

Descripción de la generación de empleo en la ciudad de Harare utilizando estadística espacial

Óscar A.¹

Universidad de Costa Rica

Resumen

El siguiente documento se centra en analizar la relación de los años de educación, la cantidad de población, la distancia que se recorre de cada distrito al punto con mayor generación de empleo y el área en kilómetros cuadrados sobre la generación de empleo al año 2002 con datos del censo aplicado en Zimbabwe. Para evaluar la relación entre las variables se utilizan tres modelos de regresión espacial: regresión ponderada geográficamente (GWR), auto regresión simultánea (SAR) y auto regresión condicional (CAR), así mismo se compara con los resultados obtenidos con un modelo de mínimos cuadrados ordinarios. Los resultados obtenidos con los modelos SAR y CAR son muy similares, estos resultados pueden emplearse para plantearse mejores estrategias de generación de empleo considerando el desplazamiento que tiene que darse a puntos de mayor generación de empleo.

Palabras clave: Estadística Espacial, Regresión, Empleo, Distancias.

Introducción

El siguiente proyecto se centrará en el análisis de datos a nivel espacial relacionados al país Zimbabwe, propiamente a la Provincia de Harare, utilizando datos del censo 2002.

En cuanto a Zimbabwe, este país está situado en el sureste del continente africano. Este país carece de costas oceánicas, limita al oeste con Botsuana, al norte con Zambia, al sur con Sudáfrica y al este

con Mozambique. Sus territorios corresponden a la antigua Rodesia del Sur, además, su idioma oficial es el inglés y el shona.

Como dato interesante Zimbabwe es el país al año 2010 con el menor IDH, su nivel es de 0.14, sin embargo, gracias a reformas que se han aplicado en los últimos años le ha permitido llegar a tener un IDH igual a 0.563 evidencia una mejora en términos generales de las condiciones sociales del país.

A partir de la información anterior es que resulta de interés analizar si existen factores que incidan en la generación de empleo, incorporando las principales variables citadas en la literatura económica. Al ser un país con alta desigualdad, el estudio se centra en la provincia de Harare, la cual concentra el 15 por ciento del total del empleo con los niveles más altos de educación en años de estudio terminados promedio, además, de ser la provincia con mayor población de todo el país.

En este estudio el problema planteado es poder observar la relación el empleo generado en función de los años de escolaridad, población total, distancia lineal de cada distrito con respecto al distrito de mayor generación de empleo. Lo anterior se desea estudiar por medio de análisis de regresión espacial utilizando tres diferentes modelos GWR, SAR y CAR.

Marco Teórico

Diversos estudios empíricos han mostrado el desempeño del mercado laboral a través del análisis de la relación entre educación y empleo, además del comportamiento de los salarios, donde se confirma una asociación positiva entre los años de estudio y el ingreso percibido (Carnoy, 1967, Mincer, 1974) Ante lo anterior, se plantea la importancia de comprender el efecto de los años de educación en un país tan subdesarrollado como Zimbabwe, tomando de referencia una de sus provincias con mayor nivel educativo.

Además, al tomar en cuenta la cantidad de población, se considera que se cuenta con mayor capital humano para poder generar mayores volúmenes de trabajo, lo cual, según López, en el corto plazo por diferentes teorías del capital humano se puede explicar cómo los cambios tecnológicos, la asignación de competencia y la falta

de experiencia pueden generar factores de desajuste en la relación educación – empleo, sin embargo, este mismo autor explica que en el largo plazo a relación se ajusta, evidenciando un efecto positivo (2009. Pag:864).

En algunos países, establecer una relación positiva entre empleo en función de educación y tamaño poblacional (Como disponible de capital humano) puede resultar complicado por la heterogeneidad de la estructura del mercado laboral que se analiza, la diversidad de puestos requiere diferentes niveles de educación o un grado técnico. Esta heterogeneidad en el capital humano y los requerimientos del mercado laboral se refleja en las oportunidades de empleo para personas alejadas a la zona específica donde se requiere el personal, según McConell y Brue, los tralados de personas aumentan la disponibilidad de recurso humano dentro de un área que le permita desplazarse (1997).

Si bien es cierto, el impacto de la distancia se debe medir en función de la disponibilidad de transporte, ya que como indica Figueroa, las grandes ciudades ganan mayor cantidad de trabajadores en función de la accesibilidad y fluidez del transporte (2005). Ante lo anterior, dentro del análisis de regresión geoespacial incorporar la variable distancia nos permitiría tener una idea del desplazamiento que se debe realizar, la deficiencia en el estudio es no contar con variables como acceso a transporte público o privado como una variable de interés para efecto del estudio realizado.

