ser aplicadas de manera individual o conjunta para la descripción en el análisis histórico y prospectivo de las elecciones en México.

En los pioneros estudios de geografía electoral mexicana destacaba el análisis estadístico basado en el agrupamiento de patrones y clústeres electorales para definir las cartografías de las tendencias políticas o los resultados electorales. Es decir, el agrupamiento estadístico de la preferencia electoral hacia cierto partido o partidos en zonas geográficas definidas, y que resulta ser excluyente entre partidos políticos, lo que indica que cada uno tiene su zona de influencia. Este tipo de análisis se han enfocado en el análisis de los comicios presidenciales (Hernández, 2015) y las elecciones en la Ciudad de México (Vilalta, 2008); algunos de estos estudios consideran importante incluir variables socioeconómicas como la pobreza, la tasa de desempleo o el nivel educativo en los espacios geográficos asociados a las cartografías del voto pues consideran que estos elementos están relacionados con las decisiones políticas. La geografía electoral en México permite conocer los comportamientos electorales, entender el diseño de las campañas políticas, las relaciones y disputas de poder en un territorio en específico y, también, permite profundizar en malas prácticas que se

presentan en los comicios (como la compra de voto y la intermediación política), principalmente en territorios conflictivos o con bajos niveles de desarrollo social.

Las elecciones, como todo proceso social, se llevan a cabo en un espacio geográfico, pero al ser un proceso institucional este espacio está delimitado y definido por información estadística producida por los gobiernos y los organismos encargados del proceso electoral. Regularmente las cartografías electorales facilitadas por las instituciones contienen datos estadísticos agregados que delimitan secciones y distritos electorales por estados y municipios. Así los métodos utilizados en la geografía electoral resultan ser estadísticos, desde descriptivos que sirven para explorar los comportamientos electorales y su relación con la geografía, hasta predictivos cuya mejor expresión son aquellos que consideran de dependencia espacial del voto. Algunas de estas estrategias para el análisis espacial fueron desarrolladas alrededor del año 1950 (Moran, 1950) y popularizadas en 1980 y años posteriores ((Anselin, 2013), (O'Loughlin & Anselin, 2008)). Su aplicación en la geografía electoral ha servido para que otras disciplinas empiecen a tomar en cuenta el espacio geográfico como una variable importante en sus análisis, es el caso de la econometría espacial, la ciencia de la planificación regional y la estadística espacial (esta última es considerada por algunos autores como un campo diferente al de la estadística) (Griffith, 2009).

Finalmente, los programas computacionales actuales y los lenguajes de programación permiten realizar el análisis exploratorio de datos espaciales, la medición y detección estadística de patrones y clústeres. Además, dichos programas potencian una visualización de los datos permitiendo representar los resultados estadísticos por medio de mapas temáticos que son fundamentales para el análisis político desde una perspectiva interdisciplinaria. Este software ha materializado el rublo de los sistemas de información geográfica (SIG). Vale decir que en la actualidad existen programas computacionales (libres y de paga) para el análisis estadístico de datos de diversa índole los cuales permiten la construcción y visualización de representaciones gráficas y mapas, es el caso de Arcgis, Qgis, Geoda y los paquetes geográficos para los lenguajes de programación R y Python entre muchos otros. En las siguientes líneas de esta tesis destacaré como la aparición de estos programas computacionales facilitan el tratamiento de una gran cantidad de datos, su análisis estadístico y la representación de resultados, que pueden servir para la solución de problemas concretos.

1.2. Geografía electoral

La geografía electoral surgió en Francia durante los años 1870-1940 (Zavala, 2012), (Forest, 2017) como una subdisciplina de la ciencia política, que fue definida como

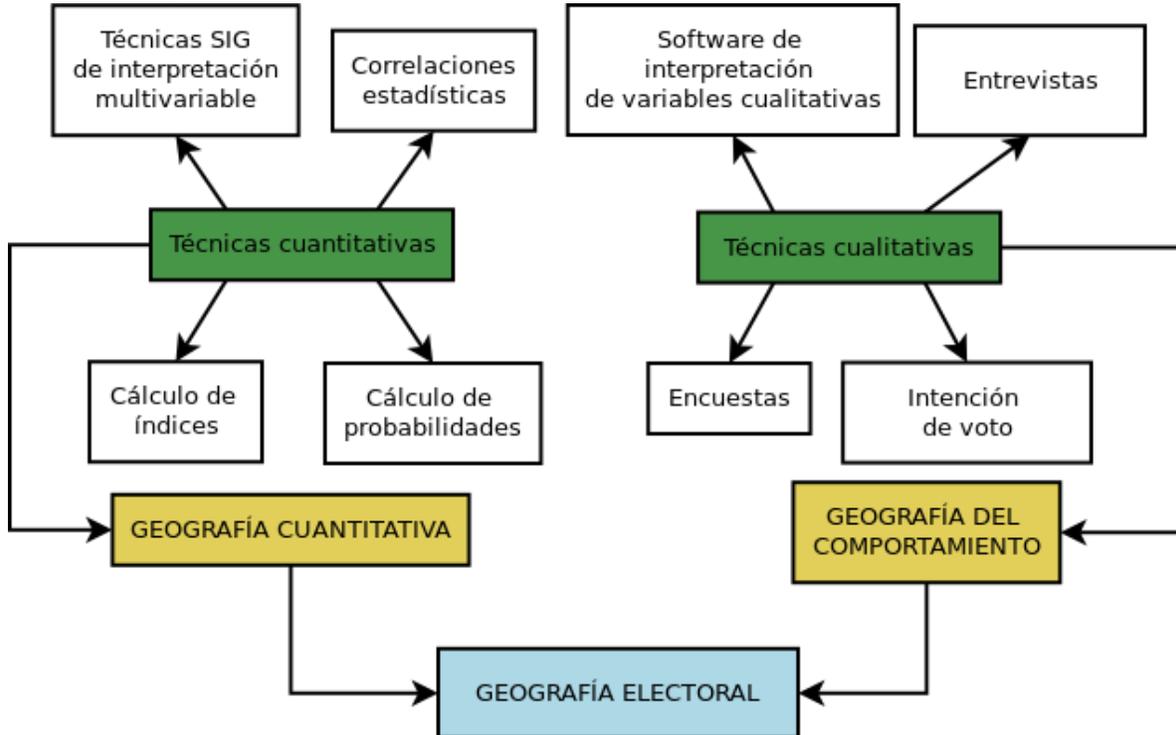
parte de la geografía política, y que aborda el estudio del sufragio desde una perspectiva espacio temporal de los resultados electorales. Cabe destacar que lo innovador de esta disciplina es que pone el acento en indagar las causas que explican los comportamientos de los votantes. Con el paso del tiempo se ha confirmado que la geografía es importante en los estudios electorales, pues los comportamientos políticos y las tendencias del voto no pueden no considerar la influencia socioespacial que se tiene con la localidad o la región en donde se desarrolla (Forest, 2017).

Tradicionalmente han existido dos enfoques dentro de la geografía electoral: el cronológico y el ecológico. El primero, de tipo estadístico y descriptivo, pretende estudiar la relación entre los resultados electorales y la geografía física del espacio sin tomar en cuenta características de la población votante. El segundo, incluye tanto técnicas cuantitativas como cualitativas para analizar variables sociales (niveles de ingreso de la población, índices de desarrollo social, educación, etcétera). Este enfoque mixto al combinar métodos cuantitativos con técnicas estadísticas construye indicadores de competitividad, correlaciones, regresiones y probabilidades.

Hoy día, el enfoque ecológico se ha potenciado con el uso de programas computacionales para el análisis de grandes cantidades de datos, que de manera prospectiva construyen escenarios virtuales en donde se pueden aplicar estrategias de investigación cualitativa, para reforzar los datos estadísticos mediante encuestas, sondeos y entrevistas. La información cuantitativa y cualitativa sistematizada en bases de datos es el insumo para representar los indicadores relacionados a una problemática político-electoral de forma visual, por medio de cartografías enfocadas al tema (mapas del voto, influencia regional por partido, niveles de competitividad partidaria, nichos de intercambio político, etcétera). Ejemplo de lo anterior son los trabajos que aplican métodos estadísticos a la geografía electoral (Vilalta, 2006), (Terron & Dillion Soares, 2008).

Para presentar esquemáticamente los métodos de la geografía electoral, donde se destaca el uso y combinación de una metodología mixta, que incluye técnicas matemáticas, etnográficas y computacionales, presento la siguiente figura (1.1), la cual es una modificación de la mostrada en (Monzón, 2009).

Figura 1.1: Diagrama de la geografía electoral con sus componentes



Fuente: (Monzón,2009)

1.3. Estadística espacial

Como mencionamos anteriormente la geografía electoral se vale de métodos estadísticos para el análisis de las elecciones. Los tipos de datos que generalmente encontramos en estudios de fenómenos relacionados con el espacio geográfico requieren de un enfoque distinto para ser analizados y para obtener resultados válidos, ya que muchos métodos estadísticos tradicionales no contemplan los efectos que tiene el espacio en sus modelos e incluso la descripción de los datos espaciales puede presentar diferencias. Los métodos estadísticos utilizados para el análisis de datos georreferenciados forman parte de la estadística espacial o análisis espacial y buscan responder a las preguntas de ¿cuánto es significativo a nivel estadístico? y ¿dónde están sus particularidades espaciales? De acuerdo con la definición sobre la estadística espacial tenemos lo siguiente:

Campo de estudio que trata de los métodos estadísticos que utilizan espacios y relaciones espaciales (tales como distancia, área, volumen, longitud, altura, orientación, centralidad u otras características espaciales)

de los datos) directamente en sus cálculos matemáticos. La estadística espacial se emplea para toda una variedad de tipos de análisis diferentes, tales como el análisis de patrones, análisis de formas, modelado de superficies y predicciones de superficies, regresión espacial, comparaciones estadísticas de datasets espaciales, modelado y predicción estadísticos de la interacción espacial, etcétera. Entre los muchos tipos de estadísticas espaciales están las estadísticas descriptivas, inferenciales, exploratorias, geoestadísticas y econométricas. (Esri, 2019a).

Así, la estadística espacial comprende el conjunto de técnicas para describir y modelar datos a nivel geográfico, evaluar patrones, distribuciones, tendencias, procesos y relaciones espaciales entre los objetos georreferenciados (Sánchez Perales, 2012). Un aspecto importante de esta rama de la estadística es que está basada en la no independencia de las observaciones; esto implica que la investigación se funde en la suposición de que las unidades más cercanas están asociadas de alguna manera de acuerdo con la ley de Tobler (Tobler, 1970). Esto hace que se diferencie de otras ramas de la estadística donde se asume que las variables están idénticamente distribuidas y que son independientes (Manterola, 2007). Por ello, algunos autores mencionan que puede ser considerada como un área de investigación distinta a la estadística tradicional en vez de una rama de ella (Getis, 1999).

1.4. Econometría espacial

La econometría espacial es la subrama de la econometría que se encarga de los efectos espaciales (autocorrelación y heterogeneidad a nivel del espacio) en modelos de regresión estadística para datos transversales y datos en diferentes tiempos, así como el conjunto de técnicas de especificación, estimación, contraste y predicción para el tratamiento de problemas espaciales. El término de econometría espacial fue acuñado por Jean Paelinck y Klaseen en la década de los setenta para designar a un campo en crecimiento de la ciencia regional que trata primordialmente con problemas de estimaciones y test encontrados en la implementación de modelos econométricos aplicados a la planeación de desarrollos urbanos, agrícolas e industriales en diferentes escalas a nivel espacial y regional (Anselin, 2013).

La econometría y estadística espaciales están estrechamente relacionadas y es común encontrar en la literatura científica que ambos conceptos se asumen erróneamente como equivalentes. Luc Anselin menciona que la diferencia entre la econometría espacial y la estadística espacial es la misma que existe entre la econometría y la

estadística (Anselin, 2013). Sin embargo, esta precisión sigue resultando poco clara, pues los métodos utilizados tienden a aparecer en ambas disciplinas en estudios aparentemente similares. Los métodos estadísticos tienden a ser categorizados como pertenecientes a uno u otro campo dependiendo de la preferencia personal del investigador (Anselin, 2013). De nuevo, esta categorización resulta ambigua y no provee una verdadera distinción entre ambas disciplinas. Anselin indica que econometría espacial trata con cuestiones sociales mientras que la estadística espacial trata con fenómenos físicos. Pero otras investigaciones sociales (Vilalta, 2008), ubican su análisis dentro de la estadística espacial y no hacen mención a la econometría. De igual modo trabajos como (Darmofal, 2006) y (Acevedo & Velásquez Ceballos, 2008) se ubican en la econometría espacial desmarcándose de la estadística del mismo tipo. Finalmente, Martori y Hoberg (Martori & Hoberg, 2008) encuentran que la principal diferencia entre ambas es la importancia que la econometría espacial pone a la modelización.

Los conceptos de dependencia y heterogeneidad espaciales llamados “efectos espaciales” serán explicados más adelante. Sin embargo, estos conceptos están relacionados con las características geográficas del objeto de estudio, que nuestro caso son las elecciones locales y la dinámica político-territorial de los espacios de la Ciudad de México en donde sobresalen ciertos comportamientos políticos.

1.5. La problemática

El primero de Julio del 2018 se celebró el proceso electoral en nuestro país en donde se eligieron, entre otros cargos, al presidente y gobernadores, diputados federales y senadores y en particular en la CDMX, al jefe de gobierno, alcaldes para cada una de las 16 alcaldías y diputados locales. Como ya es costumbre este proceso electoral no estuvo exento de acusaciones mutuas entre partidos que incluían la denuncia de actos de clientelismo político hasta actos de violencia de género y coacción del voto. Ejemplo de ello es la elección para alcalde de Coyoacán en la Ciudad de México donde las acusaciones fueron sobre la compra del voto, uso de programas sociales con fines electorales y violencia de género. Más adelante detallaremos este caso particular, ya que derivó en una serie de recursos legales en los que determinar de manera estadística las malas prácticas electorales resultará un tema fundamental.

Los cambios y fragmentación en la geografía política de la CDMX van a sufrir modificaciones radicales en las elecciones del 2018. Previo a esos comicios, en la mayoría de las demarcaciones el partido en el poder era el Partido de la Revolución democrática (PRD). Sin embargo, desde el año 2015 aumentó la competitividad

electoral con la aparición de un nuevo partido de izquierda (Movimiento de Regeneración Nacional (MORENA)), este hecho promovió una competencia más cerrada por los cargos de elección popular durante las elecciones del 2018. Derivado de la alta competitividad en dichos comicios, en la alcaldía de Coyoacán este partido fue acusado de utilizar los programas sociales “A tu lado”, “Entrega de Tabletas Electrónicas”, “Tu Unidad sin Goteras” y “Entrega de Juguetes a niños y a niñas” con fines electorales, además de ejercer violencia de género contra la candidata de la coalición MORENA-PT-PES María de Lourdes Rojo. La situación al final de la contienda llegó a tal punto que el 21 de septiembre de ese mismo año la Sala Regional de la Ciudad de México del Tribunal Electoral del Poder Judicial de la Federación (TEPJF) anuló la elección en la cual el candidato de la coalición PAN-PRD-MC Manuel Negrete había resultado triunfador.

La anulación de la elección en la alcaldía de Coyoacán fue celebrada en un principio por los partidos de oposición como el PT y MORENA, no obstante y casi 10 días después la Sala Superior del TEPJF revocó la anulación de la elección de la alcaldía de Coyoacán al concluir que no existían pruebas contundentes de la violencia de género y el uso de los programas sociales con fines electorales, pues este último punto había de probarse “fehacientemente” y sobre todo establecer con precisión si dichas violaciones a la normatividad son “determinantes” en los resultados de la elección favoreciendo al candidato ganador. Esto implica un reto empírico dado que resulta casi imposible determinar cuántos votos se deben comprar para anular una elección. Más aun cuando la ley establece que este hecho debe ser acreditado de manera objetiva y material y no con base en inferencias. Ante esta interpretación de la norma la sala superior ratificó finalmente como ganador al candidato de la coalición PAN-PRD-MC.

El caso de Coyoacán evidencia que la tarea de demostrar estadísticamente en qué medida la compra y coacción del voto mediante recursos públicos y programas sociales durante las campañas electorales es complicada. En primera instancia, por la naturaleza del problema, pues los partidos políticos no desean que se les relacione con estas actividades y por ello no es sencillo encontrar información que sirva para demostrar y cuantificar las malas prácticas electorales, más allá de las encuestas realizadas a la población, que, si bien proporcionan información, es en muchos casos insuficiente. En contraste, al analizar la literatura especializada y al utilizar técnicas de la estadística espacial aplicada al análisis de los datos proporcionados por las estadísticas y mediciones oficiales, podemos elaborar espacios mediante la construcción de indicadores, sus relaciones estadísticas, y la aplicación de modelos, que nos permitan ubicar secciones o áreas con características deseadas para inferir la posible vulnerabilidad de los habitantes de estos espacios a presentar o reproducir

este tipo de actos¹. Además, permite explorar la relación entre los beneficiarios de los programas sociales y los resultados electorales en orden de identificar patrones, que, si bien no constituyen una prueba por sí misma, sí apoyan en la toma de decisiones jurídicas en casos como los de Coyoacán. Así, el propósito de esta tesis es la de combinar técnicas de la estadística espacial con la literatura especializada sobre estos actos para identificar espacios vulnerables y estudiar la relación entre los programas sociales y los resultados electorales en Coyoacán.

1.6. Metodología e información

Los conjuntos de datos analizados en esta tesis proceden de las siguientes fuentes de información públicas; el Instituto Nacional Electoral (INE), Instituto Electoral del Distrito Federal (IEDF), Instituto Electoral de la Ciudad de México (IECM), Consejo de Evaluación para el Desarrollo Social de la Ciudad de México (EVALUA CDMX), el Sistema de Información de Desarrollo Social (SIDESO), Instituto Nacional de Estadística y Geografía (INEGI) y el Observatorio Social Universitario de la Facultad de Ciencias Políticas y Sociales de la UNAM. Los conjuntos de datos analizados son los siguientes:

- Resultados de las elecciones en Coyoacán de los años 2009, 2012, 2015 y 2018 por sección electoral.
- Padrones de beneficiarios del programa social “A tu lado” del 2017.
- Datos socioeconómicos de Coyoacán referenciados geoelectoralmente.
- Encuesta “Dinámicas y Prácticas Político-Electorales y Clientelares en la Ciudad de México. El caso de los Pedregales de Coyoacán” del año 2018.

Todos los conjuntos de datos son o pueden ser transformados en conjuntos de datos agregados por zona geográfica por lo que pueden ser analizados de manera espacial. En el caso de los resultados electorales los datos están agregados por sección electoral al igual que los datos de la encuesta. Los índices de desarrollo social y los padrones de beneficiarios del programa social están agregados por colonia y se utiliza

¹ Esta primera estrategia analítica evita caer en la falacia ecológica, pues no determinamos un comportamiento individual en relación a un marco espacial derivado del análisis estadístico. No obstante, este ejercicio nos permite ubicar divisaderos de prospección cualitativa en donde es posible que acontezcan situaciones relevantes para nuestro análisis de la intermediación y clientelismo político (Rodríguez, 2015).

un método de desagregación para estimar el número de beneficiarios por sección electoral.

Antes de estudiar la relación entre el padrón de beneficiarios del programa social “A tu lado” y los resultados electorales para el PRD en la alcaldía de Coyoacán, definimos los conceptos utilizados en este trabajo, exploramos las condiciones socioeconómicas de la demarcación, e identificamos espacios de vulnerabilidad para reproducir prácticas de intermediación y compra de voto en sus enclaves locales. Luego estudiamos los comportamientos electorales históricos durante el periodo del 2009 al 2018, que abarca cuatro elecciones para alcalde, mediante las cuales generamos una reconstrucción diacrónica a nivel cartográfico. En cada proceso, los niveles de competitividad son calculados a nivel seccional y de la demarcación, son incluidos como una característica más en los espacios vulnerables a la intermediación política. Este análisis geoestadístico permite conocer un contexto previo a las elecciones del 2018, el cual es la base para analizar la dinámica política de la alcaldía en los comicios de dicho año. Finalmente, y por medio de un modelo de regresión espacial estudiamos la relación que guardan los beneficiarios del programa social con los resultados obtenidos por el PRD en el 2018.

1.7. Objetivos

Este trabajo tiene como meta explorar estadísticamente las elecciones para alcalde en Coyoacán del año 2018. Con el objetivo de establecer espacios vulnerables a reproducir actos de intermediación o clientelismo político dentro de la demarcación. Así como estudiar la relación entre el padrón de beneficiarios del programa social “A tu lado” y los resultados electorales obtenidos por el PRD derivado de los reportes realizados durante esas elecciones. El trabajo también busca implementar diferentes técnicas estadísticas y servir indirectamente como una metodología reproducible y aplicable a otras demarcaciones de la Ciudad de México de manera que apoyen en la toma de decisiones jurídicas para validar el desarrollo democrático de las elecciones, así como para la creación de estrategias de prevención por parte de las fuerzas políticas, las autoridades electorales y los organismos de estado que se encargan de los procesos democráticos.

Capítulo 2

Métodos estadísticos en la investigación social

Los métodos cuantitativos en la investigación social corresponden a métodos estadísticos, dada la necesidad de cuantificar, describir, resumir e inferir conclusiones desde los datos del fenómeno estudiado. En el caso de los estudios electorales estos métodos se utilizan para medir y reportar los resultados de los comicios, los indicadores sociodemográficos de la elección, las particularidades territoriales de las secciones electorales y otros indicadores que se pueden construir a partir de todos los anteriores, para representar espacios de conflicto político o disputas partidarias mediante cartografías propias de la geografía electoral. La metodología y las técnicas de la estadística son utilizadas en los estudios electorales para describir el fenómeno por medio de los análisis de datos y poder explicar comportamientos políticos en el área geográfica que respondan a preguntas de interés para el investigador. Los métodos también son utilizados para estudiar relaciones de correlación entre las variables, generalmente variables socioeconómicas, como los índices de desarrollo social, el grado educativo de la población o la tasa de desempleo. En nuestro caso, correlacionamos estos datos socioeconómicos con la preferencia hacia cierto partido en orden de poder explicar los comportamientos electorales aprovechando su capacidad predictiva para próximas elecciones.

Los métodos de la estadística espacial permiten el tratamiento de datos geográficos. Goodchild (Goodchild, 1986) define cuatro tipos de objetos geográficos: los puntos, las líneas, las áreas y las retículas. Siendo los puntos datos geográficos dentro del área que se estudia que pueden representar localizaciones específicas como tiendas, personas, lugares etcétera. Las líneas, representan datos geográficos como ríos, carreteras, vías de tren entre otros. En las áreas, los datos son agregados y resumidos por zonas, las cuales pueden ser países, estados, secciones electorales, etcétera. Dichas zonas son de sumo interés para nuestra investigación debido a que mediante su uso podemos determinar la influencia territorial de la compra y coacción del voto y estimar en qué medida dichas prácticas inciden en el resultado de una elección. Finalmente, una retícula es una cuadrícula regular donde cada cuadrado dentro de la retícula se le conoce como celda o célula. Una retícula puede formarse de distintas

figuras geométricas como triángulos o hexágonos y no solo cuadrados. Las retículas de este tipo son conocidas como mosaicos regulares.

Actualmente muchos métodos se encuentran integrados en paquetes de software o se pueden programar utilizando lenguajes de programación o utilizar funciones ya programadas que permiten al investigador, ordenar, procesar, analizar y presentar los datos y resultados de una manera eficiente. Este capítulo aborda una variedad de métodos y técnicas utilizadas en este trabajo, el propósito del capítulo no es el de discutir de manera formal cada uno de los conceptos y métodos aquí utilizados, sino más bien el de proveer de una introducción suficiente para el entendimiento de estos métodos y conceptos en un contexto de aplicación en la investigación social en el campo de los estudios electorales.

2.1. Análisis exploratorio de Datos y Análisis Exploratorio de Datos Espaciales

Es difícil encontrar una definición universal para el Análisis Exploratorio de Datos (AED) y es común que se confunda con la estadística descriptiva, de modo que los investigadores desarrollan su propia definición de AED² desde las denotaciones de su nombre (Behrens, 1997). El AED se considera como un trabajo de detectives en palabras de John Tukey, pues el análisis exploratorio de datos es un trabajo sigiloso de conteo, representación gráfica y de análisis numérico. Agregando que un detective necesita herramientas y comprensión para resolver un crimen cuyas causas están ocultas en múltiples evidencias cuya relación no es explícita en la escena donde se cometió el delito (Tukey, 1997). Esto refleja el carácter del AED como un proceso de descubrimiento y organización de información dispersa y desvinculada mediante una serie de herramientas generalmente estadísticas y gráficas.

Cuando abordamos un fenómeno y queremos explorar los datos con herramientas estadísticas para observar que nos pueden contar, estamos haciendo un análisis exploratorio de datos (Andrienko & Andrienko, 2006). Es decir, tenemos principalmente una filosofía del análisis de datos en donde el investigador examina los datos sin ninguna idea preconcebida para descubrir como dichos datos pueden expresar algún indicio explicativo sobre los fenómenos que se están siendo estudiados (Martinez et al., 2005). Así pues, podemos considerar que el análisis exploratorio de datos es un proceso de investigación/exploración llevado a cabo por el investigador en orden de adquirir conocimiento desde el conjunto de datos sobre el fenómeno de

² El término en inglés es EDA de "Exploratory Data Analysis".

interés empleando una serie de herramientas estadísticas y gráficas. Por ello no es de extrañar que los investigadores realicen un AED como paso inicial en su investigación.

En lo que respecta al Análisis Exploratorio de Datos Espaciales (AEDE), es considerado como una subrama del AED que puede ser concebido como un súper-conjunto de técnicas con especial enfoque en la locación, distancia e interacción espacial (Anselin, 2012). Anselin define al AEDE como el conjunto de técnicas que describen y visualizan las distribuciones espaciales, identifican localizaciones atípicas o atípicos espaciales, descubren esquemas de asociación espacial, clústeres o hot spots y sugieren estructuras espaciales u otras formas de heterogeneidad espacial (Anselin, 1999). Es decir, el AEDE es aplicado a datos espaciales como puntos en el espacio o agregados por área y tiempo. Por lo que, tanto el AED como el AEDE, resultan ser elementos importantes en la investigación social y en la geografía electoral donde la mayoría de los datos tienen una relación con su espacio geográfico. Por lo tanto, ambas estrategias de análisis funcionan como paso predictivo en la investigación social, pero tienen que ser complementadas con otras estrategias de corte histórico o cualitativo.

2.2. Unidad espacial de análisis

Muchos de los datos y variables con las que se trabajan en el análisis de datos espaciales son medidos en relación con el área geográfica de estudio (por ejemplo: la densidad de población de cierto país, las tasas de crimen en cierto estado, los índices de desarrollo por alcaldía, etcétera.). En esta tesis la unidad espacial de análisis utilizada es la sección electoral. Las secciones electorales son las subdivisiones de los distritos electorales y es la unidad básica de la geografía electoral en México. En ella se lleva a cabo la inscripción de los ciudadanos tanto al padrón electoral como a la lista nominal la cual debe contar con un mínimo de 100 electores y un máximo de 3000 (INE, 2019).

Las áreas geográficas definidas y con las que se trabaja en su mayoría en la medición de variables y datos agregados son arbitrarias y no responden a criterios del análisis espacial (Olaya, 2019). De manera que el cambio entre una unidad de análisis y otra implica un problema conocido como “El problema de la Unidad de Área Modificable” el cual se reconoce en dos formas interrelacionadas; los efectos o problemas de escala y de zona. El primero, está relacionado con el tamaño de las unidades de área, surge al agregar el mismo conjunto de datos en distintas escalas haciendo que los resultados del análisis varíen para cada escala. El segundo, dado un conjunto de unidades de área

fijas, es decir, de tamaño fijo, al variar su forma también varían los resultados del análisis. Algunas de estas variaciones son válidas, pues responden a diferentes preguntas, sin embargo, es un aspecto importante a tomar en cuenta en la realización de investigaciones sociales en diferentes escalas geográficas.

2.3. Indicadores de competitividad electoral

2.3.1. Índice de Rae

Para el estudio de la competitividad electoral es importante considerar la fragmentación del poder, la “fuerza política” o el “peso” de cada partido en la contienda, pues el electorado³ puede estar repartido de diversas maneras ya sea equitativamente entre todos los partidos o una cantidad pequeña de ellos puede dominar la contienda dejando a los demás marginados. En ambos casos nos interesa conocer el reparto de los votantes a nivel espacial, para identificar las fuerzas políticas predominantes en la contienda electoral y en algún sitio en específico.

El índice de Rae proporciona una medida del reparto del poder, es decir, nos informa la medida en la que el electorado se halla repartido o concentrado. Su fórmula es:

$$IRAE = 1 - \sum_{i=1}^n S_i^2 \quad (2.1)$$

Donde n es el número de partidos⁴ en la contienda electoral y S_i es la proporción de votos obtenida por el partido i .

Notemos que el término $\sum_{i=1}^n S_i^2$ es en realidad el índice de Herfindahl-Hirschman (abreviado como IHH) utilizado para medir la concentración en varios contextos sobre todo económicos, pues es usado para medir la concentración de un mercado entre varias compañías (Neria, 2015).

El índice de Rae es fácil de interpretar, pues, valores cercanos a 1 indican una alta competitividad o baja concentración del electorado y valores cercanos a 0 indican una alta concentración o poca competitividad, siendo el 0 una nula competitividad. Para facilitar la interpretación y dado que el índice Rae está relacionado con el

³ El padrón electoral en México es la base de datos de todos aquellos ciudadanos que han solicitado su credencial para votar. Por su parte la lista nominal contiene a todos aquellos ciudadanos que ya cuentan con su credencial y que pueden emitir su voto.

⁴ Aunque aquí utilizamos la palabra partido no se limita a ello. Los candidatos independientes también pueden ser tomados en cuenta.

IHH utilizamos el siguiente cuadro que es una modificación de la escala mostrada en (Naldi & Flamini, 2014) para el IHH en un contexto económico.

