



Universidad Autónoma de Baja California
Facultad de Ciencias

Reporte final de proyecto

Análisis de transformación socioespacial por AGEB en la ciudad de
Ensenada.

Maximiliano Vaca Montejano
Alondra Alejandra Valdez García

Ensenada, Baja California, México

22 de mayo de 2024

Índice

1. Resumen	
2. Introducción	
3. Antecedentes	
4. Planteamiento del problema	
5. Pregunta de investigación	
6. Objetivos	
7. Marco Teórico	
7.1. Modelos lineales	
7.1.1. Modelo lineal mixto.....	
7.2 Significancia estadística y p-value.....	
7.2.1 Interpretación del p-value.....	
7.3 Residuals vs Fitted.....	
7.3.1 Interpretación.....	
8. Metodología	
8.1. Descripción de la base de datos	
8.2. Definición del ITSE	
8.3. Ajuste de modelos lineales	
8.3.1. Modelo lineal simple	
8.3.2. Modelo lineal mixto.	
8.4 Evaluación del modelo.....	
8.4.1 Análisis de residuales	
9. Descripción de los datos	
9.Resultados	
9.1. Modelo lineal simple	
9.2. Modelo lineal mixto	
10.Conclusión	
11.Anexo	

1. Resumen

La gentrificación es un fenómeno urbano caracterizado por la transformación de barrios populares en áreas de clase media o alta, resultando en el desplazamiento de los residentes originales. A pesar de los numerosos estudios, sigue faltando un análisis a gran escala y metodologías comparativas que aborden el problema de manera integral. Este trabajo se enfoca en la ciudad de Ensenada, Baja California, y utiliza métodos de minería de datos y modelos lineales mixtos para determinar el grado de gentrificación presente en el municipio.

El objetivo principal es calcular el Índice de Transformación Socioespacial (ITSE) para las Áreas Geoestadísticas Básicas (AGEB) de Ensenada. El ITSE se define utilizando ocho variables sociales y espaciales obtenidas de censos poblacionales. Además, se evaluará la idoneidad de estas variables para medir la gentrificación y se considerará la posibilidad de incluir variables relevantes o eliminar variables que aportan poca información.

Para el análisis se emplean los **p-values** para determinar la significancia de cada variable en los modelos planteados. Los resultados fueron sorprendentes, al revelar la baja significancia de variables que intuitivamente parecían de peso en la problemática como el promedio de los salarios mínimos por localidad. Mientras que se descubrió una significancia aceptable en variables que hasta el momento habían sido ignoradas para el estudio.

2. Introducción

El concepto de "gentrificación" proviene del inglés "gentrification", derivado de "gentry" (noble, burgués). Este término fue acuñado en 1964 por la socióloga Ruth Glass para describir las transformaciones en los barrios obreros del centro de Londres, donde las clases medias y altas comenzaron a desplazar a la clase trabajadora residente.

La gentrificación se refiere a la transformación urbana de un barrio o vecindario que históricamente ha sido habitado por una población de bajos ingresos. Este proceso involucra un cambio demográfico y socioeconómico que resulta en el incremento de los precios de la vivienda y la expulsión de los residentes originales. Generalmente, este fenómeno está asociado con la llegada de nuevos residentes de clase media o alta que buscan viviendas más asequibles, lo que a su vez transforma la infraestructura y la cultura del vecindario.

Díaz Parra, Apaolaza (2020) señalan que Casgrain y Janoschka (2013), siguiendo a Lees, Slater y Wyly (2008), intentan arrojar luz sobre el asunto, incluyendo cuatro elementos que deberían coincidir para hablar de gentrificación: *a)* reinversión y alza del valor del suelo, *b)* llegada de agentes con mayor capacidad de pago, *c)* cambios en las actividades y el paisaje urbano, y *d)* desplazamiento de los grupos sociales con menores ingresos.

A pesar de ser un tema de estudio desde hace 60 años, la gentrificación ha cobrado relevancia en el contexto latinoamericano a partir de la década de los 90. En la Ciudad de México, este fenómeno se ha intensificado desde 2001, impulsado por la inversión en reformas y restauraciones en el centro de la ciudad.

La gentrificación es un tema controvertido, ya que, aunque puede llevar a mejoras en la calidad de vida y a la revitalización urbana, también puede tener efectos negativos sobre los residentes originales, generando problemas sociales como la exclusión y la desigualdad económica. Este estudio se centra en medir la gentrificación en **nuestra ciudad** utilizando técnicas de minería de datos y modelación lineal, con el propósito de comprender mejor las dinámicas y consecuencias de este proceso.

3. Antecedentes

En busca de información y metodologías que involucren el fenómeno de la gentrificación, se encontraron en su mayoría una gran cantidad de artículos abordando el tema desde un enfoque cualitativo, señalando las formas en las que se ven afectados distintos grupos históricamente segregados. A pesar de la evidente importancia que tiene este tipo de análisis, el objetivo principal de este proyecto es el análisis cuantitativo. Si bien no son mayoría, nos quedamos con artículos que tuvieran un enfoque similar al que buscamos ya sea desde la metodología o los objetivos. Entre estos el que más destaca y es fuente principal del análisis aquí realizado es *Cambios socioterritoriales e indicios de gentrificación, Un método para su medición*.^[1] Ya que establece una metodología centrada en datos de censos realizados por el INEGI y una forma cuantitativa de analizar la gentrificación mediante la definición del ITSE (Índice de Transformación Socioespacial).