Metodología

Los datos utilizados se toman del censo del año 2002 aplicado a la población de Zimbabue. Se realiza una primera etapa exploratoria en todo el país de donde se selecciona la provincia con mayor generación de empleo, además, se considera de importancia el nivel educativo promedio de la región, el cual se esperaba que fuera alto. Una vez identificada la provincia con mayor nivel educativo y mayor generación de empleo, se procede a realizar la aplicación de las regresiones espaciales, las que se proceden a describir.

Regresión ponderada geográficamente (GWR)

La técnica GWR extiende la regresión global convencional al agregar un parámetro de ubicación geográfica, con el modelo reescrito como:

$$y_i = \beta_0(\mu_i, v_i) + \sum_k \beta_k(\mu_i, v_i)x_{ik} + \epsilon_i \quad (1)$$

donde y_i, x_{ik}, ϵ_i representan la variable de respuesta (dependiente), las variables predictores (independientes) y el término de error aleatorio en los puntos espaciales i . Por su parte, $\beta_0(\mu_i, v_i)$ es la intersección del modelo, y $\beta_k(\mu_i, v_i)$ es el coeficiente de la pendiente para la k -ésima variable predictiva en las coordenadas (μ_i, v_i) del i -ésimo punto. Los parámetros en la ecuación anterior podrían estimarse resolviendo la siguiente ecuación matricial:

$$\beta(\mu_i, v_i) = (X^t W((\mu_i, v_i) X^{-1}) X^t W(\mu_i, v_i) Y \quad (2)$$

Donde $\beta(\mu_i, v_i)$ representa la estimación insesgada de β , $W(\mu_i, v_i)$ es la matriz de ponderación, cuya función es garantizar que las observaciones cercanas a la ubicación específica tenga un mayor peso. En este estudio, se utiliza la siguiente función de kernel de ponderación gaussiana

$$W_{ij} = \exp\left(-\frac{d_{ij}^2}{b^2}\right) \quad (3)$$

donde d_{ij} es la distancia euclidiana entre el punto de regresión i y la observación vecina j , y b representa un ancho de banda de la función de núcleo (Shi et al., 2006; Li et al. 2010).

Autoregresivo Simultáneo (SAR)

El SAR asume que la respuesta en cada ubicación i es una función no solo de la variable explicativa en i , sino también de los valores de la respuesta en las ubicaciones vecinas j . La relación de vecindad se expresa formalmente en pesos espaciales (W) que representan una medida de la conexión (es decir, la distancia geográfica) entre las ubicaciones i y j (Anselin, 1988; Kissling and Carl, 2008). El modelo

SAR de respuesta retardada supone que el proceso autorregresivo solo se genera en la variable de respuesta:

$$Y = X\beta + \phi WY + \epsilon \quad (4)$$

donde ϕ es el parámetro autoregresivo, W es la matriz de pesos espaciales y β es un vector que representa las pendientes asociadas con los predictores en la matriz de predicción original (X).

Autoregresión Condicional (CAR)

El modelo CAR se puede escribir como:

$$Y = X\beta + \phi W(Y - X\beta) + \epsilon \quad (5)$$

con matriz de covarianza entre los residuos dada por:

$$C = X\beta + \alpha^2(I - \phi W)^{-1} \quad (6)$$

donde α^2 es la varianza de los residuos, e I es una matriz de identidad $n \times n$. C es la función solo del parámetro autorregresivo ϕ y la matriz de ponderaciones espaciales (W), donde W es simétrica, definida con una función inversa de las distancias geográficas en el estudio (Keitt et al. 2002; Kissling and Carl, 2008)

La I de Moran

La I de Moran es una herramienta útil para describir la dependencia de patrones espaciales o autocorrelación espacial de datos. El estudio utiliza la I de Moran para evaluar la autocorrelación espacial de los residuos del modelo. El índice I se calcula de la siguiente manera:

$$I = \frac{(1/WT) \sum_{j=1}^n \sum_{k=1}^n wt_{jk}(z_j - \bar{z})(z_k - \bar{z})}{(1/n) \sum_{j=1}^n (z_j - \bar{z})^2} \quad (7)$$

donde z_j y z_k denotan los valores de la variable observada en los sitios de j y k , respectivamente; wt_{jk} denota el peso de la variable. La matriz de peso WT ($n \times n$), es la suma de los pesos de wt_{jk} para una clase de distancia dada. La I de Moran es alta y positiva cuando

el valor es similar a los valores adyacentes, pero bajo cuando uno es diferente a los valores adyacentes. Un valor cero de la I de Moran indica un patrón espacial aleatorio (Moran, 1950).

Resultados

Cálculo de la Distancia

La primera parte de los resultados se enfoca en la complejidad de calcular la distancia dentro del polígono, en la figura 1, se puede observar como para determinar el punto de mayor empleo se utilizó la cantidad de trabajadores dentro de la provincia de Harare entre sus diferentes distritos.