Cuadro 2.1: Rango e interpretación para el índice de Rae

Rango	Interpretación
>0.85	Alta competencia-concentración
0.85-0.75	Media competencia-concentración
<0.75	Baja competencia-concentración

Fuente: Tomado de (Naldi & Flamini, 2014)

2.3.2. Índices de competitividad electoral para dos partidos

Dependiendo del sistema político-electoral que se esté analizando, puede darse el caso en el que dos partidos sean las principales fuerzas políticas (sino es que las únicas) en una contienda electoral dejando al resto de los partidos con una participación escasa. En un caso como este el índice de Rae mostrará concentración del poder, pero no nos informará cómo es la competitividad entre las principales fuerzas políticas.

El índice de competitividad electoral (ICE) es una medida de competitividad que solo toma en cuenta al primer y segundo lugar de la contienda. Se utiliza para medir que tan cerrada es la contienda entre los primeros lugares. En otras palabras, permite observar que tan fácil o difícil es para el partido ganador alcanzar o mantener el poder (Ibarra et al., 2009). El ICE es calculado al dividir el porcentaje de votos del segundo lugar entre el porcentaje de votos del primer lugar.

$$ICE = \frac{P_2}{P_1} \quad (2.2)$$

Donde P_1 , P_2 son los porcentajes de votos obtenidos por el primer y segundo lugar respectivamente. De nuevo el ICE toma valores entre 0 y 1, con valores cercanos a 1 indicando alta competitividad y cercanos a 0 baja competitividad.

Al igual que con el índice Rae utilizaremos una clasificación para su interpretación. La clasificación que aquí utilizamos es la propuesta por Ibarra en (Ibarra et al., 2009).

Cuadro 2.2: Rango e interpretación para el ICE

Interpretación	Rango
Competitividad escasa:	$ICE \leq 0.4$
Competitividad baja:	$0.4 < ICE \leq 0.55$
Competitividad media:	$0.55 < ICE \leq 0.70$
Competitividad alta	$0.70 < ICE \leq 0.85$
Competitividad muy alta	$0.85 < ICE \leq 1$

Fuente: Tomada de (Ibarra et al., 2009).

De manera que cuando la competitividad es escasa el segundo lugar obtenga menos del 50% de la votación obtenida por el ganador. Cuando es baja el segundo lugar obtiene una votación cercana o igual al 50%. En el caso de que sea media el segundo lugar obtiene más de la mitad de los votos del primero, pero menos o igual al 70%. Cuando el ICE es alto la votación del segundo alcanza más del 70% del primero y cuando es muy alta la cantidad de votos obtenidos por el segundo lugar está muy cerca de la votación del primero. Por lo que los votos del primero y segundo lugar tendrán una diferencia de votos pequeña. El ICE ha sido utilizado en estudios de varios procesos electorales (por ejemplo, véase: (Ibarra et al., 2009), (Neria, 2015)) en nuestro país para medir la competitividad.

Otra medida de competitividad es el margen de victoria que al igual que el ICE solo toma en cuenta al primer y segundo lugar de la contienda electoral. Esta medida es calculada tomando la diferencia de las proporciones de votos obtenidas por el primer y segundo lugar, $P_1 - P_2$.

2.4. Método de ponderación de área simple

En la investigación geográfica es común encontrarse con datos agregados en unidades de análisis diferentes a las de interés, por ejemplo, se puede estar interesado en conocer la población de cada estado de México, pero sólo contamos con la información de la población agregada por país. En decir, conocemos cuanta población hay en México más no en cada uno de sus estados. Para este tipo de problemas se han desarrollado diferentes métodos de desagregación espacial que buscan estimar las cantidades de la mejor manera. Por supuesto todas estas estrategias cometen errores debido a sus limitaciones y los supuestos que llevan a cabo (Hallisey et al., 2017). Entre estos tenemos el método de ponderación de área simple que es un procedimiento cartográfico que asume homogeneidad dentro del área geográfica. Se considera a este

método como el más simple de todos, y puede ser muy útil en zonas geográficas donde el supuesto de homogeneidad sea un supuesto realista.

El método de ponderación de área simple funciona de la siguiente manera. Consideremos un espacio geográfico dividido en n divisiones que no se traslapan entre ellas, estas unidades geográficas tendrán áreas diferentes que denotamos como A_i con $i = 1, 2, 3, \dots, n$. Ahora supongamos que estas áreas tienen una población⁵ correspondiente denotada por P_i con $i = 1, 2, 3, \dots, n$ y supongamos finalmente que el área geográfica además está dividida en m subdivisiones con áreas más pequeñas que cualquier A_i , y que tampoco se traslapan entre ellas. Las áreas de estas subdivisiones las denotamos como A'_j con $j = 1, 2, 3, \dots, m$ y A_{ij} el área compartida entre las unidades i y j . Note que si la unidad j está completamente dentro de una unidad i entonces $A_{ij} = A'_j$.

Para estimar la población de cada unidad espacial j por el método de ponderación de área simple, utilizamos la siguiente fórmula.

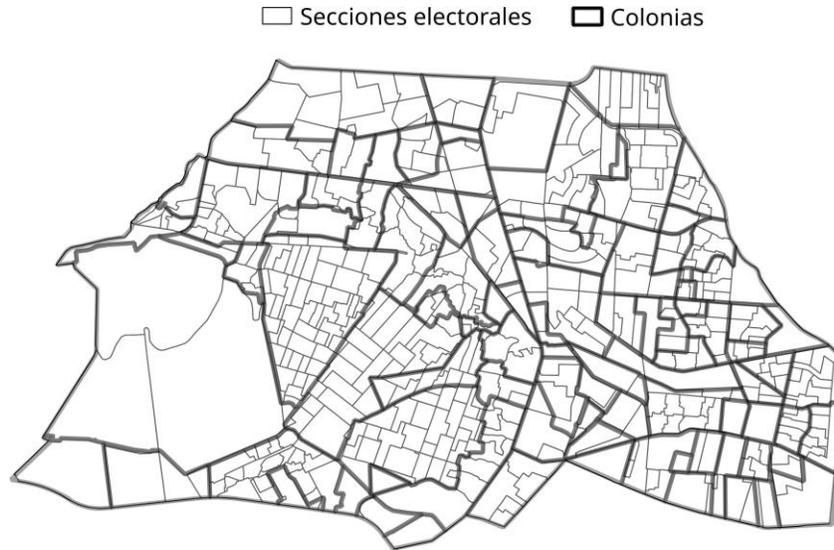
$$P_j = \sum_{i=1}^k \frac{A_{i,j}}{A_i} P_i \quad (2.3)$$

Donde P_j es la población estimada para la subdivisión j y k es el número de divisiones i que contiene una parte del área de la subdivisión j . Si $k = 1$ la suma es solo una proporción entre las áreas. La suma se lleva a cabo dado a que no todas las subdivisiones están completamente contenidas en una sola división sino por el contrario podría darse el caso de que se encuentren repartida entre dos o más divisiones. La figura 3.1 muestra el mapa de Coyoacán dividido por colonia y por secciones electorales, es claro que algunas secciones electorales no se encuentran totalmente dentro de alguna colonia.

Este método calcula la estimación al realizar una división entre el espacio geográfico y las áreas geográficas más pequeñas que la componen y luego multiplica la población por esta proporción. De manera que la población estimada es la suma de las poblaciones multiplicadas por su respectiva proporción. La ventaja que este método posee es su simplicidad; sin embargo, se debe ser cauteloso al usarlo pues el supuesto de homogeneidad no se cumple para todos los casos.

⁵ Aquí utilizamos población para ejemplificar, pero puede aplicarse a cualquier otra variable de interés que esta agregada por unidad geográfica.

Figura 2.1: Mapa de la demarcación de Coyoacán en la Ciudad de México con sus colonias y secciones electorales



2.5. Método de agrupación K-means

El algoritmo de Lloyd o k-means, es un algoritmo iterativo de agrupación muy popular que busca a partir de un conjunto de n observaciones (donde cada observación pertenece a \mathbb{R}^d con d un entero.) construir k clústeres de manera que todos los puntos se encuentren en uno de ellos minimizando el valor de una función, llamada función de potencial y denotada por φ . La función de potencial es la suma de los cuadrados de la distancia de cada punto y su centro más cercano y está dada por la expresión.

$$\varphi = \sum_{x \in X} \min_{c \in C} \|x - c\|^2 \quad (2.4)$$

Donde C es el conjunto de los k centros de cada clúster y $X \subseteq \mathbb{R}^d$ es el conjunto de las n observaciones.

El algoritmo puede resumirse en los siguientes 4 sencillos pasos (Vassilvitskii & Arthur, 2007):

1. Escoger arbitrariamente el conjunto de k centros, $C = \{c_1, c_1, \dots, c_k\}$.
2. Para cada $i \in \{1, 2, \dots, k\}$, sea el clúster C_i el subconjunto de observaciones de

X que están más cerca de c_i que de $c_j, \forall i \neq j$.

3. Para cada $i \in \{1, 2, \dots, k\}$, sea c_i el centro de masa de todos los puntos en C_i : $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$.
4. Repetir el paso 2 y 3 hasta que C no cambie más.

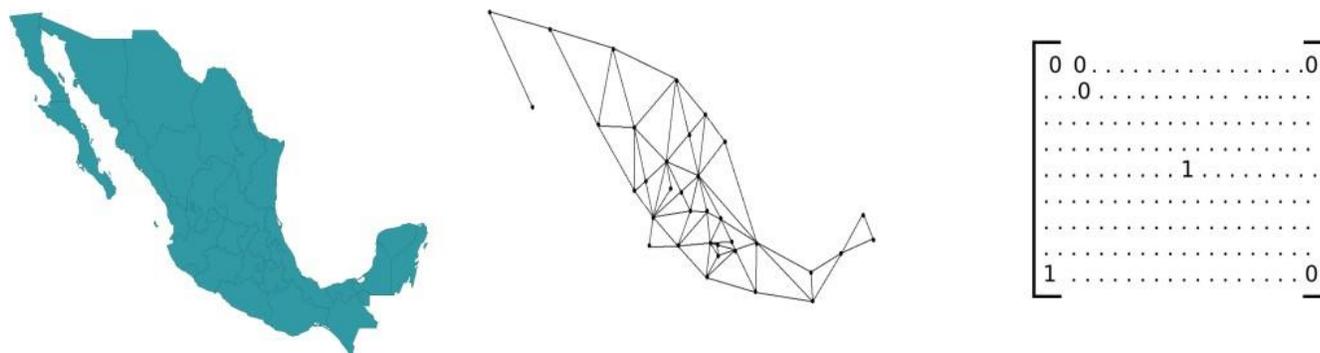
El primer paso se lleva a cabo de manera aleatoria, escogiendo de nuestras n observaciones los k clúster iniciales. El segundo paso crea los clústeres de cada iteración, calculando la distancia entre los centros y cada punto u observación de manera que cada uno de ellos se ubique en uno y sólo uno de estos clústeres. El paso tres actualiza los centros y por lo tanto el conjunto C por medio de la media de los puntos en cada clúster. Finalmente se repiten los pasos 2 y 3 hasta que este conjunto ya no tenga cambios. Puede advertirse que los centros finales de los clústeres no necesariamente pertenecen a nuestro conjunto de observaciones.

Una alternativa al algoritmo ordinario, es el algoritmo K-means++ el cual selecciona los centros iniciales de distinta manera; definiendo $D(x)$ como la distancia más corta de x al centro más cercano c_i que ya hayamos escogido. Así, primero selecciona el centro c_1 aleatoriamente de X , después selecciona los restantes $k-1$ c_i 's aleatoriamente, pero en base a su probabilidad dada por $\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$ y luego sigue con el procedimiento usual visto arriba. Una discusión más profunda y con mayor detalle entre los algoritmos k-means y k-means++ se puede consultar en (Vassilvitskii & Arthur, 2007). Estos algoritmos, se encuentran implementados en el lenguaje de programación Python (Pedregosa et al., 2011).

2.6. Matriz de contigüidad y vecindarios

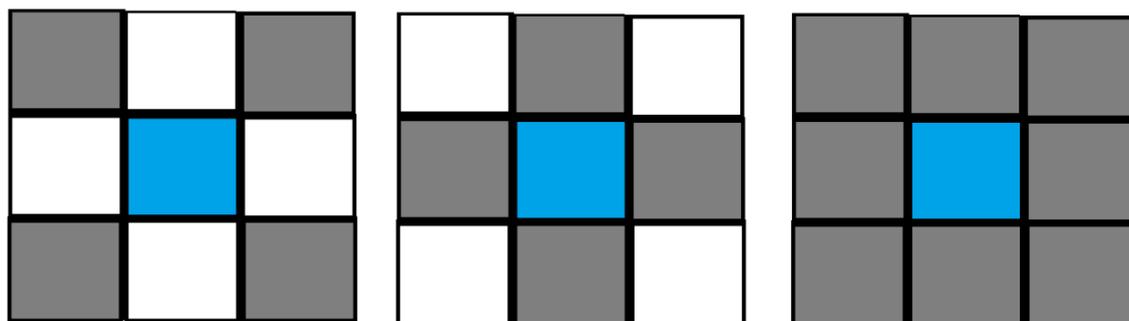
Las estadísticas usadas en la medición de la autocorrelación como el I de Moran o la C de Geary hacen uso de una matriz, llamada matriz de peso espacial o simplemente matriz de peso, para indicar la conexión entre las unidades espaciales. Estas comparten fronteras unas con otras (por ejemplo, la frontera entre países como México y Estados Unidos), indicando dicha conexión por medio de ceros y unos. Por supuesto esta matriz no es más que la matriz de adyacencia de la gráfica no dirigida que se obtiene de la conexión del conjunto de unidades espaciales. Es decir, cada renglón y columna representan una unidad espacial, los elementos de la matriz son 1 si una unidad está conectada con otra y 0 si no lo está.

Figura 2.2: La gráfica y la matriz representan de distinta manera los datos de conexión del mapa de México.



Podría ser el caso que una unidad espacial no comparta fronteras con otra, o que está sea solo un punto en alguna esquina, por ello se requiere definir de criterios de conexión entre estas unidades y sus vecinos más cercanos. Esto se logra considerando distintos criterios para vecindarios de las unidades espaciales. Los vecindarios más populares son los conocidos como los criterios del Alfil, Torre y Reina llamados así por la analogía con los movimientos que poseen las piezas del mismo nombre en el tablero de ajedrez. La figura 2.3 muestra estos tres tipos de vecindarios para la unidad central (en azul) y sus vecinos (en gris)

Figura 2.3: Ejemplos de vecindario tipo Alfil, Torre y Reina



Estos vecindarios tienen correspondencia con los vecindarios encontrados en la literatura relacionada con los autómatas celulares, aunque son nombrados de distinta manera. La matriz de pesos se simboliza con la letra W y cada una de sus entradas como w_{ij} . Así podemos asociarles un valor a las entradas de la matriz W en

concordancia con su tipo de vecindario. Siguiendo con la notación de (Smith, 2019): Para el caso de un vecindario tipo Reina, sea $N(i)$ el vecindario o la vecindad de la unidad espacial i , y sea j una unidad espacial arbitraria. Así $bnd(i)$ y $bnd(j)$ ⁶ son sus fronteras, entonces:

$$w_{i,j} = \begin{cases} 1 & \text{si } bnd(i) \cap bnd(j) \neq \emptyset \\ 0 & \text{si } bnd(i) \cap bnd(j) = \emptyset \end{cases} \quad (2.5)$$

Esta definición permite seleccionar unidades espaciales cuya frontera compartida está formada sólo por un punto.

Para el caso de un vecindario tipo Torre. Consideramos la longitud de la frontera compartida y la denotamos por $l_{i,j}$. Así:

$$w_{i,j} = \begin{cases} 1 & \text{si } bnd(i) \cap bnd(j) \neq \emptyset \text{ y } l_{i,j} > 0 \\ 0 & \text{si } bnd(i) \cap bnd(j) = \emptyset \end{cases} \quad (2.6)$$

Con esto sólo seleccionamos las unidades cuya frontera es mayor a cero. Sin considerar aquellas fronteras formadas sólo por un punto.

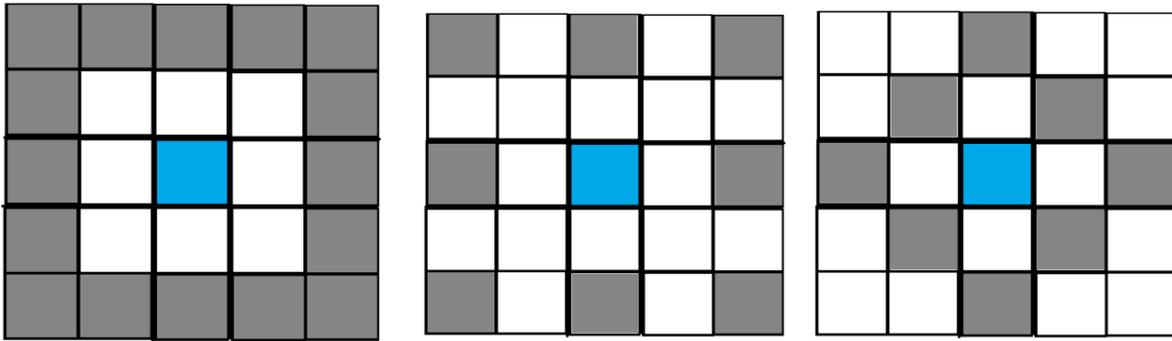
Para el vecindario de tipo Alfil tenemos:

$$w_{i,j} = \begin{cases} 1 & \text{si } bnd(i) \cap bnd(j) \neq \emptyset \text{ y } l_{i,j} = 0 \\ 0 & \text{si } bnd(i) \cap bnd(j) = \emptyset \end{cases} \quad (2.7)$$

Aparte de los distintos vecindarios, también se organiza a los vecinos dada su cercanía por medio de órdenes. De esta manera los vecinos más cercanos son considerados como vecinos de primer orden, los vecinos más cercanos diferentes a los primeros son vecinos de segundo orden y así consecutivamente. Teniendo la posibilidad de considerar vecinos de k -orden con $k \in \mathbb{N}$.

Figura 2.4: Ejemplos de vecindario tipo Reina, Alfil y Torre de segundo orden

⁶ El termino bnd es la abreviatura de la palabra en inglés "boundary" que significa frontera.



De esta manera la matriz W es llamada matriz de contigüidad de k - orden, en el caso de $k = 1$ sólo se refiere a ella como matriz de contigüidad. Aunque esta definición de la matriz W no es la única y de echo es la más sencilla. Suele ser la más mencionada y utilizada en la literatura espacial (véase: (Griffith, 2009), (Helbich et al., 2012), (Acevedo & Velásquez Ceballos, 2008), (Vilalta, 2008), (Darmofal, 2006)) junto con un vecindario tipo Reina.

Algunos autores además mencionan una variación de esta definición normalizando los renglones por medio de la siguiente fórmula:

$$w_{i,j} = \begin{cases} \frac{1}{\sum_k^n w_{i,k}} & \text{si } bnd(i) \cap bnd(j) \neq \emptyset \\ 0 & \text{c. o. c} \end{cases} \quad (2.8)$$

La suma de cada renglón de la matriz W normalizada es igual a 1. Aunque esta normalización no es estrictamente necesaria. Por último, como existe una gran variedad de formas de construir la matriz W , cabe mencionar que no existe un consenso ni una guía sobre qué criterio de construcción de la matriz W es mejor, siendo la matriz formada por 0's y 1's la opción más sencilla y utilizada.

2.7. Retardo espacial

La matriz de pesos W contiene la información de pertenencia de los vecinos de cada variable en alguna unidad espacial i dado el tipo de vecindario seleccionado. Si tomamos la suma de valores observados de cada vecino de esa unidad espacial obtendríamos el retardo espacial. Así por medio de la matriz W podemos calcular el retardo espacial como la suma ponderada de los valores de los vecinos de i . El retardo

espacial de la variable y_i utilizando una notación similar que en (Anselin, 2013), se expresa como:

$$y_{i,L} = \sum_{j=1}^n w_{i,j} y_j \quad (2.9)$$

Donde los valores $w_{i,j}$ son los valores de las entradas de la matriz W en el renglón i y columna j . Cuando la matriz de pesos es una matriz binaria con entradas 0's o 1's. El retardo espacial es la suma de los valores de los vecinos de i , como vimos antes, pues para los valores y_j que no pertenecen al vecindario de i tenemos que $w_{i,j} = 0$. En el caso de que la matriz de pesos sea normalizada, esto es $\sum_{j=1}^n w_{i,j} = 1$ se puede interpretar el retardo espacial como el promedio ponderado de los valores de los vecinos.

En notación vectorial, el retardo espacial se representa como:

$$y_L = W y \quad (2.10)$$

Donde W es la matriz de pesos de $n \times n$, y es el vector de observaciones de $n \times 1$ y y_L es el vector $n \times 1$ que contiene los retardos espaciales calculados para cada unidad espacial. El retardo espacial es importante como variable regresora en algunos modelos de regresión lineal como veremos más adelante.

2.8. Coeficientes de correlación

Recordemos que en estadística la covarianza es una medida de dependencia lineal entre dos variables aleatorias X y Y . Una covarianza positiva indica que la variable aleatoria Y crece cuando X también lo hace, una covarianza negativa indica que cuando Y disminuye X crece y por último una covarianza igual a cero indica que las variables aleatorias son no correlacionadas y no hay dependencia lineal.

La covarianza está dada por la siguiente expresión:

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] \quad (2.11)$$

Donde μ_x y μ_y son las medias de ambas variables aleatorias. Sin embargo, dado que la covarianza depende de la escala de medición es difícil determinar si una covarianza es grande o pequeña (Wackerly et al., 2008). Para resolver este problema se opta por el coeficiente de correlación de Pearson el cual elimina esta dependencia de la escala por medio de una normalización de la covarianza al dividirla entre el producto de las desviaciones estándar de ambas variables aleatorias eliminando la dependencia de la escala.

$$\rho = \frac{Cov(X,Y)}{\sigma_x \sigma_y} \quad (2.12)$$

Los estimadores de estas dos medidas son respectivamente:

$$s_{xy} = \sum_{i=1}^n (x - \bar{x})(y - \bar{y}) \quad (2.13)$$

$$r = \frac{n \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{n \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.14)$$

Donde r toma valores entre -1 y 1 siendo -1 y 1 una correlación perfecta⁷ y 0 indica que no hay correlación entre las variables. El coeficiente de correlación de Pearson realiza los siguientes supuestos; en primera los datos deben ser continuos, la relación entre las variables es lineal y la distribución conjunta de las variables es normal.

El coeficiente de correlación de Spearman es similar al coeficiente de correlación de Pearson, pero en vez de comparar los valores de las variables, el coeficiente de Spearman ranquea los datos y compara sus rangos. Es aplicable a datos continuos y ordinales, y a diferencia del coeficiente de Pearson, no requiere el supuesto de normalidad y la relación entre las variables sólo debe ser monótona⁸. En el caso de que no haya valores repetidos y por lo tanto tengan el mismo rango el coeficiente de Spearman está dado por la siguiente expresión:

$$r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)} \quad (2.15)$$

⁷ Para ver una demostración de este hecho vea (Casella & Berger, 2002).

⁸ Una relación monótona entre dos variables es aquella en la que una variable crece cuando la otra también lo hace o decrece cuando la otra crece. Sin importar si este cambio se da de manera lineal o no. Por ejemplo $f(x) = x^2$ y $f(x) = x$ en el intervalo $[0, \infty]$ son ambas funciones monótonas.

Donde n es el número de pares ranqueados y D_i la diferencia entre los rangos de un par ranqueado. Por otro lado, cuando se tiene valores repetidos se asigna el promedio de los rangos y se utiliza la siguiente expresión:

$$r_s = \frac{(n^3 - n) - 6 \sum D_i^2 - (T_x + T_y)/2}{\sqrt{(n^3 - n)^2 - (T_x + T_y)(n^3 - n) + T_x T_y}} \quad (2.16)$$

Aquí $T_x = \sum_{i=0}^g (t_i^3 - t_i)$, $T_y = \sum_{i=0}^g (t_i^3 - t_i)$ y g es el número de grupos de valores repetidos y t_i es el número de valores repetidos en el grupo. Nótese que si no se tienen valores repetidos la expresión (2.16) se convierte en la expresión (2.15).

En ambos casos tanto el coeficiente de correlación de Pearson como el de Spearman realizan una prueba de significancia, bajo la hipótesis nula de no correlación entre las variables. La hipótesis alternativa puede ser de dos o una cola, dependiendo si se busca una correlación positiva, negativa o simplemente alguna de las dos. Los lenguajes de programación como Python y R son capaces de calcular ambos estadísticos y devuelven tanto la medida de correlación como el p-valor. Finalmente, el cuadro 2.3 propuesta en (Hinkle, et al., 2002) nos permite interpretar la medida de correlación dada por ambos estadísticos en nuestro estudio.

Cuadro 2.3: Rangos e interpretación para la correlación

Interpretación	Rango
Muy alta	0.90 – 1.0
Alta	0.70 – 0.90
Moderada	0.50 – 0.70
Baja	0.30 – 0.50
Despreciable	0.0 – 0.30

Fuente: Tomado de (Hinkle, et al., 2002)

2.9. Autocorrelación espacial o dependencia espacial

El concepto de autocorrelación espacial es importante en el análisis de la presente tesis. Como ya se ha visto en las reflexiones sobre los coeficientes, el término correlación denota información redundante. Así si dos variables x e y están perfectamente correlacionadas entonces conociendo x podemos conocer y . Es decir, la información que, y nos pueda proveer, se encuentra ya en x (Griffith, 2009). El grado

de la información redundante varía como cambia la correlación entre estas variables. Así la autocorrelación espacial extiende este significado al espacio geográfico. La autocorrelación espacial o dependencia espacial es uno de los dos efectos espaciales⁹ y es entendida de manera general como el grado de asociación que tienen actividades, objetos o fenómenos en una posición geográfica con sus vecinas más cercanas (Goodchild, 1986). Esto surge al poner mayor importancia en una organización no aleatoria de estos objetos, fenómenos o actividades mediante la formación de patrones, clústeres o concentraciones. Una definición más precisa de autocorrelación espacial es la siguiente:

Autocorrelación espacial es la correlación entre valores de una sola variable estrictamente atribuible a sus posiciones relativamente cercanas sobre una superficie de dos dimensiones, introduciendo una desviación del supuesto de independencia de las observaciones de la estadística clásica. (Griffith, 2009).

La autocorrelación espacial denota dependencia entre las observaciones cercanas a lo largo de una superficie geográfica. Por lo que dichas observaciones pueden ser conocidas o predichas conociendo los valores de sus vecinos más próximos. Además, la autocorrelación espacial tiene importancia en el proceso de modelado, en primer lugar, porque provee de información que no está disponible en algún otro análisis estadístico (Goodchild, 1986); en segundo, porque al no considerar el análisis de la autocorrelación espacial en datos geográficos se acarrea severos problemas en, por ejemplo, el análisis de regresión. Así, la autocorrelación espacial implica dependencia y por lo tanto se estaría violando principalmente el supuesto de independencia de las observaciones (Vilalta, 2008). Vale decir que no parece extraño que se haya considerado a la autocorrelación espacial como un parámetro molesto que interfiere en la estimación de otros parámetros de interés del modelo (Griffith, 2009). No obstante, al mismo tiempo, la información sobre la distribución espacial que provee puede apoyar a una interpretación correcta del fenómeno bajo estudio. Es decir, la autocorrelación espacial funciona como una herramienta de diagnóstico y como un índice descriptivo de la medida en el que las cosas se encuentran distribuidas en el espacio geográfico además de medir el grado de influencia ejercido por algún fenómeno sobre su vecindario (Goodchild, 1986). Matemáticamente se denota la autocorrelación como:

$$Cov[y_i, y_j] = E[y_i y_j] - E[y_i]E[y_j] \neq 0 \quad (2.17)$$

⁹ El otro efecto es la heterogeneidad espacial que se define como la variación de las relaciones entre variables debido al cambio de una región a otra (Pace, 2009).

Para $i \neq j$. Los subíndices i, j denotan la localización de las observaciones, y es el valor de la variable aleatoria de interés (Autores, 2001), esta expresión precisamente denota la no independencia de las variables. Los términos de autocorrelación espacial y dependencia espacial se refieren al mismo fenómeno y su única diferencia es el significado conceptual que se les da, utilizando autocorrelación espacial para hacer referencia a las técnicas estadísticas y dependencia espacial más frecuentemente utilizada dentro de la econometría espacial para referirse a la explicación teórica (Vilalta, 2006).

