

**UNIVERSIDAD AUTÓNOMA DE COAHUILA
CENTRO DE INVESTIGACIONES SOCIOECONÓMICAS**

TESIS

TESIS
2021

NOEL ABRAHAM VELAZQUEZ VIVEROS

**“DIFERENCIA SALARIAL EN CIUDADES DE MÉXICO,
UN ANÁLISIS ESPACIAL;
2009 Y 2019”**



UNIVERSIDAD AUTÓNOMA DE COAHUILA

CENTRO DE INVESTIGACIONES SOCIOECONÓMICAS

MAESTRIA EN ECONOMÍA REGIONAL

TESIS

**“Diferencia salarial en ciudades de México, un análisis espacial;
2009 y 2019”**

que se presenta como requisito parcial para obtener
el grado de Maestro en Economía Regional

NOEL ABRAHAM VELAZQUEZ VIVEROS

Comité Evaluador:

Director: Dr. Luis Gutiérrez Flores

Codirector: Dr. David Castro Lugo

Lectores: Dra. Reyna Elizabeth Rodríguez Pérez

Saltillo, Coahuila
Julio de 2021

CONTENIDO

INTRODUCCIÓN	1
CAPÍTULO I: MARCO TEÓRICO – CONCEPTUAL	2
INTRODUCCIÓN DEL CAPÍTULO	2
1.1 MERCADO LABORAL.....	3
1.1.1 <i>Funcionamiento del mercado</i>	4
1.1.2 <i>Salarios</i>	5
1.1.3 <i>Diferenciación salarial</i>	7
1.2 AGLOMERACIÓN	10
1.2.1 <i>Economías de localización y de urbanización</i>	11
1.2.2 <i>La Nueva Geografía Económica</i>	12
1.2.3 <i>Relación con la fuerza laboral</i>	13
1.3 CRECIMIENTO ECONÓMICO REGIONAL DESIGUAL.....	14
1.4 MEDIDA Y PERTINENCIA DE ESTUDIO	17
CONCLUSIONES DEL CAPÍTULO	18
CAPÍTULO II: MARCO METODOLÓGICO	20
2.1. ANÁLISIS ESPACIAL.....	20
2.2. DATOS ESPACIALES	21
2.2.1. <i>Análisis Exploratorio de datos espaciales</i>	23
2.3. AUTOCORRELACIÓN ESPACIAL.....	24
2.3.1. <i>Estadístico I-Moran</i>	24
2.3.2. <i>Estadístico C-Geary</i>	27
2.4. MATRICES DE PESOS ESPACIALES	28
2.5. MODELOS DE REGRESIÓN ESPACIAL	30
2.5.1. <i>Modelo de Mínimos Cuadrados Ordinarios</i>	31
2.5.2. <i>Modelo autorregresivo espacial (SAR)</i>	31
2.5.3. <i>Modelo del Error Espacial (SEM)</i>	32
2.5.4. <i>Modelo espacial autorregresivo combinado (SAC)</i>	33
2.5.5. <i>Modelo de rezago espacial en X (SLX)</i>	34
2.5.6. <i>Modelo espacial de Durbin (SDM)</i>	35
2.6. INTERPRETACIÓN DE COEFICIENTES O EFECTOS.....	36
2.7. DATOS Y VARIABLES	36

2.7.1. ENOE	37
2.7.2. Variables	40
2.7.3. Construcción de variables	41
2.7.3.1. Años promedio de escolaridad	42
2.7.3.2. Tasa de la población con posgrado	42
2.7.3.3. Tasa de población con primaria incompleta	43
2.7.3.4. Edad promedio	43
2.7.3.5. Tasa de población joven	43
2.7.3.6. Tasa de población en tercera edad	44
2.7.3.7. Tasa Desocupación	44
2.7.3.8. Tasa de trabajadores en sector informal	44
2.7.3.9. Tasa de trabajadores en industria manufacturera	45
2.7.3.10. Salario promedio por hora	46
CAPÍTULO III: RESULTADOS	49
3.1. ANÁLISIS ESPACIAL DE LAS VARIABLES	53
3.1.1. <i>análisis visual</i>	53
3.1.2 <i>Pruebas de autocorrelación espacial</i>	62
3.2. ESTIMACIÓN DE MODELOS	71
3.2.1. <i>Modelo de mínimos cuadrados ordinarios</i>	72
3.2.2. <i>Modelos de regresión espacial</i>	74
3.2.3. <i>Contraste de resultados con el tercer trimestre del 2009</i>	80
CONCLUSIONES	85
REFERENCIAS	91
ANEXOS	97

INTRODUCCIÓN

El objetivo general que se pretende en esta investigación es identificar si existen fenómenos espaciales de propagación del ingreso en las ciudades de México, en 2009 y 2019, es decir que ciudades que presenten altos niveles de ingreso se rodean de ciudades con la misma condición. El objetivo se demuestra mediante la estimación de un modelo de econometría espacial. Para ello se describiré la importancia económica de los procesos de aglomeración, se explican los determinantes del salario y los factores que originan desigualdades salariales, se determina y contrasta la diferencia salarial a nivel de ciudad y por último identifican autocorrelaciones espaciales entre las ciudades seleccionadas.

Los países procuran mejorar su desarrollo económico y eso supone contar con un crecimiento económico sostenido y sustentable. Sin embargo, cada país cuenta con características tanto de recursos naturales, como características socio-políticas a nivel regional que determinan el crecimiento económico. Esas características únicas y específicas generan diferentes dinámicas y desigualdades en el crecimiento de estas regiones.

De acuerdo con la literatura, existe una asociación entre aglomeración y desigualdad salarial. Los salarios que se perciben en las regiones, al igual que el crecimiento de estas, es desigual y esto implican una diferenciación en el bienestar de los individuos que las habitan, manifestándose como diferenciación salarial. Resulta teórica y económicamente relevante determinarla a fin de comprender si la diferenciación salarial se propaga o limita en las ciudades más importantes de México, pues son estas las de mayor dinámica económica en el país, por lo cual se desprendería de ello el que fomenten u obstaculicen el crecimiento y desarrollo del país.

La hipótesis del trabajo es que el salario de las ciudades de México manifiesta correlación espacial positiva entre ellas. Además, mostrar que las diferencias salariales entre las ciudades del país son afectadas por la población la tasa de desocupación y de informalidad, y el promedio de años de escolaridad y edad de los trabajadores. Para comprobarla se utiliza la información de la Encuesta Nacional de Ocupación y Empleo (ENOE).

CAPÍTULO I:

MARCO TEÓRICO – CONCEPTUAL

INTRODUCCIÓN DEL CAPÍTULO

En este primer capítulo se tiene por finalidad el presentar las bases teóricas y conceptuales que guían la investigación, las cuales se encuentran divididas en cuatro apartados: 1) mercado laboral, 2) aglomeración, 3) crecimiento económico desigual y 4) medidas y pertinencia de estudio.

En el primer apartado se define el mercado laboral, se identifican elementos relevantes de la oferta y demanda de trabajo y el funcionamiento de estas. Se centra en conceptualizar el salario y el trabajo, así como la determinación del salario por medio de la curva de Phillips y la curva de salario. Posterior se muestran la clasificación de características que contribuyen a diferencias salariales.

El segundo apartado tiene por finalidad el mostrar la importancia de la aglomeración en las diferencias salariales, justificando el estudio de estas en las ciudades. Se describe cómo surgen los procesos de aglomeración y los efectos que conlleva estos, clasificando en economías de localización y de urbanización. Se relacionan los procesos y las externalidades con fuerza laboral.

El tercer tiene por finalidad el presentar cómo la dinámica y la estructura económica de las regiones presenta diferencias en el crecimiento económico y cómo el crecimiento puede contribuir a las diferencias en los salarios percibidos. Se muestra cómo la aplicación de una política económica puede contribuir a las desigualdades del crecimiento. El cuarto, y último apartado del capítulo, exhibe brevemente las formas de medir las diferencias salariales.

1.1 MERCADO LABORAL

En economía un mercado se entiende como un espacio, no necesariamente físico, en donde los agentes intercambian bienes y/o servicios determinados, tanto por una oferta como por una demanda de los mismos. Teniendo conocimiento de esto, el mercado de trabajo podría entenderse como el espacio intangible en donde interactúan la oferta y la demanda de trabajo con fin de determinar el precio, conocido como salario, que percibirán y la cantidad de trabajadores, conocida también como mano de obra. En la presente investigación se entiende como mercado de trabajo:

“... el conjunto de trabajadores en búsqueda de un puesto de trabajo (oferta de trabajo) y de empresarios de todo tipo que necesitan y requieren fuerza de trabajo (demanda de trabajo). Pero la consideración de este mercado puede ser diferente según las variables que indican en esa demanda y oferta mencionadas y según los supuestos económicos manejados.” (Fernández Díaz, et al., 1999, p. 149)

El término fue ampliamente estudiado por diversos autores a partir de la década de los sesenta, desde distintos enfoques y, originalmente sobresalió el enfoque clásico y el keynesiano (Fernández Díaz, et al., 1999). En dichas perspectivas destacan 5 características:

- I. La demanda de trabajo cambia o se desplaza comúnmente;
- II. La oferta de trabajo es heterogénea;
- III. La información que se encuentra en el mercado de trabajo es insuficiente e inadecuada;
- IV. El desempleo constituye una situación normal, y
- V. La movilidad en la mano de obra es totalmente ilimitada.

Algunos de los elementos más destacados que se estudian en el mercado de trabajo son: el nivel de desempleo, los precios, la distribución de la renta, la cantidad de mano de obra disponible, la existencia de sindicatos, la variación de los salarios, etc.

El estudio del mercado de trabajo toma importancia económica a consecuencia de las distintas estructuras económicas de las regiones, reflejadas en una interacción distinta en

el mercado de trabajo, dando como resultado diferencias tanto en la cantidad de trabajadores como en el salario de estos. Otro factor que llevó al estudio del concepto es que la manera en que se interactúa en él fue considerada como un factor determinante del crecimiento económico en las regiones (Aguilera Fernández, 2012)

Una característica destacable para la presente investigación es que tanto la oferta como la demanda de mano de obra no suele ser homogénea. La oferta de trabajo está integrada por individuos que presentan distintas características, por ejemplo, escolaridad; a fin de intercambiar su trabajo por un salario. Como cada individuo posee características particulares la oferta de trabajo es distinta para cada espacio. Por su parte, la demanda de trabajo se encuentra compuesta principalmente por empresas o instituciones que buscan trabajadores con características particulares para desempeñar las labores deseadas, por ejemplo, contar con posgrado en economía, al poseer necesidades distintas, la demanda de trabajo no puede ser homogénea para cada espacio (Ehrenberg & Smith, 2018).

El Estado han buscado intervenir en el mercado laboral a fin de conseguir el equilibrio entre la oferta y demanda de trabajo, ya que si se presentan desequilibrios no deseados pudiesen reflejarse en aspectos negativos para la sociedad, como altas tasas de desempleo que conllevarían a reducciones de la producción. Es por ello que el Estado y sus instituciones desempeñan usualmente un papel como agentes reguladores del mercado laboral para propiciar, garantizar y mejorar las condiciones laborales, así como intervenir en la determinación y la distribución del salario que se percibe (Lucifora, 1998).

1.1.1 FUNCIONAMIENTO DEL MERCADO

Tanto la oferta como demanda de trabajo se encuentran en función de los niveles actuales y futuros de la demanda de bienes y/o que utilizan el factor trabajo, de igual manera se encuentran en función de las utilidades que esperan obtener las empresas de los procesos productivos implementados y de las expectativas en cambios tecnológicos, así como políticos y sociales.

Dentro de la demanda de trabajo que realizan los empresarios, debemos determinar cómo afecta a la elasticidad un incremento en el precio de los salarios; ésta dependerá principalmente de tres razones: la dificultad de sustituir el factor trabajo por otro (nueva tecnología, por ejemplo), la proporción que ocupa el factor trabajo dentro de los costos totales de los empresarios y de la elasticidad de la demanda del producto. Tomando en cuenta esto, un aumento en el salario que otorgan los empresarios a los trabajadores producirá sólo una pequeña reducción en el nivel de empleo que demanden, es decir, la demanda de trabajo será inelástica, solo si:

- a. La demanda del producto es inelástica también,
- b. Existen pocas posibilidades de sustituir el trabajo establecido (poca posibilidad de cambios tecnológicos y del capital en los procesos de producción),
- c. Dentro de los costos totales de la empresa, el costo de la fuerza de trabajo representa una parte no importante (Fernández Díaz, et al., 1999).

Por otra parte, dentro de la oferta de trabajo se encuentran cuatro principales factores que la condicionan. La primera es la eficiencia de la fuerza de trabajo, la cual presenta el grado de calificación de los empleados, la segunda es la población total, la tercera es la distribución de ésta por edad y sexo, y la cuarta es el marco social y legal con el que cuenta el país, el cual está determinado por las instituciones que integran al Estado.

1.1.2 SALARIOS

El salario ha sido una variable económica que ha presentado relevancia a lo largo de su estudio, ya que es capaz de relacionarse con el nivel de vida de la población, así como con el crecimiento de la producción. Para la presente investigación el salario será entendido como el precio por el trabajo¹.

¹ El trabajo será entendido como “la capacidad y habilidad del ser humano para producir bienes y servicios” (Cue Mancera & Quintana Romero, 2009, p. 201)

Durante la década de los noventa, comenta Castellanos (2010), se suscitó un debate sobre cómo analizar la relación entre el desempleo y los salarios, dicho análisis podría verse a través de la curva de Phillips o de la curva de salarios. En torno a este debate Arellano y Bond (1991) pudieron concluir, mediante un modelo empírico por método generalizado de momentos, que la mejor manera de explicar la relación es a través de la curva de salarios; en el trabajo de Castellanos (2010) igualmente se llega a la misma conclusión utilizando un análisis de paneles dinámicos. Es importante destacar que ambas curvas solo presentan significancia al corto plazo, cuando los agentes económicos aun no adaptan sus expectativas, por ejemplo, a la inflación.

La curva de Phillips (1958) tiene como fin mostrar la relación que existe entre el desempleo y los salarios de una economía; se considera que esta constituye el punto de partida para analizar la relación entre estas dos variables. La curva toma vital importancia por parte de diversos autores en la década de los setenta, y fue utilizada para describir la dinámica de la oferta agregada de modelos Keynesianos. Una crítica fue que presenta una visión macroeconómica y dejaba fuera la visión de datos microeconómicos y regionales, por ello surgieron nuevas visiones como la curva de Phillips incrementada por expectativas de Phelps (1968) y la curva de oferta de Lucas (1972)².

La curva de Phillips muestra una relación negativa entre el cambio porcentual en los salarios y la tasa de desempleo. Menciona que los cambios de las tasas salariales se asocian con la inflación y que las variaciones en las tasas de desempleo de la misma con los cambios en el PIB. En base a estas asociaciones Phillips indica una relación entre un nivel elevado de producto con altas tasas de inflación y bajas tasas de desempleo (Jiménez, 2001).

El salario puede ser determinado por las características particulares de trabajador. Por parte de la demanda de trabajo es posible que exijan un alto grado de escolaridad y por parte de la oferta es posible que oferten un alto grado de escolaridad, en conjunto se puede determinar el salario. Además de la escolaridad son diversas las variables del trabajador que pueden considerarse, algunas de las más destacadas por distintos autores son: escolaridad,

² En dichas visiones los modelos parten de agentes representativos que maximizan su utilidad.

sexo, experiencia, estado civil, edad, raza, sindicalizado, etc. (Llamas Huitrón, 2019). Además de estas características individuales también pueden influir características sectoriales, entre las que destacan la actividad en que labora, el tamaño del establecimiento, la ocupación, etc. (Castro Lugo, 2005).

Considerando estas variables individuales y sectoriales se es capaz de estimar una ecuación salarial; en ella, el nivel de salario es explicado por tasas de desempleo regional, las que a su vez están determinadas por variables de las características individuales y sectoriales mencionadas. Las ecuaciones salariales fueron estudiadas en diversos trabajos por Blanchflower y Oswald (1996), durante la década de los ochentas, y tomaron relevancia en los noventa. Blanchflower y Oswald (1996) lograron encontrar que en varios países de Europa, Japón, Corea, India, Canadá y Estados Unidos la elasticidad de los salarios frente al desempleo regional era negativa (-0.1). A partir de esos estudios pudieron demostrar que si se duplicara la tasa de desempleo entonces el nivel de salario real disminuiría en 10 por ciento.

Este modelo de curva de salario no puede explicarse mediante modelos de oferta y demanda ya que en estos se obtiene que, a mayor salario, mayor será el desempleo (Castellanos, 2009). Para comprender la curva de salarios es necesario explicarla mediante modelos de negociación, bajo los cuales, a mayor número de trabajadores empleados y calificados, mayor será el poder de negociación para aumentar de salarios, ya que encontrar mano de obra calificada y capacitada podría aumentar costos y/o reducir la productividad del trabajo.

1.1.3 DIFERENCIACIÓN SALARIAL

En cuanto al término de desigualdad se entiende en la presente investigación como una diferenciación entre dos o más individuos. El término de diferenciación salarial ha sido descrito en la literatura como un aspecto negativo y un problema para la eficiencia en el mercado de trabajo (Islas Camargo & Cortez, 2004).

El salario puede diferir entre regiones a consecuencia de la estructura económica de cada una. Marshall (1979) comenta que el salario es capaz de verse determinado de manera endógena por parte de las fuerzas económicas y de manera exógena a través de factores socio-institucionales.

El estudio de las diferencias salariales ha sido visto por diversos autores y usualmente se inició al observar las diferencias a nivel de país. Los cambios en la distribución salarial suelen analizarse a *grosso modo* en tres formas: a través de cambios en la demanda, por factores de oferta y, en menor medida, por factores institucionales (Castro Lugo & Huesca Reynoso, 2007). Al analizar en mayor medida la diferencia entre países, comenta Castro (2006) que se ha encontrado ausencia en el estudio de las diferencias urbano o regional.

Variables que influyen en la diferencia salarial pueden agruparse en tres. Las primeras son características individuales al trabajador, entre las cuales se puede mencionar el sexo, la edad, la experiencia laboral, la escolaridad, etc. la segunda son características relacionadas a la empresa-institución, en esta clasificación se puede mencionar el tamaño de la empresa, la rama, prestaciones laborales, si trabaja en el sector público o privado, el tipo de contrato (usualmente si se tiene contrato indefinido se presenta mayor sueldo) y la existencia de sindicato. Y una tercera forma de diferencia salarial relacionadas a variables regionales, donde destaca la ciudad en la que se labora y el tamaño de la localidad (Gámez Arroyo & Rosas Rojas, 2015).

Wheaton (2002) coincide con Gámez y Rosas (2015) en relación a la diferenciación de salarios, mencionando que existe evidencia en la cual esta no se origina por diferenciación en la productividad de la empresa, si no que existen otros tipos de factores institucionales que la ocasionan, por ejemplo, que los trabajadores se encuentren sindicalizados. También argumenta una relación positiva entre salarios y rentas inmobiliarias.

Algunas de las teorías más sobresalientes sobre la diferenciación salarial son la teoría del capital humano, diferencias salariales compensatorias, teoría de la segregación ocupacional y la teoría de los salarios de eficiencia³.

Según la literatura se ha mostrado que existe una relación inversa entre el salario local y el desempleo en las regiones, esta hipótesis es consistente con la hipótesis de la curva salarial. Sin embargo, la hipótesis de la curva salarial describe que los salarios se determinan en el mercado de trabajo local, pero no consideran las interacciones espaciales entre los mercados locales. La hipótesis se puede cumplir al examinar un mercado de trabajo nacional, no obstante, para uno local es importante el determinar las relaciones salariales entre las regiones de los países (Buettner, 1999).

Lo que permite la interacción y dependencia salarial entre las regiones son los factores de movilidad de mano de obra, comercio internacional e instituciones. En cuanto a la contigüidad y vecindad de las regiones se comenta que:

“La contigüidad, está relacionado con la integración por factor de movilidad, ... Cualquier variación en el área y la distancia de las unidades espaciales causará que el nivel de desagregación espacial afectará la relación entre los salarios locales y el desempleo local. Por lo tanto, tener en cuenta los efectos de contigüidad podría ser útil en el intento de desentrañar esa relación...” (Buettner, 1999, p. 48).

La mayoría de los trabajos de desigualdades salariales son analizados desde una diferenciación social, por ejemplo, desigualdad por género, por grupo étnico, por edad, por tipo de contrato que se posee e inclusive en el sector privado o público en el que se encuentre el empleado etc. En el presente trabajo la desigualdad salarial es analizada desde una perspectiva espacial, viéndola como una asociación geográfica localizada exactamente en un mapa. La unidad geográfica de análisis son ciudades entre las cuales se mostrarán los efectos de las diferencias entre ellas para hacer un acercamiento a determinar si las diferencias salariales convergen o divergen.

³ Para mayor información sobre las teorías mencionadas véase (Castro Lugo, 2006)

Usualmente las diferencias en los salarios llevan a movilidad del capital humano, donde los trabajadores locales se desplazan a áreas que paguen más, sin embargo (Qiong & Chand, 2014) comentan que las diferencias salariales tienden a converger en la distribución de los salarios en un país. Además, comentan un aspecto esencial para la presente investigación, y es que la aglomeración es una causa de las disparidades salariales dentro de los países, por lo que “la aglomeración de la industria puede ser una fuerza para la divergencia de salarios entre las regiones” (Qiong & Chand, 2014, p. 98).

1.2 AGLOMERACIÓN

El deseo de una mejor calidad de vida ha propiciado que las familias busquen gestionar relaciones personales, sociales y económicas concentradas en determinados espacios. En la presente investigación, el término de ciudad será entendido por: “conjunto compacto de personas y de actividades económicas que se desarrollan sobre un espacio físico restringido o que desembocan en una polaridad reconocible” (Camagni, 2005, p. 21). Entendiendo esto, el término de aglomeración presenta una característica elemental para la ciudad y para la investigación.

La literatura menciona que existe presencia de procesos productivos más eficientes cuando existe la aglomeración. Camagni (2005) describe que una de las causas que propicia modelos concentrados eficientes radica en las economías de escala encargadas de la concentración espacial de tipo productivo, distributivo y/o financiero. Estas concentraciones suelen llevar a externalidades positivas, en mayor medida, y en ocasiones negativas.

Entorno a las externalidades positivas de la aglomeración Duranton y Puga (2004) describe tres mecanismos: compartir, igualar y aprender. El primero beneficia de tal manera que reduce los costos mediante la utilización de instalaciones o infraestructura conjunta, además de beneficios intangibles como el poder compartir el mismo proveedor de insumos, misma campaña publicitaria para los bienes o adquisición de servicios de mismo interés. El

segundo, beneficia de modo que al estar los trabajadores próximos entre empresas aumentan la probabilidad de encontrar o cambiar de trabajos productivos rápidamente y a bajos costos. Y el último mecanismo genera beneficios económicos, ya que la cercanía de las empresas y de los trabajadores induce al intercambio de información y de conocimiento (spillovers) entre ellos, propiciando innovación y aumentos en la productividad de la región aglomerada (Duranton & Puga, 2004).

Las economías de aglomeración pueden clasificarse en: economías internas a la empresa, economías externas a la empresa, pero internas a la industria o economías de localización y economías externas a la empresa y a la industria o economías de urbanización. En el siguiente apartado se describe más a detalle estas dos últimas.

1.2.1 ECONOMÍAS DE LOCALIZACIÓN Y DE URBANIZACIÓN

Para comprender estas dos clasificaciones se comparte la definición de Camagni (2005). Las economías de localización se entienden como las ventajas que se derivan de la localización concentrada de las empresas pertenecientes a la misma industria o sector productivo. Las economías de urbanización se entienden como las ventajas de un ambiente urbano, que se derivan de la presencia de infraestructuras genéricas, utilizables por todas las industrias. Cuando existe economías de escala internas a una empresa conlleva a la aglomeración de otras más y puede generar ventajas de economías de localización, que con la creciente producción generará la concentración de otras actividades distintas a la empresa y a la industria ya concentrada, entonces existirán economías externas de urbanización.

Si los beneficios de la aglomeración son causados por la localización, los beneficios son conocidos como externalidades tipo MAR⁴, por su parte, si son originados por la urbanización se conocen como externalidades tipo Jacobs (Wheaton & Lewis, 2002).

⁴ Marshall-Arrow-Romer

Las economías de localización consideran: procesos de especialización por parte de las empresas, reducción de costos de transacción, creación de una cultura industrial con combinaciones tecnológicas y organizativas y, lo más importante para la presente investigación, la formación de un mercado de mano de obra especializado (Camagni, 2005).

Las economías de urbanización presentan externalidades que se pueden agrupar en tres categorías. La primera se basa en la concentración de infraestructura y servicios públicos. La segunda es la naturaleza de un mercado de grandes dimensiones con varios nichos de especialización. La tercera son factores productivos y de mercado de la propia ciudad, entre los cuales se destacan: accesibilidad a funciones urbanas especializadas, oferta de capacidades empresariales y directivas, economías de comunicación e información y acceso a un mercado de trabajo amplio y especializado (Camagni, 2005).

En cuanto a los trabajadores se ha comentado que las economías de aglomeración proporcionan beneficios para estos, particularmente Wheaton y Lewis (2002) señala que en la teoría las economías de urbanización han distinguido que: el conocimiento de los trabajadores es transferido entre las empresas a través del espacio y en segundo lugar, que existen beneficios a escala en el mercado laboral a causa de una mayor demanda de mano de obra especializada, situación que incentiva a los trabajadores a la especialización, aumentando con esto la productividad y así las tasas de crecimiento económico de la región aglomerada.

1.2.2 LA NUEVA GEOGRAFÍA ECONÓMICA

La nueva geografía económica describe cómo el espacio puede explicar los procesos dinámicos de la formación urbana. Se plantea un modelo en donde se supone que en la economía existen dos sectores, uno de ellos es manufacturero, en donde existen rendimientos crecientes y movilidad perfecta en el espacio, debido a que las regiones urbanas tienden a dedicarse en menor medida a actividades agrícolas, los factores descritos por el sector manufacturero fomentan la aglomeración de la actividad productiva y la

formación urbana de ciudades (Fujita, et al., 2001). Para explicar la estructura de las ciudades, en una primera etapa se analiza la existencia del espacio (costos de transporte) y la diversidad (economías de escala), además la infraestructura urbana y los bienes públicos influyen de manera directa (Fujita & Krugman, 2004).

