

Informe Tarea 1

Análisis de Datos e Inferencia Causal

Integrantes del grupo:

Agustín Barria
Benjamín
Jorquera
Diego Concha
Felipe Pérez



INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE

Parte 1

Hipótesis: El nivel de pobreza de los habitantes de la comuna se relaciona con una mayor cantidad de viviendas irregulares catastradas.

La elección de esta hipótesis específicamente fue fundada bajo la justificación de que la cantidad de viviendas irregulares se relaciona directamente con el nivel de ingresos que cuentan las personas de cada comuna, es decir, bajo el supuesto de que se cuenta con mayor ingreso monetario para la estructuración, modificación y solidez del hogar de la persona. Estos ingresos se ven reflejados con el índice de pobreza que representa a cada comuna.

Las fuentes internacionales seleccionadas que avalan la hipótesis son las siguientes:

- Fuente: *Hernández, F., & Alejandra, I. (2021). El imaginario del espacio urbano en la periferia (1.ª ed., p. 16). Tláhuac, Ciudad de México. Recuperado de <https://repositorio.xoc.uam.mx/jspui/handle/123456789/24519> (En la ilustración 1 del anexo se ve la imagen referencial)*
- Fuente: *Iván Elías Ruiz Hernández. (2014). Identificación de asentamientos irregulares y diagnóstico de sus necesidades de infraestructura en Ciudad Juárez, Chihuahua, México (1.ª ed., p. 4). Chihuahua, México. Recuperado de https://www.scielo.org.mx/scielo.php?pid=S0188-46112015000200088&script=sci_arttext (En la ilustración 2 del anexo se ve la imagen referencial)*

Parte 2

Para la construcción de la base de datos se buscaron indicadores que como equipo se cree que afectan al número de viviendas irregulares en todas las comunas del país, dejando uno que se relacione con la hipótesis planteada, que es la encuesta CASEN que mide el índice de pobreza.

Para elegir los indicadores se utilizaron datos del SINIM, enfocándose en que cada indicador sea de distinta área, esto con el fin de que no tengan alta correlación entre sí y así no tener resultados sesgados a la hora de analizar los datos.

Área, subárea y nombre de indicador en base a los datos del SINIM.

Área	Sub-área	Indicador	Unidad de análisis
Administración y Finanzas Municipales	Ingresos Municipales (M\$)	Disponibilidad Presupuestaria Municipal por Habitante (M\$)	Millones (M\$)

Social y Comunitaria	Información Encuesta CASEN	Índice de Pobreza CASEN (Última Encuesta Vigente)	Porcentaje (%)
Desarrollo y Gestión Territorial	Características Territoriales	Superficie Total Comunal Km ²	Kilómetros Cuadrados (Km ²)
Caracterización de la Comuna	Geográfico Administrativa	Distancia con Respecto a la Capital Regional	Kilómetros (Kms).

La regresión lineal queda de la siguiente manera:

$$\begin{aligned}
 \text{N}^\circ \text{ viviendas irregulares} = & \beta_0 + \beta_1 * \text{Disponibilidad Presupuestaria Municipal por Habitante} + \beta_2 * \text{Indice de Pobreza CASEN} + \beta_3 * \text{Superficie Total Comunal km}^2 + \beta_4 * \text{Distancia con Respecto a la Capital Regional}
 \end{aligned}$$

Todos los datos encontrados son en base al 2020 (recomendación ayudante), a través del SINIM, a excepción de la Cantidad de viviendas irregulares por comuna que fueron obtenidos de la Información y datos de caracterización catastro 2022 del MINVU.