2.10. Indicadores de asociación espaciales globales y locales

2.10.1. I de Moran

La I de Moran es un estadístico utilizado para medir la autocorrelación espacial global de una variable. Este estadístico fue presentado por el matemático australiano Patric Moran en su trabajo “Notes on Continuous Stochastic Phenomena” entre los años 1949 y 1950. La expresión matemática que da Moran en su trabajo es la siguiente:

$$r_{1,1} = \left(\frac{mn}{2mn-m-n} \right) \frac{\sum_{i=1}^m \sum_{j=1}^{n-1} z_{i,j} z_{i,j+1} + \sum_{i=1}^{m-1} \sum_{j=1}^n z_{i,j} z_{i+1,j}}{\sum_{i=1}^m \sum_{j=1}^n z_{i,j}^2} \quad (2.18)$$

Considerando una retícula de $m \times n$ y $x_{i,j}$ como observaciones del conjunto de variables aleatorias $X_{i,j}$ con índices enteros ($i = 1, 2, 3, \dots, m; j = 1, 2, 3, \dots, n$) y $z_{i,j} = x_{i,j} - \bar{x}$ como la desviación de la media. Aunque esta es la expresión original, en la actualidad la más popular es la que incluye la matriz W y una indexación distinta para cada unidad espacial ($i = 1, 2, 3, \dots, N, j = 1, 2, 3, \dots, N$):

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{i,j} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^N \sum_{j=1}^N w_{i,j})(\sum_{i=1}^N (x_i - \bar{x})^2)} \quad (2.19)$$

La I de Moran tiene un parecido con el coeficiente de correlación de Pearson. Como vemos a continuación:

$$r = \frac{n \sum_{i=1}^n (x_i - \bar{x})(x_j - \bar{x})}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})} \sqrt{n \sum_{i=1}^n (y_i - \bar{y})}}$$

Si sustituimos las y 's por x 's y en el numerador y solo sumamos los términos para los cuales las unidades espaciales i, j son vecinas cercanas indicado por el valor $w_{i,j}$ en la matriz W obtenemos:

$$r' = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{i,j} (x_i - \bar{x})(x_j - \bar{x})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})} \sqrt{\sum_{i=1}^n (x_i - \bar{x})}}$$

Finalmente, si dividimos por el número total de pares vecinos y multiplicamos por el número de unidades espaciales, tenemos el índice de Moran

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{i,j} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^n \sum_{j=1}^n w_{i,j}) (\sum_{i=1}^n (x_i - \bar{x})^2)}$$

Cabe preguntarse cuál es la función de la matriz W en expresión (2.19) y que la hace diferente de la expresión original (2.18). Para ver la diferencia entre ambas expresiones consideremos lo siguiente:

Si tenemos una retícula de $m \times n$ con $N = mn$ unidades espaciales entonces la fórmula (2.18) requiere de una matriz cuadrada de $N \times N$. Consideremos un vecindario tipo torre para la construcción de esta matriz. Así por la fórmula (2.3) tenemos que si la unidad espacial i y j comparten frontera y la longitud de está es mayor a cero entonces $w_{i,j} = 1$ y por tanto $w_{j,i} = 1$ para toda $i, j = 1, 2, 3, \dots, N$. Así:

$$\sum_{i=1}^N \sum_{j=1}^N z_i z_j = 2 \left(\sum_{i^*=1}^m \sum_{j^*=1}^{n-1} z_{i^*,j^*} z_{i^*,j^*+1} + \sum_{i^*=1}^{m-1} \sum_{j^*=1}^n z_{i^*,j^*} z_{i^*+1,j^*} \right)$$

Figura 2.5: Ejemplo de una retícula de 3x2 con las unidades espaciales indexadas por

$$i=1,2,3,\dots,6 \text{ e } i^*=1,2,3, j^*=1,2$$

1 1,1	2 1,2
3 2,1	4 2,2
5 3,1	6 3,2

Por lo tanto, el número de 1's en la matriz W es:

$$2(m(n-1) + (m-1)n) = 2(mn - m - mn - n) = 2(2mn - m - n)$$

y

$$\sum_{i=1}^N z_i^2 = \sum_{i^*=1}^m \sum_{j^*=1}^n z_{i^*,j^*}^2$$

Finalmente vemos que sustituyendo todo lo anterior en la ecuación 3.19

$$\begin{aligned} \frac{N \sum_{i=1}^N \sum_{j=1}^N z_i z_j}{(\sum_{i=1}^N \sum_{j=1}^N w_{i,j})(\sum_{i=1}^N z_i^2)} &= \frac{2mn(\sum_{i^*=1}^m \sum_{j^*=1}^{n-1} z_{i^*,j^*} z_{i^*,j^*+1} + \sum_{i^*=1}^{m-1} \sum_{j^*=1}^n z_{i^*,j^*} z_{i^*+1,j^*})}{(2(2mn - m - n)) \sum_{i^*=1}^m \sum_{j^*=1}^n z_{i^*,j^*}^2} \\ &= \frac{mn(\sum_{i^*=1}^m \sum_{j^*=1}^{n-1} z_{i^*,j^*} z_{i^*,j^*+1} + \sum_{i^*=1}^{m-1} \sum_{j^*=1}^n z_{i^*,j^*} z_{i^*+1,j^*})}{(2mn - m - n) \sum_{i^*=1}^m \sum_{j^*=1}^n z_{i^*,j^*}^2} \end{aligned}$$

Que es el índice dado por Moran de la ecuación (2.18). De aquí vemos que el índice (2.18) es un caso especial de la fórmula (2.19) con una matriz binaria construida con un vecindario tipo Torre y que la matriz W proporciona la estructura del vecindario definido.

La interpretación de la I de Moran es bastante directa. Si esta toma valores positivos indica agrupación de unidades espaciales: altos valores cerca de altos valores, bajos valores cerca de bajos valores. Si los valores altos están cerca de valores bajos el índice será negativo (Esri, 2019a) y será aproximadamente 0 cuando haya aleatoriedad

e independencia entre las unidades espaciales (Goodchild, 1986). Más precisamente valores mayores a $\frac{-1}{N-1}$ indican autocorrelación espacial positiva y menores autocorrelación espacial negativa.

El valor $\frac{-1}{N-1}$ es la esperanza de I bajo la hipótesis de aleatoriedad y un valor cercano a la esperanza indica aleatoriedad. Por supuesto sí $n \rightarrow \infty$ entonces $E[I] \rightarrow 0$ (Goodchild, 1986).

$$E[I] = \frac{-1}{N-1} \quad (2.20)$$

Como hemos mencionado si la I de Moran resulta positiva, el valor indica que existen clústeres de valores similares, si resulta negativo indica disociación de los valores de las observaciones y si resulta cero, o un valor muy “cercano” a cero, tenemos un patrón aleatorio y por lo tanto no hay autocorrelación espacial. Sin embargo, para obtener conclusiones válidas de nuestros datos y saber que los patrones no son producidos de manera aleatoria, se realiza una prueba de significancia bajo la hipótesis nula de no autocorrelación espacial. Existen dos enfoques para realizar una prueba de hipótesis de este tipo, el enfoque de muestreo y el de aleatoriedad.

El enfoque de muestreo supone que cada una de las observaciones fue seleccionada independientemente de una población de valores idénticamente distribuida y que se distribuyen de manera normal. Es decir, nuestros resultados son uno de todos los posibles muestreos. Por definición esta muestra no debe de revelar autocorrelación espacial pues cada una de las observaciones es seleccionada independiente de las otras. Este enfoque es conveniente sólo cuando los datos se distribuyen de manera normal. Por otro lado, la hipótesis de aleatoriedad supone que la muestra es una de todas las posibles $n!$ permutaciones de los valores observados en, por ejemplo, n unidades espaciales, seleccionada aleatoriamente.

El estadístico I tiene una distribución aproximadamente normal (Goodchild, 1986), por ello se estandariza como $Z = \frac{I-E[I]}{\sqrt{Var[I]}}$ para realizar la prueba de significancia. El valor esperado, bajo ambos enfoques y bajo la hipótesis nula es el mismo, más no así su varianza. El subíndice es utilizado para diferenciar ambos momentos con respecto al enfoque utilizado, N para muestreo y R para aleatoriedad.

$$E_N[I] = \frac{-1}{n-1} \quad (2.20)$$

$$E_R[I] = \frac{-1}{n-1} \quad (2.21)$$

$$Var_N[I] = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1) - (E_N[I])^2} \quad (2.22)$$

$$Var_R[I] = \frac{n(n^2 - 3n + 3) S_1 - n S_2 + 3 S_0^2}{(n-1)(n-2)(n-3) S_0^2} - \frac{b_2((n^2 - n) S_1 - 2n S_2 + 6 S_0^2)}{(n-1)(n-2)(n-3) S_0^2} - E_R[I] \quad (2.23)$$

Donde

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{i,j}$$

$$S_1 = \sum_{i=1}^n \sum_{j=1}^n \frac{(w_{i,j} - w_{j,i})}{2}$$

$$S_2 = \sum_{i=1}^n \left(\sum_{i=1}^n w_{i,j} + \sum_{j=1}^n w_{j,i} \right)^2$$

$$b_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{\left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

La hipótesis alternativa, puede ser de una cola en el caso de autocorrelación espacial positiva ($I > 0$) y autocorrelación negativa ($I < 0$) o de dos colas en caso simplemente de algún tipo de autocorrelación ($I \neq 0$).

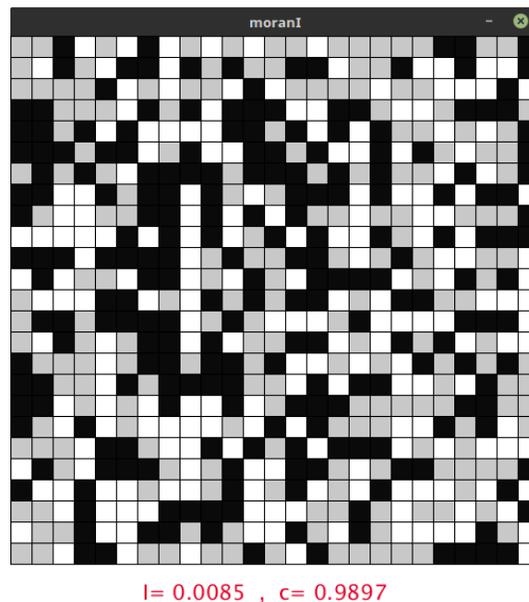
Hay que mencionar, además, la existencia de otros estadísticos para la autocorrelación espacial, como el de la c de Geary, que fue desarrollado por el estadístico irlandés Roy C. Geary y que a diferencia de la I de Moran toma valores entre 0 y 2 donde los valores entre 0 y 1 representan la presencia de clústeres mientras que valores mayores a 1

representan contrastes y un valor igual a 1 significa independencia y aleatoriedad. La c de Geary está también relacionada con la I de Moran (Goodchild, 1986), (Griffith, 2009) y ambos son considerados los índices más populares para estimar la autocorrelación espacial. La elección entre uno u otro depende del investigador ya que cada uno tiene ventajas y desventajas, aunque en general se le reconoce una leve superioridad estadística a la I de Moran (Moran, 1950). La c de Geary está dado por la siguiente expresión:

$$c = \frac{(N - 1) \sum_{i=1}^N \sum_{j=1}^N w_{i,j} (x_i - \bar{x})^2}{2 \sum_{i=1}^N \sum_{j=1}^N w_{i,j} \sum_{i=1}^N (x_i - \bar{x})}$$

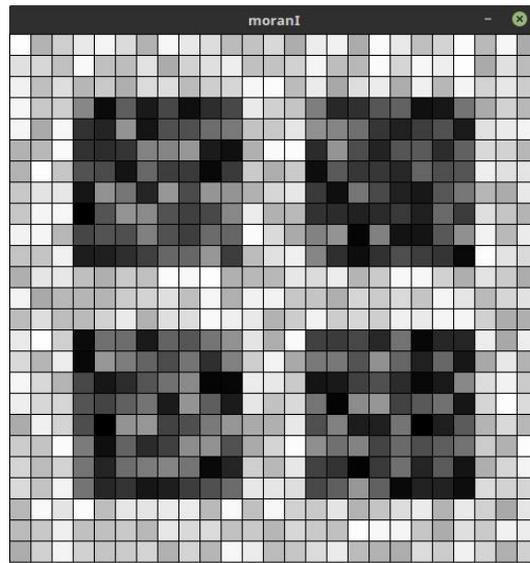
Las siguientes figuras muestran diferentes patrones en una cuadrícula de 25×25 ¹⁰. A cada configuración se le calcula ambos estadísticos. Podemos ver como los valores son diferentes para ambos casos, pero su interpretación es similar.

Figura 2.6: Vecindario tipo Torre con $I=0.0085$ y $c=0.9897$



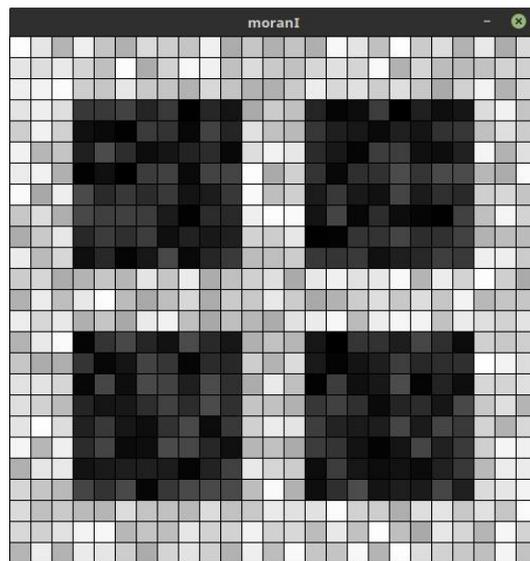
¹⁰ El código de este simulador se encuentra en los anexos de la tesis.

3.2: Vecindario tipo Torre con $I=0.6159$ y $c=0.3982$



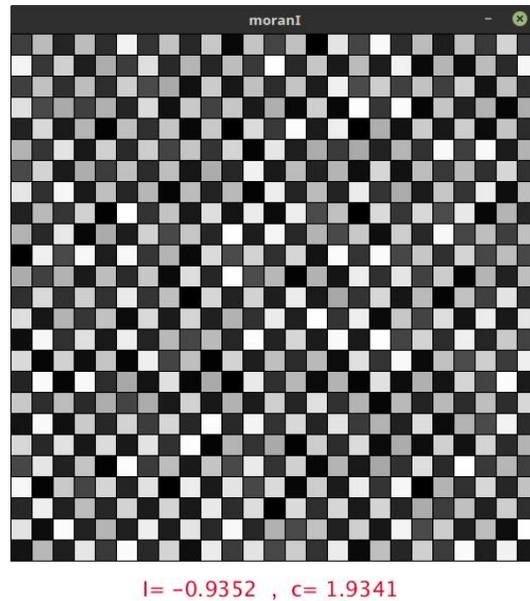
$I=0.6159$, $c=0.3982$

Figura 2.8: Vecindario tipo Torre con $I=0.7382$ y $c=0.2713$



$I=0.7382$, $c=0.2713$

Figura 2.9: Vecindario tipo Torre con $I = -0.9352$ y $c = 1.9341$



2.10.2. I de Moran local e indicadores de asociación espaciales

La I de Moran o la c de Geary son estadísticos de autocorrelación espacial globales que evalúan el grado de autorrelación del conjunto de datos completo, resumiendo los patrones espaciales en un solo estadístico; por lo tanto, no advierten de la localización de los clústeres (si es que éstos existen). En otras palabras, nos permiten saber si existen clústeres espaciales en nuestro conjunto de datos, pero no nos permite saber dónde están si es que estos clústeres en realidad fueron detectados.

A causa de esto Luc Anselin (Anselin, 1995) propuso un tipo de estadístico llamado indicador de asociación espacial local o LISA¹¹ como una manera de identificar clústeres (locaciones con valores altos rodeados de altos o bajos rodeados de bajos) y outliers espaciales (locaciones con valores altos rodeados de bajos o viceversa) al asociar los valores de una variable con los valores de sus locaciones cercanas de esa misma variable. La definición de LISA que propone Anselin es la siguiente:

¹¹ El indicador de asociación espacial es conocido como LISA por las siglas en inglés de “Local Indicator of Spatial Association”.

Definición 1. Un indicador de asociación espacial local es un estadístico que satisface las siguientes dos propiedades:

- El indicador para cada observación debe proveer de un indicio de la extensión del agrupamiento significativo de valores similares alrededor de esa observación.
- La suma de todos los indicadores en cada observación es proporcional a un indicador de asociación espacial global.

Siguiendo con la notación de (Anselin, 1995). Matemáticamente un indicador de asociación espacial se puede expresar como:

$$L_i = f(y_i, y_J)$$

Donde y_i es una variable observada en localización i , y_J son los valores observados en el vecindario J definido para y_i y f es una función de y_i y de y_J .

La segunda propiedad, relaciona a los indicadores de asociación espacial local con el indicador de asociación espacial global:

$$\sum_i^n L_i = \gamma \Lambda$$

Con γ un escalar y Λ es algún indicador de asociación global. La contraparte local de la I de Moran está dada por la expresión.

$$I_i = \frac{n(x_i - \bar{x})}{\sum_i^n (x_i - \bar{x})} \sum_j^n w_{i,j} (x_i - x_j) \quad (2.24)$$

Si tomamos la suma de los I de Moran locales de todas las observaciones podremos evidenciar que ésta, en efecto, es proporcional a su contraparte global. Haciendo:

$$\sum_i^n I_i = \sum_i^n \frac{n(x_i - \bar{x})}{\sum_i^n (x_i - \bar{x})^2} \sum_j^n w_{i,j} (x_i - x_j) = \frac{\sum_i^n \sum_j^n w_{i,j} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2}$$

Comparando con el I de Moran global de la expresión (2.19) obtenemos que el factor de proporcionalidad γ está dado por:

$$\gamma = \frac{\sum_i^n \sum_j^n w_{i,j}}{n}$$

por lo que la suma de los indicadores locales es igual al indicador global multiplicado por el factor γ .

La interpretación de la I de Moran local es similar al caso global, si $I_i > 0$ entonces la unidad espacial i está rodeada de valores similares altos o bajos por lo que forma parte de un clúster. Si $I_i < 0$ la unidad espacial está rodeada de unidades espaciales con valores disimiles por lo que la unidad i es un outlier espacial (Esri, 2019b). En este caso la significancia permite la identificación de outliers y clústeres espaciales y su visualización por medio del diagrama de dispersión de Moran y los mapas de significancia y de clusterización.

La terminología que se usa para ello, responde a cuatro casos los cuales son: Alto-Alto, Alto-Bajo, Bajo-Alto y Bajo-Bajo. El primer y último caso corresponde a clústeres y los dos restantes a outliers. Los estadísticos LISA pueden ser aplicados en numerosos campos de investigación como la economía, la geografía electoral, la demografía, medicina, etcétera.

2.11. Análisis de regresión, modelos espaciales y medidas de comparación

El análisis de regresión es una metodología estadística ampliamente utilizada en distintos campos de investigación, como la economía, las ciencias sociales, las ciencias biológicas o en la geografía electoral. Particularmente, se ha identificado tres principales propósitos para un análisis de regresión: describir, predecir y controlar. Este tipo de análisis fue, en un principio, desarrollado por Sir Francis Galton en el siglo XIX. Quien había estado estudiando la relación entre la altura de padres e hijos. Galton dio cuenta de que, aunque padres altos/bajos tenían hijos altos/bajos. Sin embargo, padres muy altos tendían a tener hijos más bajos que ellos y padres muy bajos tendían a tener hijos más altos que ellos. Es decir, la altura de sus hijos tendía a “regresar” a la altura media de la población. Galton llamó a este fenómeno “regresión hacia la media” (Neter et al., 2005), (Casella & Berger, 2002). El término de regresión sigue vigente hasta nuestros días.

2.11.1. Regresión lineal por mínimos cuadrados ordinarios

El modelo de regresión lineal considera la relación estadística lineal entre dos variables o entre una y varias variables. Cuando el modelo de regresión solo incluye dos variables se le conoce como modelo de regresión lineal simple y cuando involucra más de dos variables se le conoce como modelo de regresión lineal múltiple. El modelo lineal simple es un caso especial del modelo lineal múltiple y es establecido de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.26)$$

Y el modelo múltiple como

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (2.27)$$

Donde:

- Y_i es el valor de la variable dependiente
- $\beta_0, \beta_1, \dots, \beta_k$ son parámetros del modelo, conocidos como parámetros de regresión
- $X_i, X_{i,1}, X_{i,2}, \dots, X_{i,k}$ son los valores conocidos de las variables independientes o regresoras
- ε_i es el término de error aleatorio también conocido como perturbación

El modelo de regresión múltiple se puede expresar en forma vectorial como sigue:

$$Y = X\beta + \varepsilon \quad (2.28)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_k \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & \dots & X_{nk} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Donde:

- Y es un vector $n \times 1$ de las variables dependientes
- β es el vector $(k + 1) \times 1$ de los estimadores de regresión
- X es la matriz $n \times (k + 1)$ de los valores de las variables predictivas
- ε es el vector $n \times 1$ de los errores aleatorios

Para estos modelos se establece que cumplen con los siguientes supuestos o hipótesis básicas (Peña, 2012).

- Los errores o la perturbación tienen esperanza cero.

$$E[\varepsilon_i] = 0$$

- La varianza del error es constante o es homocedástico.

$$Var[\varepsilon_i] = \sigma^2$$

- El error se distribuye de manera normal.

$$\varepsilon_i \sim N(0, \sigma^2)$$

- Los errores ε_i son independientes entre sí.

$$E[\varepsilon_i \varepsilon_j] = 0 \quad \forall i \neq j$$

- El número de datos es mayor al número de estimadores
- Ninguna de las variables regresoras o dependientes son una combinación lineal de las restantes

Los últimos dos supuestos son sólo para el modelo de regresión múltiple. Note que el supuesto 3 implica que cada variable aleatoria y_i también se distribuye de manera normal y su varianza es la misma que la de los errores. Esto hace posible utilizar el método de estimación de máxima verosimilitud que se verá más adelante.

Como en un principio no conocemos los valores de los estimadores de regresión β_i , por ello necesitamos estimarlos a partir de los datos. Los parámetros β_i son estimados haciendo uso del método de mínimos cuadrados ordinarios. Este método considera la suma de los cuadrados de las desviaciones. Para el modelo lineal simple (2.26), expresamos esta suma como:

$$Q = \sum_{i=0}^n (Y_i - \beta_0 - \beta_1 X_1)^2 \quad (2.29)$$

Los estimadores de β_0 y β_1 son los valores b_0 y b_1 que minimizan Q . Estos estimadores se obtienen resolviendo el siguiente sistema de ecuaciones:

$$\frac{\partial Q}{\partial \beta_0} = 0$$

$$\frac{\partial Q}{\partial \beta_1} = 0$$

Si además, asumimos que $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ como se mencionó arriba, los estimadores pueden ser calculados por medio del método de máxima verosimilitud donde los estimadores de β_0 y β_1 maximizan la función de verosimilitud L .

$$L = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{(-\frac{1}{2\sigma^2}) \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_1)^2}$$

Bajo el supuesto de normalidad de los errores los estimadores obtenidos por el método de máxima verosimilitud son los mismos que los obtenidos por el método de mínimos cuadrados.

El caso de la regresión lineal múltiple es similar, aquí la función a minimizar es:

$$Q = \sum (y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_k X_{ik})^2$$

Al derivar esta ecuación con respecto a cada una de las β_i obtenemos el siguiente sistema de ecuaciones.

$$\begin{aligned} \sum y_i &= n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{i1}^2 + \dots + \hat{\beta}_k \sum X_{i1}X_{ik} \\ \sum y_i X_{i1} &= \hat{\beta}_0 \sum X_{i1} + \hat{\beta}_1 \sum X_{i1}^2 + \dots + \hat{\beta}_k \sum X_{i1}X_{ik} \\ &\vdots \\ &\vdots \\ &\vdots \\ \sum y_i X_{ik} &= \hat{\beta}_0 \sum X_{ik} + \hat{\beta}_1 \sum X_{i1}X_{ik} + \dots + \hat{\beta}_k \sum X_{ik}^2 \end{aligned}$$

Utilizando notación matricial podemos expresarlo como

$$\begin{bmatrix} 1 & \dots & \dots & 1 \\ X_{11} & \dots & \dots & X_{n1} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ X_{1k} & \dots & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & \dots & \dots & 1 \\ X_{11} & \dots & \dots & X_{n1} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ X_{1k} & \dots & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & \dots & \dots & \dots \\ \vdots & \ddots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

Que debido a los supuestos para la regresión lineal múltiple la matriz resultante del producto de matrices del lado derecho es no singular (Peña, 2012), es decir es invertible. Por lo tanto, los estimadores están dados por la siguiente expresión.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Al igual que otros trabajos como (Helbich et al., 2012) aquí utilizamos la prueba de Breusch-Pagan para contrastar la hipótesis nula de homocedasticidad en los errores.

Si el p -valor es significativo ($p < 0.5$) se rechaza la hipótesis nula. La prueba de Jarque-Bera, contrasta la hipótesis nula de normalidad en los errores y de nuevo si el p -valor es significativo se rechaza dicha hipótesis. La dependencia espacial entre los errores

es probada por medio de la I de Moran como ya se ha explicado con anterioridad. Dado que además podría haber multicolinealidad entre las variables regresoras se utiliza el factor de inflación de la varianza (VIF) para cada variable. En caso de que su valor no sobrepase el valor crítico de 5 rechazamos la multicolinealidad. Finalmente, los paquetes de software estadístico suelen devolver medidas de bondad de ajuste conocidas como R^2 y R^2_{adj} (que se leen como erre-cuadrada y erre-cuadrada ajustada). Estos estadísticos proporcionan una medida de la proporción de la variabilidad de la variable independiente que es explicada por el modelo (Miles, 2005).

La teoría sobre el análisis de regresión es muy amplia, pretender abarcar todo conllevaría a realizar todo un trabajo completo y por tanto se alejaría del propósito central de esta tesis. Por lo que solamente se ha mostrado lo “suficiente” para comprender su funcionamiento y aplicación en el caso de estudio. Así como para realizar la interpretación adecuada de los resultados que los paquetes de software normalmente devuelven. No obstante (Peña, 2012), (Yan & Gang Su, 2009) desarrollan con mayor profundidad esta teoría.

2.11.2. Modelo autorregresivo de retardo espacial

Se identifican dos tipos de dependencia espacial que pueden afectar los resultados obtenidos de la regresión lineal no espacial: la dependencia espacial “estorbo” y la “sustantiva”. La primera reduce la eficiencia del modelo y viola el supuesto de no autocorrelación entre los errores. Mientras que la segunda causada por la dependencia de la variable dependiente consigo misma en unidades cercanas, sesga los coeficientes estimados por la regresión de mínimos cuadrados (Matthews, 2006). El modelo de retardo espacial y el modelo de error espacial son alternativas que buscan corregir estos problemas.

Para comenzar, el modelo autorregresivo simple, incluye al retardo espacial de la variable Y como variable predictiva. En notación vectorial este modelo se especifica como sigue:

$$\mathbf{y} = \rho W \mathbf{y} + \boldsymbol{\varepsilon} \quad (2.30)$$

Donde ρ es el parámetro autorregresivo espacial. Este modelo es similar al utilizado en series de tiempo (de ahí su nombre) y expresa la idea de que el valor de una variable en una localización está relacionado con los valores de la misma variable en locaciones más cercanas (de Smith, et al., 2018). Si además agregamos las variables predictivas X el modelo resultante se le conoce como modelo mixto de retardo espacial o simplemente como modelo de retardo espacial.

$$\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \quad (2.31)$$

Trabajando esta expresión para que \mathbf{y} sólo aparezca del lado izquierdo de la ecuación.

$$\mathbf{y} - \rho\mathbf{W}\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}$$

$$(\mathbf{I} - \rho\mathbf{W})\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}$$

Así obtenemos

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon} \quad (2.32)$$

Cuando se tiene dependencia espacial sustantiva es común optar por esta última expresión, en la cual se agrega al modelo de regresión lineal el retardo espacial para incorporar la dependencia. Con esto se busca obtener un mejoramiento significativo con respecto a la regresión lineal por mínimos cuadrados y resolver uno de los problemas antes mencionados.

2.11.3. Modelo de error espacial

El modelo de error espacial es aplicado cuando se detecta dependencia espacial en los residuos de la regresión lineal y esta dependencia espacial no es sustantiva. Este tipo de dependencia se le conoce como “estorbo”. El modelo se especifica como sigue:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.33)$$

donde $\boldsymbol{\varepsilon} = \lambda\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u}$.

Los parámetros del modelo de retardo como el de error espacial son calculados por medio del método de máxima verosimilitud (Borrego, 2018). Para los modelos espaciales, suele utilizarse como medida de bondad de ajuste el $R^2_{Nagelkerke}$ (véase: (Helbich et al., 2012) y (Nagelkerke, 1991)) que es similar al R^2 y que se encuentra implementado en lenguaje de programación R.

Una vez que ya hayamos obtenido los resultados de la regresión lineal por mínimos cuadrados y con conocimiento de los modelos espaciales. Quisiéramos saber cuál de ambos modelos mejora los resultados de la regresión inicial y por lo tanto escoger dicho modelo. El software GeoDa y el lenguaje de programación R tienen

implementada la prueba del multiplicador de Lagrange (LM¹²) para este propósito. El LM es una prueba estadística utilizada para seleccionar el modelo de regresión espacial adecuado, una vez que hayamos obtenido los resultados de la regresión lineal por mínimos cuadrados (Anselin et al., 1996). Esta prueba es sencilla de interpretar, si el LM para el modelo de error o de retardo es significativo entonces la prueba sugiere que ese es el modelo adecuado, si ninguno de los modelos es significativo entonces no hay motivo para cambiar la regresión lineal inicial. En el caso de que la prueba sea significativa para ambos modelos se refiere a la prueba robusta del multiplicador de Lagrange y se realiza la misma interpretación.

Si el LM es significativo para alguno de los dos modelos entonces optaríamos por alguno de ellos teniendo al final, la regresión lineal inicial y el modelo espacial que hayamos escogido. Para saber cuál de ellos resulto ser mejor se utilizan dos criterios, el criterio de información de Akaike (AIC) o el de información bayesiano o de Schwarz (BIC). Estos criterios permiten realizar una selección del modelo y su principal diferencia es que el BIC penaliza más el número de parámetros. El AIC y el BIC están dados por las siguientes fórmulas de acuerdo a (Fabozzi et al., 2014).

$$AIC = -2 \log(L(\hat{\theta})) + 2k \quad (2.34)$$

$$BIC = -2 \log(L(\hat{\theta})) + k \log(n) \quad (2.35)$$

Donde

- θ : Es el vector de parámetros del modelo
- $L(\hat{\theta})$: Es la verosimilitud evaluada en los parámetros estimados del modelo candidato dado los datos
- k : Es el número de parámetros estimados del modelo candidato
- n : Es el número de observaciones

El AIC y el BIC¹³ son calculados para todos los modelos candidatos realizando la selección en base a cuál de ellos obtuvo el menor valor siendo este el “mejor” modelo

¹² Por las siglas en inglés de “Lagrange Multiplier”.