Para lograr estos factores las ciudades han aprovechado los beneficios de las externalidades ocasionadas por la aglomeración de las actividades económicas. Algunos beneficios es que las empresas ahorran costos de transporte al localizarse cerca de proveedores y de clientes; la agrupación del mercado laboral facilita el acceso a mano de obra calificada y, mayor facilidad de ser empleado y la presencia de spillovers de conocimiento entre el capital humano (Melo & Graham, 2013).

1.2.3 RELACIÓN CON LA FUERZA LABORAL

Para esta investigación es destacable la asociación de la aglomeración con la fuerza laboral y a su vez con los salarios percibidos en ciudades. Por ejemplo, Wheaton y Lewis (2002) muestran que los trabajadores del sector manufacturero perciben un salario mayor si se encuentran en mercados laborales urbanos con alta participación nacional en el empleo.

En cuanto a los trabajadores y su agrupación es relevante, para la investigación, describir la hipótesis que formula Melo y Graham (2013) ya que contribuye a la relación de las diferencias salariales en regiones a causa de la aglomeración y la mano de obra capacitada:

“Los efectos de la agrupación laboral se basa en la idea de que las industrias que usan trabajadores con habilidades similares tienden a ubicarse cerca unas de las otras para disfrutar de beneficios de fácil acceso a una fuerza laboral con habilidades relevantes... La hipótesis que se está probando es que aumentan los mercados laborales más aglomerados tanto la probabilidad como la calidad de las coincidencias entre empleadores y trabajadores... Una de las hipótesis utilizadas para evaluar la agrupación laboral consiste en evaluar un resultado positivo de la relación entre la proporción de trabajadores calificados

con educación superior y la aglomeración espacial en industrias que emplean trabajadores especializados y calificados. Este enfoque tiene dos principales deficiencias: su capacidad limitada para capturar habilidades especializadas específicas de la industria y el hecho de que también captura los efectos de aglomeración que surgen de los efectos secundarios del conocimiento” (Melo & Graham, 2013, pp. 5-7).

Otra hipótesis que mencionan es que existe un resultado positivo de la relación entre la proporción de trabajadores calificados con educación superior y la aglomeración espacial que emplean a trabajadores con alto grado de especialización y calificados. Con esta hipótesis se puede inferir que las ciudades aglomeradas tienden a presentar una mayor demanda de mano de obra con capital humano formado. Si bien las empresas demandan el un mayor capital humano, los trabajadores pueden demandar un mayor salario. El otorgar y percibir un mayor salario es más posible en regiones donde las empresas se encuentran aglomeradas, ya que los beneficios de la aglomeración le reducen costos y así pueden otorgar mayores salarios.

El aumento en la productividad de los trabajadores debería verse reflejado en ganancia para las empresas y presencia de mayores salarios, los trabajadores con búsqueda de mejores salarios también tienen incentivos a especializarse más, por lo que estos optan por profundidad en lugar de amplitud del capital humano, así generando productividad e innovación (Wheaton & Lewis, 2002, pp. 543-544). Es así como se presenta rendimientos crecientes en los mercados laborales dentro de economías de aglomeración.

1.3 CRECIMIENTO ECONÓMICO REGIONAL DESIGUAL

Siguiendo los cánones de la investigación referente a las diferencias salariales vistas desde el espacio, los determinantes del crecimiento económico de una región son diversos, pero en este trabajo se asociarán las diferencias en los salarios con el papel económico del capital humano y de los beneficios de la aglomeración. El desempeño de México en materia de crecimiento económico ha ido a la baja (en contraste con países miembros de la OCDE);

en contra parte, el papel que ha mostrado el país en materia de capital humano es positivo, supera el promedio de años de escolaridad en América Latina durante dos décadas (Levy, 2018, p. 5). Además, México se posiciona en el segundo lugar de países que presentan mayor número de horas trabajadas por semana de los países miembros de la OCDE. Levy (2018) comenta que, a pesar de un aumento en el capital humano de México, este no ha llevado a un incremento en la productividad, señala que para lograrlo es necesario, entre otras cosas, la aplicación de políticas que fomenten inversión en capital físico.

El contexto de competencia internacional ha conllevado a que ciertas regiones se rezaguen o progresen en términos económicos, el situarse en alguna de las dos categorías depende directamente del desempeño económico de la región. Al contar con dinámicas y estructuras desiguales entre sí, también los desempeños económicos de estas difieren de región a región, es por ello la necesidad mencionar cuales determinantes influyen en el desempeño y en las diferencias, algunos de estos son: “La distribución territorial de los recursos y las actividades económicas, las características de las estructuras productivas, el crecimiento económico, y el crecimiento urbano y la urbanización son importantes elementos en los que se reflejan las diferencias territoriales y que afectan notablemente los procesos de desarrollo.”⁵ (Trejo Nieto, 2013, p. 546). A Grosso modo existe la necesidad de las regiones por aumentar su capacidad productiva, atraer nuevas actividades y crear nuevos empleos.

La dinámica de cada región determina el crecimiento económico y la productividad de ellas y al ser desiguales entre las regiones conlleva a desigualdades en el crecimiento y en la productividad. Un problema es que “México asigna mal el capital humano que los trabajadores acumulan antes de ingresar en la fuerza laboral” (Levy, 2018, p. 178) y esta mala asignación conlleva las desigualdades mencionadas. Si bien los salarios pueden ser determinados por las dinámicas de la región, también existe evidencia de que no solo es la dinámica de la misma si no también determinados por la dinámica de las regiones que la rodean, en cuanto a esto: “el comprador y el vendedor del trabajo se encuentran dentro de alguna área geográfica fija, pero el precio al que el intercambio se lleva a cabo a menudo es

⁵ Las citas textuales del trabajo se realizan con traducciones propias del inglés en los documentos originales.

determinado en última instancia por otros organismos de cientos de millas de distancia” (Ross, 1947, p. 802).

Jacobs menciona que el intercambio de conocimiento, ideas e innovaciones entre personas puede generar que fluyan más rápido avances tecnológicos, los cuales conllevan efectos positivos en la productividad y por ende en el crecimiento económico, es por ello que la aglomeración de personas fomenta, de manera más eficiente, el crecimiento de las ciudades. Glaeser (1992) menciona que, si bien estos intercambios de conocimiento generan externalidades positivas para el crecimiento económico, también generan aumentos en los precios de las tierras y de los alquileres y, ambos son determinantes en las dinámicas de las ciudades. De igual manera en distintas teorías de crecimiento económico como las de Romer, Porter y Jacobs se analiza el rol que juegan los efectos indirectos de la tecnología y los spillovers de conocimiento entre distintas áreas (Glaeser, et al., 1992). Ello ayuda a comprender el por qué las ciudades tienen mayor crecimiento económico que las localidades rurales.

Las poblaciones rurales se encuentran superadas en el crecimiento de la productividad de las actividades económicas frente a las poblaciones urbanas, esta desventaja fomenta la desigualdad en los ingresos per cápita ya que el aumento de la productividad puede reflejarse en crecimiento económico y en mayores ingresos (Kuznets, 1955).

Las desigualdades regionales a consecuencia de las características y dinámicas de producción de las regiones han presentado sintonía con las políticas económicas efectuadas en el país y en las regiones. En torno a esto Trejo Nieto (2013, p. 546) comenta que el país presentó expansión durante el periodo de industrialización, particularmente con el modelo de sustitución de importaciones⁶. Durante el auge de este modelo se detectó que las

⁶ Modelo que basado principalmente en el sustituir productos manufacturados extranjeros por productos elaborados en el país mediante la aceleración de inversiones en la industria interna. Fue aplicado en la década de los cuarentas durante el gobierno del presidente Manuel Ávila Camacho.

ciudades con mayores beneficios en términos de industrialización y modernización fueron: Distrito Federal, Guadalajara y Monterrey, ciudades que, desde entonces, y hasta hoy en día, pueden ser consideradas los núcleos económicos y poblacionales del país.

Otra política económica que se reflejó en beneficios para algunas regiones fue la aplicación del Programa de Industrialización Fronteriza (PIF), aplicado en la década de los sesentas, con el cual, comenta Trejo (2013), que se logró la creación de empleos y la formación de bases industriales, este programa sentó bases para la identificación de un patrón económico dual centro-norte.

El progreso económico de algunas regiones conllevó a la generación de externalidades económicas positivas en ellas, por lo que se fomentaron procesos de urbanización en el país, los cuales propiciaron la creación de nuevas metrópolis.

1.4 MEDIDA Y PERTINENCIA DE ESTUDIO

Una herramienta muy utilizada en los análisis de econometría espacial es la utilización del estadístico I-Morán, el cual en la investigación es utilizado para determinar si existe o no relación, ya sea positiva o negativa, entre otras regiones en el espacio. Por último, se estiman modelos de econometría espacial donde se muestra sí las variables explicativas son significativas y la relación que presenten estas frente a los salarios.

En México existen diversos trabajos relacionados con la desigualdad salarial, analizados a diferentes áreas, desde localidades hasta nivel nacional, abarcando análisis de desigualdad en ciudades. Sin embargo, no se ha encontrado hasta el momento algún trabajo que proporcione un análisis espacial de la relación entre los salarios de las ciudades. Por su parte diversos países como China (Liu & Yin, 2008), Francia (Combes, Duranton, Gobillon, 2008), Alemania (Longhi, Nijkamp, y Poot, 2006), España (Tirado, Pons, Paluzie, Martínez-Galarraga, 2013) y Brasil (Kovak, 2013) han realizado estudios donde analizan relaciones salariales desde una perspectiva espacial. Con base en la ausencia de un estudio que

muestre si existen relaciones espaciales por parte de los salarios en ciudades de México, es pertinente la realización del presente trabajo.

CONCLUSIONES DEL CAPÍTULO

De acuerdo con lo anteriormente expuesto, el salario se conceptúa como el precio dado a la capacidad del ser humano para producir bienes y /o servicios. Dentro del mercado de trabajo, será el encargado de poder concebir el equilibrio entre la oferta y la demanda de trabajo, donde también se consideran los factores institucionales. Como variable de desarrollo puede considerarse influyente en el nivel de vida de los trabajadores.

Las características que llevan a que existan diferencias salariales descritas se clasifican en personales, institucionales y regionales. En las primeras sobresalen la escolaridad, la experiencia, estado civil, raza, etc. para las institucionales se mencionaron la existencia de sindicatos, el trabajar en sector público o privado, el tamaño de la empresa, etc. y en la última categoría presta vital importancia en la investigación ya que se menciona que pueden existir diferencias salariales en función del tipo de localidad y del tamaño de esta.

Para comprender como es que factores regionales influyen en las diferencias salariales se explicaron los procesos de aglomeración y los beneficios que originan esta. La agrupación de personas y actividades económicas genera externalidades positivas para la producción, reflejadas en aumentos de productividad y acumulación de diversas industrias en el espacio.

Los aumentos de la productividad y la especialización, demandan una mayor formación de capital humano en la región. Una fuerza laboral capacitada demanda un mayor salario ya que puede ofrecer una mayor productividad. A mayor aglomeración mayor oferta de servicios, por ejemplo, educativos especializados, los cuales contribuyen a la formación de mano de obra especializada. Es así como, a *grosso modo*, las regiones más aglomeradas pueden ofrecer un mayor salario por parte de las empresas y demandando mayor salario por parte de los trabajadores.

Además de la aglomeración, la estructura y dinámica productiva puede contribuir a la productividad de la misma, provocando diferencias salariales también. El crecimiento económico de una región se considera una fuerza de atracción para las actividades productivas y para la movilidad de factor de capital humano. Por consiguiente, si cada región presenta diferencias en las tasas de crecimiento y diversas estructuras y dinámicas productivas, conllevará a posibles diferencias en la percepción de salarios.

En el siguiente capítulo se presenta la metodología empleada para poder determinar las desigualdades de los salarios de ciudades de México y la metodología para determinar si existen relaciones espaciales entre dichas desigualdades.

CAPÍTULO II:

MARCO METODOLÓGICO

2.1. ANÁLISIS ESPACIAL

En las últimas dos décadas se ha observado un creciente interés por parte de investigadores y por diversas disciplinas, como economía, sociología, medicina, criminología, política, etc., por el papel que juega la ubicación de datos para explicar sin fin de fenómenos científicos (Griffith, 2003). Para satisfacer la necesidad del estudio de dichos fenómenos se ha demandado la implementación de diversas herramientas de carácter espacial que otorguen un adecuado manejo de datos.

Si bien el análisis espacial es estudiado por la mayoría de las disciplinas, se ha logrado observar que los investigadores dedicados particularmente a las ciencias sociales han mostrado un interés particular por los análisis espaciales ya que frecuentemente los fenómenos bajo estudio no presentan un marco formal o teoría previa sobre la dimensión espacial del fenómeno (Anselin, 1999).

Uno de los principales impulsores y desarrolladores de herramientas de análisis y econometría espacial es Luc Anselin, quien puede expresar el análisis espacial como el estudio cuantitativo de fenómenos que se reflejan en el espacio. Para su estudio es necesario considerar las áreas en las que se presenta, distancias, e interacciones expresadas por Tobler (Anselin, 1989, p. 2). Por su parte recordemos que se entiende por datos espaciales como la observación de una variable asociada a una localización en el espacio.

Griffith (2003) argumenta que existen tres principales razones que diferencian entre el análisis estadístico espacial frente al tradicional; el patrón espacial de la ubicación, la asociación espacial entre atributos observados en diferentes ubicaciones y la variación del

fenómeno observado ocasionada principalmente por la variación en la ubicación. Es por estas razones que se ha encontrado un creciente uso de herramientas sofisticadas para el manejo de datos espaciales en distintos campos como la geoestadística, la estadística espacial y la econometría espacial (Griffith, 2003, p. 7) .

Las metodologías de análisis de datos espaciales suelen clasificarse en dos metodologías. Una es conocida como la Minería de Datos Espaciales (MDE) que consiste a grosso modo en la utilización de técnicas de geovisualización de patrones espaciales, es conocida como un proceso de búsqueda de asociaciones y patrones espaciales a partir de un análisis exploratorio de grandes y complejos conjuntos de datos espaciales (Shekhar, et al., 2004). La segunda es conocida como Análisis Exploratorio de Datos Espaciales (AEDE) en la cual se emplean técnicas estadísticas espaciales. Ambas metodologías demandan un solo Sistema de Información Geográfica (SIG).

Anselin (1996) argumenta que con la utilización de herramientas de geovisualización con más eficientes para el uso de conjunto de datos relativamente grandes y por su parte las estimaciones de autocorrelación por parte de AEDE con muestras relativamente grandes presentan menor fidelidad y eficiencia frente a datos relativamente pequeños (Anselin, 1996). Para la presente investigación los datos utilizados no se consideran una muestra grande ya que, por ejemplo, se utilizan 36 ciudades autorrepresentativa para el 2019, por lo que la utilización de metodología de AEDE es adecuada.

2.2. DATOS ESPACIALES

La importancia de los datos espaciales para la investigación radica en identificar patrones de distribución del salario en las ciudades de México y la relación entre estos. Para iniciar el análisis de los datos espaciales, en primera instancia es necesario determinar la forma en la que se presentaran estos. Anselin (1989) argumenta que las observaciones espaciales pueden presentarse a través de tres tipos de datos: puntos, superficies o

polígonos. Cada uno demanda distintos tratamientos estadísticos, análisis de patrones de puntos, geoestadística y análisis de polígonos (Anselin, 2006, p. S3), respectivamente.

En la presente investigación los datos espaciales se asocian a las ciudades a través de polígonos. La asociación se realizó con el software “RStudio”, en primera instancia se obtuvo del INEGI un paquete de información a nivel municipal que contiene archivos en formato .bdf, .prj, .shp y .shx. en ellos se encuentra el mapa de México delimitado a nivel municipal. Para agregar los datos recabados de la ENOE se aplicaron comandos que permiten asociar las variables de interés a cada espacio del polígono de los municipios.

Cabe mencionar que, si bien cada ciudad autorrepresentativa de la ENOE se encuentra compuesta por diversos municipios, los datos fueron agregados únicamente al municipio central. Sin embargo, esto no afecta a los resultados estadísticos obtenidos ya que los polígonos se trabajan con localización discreta con atributos fijos. Esta aplicación se consideró únicamente para poder presentar los mapas e identificar la distribución de las variables asociadas a la localización fija.

Los datos espaciales tienen dos características básicas para su manejo que son la georreferenciación y la multidireccionalidad. La georreferenciación es asociada a la posición ya sea absoluta o relativa de un fenómeno en el espacio, con ella es posible identificar visualmente patrones de distribución, en la investigación es posible identificar la distribución de los salarios. Por su parte la multidireccionalidad se entiende como los vínculos y/o relaciones que se presentan entre los puntos (o polígonos) del espacio, se centra en fenómenos espaciales como interacción, dependencia, externalidades, contagio, desbordamiento o replica (Chasco Yrigoyen, 2003).

Las dos características descritas permiten que el análisis espacial pueda definirse a través de tres componentes; análisis exploratorio de datos espaciales, visualización y modelación espacial, estos componentes están fuertemente relacionados (Anselin, 2006, p. S4). En la presente investigación los datos espaciales obtenidos permiten la realización de un análisis espacial completo por los tres componentes.

2.2.1. ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES

Para comprender el AEDE es necesario saber que se compone de un conjunto de técnicas para detectar características distintivas de los datos geográficos por efectos de dependencia y heterogeneidad espacial. Anselin define el AEDE como *“una colección de técnicas que permiten describir y visualizar distribuciones espaciales, identificar localizaciones espaciales atípicas (outliers espaciales), descubrir patrones de asociación espacial (agrupamientos espaciales), clusters o puntos calientes y sugerir diferentes regímenes espaciales u otras formas de heterogeneidad espacial”* (Anselin, 1998, p. 4)

Entendiendo la definición de AEDE se intuye la importancia de los conceptos de localización, el área, topología, la distancia y la interacción espacial. Para hacer funcional, en materia estadística, estos conceptos se representan como polígonos (como es el caso de la investigación), líneas o puntos.

La principal diferencia metodológica del AEDE frente al análisis estadístico que considera Anselin (1992) es que no se cumple el supuesto fundamental de interdependencia de las observaciones. Por su parte el AEDE expresa que la dependencia condiciona la localización de los datos, identificando distribuciones en subgrupos. Dichos subgrupos muestran variaciones entre las medias, varianza u otros parámetros. Para la investigación se espera que existan subgrupos de salarios altos en las ciudades ubicadas en el norte del país y salarios bajos en las ubicadas en el sur de México.

A consecuencia de los efectos descritos, el estudio de fenómenos espaciales a través de métodos tradicionales, como pueden ser modelos econométricos de regresiones por Mínimos Cuadrados Ordinarios, suscita problemas en la estimación de los parámetros ya que suelen ser sesgados e ineficientes (Moreno & Vaía, 2000).

Para la presente investigación la dependencia espacial y la heterogeneidad espacial puede llevar a obtener nuevas preguntas de investigación e hipótesis para los salarios en las ciudades. Por parte de la dependencia identifica la existencia procesos de difusión, de intercambio y transferencia, de interacción o dispersión, en este caso se espera que las

ciudades se rodeen de aquellas que presenten salarios similares entre ellas. Para comprobar esto se aplican diversas herramientas del AEDE.

2.3. AUTOCORRELACIÓN ESPACIAL

La mayoría de estudios de análisis espacial utilizan como principal herramienta los mapas, a través de inspección visual e identificación de patrones las cuales pueden arrojar resultados útiles, pero estos son subjetivos y pueden variar entre investigadores. Al no ser una herramienta rigurosa, en la investigación se complementan herramientas como el estadístico de autocorrelación espacial I-Moran o el C-Geary, a fin de evaluar la significancia y el grado de dependencia espacial de las variables de interés.

Al hablar de autocorrelación espacial se refiere a la correlación de una variable consigo misma en el espacio, esta puede presentar valores negativos o positivos. Se encuentra autocorrelación espacial positiva cuando los valores percibidos en una región dada son altos (o bajos) y los mismos valores presentados en regiones vecinas de igual manera son altos (o bajos), cuando esto ocurre se encuentra significancia de surgimiento de agrupaciones espaciales (Anselin, 2006). En el caso contrario de autocorrelación espacial negativa se encuentran valores en una región altos (o bajos) y en las regiones vecinas se correlacionan valores bajos (o altos) identificando localizaciones atípicas (Celebioglu & Dall'erba, 2009).

La autocorrelación espacial global se evalúa mediante prueba de hipótesis de aleatoriedad espacial en la que se mide el agrupamiento global de la variable en el espacio. El rechazo de la prueba de aleatoriedad espacial sugiere la existencia de un patrón espacial, lo cual permite obtener información que explique el fenómeno de estudio.

2.3.1. ESTADÍSTICO I-MORAN

Se es capaz de evaluar la significancia y la intensidad de la autocorrelación espacial a través del estadístico I-Moran (1948), con esto se diferencia de la dependencia espacial,

la cual es una explicación teórica. En la presente investigación se estima el estadístico I-Moran Global con el que se mide la autocorrelación espacial de manera simultánea considerando la ubicación de la zona metropolitana y los valores de la variable asignados a la ubicación, determina si las aglomeraciones son efecto de dependencia espacial con vecinos o si son distribuciones espaciales casuales que se observan al azar. El indicador evalúa si el patrón de las ubicaciones y sus atributos asociados se encuentran agrupados, dispersos o aleatorios. De acuerdo con (Sergio J. & Brett D., 2010) el indicador se estima mediante la ecuación:

$$I = \left(\frac{n}{S_0} \right) \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} X_i X_j}{\sum_{i=1}^n \sum_{j=1}^n X_i X_j} \quad (1)$$

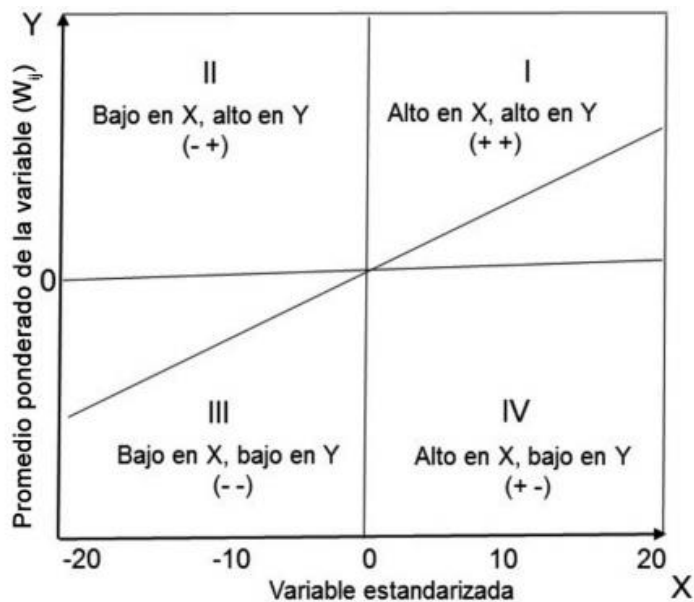
Según (Anselin, 1996) los elementos de la ecuación W_{ij} pertenecen a una matriz binaria de contigüidad asignado a la variable de análisis en el municipio i, j , n es el número de observaciones, S_0 es la suma de los elementos que componen la matriz W . Se recuerda que cuando el número de observaciones es relativamente elevado el indicador no es muy eficiente.

El I-Moran oscila entre +1 y -1. En el caso positivo es autocorrelación espacial positiva lo que quiere decir que los valores del conjunto de los datos tienden a agruparse en el espacio, en la presente investigación se referiría a que, si una zona metropolitana presenta valores elevados en su variable, las zonas metropolitanas de su periferia también presentan valores elevados de la misma variable. Para el caso en el que el I-Moran es negativo se observa autocorrelación espacial negativa, en la cual los valores asociados a una ubicación alta rechazan valores cercanos a la ubicación altos, es decir si una zona metropolitana presenta valores altos de una variable, las zonas metropolitanas de su periferia presentarían valores bajos de la misma variable. El índice será cercano a cero cuando no se encuentre evidencia de autocorrelación espacial, los valores están dispersos, es decir que existe equilibrio entre los valores negativos y los positivos de los productos cruzados.

La forma en la que se acepta la significancia del estadístico es a través de prueba de hipótesis en la cual la hipótesis nula (H_0) establece que los valores de las ubicaciones se encuentran distribuidos de forma aleatoria entre la región de estudio. La aceptación o rechazo de la H_0 se determina con el valor obtenido del P-Valuó para el intervalo de confianza deseado; si el P-Valuó del test es estadísticamente significativo se rechaza la H_0 , lo cual indica que no existe evidencia de que la distribución sea aleatoria (Anselin, 1993).

Una herramienta de visualización del estadístico I-Moran es a través del diagrama de dispersión de Moran o *Moran Scatterplot* (Anselin, 1996). “El scatterplot de Moran es una representación del promedio estandarizado de los vecinos contra la variable estandarizada con una línea de regresión con pendiente igual al I-Moran” (Pérez Mendez, 2013). El diagrama permite posicionar a las regiones en uno de 4 cuadrantes, en donde en el eje de las abscisas se representan las observaciones normalizadas de la variable para cada región y por su parte en el eje de las ordenadas se encuentran el rezago espacial de la variable.

FIGURA 1: CUADRANTES DEL DIAGRAMA DE DISPERSIÓN DE MORAN



FUENTE: (RODRÍGUEZ-LICEA, 2016)

Si la región se encuentra en el primer cuadrante correspondiente a valores “alto-alto” se interpreta que la región pose un valor alto de la variable y además las regiones vecinas

también cuentan con valores altos. En el segundo cuadrante se encuentran regiones correspondientes a valores “Bajo-Alto”, dicha región presentará valores bajos de la variable, pero se encontrará rodeada de regiones con valores altos. En el tercer cuadrante se encuentran valores “Bajos-Bajos” lo que quiere decir que la región tiene valores bajos de la variable, y sus vecinos también. Para el cuarto cuadrante existen valores “Alto-Bajo” donde la región presenta valores altos de la variable, pero sus vecinas valores bajos. Se encontrará correlación espacial positiva cuando las regiones tiendan posicionarse en los cuadrantes uno y tres, por su parte esta será negativa cuando se ubiquen en el segundo y cuarto cuadrante (Gerónimo Antonio, 2014).

