RRD

Wilmer Rojas & William Aguirre & John Esteban Londoño

2024-10-15

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'ggplot2' was built under R version 4.2.3

## Warning: package 'tibble' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## Warning: package 'readr' was built under R version 4.2.3

## Warning: package 'purrr' was built under R version 4.2.3

## Warning: package 'dplyr' was built under R version 4.2.3

## Warning: package 'stringr' was built under R version 4.2.3

## Warning: package 'forcats' was built under R version 4.2.3

## Warning: package 'lubridate' was built under R version 4.2.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(rddensity)  
library(rdrobust)

## Warning: package 'rdrobust' was built under R version 4.2.3

load("estratos\_2y3\_rdd.RData")

# Introducción

El propósito de este trabajo es estimar los efectos de la estratificación socioeconómica en la ciudad de Medellín, enfocándose específicamente en cómo dicha estratificación afecta el valor de las rentas y otros indicadores de calidad del hábitat en los diferentes barrios. La estratificación socioeconómica es una característica única del sistema colombiano, que asigna un estrato de 1 a 6 a las viviendas según el nivel socioeconómico del área en la que se encuentran, y es utilizado para definir subsidios y tarifas de servicios públicos. Este sistema, aunque útil en muchos casos, también puede tener efectos no deseados sobre los precios de la vivienda, la distribución de la riqueza y el acceso a servicios públicos y oportunidades de empleo en distintos sectores de la ciudad.

En particular, este estudio tiene como objetivo analizar cómo el hecho de que una vivienda se encuentre en un estrato determinado puede influir en el valor de su renta, así como en otros factores relacionados con la calidad del hábitat, como la densidad poblacional, la accesibilidad a servicios comerciales y la proporción de migrantes en el barrio. A través de un enfoque empírico, utilizando datos sobre los estratos 2 y 3 en Medellín, se buscará comprender mejor las implicaciones económicas y sociales de esta categorización, para eventualmente proponer recomendaciones de política pública que puedan mitigar cualquier impacto negativo que se detecte.

# Analisis exploratorio de los datos

summary(estratos\_2y3\_rdd$valor\_m2)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5220 450015 598890 764462 797218 10774862

El resumen estadístico de la variable valor\_m2 muestra una considerable dispersión en los valores de renta por metro cuadrado en los estratos 2 y 3 de Medellín. El valor mínimo es de 5,220, mientras que el valor máximo alcanza los 10,774,862, lo que refleja una gran variabilidad en los precios del suelo. El valor promedio (media) es de 764,462, mientras que la mediana, que representa el punto medio de los datos, es de 598,890, lo que sugiere que la distribución está sesgada hacia valores más altos, probablemente debido a algunos valores atípicos en el extremo superior. Además, el primer cuartil (450,015) y el tercer cuartil (797,218) indican que el 50% central de los valores se encuentra dentro de este rango.

estratos\_2y3\_rdd %>%  
 group\_by(estrato) %>%  
 summarise(  
 Min = min(valor\_m2, na.rm = TRUE),  
 Q1 = quantile(valor\_m2, 0.25, na.rm = TRUE),  
 Median = median(valor\_m2, na.rm = TRUE),  
 Mean = mean(valor\_m2, na.rm = TRUE),  
 Q3 = quantile(valor\_m2, 0.75, na.rm = TRUE),  
 Max = max(valor\_m2, na.rm = TRUE)  
 )

## # A tibble: 2 × 7  
## estrato Min Q1 Median Mean Q3 Max  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2 5220. 408894. 495030. 508415. 598890. 5362617.  
## 2 3 15419. 658780. 818916. 1129829. 1322208. 10774862.

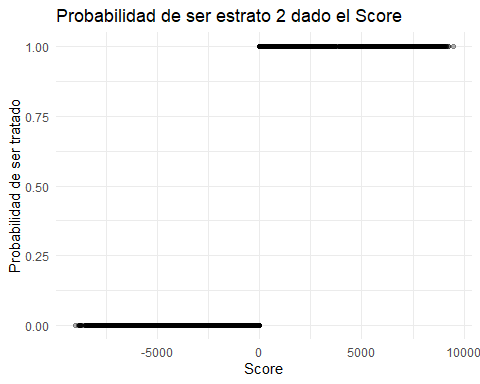
El resumen estadístico para la variable valor\_m2 en los estratos 2 y 3 muestra diferencias notables. En el estrato 2, el valor mínimo es de 5,220 y el máximo llega a 5,362,617, con una mediana de 495,030 y una media de 508,414.7, lo que indica una distribución de precios más baja y relativamente menos dispersa. En contraste, el estrato 3 presenta valores más altos, con un mínimo de 15,419 y un máximo de 10,774,862. La mediana en el estrato 3 es de 818,916.3 y la media es considerablemente más alta, con 1,129,828.8, lo que sugiere una mayor variabilidad en los precios por metro cuadrado, probablemente debido a la presencia de valores atípicos hacia el extremo superior. Estos resultados reflejan una clara diferencia en los valores de renta por metro cuadrado entre ambos estratos, con los precios siendo considerablemente más altos en el estrato 3.

estratos\_2y3\_rdd %>%  
 group\_by(estrato) %>%  
 summarise(  
 Min = min(viv\_informal, na.rm = TRUE),  
 Q1 = quantile(viv\_informal, 0.25, na.rm = TRUE),  
 Median = median(viv\_informal, na.rm = TRUE),  
 Mean = mean(viv\_informal, na.rm = TRUE),  
 Q3 = quantile(viv\_informal, 0.75, na.rm = TRUE),  
 Max = max(viv\_informal, na.rm = TRUE)  
 )