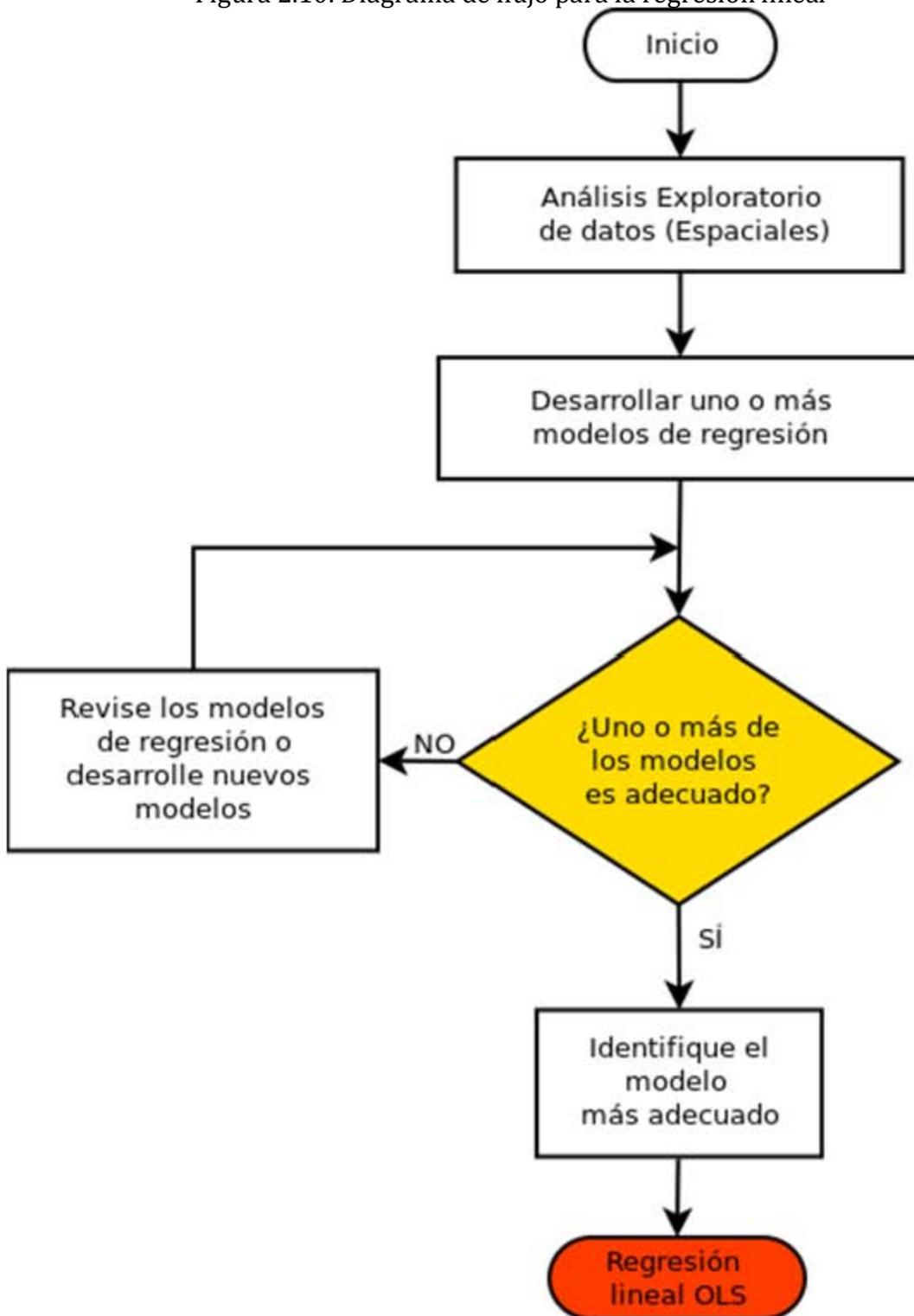
¹³ AIC y BIC por las siglas en inglés de “Akaike information criterion” y “-bayesian information criterion”.

(Fabozzi et al., 2014). Estos son los criterios que, por ejemplo, GeoDa calcula para comparar los modelos de regresión y se encuentran implementados en R.

A manera de resumen y para terminar este capítulo las figuras 2.10, 2.11¹⁴ muestran una serie de pasos para trabajar primero con la regresión lineal OLS que incluye el análisis exploratorio de datos (espaciales) y posteriormente una vez se cuente con el modelo OLS se procede a examinar si alguno de los modelos espaciales (de retardo y error) es necesario o si se mantiene la regresión original.

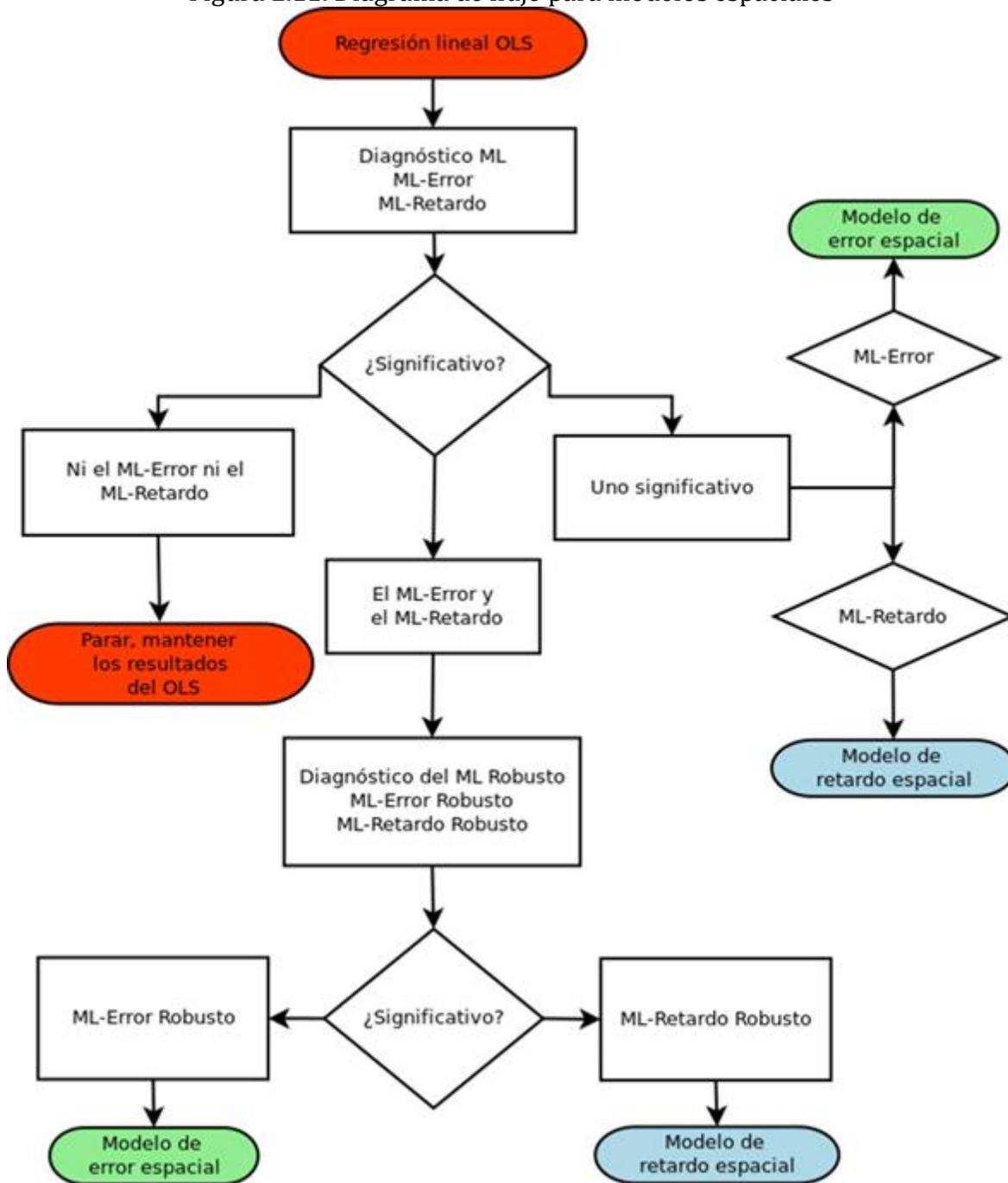
¹⁴ Las figuras 2.10 y 2.11 son recuperados de (Neter et al., 2005) y (Anselin, 2005) respectivamente con algunas modificaciones.

Figura 2.10: Diagrama de flujo para la regresión lineal



Fuente: (Neter et al.,2005).

Figura 2.11: Diagrama de flujo para modelos espaciales



Fuente: (Anselin, 2005).

Capítulo 3

Los usos electorales de la pobreza: desarrollo social, programas de atención a la pobreza y disputas electorales.

3.1. Clientelismo político y prácticas clientelares en México

Para comenzar este capítulo expondré la definición del término “clientelismo político”. Es necesario recalcar que no existe una definición universal para el clientelismo, conceptualizándose de diferente manera de acuerdo a factores como la relación entre los actores participantes y la disciplina que lo estudia. Como este trabajo está interesado principalmente en las acciones de la compra de voto y el uso de los programas sociales en elecciones locales, suscribimos una definición que abarque estas acciones, sin indagar demasiado en la interacción entre los personajes que participan en ellas. Así pues, entenderemos al clientelismo político como todas aquellas acciones de compra y coacción del voto como lo son el ofrecimiento de servicios, amenazas o condicionamiento de apoyos (como la permanencia en tal o cual programa social) e incluso actos de violencia. Todo esto en orden de obtener el voto o beneficiarse en las votaciones.

Generalmente en la literatura especializada se distingue una estructura general del clientelismo político. En ella se describe a los principales participantes de estas actividades; el patrón o líder, el cliente y el intermediario o mediador. El patrón es aquel que posee los medios para satisfacer las necesidades del cliente y que a cambio pide el voto de éste. Los recursos con los que cuenta el patrón, como bien apunta (Molina, 2014), pueden provenir de distintas fuentes ya sean públicas o privadas. Por esta razón, los patrones pueden ser políticos, líderes sociales, empresarios etcétera. En contraparte, el cliente es la persona o grupo de personas cuyas necesidades están insatisfechas y perciben en el patrón una forma de satisfacerlas. Es común identificar al cliente con aquellas personas en situación de pobreza quizá debido a las necesidades económicas que éstos poseen. Finalmente se tiene al mediador, quien

funciona como puente entre el patrón y el cliente, haciendo llegar el flujo de recursos desde el primero hacia el segundo y gestionando el apoyo del último hacia el primero (Schroter, 2010).

El clientelismo político en México ha sido estudiado en diferentes artículos académicos (Hevia de la Jara, 2010), (Schroter, 2010), (Tosoni, 2007), (Muñoz, 2006), (Baqueiro, 2014), (Molina, 2014)), los cuales muestran que dichas actividades clientelares no son ajenas a nuestro país; de manera particular, estos trabajos discuten sus causas y consecuencias, las interacciones entre patrones y clientes así como las distintas categorías de clientelismo. Además, estos escritos señalan que los actos de clientelismo político en zonas urbanas están enfocados a sectores con altos niveles de competitividad electoral en zonas donde predomina población pobre. Ya que dentro de entornos competitivos y con población con necesidades el clientelismo político y sus estructuras toman una mayor importancia, sobre todo cuando se le unen altos niveles de abstencionismo (Molina, 2014). Por ejemplo, Tosoni (Tosoni, 2007) menciona que a pesar de los avances democráticos en nuestro país después de la era del PRI, el clientelismo político no ha desaparecido y perdura en estos sectores urbanos de alta competitividad. Mientras que Molina discute que es en la pobreza donde el clientelismo político encuentra una base sólida para su desarrollo, suscribiendo junto con otros autores en que los pobres parecen tener mayor inclinación a este tipo de actos (Molina, 2014).

El uso del gasto público enfocado a la atención de la pobreza es analizado centrándose en los programas sociales destinado a estos rubros. Ejemplo de ello son los trabajos sobre el uso del programa “Progresas/Oportunidades” ((Hevia de la Jara, 2010), (Baqueiro, 2014)). Ahora bien, muchos de estos escritos son abordados desde una perspectiva cualitativa, pero existen otros como es el caso de (Vilalta, 2007) que realiza pruebas estadísticas a datos obtenidos por medio de encuestas sobre los programas sociales, incluyendo ya no solo una perspectiva cualitativa del problema; sino que además se intenta responder a las preguntas de si el contexto local determina y afecta la compra y coacción del voto y cómo lo hace. Hallando diferencias en la forma en la que la población beneficiaria y no beneficiaria reportan los actos de clientelismo en su estado.

Tomando como base la literatura revisada y aplicándola a nuestro caso de estudio podemos identificar secciones electorales de Coyoacán con características similares a los trabajos que han explicado los usos electorales del gasto social. Con ello buscamos explorar si en esta demarcación existen territorios como los que se describen en la literatura y que se identifican con una mayor presencia de actos de clientelismo

político. Así consideraremos una sección electoral como “vulnerable al clientelismo” si cuenta con las siguientes características.

- Índice de desarrollo social bajo o muy bajo
- En el territorio existe una gran cantidad de población votante.
- Indicadores de competitividad electoral altos o muy altos

Las primeras dos características son revisadas en este capítulo mientras que la competitividad se analiza en el siguiente.

3.2. Clientelismo político en Coyoacán 2018

Según las numerosas notas periodísticas del año 2018 (véase: (Bolaños, 2018), (Ruiz, 2018), (Hernández, 2018), (Ruiz & Animal Político, 2018)) que cubren desde testimonios de la población, en torno a si habían sido testigos de actos de clientelismo político durante el proceso electoral, hasta simples acusaciones entre los contendientes durante las campañas, reflejan los primeros indicios de que la población en Coyoacán estuvo expuesta a actos de clientelismo político como la entrega de tabletas, tinacos, sumas de dinero y el condicionamiento de la permanencia o ingreso en programas sociales, todo esto para obtener su voto.

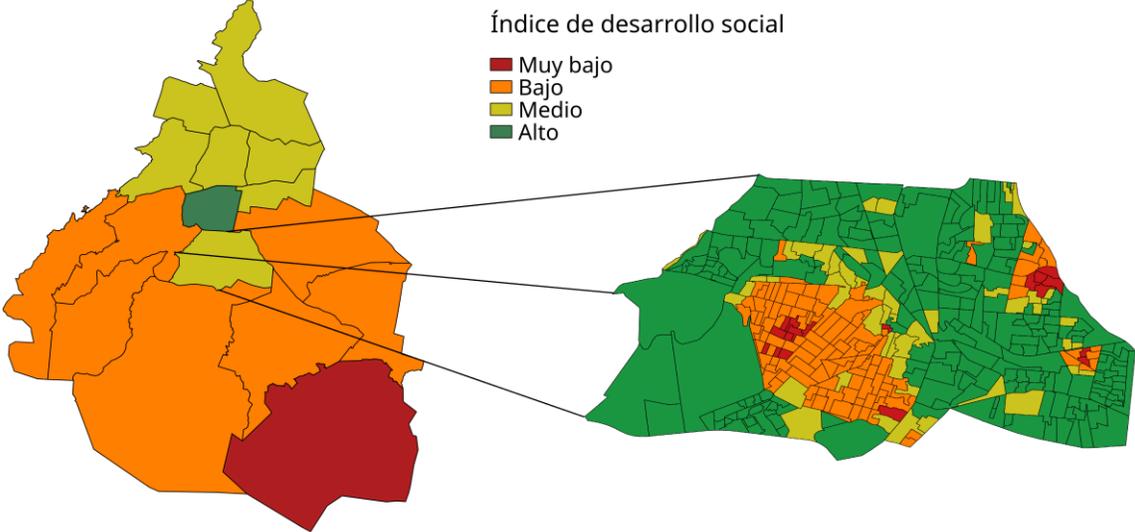
Las anécdotas plasmadas en estos trabajos, describen no sólo los tipos de apoyos que les fueron ofrecidos, sino también las formas en las que se operaba. Los programas sociales “Tu Unidad sin Goteras”, “A tu lado”, “Entrega de Tabletillas Electrónicas” y “Entrega de Juguetes a niños y a niñas” se siguieron ejerciendo durante el proceso electoral, lo anterior a pesar de la suspensión ordenada por el Instituto electoral de la Ciudad de México (IECM), incluso algunas notas denuncian la construcción de una red clientelar para el uso electoral de los programas sociales (Ruiz, 2018).

El condicionamiento de los programas sociales y la entrega de recursos en Coyoacán es lo que reportan en general las notas periodísticas sobre el tema. Con datos de la encuesta realizada por la Universidad Nacional Autónoma de México podemos obtener una idea cuantificable de lo reportado en trabajos periodísticos¹⁵. En primer lugar, Coyoacán es una de las demarcaciones con mayor índice de desarrollo social de

¹⁵ Esta encuesta fue realizada en algunas de las secciones electorales dentro de lo que llamamos como “Los Pedregales” y no en la completitud de Coyoacán. Para saber más sobre la metodología de la encuesta véase: (Varela & Ascensio, 2018).

la Ciudad de México, sin embargo, dentro de ella existen diferencias marcadas entre una zona y otra. La figura 3.1 muestra las demarcaciones de la Ciudad de México y la demarcación de Coyoacán por el índice de desarrollo social. El nivel de desarrollo no es uniforme dentro de la demarcación que, a pesar de tener un nivel de desarrollo medio, cuenta con zonas con un desarrollo bajo y muy bajo.

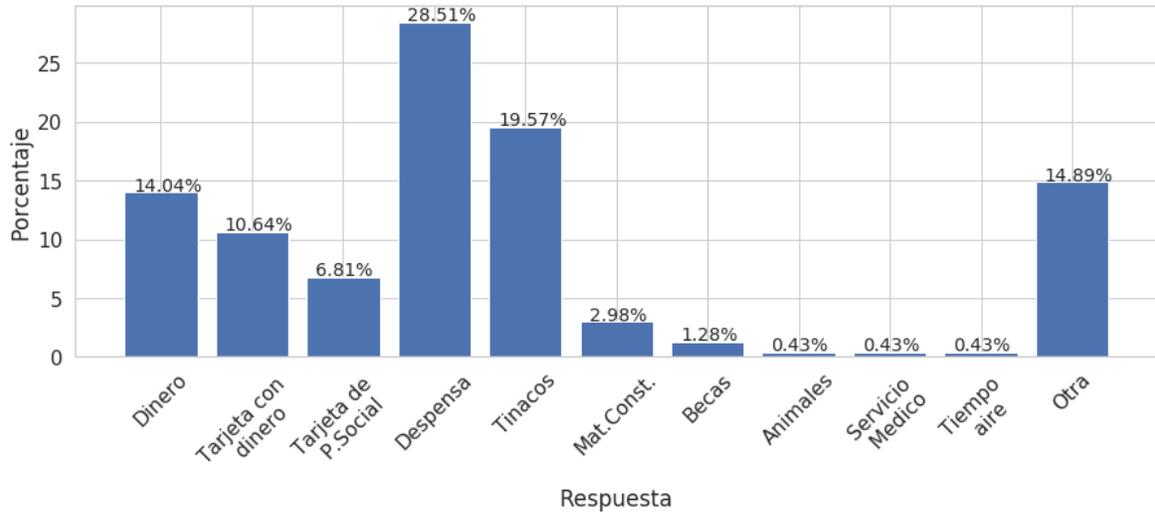
Figura 3.1: Índice de desarrollo social de cada demarcación en la Ciudad de México y de Coyoacán



Fuente: Elaboración propia en base a información de Evalúa CDMX.

De los datos de la encuesta realizada por la FCPyS cerca del 30% de los entrevistados afirma haber recibido el ofrecimiento de algún apoyo. Siendo la despensa el apoyo que más se ofreció, seguido de tinacos y el ofrecimiento de dinero. También se registra el ofrecimiento de material de construcción, tarjetas de programas sociales y becas como se ve en la figura 3.2.

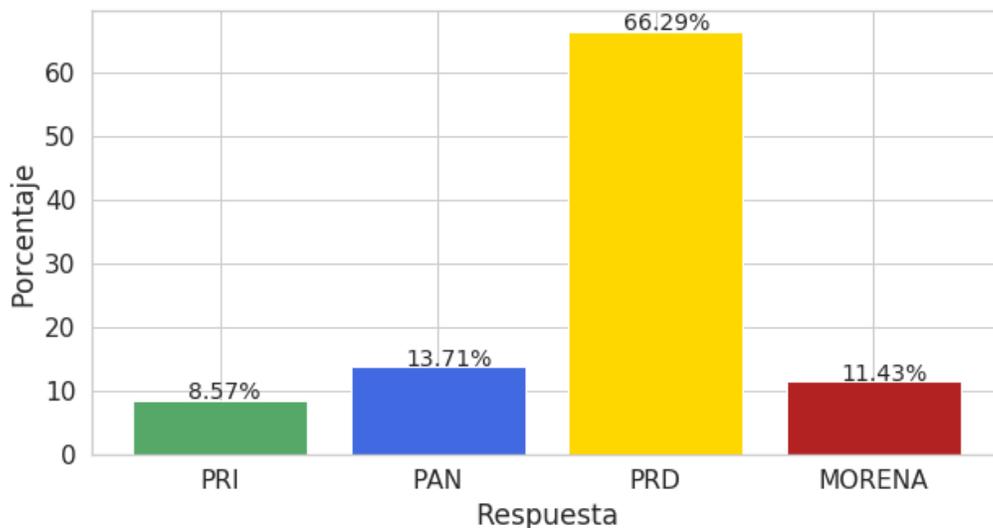
Figura 3.2: Tipo de apoyos ofrecidos



Fuente: Elaboración propia en base a información de (Varela & Ascensio, 2018).

De los 10 partidos políticos que participaron en la contienda electoral sólo 4 fueron señalados de ofrecer algún tipo de apoyo. Dentro de ellos el que más destaca es sin lugar a dudas el Partido de la Revolución Democrática (PRD), siendo el que más fue mencionado en la encuesta y en los reportes periodísticos. El partido Movimiento Regeneración Nacional (MORENA), el Partido Acción Nacional (PAN) y el Partido de la Revolución Institucional (PRI) también fueron señalados, aunque en menor medida.

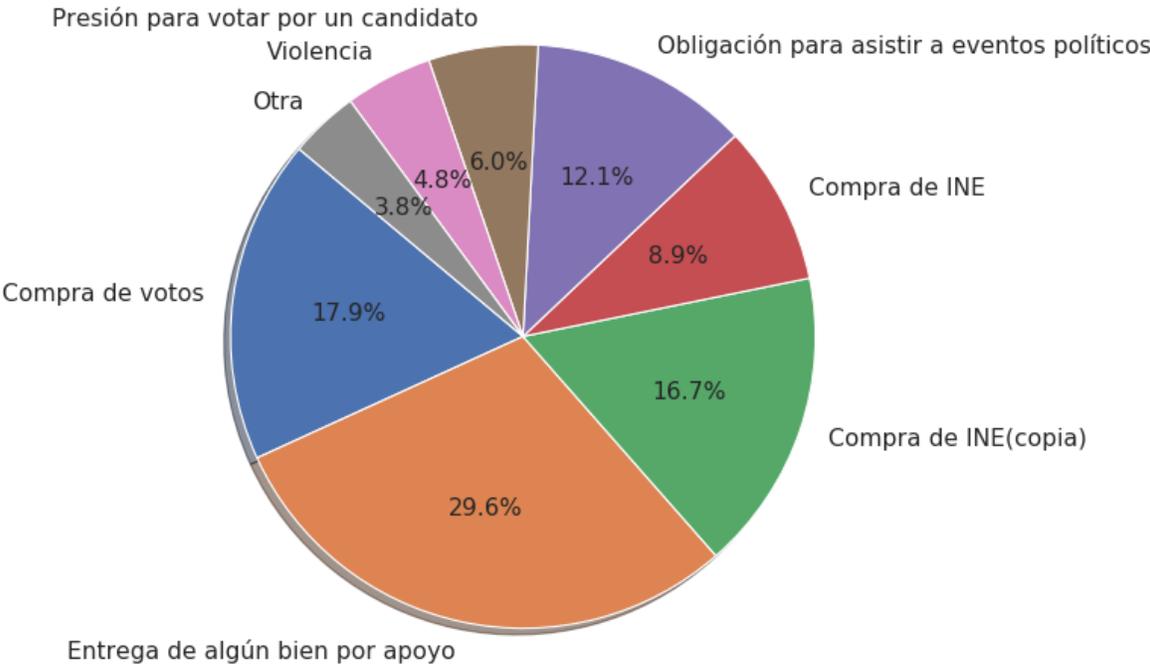
Figura 3.3: Partidos que fueron vistos cometiendo actos de clientelismo político



Fuente: Elaboración propia en base a información de (Varela & Ascensio, 2018).

Entre los actos de clientelismo que se visualizaron en la demarcación, los datos arrojan que la entrega de recursos o bienes fue el acto que más se pudo observar. Seguido de la compra de voto y la compra de la copia del INE. El gráfico muestra las opciones que los encuestados podían seleccionar y sus porcentajes. Es evidente la variedad de actividades clientelares que se reportan, así como el registro de actos de violencia.

Figura 3.4: Actos de clientelismo político vistos en Coyoacán



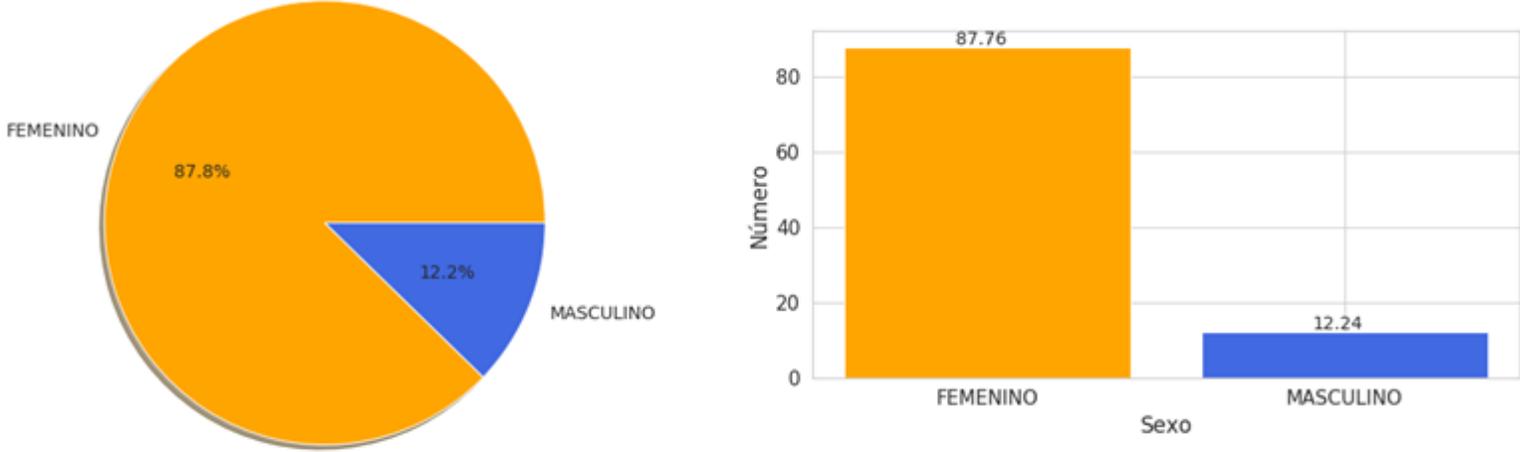
Fuente: Elaboración propia en base a información de (Varela & Ascensio, 2018).

Por otra parte, los padrones de beneficiarios de los programas sociales, nos permiten saber en dónde se ubican los beneficiarios y cómo se les beneficia. En orden de encontrar alguna relación entre el número de beneficiarios y las zonas de apoyo partidista. Se examinó el padrón de beneficiarios del programa de transferencias unitarias “A tu lado” para el 2017, el cual está enfocado a personas de escasos recursos que vivan en la demarcación de Coyoacán y que sean mayores de edad, según la gaceta oficial de la Ciudad de México del 2019 (Gobierno de la Ciudad de México, 2019b). Según el padrón de este programa para finales del 2017 se contaba con 13613 beneficiarios que representó el 2.4% del padrón electoral del 2018 y con un presupuesto de 54,996,520.00 de pesos. A pesar de ello en la gaceta oficial de la

Ciudad de México y acorde a lo que se menciona en (Ruiz, 2018), la capacidad de 13613 beneficiarios fue modificada para el año 2018 con un presupuesto de 109,993,040.00 de pesos y con una capacidad máxima de 27226 beneficiarios que representaría el 4.81% del padrón electoral de ese año.

Con un aumento así y en año electoral, se podría dar no sólo el condicionamiento de permanencia en el programa social sino además la oferta de inclusión a éste. Según datos oficiales (Gobierno de la Ciudad de México, 2019a), el programa “A tu lado” está dirigido a la población de Coyoacán en situación de pobreza o en proceso de empobrecimiento y se les apoya por medio de una tarjeta electrónica donde se les proporciona un apoyo de 4040 pesos. La gaceta además indica que el programa busca beneficiar indirectamente a 81,678 personas. Dado que quienes reciben la tarjeta electrónica son identificados dentro de un núcleo familiar. Sin embargo, en los archivos públicos del Consejo de desarrollo social de la Ciudad de México y a la fecha en que esto se escribe (año 2019), no se encuentran datos sobre la publicación del padrón de beneficiarios del programa para el año del 2018. Además, el padrón de beneficiarios del 2017 muestra que, aunque en la gaceta dice en sus reglas de operación que el programa social se ejecutará con una perspectiva de equidad de género, el 87.8% de los beneficiarios son mujeres de entre 18 y 68 años y sólo el 12.2% son hombres en el mismo rango de edades.

Figura 3.5: Beneficiarios del programa social “A tu lado” por sexo



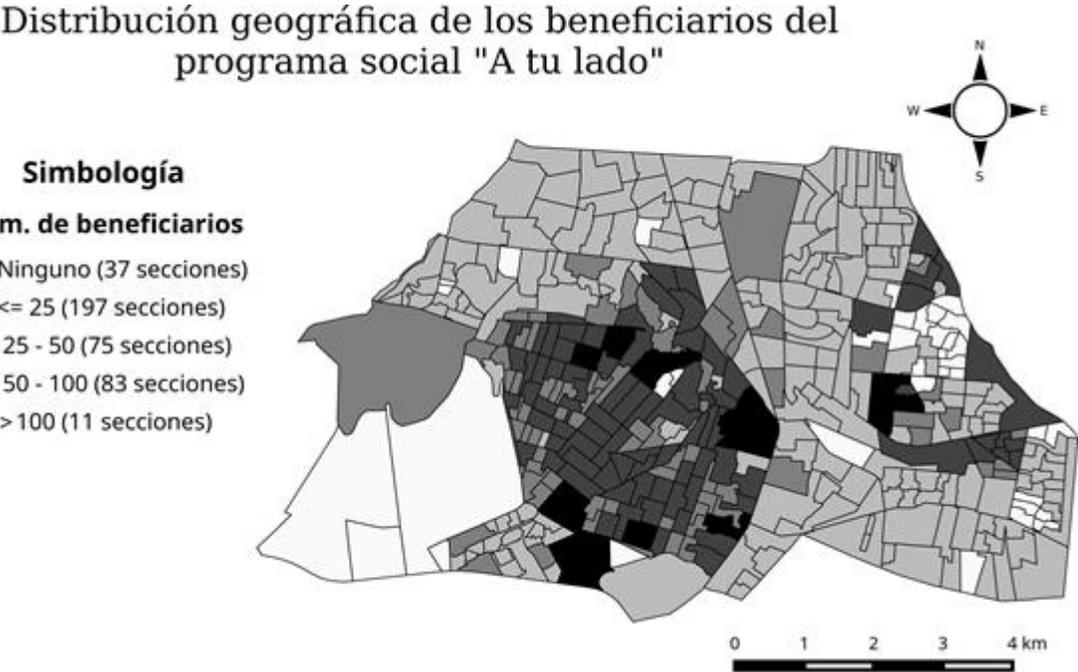
Fuente: Elaboración propia en base a información del SIDESO.