2.3.2. ESTADÍSTICO C-GEARY

Otra prueba estadística de autocorrelación espacial es el índice de Geary (C). El índice mide el grado de asociación entre los puntos ponderados y el resto de ellos asignados a un radio de distancia desde el uno original, por lo que es posible obtener índice asociado a cada punto, ello permite detectar *hot spots* o puntos de calor (Ord & Getis, 1995). El índice mantiene como hipótesis nula la presencia de no autocorrelación espacial se estima con la siguiente ecuación:

$$G_i = \frac{(N-1) [\sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2]}{2(\sum_{i=1}^n \sum_{j=1}^n W_{ij})(\sum_{j=1}^n (X_i - \bar{X})^2)} \quad (2)$$

En la ecuación, N representa la población, X_i el valor otorgado a la variable ubicada en el punto, X_j el valor de la misma variable en otro punto, \bar{X} la media del valor de la variable en el espacio de estudio y W_{ij} el valor ij de la matriz de pesos espaciales.

Los resultados del índice oscilan entre 0 y 2; si el resultado es <1 se estará en presencia de autocorrelación espacial positiva, en este indicador se interpretará como que el punto se rodea de valores altos. Si el valor obtenido de indicador es igual a 1 no existe

autocorrelación espacial. Y, por último, si el valor es >1 significara que el punto tiene asociación espacial con valores bajos, en este caso autocorrelación espacial negativa (Griffith, 1987).

2.4. MATRICES DE PESOS ESPACIALES

Para la creación de la matriz W es necesario saber que existen 2 criterios de asociación espacial, el primero es por vecindad la cual se entiende como la existencia de fronteras entre las unidades espaciales o que estas se encuentran posicionadas dentro de una distancia asignada, el segundo es por distancia entre las unidades espaciales el cual se basa en geoestadística. Para la presente investigación se decidió trabajar con el criterio de vecindad en base a distancia asignada a causa que es de los criterios más utilizados en investigaciones revisadas debido a que permite un mejor trabajo de unidades espaciales como polígonos. La matriz W es llamada matriz de pesos espaciales, de retardos o de contigüidad, y es definida como “una matriz cuadrada no estocástica cuyos elementos reflejan la intensidad de la interdependencia existente entre cada par de municipios i y j ” (Moreno & Vayá, 2000, p. 23). Anselin (1989) comenta que no existe una forma generalmente aceptada para la construcción de la matriz, al menos que los valores no sean negativos. La forma más utilizada se encuentra basada en la continuidad o vecindad física de las unidades, donde los elementos que componen la matriz son igual a 1 si las unidades i y j comparten una vecindad establecida y este elemento será igual a 0 de lo contrario, para este caso la diagonal de la matriz estará integradas por ceros ya que la unidad espacial no será considerada vecina de sí misma. Explicado esto la forma de la matriz es la siguiente:

$$W = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{bmatrix} \quad (3)$$

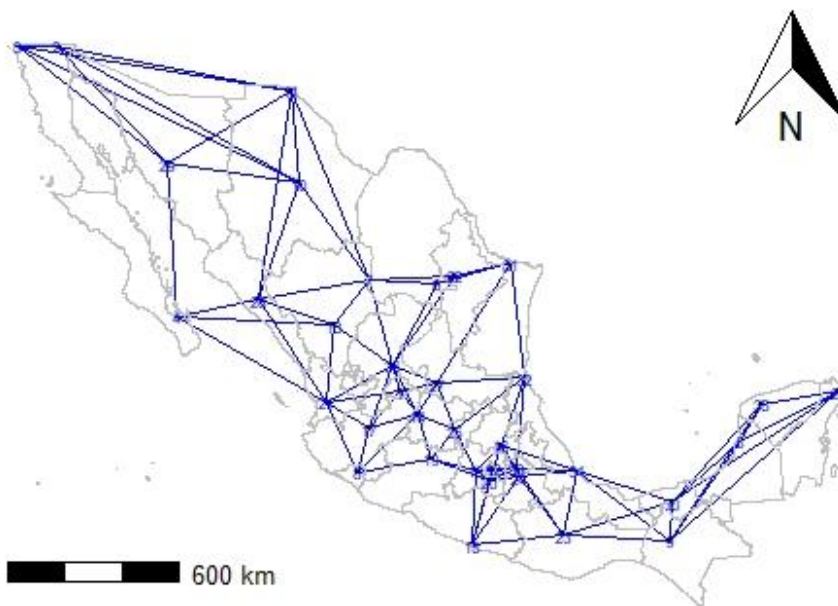
Las conceptualizaciones más utilizadas se basan en la distancia fija, distancia inversa, tiempo de viaje, vecinos K más próximos y contigüidad física:

- Criterio lineal: las unidades i - j son vecinas si comparten el lado izquierdo o derecho

- Criterio torre (rock): existe contigüidad de i con j si estos comparten algún lado
- Criterio alfil (bishop): las unidades i - j son vecinas si comparten algún vértice
- Criterio reina (queen): son vecinas las unidades i - j que compartan un lado o vértice
- Criterio de distancia mínima: tras establecer un radio de influencia γ , todo j compartirá vecindad con i si se encuentra dentro del radio de influencia γ
- Criterio de K-ésimo vecino: cada unidad espacial i cuenta con el mismo número de vecinos asignados (j) siempre que estos sean los más cercanos, determinado por la cercanía de las unidades.

En la presente investigación se determinó trabajar con el criterio de vecindad del K-ésimo vecino más cercano ya que las ciudades estudiadas no presentan colindancia física entre ellas y además las distancias en las que se posicionan unas de otras presentan una alta varianza, por lo que ciudades lejanas no contarían con vecinos. La matriz de pesos espaciales utilizada cuenta con 36 regiones correspondientes a las ciudades auto representativas. y cada ciudad tendrá 4 vecinos. El mapa de conexiones utilizado es el siguiente:

MAPA 1: CONEXIONES DE VECINDAD EN CIUDADES DE MÉXICO CON CRITERIO K-ENESIMO VECINO = 4



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO EN DATOS DEL INEGI

En el mapa se aprecia que existe una distribución homogénea de vecindades para evitar sesgos entre las ciudades que se encuentren muy alejadas y aquellas que estén aglomeradas en el centro.

2.5. MODELOS DE REGRESIÓN ESPACIAL

Para un análisis detallado del ingreso y las variables consideradas determinantes, se realizan, describen e interpretan distintos modelos de econometría espacial y uno de regresión simple, estos tienen por finalidad realizar estimaciones, presentar impactos y coeficientes.

La principal diferencia entre Modelo de Mínimos Cuadrados Ordinarios y los modelos de econometría espacial radica en que los últimos incluyen un rezago espacial con el que se consiguen estimaciones más eficientes para los casos en los que los datos presenten dependencia entre las variables (especialmente dependencia espacial). Además, se consideran más eficientes en presencia de heterogeneidad espacial (ambos casos rompen los supuestos de los MCO por lo que generarían estimaciones sesgadas) (Anselin, 2006).

Los modelos que son capaces de generar coeficientes insesgados en caso de dependencia por fuentes exógenas y/o endógenas son: Modelos Espacial Autorregresivo (SAR), Modelo de Error Espacial (SEM), Modelo Espacial General (SAC) y Modelo Espacial de Durbin (SDM) (LeSage, 2008).

En muchas ocasiones, la dependencia espacial de las variables puede ser explicada por los fenómenos económicos por lo que, para evitar sesgos, se agrega el rezago espacial incorporando un proceso autorregresivo ya sea en la variable dependiente o en el término de error (Elhorst, 2003). En caso de no agregar el proceso autorregresivo y exista dependencia espacial, los coeficientes serán inconsistentes.

En la forma matemática del proceso de dependencia es necesario agregar un parámetro escalar ρ , llamado coeficiente autorregresivo, y una matriz de pesos espaciales de $n \times n$ llamada W (LeSage & Pace, 2009). La forma de la ecuación es la siguiente:

$$Y = \alpha l + \rho W Y + \varepsilon \quad (4)$$

2.5.1. MODELO DE MÍNIMOS CUADRADOS ORDINARIOS

Este modelo lineal es el más utilizado en la econometría, debido a que es considerado de los más eficientes, siempre y cuando se cumplan los supuestos clásicos de los MCO. De ser así los parámetros y estimaciones que resultan son insesgadas y eficientes. La siguiente ecuación presenta la forma general del modelo de regresión:

$$Y = X\beta + \varepsilon \quad (5)$$

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

En la ecuación, la variable dependiente se presenta con Y , y corresponde a un vector de $n \times 1$. X corresponde a una matriz de $n \times k$ que muestra las observaciones de las variables independientes del modelo. β es el parámetro asociado a las variables independientes. ε es un vector de $n \times 1$ que corresponde a las perturbaciones no observables.

2.5.2. MODELO AUTORREGRESIVO ESPACIAL (SAR)

En este modelo, con la finalidad de verificar si existe efecto sobre la variable dependiente, se les aplica un proceso autorregresivo a las variables independientes. Este tipo de modelos busca explicar cómo las variaciones de Y son afectadas por el promedio de la misma variable independiente de las regiones vecinas y un grupo de variables independientes (LeSage & Pace, 2009). Las ecuaciones del modelo y el proceso generador de datos son:

$$Y = \sigma_n + \rho WY + X\beta + \varepsilon \quad (6)$$

$$Y = (I_n - \rho W)^{-1} (\sigma_n + X\beta) + (I_n - \rho W)^{-1} \varepsilon$$

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

En la ecuación anterior, Y corresponde al vector de observaciones de la variable dependiente, de tamaño $n \times 1$. X corresponde a una matriz de observaciones de las variables exógenas de tamaño $n \times k$ asociada a su parámetro β . W corresponde a una matriz de pesos espaciales de tamaño $n \times k$. ρ es el coeficiente espacial autorregresivo. ε es el vector correspondiente a las perturbaciones no observables de tamaño $n \times 1$. I_n es un vector de unos.

Para una correcta interpretación de los coeficientes, esta no debe ser directa, como efectos marginales, debido a la endogeneidad causada por la dependencia espacial; es por ello que se requiere estimar impactos directos e indirectos.

En los casos de dependencia espacial ($\rho \sim 0$), las variaciones en una región se difunden a las regiones vecinas, posteriormente las variaciones en los vecinos repercuten nuevamente en la región, de modo tal que se generan impactos directos (variaciones en la variable dependiente son producto de influencias locales) e indirectos (variaciones en la variable dependiente son producto de influencias vecinas) (LeSage & Pace, 2009). Este fenómeno de interacción endógena entre regiones se le denomina *spillover espacial*.

2.5.3. MODELO DEL ERROR ESPACIAL (SEM)

En este modelo, supone que no existe dependencia espacial en las variables dependientes ni independientes, si no que esta se puede contenerse en variables no observables que influyen en el comportamiento. Obedeciendo a la omisión de variables no esenciales captadas en el término de error, plantea que la dependencia espacial se encuentra acaecida en el término de error (LeSage & Pace, 2009) . Las ecuaciones del son:

$$Y = \sigma_n + X\beta + u \quad (7)$$

$$u = \lambda W u + \varepsilon$$

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

En la ecuación anterior, Y corresponde al vector de observaciones de la variable endógena, de tamaño $n \times 1$. X corresponde a una matriz de observaciones de las variables exógenas de tamaño $n \times k$ asociada a su parámetro β . ε es el vector correspondiente a las perturbaciones no observables de tamaño $n \times 1$. I_n es un vector de unos. λ es el parámetro autorregresivo de los errores espacialmente correlacionados

Para la correcta interpretación de los coeficientes estimados, esta se realiza de manera directa, como efectos marginales, al igual que los modelos MCO, ya que no existen procesos endógenos que generen dependencia entre las variables (LeSage & Pace, 2009).

2.5.4. MODELO ESPACIAL AUTORREGRESIVO COMBINADO (SAC)

El modelo considera dependencia espacial tanto en la variable dependiente como en el término de error. Al presentar dependencia en las variables, la interpretación no puede ser directa, por lo que requiere la estimación de los efectos directos e indirectos. Este modelo se considera equivalente a otro tipo de modelos espaciales autorregresivos de media móvil (SARMA) (LeSage & Pace, 2009). En este modelo tanto los parámetros como los coeficientes son iguales a los de los modelos descritos anteriormente. Las ecuaciones del modelo y el proceso generador de datos son:

$$Y = \sigma I_n + \rho W_1 Y + X\beta + u \quad (8)$$

$$u = \theta W_2 u + \varepsilon$$

$$Y = (I_n - \rho W_1)^{-1} (\sigma I_n + X\beta) + (I_n - \rho W_1)^{-1} (I_n - \theta W_2)^{-1} \varepsilon$$

En la ecuación anterior, Y corresponde al vector de observaciones de la variable dependiente, de tamaño $n \times 1$. X corresponde a una matriz de observaciones de las variables exógenas de tamaño $n \times k$ asociada a su parámetro β . W_1 y W_2 corresponde a una matriz

de pesos espaciales de tamaño $n \times k$. ρ y θ son parámetros del coeficiente espacial autorregresivo. ε es el vector correspondiente a las perturbaciones no observables de tamaño $n \times 1$. I_n es un vector de unos.

En este modelo, es importante conocer que la matriz $W1$ y $W2$ pueden ser iguales, la interpretación de los parámetros no se efectúa de manera directa, a causa de la endogeneidad en el proceso autorregresivo.

2.5.5. MODELO DE REZAGO ESPACIAL EN X (SLX)

La principal característica de este modelo es que considera los cambios en la variable endógena se originan por cambios en la misma región como en las regiones vecinas. Es así como las características de los vecinos se consideran una variable explicativa, por este motivo el modelo parametriza la matriz de pesos espaciales (LeSage & Pace, 2009). En este modelo tanto los parámetros como los coeficientes son iguales a los de los modelos descritos anteriormente. Las ecuaciones del modelo son:

$$Y = \sigma I_n + X\beta_1 + \beta_2 + \varepsilon \quad (9)$$

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

En la ecuación anterior, Y corresponde al vector de observaciones de la variable dependiente, de tamaño $n \times 1$. X corresponde a una matriz de observaciones de las variables exógenas de tamaño $n \times k$ asociada a su parámetro β_1 . WX corresponde al promedio espacial de las características de la región vecina, asociada a su parámetro β_2 . ε es el vector correspondiente a las perturbaciones no observables de tamaño $n \times 1$. I_n es un vector de unos.

La interpretación de los coeficientes del modelo se efectúa de manera directa a causa de que no existe un proceso autorregresivo por endogeneidad espacial. Tanto el efecto directo como indirecto son capturados en los coeficientes los cuales pueden estimarse

eficientemente a través de MCO, interpretándolos como efectos marginales (LeSage & Pace, 2009).

2.5.6. MODELO ESPACIAL DE DURBIN (SDM)

El modelo SDM, es considerado un modelo efectuado por manipulaciones en el proceso generador de datos de los modelos SEM y SAR. El modelo presenta rezago espacial las variables independientes y en la dependiente y busca capturar los *spillovers* entre las regiones vecinas (LeSage & Pace, 2009). Las ecuaciones del modelo y el proceso generador de datos son:

$$Y = \sigma I_n + \rho WY + X\beta + WX_\gamma + \varepsilon \quad (10)$$

$$Y = (I_n - \rho W)^{-1} (\sigma I_n + X\beta + WX_\gamma + \varepsilon)$$

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

En la ecuación anterior, Y corresponde al vector de observaciones de la variable dependiente, de tamaño $n \times 1$. X corresponde a una matriz de observaciones de las variables exógenas de tamaño $n \times k$ asociada a su parámetro β . W corresponde a una matriz de pesos espaciales de tamaño $n \times k$. ρ y γ son los parámetros del coeficiente espacial autorregresivo de la variable dependiente y la explicativa, respectivamente. ε es el vector correspondiente a las perturbaciones no observables de tamaño $n \times 1$. I_n es un vector de unos.

LeSage & Pace (2009), argumentan que el modelo SDM suele superar en eficiencia los modelos SAC, SEM y SAR. Comentan que se capturan mejor los efectos de la variable omitida, creando estimadores insesgados al agregar rezagos espaciales tanto en la variable dependiente como en las independientes. La interpretación de los parámetros se realiza a través de los efectos directos e indirectos, a consecuencia de la endogeneidad espacial. Los efectos indirectos contempla efectos endógenos y exógenos de manera conjunta, por lo que se representan mejor los *spillovers* espaciales (LeSage & Pace, 2009).

2.6. INTERPRETACIÓN DE COEFICIENTES O EFECTOS

Como ya se describió, algunos modelos permiten una interpretación directa siempre y cuando se cumplan los supuestos de MCO, pero aquellos modelos en los que exista endogeneidad espacial se necesita la estimación e interpretación de efectos directos e indirectos.

Los modelos que permiten una interpretación directa como derivadas parciales $\left(\frac{\partial y_i}{\partial x_{i\gamma}} = \beta_\gamma\right)$ a causa de la exogeneidad de las variables, lo cual significa que variaciones en la variable dependiente no generan efectos en las variables de otras regiones $\left(\frac{\partial y_i}{\partial x_{j\gamma}} = 0 \quad \forall j \neq i\right)$.

2.7. DATOS Y VARIABLES

Para el caso de México, los datos públicos relacionados al empleo se obtienen de la Encuesta Nacional de Ocupación y Empleo (ENOE). Los datos recolectados de ella son presentados en 2 secciones; la primera muestra una base de datos de corte transversal para el año 2019; ya que es el año más actual y con el que se muestran más observaciones (36 ciudades), dicho año se contrasta con el año 2009, con fin de mostrar los efectos ocasionados por la crisis económica del 2008. La segunda parte presenta un análisis descriptivo de las variables que abarca del año 2005 al 2019, se utiliza este periodo de análisis ya que son los años que se presentan en la ENOE.

En cuanto a la temporalidad descrita, es importante mencionar que a ENOE presenta sus datos de forma trimestral, sin embargo, para el análisis realizado no se consideraron los 4 trimestres que integran cada año. Si bien, poder utilizar los 4 trimestres aumentaría el número de datos, no se optó por esto debido a que la variabilidad de los datos que se presentan cada trimestre es baja (Castro Lugo, 2005). En base a esto se optó por trabajar con el tercer trimestre de cada año, correspondiente a los meses de julio, agosto y

septiembre, ya que Castro (2005) argumenta que es este trimestre el que presenta menor estacionalidad de las variables a comparación de los demás.

Las variables que se considera como variable dependiente es el ingreso promedio por hora.

Las variables consideradas independientes son:

- Población total
- Años promedio de escolaridad
 - Tasa de población con posgrado
 - Tasa de población con primaria incompleta
- Edad promedio
 - Tasa de población joven
 - Tasa de población en tercera edad
- Tasa de desocupación
- Tasa de trabajadores trabajando en el sector informal
- Tasa de trabajadores trabajando en industria manufacturera

2.7.1. ENOE

La ENOE es una encuesta actual en México capaz de presentar información de manera trimestral y consecutiva hasta la fecha. Actualmente, es considerada la encuesta levantada en hogares más grande del país. Surge en enero del 2005 como una herramienta un nuevo modelo de la encuesta sobre la fuerza de trabajo en México, finalizando las anteriores encuestas de fuerza laboral con modelos de capacitación y procesamiento aplicados durante 20 años correspondientes a la Encuesta Nacional de Empleo Urbano (ENEU) seguida por la Encuesta Nacional de Empleo (ENE) en donde aquélla quedó integrada.

La implementación de ella continuo con las consideraciones conceptuales aplicadas en la ENE y la ENUE. Con la creación del nuevo modelo de la ENOE se esperaba dar respuesta a las necesidades de los cambios operados en el mercado a partir de procesos de calidad.

Para el origen se aplicaron recomendaciones internas (institutos, gobiernos federales y gobiernos estatales) y externas por parte de organismos internacionales⁷.

La ENOE es una encuesta realizada a través del INEGI, la cual expone como objetivo general el “Obtener información estadística sobre las características ocupacionales de la población de 15 años y más a nivel nacional, entidad federativa y ciudades, así como de variables demográficas y económicas para el análisis de la fuerza de trabajo, la toma de decisiones, el diseño y seguimiento de las políticas laborales.” (INEGI, 2019). Dentro de sus objetivos específicos resultan de interés para la investigación los siguientes:

- Captar información sobre el tamaño, composición y distribución de la fuerza de trabajo y de los niveles de participación en la actividad económica.
- Contar con cifras del volumen de población desocupada, sus características y de los niveles de desocupación de las personas.
- Recabar datos sobre la inserción laboral, calificación y relaciones laborales de la fuerza de trabajo.
- Conocer las condiciones laborales: salarios, horas trabajadas, prestaciones, seguridad social y forma de contratación de los trabajadores.
- Disponer de estimaciones de la subocupación e informalidad laboral de las personas que actualmente están ocupadas.
- Proveer de información para el análisis de la dinámica y tendencias del mercado laboral mexicano (INEGI, 2019)

La clasificación de la encuesta se encuentra ordenada y cinco tablas de datos para almacenar la información recabada, cada una muestra información particular utilizada de acuerdo al fin de interés, las cinco se relacionan entre sí a través de campos de identificación comunes, iguales entre ellas, denominados campos de identificación o llave primarias. Las tablas son: la tabla de vivienda (VIVT), la de hogares (HOGT), la de las características

⁷ El diseño conceptual de la encuesta va acorde a las recomendaciones internacionales establecidas por la Organización Internacional del Trabajo (OIT) y los lineamientos de la Organización para la Cooperación y el Desarrollo Económicos (OCDE). La ENOE toma como base las resoluciones en materia de estadísticas del trabajo de la 13a, 15a, 16a y 18a Conferencia Internacional de Estadísticos del Trabajo (CIET) (INEGI, 2019).

sociodemográficas de los residentes (CS), la del cuestionario de ocupación y empleo de las personas de 12 años y más (COE1T y COE2T)⁸.

Todas las tablas permiten recabar información a distintos niveles de profundidad, en primera instancia existen salidas fijas de información que se encuentran en los tabulados básicos de las tablas y en los indicadores estratégicos. Por su parte, los indicadores estratégicos tienen aplicaciones como InfoLABORAL, o Tabulados interactivos, los cuales dan posibilidad a los usuarios de utilizar variables precodificadas y cruce de demás variables.

Además de esta información, a los usuarios de la encuesta les es posible el generar información a partir del nivel de microdatos, construyendo a partir de campos específicos llaves únicas con el fin de otorgar una identidad digital irrepetible al microdato (INEGI, 2010).

Otro nivel de información que es posible trabajar en la encuesta son los campos precodificados, considerados por INEGI (2010) como un segundo nivel de información; con ellos es posible identificar a que grupo de población (ocupado, desocupado, no económicamente activo) pertenece cada microdato y, a su vez, a qué subconceptos corresponde dentro de estos grupos, por ejemplo, si es ocupado, a través de la precodificación, se precisa si el microdato cumple con las condiciones de un subocupado, un trabajador en el sector informal o el de un asalariado, así como otros rasgos más que se vierten y presentan en los indicadores estratégicos, tales como el sector en el que se realiza su actividad (INEGI, 2010).

Con fin de ampliar la muestra a un estimado total de la población nacional, para la investigación es de vital relevancia la aplicación del factor expansivo que se presenta en las observaciones de la ENOE. La encuesta lo define como *“el coeficiente que le otorga determinado peso en la muestra en función de su representatividad de otros tantos casos*

⁸ Las tablas pueden consultarse en la página <http://www.beta.inegi.org.mx/proyectos/enchogares/regulares/enoe/> donde además se presentan documentos de interés para el manejo de las bases de manera mas detallada.

similares a él, tomando en cuenta su estrato socioeconómico y lugar de residencia.” (INEGI, 2010, p. 1)

La encuesta presenta dos versiones, diferenciadas por las características de sus factores de expansión, la finalidad es ofrecer las mismas bases de datos ajustadas al insumo demográfico oficial más reciente, de manera que pueda formar series homogéneas más largas (INEGI, 2010). La primera versión es denominada “serie original” en la cual su factor expansivo se encuentra ajustado a las proyecciones demográficas que tienen como base el Segundo Censo General de Población y Vivienda de 2005, la serie abarca el periodo de tiempo del primer trimestre de 2005 y cierra en el cuarto trimestre de 2010. La segunda versión de datos presenta el factor expansivo ajustado a las proyecciones demográficas que tienen como base el Censo de Población y Vivienda de 2010, utilizada del primer cuatrimestre del 2011 a la fecha.

2.7.2. VARIABLES

Las variables que presentan las bases de datos son: ingresos por hora, población total en la ciudad, promedio de la edad, tasa de jóvenes, tasa de adultos mayores, promedio de años escolaridad, tasa de población con posgrado, tasa de población con primaria incompleta, tasa de desocupación y tasa de informalidad.

La presentación de estas variables se encuentra en 3 secciones, la primera tiene por finalidad el contextualizar un panorama nacional. Las siguientes se desagregan a nivel de ciudad auto representativa, primero un análisis descriptivo para los años disponibles (2005-2019) seguido de análisis de corte transversal para el 2019, comparándolo con el año 2009.

Para la obtención de las variables mencionadas se realizaron técnicas de reconstrucción de variables y métodos para comparación de ellas. Se inicio el manejo de las variables

extrayendo de la tabla de datos de SOCIODEMOGRÁFICO. Se extrajo el tercer trimestre correspondiente a cada año⁹.

2.7.3. CONSTRUCCIÓN DE VARIABLES

La reconstrucción de las variables de interés se realizó a través del software “STATA”. Con el uso del mismo, se utilizó el factor expansivo (FAC) presente para cada individuo, con el fin de lograr una estimación total en las ciudades auto representativas. La estimación se realizó con el comando [fw=fac]. Además de la ampliación de las observaciones, el FAC permitió el obtener la población total, consiguiendo desagregarla a cada ciudad y a nivel nacional. La siguiente tabla muestra el contraste entre el número de observaciones proporcionadas inicialmente por la encuesta (muestra) y el número total de observaciones al aplicar el factor expansivo (población total)¹⁰.

TABLA 1: CONTRASTE NACIONAL ANUAL DEL NÚMERO DE OBSERVACIONES SIN Y CON APLICACIÓN DE FAC

AÑO	SIN FAC	CON FAC	% DE MUESTRA	VECES AUMENTADA LA MESTRA
2005	421,751	107,306,264	0.3930%	254.43
2006	423,305	108,577,054	0.3899%	256.50
2007	418,327	109,974,218	0.3804%	262.89
2008	410,219	111,498,454	0.3679%	271.80
2009	402,919	113,042,774	0.3564%	280.56
2010	405,533	114,468,134	0.3543%	282.27
2011	399,716	115,857,912	0.3450%	289.85
2012	397,893	117,226,029	0.3394%	294.62
2013	394,472	118,564,077	0.3327%	300.56

⁹ Cada base de datos extraída contaba con un total de 104 columnas, al depurar las bases de datos se contó con únicamente 24 columnas. El manejo de depuración de la base de datos originales se realizó a través del software Excel. Los Mnemónico que se utilizaron y la descripción se presentan en el anexo 2.