4. Planteamiento del Problema

A pesar de los numerosos estudios sobre gentrificación, falta un análisis a gran escala y metodologías comparativas y sistemáticas que aborden el problema de manera integral. La ausencia de políticas efectivas para mitigar los efectos negativos de la gentrificación, como señala Harvey (en Martí y Salazar, 2016), se debe en parte a la carencia de información sistemática y longitudinal que evalúe estos procesos en distintas ciudades.

Para abordar eficazmente la gentrificación, es crucial anticipar y prever su ocurrencia en lugar de mitigar daños ya establecidos. Las políticas anti deslizamiento tempranas, destacadas por Marcuse (1985), Levy, Comey y Padilla (2006), y Bates (2013), subrayan la importancia de anticipar estos procesos para evitar consecuencias irreversibles (Blanco y Apaolaza, 2016).

Se requiere una aproximación sistemática que use variables significativas en un amplio territorio y a lo largo del tiempo, capturando así la dinámica de la gentrificación. Los censos de población son ideales para este tipo de análisis.

El objetivo de este proyecto es proporcionar una visión integral de los espacios urbanos gentrificados o con potencial de transformación. Este análisis será útil para la elaboración de políticas públicas y acciones específicas que reduzcan los efectos negativos de la gentrificación en la comunidad.

5. Pregunta de investigación

Al revisar el artículo en el que basamos nuestra metodología notamos que para definir el ITSE (Índice de transformación socioespacial) se utilizan 8 variables de los censos poblacionales, categorizables como variables sociales y espaciales debido a la naturaleza del fenómeno, que involucra ambos frentes. Cada una de estas variables está directamente relacionada con la gentrificación ya sea de forma negativa o positiva según el artículo.

Además de seguir la metodología descrita para el caso de estudio específico de Ensenada, el objetivo de este proyecto es comprobar y darnos una idea de que tan relacionadas están las variables elegidas con la gentrificación, si es posible que haya otras variables que nos sean de utilidad o si a nuestra elección le sobran variables que no aportan demasiado al índice.

El principal problema para nuestro objetivo es la falta de datos específicos que están relacionados con las variables que queremos poner a prueba.

6. Objetivos

Calcular el Índice de transformación socioespacial (ITSE) por AGEB en Ensenada:



- Utilizar los censos de población y vivienda para obtener los datos relevantes para el estudio.
- Aplicar los criterios de variables para la medición de gentrificación a través de un sistema de puntos.

Caracterizar el problema por medio de un modelo lineal ■

- Utilizar las variables planteadas para ajustar un modelo lineal que sea de utilidad para la problemática ■
- Realizar un análisis comparativo de modelos ■
- Analizar los residuales par verificar el correcto ajuste del modelo ■

Evaluar la idoneidad de las variables definidas ■

- Comprobar la relevancia y relación de las variables con la gentrificación en Ensenada.
- Determinar cuáles variables tienen efecto en la gentrificación y en qué medida.

7. Marco Teórico

7.1. Modelo lineal

Los modelos lineales son una herramienta fundamental en estadística y análisis de datos, utilizados para describir la relación entre una variable dependiente y una o más variables independientes. La simplicidad y flexibilidad de los modelos lineales los hacen adecuados para una amplia gama de aplicaciones en diversas disciplinas.

Un modelo lineal básico se expresa generalmente de la forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

Figura 1.1



donde:

- Y es la variable dependiente.
- X_1, X_2, \dots, X_p son las variables independientes.
- β_0 es el intercepto.
- $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes de las variables independientes.
- ϵ es el término de error.

con valor esperado $E(Y_i) = \mu_i$ y ϵ_i variables aleatorias con media igual a cero.

7.1.1 Modelo lineal mixto

El modelo lineal mixto (MLM) es una extensión del modelo lineal clásico que permite incorporar tanto efectos fijos como aleatorios. Los efectos fijos son constantes para toda la población, mientras que los efectos aleatorios varían entre diferentes grupos o niveles jerárquicos. Esta característica hace que los modelos lineales mixtos sean especialmente útiles para datos que tienen una estructura jerárquica o agrupada.

Un modelo lineal mixto puede expresarse como:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_i + \epsilon_{ij}$$

Figura 1.2

donde:

- Y_{ij} es la variable dependiente para el j -ésimo caso en el i -ésimo grupo.
- X_{ij} es la variable independiente para el j -ésimo caso en el i -ésimo grupo.
- β_0 y β_1 son los coeficientes fijos.
- u_i es el efecto aleatorio para el i -ésimo grupo.
- ϵ_{ij} es el término de error.

Los modelos lineales mixtos son particularmente valiosos en situaciones donde los datos tienen dependencias dentro de los grupos, como en estudios longitudinales, estudios multinivel y datos anidados. Permiten modelar estas dependencias y proporcionar estimaciones más precisas de los efectos de interés.