La encuesta proporciona una mirada a las actividades clientelares en Coyoacán, y sigue la línea que traza la literatura especializada. Entre los actos de clientelismo

político encontramos actos de compra de voto y del uso de programas sociales. Para este último, el padrón revela una desigualdad muy marcada entre hombres y mujeres, siendo casi el 90% del padrón de beneficiarios conformado por mujeres contradiciendo lo especificado en sus reglas de operación.

Una aproximación de la distribución geográfica de los beneficiarios en la alcaldía se obtiene aplicando el método de desagregación de ponderación de área simple explicado en el capítulo 2 a las colonias y secciones electorales. La figura 3.6 muestra el mapa con la aproximación obtenida. Es claro que los beneficiarios no se distribuyen de manera aleatoria y uniforme a través de toda la demarcación. Pues se tiene colonias con ningún beneficiario y un tanto más con un gran número de ellos¹⁶.

Figura 3.6: Beneficiarios del programa social “A tu lado” por sección electoral



Fuente: Elaboración propia en base a información del SIDESO.

3.3. Espacios vulnerables, marginados y votantes

La construcción de las secciones electorales se realiza de forma que el padrón de beneficiarios sea más o menos el mismo para todas las secciones. El análisis de datos

¹⁶ Hay que mencionar que no fue posible ubicar algunos beneficiarios que estaban dentro del padrón. Por lo que hay una diferencia entre el número total estimado y el número de beneficiarios en el padrón.

refleja que existen secciones con mayor población votante y que por lo tanto representan un mayor peso electoral a nivel local. Stokes y otros autores realizan una clasificación de distritos electorales en base a sus características formando categorías no excluyentes muy útiles. Por ejemplo, ellos llaman “distrito pivote” a aquellos distritos que pueden hacer ganar o perder el control del gobierno a un partido político y “distritos marginales” a aquellos donde la diferencia de votos es pequeña entre los primeros lugares. Además, Molina (Molina, 2014) menciona que la categorización de distritos permite al patrón seleccionar que distritos se van a beneficiar, en otras palabras, que distritos van a presentar actos de compra y coacción del voto.

Siguiendo con lo anterior, Coyoacán es una alcaldía que no es homogénea al hablar del índice de desarrollo social para cada sección electoral. Ni tampoco lo es en cuanto el número de población votante (véase: figura 3.1). Los datos indican que existen secciones con grandes diferencias. Pues tenemos un mínimo de 246 y un máximo de 4565 personas registradas en la lista nominal de sus respectivas secciones electorales. Aquí nos encontramos con el problema de decidir qué número de votantes se podría considerar importante. Por ejemplo, ¿una sección con 2000 votantes es el doble de importante que otra con 1000 y 4 veces más importante que otra con 500 votantes? y ¿cómo relacionamos esto con los diferentes niveles de desarrollo social? Por supuesto, las respuestas dependerán del tipo de investigación. En este caso, nos centraremos en las secciones pobres (IDS bajo o muy bajo) y con más de 1500 personas en lista nominal. Utilizando el método de clusterización de K-means podemos agrupar las secciones de acuerdo a sus niveles de IDS y el número de votantes en su lista nominal. También podemos utilizar el método del “codo”¹⁷ para seleccionar un número k de clústeres adecuado. Este método funciona graficando diferentes números de k en contra del valor de la función potencial. En la gráfica resultante donde se vea un punto de quiebre o un “codo” (de ahí su nombre) es el valor que se escoge. Aquí el valor k sugerido es el valor de $k = 2$.

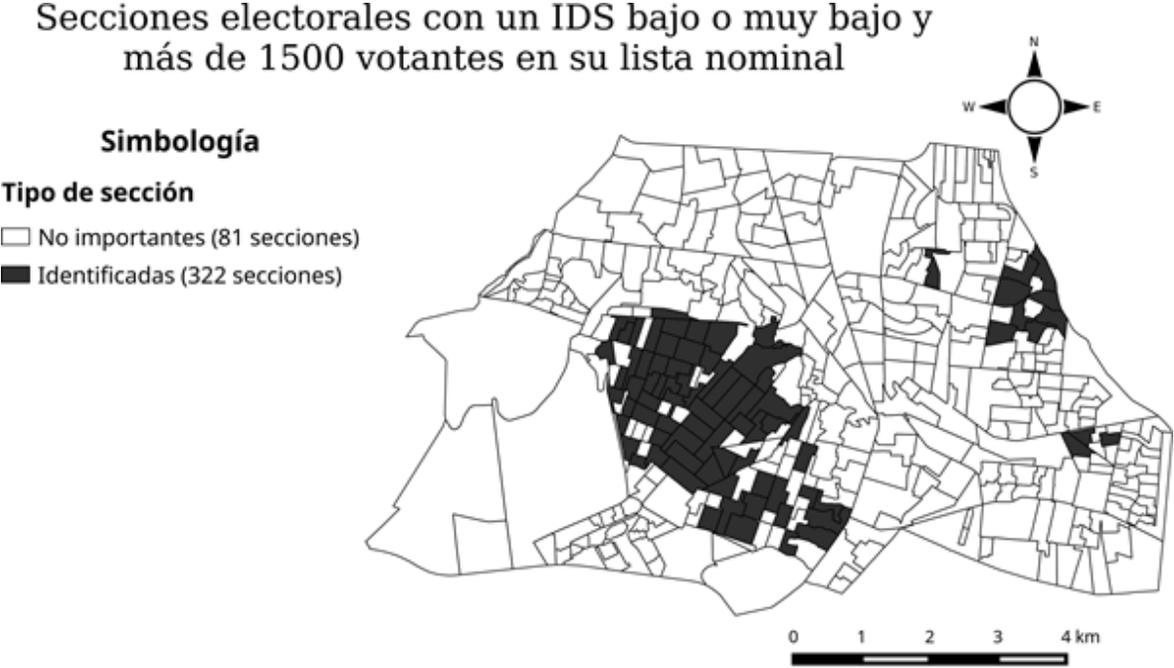
El primero de los clústeres no incluye ninguna sección con más de 1500 personas en su lista nominal y su IDS es en general alto. El clúster número 2 es el clúster que incluye mayor cantidad de población (>1500) combinado con IDS de bajos a altos. Centrándonos sólo en secciones con un IDS bajo y muy bajo obtenemos las secciones mostradas en la figura 3.7. Las secciones identificadas cumplen con estos dos requerimientos.

Podemos observar que muchas de las secciones que hemos identificados son vecinas unas de otras y muy pocas se encuentran aisladas. En la figura 3.7 resalta el manchón

¹⁷ En inglés “Elbow method”.

negro en el centro de Coyoacán, aunque aún es necesario agregar la variable de competitividad que se realizará en el siguiente capítulo.

Figura 3.7: Secciones identificadas



Fuente: Elaboración propia en base a información del IECM y del INEGI.

Capítulo 4

Geografías de poder; comportamientos electorales en Coyoacán 2009 – 2018

Data will talk to you if you are willing to listen.

Jim Bergeson

Los comportamientos electorales son una parte importante del contexto que enmarcó las elecciones del 2018, con ello podemos conocer los cambios de poder y las nuevas configuraciones llevadas a cabo durante las últimas elecciones y que convergen en los sucesos de la última elección en Coyoacán. En este contexto es importante por el hecho de que permite construir la historia del poder, la identificación de los partidos principales, la competitividad entre ellos y la distribución de los espacios geográficos por medio del voto y el cálculo de indicadores. De ahí que examinar los comportamientos electorales se vuelve imperativo en orden de obtener información suficiente para la correcta interpretación y entendimiento de las características que cualquier elección haya presentado.

En el caso de Coyoacán (al igual que otras demarcaciones de la ciudad de México) los comportamientos electorales cuentan dos historias; una donde los protagonistas son los “partidos tradicionales” (PAN, PRD, PRI, etcétera.) y otra que incluye a las “nuevas opciones políticas” (MORENA, Partido Encuentro Social (PES), Partido Humanista (PH)). En vista de ello el periodo 2009- 2018 abarca dos elecciones con los partidos tradicionales y dos elecciones con nuevas opciones políticas. Durante este periodo de tiempo la demarcación de Coyoacán estaba conformada por 403 secciones electorales, las cuatro elecciones que se celebraron en esta demarcación fueron ganadas por el PRD, ya sea en coalición con otros partidos de menor poder político o en solitario. Ante lo cual podemos establecer que no ha existido alternancia en esta localidad.

En paralelo, los niveles de competitividad nos hablan del reparto del poder (o de su concentración) y de los principales partidos, pues, aunque hay un gran número de partidos políticos siempre existirá el caso en el que un reducido grupo (incluso un solo partido) concentre la mayoría del poder arrojando así a los verdaderos

contrincantes. Sumado a lo anterior, remarco que cuando se observan ciertos niveles de competitividad, el clientelismo político toma relevancia (Molina, 2014); lo anterior reafirmar que los niveles de competitividad tienen un papel importante de nuestro análisis.

Cuadro 4.1: Partidos participantes en diferentes elecciones

Partidos participantes en elecciones en Coyoacán			
2009	2012	2015	2018
PAN	PAN	PAN	PAN
PRI	PRI	PRI	PRI
PRD	PRD	PRD	PRD
PT	PT	PT	PT
PVEM	PVEM	PVEM	PVEM
Convergencia	MC	MC	MC
NA	NA	NA	NA
PSD	-	-	-
-	-	MORENA	MORENA
-	-	PH	PH
-	-	PES	PES

Fuente: Elaboración propia en base a información del IEDF y IECM.

Para Coyoacán el número de partidos políticos que participaron en las diferentes elecciones osciló entre 7 y 10 de los cuales solo tres y dos fuerzas partidarias fueron competitivas. Estos procesos registraron una participación electoral del 45.4%, 70.1%, 47.6% y del 73.9% respectivamente existiendo mayor abstencionismo en las elecciones que no coinciden con la elección presidencial. Así mismo los análisis territoriales muestran los espacios de competitividad y de dominio partidista, incluyendo la variación geográfica que se da junto con el cambio de las tendencias electorales proveyendo de una nueva dimensión para el análisis estadístico.

El cuadro 4.1 muestra a los diferentes partidos que han participado en las cuatro elecciones analizadas. El partido Verde Ecologista de México (PVEM), el Partido

del Trabajo (PT), el Partido Social Demócrata (PSD), Nueva Alianza y Movimiento Ciudadano¹⁸ (MC) aparecen junto a los partidos ya mencionados.

4.1. Comportamientos electorales 2009-2012. Los partidos tradicionales

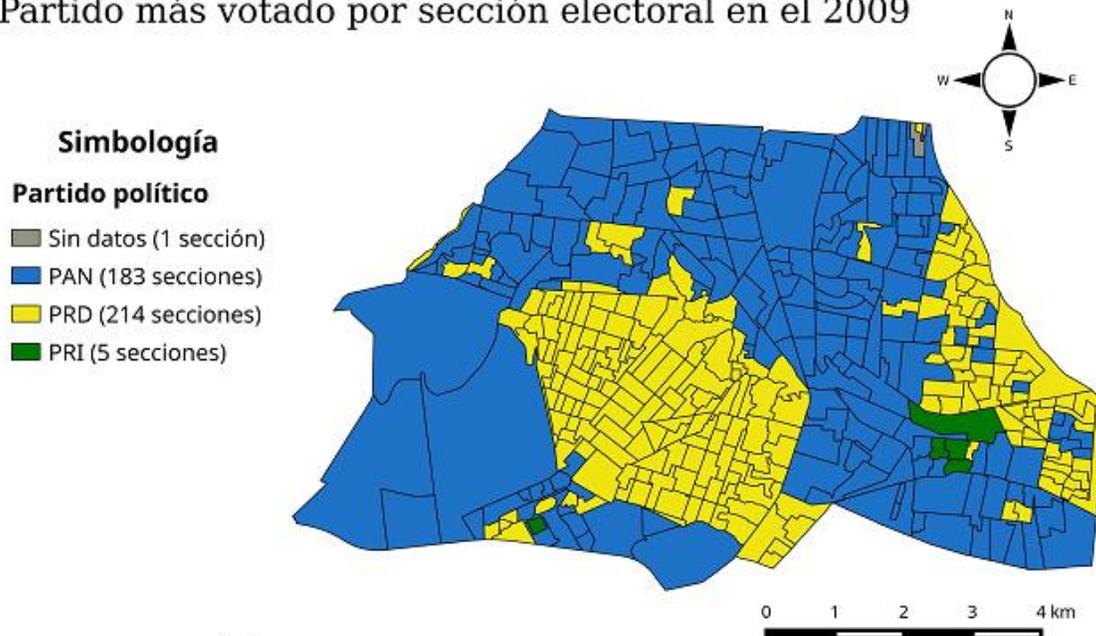
En 2009 el número de partidos políticos participantes en la elección fue de ocho. Ahora bien, solo el PAN, el PRD y el PRI ganaron alguna de las 403 secciones electorales. El PAN y el PRD obtuvieron una votación muy cercana con apenas una diferencia del 1%; el PRD ganó 31 secciones más que el PAN y entre los dos se repartieron un total de 397 secciones de las 403, siendo ambos las fuerzas dominantes de esta contienda. La figura 4.1 muestra una marcada separación geográfica entre los espacios ganados por el PAN y el PRD además de las secciones que ganó el PRI. Es claro que en esta elección cada partido tuvo conformado su territorio donde electoralmente dominan, y al mismo tiempo se excluyen geográficamente en concordancia con lo que se ha observado en otras elecciones (Vilalta, 2008).

Los niveles de competitividad que se midieron resultaron en un nivel medio según el índice de RAE, pues más de la mitad de los partidos políticos en la contienda no ganaron ninguna sección electoral y la votación que obtuvieron fue pequeña. No obstante, el indicador de competitividad electoral calculado resultó en un nivel muy alto entre los dos primeros lugares, es decir el poder se encuentra concentrado y disputado entre los dos primeros lugares. El resto de los partidos quedo marginado, con muy poco poder político y sin posibilidades de ganar la elección. A nivel de sección electoral (ver. 4.2) el ICE se distribuyó en secciones de escasa y muy alta competitividad identificando espacios en disputa que contrastan con aquellos donde los niveles son bajos y escasos. Estos últimos concuerdan con los espacios definidos por la figura 4.1 en otras palabras el PAN y el PRD mantiene espacios de control con un ICE bajo y de disputa con un ICE alto.

¹⁸ Antes Convergencia.

Figura 4.1: Secciones ganadas por partido en la elección para jefe delegacional del 2009

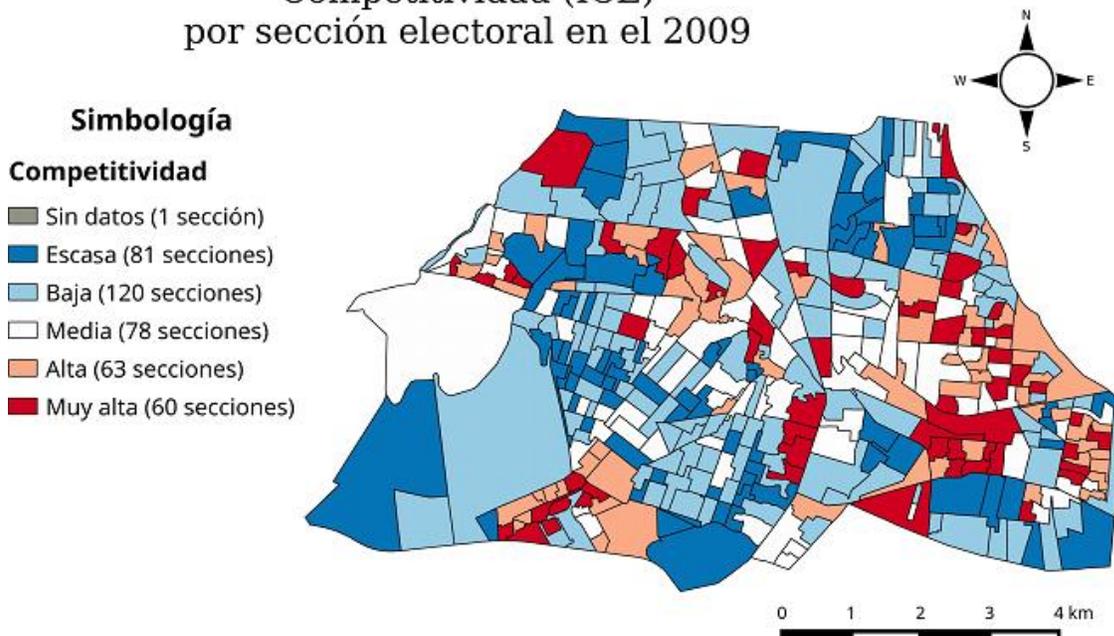
Partido más votado por sección electoral en el 2009



Fuente: Elaboración propia e base a información del IEDF.

Figura 4.2: Competitividad (ICE) por secciones electorales

Competitividad (ICE) por sección electoral en el 2009

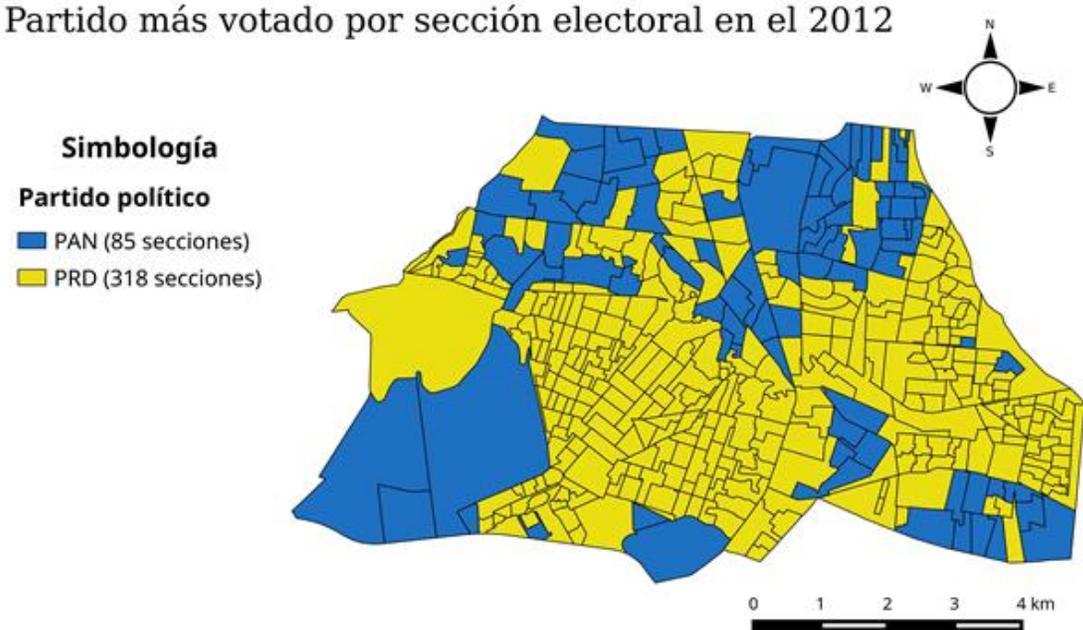


Fuente: Elaboración propia en base a información del IEDF.

En cuanto al proceso electoral del 2012, la lucha vista entre el PRD y el PAN en el 2009 continuó, el PRD tuvo una expansión en los territorios del PAN y la distancia que se mostraba entre estos dos partidos en el pasado proceso electoral se incrementó considerablemente pues paso del 1% en el 2009 a 15.5% en el 2012. Teniendo en cuenta el fortalecimiento que el PRD tuvo en esta elección, ya que gano un mayor número de secciones electorales en comparación con el proceso electoral pasado (ver. 4.3). El PRI en coalición con el PVEM no lograron ganar ninguna sección electoral, ni alcanzar al PAN, que a pesar del incremento del PRD siguió como segunda fuerza política, aunque con un debilitamiento considerable ya que solo ganó 85 secciones electorales (98 secciones menos con respecto a las ganadas en el 2009). El número de partidos que participaron en la contienda fue de siete, pero de nuevo el resto de los partidos tuvo una participación marginal.

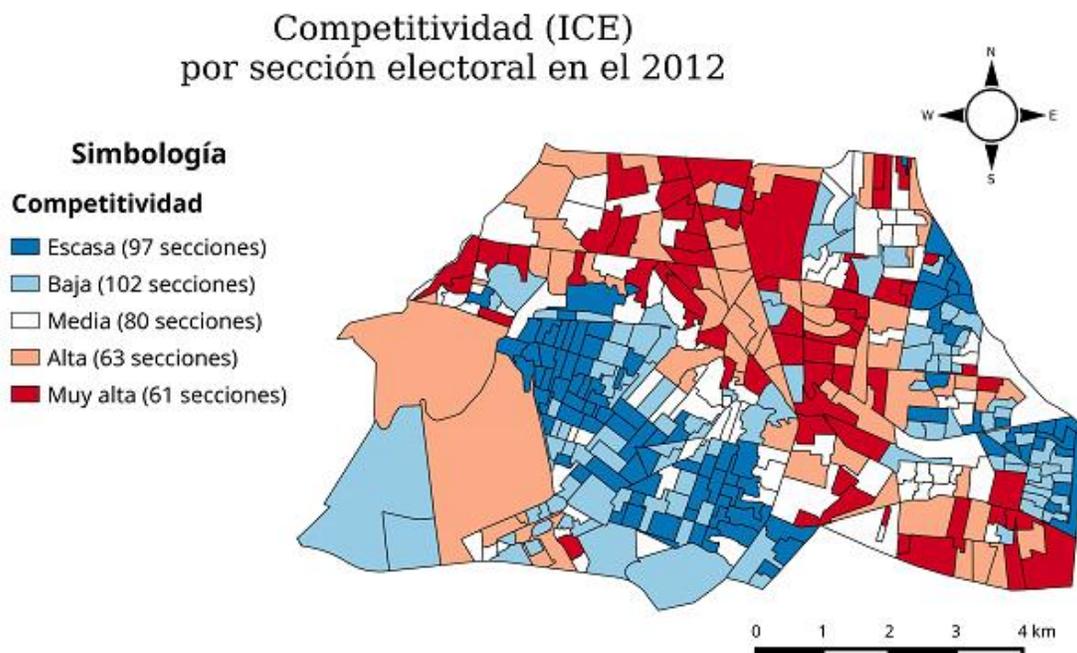
Los niveles de competitividad para este año resultaron en un cambio del ICE, que pasó de un 0.970 en el 2009 a un 0.623 en el 2012, un descenso del 0.347 en la competitividad entre los primeros dos partidos. Mientras que el índice Rae que registraba una competitividad media en el año 2009 ahora registraba una baja competitividad (de 0.755 a 0.73). Una posible explicación puede ser atribuida a la expansión que experimentó el PRD en la demarcación, ocupando territorios en los que el PAN y el PRI tenían presencia.

Figura 4.3: Secciones ganadas por partido en la elección para jefe delegacional del 2012



Fuente: Elaboración propia en base a información del IECM.

Figura 4.4: Competitividad en cada sección según el índice de competitividad electoral ICE en el 2012



Fuente: Elaboración propia en base a información del IECM.

Una comparación entre las figuras 4.3 y 4.1 demuestra la expansión territorial del PRD y el cambio geográfico en Coyoacán que corresponde al descenso de los índices de competitividad electoral y al incremento en el margen de victoria. En forma global, por cada sección electoral la competitividad entre el PAN y el PRD mostró mayores valores en secciones del PAN como muestra la figura 4.4. Las secciones ganadas por el PRD tienen una baja competitividad lo que indica un control casi total en esas secciones.

Al final de la primera mitad de los procesos electorales del 2009 y 2012, se muestra que la disputa electoral por Coyoacán se llevó a cabo principalmente por el PAN y el PRD con una participación marginal por parte de los demás partidos políticos. Lo que evidencia la concentración de poder en sólo dos partidos, pero ésta se incrementó para el año 2012, debido al fortalecimiento del PRD. Geográficamente este reparto del poder arroja una configuración geográfica de espacios partidistas irregulares pero que aparentan tener espacios de control definidos donde la competitividad es baja. El cambio territorial dado del 2009 al 2012 es acompañado de una baja en los niveles de competitividad en la demarcación conservando niveles altos en las secciones electorales disputadas. Así, para el final del periodo 2009-2012 el PRD es el partido dominante de Coyoacán con ningún contrincante con suficiente poder político para

ganar la elección; el PAN es la segunda fuerza y el resto de los partidos mantiene una participación despreciable.

4.2. Comportamientos electorales 2015-2018. Nuevas opciones políticas

Las elecciones en la segunda mitad del periodo 2009-2012 estuvieron marcadas por el ingreso y participación de nuevas opciones políticas. La aparición de tres partidos políticos nuevos supuso un cambio en la geografía y repartición del poder en Coyoacán, pues ahora había un mayor número de partidos en la contienda que modificaban el panorama visto en las pasadas elecciones. El partido Movimiento Regeneración Nacional (MORENA), el partido Encuentro Social (PES) y el Partido Humanista (PH) se integraron en el 2015 convirtiendo a la elección de ese año en una muy competitiva, registrando un ICE de 0.961 y un IRAE de 0.816, niveles comparables con la elección del 2009. Conviene subrayar que no todos los partidos de recién ingreso tuvieron una igual participación. De los tres, MORENA demostró ser tan competitivo como las principales fuerzas políticas, en contraste las votaciones del PH y el PES fueron insignificantes. Pero lo que diferencia a esta contienda de las anteriores es que marca un punto de inflexión que cambió las preferencias electorales. Las principales fuerzas habían sido hasta el 2012 el PRD y el PAN, pero después del 2015 el PAN había sido desplazado por MORENA que se acercaba amenazante al PRD. El margen de victoria entre estos dos partidos fue de apenas 1 % en el 2015.

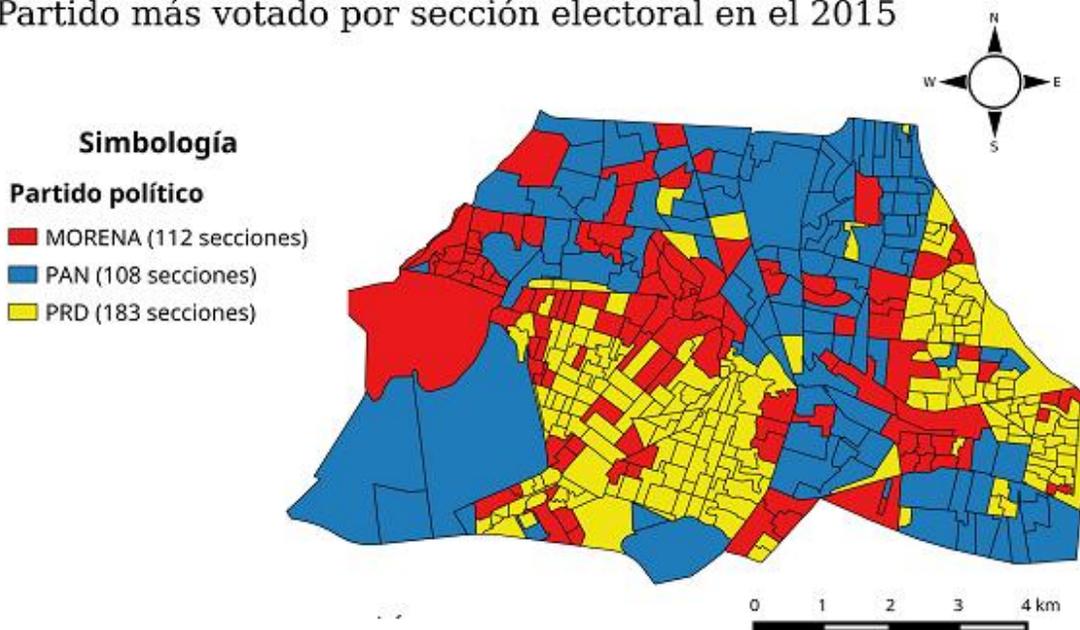
Los niveles de competitividad se vieron modificados con respecto al proceso electoral pasado tanto la fragmentación del poder (de 0.73 a 0.816), el margen de victoria (de 0.155 a 0.01) y la competitividad entre los dos primeros lugares (de 0.623 a 0.961). Con un promedio de competitividad alto según el ICE y el índice de Rae en las secciones electorales. La participación en la demarcación fue del 47.5% una de las más bajas registradas en nuestro periodo de tiempo. El ingreso tan competitivo de MORENA significó un duro golpe para la estabilidad y control del PRD en la demarcación; así mismo geográficamente MORENA ocupó territorio principalmente del PRD y del PAN reconfigurando el reparto del poder y la geografía electoral de Coyoacán que quedó repartida entre tres partidos políticos (como se muestra en la figura 5.5). El PRD ganó 183 secciones, Morena 112 secciones y el PAN 108 secciones, lo que significó una pérdida de 135 secciones para el PRD y una ganancia de 23 secciones electorales para el PAN. En particular las secciones de MORENA no se encuentran conglomeradas como fue el caso del PAN y el PRD durante el 2009 y el 2012 y no podemos decir que estos partidos se excluyen geográficamente. El resto de los partidos políticos nuevos no fueron tan competitivos como lo fue MORENA y su

participación fue tan escasa como la del resto de los partidos tradicionales y sin posibilidad de ganar alguna sección electoral.