¹⁰ Se presentan los datos a nivel nacional. En promedio, del 2005 al 2019, la muestra presentaba únicamente el 0.3455% de la población total estimada. En promedio, el factor expansivo ampliaba casi 290 veces las observaciones totales de la ENOE.

2014	405,803	119,879,581	0.3385%	295.41
2015	401,825	121,172,342	0.3316%	301.56
2016	391,934	122,434,046	0.3201%	312.38
2017	392,178	123,675,351	0.3171%	315.36
2018	390,837	124,892,308	0.3129%	319.55
2019	405,449	125,994,560	0.3218%	310.75

FUENTE: ELABORACIÓN PROPIA CON BASE A DATOS DE LA ENOE

2.7.3.1. AÑOS PROMEDIO DE ESCOLARIDAD

Como variables proxys a la formación de capital de humano se reconstruyeron 3 variables: años promedio de escolaridad, tasa de población con posgrado y tasa de la población con primaria incompleta. Para la obtención de la primera se utilizó el comando “*tab cd_a [fw=fac] if anios_esc<25 & eda>14 & eda<98, sum (anios_esc)*”, en el cual se indicó considerar únicamente los años de escolaridad de la población de 15 años en adelante. De la tabla dada por el comando se extrajo la información correspondiente a la media por ciudad.

2.7.3.2. TASA DE LA POBLACIÓN CON POSGRADO

Para la obtención de la tasa de la población con posgrado, en primera instancia se obtuvo la población con posgrado por ciudad mediante el comando “*tab cd_a [fw=fac] if cs_p13_1>07 & cs_p13_1<10*”, con el cual se indica que considere la población únicamente si aprobó el grado de maestría o doctorado en la escuela, la condición “>10” indica que no se considere la población que presenta como respuesta “no sabe”. Al obtener la población con posgrado y la población total, por ciudad, se realizó el cociente de para finalmente obtener la tasa de población con posgrado por ciudad.

2.7.3.3. TASA DE POBLACIÓN CON PRIMARIA INCOMPLETA

En cuanto a la tasa de población con primaria incompleta se realizó un procedimiento similar, el comando utilizado es “`tab cd_a [fw=fac] if niv_ins=1 & eda>14`”. El comando señala que se obtenga la población por ciudad si el nivel de instrucción es igual a 1, es decir que se cuente con primaria incompleta y además la edad sea mayor a 14 años. Posteriormente se obtiene el cociente entre la población con primaria incompleta ya la población total, por ciudad.

2.7.3.4. EDAD PROMEDIO

Las variables consideradas proxy a la experiencia son: promedio de la edad, tasa de población entre los 15 y 34 años (jóvenes) y tasa de la población mayor a 65 años (adultos mayores). Para la obtención de la edad media por ciudad, únicamente se consideró a la población de 15 años en adelante. El comando utilizado para la reconstrucción de la media de la edad fue: “`tab cd_a [fw=fac] if eda>14 & eda<98, sum (eda)`”. De la tabla obtenida con el comando se extrajeron los datos correspondientes a la media por ciudad.

2.7.3.5. TASA DE POBLACIÓN JOVEN

La estimación de la tasa de jóvenes se realizó aplicando el comando “`tab cd_a [fw=fac] if eda12c <5 & eda>14`”, en donde se especifica que se obtenga la población por ciudad considerando que la variable 12c sea menor a 5; lo que indica que sea población menor a 35 años de edad, y que la variable de eda sea mayor a 14 años. Posteriormente a conseguir la población entre 15 y 35 años de edad, se obtuvo el cociente entre esta y la población total de cada ciudad.

2.7.3.6. TASA DE POBLACIÓN EN TERCERA EDAD

Para la variable de tasa de adultos mayores el comando utilizado fue `"tab cd_a [fw=fac] if eda12c==11"` con el que se obtiene la población por ciudad que cumple el criterio de que la variable `eda12c` sea igual a 11, donde el 11 equivale a la población de 65 años y más. La tasa se obtiene de realizar el cociente entre la población obtenida con la población total por ciudad.

2.7.3.7. TASA DESOCUPACIÓN

Con el fin de corroborar la relación entre el salario y la tasa de desempleo, según consideraciones teóricas de la curva de salario expuestas por diversos autores como Blanchflower & Oswald (1996), Jiménez (2001) y Castellanos (2009), la presente investigación considera como variable la tasa de desempleo.

La tasa de desempleo fue una variable reconstruida de la ENOE mediante la aplicación del comando: `"tab cd_a [fw=fac] if clase2==2"`. El criterio de que `clase2=2` hace referencia a que se considere la población desocupada, al cumplirse este, junto con el resto del comando se permite obtener la población desocupada por ciudad. Para la obtención de la tasa de desocupación es necesario obtener el cociente de la población económicamente activa (PEA) y la población desocupada. Para obtener la PEA se utiliza el comando `"tab cd_a [fw=fac] if clase1==1"`, el cual cumple el criterio (`clase1==1`) para obtener la obtención. Posterior se realiza el cociente y se consigue la tasa de desocupación desagregada por ciudad.

2.7.3.8. TASA DE TRABAJADORES EN SECTOR INFORMAL

A fin de mostrar la relación existente entre el salario percibido en las ciudades con la informalidad laboral se obtiene una variable proxy al empleo informal¹¹. Para la

¹¹ También conocido como empleo atípico por la Organización Internacional del Trabajo (OIT).

construcción de esta tasa se consideró a los empleados asalariados informales, quienes son aquellos que no cuentan con contratos formales, beneficios o protección social y de salud que trabajan en empresas informales o formales (OIT, 2002). En concreto para cumplir con esto se considera de la ENOE a aquellos asalariados que no cuentan con ningún contrato escrito o no cuentan con seguridad social.

Para estimar la informalidad con los criterios mencionados se utilizó el comando *"tab cd_a [fw=fac] if tip_con==5 | seg_soc==2"*. Dicho comando otorga la población que cumple los criterios de no presentar contrato escrito cuando la variable "tip_con" sea igual a 5 o el criterio de que el trabajador no presente acceso a instituciones de salud cuando la variable "seg_soc" sea igual a 1. Posteriormente se obtiene la población ocupada con el comando *"tab cd_a [fw=fac] if clase2==1"*. Ya contando con la población asalariada informal y la población ocupada se realiza el cociente de ambas para finalmente contar con la tasa de informalidad por ciudad.

2.7.3.9. TASA DE TRABAJADORES EN INDUSTRIA MANUFACTURERA

Con el fin de representar la estructura económica de las ciudades a través de los trabajadores, se utiliza, como proxy, la tasa de trabajadores laborando en la industria manufacturera, esperando que estos trabajadores en la industria presenten un mayor salario y con ello las ciudades con altas tasas muestren mayores salarios.

En la construcción de la variable de trabajadores en industria manufacturera se aplicó el comando: *"tab cd_a [fw=fac] if rama==2"*. Este comando permite obtener el número de trabajadores que se encuentran laborando en la rama=2, donde 2 se atribuye a aquellos trabajadores que se encuentren laborando en la manufactura. Posteriormente, para la obtención de la tasa se realizó el cociente del número de trabajadores en manufactura y la población ocupada para cada ciudad.

2.7.3.10. SALARIO PROMEDIO POR HORA

Para obtener la variable de mayor interés para la investigación, el promedio del ingreso percibido por hora, se realizó la construcción de datos a partir la ENOE. La reconstrucción de datos de realizo a consecuencia que la variable “ing_x_hrs” mostraba ausencia de dato en casos donde el trabajador si presentaba ingreso (la variable “ing7c” presentaba valores distintos a 0) y duración de la jornada laboral (la variable “dur9c” presentaba valores distintos a 0).

Para estimar observaciones se realizó un ejercicio de aproximación a los datos categóricos utilizados en las variables de ing7c y dur9c. En la variable ing7c, el valor otorgado¹² reflejaba el número de salarios mínimos que percibía el trabajador. Posteriormente el valor se multiplicaba por el salario mínimo vigente actual para el año y para la zona en la que se ubicaba la ciudad autor representativa, este dato la ENOE lo presenta en la variable denominada “salario”. Obteniendo el ingreso mensual en términos monetarios paso a realizarse una división entre 30¹³, para obtener el ingreso diario por trabajador. El número de salarios minimos asignado, según su código valido para captura, se buscó que reflejara el promedio entre los parámetros fijados por la ENOE, estos se exponen en la siguiente tabla:

TABLA 2: NÚMERO DE SALARIOS MÍNIMOS ASIGNADOS SEGÚN LA CLASIFICACIÓN DE LA VARIABLE ING7C DE LA ENOE.

Código valido de ing7c	Descripción del código valido	Número de salarios Mínimos asignados al código
1	Hasta un salario mínimo	1
2	Más de 1 hasta 2 salarios mínimos	1.5
3	Más de 2 hasta 3 salarios mínimos	2.5

¹² En la ENOE se encuentra como “Códigos validados para captura”

¹³ De dividió en 30 ya que fue el número de días que reflejaba el resultado más próximo al expuesto en la variable ing_x_hrs.

4	Más de 3 hasta 5 salarios mínimos	4
5	Más de 5 salarios mínimos	7
6	No recibe ingresos	0
7	No especificado	0

FUENTE: ELABORACIÓN PROPIA CON BASE EN LA ENOE

Para obtener las horas trabajadas al día por trabajador se dividió el número de horas de la jornada laboral entre 7 días de la semana¹⁴. Al igual que en la asignación de ing7c se buscó que las horas de la jornada fueran el promedio de la descripción del código válido. La asignación de valores a la duración de la jornada en la variable “dur9c” fue la siguiente:

TABLA 3: DURACIÓN DE LA JORNADA ASIGNADA SEGÚN LA CLASIFICACIÓN DE LA VARIABLE DUR9C DE LA ENOE.

Código valido dur9c	Descripción del código valido	Horas semanales de la jornada asignadas
1	Ausentes temporales con vínculo laboral	0
2	Menos de 15 horas	10
3	De 15 a 24 horas	20
4	De 25 a 34 horas	30
5	De 35 a 39 horas	37
6	De 40 a 48 horas	44
7	De 49 a 56 horas	52
8	Más de 56 horas	60

¹⁴ De dividió entre 7 ya que fue el número de días que reflejaba el resultado más próximo al expuesto en la variable ing_x_hrs.

FUENTE: ELABORACIÓN PROPIA CON BASE EN LA ENOE

Finalmente, para obtener el ingreso por hora se realizó el cociente entre el salario percibido al día y las horas trabajadas al día. Este cálculo fue utilizado para remplazar el valor “0” en la variable “ing_x_hrs”. Para contextualizar los cambios, en el año 2019 a la variable ing_x_hrs se le agregaron 18,614 datos, los cuales representaron un aumento de casi el 4.6% en las observaciones, se pasó de 121’480 datos a 140’464. El aumento del número de observaciones de la variable se realizó a todos los años.

La obtención del promedio de ingreso por hora por ciudad se obtuvo a través del comando “tab cd_a [fw=fac] if ing>4 & ing<2000, sum (ing)”. En él se aplicó el criterio que el ingreso por hora mínimo considerado fuese mayor a \$2 y que el ingreso por hora máximo considerado fuese de \$2’000 por trabajador, esto con fin de eliminar los outlayers de la muestra. De la tabla adquirida por el comando se extrajo la media por ciudad auto representativa.

El ingreso expuesto por la ENOE se encuentra a precios corrientes, por lo que para realizar un análisis comparativo en la evolución del ingreso fue necesario expresar el ingreso en precios constantes. Para llevar el ingreso a precios constantes, se utilizó el Índice de Precios al consumidor (IPC). El IPC considerado en la investigación fue el observado durante el mes de agosto de cada año, debido a que es el mes intermedio del tercer trimestre (julio, agosto, septiembre) utilizado en la presente investigación.

Después de obtener los IPC de cada año para el mes de septiembre, se realizó un cociente de estos, el año base considerado fue el más actual, 2019. El valor resultante se multiplica por el promedio del ingreso por hora observado en el tiempo t .

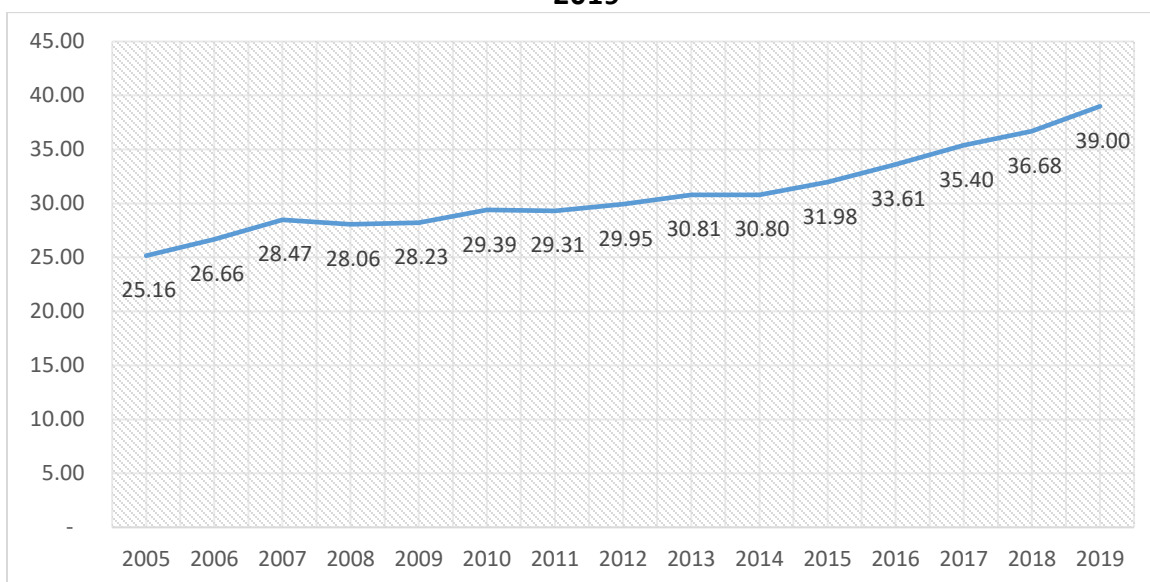
CAPÍTULO III:

RESULTADOS

Con fin de mostrar un panorama del comportamiento de las variables analizadas a través del tiempo se muestran la evolución de estas en el periodo de datos que comprende del 2005 al 2019 a nivel nacional.

En la siguiente grafica se presenta la evolución de la variable de ingreso promedio por hora a nivel nacional del 2005 al 2019. Los resultados a nivel nacional muestran como el ingreso presenta tendencia a la alta, a excepción de una ligera reducción en el año 2008 de \$0.41 pesos. La reducción se puede atribuir a la crisis económica mundial que se atravesaba en dicho año. Otra ligera reducción se muestra en el año 2011.

GRAFICA 1: EVOLUCIÓN DEL INGRESO MEDIO POR HORA EN MÉXICO DEL 2005 AL 2019

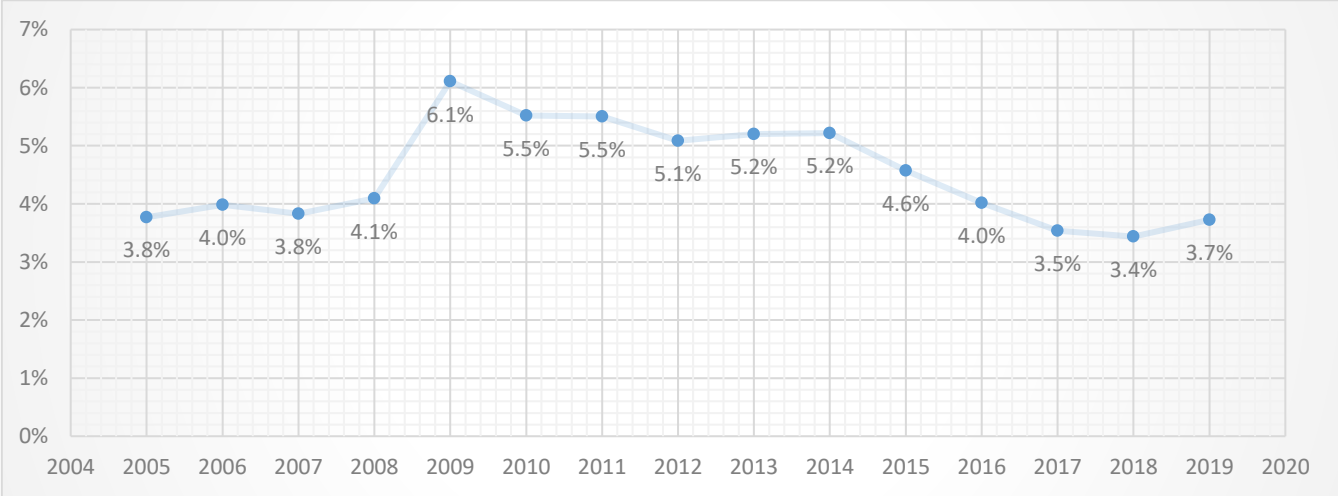


FUENTE: ELABORACIÓN PROPIA CON BASE EN DATOS DE LA ENOE

En la gráfica 2 se muestra el comportamiento de la tasa de desocupación para México en el periodo 2005 al 2019. Se logra observar que los resultados no presentan una tendencia. La oscilación más significativa se observa del 2008 al 2009, la cual representa un aumento de

la tasa de desocupación del 2%, este incremento es principalmente provocado por la contracción económica mundial a consecuencia de la crisis económica del 2008. Posterior al 2009 la tasa de desocupación presenta una tendencia a la baja, a excepción del 2013 que aumento en 0.1% y 2019 que aumento en 0.3%.

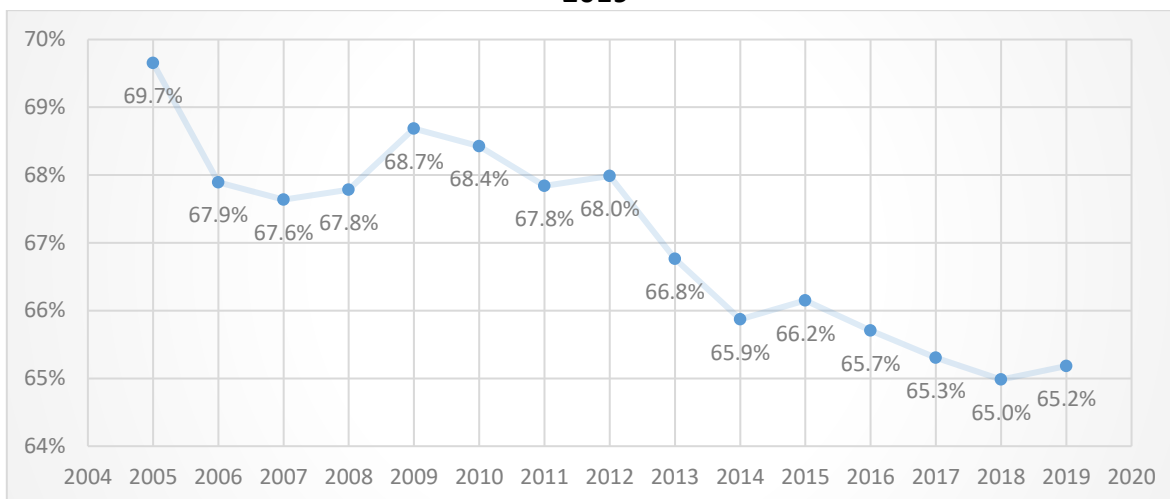
GRAFICA 2: EVOLUCIÓN DE LA TASA DE DESOCUPACION EN MÉXICO DEL 2005 AL 2019



FUENTE: ELABORACIÓN PROPIA CON BASE EN DATOS DE LA ENOE

La grafica 3 muestra la evolución de la tasa de informalidad para México del 2005 al 2019, Se observa una tendencia a la baja de esta. Es destacable un aumento en la tasa de informalidad del 2008 al 2009 de 0.9%, dicho incremento puede atribuirse a la crisis económica mundial del 2008. También se observan aumentos en la tasa de informalidad en 2012, 2015 y 2019 de 0.2%, 0.3% y 0.2% respectivamente.

GRAFICA 3: EVOLUCIÓN DE LA TASA DE INFORMALIDAD EN MÉXICO DEL 2005 AL 2019



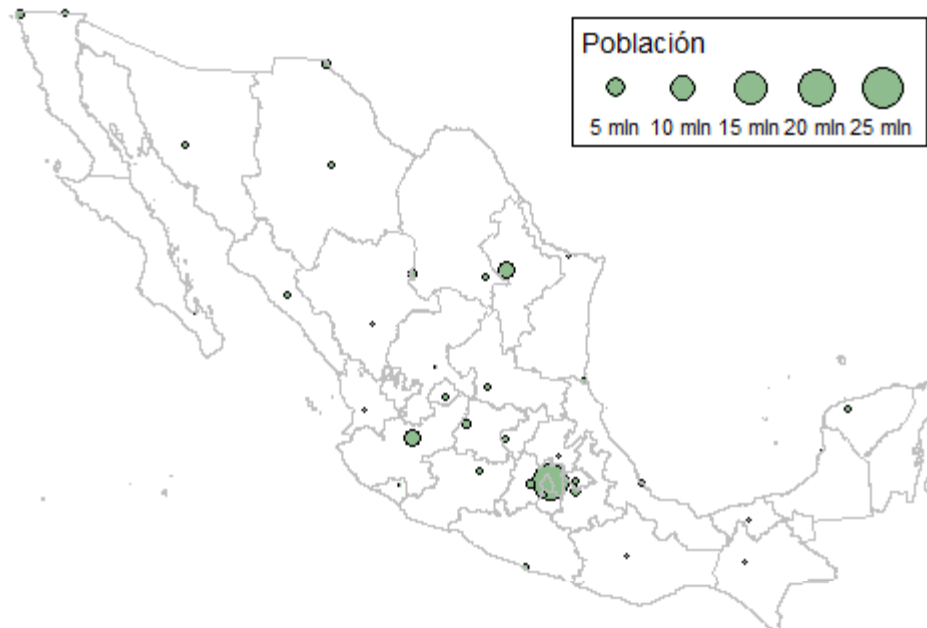
FUENTE: ELABORACIÓN PROPIA CON BASE EN DATOS DE LA ENOE

En cuanto a las demás variables analizadas, las variaciones no son significativas. La población nacional crece de manera constante, la edad promedio en México igual crece a tasas constantes, la tasa de población en tercera edad aumenta, mientras que la tasa de población joven se reduce. En cuanto a los años promedio de escolaridad, estos aumentan y la tasa de población con posgrado igual aumenta, en contra parte la tasa de población con primaria incompleta se reduce en el país.

Para la contextualización de las variables de manera visual se presentan mapas, en los cuales se muestra la información más reciente correspondiente al tercer trimestre del año 2019. En este año se presentan 36 ciudades representativas del país.

El siguiente mapa logra mostrar la población total de cada ciudad representativa:

MAPA 2: POBLACIÓN TOTAL DE CIUDADES DE MÉXICO EN 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El mapa 2 consigue mostrar la población de las ciudades por tamaño de punto, se aprecia que las ciudades con mayor población en México es la ciudad de México ubicada en el centro con 20,267,247 habitantes, seguida de la ciudad de monterrey ubicada al noroeste con 4,337,717 habitantes y en tercer lugar la ciudad de Guadalajara en el oeste del país con 4,315,425 habitantes.¹⁵

La desviación estándar de la población de las ciudades de México es alta, superior a los 3 millones de habitantes; la ciudad de México es casi 5 veces mayor a la ciudad en segunda posición. Siguiendo la teoría en que ciudades aglomeradas tienen incentivos percibir mayores ingresos en base a la escala, se puede intuir en primera instancia que la ciudad de México debería de ser una de las ciudades con mayor salario, sin embargo, más adelante la evidencia presentara lo contrario.

¹⁵ Para visualización del número de habitantes por ciudad ver anexo 1.

3.1. ANÁLISIS ESPACIAL DE LAS VARIABLES

Para la realización de los análisis espaciales se presenta en primera instancia el mapa de la distribución de los valores de las variables, posteriormente se muestran las pruebas de autocorrelación espacial y un diagrama de dispersión de Moran o *Moran Scatterplot* (Anselin, 1996).

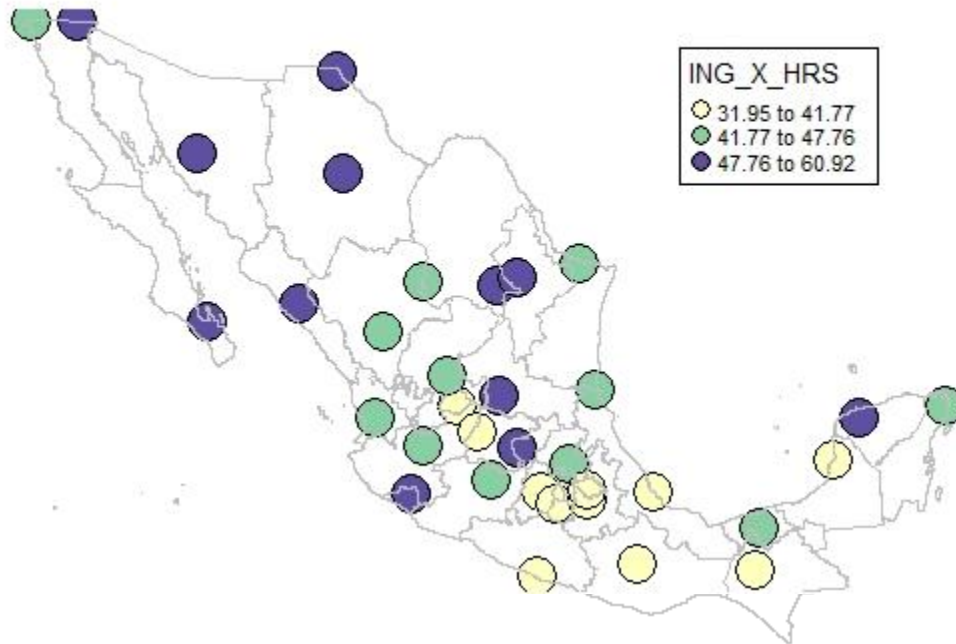
Los mapas de distribución de las variables muestran los resultados agrupados en tres categorías¹⁶, con fin de una mayor facilidad para identificar la distribución. El mapa muestra burbujas posicionadas en los centroides de las ciudades analizadas. El color de las burbujas se encuentra relacionado con los valores de la variable relacionada; el color azul corresponde a valores altos de la variable, el color verde corresponde a valores medios de la variable y el color amarillo corresponde a valores bajos de la variable, los cuales son relacionados a valores altos. La elaboración de los mapas fue mediante el software Rstudio con paqueterías de “tm_shape” y “tm_bubbles”.