7.2 Significancia estadística y p-value

El p-value, o valor p es utilizado ampliamente para evaluar la significancia de los resultados obtenidos en un análisis estadístico. El valor p ayuda a determinar la evidencia contra una hipótesis nula, proporcionando una medida cuantitativa de dicha evidencia.

El p-value es la probabilidad de obtener un resultado igual o más extremo que el observado, bajo la suposición de que la hipótesis nula (H_0) es verdadera. En términos matemáticos, se puede definir como:

$$p = P(\text{datos observados} | H_0 \text{ es verdadero})$$

7.2.1 Interpretación del p-Value

El p-value se utiliza para decidir si se debe rechazar la hipótesis nula en favor de la hipótesis alternativa (H_1). La interpretación del p-value se basa en un umbral predefinido, denominado nivel de significancia o tolerancia α comúnmente fijado en 0.05 o 0.01. La interpretación estándar del p-value es la siguiente:

- $p\text{-value} \leq \alpha$: Se rechaza la hipótesis nula. Existe suficiente evidencia estadística para apoyar la hipótesis alternativa.
- $p\text{-value} > \alpha$: No se rechaza la hipótesis nula. No hay suficiente evidencia estadística para apoyar la hipótesis alternativa.

Es importante destacar que un p-value bajo no implica necesariamente que la hipótesis alternativa sea verdadera, sino que hay suficiente evidencia contra la hipótesis nula.

7.3 Residuals vs Fitted

Las gráficas de residuos vs valores ajustados (residuals vs fitted plots) son una herramienta para evaluar la adecuación de un modelo. Estas gráficas permiten visualizar la relación entre los residuos y los valores ajustados, proporcionando información crucial sobre la validez de los supuestos del modelo y la identificación de posibles problemas.

Los valores ajustados \hat{y} son las predicciones obtenidas del modelo estadístico en un modelo de regresión lineal figura 1.1. Los residuos son las diferencias entre los valores observados y_i y los valores ajustados \hat{y}_i .

La gráfica residuals vs fitted es una representación gráfica de los residuos frente a los valores fijados. Esta gráfica es esencial para diagnosticar varios aspectos del modelo de regresión:

1. Linealidad:

Para un modelo de regresión lineal, se espera que los residuos no muestran patrones sistemáticos cuando se trazan contra los valores ajustados. Si se observa una tendencia o patrón, puede indicar que la relación entre las variables no es lineal y que quizás un modelo no lineal sería más apropiado.

2. Normalidad de los Residuos:

Aunque la gráfica residuals vs fitted no evalúa directamente la normalidad de los residuos, un patrón claro o una estructura en los residuos podría sugerir desviaciones de la normalidad. La normalidad de los residuos se evalúa mejor con otras gráficas, como el gráfico Q-Q (cuantil-cuantil).

3. Identificación de Outliers:

Los puntos atípicos (outliers) pueden identificarse como residuos que se alejan significativamente de la línea horizontal. Estos puntos pueden tener un impacto desproporcionado en la estimación de los coeficientes del modelo y deben ser investigados.

7.3.1 Interpretación de la Gráfica **Residuals vs Fitted**

La correcta interpretación de la gráfica residuals vs fitted involucra varios pasos:

1. Distribución Aleatoria de los Residuos:



Los residuos deben **distribuirse de manera aleatoria** alrededor de la línea horizontal que representa cero. Esto indica que no hay patrones sistemáticos y que el modelo es adecuado en términos de linealidad y homocedasticidad.

2. Ausencia de Patrones Sistemáticos:

La presencia de un patrón en forma de curva sugiere una relación no lineal entre las variables independientes y la variable dependiente, lo que podría indicar la necesidad de transformar las variables o de utilizar un modelo no lineal.

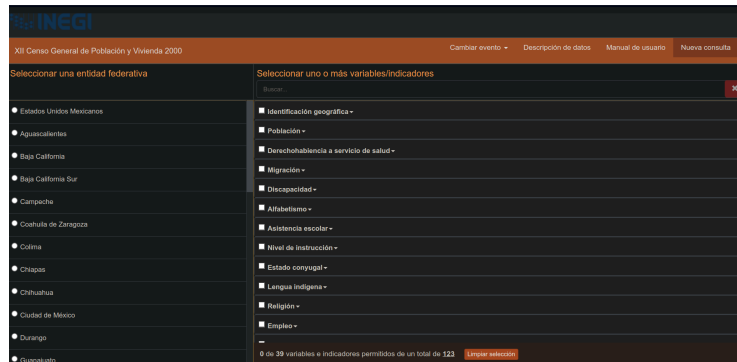
3. Varianza Constante de los Residuos:

Si la variabilidad de los residuos aumenta o disminuye con los valores ajustados, indica heterocedasticidad. Esto sugiere que la varianza de los errores no es constante y puede requerir técnicas como la transformación de las variables o el uso de estimadores robustos.

8. Metodología

8.1. Descripción de la base de datos

Los datos que vamos a utilizar fueron extraídos de las bases de datos públicas ofrecidas por el Instituto Nacional de Estadística y Geografía (INEGI), correspondientes a los censos poblacionales de 2010 y 2020 para el ITSE, y el de 2000 para el análisis de las variables.