A su vez el mapa de competitividad se distribuyó de una manera diferente con 238 secciones con una competencia alta o muy alta. Es decir, el 59% de las secciones electorales tenían niveles altos o muy altos en el 2015 (ver. 4.6). Esto va de la mano con la medición del índice de Rae del 0.816 dado que el electorado ahora se repartía en tres partidos políticos competitivos y entre siete partidos menores. Al final de esta elección el ingreso de MORENA junto al PH y al PES se tradujo en un cambio en el reparto del poder y la configuración geográfica de Coyoacán, presionando a las principales fuerzas políticas que ya existían y por primera vez desde el 2009, el PRD estuvo en posibilidades de perder la elección. Este panorama un tanto desesperado vuelve verosímil y posible la aparición de actos clientelares en la siguiente elección dado que el partido en el poder ya no cuenta con la seguridad, el afianzamiento y el dominio que mostró en los procesos electorales pasados. La elección del 2015 es el primer indicio de un contexto competitivo y donde el uso de programas sociales y de actos de clientelismo político toman relevancia.

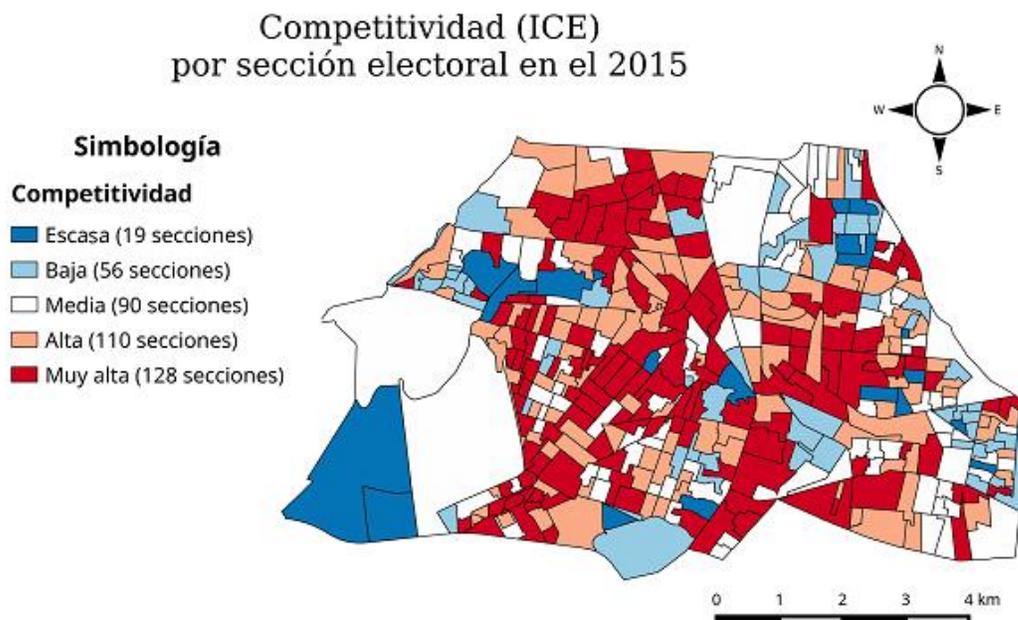
Figura 4.5: Secciones ganadas por partido en la elección para jefe delegacional del 2015

Partido más votado por sección electoral en el 2015



Fuente: Elaboración propia en base a información del IECM.

Figura 4.6: Competitividad en cada sección según el índice de competitividad electoral ICE en el 2015



Fuente: Elaboración propia en base a información del IECM.

Avanzando en nuestro razonamiento, para el 2018 la competitividad ICE se ubicó en 0.748 es decir una reducción del 26.8% con respecto al 2015. Esto se explica con el incremento de la preferencia a favor de MORENA que se convirtió en la principal fuerza política de la demarcación superando al PRD. En esta elección las coaliciones jugaron un papel principal en contraste con las elecciones pasadas pues gracias a la alianza entre el PAN (la tercera fuerza política) y el PRD (ahora la segunda), no hubo alternancia en el 2018. A pesar de los buenos resultados obtenidos por MORENA en solitario y en coalición con el PT y el PES. De cuadro 4.2 podemos observar los porcentajes de votaciones obtenidos por cada fuerza política en el 2018 incluyendo las opciones presentes en la boleta electoral y por partido y coalición. MORENA superó a las principales fuerzas políticas de la demarcación enfrentándose principalmente con el PRD y cuyo margen de victoria ahora fue del 8.1% a favor de MORENA. Geográficamente estos resultados se traducen como una continuación a la expansión de MORENA por territorios del PAN y mayormente del PRD, quitándole un gran número de secciones a los perredistas. El mapa 4.7 muestra los territorios ganados por MORENA y los demás partidos políticos, el PAN conservó muchas de sus secciones electorales y el PRD perdió una gran cantidad de territorio ganando apenas 45 secciones (casi seis veces menos que MORENA). Lo anterior no quiere decir que a mayor territorio mayor voto, si bien el PAN obtuvo más del doble del territorio que el PRD, en la votación fue el PRD el que superó al PAN (ver 4.2 y 4.7). Si revisamos la lista nominal las secciones con mayor número de población en la lista son secciones que se

compiten entre MORENA y el PRD, el cuadro 4.3 muestra algunas de las características de las secciones categorizadas por los partidos que ocuparon los primeros lugares en ellas. Como era de esperarse los partidos de MORENA, PRD y PAN compiten la mayoría de las secciones incluyendo una minoría donde también aparece el PRI.

Cuadro 4.2: Porcentaje de votos obtenidos por cada partido y coalición en la elección de alcalde en Coyoacán 2018

Coyoacán 2018			
Partido o coalición	% de votación	% de votación	% de votación
PAN	19.92	20.5	-
PRI	9.55	9.55	9.55
PRD	23.31	23.91	-
PT	1.75	2.28	-
PVEM	3.1	3.1	3.1
MC	2.54	2.54	-
NA	1.71	1.71	1.71
MORENA	31.52	32.12	-
PES	1.15	1.63	-
PH	2.22	2.22	2.22
Común (MORENA-PT-PES)	1.13	-	-
Común (MORENA-PT)	0.28	-	-
Común (PT-PES)	0.03	-	-
Común (MORENA-PES)	0.17	-	-
Común (PAN-PRD-MC)	1.04	-	-
Común (PAN-PRD)	0.39	-	-
Común (PAN-MC)	0.07	-	-
Común (PRD-MC)	0.11	-	-
Coalición PAN-PRD-MC	-	-	47.38
Coalición PAN-PRD-MC	-	-	36.03

Fuente: Elaboración propia en base a información del IECM.

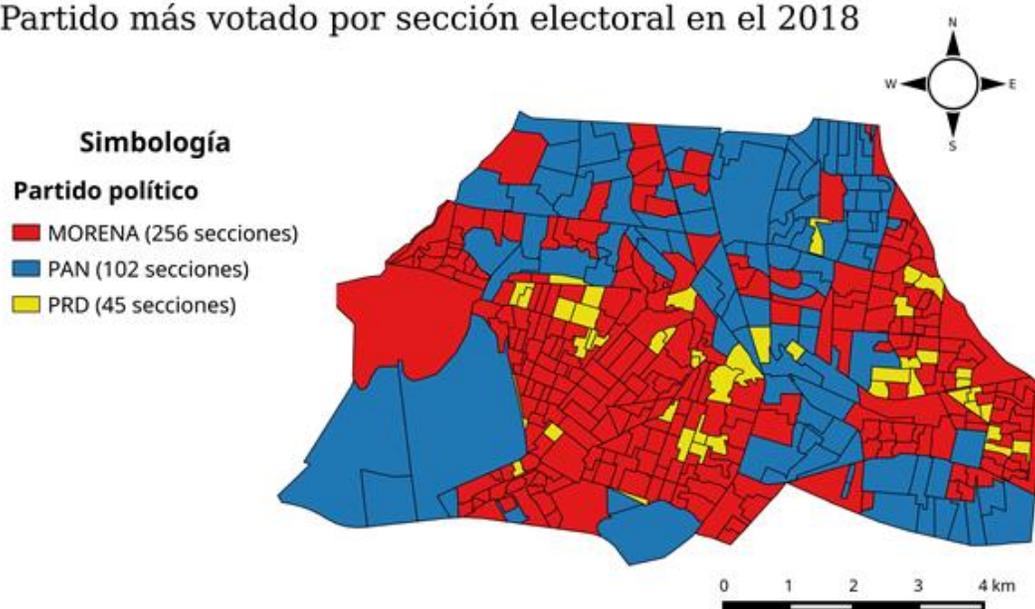
Cuadro 4.3: Indicadores promedio de las secciones según los primeros lugares

Coyoacán 2018						
Primeros lugares	Promedio en lista nominal	Promedio de votos emitidos	Número de secciones	IDS	ICE	IRAE
MORENA-PRD	1576.15	1146.85	223	0.807	0.801	0.754
MORENA-PAN	1177.54	892.75	165	0.948	0.753	0.770
MORENA-PRI	409.0	227.0	1	0.846	0.570	0.774
PAN-PRI	1108.83	850.83	6	0.938	0.381	0.661
PRD-PAN	1392.00	1045.87	8	0.825	0.748	0.771

Fuente: Elaboración propia en base a información del IECM.

Figura 4.7: Secciones ganadas por partido en la elección para alcalde del 2018

Partido más votado por sección electoral en el 2018



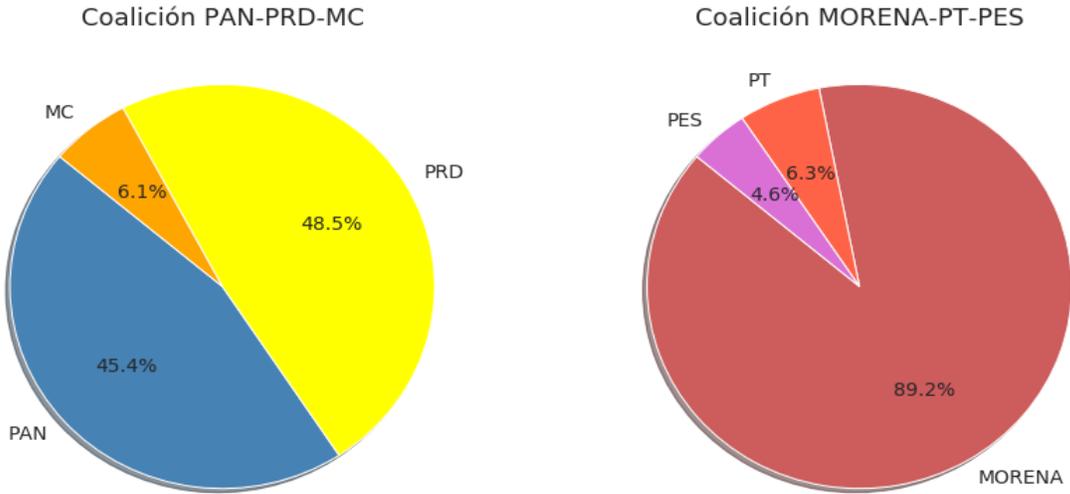
Fuente: Elaboración propia en base a información del IECM.

En definitiva, el crecimiento de poder político que obtuvo MORENA del 2015 al 2018 en términos de votaciones por partidos políticos individuales está por encima de cualquier otro competidor en el proceso electoral del 2018, estos resultados fácilmente dan como ganador a MORENA. No obstante, en términos de coaliciones, la

coalición PAN-PRD-MC superó a la coalición formada por MORENA-PT-PES por un 11.32% de la votación en general y ganó la mayoría de las secciones electorales. MORENA proporcionó casi el 90% de la votación en su coalición mientras que el PAN y el PRD proporcionaron cantidades semejantes (45.4% y el 48.5% respectivamente) (ver gráfico 4.8). El resto de los partidos políticos participantes no obtuvo más del 10% del total de la votación y no ganaron ninguna sección electoral. Debido al papel que las coaliciones jugaron en esta elección los niveles de competitividad entre partidos y coaliciones no poseen los mismos comportamientos geográficos, así por ejemplo un partido puede dominar a los demás en cierta zona, pero no así a su coalición.

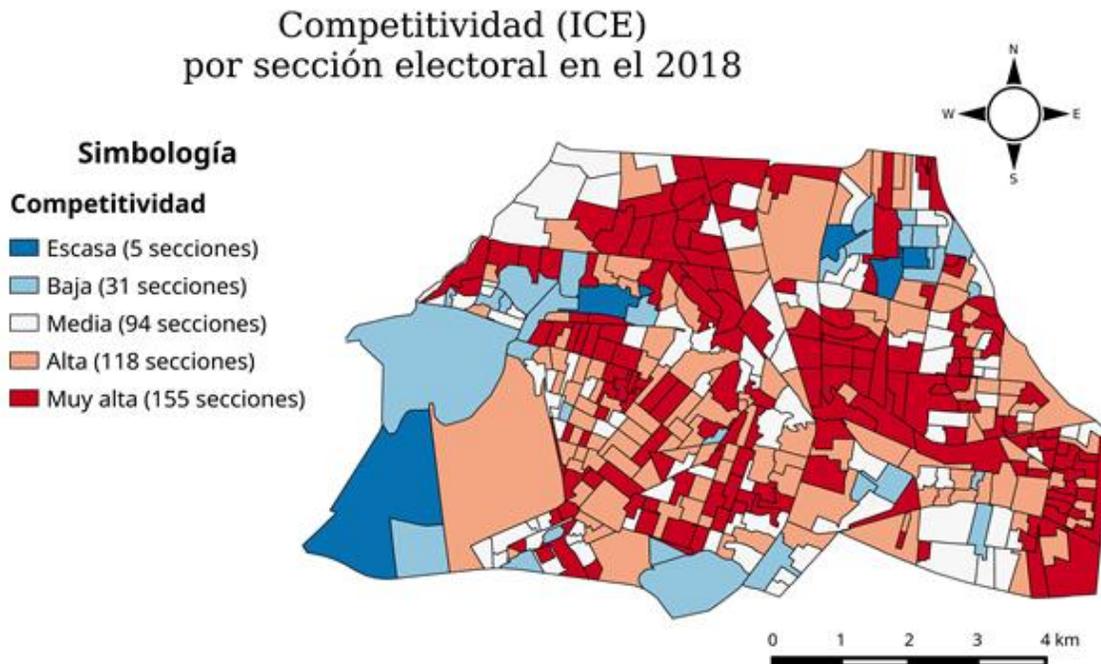
Los niveles de competitividad calculados para los partidos y coaliciones revelan que éstas últimas modificaron los niveles existentes entre partidos aumentando o reduciendo el margen de victoria, pero manteniendo secciones en donde las votaciones de ambas coaliciones son muy parejas. Las figuras 4.9 y 4.10 muestran la competitividad medida para los partidos y las coaliciones en la demarcación. Estas figuras muestran además las diferencias entre las competitividades y cómo las coaliciones tienen un agrupamiento de secciones con altos niveles en el centro de Coyoacán. Estas secciones ya registraban altos niveles entre partidos por lo que podemos consideradas como espacios de alta competitividad electoral.

Figura 4.8: Contribución de cada partido en proporción a su coalición



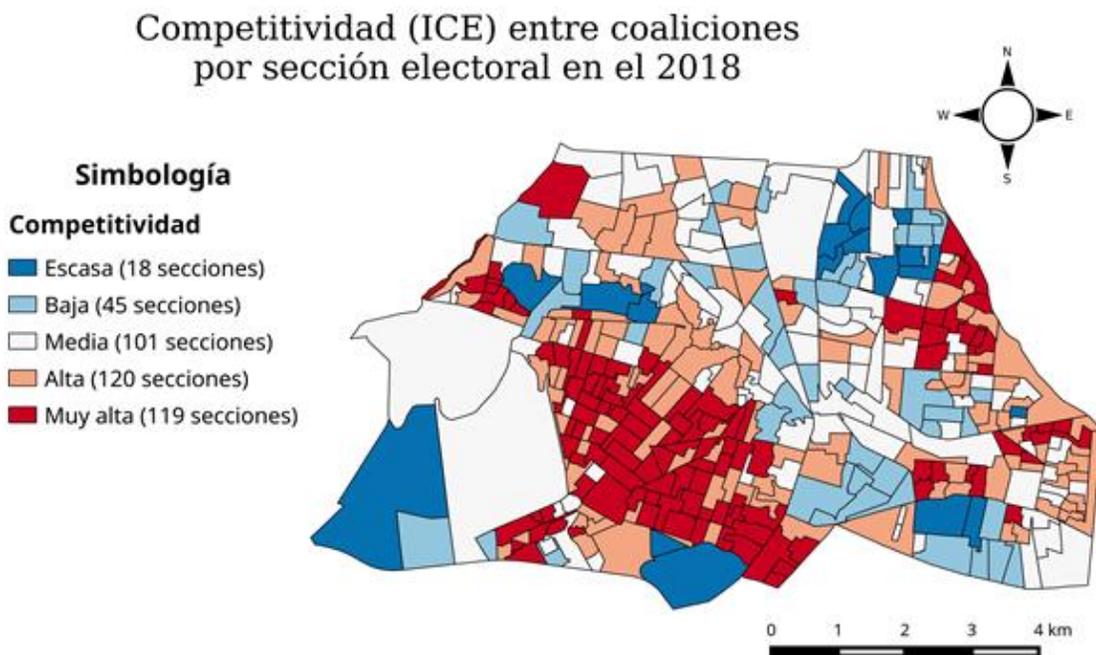
Fuente: Elaboración propia en base a información del IECM.

Figura 4.9: Competitividad entre partidos en Coyoacán 2018



Fuente: Elaboración propia en base a información del IECM.

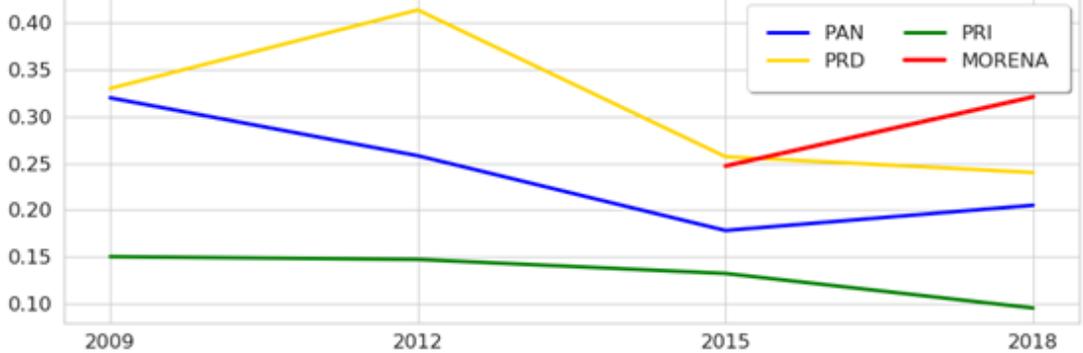
Figura 4.10: Competitividad entre las coaliciones en Coyoacán 2018



Fuente: Elaboración propia en base a información del IECM.
Como se afirmó arriba la preferencia electoral sufrió varios cambios dentro del periodo 2009-2018. En las figuras 4.11, 4.12 se observan los cambios de la preferencia

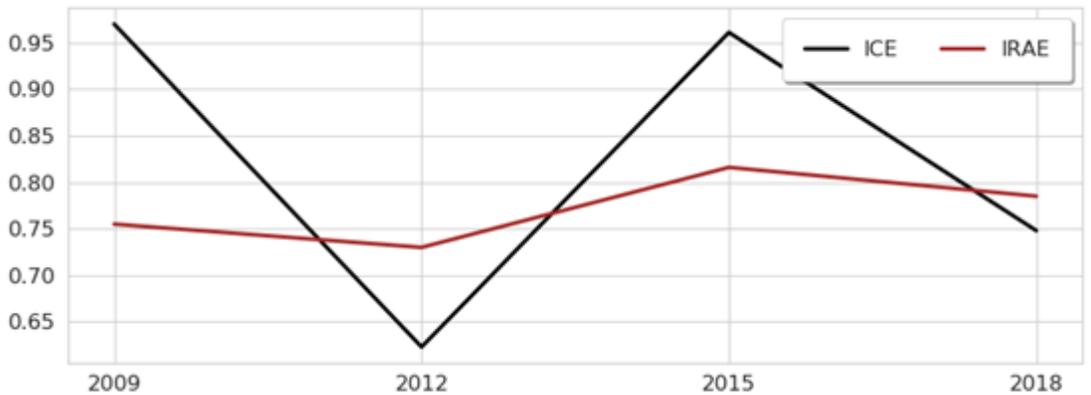
de los principales partidos políticos en Coyoacán y la variación de la competitividad en cada elección. La preferencia del PRI, PRD y PAN van descendiendo mientras que MORENA, que inició con una preferencia considerablemente alta en el 2015, no hizo más que crecer para el proceso electoral 2018. Territorialmente este cambio de preferencia es también evidente en zonas de influencia de cada partido que se vieron afectadas, reconfigurando la geografía electoral de Coyoacán y el reparto del poder en la demarcación. Los niveles de competitividad fueron mayores en las elecciones que no coincidieron con la elección presidencial y acorde con lo visto el índice IRAE tuvo un pequeño aumento con el ingreso de MORENA, el PES y el PH. Para finales del 2018 MORENA fue la principal fuerza política de la demarcación, superando a los partidos tradicionales del PAN y el PRD, que sólo al unir fuerzas para el proceso electoral del 2018 pudieron evitar que MORENA ganara la contienda.

Figura 4.11: Cambio en las preferencias electorales en el periodo 2009-2018



Fuente: Elaboración propia en base a información del IECM y del IEDF.

Figura 4.12: Variación en los índices de competitividad durante el periodo 2009-2018

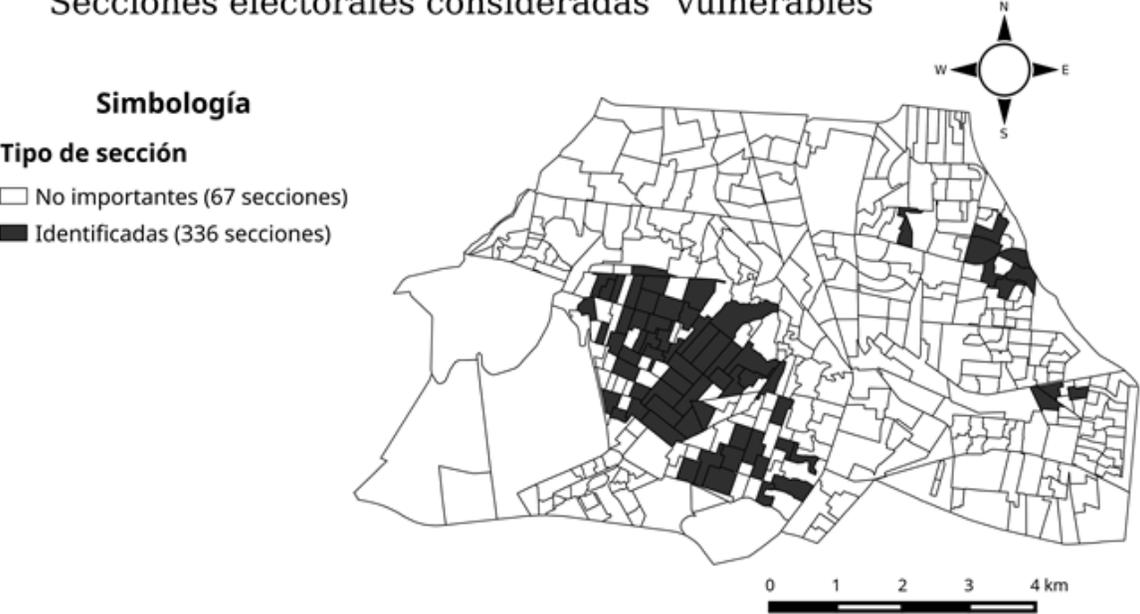


Fuente: Elaboración propia en base a información del IECM y del IEDF.

Continuando con el trabajo realizado en el capítulo anterior, agregamos la variable de competitividad para la identificación de espacios vulnerables. Obtenemos que

efectivamente Coyoacán en el 2018 tuvo espacios vulnerables para que se presentaran actos de clientelismo político como los trabajos periodísticos señalaban, esto último de acuerdo a las características que la literatura especializada indica. La figura 4.13 presenta los espacios de vulnerabilidad identificados con las características definidas en el capítulo 3. De nuevo, resalta el agrupamiento en el centro de Coyoacán en negro con algunas secciones en blanco, que no tienen una o más de los tres aspectos definidos para estos espacios.

Figura 4.13: Espacios de vulnerabilidad finales
Secciones electorales consideradas "vulnerables"



Fuente: Elaboración propia en base a información del IECM y del INEGI.

Capítulo 5

Análisis estadístico y representación de datos: los usos electorales del gasto social en la Alcaldía de Coyoacán 2018

All models are wrong but some are useful.

George E.P Box.

5.1. Datos y variables

Los reportes realizados durante las campañas electorales del 2018 en Coyoacán señalan el uso de los programas sociales en la demarcación por parte del partido en el poder. Además de unas modificaciones que permitieron a estos programas trabajar de manera diferente a como lo hacían, por ejemplo, aumentando la capacidad de beneficiarios, aumentando el presupuesto destinado a los programas y eliminando requisitos para que más personas puedan acceder a ellos. Habría que decir también que la literatura sobre el clientelismo político en México refiere al uso de los programas sociales con fines electorales más como una regla en general que como una excepción. Con el objetivo de obtener una conclusión desde los datos se analiza la relación entre estos dos elementos; la preferencia del voto para el PRD y el número de beneficiarios del programa social “A tu lado” así como otros factores. El estudio se enfoca sólo a este programa social pues según sus reglas de operación desde su creación ha estado dirigido a personas mayores de edad y por lo tanto con capacidad de emitir su voto.

El número de observaciones de nuestros datos es de $n = 403$, que representan las 403 secciones electorales de Coyoacán las cuales también son nuestra unidad espacial base. Para estas secciones se calcularon todas las variables. En este capítulo identificamos como variable dependiente (VD) la preferencia del voto hacia el PRD que es calculada por medio de la proporción entre el voto obtenido por el PRD y el

total de los votos válidos emitidos durante la elección del 2018¹⁹. Las variables independientes (VI) consideradas son: las proporciones de beneficiarios, de población no católica y de viviendas habitadas sin ningún bien, además de la proporción de personas analfabetas los índices de competitividad y de desarrollo social. Estas variables son continuas y calculadas a partir de los resultados de la elección para alcalde en Coyoacán, del padrón de beneficiarios del programa social para el 2017 y de las estadísticas censales a escalas geoelectorales. El cuadro 5.1 muestra la descripción de cada una de las variables calculadas.

Cuadro 5.1: Descripción de las variables.

Variable	Descripción
Independiente	
PRD	Proporción de votos obtenidos por el PRD
Dependiente	
Beneficiarios	Proporción de beneficiarios
No católica	Proporción de población no católica (Evangelistas, protestantes etcétera.).
Viviendas	Proporción de viviendas habitadas sin ningún servicio (internet, televisión, automóvil etcétera)
Analfabetas	Proporción de personas analfabetas
ICE	Nivel de competitividad por partidos según el ICE
ICEco	Nivel de competitividad por coaliciones según el ICE
IRAE	Nivel de competitividad por el índice de Rae
IDS	Índice de desarrollo social

5.2. Análisis exploratorio de datos espaciales

La variable beneficiarios fue construida a partir de los datos del padrón de receptores del programa social. Se debe agregar que los habitantes que reciben un programa de transferencia directa están georreferenciados por medio del nombre de la colonia

¹⁹ Consideramos como voto válido a todos los votos emitidos que no son nulos.

donde habitan, a causa de esto se estima el número de beneficiarios de cada sección electoral aplicando el método de ponderación de área simple (explicado en el capítulo 2) a cada colonia de Coyoacán. La variable final es calculada tomando el cociente entre la estimación obtenida y el número total de habitantes, la variable es entonces, la proporción de beneficiarios de cada sección electoral como se mencionó antes. El resto de las variables fueron calculadas de manera análoga, aunque más sencillo puesto que los datos originales ya se encontraban medidos por sección electoral y no fue necesario aplicar ningún método de desagregación.

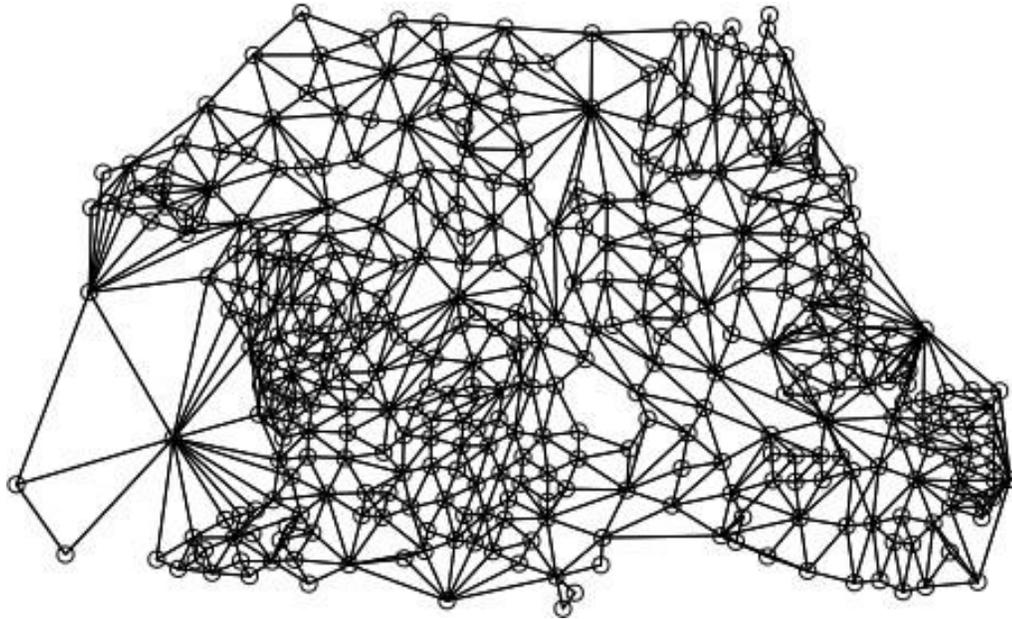
Todas las VI mostraron estar correlacionadas significativamente con la VD (cuadro 5.2) según el coeficiente de Spearman, las correlaciones son positivas y negativas, las variables de beneficiarios y analfabetas son las que más están correlacionadas positivamente así mismo el IDS tiene una correlación negativa moderada con la preferencia del PRD.