Cabe destacar que para el indicador y variable “Superficie total comunal en km²” había ciertos datos los cuales tenían de valor: *Sin dato oficial*, por ende, para no tener problemas con estos valores se realizó una investigación sobre estos en BCN (Biblioteca del Congreso Nacional de Chile), otorgándoles el valor correspondiente. De igual maneras surgieron problemas con el indicador “Distancia con respecto a la capital regional” que tenía valores con: *No recepcionado*, para solucionar esto buscamos la distancia correcta a través de Google Maps. Por último, encontramos los mismos valores de *No recepcionado* en el indicador Disponibilidad presupuestaria municipal por habitante, la solución en este caso fue otorgarles el valor promedio de este indicador a aquellos valores, dado que tan solo eran 4 comunas las que constataban este problema.

Parte 3:

Cabe destacar que para el indicador y variable “Superficie total comunal en km²” había ciertos datos los cuales tenían de valor: *Sin dato oficial*, por ende, para no tener problemas con estos valores se realizó una investigación sobre estos en BCN (Biblioteca del Congreso

Nacional de Chile), otorgándoles el valor correspondiente. De igual manera surgieron problemas con el indicador “Distancia con respecto a la capital regional” que tenía valores con: *No recepcionado*, para solucionar esto buscamos la distancia correcta a través de Google Maps. Por último, se encontraron los mismos valores de “*No Recepcionado*” en el indicador “Disponibilidad presupuestaria municipal por habitante”, la solución en este caso fue otorgarles el valor promedio de este indicador a aquellos valores, dado que tan solo eran 4 comunas las que constataban este problema.

Luego, con el fin de analizar la hipótesis, se agregó a la base de datos la columna “Porcentaje de población rural”, para de esta forma considerar la mediana de esta variable y diferenciar por comunas con un alto y bajo porcentaje de ruralidad. Específicamente, se desarrolló de mejor manera este análisis, haciendo un foco hacia los gráficos que se iban realizando uno para cada variable o indicador diferente, manteniendo fijo en el eje Y o variable dependiente a la variable “Cantidad de viviendas irregulares”. Se mostrarán estos gráficos en el Anexo.

Los resultados no fueron los esperados. Los datos no coincidían con las estimaciones previstas. Ejemplo, en el gráfico de Cantidad de Viviendas irregulares vs Disponibilidad presupuestaria municipal por habitante (Anexo, gráfico 1) se observó una curva extraña e inesperada (a mayor disponibilidad presupuestaria municipal por habitante había menores cantidades de viviendas irregulares). Es por esto que se realizó una indagación profunda de la columna de este indicador en la base de datos. En ella se visualizó la relación de estas variables, observando alguno de los datos, específicamente la comuna de Iquique que cuenta con un Índice de pobreza CASEN 4,38 tiene menor disponibilidad presupuestaria municipal por habitante 292,4 vs los 642,78 que considera la comuna Pozo Almonte, la cual cuenta con Índice de Pobreza CASEN 10,88. De esto, se obtuvo de conclusión que por un lado positivo no existe correlación entre las variables independientes (indicadores) y por el lado de la relación con la variable independiente se identificó que lo esperado estaba erróneo, para esto se investigó y encontró que en algunas comunas se predominaba el presupuesto municipal en otros servicios o bienes tanto para habitantes como en común, en vez de atribuirlo a las viviendas. (Párrafo de conclusión)

Se concluyó que existen sesgos que no permiten llegar al EDA estimado (comportamiento de los grafos). Se tomó la decisión de trabajar con las 4 regiones que poseen la mayor cantidad de viviendas irregulares¹ dentro de Chile: Antofagasta, Tarapacá, Valparaíso y Metropolitana, para ver el comportamiento que ocurre dentro de sus comunas, además, usando el indicador “Porcentaje de Ruralidad” se redujeron los datos tomando el percentil 80 de los datos para trabajar con las comunas que poseen mayor porcentaje de superficie urbana. (también se mostrarán cada uno de los gráficos en base a los acotamientos correspondientes en el anexo).