<https://www.inegi.org.mx/app/scitel/Default?ev=3>



La interfaz que permite acceder a los datos cuenta con selectores para filtrar las variables requeridas por estado y categoría. Las categorías en las que estamos interesados son principalmente: Identificación geográfica, Nivel de instrucción, Vivienda y Bienes electrodomésticos. Para el análisis y comprobación de variables nos aprovecharemos del hecho de que en el censo del año 2000 contamos con datos, en la sección de Empleo, correspondientes a *Población ocupada que recibe X y hasta X salarios mínimos mensuales de ingreso por trabajo*. El único inconveniente en este censo se encuentra en que la unidad más pequeña de análisis geográfico es la localidad, aún no se divide por AGEB.

Otra base de datos utilizada es proveniente, de nuevo, del Instituto Nacional de Estadística y Geografía, aunque esta vez a través de una librería de python llamada INEGIpy

“INEGIpy es una librería para interactuar fácilmente con los datos del Instituto Nacional de Geografía y Estadística (INEGI) desde Python. Su propósito es apoyar en la creación de consultas automatizadas y en el acceso a la información para el análisis de datos.”

<https://pypi.org/project/INEGIpy/>



La cual fue de mucha utilidad para crear relaciones entre los datos trabajados y sus respectivos polígonos geográficos al momento de realizar el mapa sobre el cual se visualizará el ITSE.

Las variables extraídas de los censos [2010](#) y [2020](#) fueron:

- LOC: Clave de localidad
- NOM_LOC: Localidad
- AGEB: Clave de AGEB
- MZA: Clave de manzana
- POBTOT: Población total

- P_0A2: Población de 0 a 2 años
- P_3A5: Poblacion de 3 a 5 años
- POB65_MAS: Población de 65 años y más
- GRAPROES: Grado promedio de escolaridad
- PROM_HNV: Promedio de hijos nacidos vivos
- VIVTOT: Total de viviendas
- PRO_OCUP_C: Promedio de ocupantes por cuarto
- PROM_OCUP: Promedio de ocupantes por vivienda
- VPH_PC: Viviendas que disponen de computadora
- PNACOE: Población nacida en otra entidad

Mientras que las del [censo del 2000](#) fueron:

Variables dependientes (salarios mínimos):

- P_1SM (Población ocupada que recibe menos de un salario mínimo mensual de ingreso por trabajo)
- P1_2SM (Población ocupada que recibe 1 y hasta 2 salarios mínimos mensuales de ingreso por trabajo)
- P2_5SM (Población ocupada que recibe 2 y hasta 5 salarios mínimos mensuales de ingreso por trabajo)
- P6_10SM (Población ocupada que recibe 6 y hasta 10 salarios mínimos mensuales de ingreso por trabajo)
- P10_SM (Población ocupada con más de 10 salarios mínimos mensuales de ingreso por trabajo)

Variables independientes (potenciales influencias):

- P_TOTAL (Población total)
- POB0_4 (Población de 0 a 4 años)
- POB18_ (Población mayor de 18 años)
- PNACOE (Población nacida en otra entidad)
- GRADOESCO (Grado promedio de escolaridad)
- PCASADA12_ (Proporción de personas casadas mayores de 12 años)
- T_VIVHAB (Total de viviendas habitadas)
- VIVPARHAB (Viviendas particulares habitadas)
- PRO_OCV (Promedio de ocupantes por vivienda particular habitada)
- PRO_OVP (Promedio de ocupantes por vivienda)
- VP_TV (Viviendas particulares habitadas con TV)
- TOTHOG (Total de hogares)

Los datos del censo del 2000 contienen datos hasta nivel local de las siguientes entidades: Aguascalientes, Baja California, Baja California Sur, Campeche, Coahuila de Zaragoza, Colima y Chiapas. Las cuales se seleccionan automáticamente al elegir “Estados Unidos Mexicanos” como opción en el menú de consulta.

8.2. Definición del ITSE

Para la definición del índice de transformación socioespacial (ITSE) se tomaron en cuenta dos los componentes sociales y espaciales que conforman el fenómeno de la gentrificación. Para caracterizar ambos componentes se emplearon ocho variables

Variables sociales

1. Población total (PT)
2. Niños de 0-4 años (N)
3. Adultos mayores de 65 años (AM)
4. Grado promedio de escolaridad (E)
5. Promedio de hijos nacidos vivos (H)

Variables espaciales

6. Número de viviendas (V)
7. Promedio de ocupantes por cuarto (OC)
8. Viviendas particulares habitadas que disponen de computadora ©

Una vez definidas las variables a utilizar se calculó el promedio de sus valores, el cual fue utilizado como referencia. Adicionalmente a las variables anteriores se agregó un promedio de salarios mínimos por localidad, esta variable no forma parte del ITSE definido en *Cambios socioterritoriales e indicios de gentrificación, Un método para su medición*. [1], pero se consideró que era una variable valiosa que podía tener una alta relación con el fenómeno.

8.3. Ajuste de modelos lineales

El análisis por medio de modelos lineales se hará sobre los datos del censo del año 2000, esto debido a que contienen información sobre ingresos en salarios mínimos, lo cual utilizaremos como variable dependiente para observar las posibles influencias que tiene el resto de variables

8.3.1. Modelo lineal simple

Armamos el modelo con la siguiente pregunta en mente:

- ¿Qué tanto peso tiene el resto de variables en el promedio de salarios mínimos?
- Las variables que elegimos son lo suficientemente significativas para describir el fenómeno?