Cuadro 5.2: Correlaciones de las variables con la preferencia del PRD.

Variable	r_s	p -val	Interpretación
Analfabetas	0.659	0.000	Moderada
Beneficiarios	0.539	0.000	Moderada
Viviendas	0.420	0.000	Baja
ICE	0.380	0.000	Baja
ICEco	0.373	0.000	Baja
No católica	0.324	0.000	Baja
IRAE	-0.279	0.000	Escasa
IDS	-0.734	0.000	Alta

En segunda instancia una autocorrelación espacial significativa (un I de Moran significativo) en las VD y en la VI, es una primera pista que sugiere el uso de modelos espaciales. Precisamente los estimadores del modelo de regresión lineal por mínimos cuadrados pueden verse afectados por la dependencia espacial sustantiva o molesta en los errores. Por esta razón se define una matriz de pesos para Coyoacán bajo el criterio de contigüidad de primer orden tipo Reina. El grafo construido que conecta las secciones electorales se muestra en la figura 5.1. El índice de Moran fue calculado para las VD y VI resultando en una autocorrelación positiva para todas ellas.

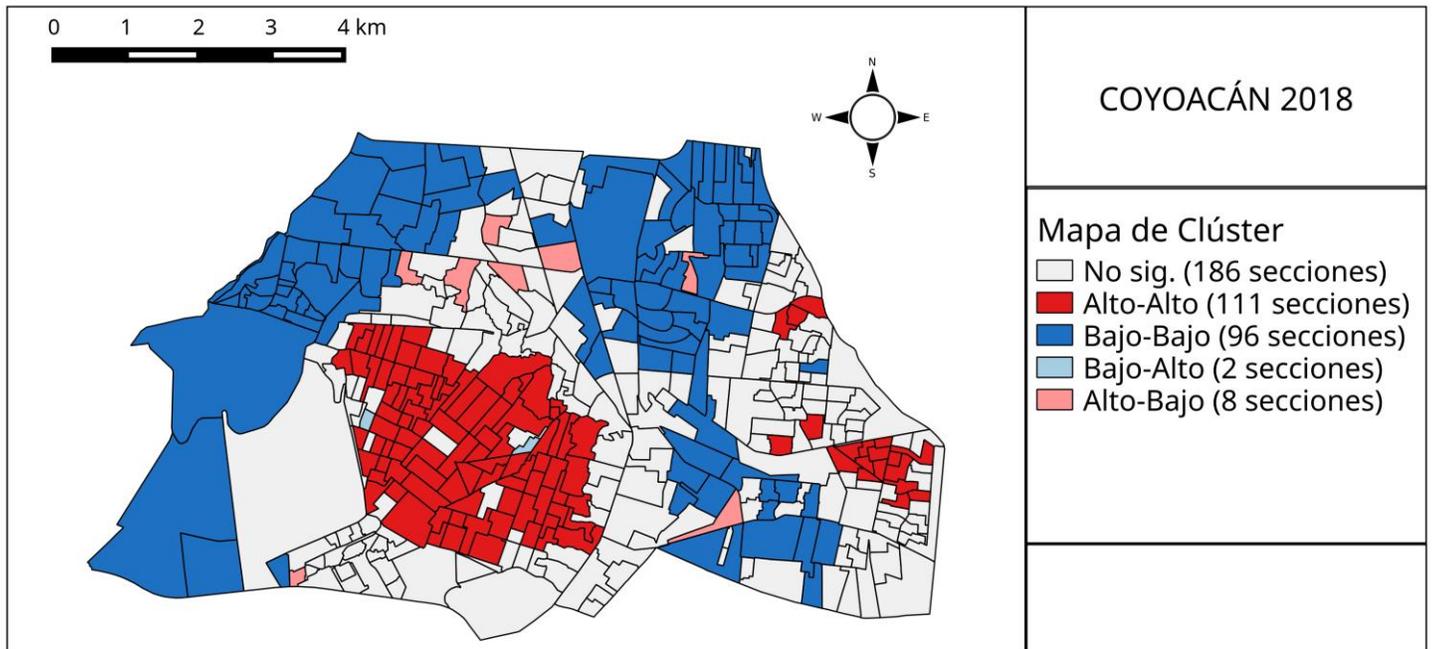
Figura 5.1: Grafo construido para Coyoacán con un vecindario tipo Reina



Cuadro 5.3: Estadístico I de Moran global para las variables dependientes e independiente

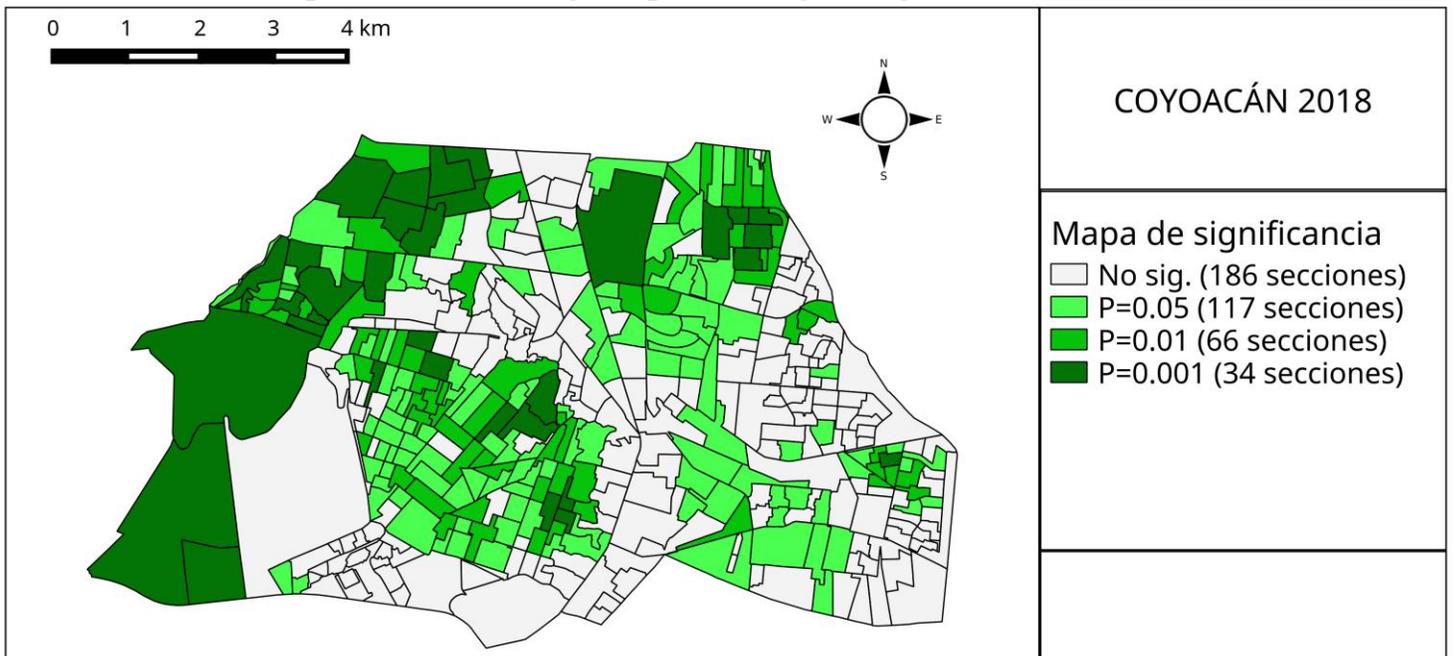
Variable	<i>I</i>	<i>p-val</i>
PRD	0.637	0.000
Analfabetas	0.611	0.000
Beneficiarios	0.531	0.000
Viviendas	0.507	0.000
ICE	0.299	0.000
ICEco	0.489	0.000
No católica	0.300	0.000
IRAE	0.452	0.000
IDS	0.777	0.000

Figura 5.2: Clústeres para la preferencia del PRD



Fuente: Elaboración propia en base a información del IECM.

Figura 5.3: Secciones por significancia para la preferencia del PRD



Fuente: Elaboración propia en base a información del IECM.

La preferencia del PRD está autocorrelacionada significativamente, esto acorde a los estudios espaciales realizados en nuestro país en diferentes elecciones, en donde el voto hacia cierto partido tiende a formar agrupaciones. Además, la competitividad entre coaliciones está más autocorrelacionada espacialmente que la competitividad entre partidos, medidas con el mismo indicador. Por su parte el índice de Moran local indica, en efecto, la formación de clústeres para el PRD y outliers espaciales, esto es, la preferencia del PRD define espacios favorables distinguibles de aquellos espacios en donde no le favorecen (véase: figura 5.2 y figura 5.3).

5.3. Modelos de regresión

El modelo de regresión lineal que utiliza todas las variables presenta múltiples problemas. El primero es que muestra heterocedasticidad y los errores no se distribuyen de manera normal, además de los problemas de multicolinealidad, por lo que se han de eliminar variables. Comprobando los 256 modelos diferentes²⁰ que se pueden construir sólo con estas variables dejando fija la variable de beneficiarios. El mejor modelo con base en el AIC y el BIC y que no padecen de los problemas antes mencionados, está conformado por las VI de beneficiarios, viviendas, no católica y analfabetas. Este modelo es estadísticamente significativo de acuerdo a la prueba F con p -valor=0.000 y explica alrededor del 50% de la variación de la preferencia del PRD en Coyoacán. La proporción de beneficiarios es un predictor significativo al igual que las otras tres variables independientes, el modelo revela no tener problemas de multicolinealidad pues todos los VIF están por debajo del valor crítico de 5, ni de heterocedasticidad (Breusch-Pagan=4.8522, p =0.3028) y los residuos están normalmente distribuidos (Jarque-Bera=2.2044, p =0.3321).

²⁰ Los algoritmos para realizar este procedimiento se pueden encontrar en los anexos de la tesis.

Figura 5.4: Histograma y diagrama de dispersión de los residuos para el modelo OLS

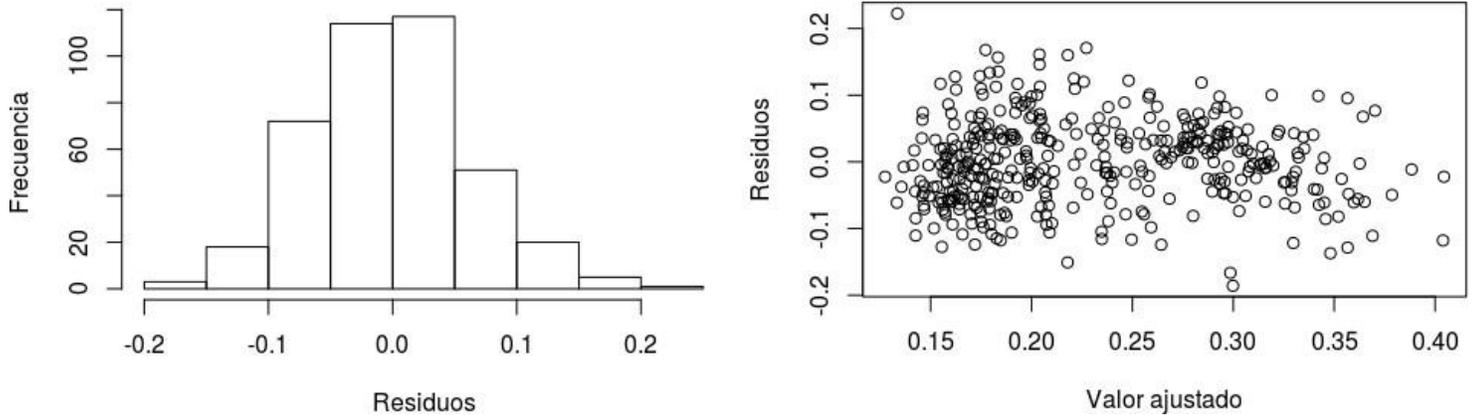
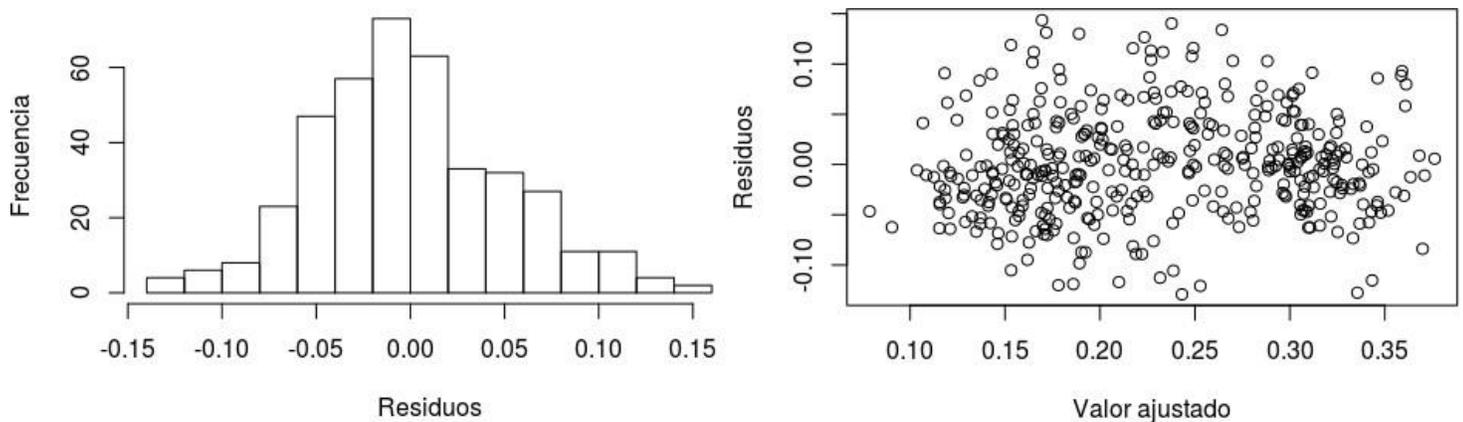


Figura 5.5: Histograma y diagrama de dispersión de los residuos para el modelo de retardo espacial



La independencia espacial de los residuos es evaluada por medio de la I de Moran global. Este estadístico revela que los residuos están autocorrelacionados espacialmente ($I = 0.3058, p = 0.000$) violando este supuesto. Como vimos antes, la VD y las VI mostraron autocorrelación espacial positiva, por lo que la dependencia en los residuos puede ser dependencia espacial sustantiva. Consideremos ahora el multiplicador de Lagrange (ML) el cual recomienda el uso del modelo de autorregresivo de retardo espacial (ML robusto = 46.5750, $p = 0.000$). Los resultados de este modelo revelan de nuevo que todas las variables son predictores significativos, la prueba Jarque-Bera muestra que los residuos se distribuyen de manera normal, además de ser homocedásticos y la prueba de la I de Moran no es significativa ($I = -0.0130, p = 0.6377$). Es por esto que el modelo de retardo espacial “captura” de manera

exitosa la dependencia espacial de los residuos, como se aprecia en el coeficiente autorregresivo $\rho=0.610$ significativo (valor=13.984 y $p=0.0000$). El cuadro 5.4 muestra un resumen y comparativa de ambos modelos.

Cuadro 5.4: Modelos de Regresión

Variable	OLS			Retardo espacial		
	Coeficiente	<i>t</i> -valor	<i>p</i> -valor	Coeficiente	<i>z</i> -valor	<i>p</i> -valor
β_0	0.111	11.783	0.000	0.035	3.539	0.000
Beneficiarios	1.320	6.901	0.000	0.560	3.535	0.000
Analfabetas	4.154	10.466	0.000	2.187	6.077	0.000
Viviendas	4.093	2.753	0.006	2.824	2.353	0.018
No católica	0.749	4.940	0.000	0.313	2.550	0.010
ρ				0.6090	13.984	0.000
		Valor	<i>p</i> -valor		Valor	<i>p</i> -valor
I de Moran		0.305	0.000		-0.013	0.637
		Valor			Valor	
AIC		-1060.929			-1197.419	
BIC		-1036.965			-1169.461	
R_{adj}^2		0.5037				
$R_{Nagelkerke}^2$					0.652	

El modelo de retardo espacial representa una mejora con respecto al modelo de regresión lineal por mínimos cuadrados de acuerdo al AIC y al BIC. Así mismo los residuos ahora cumplen con el supuesto de independencia al no ser significativa la prueba del índice de Moran.

Conclusiones

A lo largo de esta tesis se ha perseguido el objetivo de explorar los señalamientos realizados por los trabajos periodísticos, así como los comportamientos político-electorales en Coyoacán, que derivaron en la anulación de la elección y la posterior ratificación del ganador. El análisis se realizó desde una perspectiva geográfica-estadística y cualitativa debido a tres factores esenciales: 1) la naturaleza de los datos examinados, 2) la georreferenciación que acompaña a todo proceso electoral y su relación con los resultados y 3) las limitaciones de sólo utilizar técnicas cuantitativas para el análisis electoral. Con una metodología mixta se buscó encontrar patrones entre los comportamientos geográficos y enlazar lo que se percibió desde la ciudadanía con los resultados del análisis de datos.

Esta tesis evidenció que las técnicas y los modelos descritos en el capítulo 2 permitieron cuantificar varios aspectos del proceso electoral. Sin embargo, quizá su mayor contribución a este trabajo fue la de incluir la variable geográfica en el análisis exploratorio de datos y así descubrir comportamientos no divisables con técnicas y modelos no espaciales. En síntesis, estas técnicas permitieron tratar con datos geográficos, construir variables nuevas a partir de las ya existentes y obtener conocimientos útiles desde éstas. También evidenció que ignorar el aspecto geográfico de los datos, no sólo excluye información relevante para el estudio, sino que interfiere con los resultados obtenidos por medio de técnicas no espaciales. En nuestro caso, permitieron divisar el cambio en las configuraciones de poder y el reparto territorial en la alcaldía de Coyoacán, identificar espacios de vulnerabilidad, cuantificar los niveles de competitividad a distintas escalas, obtener una aproximación de la distribución de los beneficiarios y agregar la variable espacial en el análisis de regresión mejorando los resultados iniciales.

La evidencia confirma que durante el proceso electoral 2018 fue posible observar actos de compra y coacción del voto en la demarcación, como lo señala la encuesta analizada en el capítulo 3. Arrojando un indicio de la magnitud de los actos reportados por los trabajos periodísticos²¹. Por ello, en Coyoacán fue posible con base a los datos identificar espacios de marginación y concentración de los votantes que registraron altos niveles de competitividad electoral. Esto es, secciones electorales con índices de desarrollo bajo, con una cantidad de votantes considerablemente alta y que

²¹Por lo menos en algunas secciones electorales de dos áreas identificadas a nivel espacial, afectando con ello el resultado de la votación.

estuvieron duramente disputadas en la contienda. Geográficamente estas secciones son contiguas, lo que señala la importancia de la espacialidad en los estudios electorales y la influencia de la posición en los resultados.

En específico, la información de pasadas elecciones reveló que en Coyoacán desde el proceso electoral 2009, la elección para alcalde se había definido sólo entre los partidos del PRD y del PAN. En el 2015 se pudo observar un anuncio de cambio de poderes dado que las principales fuerzas políticas se habían debilitado. Así que para el 2018 se llegó con el partido político MORENA con claras oportunidades de ganar la elección. El análisis de los resultados de este proceso señala que los niveles de competitividad habían descendido en comparación con el 2015, por lo que MORENA no sólo iba a ganar la elección, sino que además lo iba a hacer con amplio margen ante cualquiera de sus competidores. Asimismo, la coalición PRD-PAN-MC benefició al partido en el poder ayudándolo a mantener su estatus al superar a la coalición contraria MORENA-PT-PES, pero sin dejar implicar una contienda con altos niveles de competitividad en secciones electorales estratégicas. Más en concreto, se tiene un contexto de cambio de poderes, más de la mitad del territorio de Coyoacán con niveles de competitividad altos y muy altos, una disminución de preferencia electoral hacia la primera fuerza política, que por primera vez dentro del periodo estudiado perdía ante una nueva opción en ascenso, y una demarcación con marcada separación entre los estratos de desarrollo social.

Lo anterior, es evidencia de un contexto complejo que suele estar acompañado de actos de compra y coacción del voto de acuerdo a los estudios previos revisados.²² Análogamente del análisis de regresión para explicar el voto del PRD, se remarca que la espacialidad de los procesos electorales sí afecta los modelos no espaciales, como lo pudimos observar con la regresión múltiple, por lo que en concordancia con lo que dice Vilalta, el uso de técnicas espaciales con datos electorales es recomendable. En cuanto a los resultados de estos modelos, la preferencia del PRD presentó clústeres espaciales similares a los encontrados en los capítulos 3 y 4 (ver figuras 5.2 y 5.3 del capítulo 5), lo que nos habla de espacios en donde el PRD tuvo una fuerte presencia electoral y en donde perdió fuerza política o nunca la tuvo. Aunado a esto, la variable de Beneficiarios resultó estar correlacionada positivamente con la preferencia de los partidos del PRD y MORENA, y negativamente con la del PAN y otros partidos menores. Más aún, la variable de beneficiarios resultó ser un predictor estadísticamente significativo (en ambos modelos de regresión) para la preferencia del voto del PRD en Coyoacán, lo que sugiere una relación positiva beneficiosa para el partido. Por supuesto, no se puede concluir con estos resultados que el PRD haya

²² Estos estudios se detallan en el capítulo 3.

utilizado el programa social “A tu lado” con fines electorales, ni que los beneficiarios necesariamente voten por el PRD. Pero sí establece un indicio cuantitativo de la relación positiva entre ambas variables, que junto con los resultados de los espacios de vulnerabilidad construidos y el contexto político electoral descrito, sugieren una atmósfera propicia para actos de clientelismo que debería de ser considerado en la declaración de validez de los comicios en casos como los de Coyoacán.

Para cerrar esta tesis es preciso mencionar algunas cuestiones de interpretación derivadas de nuestro análisis. Si bien, se trabajó en su mayoría a una escala seccional y los resultados son válidos a esta escala, evitamos en todo momento inferir sobre comportamientos a niveles individuales. Hacer esto conlleva a caer en la famosa falacia ecológica. Por ejemplo, si bien los resultados indican una relación positiva entre los beneficiarios y la preferencia electoral hacia el PRD a nivel seccional, puede no ser el caso a nivel individual. Para intentar superar esta barrera se incluyó datos de una encuesta para obtener una mirada a un nivel de análisis “micro-sociológico”. No obstante, este trabajo debe entenderse como parte de un análisis que puede profundizarse y no como un trabajo concluyente.

Finalmente, debe subrayarse el papel que las matemáticas aplicadas y la computación jugaron en esta tesis. Pues no basta con analizar los datos extraídos de las fuentes de información. Se debe construir variables nuevas para cuantificar fenómenos sociales difíciles de observar, modelos matemáticos para explicar los comportamientos electorales, programas computacionales para el tratamiento de datos y representaciones gráficas que incluyen la elaboración de mapas temáticos. Sin mencionar, el conocimiento necesario para la correcta interpretación de éstos. Por lo que estas disciplinas son esenciales en problemáticas sociales, como las que se expresan con frecuencia en nuestro país. Sin temor a equivocarnos las matemáticas aplicadas pueden llegar a conformar herramientas para el análisis de procesos electorales dudosos, aportando elementos e información útiles para los tomadores de decisiones, mejorando las herramientas y los procesos de análisis de la democracia nacional.

Bibliografía

- Acevedo, I. & Velásquez Ceballos, H. (2008) Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Ecos de Economía: A Latin American Journal of Applied Economics*, 12(27), pp.9-2.
- Alemán, V., 2018. Anulan elección en alcaldía de Coyoacán donde ganó Negrete. *Excelsior*, 22 Septiembre.
- Andrienko, G. & Andrienko, N. (2006) *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Berlin: Springer Science & Business Media.
- Anselin, L. (1995) Local indicators of spatial association-lisa. *Geographical analysis*, 27(2), pp.93-115.
- Anselin, L. (1999) Interactive techniques and exploratory spatial data analysis. *Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. Geographical Information Systems: principles, techniques, management and applications*, (1), pp.251-264.
- Anselin, L. (2005) *Exploring spatial data with geodtm: a workbook* [Online]. University of Illinois. Disponible en: <https://geodacenter.asu.edu/system/files/geodaworkbook.pdf> [Visitado el 15 de Mayo del 2019].
- Anselin, L. (2012) From SpaceStat to CyberGIS: Twenty years of spatial data analysis software. *International Regional Science Review*, 35(2), pp.131-57.
- Anselin, L. (2013) *SPATIAL ECONOMETRICS: METHODS AND MODELS*. Springer Science & Business Media.
- Anselin, L., K.Bera, A., Florax, R. & Yoond, M.J. (1996) Simple diagnostic tests for spatial dependence. *Regional science and urban economics*, 26(1), pp.77-114.
- Autores, V. (2001) *A companion to theoretical econometrics*. Oxford: Blackwell.
- Baqueiro, M.J. (2014) *Efectos electorales de programas sociales en México*. Tesis de Licenciatura. Ciudad de México: Centro de Investigación y Docencia Económicas.
- Behrens, J.T. (1997) Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), pp.131-60.
- Bolaños, Á., 2018. Rechaza Morena programas clientelares para Coyoacán. *La Jornada*, 17 Noviembre.
- Borrego, J. (2018) *Modelos de Regresión para Datos Espaciales*. Tesis de grado. Sevilla, España: Universidad de Sevilla.
- Briggs,. (2019) *The Standard Regression Model and its Spatial Alternatives*. [Online]. Disponible en: <http://www.utdallas.edu/~briggs/> [Visitado el 12 Mayo 2019].
- Casella, & Berger, R.L. (2002) *Statistical Inference*. 2nd ed. Belmont, California, United States of America: Brooks/Cole Cengage Learning.
- Corona, S., 2018. Regresan alcaldía de coyoacán a Manuel Negrete. *EL Economista*, 30 Septiembre.
- Darmofal, D. (2006) Spatial econometrics and political science. *n Annual Meeting of Southern Political Science Association*, pp.5-7.

- De Smith, M.J., Goodchild, M.F. & Longley, P.A. (2018) *Geospatial Analysis - A comprehensive guide - 2018* [Online]. Disponible en: <https://www.spatialanalysisonline.com/HTML/index.html> [Visitado el 1 Abril del 2019].
- De Smith, M.J., Longley, & Frank,. (2007) *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Leicester: Troubador publishing ltd.
- Esri. (2019a) *Esri Support GIS Dictionary* [Online]. Disponible en: <https://support.esri.com/es/other-resources/gis-dictionary> [Visitado el 11 Junio del 2019].
- Esri. (2019b) *Cómo funciona Análisis de cluster y de valor atípico (I Anselin local de Moran)* [Online]. Disponible en: <https://pro.arcgis.com/es/pro-app/tool-reference/spatial-statistics/h-how-cluster-and-outlier-analysis-anselin-local-m.htm> [Visitado el 2 Febrero del 2019].
- Fabozzi, F.J., Focardi, S.M., Rachev, S.T. & Arshanapalli, B.G. (2014) *The basics of financial econometrics: Tools, concepts, and asset management applications*. New Jersey: John Wiley & Sons.
- Forest, B. (2017) Electoral geography: From mapping votes to representing power. *Geography Compass*, 12(1), pp.1-17.
- Getis, A. (1999) Spatial statistics. *Geographical information systems*, 1, pp.239-51.
- Gobierno de la Ciudad de México. (2019a) *SIDESO* [Online]. Disponible en: <http://www.sideso.cdmx.gob.mx/documentos/2018/padrones/delegaciones/coyoacan/A%20tu%20lado.pdf> [Visitado el 21 Marzo del 2019].
- Gobierno de la Ciudad de México. (2019b) *SIDESO* [Online]. Disponible en: <http://www.sideso.cdmx.gob.mx/index.php?id=785> [Visitado el 2 Marzo del 2019].
- Goodchild, M.F. (1986) *Spatial Autocorrelation*. Norwich : Geo books.
- Griffith, D.A. (2009) *Spatial Autocorrelation* [Online]. Disponible en: <https://booksite.elsevier.com/brochures/hugy/SampleContent/Spatial-Autocorrelation.pdf> [Visitado el 3 Marzo del 2019].
- Hallisey, E. et al. (2017) Transforming geographic scale: a comparison of combined population and areal weighting to other interpolation methods. *International Journal of Health Geographics*, 16(29), pp.16-29.
- Helbich , M., Leitner , M. & Kapusta,. (2012) Geospatial examination of lithium in drinking water and suicide mortality. *International Journal of Health Geographics*, 11(1), pp.11-19.
- Hernández, S., 2018. A pesar de restricción, Coyoacán defiende reparto de tarjetas. *EL UNIVERSAL*, 30 Mayo.
- Hernández, V. (2015) Análisis geoespacial de las elecciones presidenciales en México, 2012. *EURE*, 41(122), pp.185-207. Disponible en: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0250-71612015000100009 [Visitado el 25 Mayo del 2019].
- Hevia de la Jara,. (2010) Uso político de programas sociales y nuevos intermediarios institucionales: el Programa Progreso/Oportunidades en el sur de Veracruz. *Desacatos*, (34), pp.119-32.
- Hinkle, D.E., Wiersma, , & Stephen, G.J. (2002) *Applied Statistics for the Behavioral Sciences*. 1st ed. Independence, Kentucky, USA: Wadsworth Publishing.