## # A tibble: 2 × 7  
## estrato Min Q1 Median Mean Q3 Max  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2 0 0 0.0278 0.0698 0.0667 1  
## 2 3 0 0.0102 0.0317 0.0527 0.0625 1

# Comprobación si es nitido

ggplot(estratos\_2y3\_rdd, aes(x = score, y =treat )) +  
 geom\_point(alpha = 0.3) +   
 labs(title = "Probabilidad de ser estrato 2 dado el Score",  
 x = "Score", y = "Probabilidad de ser tratado") +  
 theme\_minimal()



El tipo de regresión discontinua es nitida dado que la probabilidad de recibir subsidio depende del score (distancia a la frontera), esta probabilidad cambia abruptamente cuando la distancia a la frontera es mayor o igual a cero, las manzanas son clasificadas como estrato 2 y reciben subsidio.

# test de MaCrary

mccrary\_test <- rddensity(estratos\_2y3\_rdd$score)  
summary(mccrary\_test)

##   
## Manipulation testing using local polynomial density estimation.  
##   
## Number of obs = 8106  
## Model = unrestricted  
## Kernel = triangular  
## BW method = estimated  
## VCE method = jackknife  
##   
## c = 0 Left of c Right of c   
## Number of obs 3318 4788   
## Eff. Number of obs 550 1380   
## Order est. (p) 2 2   
## Order bias (q) 3 3   
## BW est. (h) 945.461 1267.751   
##   
## Method T P > |T|   
## Robust -1.3321 0.1828

## Warning in summary.CJMrddensity(mccrary\_test): There are repeated observations.  
## Point estimates and standard errors have been adjusted. Use option  
## massPoints=FALSE to suppress this feature.

##   
## P-values of binomial tests (H0: p=0.5).  
##   
## Window Length / 2 <c >=c P>|T|  
## 8.293 20 69 0.0000  
## 16.585 42 87 0.0001  
## 24.878 51 96 0.0003  
## 33.170 54 104 0.0001  
## 41.463 57 107 0.0001  
## 49.755 57 108 0.0001  
## 58.048 58 115 0.0000  
## 66.341 61 120 0.0000  
## 74.633 69 123 0.0001  
## 82.926 73 128 0.0001

El test arroja resultados basados en un análisis de discontinuidad alrededor de un punto de corte (c = 0). Se analizaron 8,106 observaciones, divididas en 3,318 a la izquierda del punto de corte y 4,788 a la derecha. El número efectivo de observaciones utilizadas para la estimación fue menor: 550 a la izquierda y 1,380 a la derecha, debido a la técnica de suavizado. El método aplicado fue un kernel triangular, y el ancho de banda estimado (h) fue de 945.461 y 1267.751 para cada lado del punto de corte.

El test robusto arrojó un valor T de -1.3321 con un valor p de 0.1828, lo que indica que no se encontró evidencia suficiente para rechazar la hipótesis nula de que no hay manipulación en la densidad alrededor del punto de corte. En otras palabras, los resultados sugieren que no hubo un cambio abrupto en la distribución de las observaciones cerca del umbral, lo que fortalece la validez del diseño de regresión discontinua.

# Efecto causal

rd\_result <- rdrobust(  
 y = estratos\_2y3\_rdd$valor\_m2,   
 x = estratos\_2y3\_rdd$score   
)  
  
summary(rd\_result)

## Sharp RD estimates using local polynomial regression.  
##   
## Number of Obs. 8106  
## BW type mserd  
## Kernel Triangular  
## VCE method NN  
##   
## Number of Obs. 3318 4788  
## Eff. Number of Obs. 365 660  
## Order est. (p) 1 1  
## Order bias (q) 2 2  
## BW est. (h) 585.635 585.635  
## BW bias (b) 1229.522 1229.522  
## rho (h/b) 0.476 0.476  
## Unique Obs. 3318 4743  
##   
## =============================================================================  
## Method Coef. Std. Err. z P>|z| [ 95% C.I. ]   
## =============================================================================  
## Conventional-119964.830 14924.696 -8.038 0.000[-149216.697 , -90712.963]  
## Robust - - -7.123 0.000[-144790.036 , -82302.271]  
## =============================================================================

El coeficiente estimado indica la magnitud del cambio en el valor del metro cuadrado cuando se cruza el umbral entre los estratos. Si este coeficiente es negativo, sugiere que a medida que los hogares se acercan al estrato 2, el valor de la vivienda tiende a disminuir, lo que puede reflejar la percepción de menor calidad del hábitat en estratos más bajos o la presencia de desigualdades económicas en Medellín. La significancia estadística de los resultados, indicada por un valor p bajo, refuerza la idea de que la distancia a la frontera estratificada tiene un impacto real en el valor de la propiedad.

# Recomendación de política

Se recomienda implementar políticas económicas que incluyan inversiones en infraestructura y servicios públicos en los barrios de estrato 2 de Medellín, ya que el análisis revela un efecto negativo en el valor del metro cuadrado al cruzar el umbral entre los estratos. Estas políticas podrían incluir subsidios para la renovación de viviendas, desarrollo de espacios comunitarios, programas de capacitación laboral y incentivos fiscales para empresas que inviertan en la zona, con el objetivo de mejorar la calidad del hábitat, reducir desigualdades económicas y fomentar un desarrollo urbano más equitativo y sostenible.