Para esto definimos un promedio de salarios mínimos en cada fila (localidad) como *PROM_SM*

La fórmula de nuestro modelo lineal simple será:

$$\begin{aligned} PROM_SM \sim & P_TOTAL + POB0_4 + POB18_ + PNAOENT + GRADOESCO \\ & + PCASADA12_ + T_VIVHAB + VIVPARHAB + PRO_OCVP \\ & + PRO_OVP + VP_TV + TOTHOG \end{aligned}$$

8.3.2. Modelo lineal mixto

Anteriormente habíamos definido un promedio de salarios mínimos y usado un modelo lineal simple, el cual asume independencia y ningún tipo de estructura de correlación entre las variables, lo cual es algo torpe considerando la estructura de división geográfica entre los datos, como lo son las entidades, municipio y localidad.

En contraparte, el utilizar un modelo lineal mixto nos permite incorporar efectos aleatorios y tomar en cuenta la estructura jerárquica de los datos con los que estamos trabajando (localidades dentro de municipios y municipios dentro de entidades) obteniendo información sobre la variabilidad entre los grupos y dentro de estos.

Elegimos un modelo mixto como segunda opción debido a que resulta natural pensar en una correlación de los salarios mínimos y su variabilidad dentro de las entidades, municipios y localidades, atribuible a factores del propio municipio o entidad.

Para poner en marcha el modelo hace falta definir bien cuáles son los factores que tomaremos como aleatorios y fijos.

Factores Aleatorios y Fijos:

- **Factores Fijos:** Variables que esperamos que influyan sistemáticamente en los salarios mínimos, que son: P_TOTAL, POB0_4, POB18_, PNACOENT, GRADOESCO, PCASADA12_, T_VIVHAB, VIVPARHAB, PRO_OCPV, PRO_OVP, VP_TV, TOTHOG.
- **Factores Aleatorios:** ENTIDAD, MUN, LOC como niveles de agrupamiento posibles que representan variabilidad no explicada entre estas categorías de otra forma más que por diferencias geográficas.

Existe la hipótesis de que perdemos información al definir un promedio de salarios mínimos en lugar de tomar cada uno de forma independiente, por lo que vamos a probar dos modelos: uno modelando P_1SM y otro PROM_S.

Modelos:

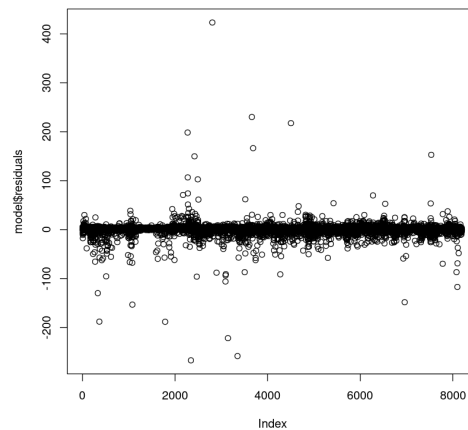
$$P_{1SM} \sim P_{TOTAL} + POB0_4 + POB18_ + PNACOENT + GRADOESCO + PCASADA12_ + T_VIVHAB + VIVPARHAB + PRO_OCPV + PRO_OVP + VP_TV + TOTHOG, random = \sim 1 | ENTIDAD/MUN/LOC$$

$$PROM_{SM} \sim P_{TOTAL} + POB0_4 + POB18_ + PNACOENT + GRADOESCO + PCASADA12_ + T_VIVHAB + VIVPARHAB + PRO_OCPV + PRO_OVP + VP_TV + TOTHOG, random = \sim 1 | ENTIDAD/MUN/LOC$$