3.1.1. ANÁLISIS VISUAL

Para el análisis espacial se presenta el mapa del ingreso medio por hora en las 36 ciudades autor representativas de la ENOE para el año 2019. En el mapa se muestra que las ciudades que perciben salarios bajos corresponden a salarios medios de 31.95 pesos por hora a 41.77 pesos por hora, estas ciudades se identificadas con burbujas color amarillo. Las ciudades que corresponden a salarios medios se encuentran en el intervalo de 41.777 pesos por hora a 47.76 pesos por hora, estas se identifican con burbujas color verde. Los salarios considerados altos corresponden a salarios superiores a 47.76 pesos por hora hasta 60.92 pesos por hora, este parámetro corresponde a burbujas color azul.

¹⁶ Los límites de las tres categorías de efectuaron mediante el comando “style = “quantile”” del software, el cual divide el número de observaciones en las categorías definidas (tres en este caso) de manera tal que cada categoría posea el mismo número de observaciones, en este caso cada categoría posee 12 ciudades.

MAPA 3: CLASIFICACIÓN DEL INGRESO MEDIO POR HORA EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



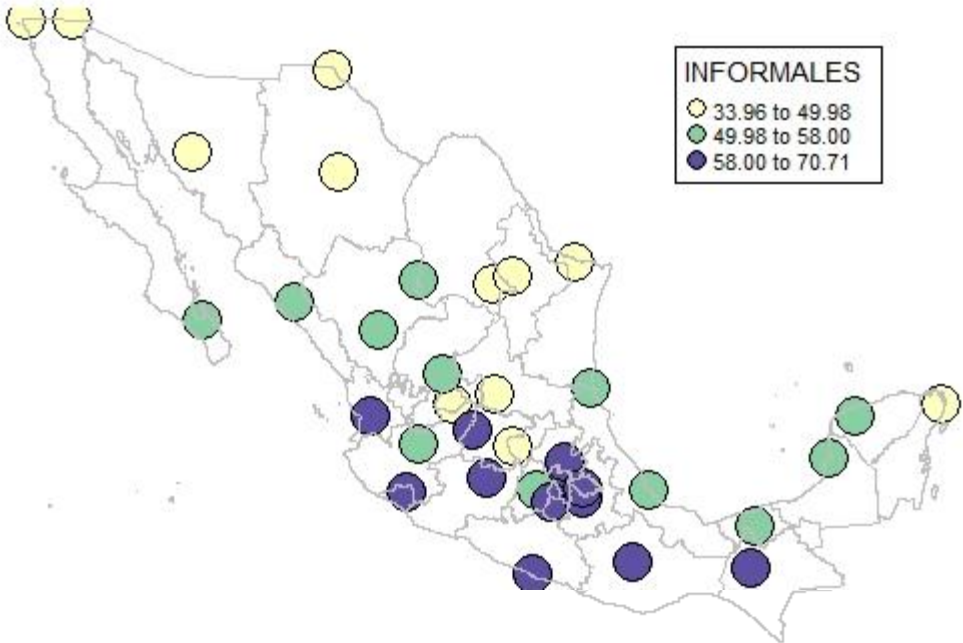
FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El mapa anterior muestra una aproximación a la agrupación de los salarios medios por hora de las ciudades, se percibe que los salarios bajos tienen a agruparse en la zona centro u sur del país, por su parte se observa que los salarios altos se posicionan en la zona norte del país, a excepción de los salarios percibidos en la ciudad de Mérida ubicada en el sur, la ciudad de Colima en el suroeste y Querétaro posicionado en el centro del país. Por su parte los salarios medios se encuentran distribuidos en gran parte del país, desde el norte con Tijuana y en el sur con Cancún. A grandes rasgos se visualiza agrupaciones salariales.

En el caso de la variable informalidad, se realiza el mismo tipo de mapa que en la variable de ingreso medio por hora. El siguiente mapa presenta la tasa de la población que se

encuentra trabajando informalmente¹⁷ según los datos obtenidos por la ENOE en el tercer trimestre del 2019. Las ciudades representadas con puntos amarillos corresponden a aquellas que tienen tasas de informalidad “baja” de 33.96% a 49.98%, las ciudades con puntos verdes tasas “medias” de 49.98% a 58% y las ciudades con tasas de informalidad “altas” se representan con el color azul y van de 58% a 70.71%.

MAPA 4: CLASIFICACIÓN DE LA TASA DE INFORMALIDAD EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

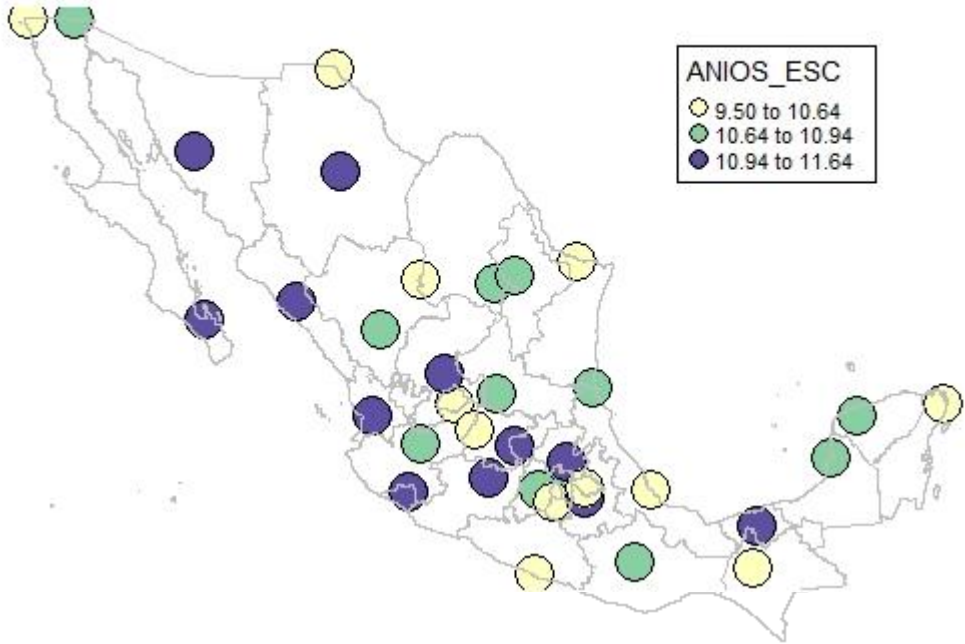
El mapa anterior permite observar que la informalidad presenta un patrón en las ciudades analizadas. En el norte se tienden a encontrar tasas de informalidad bajas, o en contra parte altas tasas de formalidad laboral. Por su parte se observa que los puntos azules

¹⁷ Revisar capítulo 2 para recordar los criterios que determinan si el trabajador se encuentra trabajando en informalidad.

correspondientes a altas tasas de trabajadores informales se encuentran en el sur del país, o de forma contraria en el sur se encuentran las ciudades con bajas tasas de trabajadores formales.

Para visualizar la distribución de la variable de años promedio de escolaridad se presenta el siguiente mapa. Los puntos color amarillo representan la ubicación de las ciudades que cuentan con un promedio de años de escolaridad de 9.5 años a 10.64 años de escolaridad, este intervalo se considerara “bajo”. Las ciudades consideradas con años de escolaridad “medios” se representan con color verde y se encuentran en el intervalo de 10.64 años a 10.94 años. Las ciudades con color azul representan años promedio de escolaridad “altos” que se encuentran en el intervalo de 10.94 años a 11.64 años promedio de escolaridad.

MAPA 5: CLASIFICACIÓN DE LOS AÑOS PROMEDIO DE ESCOLARIDAD EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

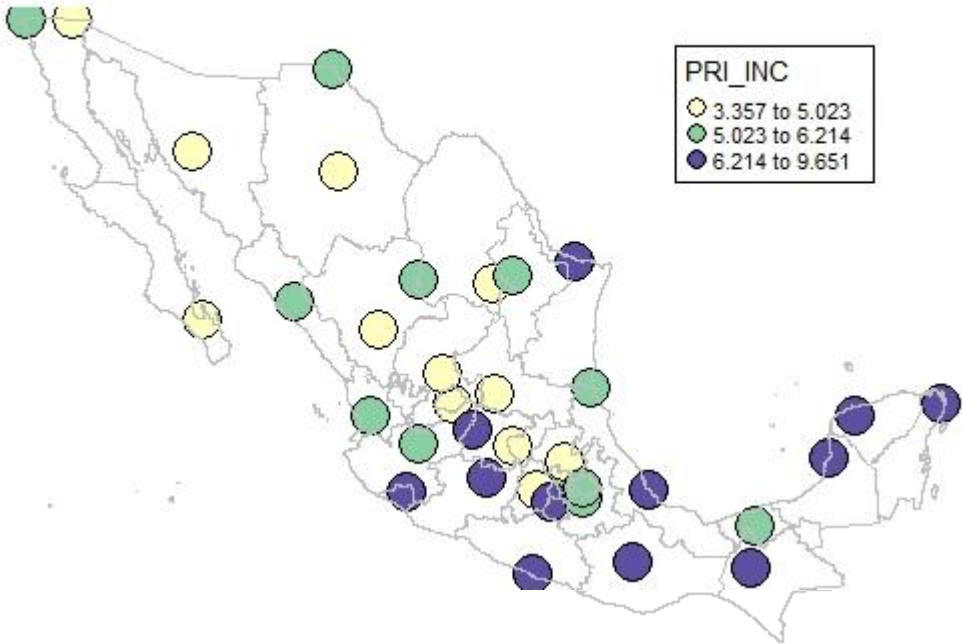


FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

En el mapa anterior no se logra apreciar un patrón en la distribución de los años promedio de escolaridad en las ciudades analizadas. La mayor agrupación de esta variable se presenta para el intervalo de años promedio de escolaridad altos y se visualiza en el oeste y centro del país. Por su parte los intervalos bajos y medios no muestran una agrupación visible.

Una variable relacionada, de manera indirecta en la ENOE, con los años promedio de escolaridad es la tasa de población con primaria incompleta. Esta tasa va desde 3.357% en Saltillo hasta el 9.651% en Acapulco. Las ciudades que se identifican con puntos amarillos corresponden a intervalos “bajos” y abarcan ciudades de 3.357% de población con primaria incompleta a ciudades con 5.023% de población con primaria incompleta. Las ciudades con intervalos “medios” van del 5.023% al 6.214% de población con primaria incompleta y se identifican con puntos color verde. Los puntos color azul corresponden a ciudades con “altas” tasas de primaria incompleta y van del 6.214% al 9.651%.

MAPA 6: CLASIFICACIÓN DE LA TASA DE POBLACIÓN CON PRIMARIA INCOMPLETA EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

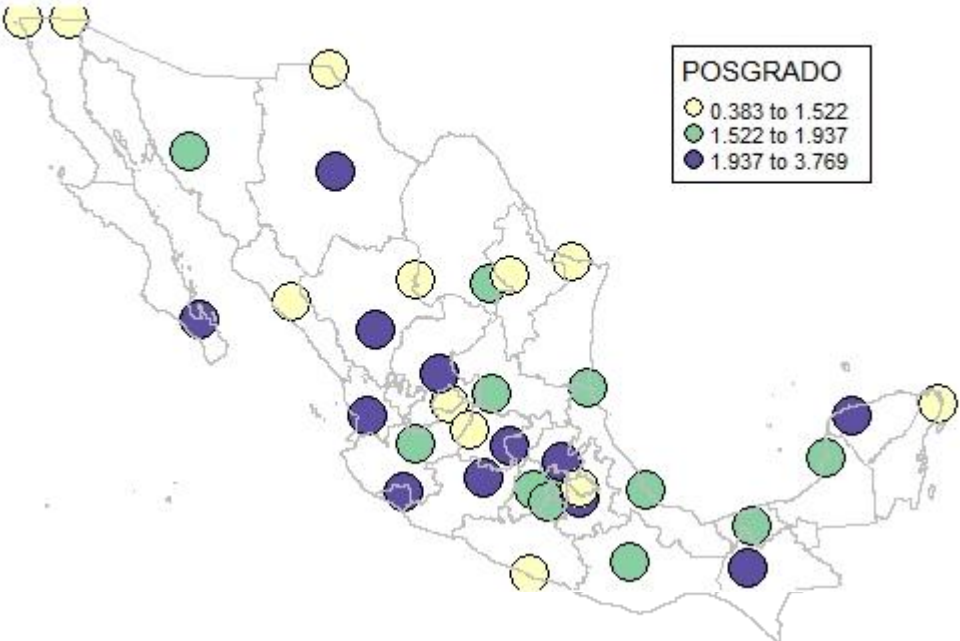


FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

En cuanto a la distribución de la tasa de población con primaria incompleta se observa que las ciudades con altas tasas de primaria incompleta tienden a concentrarse en el sur del país, a excepción de la ciudad de Tampico con el 6.12% de su población con primaria incompleta. Las ciudades con tasas bajas y medias parecen mostrar una distribución aleatoria en el centro y norte del país.

El siguiente mapa muestra la clasificación de la tasa de población con posgrado. La tasa observada en las ciudades analizadas va de 0.383% en la ciudad de Reynosa hasta el 3.769% de población con posgrado en la ciudad de Zacatecas. Los puntos amarillos corresponden a ciudades con tasas “bajas” correspondientes a el intervalo de 0.383% a 1.522% de población con posgrado. Las ciudades con intervalos “medios” van de 1.522% a 1.937% de población con posgrado y se muestran con puntos color verde. El intervalo “alto” se identifica en ciudades con puntos color azul y va de 1.937% a 3.769% de la población con posgrado.

MAPA 7: CLASIFICACIÓN DE LA TASA DE POBLACIÓN CON POSGRADO EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

Visualmente, el mapa anterior no parece mostrar algún patrón en la distribución de la tasa media de población con posgrado en las ciudades analizadas. Se muestra que las ciudades que se ubican en la zona fronteriza con Estados Unidos cuentan con tasas bajas de posgrado.

En conjunto, al analizar la tasa de población con primaria incompleta y la tasa de población con posgrado se observan comportamientos interesantes. Algunas ciudades presentan valores “altos” en ambas variables, y de igual manera algunos valores “bajos” en ambas variables. Las ciudades con valores bajos en ambas tasas son las ciudades de Mexicali y Aguascalientes, estas ciudades presentan bajas tasas de primaria incompleta pero también bajas tasas de población con posgrado, con esta información se puede dar una aproximación a concluir que en dichas ciudades tanto la estructura económica, como las condiciones que marca el estado, dan oportunidad a la población de realizar estudios básicos, sin embargo, no incentivan a estudios especializados.

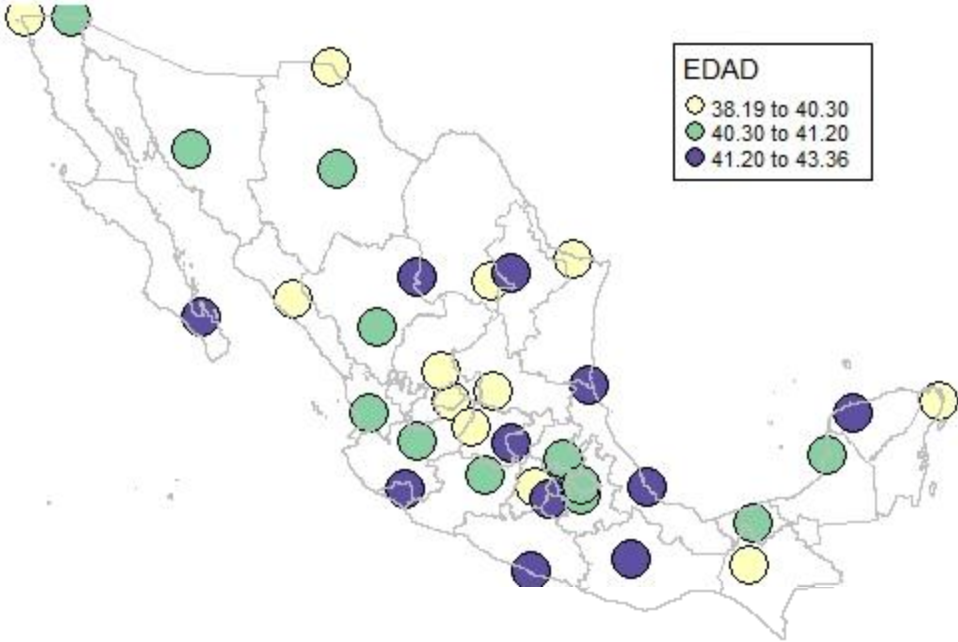
Las ciudades que muestran altos valores en ambas tasas son las de Mérida, Tuxtla Gutiérrez, Colima y Morelia; estas ciudades presentan altas tasas de población con primaria incompleta, pero de igual manera altas tasas de población con posgrado, con esto se podría concluir que en dichas ciudades tanto el estado como la estructura económica no incentivan a concluir estudios básicos, y el alto grado de posgrados puede atribuirse a un incentivo de universidades para atraer población de otras ciudades.

En el caso de la variable Edad, proxy a la experiencia, se presenta el siguiente mapa con la distribución de la edad promedio de la población. Esta variable se desglosa en población joven y población en tercera edad, sin embargo, para fin de este apartado no se presentan los mapas en la sección¹⁸ ya que la distribución observada visualmente parece ser aleatoria. La edad promedio en las ciudades analizadas va de los 38.19 años hasta los 43.36 años, lo que indica que las poblaciones en las ciudades varían en intervalos medios de casi 5 años.

¹⁸ Para revisar la distribución de las variables de jóvenes y tercera edad, revisar los mapas en anexos.

Las ciudades consideradas con intervalos “bajos” de la edad promedio se presentan con color amarillo y el intervalo se encuentra entre los 38.19 años a 40.3 años, estas ciudades se determinarán como ciudades jóvenes. El intervalo considerado “medio” de la edad promedio va de los 40.3 años a los 41.2 años y se presentan en ciudades con puntos verdes. Los puntos azules corresponden a ciudades con intervalos “altos” en la edad promedio, va de 41.2 años a los 43.36 años y son consideradas ciudades adultas.

MAPA 8: CLASIFICACIÓN DE LA EDAD PROMEDIO EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

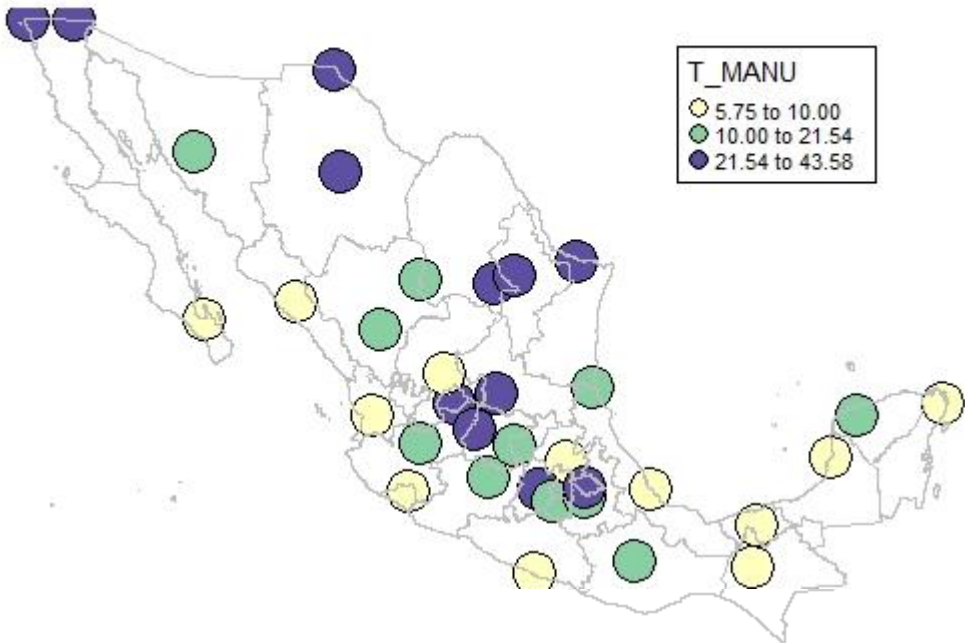


FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

La distribución de la edad promedio en las ciudades del país no parece presentar un patrón geográfico, visualmente. Se observa que en el sur del país no se encuentran casi ciudades jóvenes, solo la ciudad de Tuxtla Gutiérrez y Cancún. En el norte del país las únicas dos ciudades adultas que se observan son las de Torreón y Monterrey.

La distribución de los valores de variable que presenta la tasa de trabajadores en la industria manufacturera se presenta en el siguiente mapa. Los valores altos de la variable se atribuyen a puntos azules y van del 21.54% hasta el 43.58%, los valores medios se muestran con puntos verdes y van del 10% al 21.54% y por último los valores bajos de la variable se observan con puntos amarillos que van del 5.75% al 10%.

MAPA 9: CLASIFICACIÓN DE TASA DE TRABAJADORES EN INDUSTRIA MANUFACTURERA EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

La ciudad con mayor tasa de trabajadores en manufactura es ciudad Juárez con el 43.58% de trabajadores en la manufactura, seguida se posiciona la ciudad de Reynosa con el 39.83%. las ciudades con menor tasa son la ciudad de Cancún con tan solo 5.75% de trabajadores en manufactura, seguido se encuentra la ciudad de Acapulco con el 6.48% y le

sigue la ciudad de La Paz con el 7%. Estas ciudades presentan las menores tasas debido a que las ciudades se tienden a especializar en actividades turísticas.

En cuanto a los patrones visuales del mapa se puede observar que las ciudades con altas tasas se posicionan en el norte del país, debido a la estructura económica de ellas y la cercanía geográfica con Estados Unidos que presenta una alta demanda de productos manufacturados mexicanos. Por su parte, el mapa permite observar que las ciudades con tasas bajas se encuentran en el sur del país, característica atribuida a que dicha región es abundante en recursos naturales y turísticos. Se concluye que el mapa sí permite encontrar un patrón visual de las tasas de trabajadores en industria manufacturera para las ciudades analizadas del país.

3.1.2 PRUEBAS DE AUTOCORRELACIÓN ESPACIAL

Para corroborar la significancia de agrupaciones espaciales en los salarios se efectúan las pruebas de autocorrelación espacial, la prueba de hipótesis del estadístico I-Moran arrojó los siguientes resultados para los ingresos medios por hora de las 36 ciudades analizadas en el 2019.

Moran I test under randomisation

data: ciudades\$ING_X_HRS
weights: w

Moran I statistic standard deviate = 6.4087, **p-value = 7.339e-11**
alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.599772992	-0.028571429	0.009612981

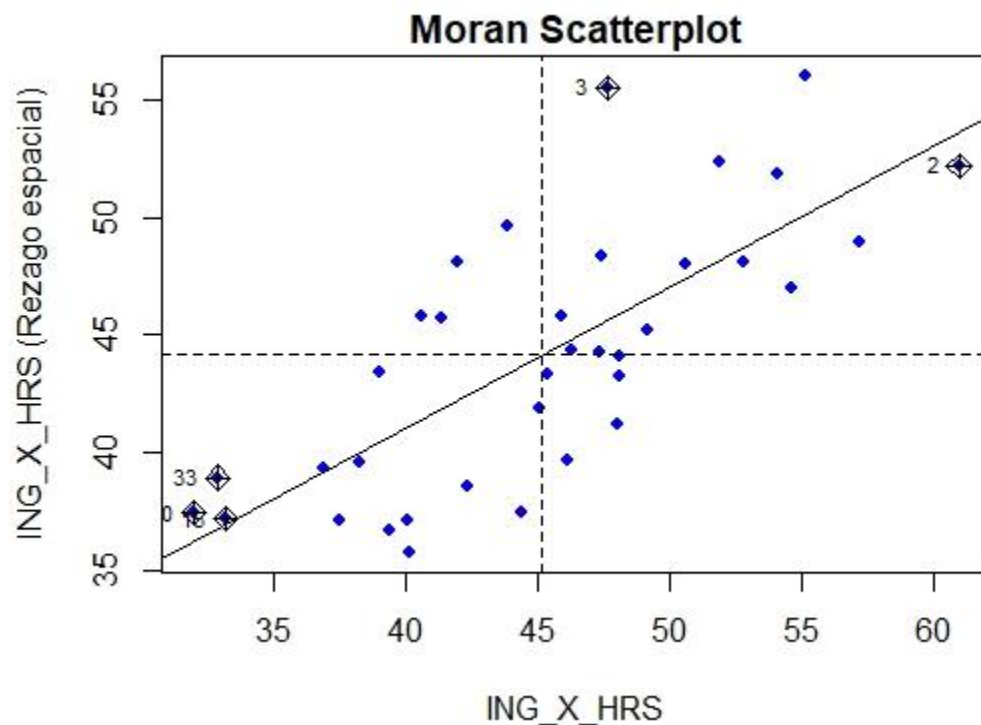
En la prueba se resalta la significancia del “p-value”, y el valor del estadístico de autocorrelación espacial “Moran I statistic”. La significancia observada es de 7.339e-11, lo cual indica que la prueba resulta significativa para intervalos de confianza superiores al 99.99%, esto permite inferir que se rechaza la hipótesis nula (H0) de no autocorrelación

espacial. La prueba de significancia concluye que existe autocorrelación espacial entre los salarios medios de las 36 principales ciudades de México.

El valor observado del estadístico I-Moran es cercano a 0.6, lo que indica que existe una autocorrelación espacial positiva. La prueba demuestra que las ciudades que perciben salarios altos, presentan vecindad con ciudades que también los perciben, y de forma contraria, las ciudades con salarios bajos se rodean de salarios con las mismas cualidades salariales.

Para visualizar la autocorrelación espacial del ingreso promedio por hora se presenta el diagrama de dispersión de Moran "Moran Scatterplot" en el que dependiendo de la ubicación de las observaciones en los cuadrantes se observa el tipo de autocorrelación que existe entre las mismas.