A continuación, se tiene el análisis EDA de los datos, en base a la reducción de datos aplicada:

Del gráfico 1 ya se comentó anteriormente, para el gráfico 2 del anexo se concluyó que no se puede identificar una tendencia uniforme debido a la existencia de outliers, por lo que no se puede encontrar la relación esperada entre las variables (a mayor índice de pobreza mayor es la cantidad de viviendas irregulares)

Del gráfico 3 nuevamente no se obtuvo la relación esperada que a mayor cantidad de superficie, mayor cantidad de viviendas irregulares, esto pudo haber ocurrido ya que al tomar todas las regiones hay una gran diferencia de tamaño en las regiones del sur del país (tienen mayor superficie) a comparación con las regiones donde hay mayor cantidad de viviendas irregulares. Provocando nuevamente outliers.

En el gráfico 4 se podría esperar que a mayor distancia de la capital regional, menor será el número de viviendas irregulares, dado a la urbanización que concentran las capitales de las regiones y por tanto abarcan más viviendas irregulares por más que se considere que en sectores más rurales (mas alejados de la capital regional) habrán más viviendas de ese tipo, sin embargo al existir en estos sectores una mayoría de terrenos con due los cuales no se consideran como viviendas irregulares por mas de que tengan una vivienda en condiciones precarias dentro de él.

Luego de hacer el recorte en la base de datos (donde se dejaron solo las regiones de Valparaíso, Metropolitana, Antofagasta y Tarapacá) los datos de los gráficos se hicieron más coherentes, pudiéndose obtener lo siguiente: Del gráfico 5 se logro obtener que a menor disponibilidad presupuestaria por comuna, mayor es la cantidad de viviendas irregulares, lo cual se puede deber a que los municipios pueden aportar con menos dinero a la construcción de viviendas.

Del gráfico 6 la cantidad de viviendas irregulares se ha de mantener aunque aumente el índice de pobreza esto se puede deber a que como la encuesta es CASEN no todo los habitantes de Chile han de responderla, provocando un error en los datos. Luego de esto en el gráfico 7 nuevamente ocurre lo mismo que en el gráfico 3, obteniéndose los mismos resultados. Finalmente, del gráfico número 8 se desprende lo mismo que en el gráfico número 4, obteniéndose las mismas conclusiones a pesar del recorte de base de datos hecho.

Parte 4:

A continuación, se procede a realizar regresiones lineales que permitan hacer inferencias con la hipótesis planteada.

Se hicieron 5 regresiones en total, donde en cada una se iba agregando un indicador más para ver cómo va cambiando el comportamiento del índice de pobreza CASEN.

Regresión 1: Viviendas irregulares por comuna vs. índice de pobreza CASEN.

El coeficiente del índice de pobreza CASEN es -0.18331 ($p = 0.0606$), lo que sugiere una asociación negativa no significativa entre el índice de pobreza y la cantidad de viviendas irregulares por comuna.

Regresión 2: Viviendas irregulares por comuna vs. índice de pobreza CASEN + Disponibilidad presupuestaria municipal por habitante.

Al agregar la variable de disponibilidad presupuestaria municipal por habitante, el coeficiente del índice de pobreza CASEN disminuye a -0.200232 ($p = 0.0451$), lo que indica una asociación negativa significativa entre el índice de pobreza y la cantidad de viviendas irregulares por comuna.

Regresión 3: Viviendas irregulares por comuna vs. índice de pobreza CASEN + % de población rural.

Al introducir la interacción entre el índice de pobreza CASEN y el porcentaje de población rural, el coeficiente del índice de pobreza CASEN se reduce aún más a -0.147459 ($p = 0.220$), aunque sigue siendo no significativo. La interacción en sí misma tampoco es significativa.

Regresión 4: Viviendas irregulares por comuna vs. índice de pobreza CASEN + % de población rural + Distancia con respecto a la capital regional.

Agregando la distancia con respecto a la capital regional, el coeficiente del índice de pobreza CASEN disminuye aún más a -0.139705 ($p = 0.251$), y sigue siendo no significativo. Ninguna de las variables adicionales es significativa en esta regresión.