8.4. Evaluación del modelo

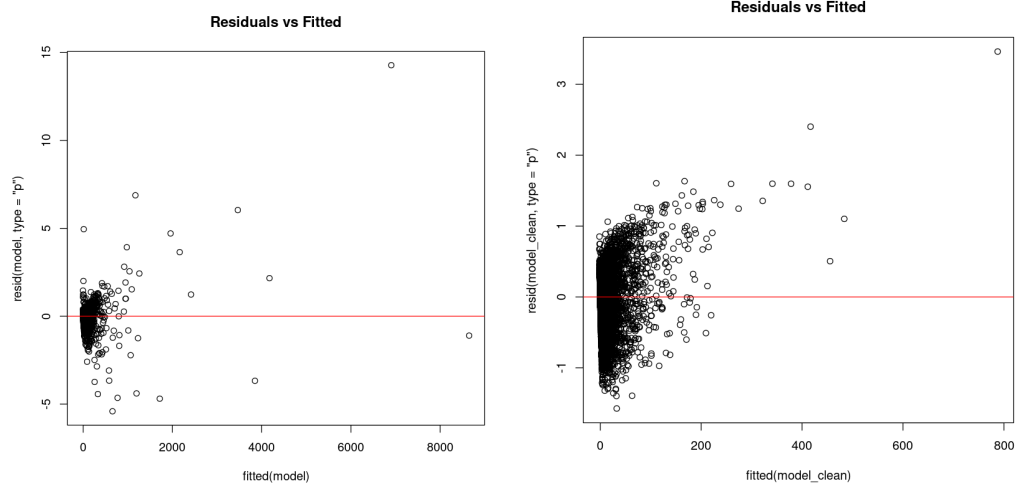
8.4.1. Análisis de residuales

- Modelo lineal simple



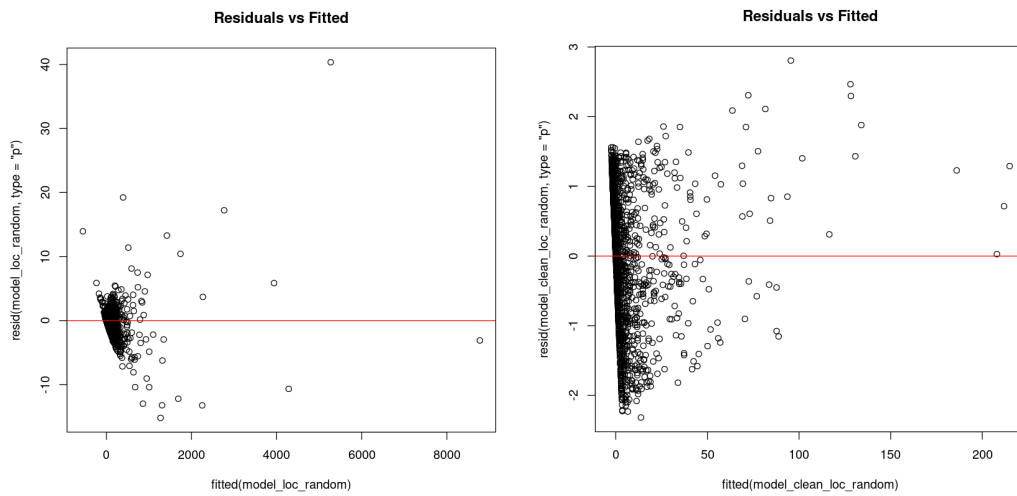
Los residuales se distribuyen casi uniformemente alrededor del 0

- Modelo 1 salario mínimo:



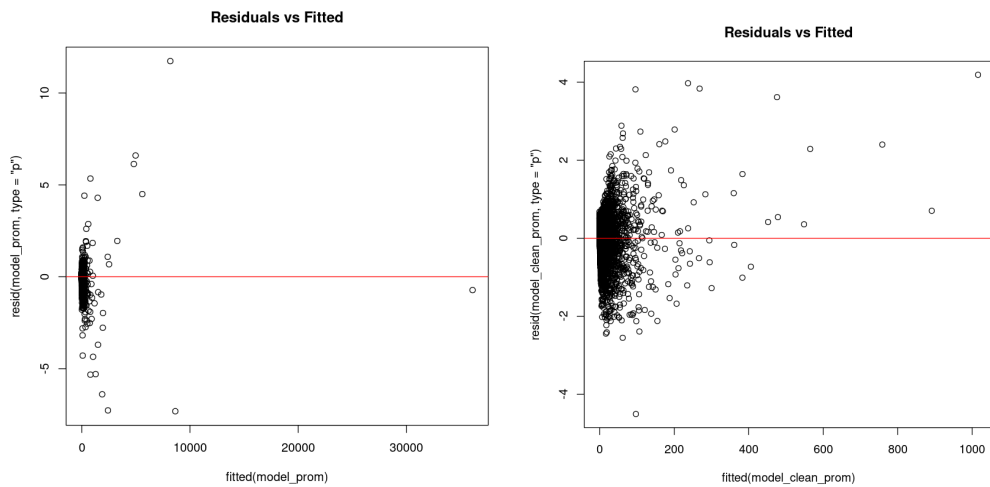
Los residuales parecen agruparse alrededor del 0, aunque de manera no muy uniforme y el ajuste no ayuda.

- Ajuste del modelo mixto considerando LOC como efecto aleatorio:



Al considerar LOC como efecto aleatorio y borrando outliers el modelo luce mejor, se distribuye de forma verticalmente uniforme alrededor del 0.

- Ajuste del modelo definiendo promedio de salarios mínimos



Es el que mejor luce hasta ahora, los datos se distribuyen de forma verticalmente uniforme alrededor del 0, un poco agrupados hacia la izquierda.

9. Resultados

9.1 Modelo lineal simple

Una vez aplicado el modelo lineal simple se obtienen los siguientes resultados:

```
Call:
lm(formula = PROM_SM ~ ., data = data2k)

Residuals:
    Min       1Q   Median       3Q      Max
-267.31  -0.73    1.21    2.62   423.36

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4266207  0.7558336   0.564 0.572472
P_TOTAL      -0.0945230  0.0026175 -36.112 < 2e-16 ***
POB0_4        0.4535314  0.0075789  59.841 < 2e-16 ***
POB18         0.1786059  0.0042411  42.113 < 2e-16 ***
PNACOENT      0.0020041  0.0006059   3.308 0.000945 ***
GRADOESCO    -0.1361300  0.0794439  -1.714 0.086652 .
PCASADA12_   -0.1438141  0.0045822 -31.386 < 2e-16 ***
T_VIVHAB      0.5195937  0.0188506  27.564 < 2e-16 ***
VIVPARHAB    -1.0732385  0.0297907 -36.026 < 2e-16 ***
PRO_OCVP      0.3488696  0.1615722   2.159 0.030862 *
PRO_OVP      -0.5714624  0.1390296  -4.110 3.99e-05 ***
VP_TV         0.1351991  0.0039377  34.335 < 2e-16 ***
TOTHOOG       0.6789565  0.0192538  35.263 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.67 on 8184 degrees of freedom
Multiple R-squared:  0.9992,    Adjusted R-squared:  0.9992
F-statistic: 8.287e+05 on 12 and 8184 DF,  p-value: < 2.2e-16
```