FIGURA 2: DIAGRAMA DE MORAN PARA EL INGRESO PROMEDIO POR HORA EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

En la figura anterior se pudo observar que el mayor número de ciudades se posicionaron en la diagonal que atraviesa el diagrama, agrupándose en el cuadrante I (superior derecho) correspondiente a valores alto-alto y III (inferior izquierdo) correspondiente a valores bajo-bajo. La pendiente de la diagonal observada corresponde al valor del estadístico I-moran por lo que los resultados del diagrama son coherentes con el valor obtenido del estadístico, corroborando la existencia de autocorrelación espacial positiva.

Al igual que los ingresos, otra variable que con el análisis visual presento agrupaciones espaciales, fue la informalidad. Para corroborar la significancia y el grado de agrupación espacial de la informalidad en las ciudades se le realizó a la variable la prueba del estadístico I de Moran, el cual arrojó los siguientes resultados:

Moran I test under randomisation

data: ciudades\$INFORMALES
weights: w

Moran I statistic standard deviate = 6.1104, **p-value = 4.969e-10**

alternative hypothesis: greater

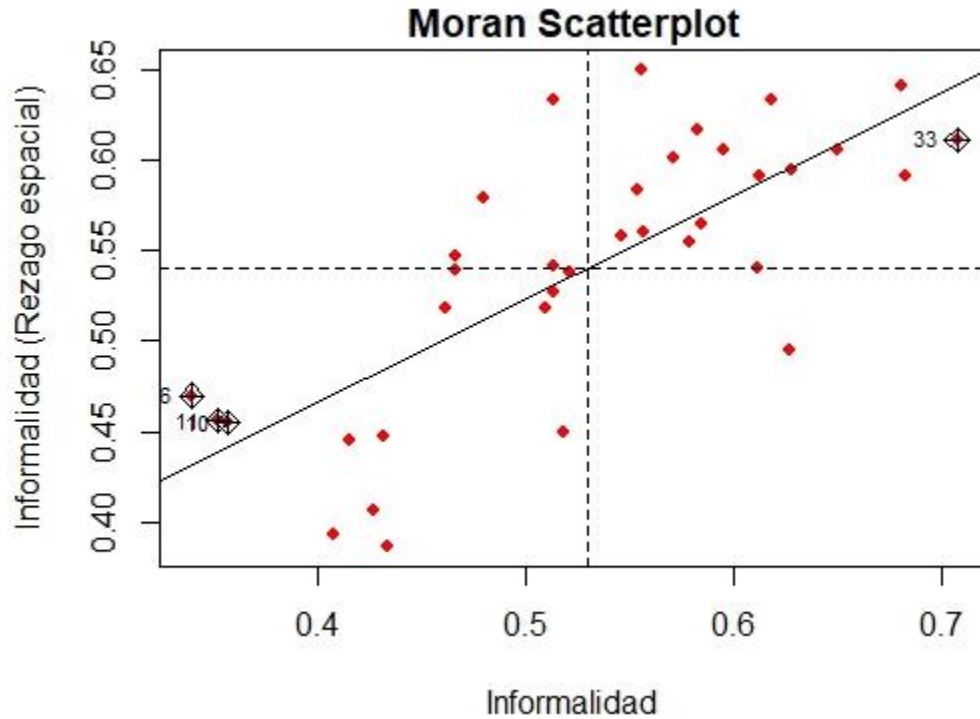
sample estimates:

Moran I statistic	Expectation	Variance
0.573236090	-0.028571429	0.009700103

Se resalta el valor del "P-Valuó", el cual permite rechazar la hipótesis nula (H0) de no autocorrelación espacial el valor es de 4.969e-10 por lo que la hipótesis se rechaza a intervalos de confianza superiores al 99.99%. El valor observado del estadístico I-Moran es de 0.5732 por lo que se concluye la existencia de autocorrelación espacial positiva para la informalidad en las ciudades analizadas, aquellas que presenten altos niveles de informalidad se rodean de ciudades de las mismas cualidades y a la inversa aquellas que presenten bajos niveles de informalidad tienen vecindad con ciudades con bajos niveles de informalidad.

Para visualizar lo mencionado se realiza el diagrama de Moran para la variable de informalidad en la siguiente figura:

FIGURA 3: DIAGRAMA DE MORAN PARA LA INFORMALIDAD EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

En el diagrama de moran se observa que la mayoría de las ciudades se posicionan en el cuadrante I correspondiente a valores altos-altos, lo cual indica que ciudades con altos niveles de informalidad tienden a ser vecinas de ciudades con la misma característica. Otra parte de ciudades se posicionan en el cuadrante III correspondiente a valores bajos-bajos.

Para el caso de la variable que presenta los años promedio de escolaridad en las ciudades se observó en el mapa que visualmente no parece identificarse un patrón en la distribución de las ciudades, para corroborar esto se realiza la prueba del estadístico I-Moran.

Moran I test under randomisation

data: ciudades\$ANIOS_ESC
weights: w

Moran I statistic standard deviate = -0.72133, **p-value = 0.7646**

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
-0.099012643	-0.028571429	0.009536527

En los resultados de la prueba se observa que el valor obtenido del “P-Valué” es de 0.7646, lo que indica que no se rechaza la hipótesis nula (H0) de no autocorrelación espacial. Se demuestra la coherencia con la conclusión visual obtenida del mapa correspondiente a los años de escolaridad; no existe autocorrelación espacial para los años promedios de escolaridad en las ciudades. El resultado demuestra que existe patrón aleatorio en la variable; las ciudades que presentan altos niveles de años promedios de escolaridad, no necesariamente se rodean de ciudades con el mismo patrón¹⁹.

Para el caso de la variable correspondiente a la tasa de población con primaria incompleta el mapa presento un patrón de altas tasas en el Sur del país y distribución aleatoria de tasas bajas y medias en el centro y norte. Para corroborar el patrón se realiza la prueba del estadístico I-Moran, la cual arrojó los siguientes resultados:

Moran I test under randomisation

data: ciudades\$PRI_INC
weights: w

Moran I statistic standard deviate = 3.262, **p-value = 0.0005532**

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.289202849	-0.028571429	0.009490111

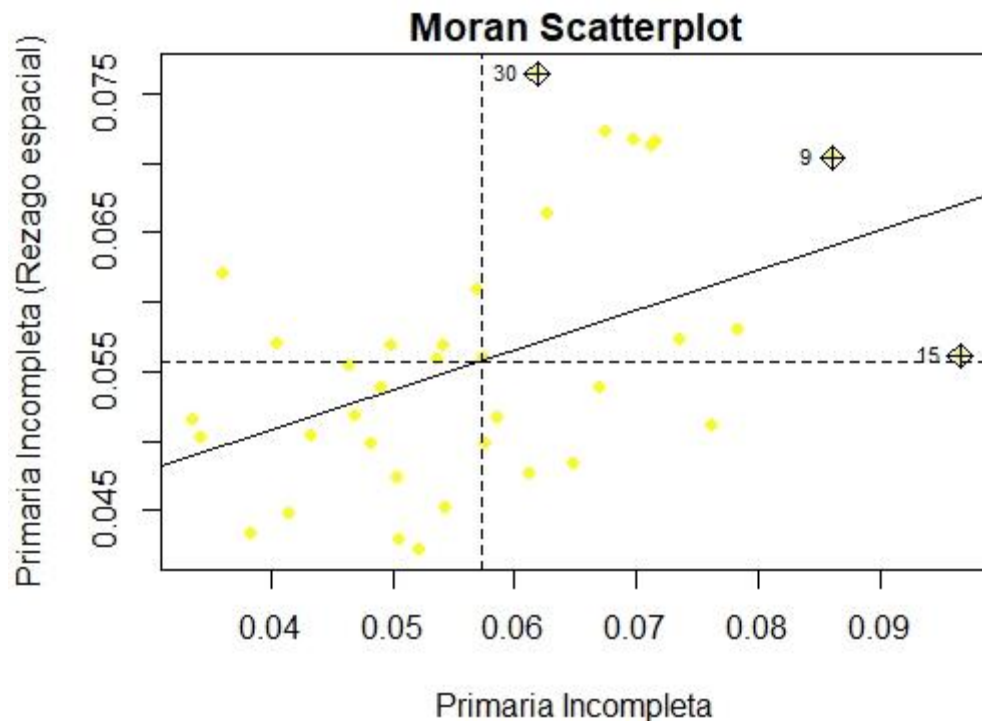
La prueba presenta un “P-Valué” del 0.0005532, lo que indica que a intervalo superior al 99.95% de confianza la H0 de no autocorrelación espacial se rechaza. El valor del estadístico I-Moran observado es de 0.2892. Con los resultados descritos, se permite concluir la

¹⁹ Para corroborar esto de forma visual se realizó en diagrama de dispersión de Moran el cual se puede consultar en anexos.

existencia de autocorrelación espacial positiva y débil en la tasa de primaria incompleta para las ciudades.

Aquellas ciudades que presenten altas tasas de primaria incompleta presentan una ligera tendencia a rodearse de ciudades con la misma característica y de manera inversa aquellas que presenten bajas tasas de primaria incompleta tienden a rodearse de ciudades con la misma característica. Como se mencionó se observó en el mapa una tendencia en concentración de las tasas altas de primaria incompleta, sin embargo, no se encontró patrón de concentración den las tasas medias y bajas, esto tiene coherencia con el valor “bajo” obtenido en el estadístico. Para contribuir a las conclusiones se analiza el diagrama de Moran:

FIGURA 4: DIAGRAMA DE MORAN PARA LA TASA DE PRIMARIA INCOMPLETA EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El diagrama permite observar una alta dispersión de las ciudades, además la mayor concentración de ellas se agrupa en el cuadrante III correspondiente a valores bajo-bajo y el cuadrante I correspondiente a valores altos-altos. La alta dispersión de las ciudades en el diagrama presenta coherencia con el bajo valor obtenido por el estadístico de Moran de 0.2892, el cual equivale a la pendiente de la línea en el diagrama. El diagrama permite concluir que la mayor concentración corresponde a tasas bajas de la tasa de primaria incompleta, y no a las tasas altas como se concluía en el mapa con un análisis visual.

Este resultado demuestra la importancia de un análisis completo y estadístico para la identificación de patrones espaciales, corroborando que el análisis visual mediante mapas no siempre permite a investigador arrojar conclusiones certeras.

En el caso de la variable de la tasa de trabajadores en la industria manufacturera se observó patrón de altas tasas en el norte y de bajas tasas en el sur, para corroborar la existencia del patrón el resultado del-Moran es el siguiente:

Moran I test under randomisation

data: ciudades\$T_MANU
weights: w

Moran I statistic standard deviate = 3.0214, **p-value = 0.001258**

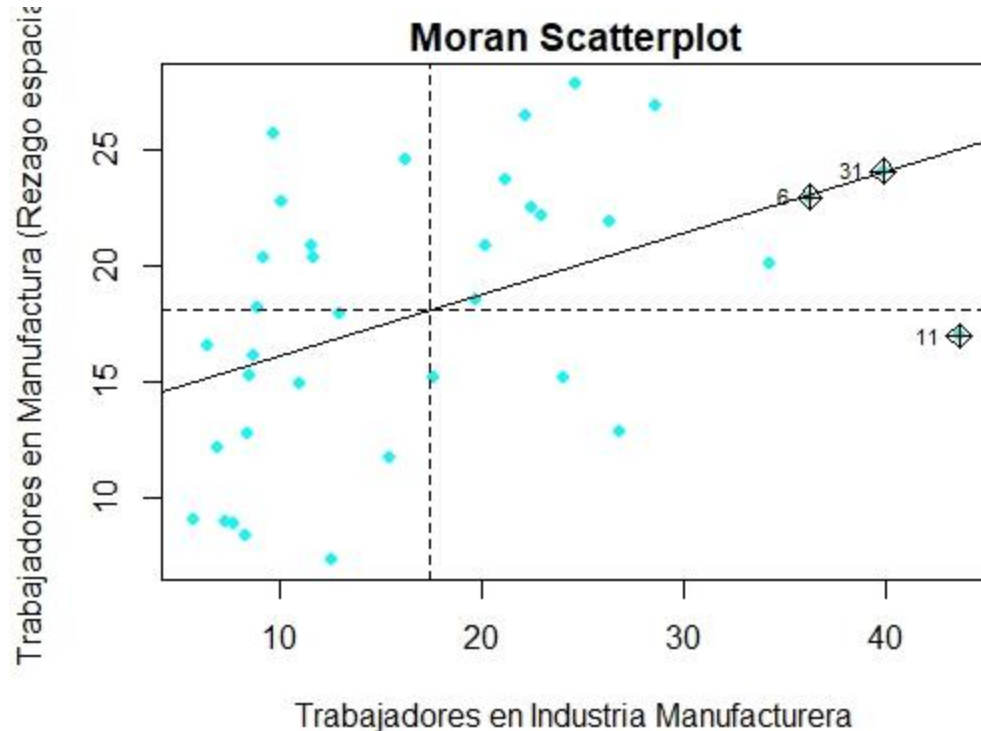
alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.266001138	-0.028571429	0.009505188

Al observar los resultados de la prueba, resulta ser significativa al presentar un “P-Value” del 0.001258, esto permite rechazar la hipótesis nula (H0) con intervalos de confianza superiores al 99%. El valor del estadístico I-Moran es de 0.266001138, por lo que se está en el caso de autocorrelación espacial positiva, a pesar de que el valor es bajo, si se permite concluir que aquellas ciudades que presenten altas tasas de trabajadores en industria manufacturera tiendan a rodearse de ciudades con la misma característica, y ciudades con bajas tasas tienden a rodearse de ciudades con tasas bajas. Con fin de contribuir a esta conclusión se presenta el diagrama de Moran para la variable analizada:

FIGURA 5: DIAGRAMA DE MORAN PARA LA TASA DE TRABAJADORES EN LA INDUSTRIA MANUFACTURERA EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El diagrama anterior muestra dispersión en los puntos lo cual muestra coherencia con el valor bajo obtenido en el I-Moran, a pesar de ellos se visualiza una tendencia a concentrarse en los cuadrantes I y III, atribuyendo a la presencia de autocorrelación espacial positiva.

Para el caso de las variables restantes (población con posgrado, edad, jóvenes, tercera edad, población desocupada), los mapas no presentaron algún patrón espacial en sus valores para las ciudades. Al momento de realizar la prueba con estadístico I-Moran²⁰ el valor observado del "P-Value" no permite rechazar la H0, por lo que se concluye que no existe autocorrelación espacial para el resto de las variables. En el diagrama de dispersión de

²⁰ Para consultar los resultados de la prueba I-Moran se pueden revisar en anexos.

Moran se logra observar coherencia con la conclusión de los mapas y de la prueba I-Mora, ya que las ciudades presentaron dispersión en conjunto.

Para continuar con un análisis espacial de autocorrelación completo se presentan los resultados del estadístico C-Gaery en el siguiente cuadro:

CUADRO 1: RESULTADOS DEL ESTADISTICO C-GEARY, POR VARIABLE, PARA EL TERCER TRIMESTRE DEL 2019

Variable	C -Geary	P valor	H0
<i>ING_X_HRS</i>	0.3799873	0.0000	<i>Se rechaza</i>
<i>INFORMALES</i>	0.4104087	0.0000	<i>Se rechaza</i>
<i>PRI_INC</i>	0.6115118	0.0002	<i>Se rechaza</i>
<i>T_MANU</i>	0.6733595	0.0015	<i>Se rechaza</i>
<i>DESOCUPADO</i>	1.0196966	0.5672	<i>No se rechaza</i>
<i>ANIOS_ESC</i>	1.1255351	0.8743	<i>No se rechaza</i>
<i>POSGRADO</i>	1.1330234	0.8720	<i>No se rechaza</i>
<i>POB</i>	1.2433968	0.8702	<i>No se rechaza</i>
<i>EDAD</i>	0.8716315	0.1157	<i>No se rechaza</i>
<i>JOVENES</i>	0.9227625	0.2410	<i>No se rechaza</i>
<i>TERC_EDA</i>	0.8088498	0.0421	<i>Se rechaza</i>

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

Recordando que la hipótesis nula (H0) es la no existencia de autocorrelación espacial, los resultados observados permiten concluir la presencia de autocorrelación espacial para las siguientes variables: ingreso medio por hora, tasa de trabajadores en sector informal, tasa de población con primaria incompleta, tasa de trabajadores en industria manufacturera y tasa de población en tercera edad²¹.

²¹ La variable de tasa de población en tercera edad rechaza la hipótesis a un intervalo del 95% de confianza, mientras que el resto de variables con H0 rechazada superan el 99% de confianza

El resultado del estadístico I-Moran difiere con la prueba C-Gaery en el resultado de la tasa de población en tercera edad. Este resultado podría explicarse a la diferencia en la interpretación, ya que en esta prueba se interpreta el resultado, menor a uno (0.8), de manera tal que las ciudades se rodean, estrictamente, de valores altos en tasas de tercera edad. Por su parte en el I-Moran las ciudades tendrían que rodearse de ciudades con la misma característica, ya sean valores altos o bajos de sus tasas de población en tercera edad.

Al observar el valor dado al C-Gaery, en los casos que se rechaza H_0 , se observa que todas las variables presentan autocorrelación espacial positiva ($C\text{-Gaery} < 1$). El resultado se interpreta, en dichas variables, que las ciudades tienden a rodearse de ciudades con valores altos otorgados a cada variable. Las ciudades que no rechazaron H_0 , se observa que el resultado es cercano a uno, lo cual se interpreta como no autocorrelación espacial.

En el caso de la variable de ingreso medio por hora, se observó el resultado más lejano a uno de la prueba, lo que indica que es la variable que cuenta con mayor certeza de autocorrelación espacial. De igual manera es la variable con la que se puede rechazar H_0 con mayor confianza, ya que posee el menor "P-Valu "; el intervalo de confianza es superior al 99.9999%. Otra variable que presenta alta confianza en autocorrelaci n espacial es la tasa de trabajadores informales.

3.2. ESTIMACI N DE MODELOS

Siguiendo a Anselin (2005) el an lisis econom trico se realiza con la estimaci n de un modelo lineal de m nimos cuadrados ordinarios, posterior de este se relajan los supuestos de independencia entre las observaciones para con esto estimar los modelos de regresi n espacial. Los resultados analizados se estiman para el tercer trimestre del 2019, y con fin de contrastar se realiza una segunda estimaci n de resultados para el a o 2009.

3.2.1. MODELO DE MÍNIMOS CUADRADOS ORDINARIOS

En cuadro 2 se presentan los resultados de los coeficientes obtenidos al llevar a cabo dos regresiones que buscan dar explicación al Ingreso promedio por hora a través de dos conjuntos distintos de variables explicativas. En el modelo 1 se aprecia que los coeficientes son tan pequeños que un cambio de una unidad en cualquiera de ellos no tendría poder explicativo sobre la variable dependiente. Para el caso del modelo 2 a pesar de ser menos variables los coeficientes son significativos, a excepción de uno, y con valores mayores.

CUADRO 2: COMPARATIVO DE MODELOS DE MINIMOS CUADRADOS ORDINARIOS PARA EL TERCER TRIMESTRE DEL 2019

Modelo 1		Modelo 2	
Intercepto	4.411e+01	Intercepto	11.7125
Población	-2.229e-09	Informales	-0.4421***
Edad	-5.970e-01	Años de	5.9005**
		Escolaridad	
Jóvenes	-5.970e-01	Desocupado	-1.2929*
Años de	7.014e+00	Tercera Edad	-0.1497
Escolaridad			
Posgrado	-4.093e-01		
PRI_INC	2.409e-01		
Desocupado	-1.270e+00		
Informales	-4.540e-01***		
T_Manu	1.909e-02		

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El modelo más adecuado para la explicación del ingreso promedio por hora es el modelo 2. Para justificar la selección del modelo se lleva a cabo un análisis de correlación para evitar problemas de multicolinealidad y/o colinealidad presentado en cuadro 3.

La multicolinealidad incide sobre la eficiencia de los estimadores, generando varianzas y covarianzas grandes que perjudican la precisión de los coeficientes beta asociados a las variables explicativas, provocando signos y significancia estadística erróneos. Para seleccionar un modelo eficiente se usó el factor de inflación de varianza (VIF) a fin de identificar las variables que generan la multicolinealidad. La premisa de este método es que a medida que la colinealidad aumenta la varianza del estimador crece, con respecto a esto un $VIF > 10$ indica un alto grado de multicolinealidad, mientras que el VIF cuando está más cerca de 1 apunta a la ausencia de este problema.

CUADRO 3: FACTOR DE INFLACIÓN DE LA VARIANZA EN MODELOS DE MINIMOS CUADRADOS ORDINARIOS PARA EL TERCER TRIMESTRE DEL 2019

Modelo 1			Modelo 2		
Variable	VIF	Diagnóstico	Variable	VIF	Diagnóstico
Población	1.2172	0	Informales	1.4697	0
Edad	6.1916	1	Años de	1.1861	0
			Escolaridad		
Jóvenes	4.3659	1	Desocupado	1.1166	0
Años de	7.3736	1	Tercera	1.4779	0
Escolaridad			Edad		
Posgrado	3.0620	1			
PRI_INC	4.8609	1			
Desocupado	1.1609	0			
Informales	2.5606	0			
T_Manu	4.1245	1			
R cuadrado ajustado: 0.5061			R cuadrado ajustado: 0.5768		

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El indicador de diagnóstico señala la presencia de colinealidad en caso de ser 1 y ausencia de colinealidad en caso de ser 0. El modelo 1 posee varias variables con diagnóstico 1, por lo que se concluye presencia de colinealidad en la edad, tasa de jóvenes, años de escolaridad, tasa de posgrado, tasa de población con primaria incompleta, y tasa de población en actividades manufactureras. Además, se aprecia que en el modelo 1, aunque el VIF de ninguna variable explicativa llega a 10, se observan algunas como edad y años de escolaridad con valores elevados.

En el modelo 2 se observan valores del VIF cercanos a la unidad lo cual indica la ausencia de multicolinealidad. El modelo, en el diagnóstico presentó valor conjunto de 0, lo cual indica ausencia de colinealidad.

Por último, otro indicador que contribuye a la selección del modelo 2 es el R cuadrado ajustado, debido a que este valor es mayor en dicho modelo, por lo cual indica que estas variables explican mejor a la variable dependiente. Traducido en la investigación las variables que mejor explican el ingreso por hora son los años promedio de escolaridad, la tasa de desocupación, la tasa de informalidad y la tasa de población en tercera edad.

Además de estos modelos se realizaron estimaciones combinando diferentes conjuntos de variables obteniendo resultados similares al modelo 1, por lo cual se llegó a la conclusión de proceder con el modelo 2 para el análisis espacial.

3.2.2. MODELOS DE REGRESIÓN ESPACIAL

A continuación, para analizar que no exista presencia de dependencia espacial entre los errores se lleva a cabo la prueba de I Moran a los residuos de la regresión de MCO. Siguiendo los resultados de la prueba presentados en el cuadro 4, los valores del p-valor indica que no es significativo, por lo cual se acepta la hipótesis nula evidenciando la ausencia de dependencia espacial entre los residuos.

**CUADRO 4: I MORAN A LOS RESIDUOS DE LA REGRESIÓN DE MCO PARA EL TERCER TRIMESTRE DEL 2019
MODELO MCO**

I DE MORAN	Estadístico	P-valor
	0.090402707	0.1803

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

Debido al resultado anterior se descarta el uso del modelo del error espacial (SDM), continuando con la metodología empleada en Anselin (2005) para la especificación del modelo, se plantea llevar a cabo la prueba del Multiplicador de Lagrange. Esta prueba parte de la hipótesis nula de que el mejor modelo es el de MCO y alternativamente a esta ubica los modelos de rezago espacial (SAR) o el modelo del error espacial (SEM).

Los resultados presentados en el cuadro 5 muestran que el modelo SAR es el más significativo por lo que se ubica a este como el modelo más adecuado para modelar los efectos espaciales.

CUADRO 5: PRUEBA DEL MULTIPLICADOR DE LAGRANGE PARA EL TERCER TRIMESTRE DEL 2019

Modelo	Estadístico	P-Valor
LMerr	0.60071	0.4283
RLMMerr	3.9135	0.0479
LMlag	9.7284	0.001814
RLMlag	13.041	0.0003047
SAC	13.642	0.001091

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

Además de esto el modelo SAC resultó ser significativo al igual que el modelo SDM por lo cual se realiza la estimación de todos los modelos buscando un análisis a más profundidad.

Como se puede ver en el cuadro 6 todos los coeficientes de fuerza de autocorrelación RHO y lambda son distintos de 0 por lo cual si se sigue los planteado en Elhorst (2010) señalando que el modelo apropiado sería el SDM debido a que sostiene bien el proceso generador de datos que corresponde a un modelo de rezago espacial, esta es la única forma de producir coeficientes insesgados. En segundo lugar, permite la obtención de efectos de spillover globales. Sin embargo, el RHO del modelo SDM no es significativo lo que lo vuelve cero y no puede ser empleado.

Continuando con el análisis del resto de modelos el modelo SAR posee el loglikelihood más elevado con respecto al modelo SAC, y el criterio de Akaike más bajo. El coeficiente RHO en este modelo es significativo y posee las mismas variables explicativas con grado de significancia igual al modelo SAC.

CUADRO 6: COMPARATIVO DE MODELOS ESPACIALES PARA EL TERCER TRIMESTRE DEL 2019

	SAR	SEM	SAC	SLX	SDM
CONSTANTE	-26.667147	7.14610	-32.047972*	-36.956110	-36.602534
INFORMALES	-0.227289**	-0.18520*	-0.220827**	-0.156607	-0.152070
AÑOS DE ESCOLARIDAD	5.813168***	4.77665***	5.867516***	6.868511***	6.748505
DESOCUPADOS	-0.823960	-0.59585	-0.833273	-0.936236	-0.872696
TERCERA EDAD	0.085797	0.09846	0.124660	-0.005563	0.032782
LAG.INFORMALES				-0.405925*	-0.372991*
LAG.AÑOS DE ESCOLARIDAD				4.763164	4.034385
LAG.DESOCUPADOS				-1.215018	-0.924231
LAG.TERCERA EDAD				-0.442346	-0.337592
\RHO	0.54408***		0.63534***		0.091922
\LAMBDA		0.70531**	-0.66847*		
NUM. OBS.	36	36	36	36	36
PARÁMETROS	7	7	8	7	11
LOG LIKELIHOOD	-94.96604	-100.3571	-95.25173		-94.27628
AIC (MODELO LINEAL)	216.95	216.95	216.95		208.66
AIC (MODELO ESPACIAL)	205.93	214.71	206.5		201.55
LR TEST: STATISTIC	11.014	4.2317	14.442		0.11007
LR TEST: P-VALUE	0.00090442	0.039676	0.00073096		0.74006

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

En resumen, la elección del modelo:

- El modelo MCO fue descartado a raíz de la dependencia espacial identificada con el I de Moran.
- La prueba de I de Moran a la regresión MCO evidenció que no había dependencia espacial en los errores por lo que se descartó el modelo SEM.
- Debido a que el modelo SDM no posee un rho significativo, fue descartado.
- Del análisis de las regresiones para cada uno de los modelos espaciales, el loglikelihood de los modelos y el criterio de Akaike confirmaron lo visto con la prueba de Lagrange arrojando al modelo SAR como el más indicado.