- Ibarra, J.F., Somuano Ventura, M.F. & Yunuen Ortega, R. (2009) La competencia electoral y su impacto en el desempeño hacendario de los municipios en México. *Foro Internacional*, XLVI(3), pp.465-92.
- INE. (2019) *Glosario Electoral* [Online]. Disponible en: <https://centralectoral.ine.mx/2018/06/08/glosario-electoral-seccion-electoral/> [Visitado el 12 Junio 2019].
- Jonathan, M. (2014) *Clientelismo político y procesos electorales: El voto de los marginados en tijuana, bc 2004-2013*. Tesis de Maestría. Tijuana: Colegio de la Frontera Norte Colegio de la Frontera Norte.
- Krehbiel, E. (1916) Geographic Influences in British Elections. *Geographical Review*, 2(6), pp.419-32.
- LeSage, J.P. (1999) *Spatial Econometrics using MATLAB* [Documento electrónico]. Toledo: University of Toledo Disponible en: <https://pdfs.semanticscholar.org/9c6d/e738660a121ed2d16c934e29cda07f93b19c.pdf> [Visitado el 10 Marzo del 2019].
- Manterola, A.L. (2007) *Métodos de estadística espacial para evaluar la influencia de factores medioambientales sobre la incidencia y mortalidad por cáncer*. Tesis de Doctorado. Girona: Universitat de Girona.
- Martinez, W.L., Martinez, A.R., Solka, J. & Martinez, A. (2005) *Exploratory Data Analysis with MATLAB*. London, UK: CHAPMAN & HALL/CRC.
- Martori, J.C. & Hoberg, K. (2008) Nuevas técnicas de estadística espacial para la detección de clusters residenciales de población inmigrante. *Scripta Nova*, XII(263).
- Matthews, S.A. (2006) *GeoDa and Spatial Regression Modeling* [Online]. Disponible en: http://ibis.geog.ubc.ca/~brian/workshop/GeoDa_Spatial_Regression.pdf [Visitado el 2 Mayo del 2019].
- Miles, J. (2005) R Squared, Adjusted R Squared. *Encyclopedia of Statistics in Behavioral Science*, 4, p.1655–1657.
- Molina, J. (2014) *Clientelismo político y procesos electorales: El voto de los marginados en tijuana, bc 2004-2013*. Tesis de Maestría. Tijuana, BC.: Colegio de la Frontera Norte.
- Monzón, N. (2009) Geografía electoral. Consideraciones teóricas para el caso argentino Cuadernos de Geografía. *Revista Colombiana de Geografía*, (18), pp.119-28.
- Moran, P. (1950) Notes on continuous stochastic phenomena.. *Biometrika* , 3(1/2), pp.17-23.
- Muñoz, E.H. (2006) El clientelismo en México: los usos políticos de la pobreza. *Espacios Públicos*, 9(17), pp.118-40.
- Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), pp.691-92.
- Naldi, M. & Flamini, M. (2014) *The CR4 Index and the Interval Estimation of the Herfindahl-Hirschman Index: An Empirical Comparison* [Online]. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2448656 [Visitado el 10 Marzo del 2019].

- Neria, J.A.T.B.y.B.L. (2015) Competencia electoral municipal. El caso de los municipios de Hidalgo (1996-2014). *Revista Mexicana de Estudios Electorales* , (15), pp.56-77.
- Neter, J., Kutner, M.H., Nachstheim, C.J. & Wasserman, W. (2005) *Applied linear statistical models*. McGraw-Hill Irwin.
- O'Loughlin, J. & Anselin, L. (2008) Bringing geography back to the study of international relations: Spatial dependence and regional context in Africa, 1966–1978. *International Interactions Empirical and Theoretical Research in International Relations*, 17(1), pp.29-61.
- Olaya, V. (2019) *Sistemas de Información Geográfica* [Online]. Disponible en: <https://volaya.github.io/libro-sig/> [Visitado el 16 Junio del 2019].
- Ord, A.G.a.J.K. (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), pp.189-206.
- Pace, J.P.L.y.K. (2009) *Introduction to Spatial Econometrics*. Boca Raton, Florida: Chapman and Hall/CRC.
- Pedregosa, F. et al. (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, (12), pp.2825-30.
- Peña, D. (2012) *Regresión y diseño de experimentos*. Alianza.
- Peña, D. (2019) *LAS MATEMÁTICAS EN LAS CIENCIAS SOCIALES* [Online]. Disponible en:<http://www.encuentros multidisciplinarios.org/Revistan%C2%BA23/Daniel%20Pe%C3%B1a%20S%C3%A1nchez%20de%20Rivera.pdf> [Visitado el 15 Abril del 2019].
- Prescott, J.R.V. (1959) THE FUNCTION AND METHODS OF ELECTORAL GEOGRAPHY. *Annals of the Association of American Geographers*, 49(3), pp.296-304.
- Raccanello, , K. & Cerón Vargas , j.A. (2018) Índice de desarrollo social de la ciudad de México como herramienta de focalización de la política social. *Retos de la Dirección*, 12(2), pp.64-86.
- Rey, L.A.a.S. (1991) Properties of tests for spatial dependence in linear regression models. *Geographical analysis* , 23(2), pp.112-31.
- Rhoades, S.A. (1993) The Herfindahl-Hirschman index. *Federal Reserve Bulletin*, (Mar), pp.188-89.
- Rodríguez, E. (2015) La delimitación cuantitativa de divisaderos de observación etnográfica: una herramienta metodológica para el estudio antropológico del estado. *Nueva antropología*, 28(83), pp.73-97.
- Rozas , V. & Camarero, J. (2005) Técnicas de análisis espacial de patrones de puntos aplicadas en ecología forestal. *Investigación agraria. Sistemas y recursos forestales*, (14), pp.79-97.
- Ruiz , C. & Animal Político, 2018. Delegación Coyoacán duplica gasto en programas sociales y aumenta beneficiarios en pleno proceso electoral. *Verificado*, 18 Mayo.
- Ruiz, , 2018. Ofrecieron despensas, dinero y tabletas para promover el voto por el PRD en Coyoacán. *Animal Político*, 25 Septiembre.
- Sánchez , L.L. (2006) Métodos para el análisis espacial. Una aplicación al estudio de la geografía de la pobreza. In *II Congreso de la Asociación Latinoamericana de Población*. Guadalajara, 2006.

- Sánchez Perales , C. (2012) *Análisis y visualización de información mediante estadística espacial: Geostads*. Tesis de Licenciatura. Universidad Nacional Autónoma de México.
- Santana, R. et al. (2019) *MATEDA: A suite of EDA programs in Matlab* [Online]. Disponible en: <http://www.cs.bham.ac.uk/~axk/EC/eda/Mateda2.0/Mateda2.0-UserGuide.pdf> [Visitado el 9 Marzo del 2019].
- Schroter, B. (2010) Clientelismo político: ¿existe el fantasma y cómo se viste? *Revista mexicana de sociología*, 72(1), pp.141-75.
- Scolastika, M., Wardono, Masrukan & Fatkhurokh. (2017) THE ArcVIEW AND GeoDa APPLICATION IN OPTIMIZATION OF SPATIAL REGRESSION ESTIMATE. *Journal of Theoretical and Applied Information Technology*, 95(5), pp.1102-15.
- Smith, T.E. (2019) *SPATIAL WEIGHT MATRICES* [Online]. Disponible en: https://www.seas.upenn.edu/~ese502/lab-content/extra_materials/SPATIAL%20WEIGHT%20MATRICES.pdf [Visitado el 1 Abril 2019].
- Sonnleitner, W. (2013) Explorando las dimensiones territoriales del comportamiento político: reflexiones teórico-metodológicas sobre la geografía electoral. *Estudios Sociológicos de El Colegio de México*, XXXI, pp.97-142.
- Sprent, P. & Smeeton, N.C. (2007) *Applied Nonparametric Statistical Methods*. Chapman and Hall/CRC.
- Terron, S.L. & Dillion Soares, G.A. (2008) Dois Lulas: a geografia eleitoral da reeleição (explorando conceitos, métodos e técnicas de análise geoespacial). *Opinião Pública*, 14(2).
- Tiebei, L., Pullar, D., Corcoran, J. & Stimson, R. (2007) A comparison of spatial disaggregation techniques as applied to population estimation for South East Queensland (SEQ), Australia. *Applied GIS*, 3(9), pp.1-16.
- Tobler, W.R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, pp.234-40.
- Tosoni, M.M. (2007) Notas sobre el clientelismo político en la ciudad de México. *Perfiles latinoamericanos*, 14(29), pp.47-69.
- Tukey, J.W. (1997) *Exploratory Data Analysis*. ADDISON-WESLEY PUBLISHING COMPANY.
- Varela, D.S. & Ascensio, D.C. (2018) *Resultados de la encuesta: "dinámicas y prácticas político-electorales y clientelares en la ciudad de México. el caso de los pedregales de coyoacán"*. Reporte. Ciudad de México: Universidad Nacional Autónoma de México.
- Vassilvitskii, & Arthur, D. (2007) k-means++: The advantages of careful seeding. *In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp.1027-35.
- Vilalta, C.J. (2006) Sobre la espacialidad de los procesos electorales urbanos y una comparación entre las técnicas de regresión OLS y SAM. *Estudios Demográficos y Urbanos*, 21(1), pp.83-122.
- Vilalta, C.J. (2007) *Compra y coacción del voto en México: variaciones estatales y diferencias de opinión entre beneficiarios y no beneficiarios de programas sociales*. Documento de trabajo. Programa de las Naciones Unidas para el Desarrollo.

- Vilalta, C.J. (2008) ¿Se pueden predecir geográficamente los resultados electorales? Una aplicación del análisis de clusters y outliers espaciales. *ESTUDIOS DEMOGRÁFICOS Y URBANOS*, 23(3), pp.571-613.
- Wackerly, D.D., Mendenhall III, W. & Scheaffer, R.L. (2008) *Estadística Matemática con Aplicaciones*. 7th ed. CENGAGE Learning.
- Wu, Jelinski , D.E. & Jianguo. (1996) The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, 11(3), pp.129-40.
- Yan, X. & Gang Su, X. (2009) *Linear Regression Analysis: Theory and Computing*. Toh Tuck Link, Singapore: World Scientific.
- Zavala, R.G.B. (2012) Génesis de la geografía electora. *Génesis de la geografía electoral. Espacialidades. Revista de temas contemporáneos sobre lugares, política y cultura*, 2(1), pp.81-95.

Anexos

Anexo 1. Glosario

Notación matemática

$=$: Igual.
 \neq : Desigual o diferente.
 \sim : Aproximadamente, se distribuye como.
 \forall : Para todo.
 \leq : Menor o igual que.
 \geq : Mayor o igual que.
 $<$: Menor que.
 $>$: Mayor que.
 \in : Pertenece a, está en.
 \cap : Intersección de conjuntos.
 \cup : Unión de conjuntos.
 \emptyset : Conjunto vacío.
 \rightarrow : Tiende a, se aproxima a.
 \mathbb{N} : Conjunto de números naturales.
 \mathbb{Z} : Conjunto de números enteros.
 \mathbb{Q} : Conjuntos de números racionales.
 \mathbb{R} : Conjunto de números reales.

Glosario de términos

C

Ciencia regional: Es el campo de estudio multidisciplinario que se encarga de estudiar los procesos socioeconómicos en un contexto regional.

Colonia: Es la denominación que se le da a los vecindarios o barrios en México.

Competitividad electoral: Capacidad de competir de los partidos políticos o candidatos para alcanzar la mayor cantidad de votos en un proceso electoral.

D

Dependencia espacial: Es la relación de las unidades espaciales debido estrictamente a su posición geográfica.

Distrito electoral: Es una de las divisiones electorales en las que México se encuentra repartido. En nuestro país existen 300 distritos uninominales en los cuales se elige a un diputado y están diseñadas por medio de un proceso llamado distritación para que cada una tenga más o menos la misma cantidad de población.

F

Falacia ecología: Es una falacia deductiva que surge al inferir sobre la naturaleza de los individuos a partir del grupo al que pertenecen. Por ejemplo, inferir que los individuos de cierta región votaron por un partido político solo porque en esta región gano dicho partido.

I

Indicador: Dato o información que sirve para conocer o medir la intensidad de un fenómeno.

Índice de desarrollo social: En México es un indicador utilizado para medir el grado de desarrollo de la población en unidades territoriales como municipios, colonias, manzanas, etcétera.

M

Manzana: Se le llama manzana a al territorio conformado por un grupo de viviendas, edificios o construcciones y de forma poligonal generalmente delimitado por cuatro calles.

P

Partido político: Es la organización ciudadana que promueve su participación y acceso al ejercicio del poder público, de acuerdo a las ideas y principios que estos postulan.

Programa social: Es una serie de acciones (comúnmente desde el gobierno) dirigidas a su población objetivo con la finalidad de servir de apoyo ya sea con recursos monetarios, en especie o de alguna otra forma.

S

Sección electoral: Es una subdivisión territorial de los 300 distritos electorales y la unidad básica de la geografía electoral en nuestro país.

U

Unidad geográfica/espacial/ de área: Es la división territorial de un país o territorio de acuerdo a diferentes criterios de división (económicos demográficos, geográficos, etcétera), que pueden ser desde estados, municipios, colonias, secciones electorales entre muchos otros.

Anexo 2. Software libre utilizado

Qgis

Qgis es un sistema de información geográfica, software libre y de código abierto. Es uno de los proyectos de Open Source Geospatial Foundation (OSGeo) disponible para distintos sistemas operativos entre ellos Windows, MacOS y Linux. Todos los mapas que se presentaron en este trabajo fueron realizados en la versión de Qgis 2.8.

GeoDa

Geoda es un software libre y de código abierto desarrollado por Luc Anselin y su equipo con el objetivo de proveer un programa amigable para los usuarios en el cual se pueda realizar análisis exploratorios de datos y que funcione como una introducción al análisis de datos espaciales. Al igual que Qgis, GeoDa está disponible para los sistemas operativos de Windows, MacOS y algunas distribuciones de Linux. Actualmente cuenta con alrededor de 300,000 usuarios activos principalmente en los Estados Unidos, China y Brasil. Los mapas LISA fueron elaborados en este software en la versión 1.12.

Python y R

Python y R son lenguajes de programación, muy conocidos actualmente para el desarrollo de aplicaciones y el análisis de datos. Ambos cuentan con una gran comunidad que introduce nuevos elementos cada cierto tiempo. Todas las gráficas que aquí se presentaron fueron realizados en Python y R y el manejo de datos, su limpieza y transformación fueron realizadas en Python. Los análisis estadísticos como la regresión lineal múltiple fueron realizadas en R, así como el modelo de retardo espacial y las pruebas de autocorrelación. Se trabajó con Python en un notebook de Jupyter y en R en el software de RStudio. Ambos entornos son de licencia libre.

Simulador y códigos utilizados

El siguiente simulador fue desarrollado en Processing bajo el lenguaje de programación Python. El programa simula una cuadrícula de 25x25 para cuatro diferentes configuraciones. Cada cuadrado en la cuadrícula es inicializado con diferentes valores y colores que denotan el parecido entre ellas. El programa también calcula el índice de Moran y la c de Geary para cada configuración. Los resultados de este programa fueron mostrados en el capítulo 2 (véase: figuras: 2.6, 2.7, 2.8 y 2.9).

```

1. """
2. Simulador del índice de Moran
3. Utiliza la I de Moran dada por Moran en su artículo y la versión que
4. incluye la matriz W construida bajo un vecindario tipo Torre.
5. Utiliza una retícula de 25x25
6. """
7. from random import uniform
8.
9. # Tamaño de la retícula
10. nCols=25
11. nRows=25
12.
13. # Inicialización
14. def setup():
15.     background(255)
16.     global nCols,nRows,grid
17.     size(500,550)
18.     cond=0
19.     evalua(cond)
20.
21. # Dibuja nuestra retícula
22. def draw():
23.     global nCols,nRows,grid
24.     #valor=int(random(0,4))
25.     #evalua(valor)
26.
27.
28. def mousePressed():
29.     valor=int(random(0,4))
30.     evalua(valor)
31.
32. def evalua(condicion):
33.     background(255)
34.     grid=makeGrid()
35.     media=0
36.     value=0
37.     id=0
38.     for i in range(nRows):
39.         for j in range(nCols):
40.             c,value=condition(condicion,i,j)
41.             grid[i][j]=Cell(j*20,i*20,20,20,c,value,id)
42.             grid[i][j].colour=c
43.             grid[i][j].display()
44.             media+=value
45.             id+=1
46.     res,lista=vecindario(grid)
47.     media=media/(nRows*nCols)
48.     i_0,i_1,c=idemoran(grid,media,res,lista)
49.     #print((lista))
50.     s = "I= "+str(round(i_1,4))+" , "+"c= "+str(round(c,4))

```

```

51. fill(230,0,38);
52. textSize(20)
53. text(s, 125, 515,500, 50)
54.
55.#Reticula de 25x25
56.def makeGrid():
57.   global nCols, nRows
58.   grid=[]
59.   for i in range(nRows):
60.     grid.append([])
61.     for j in range(nCols):
62.       grid[i].append(0)
63.   return (grid)
64.
65.
66.#Clase célula
67.class Cell():
68.   def __init__(self,X,Y,W,H,C,L,ID):
69.     self.x=X
70.     self.y=Y
71.     self.w=W
72.     self.h=H
73.     self.c=C
74.     self.l=L
75.     self.id=ID
76.   def display(self):
77.     stroke(0)
78.     fill(self.c)
79.     rect(self.x,self.y,self.w,self.h)
80.
81.#Crea aleatoriedad o patrones
82.def condition(p,x,y): #p es la condición, x,y son las coordenadas
83.   #1 Condición inicial aleatoria
84.   #0 Condición de patrón
85.   v=0
86.   #Versión totalmente aleatoria, I=0
87.   if p==0:
88.     co=uniform(0,1)
89.     if co<0.3:
90.       #Color blanco valor 0
91.       v=co*10
92.       co=255
93.     elif (co>=0.3 and co<0.6):
94.       #Color gris valor 5
95.       v=co*10
96.       co=200
97.     else:
98.       #Color negro valor 10
99.       v=co*10

```

```

100.     co=10
101.
102. #Versión no aleatoria. Clúster no muy marcados I>0
103. elif p==1:
104.     if ((x>=3 and x<11) or (x>=14 and x<22)) and ((y>=3 and y<11) or (y>=14 and y<22)):
105.         #Colores negros y grises
106.         v=uniform(0,0.6)
107.         co=255*v
108.     else:
109.         #Colores blancos y grises claros
110.         v=v=uniform(0.66,1)
111.         co=255*v
112.
113. #Versión no aleatoria. Clusters muy marcados I>0.5
114. elif p==2:
115.     if ((x>=3 and x<11) or (x>=14 and x<22)) and ((y>=3 and y<11) or (y>=14 and y<22)):
116.         #Colores negros y grises
117.         v=uniform(0,0.3)
118.         co=255*v
119.     else:
120.         #Colores blancos y grises claros
121.         v=v=uniform(0.66,1)
122.         co=255*v
123. elif p==3:
124.     if (x+y)%2==0:
125.         v=v=uniform(0,0.3)
126.         co=255*v
127.     else:
128.         v=v=uniform(0.66,1)
129.         co=255*v
130.
131. return(co,v)
132.
133.
134.
135.#Calcula la I de Moran
136.def idemoran(reticula,promedio,matrizw,new):
137.     global nCols,nRows
138.     suma1=0 #Primer término
139.     suma2=0 #Segundo término
140.     deno=0#denominador
141.
142.     """
143.     I de Moran con la fórmula original
144.     """
145.
146.     #Calcula el primer término de la expresión
147.     for i in range(nRows):

```

```

148.     for j in range(nCols-1):
149.         suma1+=(reticula[i][j].l-promedio)*(reticula[i][j+1].l-promedio)
150.
151. #Calcula el segundo término de la expresión
152. for i in range(nRows-1):
153.     for j in range(nCols):
154.         suma2+=(reticula[i][j].l-promedio)*(reticula[i+1][j].l-promedio)
155.
156. #Calcula el denominador de la expresión
157. for i in range(nRows):
158.     for j in range(nCols):
159.         deno+=(reticula[i][j].l-promedio)**2
160.
161. I_0=((nRows*nCols)*(suma1+suma2))/((2*nRows*nCols-nRows-nCols)*deno)
162.
163. """
164. I de Moran y c de Geary calculado con la matriz W
165. """
166. #Calculo del numerador y denominador
167. numerador=0
168. denominador=0
169. sumaW=0
170. sumac=0
171. N=nCols*nCols
172. for i in range(N):
173.     for j in range(N):
174.         numerador+=matrizw[i][j]*(new[i]-promedio)*(new[j]-promedio)
175.         sumaW+=matrizw[i][j]
176.         sumac+=matrizw[i][j]*((new[i]-new[j])**2)
177.         denominador+=(new[i]-promedio)**2
178. I_1=(N*(numerador)/(sumaW*(denominador))) #Expresión de la I de Moran
179. c=((N-1)*sumac)/(2*sumaW*(denominador)) #Expresión de la c de Geary
180.
181. """
182. C de Geary
183. """
184.
185.
186.
187. return(I_0,I_1,c)
188.
189.
190.#Función que llena la matriz W con 0 o 1
191.def vecindario(mapa):
192.     global nRows, nCols
193.     N=nRows*nCols #Número de columnas de la matriz W
194.     W=[] #Se crea la matriz W solo con ceros
195.     new=[]

```

```

196. for p in range(N):
197.     W.append([0]*N)
198. index=0
199.
200. #Proceso que evalua un vecindario tipo torre y llena la matriz W con 1's
201. for i in range(nRows):
202.     for j in range(nCols):
203.         new.append(mapa[i][j].l)
204.         #Evalua la esquina superior izquierda
205.         if (i==0 and j==0):
206.             v1=mapa[i+1][j].id
207.             v2=mapa[i][j+1].id
208.             for k in [v1,v2]:
209.                 W[index][k]=1
210.             index+=1
211.
212.         #Evalúa la esquina superior derecha
213.         if ((i==0) and j==(nCols-1)):
214.             v1=mapa[i+1][j].id
215.             v2=mapa[i][j-1].id
216.             for k in [v1,v2]:
217.                 W[index][k]=1
218.             index+=1
219.
220.         #Evalúa la esquina inferior izquierda
221.         if (i==(nRows-1) and j==0):
222.             v1=mapa[i-1][j].id
223.             v2=mapa[i][j+1].id
224.             for k in [v1,v2]:
225.                 W[index][k]=1
226.             index+=1
227.
228.         #Evalúa la esquina inferior derecha
229.         if (i==(nRows-1) and j==(nCols-1)):
230.             v1=mapa[i-1][j].id
231.             v2=mapa[i][j-1].id
232.             for k in [v1,v2]:
233.                 W[index][k]=1
234.             index+=1
235.
236.         #Evalúa la parte superior
237.         if (i==0) and (j>0 and j<(nCols-1)):
238.             v1=mapa[i+1][j].id
239.             v2=mapa[i][j-1].id
240.             v3=mapa[i][j+1].id
241.             for k in [v1,v2,v3]:
242.                 W[index][k]=1
243.             index+=1

```

```

244.
245.     #Evalúa la parte inferior
246.     elif (i==(nRows-1)) and (j>0 and j<(nCols-1)):
247.         v1=mapa[i-1][j].id
248.         v2=mapa[i][j-1].id
249.         v3=mapa[i][j+1].id
250.         for k in [v1,v2,v3]:
251.             W[index][k]=1
252.             index+=1
253.
254.     #Evalúa la parte izquierda
255.     elif (i>0 and i<(nRows-1)) and (j==0):
256.         #Obtiene los valores
257.         v1=mapa[i-1][j].id
258.         v2=mapa[i+1][j].id
259.         v3=mapa[i][j+1].id
260.         #Barre la matriz
261.         for k in [v1,v2,v3]:
262.             W[index][k]=1
263.             index+=1
264.
265.     #Evalúa la parte derecha
266.     elif ((i>0 and i<(nRows-1)) and (j==(nCols-1))):
267.         #Obtiene los valores
268.         v1=mapa[i-1][j].id
269.         v2=mapa[i+1][j].id
270.         v3=mapa[i][j-1].id
271.         #Barre la matriz
272.         for k in [v1,v2,v3]:
273.             W[index][k]=1
274.             index+=1
275.
276.     #Solo evalua el centro
277.     elif (i>0 and j>0) and (i<(nRows-1) and j<(nCols-1)):
278.         #Obtiene los valores
279.         v1=mapa[i-1][j].id
280.         v2=mapa[i+1][j].id
281.         v3=mapa[i][j-1].id
282.         v4=mapa[i][j+1].id
283.         #Barre la matriz
284.         for k in [v1,v2,v3,v4]:
285.             W[index][k]=1
286.             index+=1
287.
288.     return(W,new)
289.
290.

```

Códigos en Python

Códigos de programación en el lenguaje de programación Python. La mayoría de estos códigos utilizan la estructura conocida como “dataframe” por lo que es recomendable estar familiarizado con este lenguaje.

Librerías utilizadas.

```
1. import numpy as np
2. import pandas as pd
3. import matplotlib.pyplot as plt
4. import seaborn as sns
5. import pysal as ps
6. import scipy as sc
7. import statsmodels.formula.api as smf
8. import statsmodels.stats.api as sms
9. import statsmodels.api as sm
10. import statsmodels.stats.stattools as sss
11. import statsmodels.stats.diagnostic as sdd
12. %matplotlib inline
```

Lista de Funciones: “votaciones”, función que calcula el número de votos de cada partido político. Y crea una tabla donde despliega los resultados. “ICE”, función que calcula el índice de competitividad electoral por sección electoral. “irae”, función que calcula el índice de competitividad de Rae por sección electoral. “ganador”, función que obtiene el partido más votado por sección electoral. “totales”, función que calcula la votación total de cada partido político.

```
1. def votaciones_f(df,partido):
2.     totales=[]
3.     nombres=[]
4.     total=0
5.     for i in partido:
6.         totales.append(sum(df[i]))
7.         nombres.append(i)
8.     tabla=pd.DataFrame(nombres)
9.     tabla.columns=["FUERZA POLÍTICA"]
10.    tabla["NUM VOTOS"]=totales
11.    return(tabla)
12.
13. #Competitividad por partidos ICE
14. def ICE(data,lista):
15.     ice=[]
16.     for i in range(len(data)):
17.         z=sorted(list(data[lista].loc[i]),reverse=True)
18.         if z[0]!=0:
19.             ice.append(z[1]/z[0])
20.     return(ice)
21.
```

```

22.#Competitividad IRAE
23.def irae(data,names,t):
24. lista=[]
25. IRAE=[]
26. #Calcula los cocientes de los partidos o candidatos
27. for i in names:
28. lista.append(list((round(df[i]/df[t],4)**2)))
29. #Calcula el índice RAE
30. for j in range(len(lista[0])):
31. a=0
32. for h in range(len(lista)):
33. a+=lista[h][j]
34. IRAE.append(round(1-round(a,4),4))
35. return(IRAE)
36.
37.
38.def ganador(data,partido):
39. fina=[]
40. for i in range(len(data)):
41. aux=[]
42. for j in partido:
43. aux.append([data[j].loc[i],j])
44. final.append(sorted(aux,reverse=True)[0])
45.
46. return([p[1] for p in final])
47.
48.def totales(data,partidos):
49. t=0
50. for i in partidos:
51. t+=sum(data[i])
52. return(t)

```

El siguiente programa fue utilizado para generar todas las combinaciones posibles con las variables dadas pudiendo dejar una variable fija.

```

1.def modelos(variables,dt,obj,fija):
2. variables2=[]
3. for j in potencia(variables)[1:]:
4. if fija in j:
5. variables2.append(j)
6. modelo=[]
7. for i in variables2:
8. X=sm.add_constant(dt[i])
9. reg=smf.OLS(dt[obj],X,data=dt)
10. res = reg.fit()
11. aic=res.aic
12. bic=res.bic

```

```

13.  jb=sss.jarque_bera(res.resid, axis=0)
14.  bp=ssd.het_breuschpagan(res.resid,res.model.exog)
15.  if jb[1]>=0.05:
16.      if bp[1]>=0.05:
17.          modelo.append([aic,bic,i])
18.  modelo=sorted(modelo)
19.  return(modelo)
20.

```

Códigos en R

Los siguientes códigos fueron programados en el lenguaje de programación R. Y son utilizados para realizar el análisis de regresión espacial y no espacial. Los programas se encuentran en orden en el que son ejecutados. Aquí se realiza la carga del archivo, el modelado y las pruebas estadísticas como la *I* de Moran o las pruebas de homocedasticidad y normalidad.

```

1.library(spdep)
2.library(maptools)
3.library(sf)
4.library('rgdal')
5.library(lmtest)
6.library(fBasics)
7.
8.df<-readShapePoly(paste(ruta,'/mapa_final.shp',sep=""))
9.require(sp)
10.df@data[!complete.cases(df@data),]
11.
12.
13.#Regresión lineal por mínimos cuadrados ordinarios
14.model<-lm(PRDval2018~X.Benefici+casas_sin_+Pnocatolic+Analfabeta, data=df@data)
15.
16.#Resumen con R^2, R^2 ajustada, coeficientes y prueba F
17.summary(model)
18.
19.#Checamos que la media de los residuos sea cero
20.mean(model$resid)
21.
22.#Prueba de normalidad
23.#Null Hypothesis: Skewness and Kurtosis are equal to zero
24.jarqueberaTest(model$resid)
25.
26.#Prueba de heterocedasticidad
27.bptest(PRDval2018~Analfabeta+casas_sin_+Pnocatolic+X.Benefici, varformula = NULL, studentize = TRUE, data=df@data)
28.
29.#Construimos vecindarios tipo reina
30.queen<-poly2nb(df, queen=TRUE)
31.
32.W<-nb2listw(queen, style="W", zero.policy=TRUE)
33.summary(W)
34.

```

```

35.#Imprimimos el grafo de conexión de Coyoacán
36.plot(W,coordinates(df))
37.
38.#Prueba de Moran para el modelo OLS
39.moran<-lm.morantest(model, W, alternative="two.sided")
40.print(moran)
41.
42.#Prueba del Multiplicador de Lagrange
43.LM<-lm.LMtests(model, W, test="all")
44.print(LM)
45.
46.#Modelo de retardo espacial
47. modelsar<-lagsarlm(PRDval2018~Analfabeta+casas_sin_+Pnocatolic+X.Benefici, data=df@data, W,)
48.  summary(modelsar,Nagelkerke = TRUE)
49.
50.#Prueba de Moran para el modelo de autocorrelación espacial
51.moran.test(modelsar$residuals,W)
52.#Prueba de Moran para el modelo OLS
53.moran.test(model$residuals,W)
54.
55.#JarqueBera test
56.jarqueberaTest(modelsar$resid)
57.
58.#Heterocedasticidad
59.bptest.sarlm(modelsar, varformula=NULL, studentize = TRUE, data=list())
60.
61.VIF(model)

```