Al observar los resultados y los pesos asignados a cada regresor, nos damos cuenta de que la mayoría de variables utilizadas son significativas, salvo algunas excepciones interesantes:

- GRADOESCO (Grado promedio de escolaridad) no es una variable significativa en el modelo para describir el promedio de salarios mínimos por localidad. Es un resultado curioso ya que es una variable que se encontraba en las elegidas en la metodología del artículo [1] para definir el ITSE. Aunque recordemos que el promedio de salarios mínimos, aun siendo una variable valiosa con la que no contamos en 2010 y 2020, no es el único factor determinante para medir la gentrificación.
- PNACOENT (Población nacida en otra entidad) es una variable significativa, lo cual suena evidente pero nos es útil comprobar su relación a través del modelo debido a que es una variable que no se encontraba en la lista original para definir el ITSE.
- PRO_OCVP (Promedio de ocupantes por vivienda particular habitada) a pesar de ser una variable significativa, lo es menos que las otras, incluso que PNACOENT que no se encontraba en las variables elegidas.

9.2 Modelo lineal mixto

Una vez aplicado el modelo lineal mixto se obtienen los siguientes resultados:

P_1SM ~

	Value	Std. Error	DF	t-value	p-value
(Intercept)	-5.362631	3.497574	8033	-1.53324	0.1253
P_TOTAL	0.408600	0.009643	8033	42.37372	0.0000
POB0_4	-1.312213	0.028025	8033	-46.82230	0.0000
POB18_	-0.559552	0.015709	8033	-35.61969	0.0000
PNACOENT	-0.076409	0.002229	8033	-34.27729	0.0000
GRADOESCO	0.247493	0.324532	8033	0.76261	0.4457
PCASADA12_	0.141220	0.016899	8033	8.35674	0.0000
T_VIVHAB	-1.631986	0.069746	8033	-23.39882	0.0000
VIVPARHAB	-0.382625	0.109667	8033	-3.48897	0.0005
PRO_OCVP	0.574152	0.643519	8033	0.89221	0.3723
PRO_OVP	-1.431402	0.535673	8033	-2.67215	0.0076
VP_TV	-0.972231	0.014808	8033	-65.65636	0.0000
TOTHOG	3.122157	0.070449	8033	44.31778	0.0000

Al observar los resultados, resulta evidente que la mayor parte de las observaciones respecto al modelo lineal simple se conservan:

- GRADOESCO (Grado promedio de escolaridad) resalta como la variable menos significativa del modelo.
- PRO_OCVP (Promedio de ocupantes por vivienda particular habitada) resalta como la segunda variable menos significativa del modelo.

PROM_SM ~

	Value	Std. Error	DF	t-value	p-value
(Intercept)	-0.1134588	1.0899977	8033	-0.10409	0.9171
P_TOTAL	-0.0947080	0.0026392	8033	-35.88460	0.0000
POB0_4	0.4562863	0.0076813	8033	59.40207	0.0000
POB18_	0.1766101	0.0043076	8033	41.00004	0.0000
PNACOENT	0.0014056	0.0006099	8033	2.30441	0.0212
GRADOESCO	-0.1910846	0.0895430	8033	-2.13400	0.0329
PCASADA12_	-0.1456282	0.0046320	8033	-31.43974	0.0000
T_VIVHAB	0.5122194	0.0190832	8033	26.84131	0.0000
VIVPARHAB	-1.0690868	0.0300014	8033	-35.63457	0.0000
PRO_OCVP	0.3492266	0.1780192	8033	1.96174	0.0498
PRO_OVP	-0.3765859	0.1473805	8033	-2.55520	0.0106
VP_TV	0.1384217	0.0040614	8033	34.08267	0.0000
TOTHOG	0.6873781	0.0192580	8033	35.69304	0.0000

Obtenemos resultados similares a los dos modelos anteriores:

- GRADOESCO y PRO_OCVP destacan como las variables menos significativas para describir el promedio de salario mínimo.

Aunque obtenemos una excepción:

- PNACOENT (Población nacida en otra entidad) se muestra como una variable poco significativa, concordando con lo establecido en el artículo del que se extrajo la metodología. Recordemos que el modelo mixto es el que mejor tiene en cuenta la naturaleza jerarquizada de los datos, marcada por su origen geográfico.