Observando el coeficiente de RHO del modelo SAR, el cual es 0.54408, indica que, si un área está rodeada por otras regiones, entonces la influencia de cada región puede ser medida por 0.54408 veces la variable dependiente alrededor de esta.

En el modelo SAR, para la correcta interpretación de los coeficientes es necesaria la estimación de los efectos directos e indirectos, los resultados de estos se presentan en el siguiente cuadro.

CUADRO 7: ESTIMACIÓN DE LOS EFECTOS DEL MODELO SAR PARA EL TERCER TRIMESTRE DEL 2019

	<i>DIRECT</i>	<i>INDIRECT</i>	<i>TOTAL</i>
INFORMALID	-0.24320999	-0.25531255	-0.4985225
ANIOS_ESC	6.22037325	6.52991017	12.7502834
DESOCUPADO	-0.88167699	-0.92555082	-1.8072278
TERC_EDA	0.09180665	0.09637511	0.1881818

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

Los resultados del cuadro 7 muestran que la tasa de informalidad y la tasa de desocupación presentan una relación negativa con el ingreso por hora. En caso contrario, las variables de

años de escolaridad y tasa de población en tercera edad contribuyen de manera positiva al ingreso por hora.

La variable que mayor efecto tiene sobre el ingreso por hora es la de años promedio de escolaridad. El efecto directo de esta variable refleja que tras el aumento de un año en el promedio de años de escolaridad en la ciudad i , el ingreso promedio por hora aumentara en \$6.22 pesos en la ciudad i . El efecto indirecto muestra que, si los años promedio de escolaridad de las ciudades vecinas aumenta en un año, el ingreso promedio por hora de la ciudad i aumentara en \$6.53 pesos. El efecto total señala que, si se presenta un aumento de un año en el promedio de años de escolaridad para la ciudad i , y para sus ciudades vecinas, el ingreso promedio por hora de la ciudad i se incrementará en \$12.75 pesos.

La tasa de desocupación presenta una relación inversa con el salario en el modelo SAR. El efecto directo indica que, si existe un aumento de un punto porcentual en la tasa de desocupación en la ciudad i , el ingreso promedio por hora de la misma ciudad se reducirá en \$0.88 pesos. El efecto indirecto señala que, si se presenta un aumento de un punto porcentual en la tasa de desocupación en las ciudades vecinas de i , el ingreso promedio por hora de la ciudad i se reducirá en \$0.925 pesos. El efecto total señala que, si se presenta un aumento de un punto porcentual de la tasa de desocupación para la ciudad i , y para sus ciudades vecinas, el ingreso promedio por hora de la ciudad i se reducirá en \$1.807 pesos.

El efecto directo de la tasa de informalidad señala que, si esta aumenta en un punto porcentual en la misma ciudad, el ingreso por hora de la ciudad se reducirá en \$0.24 pesos. El efecto indirecto señala que, si aumenta un punto porcentual la tasa de informalidad en las ciudades vecinas, el ingreso por hora de la propia ciudad se verá reducido en \$0.255 pesos. El efecto total al ser la suma de ambos efectos indica que, si la tasa de informalidad aumenta en un punto porcentual en la misma ciudad y en las ciudades vecinas, el ingreso por hora de la propia ciudad se reducirá en \$0.4985 pesos.

Por último, la relación con la tasa de población en tercera edad se presenta positiva. El efecto directo de esta señala que, si existe un aumento en un punto porcentual de esta, en la ciudad i , el aumento en el ingreso por hora, de la misma ciudad i , será de \$0.0918 pesos.

El efecto indirecto indica que, si aumenta un punto porcentual la tasa de informalidad de las ciudades vecinas de la ciudad i , el ingreso por hora de la ciudad i se incrementara en \$0.096 pesos. El efecto total, al ser la suma de ambos efectos, muestra que tras un aumento de un punto porcentual de la tasa de informalidad en la ciudad i y en sus ciudades vecinas, el ingreso promedio por hora reflejara un incremento de \$0.188 pesos.

Se observa que en las cuatro variables el efecto indirecto es mayor que el directo, lo que indica que existe una mayor dependencia del ingreso por hora hacia las ciudades vecinas que la propia ciudad.

3.2.3. CONTRASTE DE RESULTADOS CON EL TERCER TRIMESTRE DEL 2009

Para poder comparar el comportamiento de las variables en el tiempo, se optó por realizar un comparativo con el año 2009. La selección de este año se atribuye a que es un año que presento oscilaciones destacables en el comportamiento de las variables, tal como se presentó al inicio de este capítulo.

En el cuadro 7 se presentan los resultados del modelo seleccionado para el año 2019, en base a datos del 2009.

CUADRO 8: MINIMOS CUADRADOS ORDINARIOS PARA EL TERCER TRIMESTRE DEL 2009

	<i>ESTIMATE</i>	<i>PR(> T)</i>
INTERCEPT	10.4482	0.5922
INFORMALID	-0.1962	0.0731
ANIOS_ESC	3.4574	0.0159 *
DESOCUPADO	0.3619	0.3992
TERC_EDA	-0.5711	0.3235

ADJUSTED R-SQUARED: 0.4137

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

Se aprecia que la selección de variables para este año no es adecuada, debido a que la significancia de estas solo se consigue para la variable que muestra los años promedio de escolaridad. Otro indicador que presenta que este modelo no es eficiente es el R cuadrado, el cual es de 0.393, lo que concluye que estas variables no explican el ingreso por hora de las ciudades.

Una posible causa que ocasiona que la selección de las variables explicativas del modelo no sea adecuada es que, a diferencia del 2019, este año presentó un mayor número de oscilaciones en los valores de las variables a consecuencia de los efectos causados por la crisis económica mundial del 2008.

Para el MCO del 2009, se observa que las variables en general no presentan problemas de colinealidad al analizar el diagnóstico ni problemas de multicolinealidad con el factor de inflación de la varianza.²²

A pesar de que el modelo de MCO del 2009 muestra diferencias en la significancia de las variables explicativas, con respecto al 2019, se continua el mismo análisis que en 2019, a fin de poder comparar si los efectos espaciales y la selección del modelo espacial muestran consistencia en datos para estos años.

Para continuar con el análisis espacial, lo primero es analizar que no exista presencia de dependencia espacial entre los errores por lo que se lleva a cabo la prueba de I Moran a los residuos de la regresión de MCO para 2009.

²² Véase anexo 12

**CUADRO 9: I MORAN A LOS RESIDUOS DE LA REGRESIÓN DE MCO PARA EL TERCER TRIMESTRE DEL 2009
MODELO MCO**

I DE MORAN	Estadístico	P-valor
	0.150198732	0.04922

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El resultado de la prueba revela que la hipótesis se rechaza a intervalos de confianza menores al 95.078%. Por otra parte, para intervalos de confianza superiores al 95% la hipótesis nula de ausencia de dependencia espacial entre los residuos se acepta. En contraste con el año 2019 se observan diferencias en la significancia de la prueba ya que para 2009 es necesario un menor margen de error para poder aceptar la ausencia de dependencia espacial entre los residuos.

Si aceptamos márgenes de error menores al 0.05, se descarta la aplicación de modelos SDM y se plantea llevar a cabo la prueba del Multiplicador de Lagrange, al igual que para el 2019. En el cuadro 9 se presentan los resultados de dicha prueba.

CUADRO 10: PRUEBA DEL MULTIPLICADOR DE LAGRANGE PARA EL TERCER TRIMESTRE DEL 2009

Modelo	Estadístico	P-Valor
LMerr	1.8119	0.1783
RLMMerr	1.1275	0.2883
LMlag	5.4803	0.01923
RLMlag	4.7959	0.02853
SAC	6.6078	0.03674

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

Para la identificación del modelo, en base a los resultados del cuadro 9, se determina el modelo SAR como el más adecuado para explicar los efectos espaciales ya que son los modelos que mayor significancia presentaron. Por otra parte, los modelos SEM se observa que no son significativos. Estos resultados presentan consistencia con los presentados en el 2019. Los resultados del modelo SAR para el 2009 se presentan en el cuadro 10.

CUADRO 11: MODELO SAR PARA EL TERCER TRIMESTRE DEL 2009

	<i>ESTIMATE</i>	<i>PR(> T)</i>
INTERCEPT	10.4482	0.5922
INFORMALID	-0.1962	0.0731
ANIOS_ESC	3.4574	0.0159 *
DESOCUPADO	0.3619	0.3992
TERC_EDA	-0.5711	0.3235
RHO: 0.46478	P-VALUE: 0.022324	

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

El coeficiente RHO es el coeficiente de fuerza de autocorrelación espacial, el cuadro 10 este resultado ser significativo por lo que se acepta la utilización de un modelo SAR, tal como en el 2019. Para una correcta interpretación de los coeficientes del modelo SAR, esta no debe ser directa, a causa de la endogeneidad consecuencia por la dependencia espacial; es por ello que se requiere estimar impactos directos e indirectos presentados en el cuadro 11

CUADRO 12: ESTIMACIÓN DE LOS EFECTOS DEL MODELO SAR PARA EL TERCER TRIMESTRE DEL 2009

	<i>DIRECT</i>	<i>INDIRECT</i>	<i>TOTAL</i>
INFORMALID	-0.0826	-0.0650	-0.147578
ANIOS_ESC	4.6498	3.6623	8.3120734
DESOCUPADO	0.4102	0.3231	0.7333722
TERC_EDA	-0.5269	-0.4150	-0.9419611

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DEL SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

De acuerdo con los resultados las variaciones en el ingreso una ciudad se difunde a las regiones vecinas, posteriormente las variaciones en las ciudades vecinas repercuten nuevamente en la misma ciudad, de modo tal que se generan impactos directos (variaciones en el ingreso son producto de influencias locales) e indirectos (variaciones en el ingreso son producto de influencias vecinas) (LeSage & Pace, 2009). Por lo que en el cuadro 11 se presentan los spillover espacial.

Se logra observar que la tasa de informalidad y la tasa de población en tercera edad repercuten de manera negativa al ingreso, y los años de escolaridad y la tasa de desocupación de manera positiva.

La variable que mayor influencia presenta al ingreso promedio por hora es la de años promedio de escolaridad. Si los años promedio de escolaridad de la ciudad aumentan un año, el ingreso promedio por hora incrementara en \$4.65 pesos en la misma ciudad, esto nos indica el efecto directo. El efecto indirecto, muestra que, si los años promedio de escolaridad de las ciudades vecinas incrementen en un año, el ingreso promedio por hora de la ciudad incrementa en \$3.66 pesos. El efecto total señala que si los años promedio de escolaridad, de la propia ciudad y de sus vecinas incrementan en un año, el ingreso de la ciudad incrementara en un total de \$8.31 pesos.

CONCLUSIONES

El objetivo de la investigación es demostrar la existencia de fenómenos espaciales, como la propagación del ingreso, en las ciudades de México. Para el cumplimiento del objetivo se planteó la formulación de estrategias de análisis espacial. Se determinó que el ingreso dependía de variables con justificaciones teóricas y se descartaron algunas de estas en base a metodologías estadísticas. Se seleccionó un modelo en el cual el ingreso promedio por hora de las ciudades depende de los años promedio de escolaridad, la tasa de desocupación, la tasa de informalidad y la tasa de población en tercera edad.

Se logró concluir que los ingresos que se perciben en las ciudades analizadas presentan oscilaciones entre ellas, y de manera visual se pudo comprobar que esto ocurre. Los resultados obtenidos tienen consistencia con las hipótesis de Aguilera Fernández (2012) en la que afirma que distintas estructuras económicas de las regiones, reflejan interacción distinta en el mercado de trabajo, dando como resultado diferencias tanto en la cantidad de trabajadores como en el salario de estos.

Tal como comenta Marshall (1979) el salario es capaz de verse determinado de manera endógena por parte de las fuerzas económicas y de manera exógena a través de factores socio-institucionales, la investigación permitió demostrar esto. La selección de las variables se realizó en base a pruebas de colinealidad y multicolinealidad, realizando distintas estimaciones hasta obtener el modelo más eficiente para la determinación del ingreso.

Al analizar la oferta y demanda de trabajo, se consiguió demostrar que estas no son homogéneas en las regiones, tal como lo describe Ehrenberg & Smith (2018). Siguiendo con lo expuesto en el trabajo de Ehrenberg & Smith (2018), una variable que contribuye a la oferta de trabajo son los años promedios de escolaridad y se presentó que existe ausencia de homogeneidad en las ciudades de México.

Se consiguió presentar el resultado de variables que buscaran acercarse al cumplimiento de los cuatro principales factores que contribuyen a la determinación de la oferta de trabajo,

expuestos por Fernández Díaz, et al., (1999). Para el primer factor (eficiencia de la fuerza de trabajo) se utilizó la variable años promedio de escolaridad. Para la segunda (población total) se utilizó la población de cada ciudad. la tercera es la distribución de ésta por edad y sexo, se utilizó la variable de edad desagregada a tasa de jóvenes y tasa de población en tercera edad. Y para la cuarta (marco social y legal) se trató de utilizar como proxy la tasa de informalidad y desocupación de las ciudades.

Los cuatro factores descritos anteriormente se demostraron, a través de análisis espaciales visuales con mapas, que existe ausencia de homogeneidad en el espacio. Incluso en las variables de ingreso por hora, tasa de informalidad, años de escolaridad, tasa de primaria incompleta, tasa de trabajadores en sector manufacturero, se detectó la presencia de autocorrelación espacial con el estadístico I-Moran.

Para el debate realizado en la década de los noventa que expone Castellanos (2010) sobre como analizar la relación entre el desempleo y los salarios se presentó la relación entre el ingreso y la tasa de desempleo. A pesar de que Castellanos (2010) y Arellano y Bond (1991) concluyeron que la relación entre estas variables se explica con la curva de salario, en esta investigación no se pudo concluir el cumplimiento de esta para 2009.

Otro planteamiento para explicar la relación entre el ingreso y la tasa de desempleo es el formulado con la curva de Phillips (1958). En esta curva se muestra una relación negativa entre el cambio porcentual en los salarios y la tasa de desempleo.

Para el periodo del 2019 se logra consistencia con los resultados obtenidos y lo expuesto en la curva de phillips, ya que en el modelo de MCO, el aumento de un punto porcentual en la tasa de desempleo, ocasionaría que el ingreso promedio por hora se redujera en \$1.29 pesos.

A pesar del cumplimiento en el 2019, para el año 2009 se obtuvieron resultados inversos; existe una relación positiva entre ingreso promedio por hora y la tasa de desempleo. Aunque cabe destacar que en el año 2009 la variable de tasa de desocupación no resulto

ser significativa, esta diferencia puede atribuirse a efectos coyunturales en las dinámicas económicas de las ciudades provocados por la crisis económica mundial del 2008.

Además, la discrepancia puede atribuirse a la principal crítica a la curva de Phillips, ya que presenta una visión macroeconómica y dejaba fuera la visión de datos microeconómicos y regionales, caso que no es aplicable en la investigación ya que se analizó el ingreso de manera regional en las ciudades.

Para la determinación del salario, se expuso que algunas de las principales variables a considerar son: escolaridad, sexo, experiencia, estado civil, edad, raza, sindicalizado, etc. (Llamas Huitrón, 2019). Además de estas características individuales también pueden influir características sectoriales, entre las que destacan la actividad en que labora, el tamaño del establecimiento, la ocupación, etc. (Castro Lugo, 2005). Al considerar estas características individuales y sectoriales se es capaz de estimar ecuaciones salariales donde el nivel de salario es explicado por las tasas de desempleo de cada región, que a su vez estas están definidas por las características individuales y sectoriales según Blanchflower y Oswald (1996).

En esta perspectiva de determinación de salarios ya se consideran los efectos individuales, sectoriales e incluso regionales, que les permiten a las ecuaciones salariales identificar una relación negativa entre el salario y la tasa de desocupación. Es necesario explicarlas mediante modelos de negociación, bajo los cuales, a mayor número de trabajadores empleados y calificados, mayor será el poder de negociación para aumentar de salarios.

La limitante de esta hipótesis que comenta Buettner (1999) es que la hipótesis de la curva salarial describe que los salarios se determinan en el mercado de trabajo local, pero no consideran las interacciones espaciales entre los mercados locales. La investigación logro mostrar la importancia de estas interacciones espaciales.

Al igual que en el caso de la curva de Phillips se demostró consistencia con lo estipulado por las curvas de salario para el 2019, sin embargo, para 2009 no fue así, como se describió anteriormente.

Las pruebas econométricas aplicadas en la investigación lograron mostrar que las diferencias en el ingreso de las ciudades de México son causadas por los factores que mencionan Wheaton (2002) y Gámez y Rosas (2015), originadas para el 2019 por:

- Los años promedio de escolaridad
- Tasa de informalidad
- Tasa de desocupación
- Tasa de población en tercera edad
- La variable espacial, ciudad donde se labora

La variable espacial es el principal factor que se analizó en la investigación. La dependencia del ingreso entre las ciudades puede atribuirse a movilidad de mano de obra, comercio y diferencias entre instituciones.

La hipótesis formulada por Qiong & Chand (2014) que argumenta que la aglomeración de la industria puede ser una fuerza para la divergencia de salarios entre las regiones no pudo ser demostrada debido a que la variable de población en la ciudad (utilizada como una proxy a la aglomeración) y la variable de tasa de trabajadores en sector manufacturero no resultaron ser significativas para el modelo, por lo cual se decidió que se descartarían.

Los efectos positivos a consecuencia de la aglomeración, como externalidades positivas, aumento de productividad, crecimiento económico y presencia de spillovers de conocimiento, expuestos por Gleaser (1992), Wheaton & Lewis (2002), Duranton & Puga (2004), Cagmani (2005) y Melo & Graham (2013), no lograron verse reflejadas en la determinación del ingreso de las ciudades, para esta investigación. Una sugerencia para la continuación del tema, es la inclusión de otra variable proxy para el grado de aglomeración en las ciudades, ya que la variable de población total no resulto ser significativa para este ejercicio.

A pesar de que las variables de población total en la ciudad, edad promedio, tasa de jóvenes, tasa de población con posgrado, tasa de población con primaria incompleta y la tasa de trabajadores en la industria manufacturera, fueron descartadas en el modelo para evitar

problemas colinealidad y multicolinealidad, el modelo más eficiente logro mostrar resultados espaciales que contribuyen a la comprobación de las teorías mencionadas.

El modelo seleccionado indica que el ingreso promedio por hora de las ciudades se encuentra en función de los años promedio de escolaridad, la tasa de desocupación, la tasa de informalidad y la tasa de población en tercera edad.

Para el tercer trimestre del 2019 el modelo MCO indica que la única variable que presenta relación positiva con el ingreso es la de años de escolaridad, los resultados indicaron que, tras un aumento de un año de escolaridad en las ciudades de México, el ingreso por hora aumentará en \$5.9 pesos. Para efectos del tercer trimestre del 2009, los resultados del modelo de MCO únicamente toma como variable significativa a los años de escolaridad y se indica que, tras un aumento de un año de escolaridad para las ciudades de México, el ingreso por hora aumentará e \$3.46 pesos.

Sin embargo, para ambos años se logró demostrar que la aplicación de modelos de MCO presentan coeficientes sesgados. El resultado del estadístico I de Moran mostros que para ambos años existe presencia de dependencia espacial entre los errores, por lo que para la correcta interpretación de los coeficientes se realizaron modelos de regresión espacial.

Con base en un análisis de multiplicador de Lagrange, parámetros de coeficiente de fuerza de autocorrelación espacial RHO, loglikelihood y criterio de Akaike, el modelo más eficiente para el proceso generador de datos es el modelo tipo SAR, para ambos años analizados.

La estimación de los efectos directos, indirectos y totales del modelo SAR para el año 2009 y 2019 presentaron consistencia en determinar que la variable que mayor impacto tiene sobre el ingreso de las ciudades son los años de escolaridad. Otro aspecto con consistencia es que tanto los años de escolaridad como la tasa de informalidad presentaron una relación positiva y negativa, respectivamente, para ambos años.

Se demostró inconsistencia en la relación del ingreso promedio por hora con la tasa de desocupación y la tasa de población en tercera edad, para los años analizados. En 2009 la relación con la tasa de desocupación era positiva, sin embargo, para el 2019 esta era

negativa. Y la relación del ingreso con la tasa de población en tercera edad para 2009 era negativa, pero para el 2019 fue positiva.

Otra diferencia destacable entre 2009 y 2019 es que en el año 2009 los efectos directos fueron superiores a los efectos indirectos, lo cual puede concluir que el ingreso promedio por hora depende más de características internas en la ciudad, en 2009. En contra parte, para el 2019, los efectos indirectos observados en las variables fueron superiores a los efectos directos, por lo que se concluye que el ingreso promedio por hora es más afectado por cambios en las variables de las ciudades vecinas que los cambios de la variable de la misma ciudad. Con el paso de los años la integración y dependencia espacial, puede decirse que ha aumentado.

Por último, si bien la investigación arrojo resultados destacables, también abre campo a continuar con el tema en otras investigaciones. Por ejemplo, fuese deseable ampliar el número de observaciones y desagregar el análisis a nivel municipal, sin embargo, la disponibilidad de datos es aún limitante en México.

De igual manera, otro aporte a la investigación pudiese ser la continuación mediante análisis de paneles espaciales, ya que en la presente se obtuvo la serie de tiempo publicada por la ENOE, correspondiente del 2005 al 2019. El panel espacial podría mostrar el comportamiento de las variables en el espacio año con año y contribuir con nuevos resultados y conclusiones.

REFERENCIAS

Aguilera Fernández, A., 2012. *Desigualdad salarial en la zona metropolitana de Zacatecas y Aguascalientes a partir de la apertura comercial: un estudio comparativo*. Saltillo: UNIVERSIDAD AUTONOMA DE COAHUILA.

Anselin , L., 1999. The Future of Spatial Analysis in the Social Sciences. *GeoeraDhic Information Sciences* , V(2), pp. 67-76.

Anselin, L., 1989. ¿QUÉ ES ESPECIAL SOBRE LOS DATOS ESPACIALES?PERSPECTIVAS ALTERNATIVAS EN EL ANÁLISIS DE DATOS E. *National Center for Geographic Information adn Analysis*.

Anselin, L., 1993. *Exploratory spatial data analysis and geographic information systems..* 9329 ed. s.l.:Research Paper.

Anselin, L., 1996. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. *Spatial Analytical Perspectives on GIS*, pp. 111-126.

Anselin, L., 1998. Exploratory spatial data analysis in a geocomputational environment.. *conference on GeoComputation*, pp. 17-19.

Anselin, L., 2006. How (Not) to Lie with Spatial Statistics. *American journal of preventive medicine*, XXX(25), pp. S3-S6.

Arellano, M. & Bond, S., 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, Issue 58, pp. 277-297.

Blanchflower, D. G. & Oswald, A. J., 1996. *The wage curve*. Segunda ed. s.l.:Massachusetts Institute of Technology.

Buettner, T., 1999. The effect of unemployment, aggregate wages, and spatial contiguity on local wages: An investigation with German district level data. *Centre for European Economic Research*, Volumen LXXVIII, pp. 47-67.

- Camagni, R., 2005. *Economía Urbana*. Primera ed. Barcelona: Antoni Bosch.
- Castellanos, S. G., 2009. Desempleo y determinación de salarios en la industria manufacturerade México. *economía mexicana nueva epoca* , XIX(1), pp. 171-198.
- Castro Lugo, D., 2005. *Salarios y desigualdad territorial en las áreas urbanas de México, 1992-2002*. Bellaterra: UNIVERSIDAD AUTÓNOMA DE BARCELONA.
- Castro Lugo, D., 2006. Disparidad salarial urbana en México, 1992-2002. *Estudios Sociales*, XV(29), pp. 119-153.
- Castro Lugo, D. & Huesca Reynoso, L., 2007. Wage inequality in Mexico: a review. *Papeles de población*, III(54), pp. 225-264.
- Celebioglu, F. & Dall'erba, S., 2009. Spatial disparities across the regions of Turkey: an exploratory spatial data analysis. *Springer-Verlag*, Volumen 45, p. 379–400.
- Chasco Yrigoyen, C., 2003. *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Tesis Doctoral ed. Madrid: Comunidad de Madrid.
- Cue Mancera, A. & Quintana Romero, L., 2009. *Introducción a la Microeconomía: Un enfoque integral para México*. Primera ed. México: Grupo Editorial Patri.
- Duranton, G. & Puga, D., 2004. MICRO-FOUNDATIONS OF URBAN AGGLOMERATION ECONOMIES. *Handbook of Regional and Urban Economics*, Volumen IV, pp. 2064-2117.
- Ehrenberg, R. G. & Smith, R. S., 2018. *"Modern Labor Economics" Theory and Public Policy*. Decimotercera ed. New York: Routledge Taylor & Francis.
- Elhorst, J. P., 2003. Specification and estimation of spatial panel data models. *International regional science review*, III(26), p. 244–268.
- Elhorst, J. P., 2010. Applied spatial econometrics: raising the bar. *Spatial economic analysis*, I(5), p. 9–28.

Fernández Díaz, A., Parejo Gámir, J. & Rodríguez, S. L., 1999. *POLÍTICA ECONÓMICA*. Madrid: McGRAW-HILL.

Fujita, M. & Krugman, P., 2004. The new economic geography: Past, present and the future. *Fifty Years of Regional Science*, pp. 139-164.

Fujita, M., Krugman, P. & Venables, A. J., 2001. *The Spatial Economy: Cities, Regions, and International Trade*. Primera ed. London: Massachusetts Institute of Technology.

Gámez Arroyo, J. & Rosas Rojas, E., 2015. DETERMINANTES DE LA DIFERENCIACIÓN SALARIAL EN MÉXICO. *multidisciplina*, Issue 20, pp. 53-75.