Regresión 5: Viviendas irregulares por comuna vs. índice de pobreza CASEN + % de población rural + distancia con respecto a la capital regional + Superficie total comunal en km².

Al incluir también la superficie total comunal en kilómetros cuadrados, el coeficiente del índice de pobreza CASEN cambia a -0.07853 ($p = 0.5151$), y sigue siendo no significativo. Sin embargo, la superficie total comunal en km² es significativa ($p = 0.0476$).

Conclusión:

En general, el efecto del índice de pobreza CASEN en la cantidad de viviendas irregulares por comuna parece disminuir a medida que se agregan más variables explicativas al modelo. Sin embargo, ninguno de los modelos muestra una asociación significativa entre el índice de pobreza y la cantidad de viviendas irregulares por comuna. Otros factores podrían estar influyendo más en esta relación, es decir, puede haber variables o indicadores adicionales que no están incluidos en el análisis y que podrían tener un impacto significativo en la cantidad de viviendas irregulares por comuna.

Supuestos estadísticos:

Se procede a ver si se cumplen los supuestos de homocedasticidad y multicolinealidad respecto a la quinta regresión lineal, ya que contiene todos los indicadores. Sobre la homocedasticidad se calculó usando el test de Breusch-Pagan, y el valor p asociado es 0.195, que es mayor que 0.05, por lo que no hay suficiente evidencia para rechazar la hipótesis nula de homocedasticidad. Sobre la Multicolinealidad, según los resultados del VIF dicen que la variable Porcentaje de población rural y la interacción “Índice de pobreza CASEN” con “Porcentaje de población rural” tienen VIFs altos, lo que indica la presencia de multicolinealidad significativa. Cabe destacar que la razón de que los gráficos no cumplan lo esperado se puede deber a que las regresiones y los datos no cumplen con todos los supuestos estadísticos.

Parte 5:

Con base en los análisis realizados, la investigación arrojó una serie de conclusiones significativas respecto a la relación entre la cantidad de viviendas irregulares catastradas y diversos indicadores socioeconómicos en las comunas de Chile. En primer lugar, se confirmó la hipótesis inicial de que existe una asociación entre el nivel de pobreza de los habitantes de una comuna y la cantidad de viviendas irregulares presentes en ella. Sin embargo, este vínculo no se reveló como significativo en todos los modelos analizados, sugiriendo la influencia de otros factores no considerados en el estudio.

Es importante destacar que, si bien la disponibilidad presupuestaria municipal por habitante mostró una correlación inesperada con la cantidad de viviendas irregulares, este hallazgo puede explicarse por las diferencias en las prioridades de asignación de recursos en cada comuna. Por otro lado, la falta de una relación clara entre la superficie total comunal y la cantidad de viviendas irregulares puede atribuirse a la diversidad geográfica del país y la concentración de estas viviendas en áreas urbanas más densamente pobladas.

Asimismo, se observó que a menor disponibilidad presupuestaria por comuna, se registró una mayor cantidad de viviendas irregulares, lo que sugiere limitaciones en la capacidad de los municipios para financiar la construcción de viviendas adecuadas. Por último, aunque el índice de pobreza CASEN no mostró una asociación significativa con la cantidad de viviendas irregulares por comuna en todos los modelos, la inclusión de variables adicionales demostró que otros factores podrían estar influyendo en esta relación.

En conclusión, los hallazgos subrayan la complejidad de los determinantes de la informalidad habitacional en Chile y la necesidad de considerar una amplia gama de factores socioeconómicos y geográficos para comprender completamente este fenómeno. Se destacan posibles áreas de investigación futura, como la exploración de variables adicionales que puedan influir en la cantidad de viviendas irregulares, así como el desarrollo de estrategias de política pública más efectivas para abordar este problema en las comunas del país.