10. Conclusión

El trabajo y análisis realizado en este proyecto para verificar la relación que existe entre las variables escogidas para definir el ITSE y la gentrificación (desde el lado económico) aprovechando los datos disponibles del censo de población y vivienda del año 2000 acerca de los ingresos en salario mínimo por localidad, concluye en una puesta en duda acerca de la fiabilidad al momento de escoger el grado promedio de escolaridad GRADOESCO (2000) o GRAPROES (2010 y 2020) como una variable relacionada al fenómeno de la gentrificación, por lo menos desde el frente económico del tema, aunque haya resultado intuitivo al pensar en ello. Lo mismo va para PRO_OCVP (promedio de ocupantes por vivienda), aunque con un menor grado de desconfianza.

Por otra parte, el resto de las variables parecen estar relacionadas significativamente con el fenómeno.

También valdría la pena tener en consideración la variable PNACOENT (Población nacida en otra entidad), que intuitivamente tiene sentido, es significativa en la mayoría de los modelos y no se había tomado en cuenta.

11. Anexo

Análisis descrito en este proyecto:

https://github.com/MaximilianoVM/MLR_Proyecto_AGEBS.git

Análisis previos ITSE enseñada, K-means clustering:

<https://github.com/MaximilianoVM/AGEBS-Ensenada.git>

Referencias

1. Bournazou Marcou, E. D. (2015, agosto). Cambios socioterritoriales e indicios de gentrificación: Un método para su medición. Academia XXII, Recuperado de <https://www.revistas.unam.mx/index.php/aca/article/view/51982>
2. Instituto Nacional de Estadística y Geografía. (n.d.). Área Geoestadística Básica (AGEB). Recuperado el 9 de abril de 2023, recuperado de <https://www.inegi.org.mx/app/glosario/default.html>
3. Instituto Nacional de Estadística y Geografía. (2010). Censos de Población y Vivienda. Recuperado el 21 de febrero de 2023, recuperado de <https://www.inegi.org.mx/programas/ccpv/2010/#Herramientas>

4. Cruz Flores, A. (2022). Impactos socioeconómicos y territoriales de las políticas para el desarrollo urbano: ¿gentrificación en la zona de Tacubaya CDMX 2010-2021? [Tesis de maestría, El Colegio de la Frontera Norte, A.C.]. Tijuana, B.C. Recuperado de <https://www.colef.mx/posgrado/wp-content/uploads/2022/10/TESIS-Cruz-Flores-Alfredo-MDR.pdf>
5. Koh, L. (2017). Can Unsupervised Clustering Methods Identify Gentrification? Recuperado de <https://leyankoh.wordpress.com/2017/12/12/can-unsupervised-clustering-methods-identify-gentrification>
6. Corrigan, A. E., Curriero, F. C., & Linton, S. L. (2021). Characterizing clusters of gentrification in metro Atlanta, 2000 to 2016. *Applied Geography*, 137, 102597. <https://doi.org/10.1016/j.apgeog.2021.102597> recuperado de <https://pure.johnshopkins.edu/en/publications/characterizing-clusters-of-gentrification-in-metro-atlanta-2000-t>
7. Blanco, J. y Apaolaza, R. (2016). Políticas y geografías del desplazamiento. Contextos y usos conceptuales para el debate sobre la gentrificación. *Revista INVI*, 31(88), 73-98. Recuperado de <http://www.revistainvi.uchile.cl/index.php/INVI/article/view/1085>
8. Marcuse, P. (1985). To control gentrification: Anti-displacement zoning and planning for stable residential districts. *Review of Law and Social Change*, 13(4), 931-952. Recuperado de <https://socialchangenyu.com/review/to-control-gentrification-anti-displacement-zoning-andplanning-for-stable-residential-districts/>
9. Levy, D., Comey, J. y Padilla, S. (2006). *In the face of gentrification: Case studies of local efforts to mitigate displacement*. Washington, DC: The Urban Institute. Recuperado de <https://www.urban.org/sites/default/files/publication/50791/411294-In-the-Face-of-Gentrification.PDF>
10. Apaolaza, R. (2018). *Territorio, transporte y capitales. Dinámicas y efectos del aislamiento socio territorial sobre los jóvenes residentes de los nuevos asentamientos periféricos del Área Metropolitana de Buenos Aires* (Tesis doctoral, Universidad de Buenos Aires, Facultad de Filosofía y Letras). Recuperado de <http://repositorio.filo.uba.ar/handle/filodigital/4586>
11. Martí, M. y Salazar, M. (2016). Entrevista a David Harvey sobre gentrificación: Hábitat III tiene una posición neoliberal. *El Canelazo de la Ciudad*, 5, 10-15. Recuperado de <https://elcanelazodelaciudad.files.wordpress.com/2016/04/canelazo-5.pdf>

12. Díaz Parra, I., Apaolaza, R. (2020). Una propuesta metodológica para identificar la gentrificación a partir de los censos de población. *Estudios Demográficos y Urbanos, 35*(3). <https://doi.org/10.24201/edu.v35i3.1883>
13. Casgrain, A. y Janoschka, M. (2013). Gentrificación y resistencia en las ciudades latinoamericanas. El ejemplo de Santiago de Chile. *Andamios. Revista de Investigación Social*, 22, 19-44. Recuperado de <http://www.scielo.org.mx/pdf/anda/v10n22/v10n22a3.pdf>