Gerónimo Antonio, V. M., 2014. *Análisis Exploratorio Espacial del Índice de Desarrollo Humano en México*. Saltillo: UNIVERSIDAD AUTÓNOMA DE COAHUILA.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A. & Shleifer, A., 1992. Growth in Cities. *Journal of Political Economy*, 100(6), pp. 1126-1152.

Griffith, D., 1987. *Spatial Autocorrelation: A Primer. Resource Publications in Geography*. Primera ed. Washington, DC: The Association of American Geographers..

Griffith, D. A., 2003. *Spatial autocorrelation and spatial filtering: gaining understanding through theory*. Primera ed. New York: Springer.

INEGI, 2010. *Instituto Nacional de Estadística y Geografía*. [En línea] Available at: https://www.inegi.org.mx/contenidos/programas/enoe/15ymas/doc/con_basedatos_proy2010.pdf

[Último acceso: 17 2020].

INEGI, 2019. *inegi.org.mx*. [En línea] Available at: https://www.inegi.org.mx/contenidos/programas/enoe/15ymas/doc/resultados_ciudades

[_enoe_2019_trim1.pdf](#)

[Último acceso: 8 Junio 2020].

INEGI, 2020. *ENOE Estructura de la base de datos*. [En línea]
Available at:
https://www.inegi.org.mx/contenidos/programas/enoe/15ymas/doc/fd_c_bas_amp_15ymas.pdf

[Último acceso: 9 Junio 2020].

Islas Camargo, A. & Cortez, W., 2004. Convergencia salarial entre las principales ciudades mexicanas: un análisis de cointegración. *EconoQuantum*, Volumen I, pp. 25-47.

Jiménez, F., 2001. *Macroeconomía enfoques y modelos. Tomo I*. 1 ed. Lima: Pontificia Universidad Católica de Perú.

Kuznets, S., 1955. Economic Growth and Income Inequality. *The American Economic Review*, XLV(1), pp. 1-28.

LeSage, J. P., 2008. An introduction to spatial econometrics. *Revue d'économie industrielle*, Issue 123, p. 19-44.

LeSage, J. & Pace, R. K., 2009. *Introduction to spatial econometrics*. Boca Raton: Crc press.

Levy, S., 2018. *Esfuerzos mal recompensados: la elusiva búsqueda de la prosperidad en México*. Primera ed. Washington, D.C.: Banco Interamericano de Desarrollo.

Llamas Huitrón, I., 2019. *Educación y dinámica socioeconómica en México*. Primera ed. Ciudad de México: Universidad Autónoma Metropolitana.

Lucifora, C., 1998. Wage Inequalities and Low Pay: The Role of Labour Market Institutions. *Istituto di Economia dell'Impresa e del Lavoro*, Volumen I, pp. 1-35.

Marshall, A., 1979. Notas sobre la determinación del salario. *Desarrollo Económico*, XIX(75), pp. 377-392.

Melo, P. C. & Graham, D. J., 2013. Testing for labour pooling as a source of agglomeration economies: Evidence for labour markets in England and Wales. *Papers in Regional Science*, pp. 1-22.

Moreno, R. & Vayá, E., 2000. *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*. Barcelona: Universidad de Barcelona.

Moreno, R. & Vaía, E., 2000. *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*. Primera ed. Barcelona: Universitat de Barcelona.

OIT, 2002. *Women and men in the informal economy: a statistical picture, employment sector*. Italia: International Labour Office.

Ord, J. K. & Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application.. *Geographical analysis*, IV(27), p. 286–306.

Pérez Mendez, M. A., 2013. *Análisis exploratorio espacial del índice de marginación en México, 2010*. Saltillo: UNIVERSIDAD AUTÓNOMA DE COAHUILA.

Qiong, H. & Chand, S., 2014. Spatial spillovers of regional wages: Evidence from Chinese provinces. *China Economic Review*, Issue 32, pp. 97-109.

Rodríguez-Licea, G. G.-S. J. A. & H. J., 2016. Identificación de conglomerados para impulsar las cadenas productivas de carne en México. *Agronomía mesoamericana*, II(27), pp. 353-365.

Ross, A. M., 1947. The Dynamics of Wage Determination under Collective Bargaining. *The American Economic Review*, XXXVII(5), pp. 793-822.

Sergio J., R. & Brett D., M., 2010. US Regional Income Convergence: A Spatial Econometric Perspective. *Regional Studies*, pp. 143-156.

Shekhar, S., Zhang, P., Huang, Y. & Raju Vatsavai, R., 2004. Trends in Spatial Data Mining. *Department of Computer Science and Engineering, University of Minnesota*, pp. 4-192.

Trejo Nieto, A., 2013. Las economías de las zonas metropolitanas de México en los albores del siglo XXI. *ESTUDIOS DEMOGRÁFICOS Y URBANOS*, XXVIII(3), pp. 545-591.

Wheaton, W. C. & Lewis, M. J., 2002. Urban Wages and Labor Market Agglomeration. *Journal of Urban Economics*, Issue 51, pp. 542-562.

ANEXOS

ANEXO 1: DATOS DE VARIABLES POR CIUDAD PARA EL 2019

CIUDAD	POB	EDAD	JOVENES	TERC_EDA	ANIOS_ESC	POSGRADO	PRI_INC	DESOCUPADOS	INFORMALES	INFORMALIDAD	INGRESO
01 México	20,267,247	42.4	30.78%	10.37%	10.67	1.80%	5.37%	5.06%	3.97%	58.29%	39.39
02 Guadalajara	4,315,425	40.5	33.99%	8.85%	10.65	1.91%	5.70%	3.63%	2.17%	55.43%	47.33
03 Monterrey	4,337,717	41.8	31.13%	9.64%	10.70	1.36%	5.04%	3.35%	0.66%	43.19%	54.59
04 Puebla	2,161,801	40.7	33.33%	8.70%	10.95	2.18%	5.41%	4.66%	1.95%	61.81%	40.10
05 León	1,521,922	39.2	33.64%	6.98%	9.50	1.02%	7.61%	5.03%	3.12%	62.71%	41.39
06 Torreón (a partir de 2017)	1,137,259	41.3	30.24%	8.96%	10.54	1.15%	5.06%	4.79%	0.64%	51.84%	47.38
07 San Luis Potosí	1,083,078	40.3	33.79%	8.41%	10.94	1.56%	4.82%	4.49%	1.94%	46.16%	48.12
08 Mérida	1,092,297	42.0	31.55%	10.66%	10.72	2.26%	7.13%	2.20%	1.91%	57.85%	48.02
09 Chihuahua	915,566	40.9	32.11%	8.74%	11.31	2.60%	3.43%	3.78%	0.83%	35.73%	54.10
10 Tampico	827,637	42.8	29.00%	10.55%	10.67	1.53%	6.12%	4.48%	0.92%	54.63%	45.33
12 Veracruz	700,350	43.4	30.03%	12.32%	10.38	1.64%	7.83%	4.38%	2.09%	55.60%	36.92
13 Acapulco	724,164	42.1	30.88%	10.79%	9.96	0.93%	9.65%	2.69%	5.29%	68.26%	33.17
14 Aguascalientes	887,222	39.8	33.69%	7.76%	10.60	1.11%	4.64%	3.98%	1.65%	46.68%	40.59
15 Morelia	679,754	40.5	35.39%	9.03%	11.04	2.29%	7.36%	3.27%	2.61%	58.46%	46.11
16 Toluca	1,394,086	40.1	33.42%	7.72%	10.86	1.55%	4.99%	5.26%	1.71%	51.39%	37.48
17 Saltillo	875,873	39.1	34.12%	6.65%	10.79	1.57%	3.36%	5.39%	0.46%	33.96%	50.61
18 Villahermosa	433,122	41.2	32.88%	8.77%	11.22	1.88%	6.18%	7.37%	3.17%	57.08%	42.35
19 Tuxtla Gutiérrez	672,720	39.8	33.51%	7.66%	10.64	2.10%	8.60%	5.98%	4.31%	61.24%	38.25
20 Ciudad Juárez	1,493,539	38.3	33.75%	5.69%	10.03	1.17%	5.44%	2.68%	1.08%	35.26%	51.88
21 Tijuana	1,616,075	38.8	35.43%	6.25%	10.18	0.68%	5.22%	2.45%	0.89%	43.39%	47.63
24 Culiacán	753,210	39.9	33.49%	7.59%	11.31	1.50%	5.76%	3.74%	2.45%	52.18%	52.78

25	Hermosillo	829,761	40.4	33.21%	7.83%	11.38	1.84%	3.85%	5.32%	1.94%	42.72%	55.16
26	Durango	581,970	40.3	33.14%	7.99%	10.89	1.99%	4.69%	4.40%	2.10%	51.41%	41.96
27	Tepic	446,733	41.0	31.86%	9.05%	11.19	2.38%	5.85%	3.86%	2.11%	61.18%	45.90
28	Campeche	267,869	41.1	31.83%	9.07%	10.70	1.68%	6.98%	3.25%	2.01%	55.68%	39.00
29	Cuernavaca	856,837	42.2	31.60%	10.88%	10.34	1.56%	6.71%	3.19%	3.09%	64.98%	31.95
31	Oaxaca	557,234	41.2	31.60%	9.23%	10.93	1.80%	7.17%	3.01%	3.08%	68.08%	40.14
32	Zacatecas	276,418	39.3	33.41%	6.32%	11.64	3.77%	4.06%	4.31%	1.32%	51.41%	46.30
33	Colima	302,913	42.3	30.55%	10.85%	11.00	2.29%	6.28%	4.24%	3.29%	62.75%	49.12
36	Querétaro	922,756	41.2	32.35%	8.93%	11.42	2.07%	3.60%	4.45%	1.94%	46.68%	48.07
39	Tlaxcala	766,763	40.3	33.09%	8.38%	10.23	0.84%	5.74%	3.95%	1.41%	70.71%	32.86
40	La Paz	290,077	41.5	31.59%	7.88%	11.26	2.29%	4.34%	4.26%	1.99%	50.96%	57.17
41	Cancún	851,472	38.2	36.27%	4.31%	10.26	0.97%	6.75%	4.22%	3.70%	48.02%	45.10
43	Pachuca	396,683	41.0	33.17%	9.07%	11.20	2.73%	4.91%	3.88%	2.87%	59.56%	44.39
44	Mexicali	823,386	40.8	32.47%	7.78%	10.94	1.46%	4.15%	3.14%	1.35%	40.76%	60.92
46	Reynosa	666,346	38.4	33.28%	5.53%	9.87	0.38%	6.48%	2.99%	0.29%	41.58%	43.83

ANEXO 2: DESCRIPCIÓN DE MNEMÓNICO DE LA ENOE

Num.	Mnemónico	Descripción del mnemónico	Códigos validados para captura	Descripción del código válido
1	r_def	Resultado definitivo de la entrevista	00 Tipo "A" 01 02 03 04 05 14 15 Tipo "B" 06 07 08 09 Tipo "C" 10 11 12 13	Entrevista completa Vivienda habitada Nadie en el momento de la entrevista Ausente temporal Se negó a dar información Informante inadecuado Otro motivo El hogar se mudó Entrevista suspendida Vivienda deshabitada Adecuada para habitarse De uso temporal Inadecuada para habitarse De uso temporal para fines diferentes de habitación Vivienda fuera de muestra Demolida Cambió de sitio (móvil) Uso permanente para fines diferentes a los de habitación Otro motivo
2	loc	Localidad	0001-9998	Número de la localidad
3	mun	Municipio	001-575	Número de municipio según entidad
7	t_loc	Tamaño de localidad	1 2 3 4	Localidades mayores de 100 000 habitantes Localidades de 15 000 a 99 999 habitantes Localidades de 2 500 a 14 999 habitantes Localidades menores de 2 500 habitantes
8	cd_a	Ciudad autorrepresentada	01 02 03 04 05 06 07 08 09 10 12 13 14 15 16 17 18 19 20 21 24 25	México Guadalajara Monterrey Puebla León Torreón (a partir de 2-17) San Luis Potosí Mérida Chihuahua Tampico Veracruz Acapulco Aguascalientes Morelia Toluca Saltillo Villahermosa Tuxtla Gutiérrez Ciudad Juárez (a partir de 1-19) Tijuana Culiacán Hermosillo

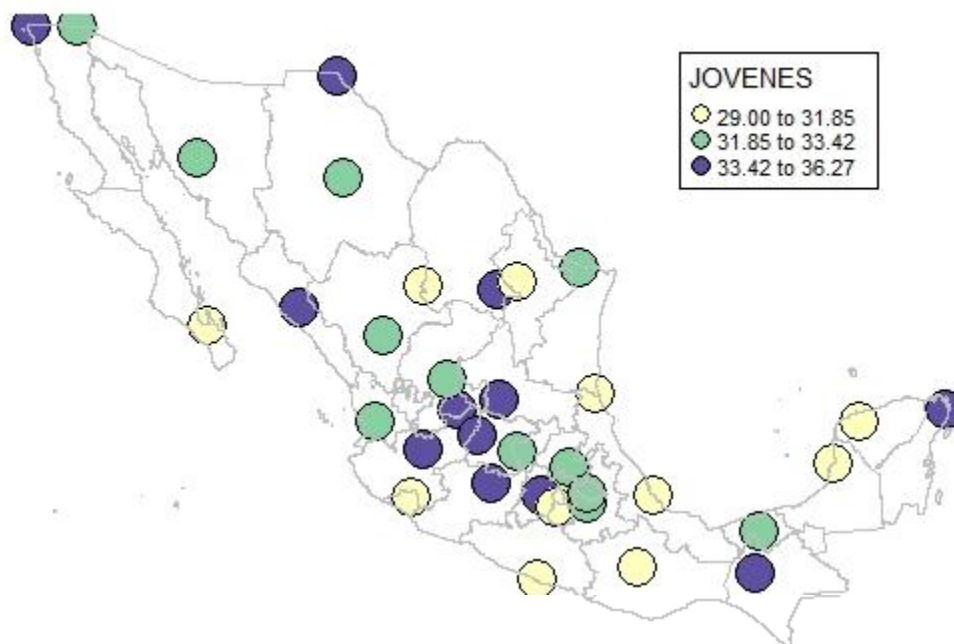
			26	Durango
			27	Tepic
			28	Campeche
			29	Cuernavaca
			30	Coatzacoalcos
			31	Oaxaca
			32	Zacatecas
			33	Colima
			36	Querétaro
			39	Tlaxcala
			40	La Paz
			42	La Ciudad del Carmen
			43	Pachuca
			44	Mexicali (a partir de 1-19)
			46	Reynosa (a partir de 1-19)
			52	Tapachula
			81	Complemento urbano-rural
			82	Complemento urbano-rural
			83	Complemento urbano-rural
			84	Complemento urbano-rural
			85	Complemento urbano-rural
			86	Complemento urbano-rural Paz
9	ent	Entidad	01-32	Número de entidad
	c_res	Condición de residencia	1	Residente habitual
			2	Ausente definitivo
			3	Nuevo residente
23	eda	Edad	00	Menores de 1 año
			01-96	Número de años cumplidos del integrante del hogar
			97	97 años y más
			98	Edad no especificada para mayores (12 años y más)
			99	Edad no especificada para menores (00 a 11 años)
29	cs_p13_1	Pregunta 13 ¿Hasta qué grado aprobó ...en la escuela?	00	Ninguno
			01	Preescolar
			02	Primaria
			03	Secundaria
			04	reparatoria o bachillerato
			05	Normal
			06	Carrera técnica
			07	Profesional
			08	Maestría
			09	Doctorado
			99	No sabe
45	salario	Salario mínimo mensual	Valor	Salario mínimo mensual
46	fac	Factor de expansión	Valor	
47	clase1	Clasificación de la población en PEA y PNEA	1	Población económicamente activa
			2	Población no económicamente activa

48	clase2	Clasificación de la población ocupada	1 2	Población ocupada Población desocupada
	pos_ocu	Clasificación de la población ocupada por posición en la ocupación	1 2 3 4 5	Trabajadores subordinados y remunerados Empleadores Trabajadores por cuenta propia Trabajadores sin pago No especificado
51	seg_soc	Clasificación de la población ocupada por condición de acceso a instituciones de salud	1 2 3	Con acceso Sin acceso No especificado
52	rama	Clasificación de la población ocupada por sector de actividad económica	6 1 2 3 4 5 7	Agropecuario Construcción Industria manufacturera Comercio Servicios Otros No especificado
54	ing7c	Clasificación de la población ocupada por nivel de ingreso	1 2 3 4 5 6 7	Hasta un salario mínimo Más de 1 hasta 2 salarios mínimo Más de 2 hasta 3 salarios mínimos Más de 3 hasta 5 salarios mínimos Más de 5 salarios mínimos No recibe ingresos No especificado
55	dur9c	Clasificación de la población ocupada por la duración de la jornada	1 2 3 4 5 6 7 8 9	Ausentes temporales con vínculo laboral Menos de 15 horas De 15 a 24 horas De 25 a 34 horas De 35 a 39 horas De 40 a 48 horas De 49 a 56 horas Más de 56 horas No especificado
73	remune2c	Población subordinada y remunerada	1 2	Subordinados y remunerados-asalariados Subordinados y remunerados-con percepciones no salariales
74	pre_asa	Población subordinada y remunerada por prestaciones laborales (sin considerar el acceso a las instituciones de salud)	1 2 3	Con prestaciones Sin prestaciones No especificado
75	tip_con	Clasificación de la población subordinada y remunerada por disponibilidad de contrato escrito	1 2 3 4 5 6	Con contrato escrito Temporal De base, planta o por tiempo indefinido Contrato de tipo no especificado Sin contrato escrito No especificado

80	niv_ins	Clasificación de la población de 5 años y más por nivel de instrucción	1 2 3 4 5	Primaria incompleta Primaria completa Secundaria completa Medio superior y superior No especificado
83	eda12c	Clasificación de la población de 15 años y más: Grupo de edad 12 claves	1 2 3 4 5 6 7 8 9 10 11 12	De 15 a 19 años De 20 a 24 años De 25 a 29 años De 30 a 34 años De 35 a 39 años De 40 a 44 años De 45 a 49 años De 50 a 54 años De 55 años 59 años De 60 a 64 años De 65 años y más No especificado
87	anios_esc	Años de escolaridad	1 a 24	Un año de escolaridad hasta 24 años
90	ing_x_hrs	Promedio de ingreso por hora trabajada	(11 enteros, un punto y 5 decimales)	El dato promedio registrado se obtiene de dividir el ingreso mensual entre el número de horas a la semana

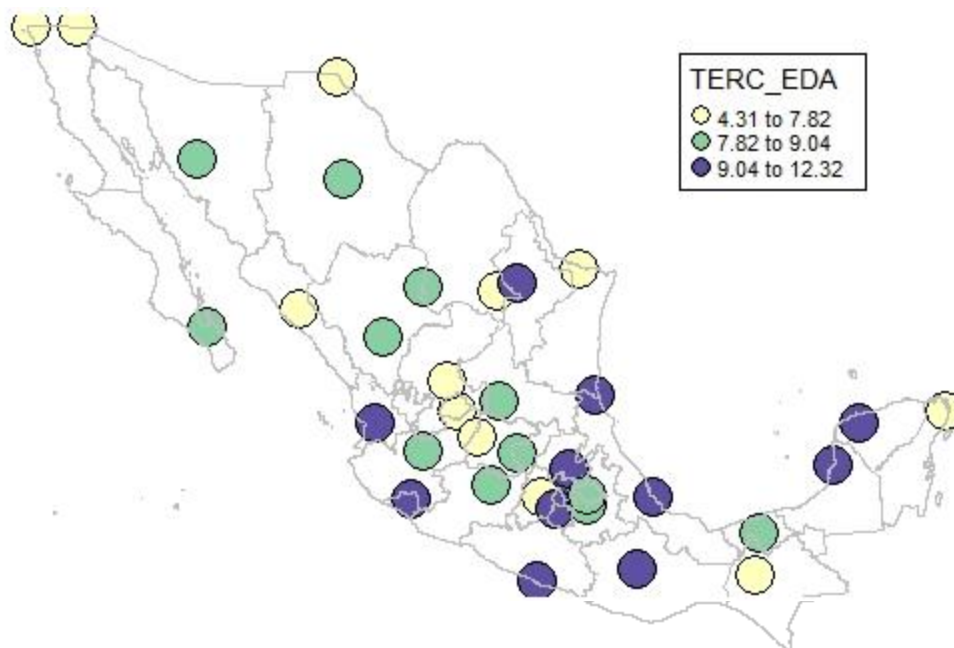
FUENTE: ELABORACIÓN PROPIA CON EXTRACCIÓN DEL DOCUMENTO “ENOE ESTRUCTURA DE LA BASE DE DATOS” (INEGI, 2020)

ANEXO 3: CLASIFICACIÓN DE LA POBLACIÓN JOVEN EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



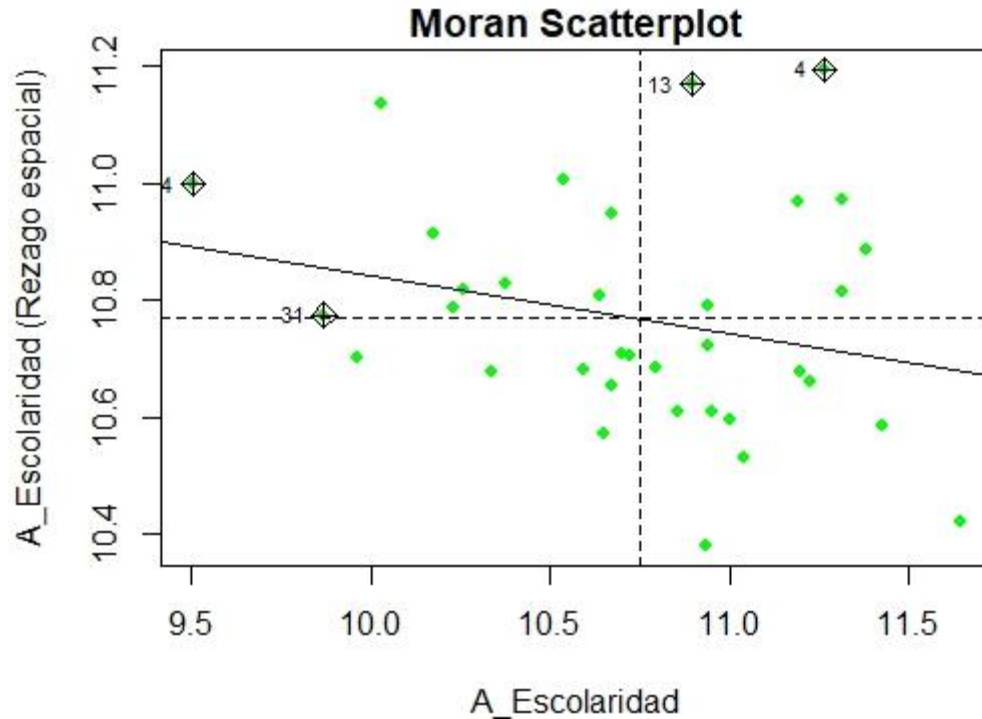
FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 4: CLASIFICACIÓN DE LA POBLACIÓN EN TERCERA EDAD EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 5: DIAGRAMA DE MORAN PARA AÑOS PROMEDIO DE ESCOLARIDAD EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019



FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 6: CUADRO DE RESULTADOS DEL ESTADISTICO I-MORAN PARA LA POBLACIÓN CON POSGRADO EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

Moran I test under randomisation

data: ciudades\$POSGRADO
weights: w

Moran I statistic standard deviate = -0.66589, p-value = 0.7473
alternative hypothesis: greater
sample estimates:

Moran I statistic	Expectation	Variance
-0.092233725	-0.028571429	0.009140156

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 7: CUADRO DE RESULTADOS DEL ESTADISTICO I-MORAN PARA LA EDAD PROMEDIO EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

Moran I test under randomisation

data: ciudades\$EDAD
weights: w

Moran I statistic standard deviate = 0.27998, p-value = 0.3897

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
-0.001074298	-0.028571429	0.009645141

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 8: CUADRO DE RESULTADOS DEL ESTADISTICO I-MORAN PARA LA TASA DE POBLACIÓN JOVEN EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

Moran I test under randomisation

data: ciudades\$JOVENES
weights: w

Moran I statistic standard deviate = -0.65899, p-value = 0.745

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
-0.092852229	-0.028571429	0.009514978

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 9: CUADRO DE RESULTADOS DEL ESTADISTICO I-MORAN PARA LA TASA DE POBLACIÓN EN TERCERA EDAD EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

Moran I test under randomisation

data: ciudades\$TERC_EDA
weights: w

Moran I statistic standard deviate = 0.80327, p-value = 0.2109

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.049617123	-0.028571429	0.009474539

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 10: CUADRO DE RESULTADOS DEL ESTADISTICO I-MORAN PARA LA POBLACIÓN TOTAL EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

Moran I test under randomisation

data: ciudades\$POB
weights: w

Moran I statistic standard deviate = -0.716, p-value = 0.763

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
-0.057378727	-0.028571429	0.001618745

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 11: CUADRO DE RESULTADOS DEL ESTADISTICO I-MORAN PARA LA POBLACIÓN DESOCUPADA EN CIUDADES DE MÉXICO DURANTE EL TERCER TRIMESTRE DEL 2019

Moran I test under randomisation

data: ciudades\$DESOCUPADO
weights: w

Moran I statistic standard deviate = -0.78967, p-value = 0.7851

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
-0.104238557	-0.028571429	0.009181623

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE

ANEXO 12: FACTOR DE INFLACIÓN DE LA VARIANZA EN MODELOS DE MINIMOS CUADRADOS ORDINARIOS PARA EL TERCER TRIMESTRE DEL 2009

	VIF	TOL	Wi	Fi	Leamer	CVIF	Klein	IND1	IND2
INFORMALID	1.5851	0.6309	5.4614	8.4847	0.7943	2.6474	0	0.0676	1.3286
ANIOS_ESC	1.4053	0.7116	3.7826	5.8766	0.8436	2.3470	0	0.0762	1.0380
DESOCUPADO	1.5494	0.6454	5.1281	7.9668	0.8034	2.5878	0	0.0691	1.2763
TERC_EDA	1.1102	0.9008	1.0282	1.5974	0.9491	1.8541	0	0.0965	0.3572

1 --> COLLINEARITY is detected by the test

0 --> COLLINEARITY is not detected by the test

INFORMALID , DESOCUPADO , TERC_EDA , coefficient(s) are non-significant may be due to multicollinearity

FUENTE: ELABORACIÓN PROPIA A TRAVÉS DE SOFTWARE RSTUDIO CON BASE EN DATOS DE LA ENOE