Anexo

Gráfico 1:

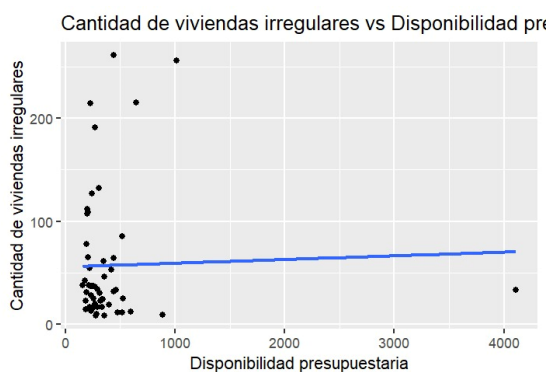


Gráfico 2:

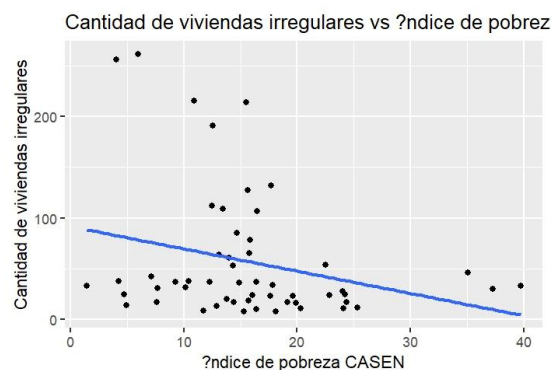


Gráfico 3:

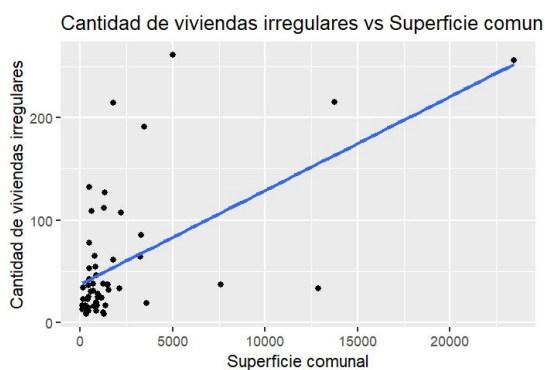


Gráfico 4:

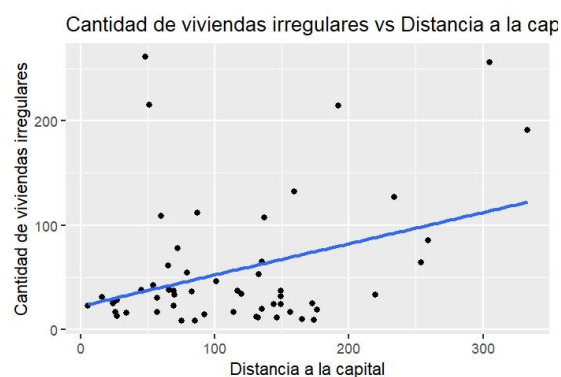


Gráfico 5:

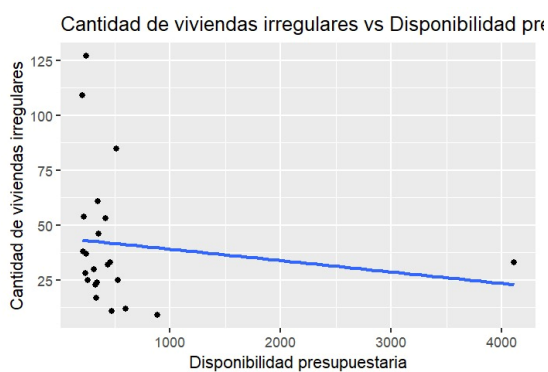


Gráfico 6:

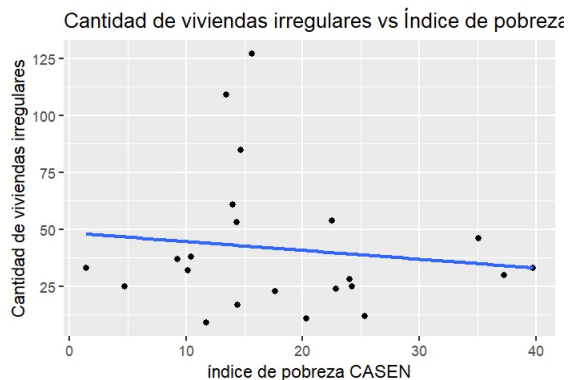


Gráfico 7:

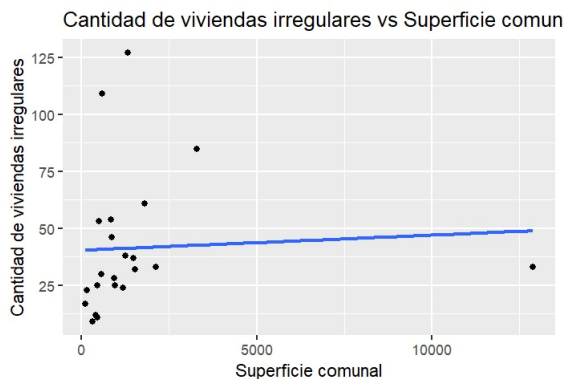


Gráfico 8:

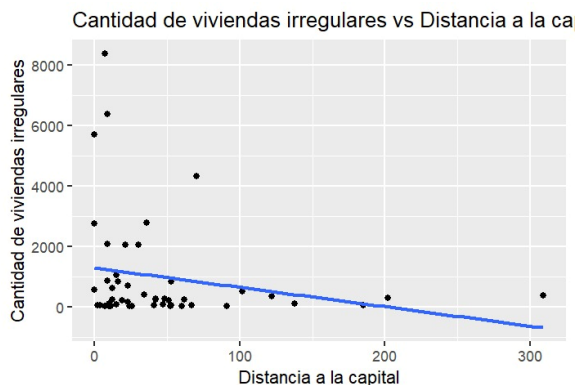


Ilustración 1:

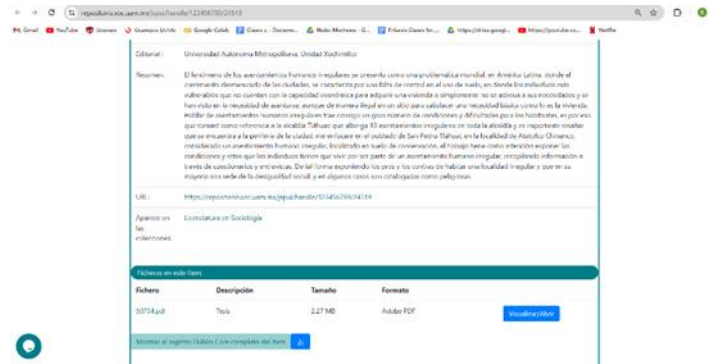


Ilustración 2:

Identificación de asentamientos irregulares y diagnóstico de sus necesidades de infraestructura en Ciudad Juárez...

de propiedad de la tierra; sin éste, se considera esa vivienda como AI sin importar los demás aspectos que maneja la ONU. La metodología planteada en esta investigación sugiere su replicación en diferentes ciudades, por lo cual se utiliza la definición de la UN-HABITAT de AI.

MARCO TEÓRICO

En primer lugar se tienen que tomar en cuenta los

Las bases teóricas de la marginalidad urbana y cultura de la pobreza indican elementos socioeconómicos que caracterizan a los AI. El Consejo Nacional de Población (CONAPO) identifica a la marginalidad urbana como un problema social asociado con las privaciones y la carencia de oportunidades de la población a los servicios como la educación, salud e ingreso; asimismo, a la dificultad para el acceso a los servicios básicos urbanos como agua potable, drenaje y electricidad (Peña, 2003). De acuerdo con Peña (2003), estas carencias o condiciones de