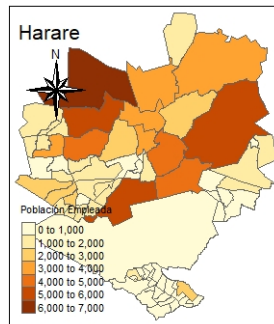


FIGURA 1: *Empleo por Distrito.*

Los distritos de más al norte de Harare son los que mayor concentración de empleo poseen, se determinó que la distancia se calcularía en función del distrito dentro de Harare que mayor concentración de personas empleadas posee. En la figura 2 se puede observar como la mayor distancia en recorrido es aproximadamente 35 kilómetros de distancia entre el punto más lejano y el sitio de mayor generación de empleo.

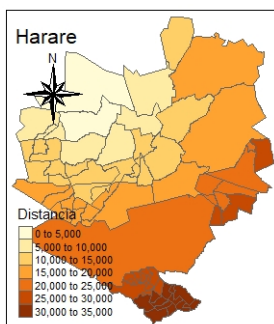


FIGURA 2: Distancia con respecto al distrito con mayor Empleo.

0.1. Desempeño de los modelos aplicados

Primero se evalúa la prueba de I de Moran, la cual se construye con una matriz de distancias basada en el criterio de la torre, el cual aplicado a los residuos del modelo OLS indica que hay evidencia para rechazar la hipótesis nula de independencia entre los distritos de la provincia de Harare.

Evaluando el desempeño de los modelos el modelo GWR presenta un menor valor en el Criterio de Información de Akaike y en comparación con el modelo OLS presenta un R^2 ajustado más alto, sin embargo, el modelo GWR presenta en la variable distancia una relación positiva, con lo que se estaría diciendo que a una mayor distancia mayor el empleo, lo cual no ocurre en el modelo SAR y CAR, ambos con coeficientes y criterios de información de Akaike muy similares.

Variable	OLS			SAR			CAR			GWR	
	Coef	S.E.	P(z)	Coef	S.E.	P(z)	Coef	S.E.	P(z)	Coef	S.E.
(Intercept)	-1,22	1,22	1,3E-13	-1,079	1,3809	5,33E-15	-1,079	1,342	8,88E-16	-1,1373	1,18
pop	5,37	4,64	2,00E-16	5,522	4,2432	2,20E-16	5,498	4,2735	2,20E-16	5,424	0,00448
years_scho	1,2	1,19	2,65E-15	1,173	1,1354	2,20E-16	1,1713	1,3187	2,20E-16	1,212	114,91
Distance	2,49	9,82	0,801	-1,302	1,3519	9,23E-01	-1,383	1,3008	9,16E-01	2,7983	0,00949
areaKM2	1,25	2,51	4,28E-06	7,2452	2,283	1,56E-03	8,1488	2,3741	5,98E-04	1,1988	2,4119
AIC	1167,67			1163,1			1163,4			1154,63	
R2*	0,86									0,872	

A partir de los resultados, se escogió el modelo GWR para representar los resultados de manera gráfica, se escogió por tener el menor

criterio de información. En la figura 3 se puede observar los resultados estimados de la variable empleo, la cual según el mapa sigue el mismo comportamiento en términos de su concentración dentro de Harare, sin embargo, la estimación subvalora la cantidad de empleo en cada distrito con respecto a los valores verdaderos.

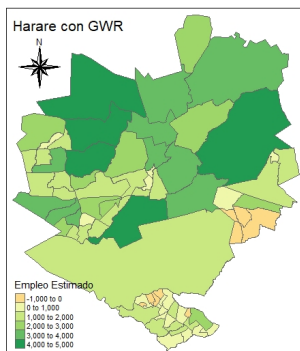


FIGURA 3: *Empleo estimado con el modelo GWR.*

Discusión de resultados

Si bien aplicar este tipo de regresiones espaciales para estimar el empleo, no es la finalidad del trabajo, por medio de esta regresiones se logró determinar la relación entre las variables educación, población, área del distrito con respecto a la cantidad de empleo por distrito satisfactoriamente, aplicando estadística espacial para lograrlo, sin embargo, se esperaba que la variable distancia, por su forma de calcularla presentara una relación negativa y en el mejor modelo identificado esto no ocurrió, lo anterior da pie a realizar ajustes pertinentes en futuras investigaciones, además de contemplar variables adicionales que a la fecha del presente estudio no se tenían a mano, por ejemplo cantidad de personas que se transportan en servicio público y empleo proveniente de otras zonas o distritos.

Conclusiones

- A partir del estudio realizado se pudo ver que las variables educación, población y área en metros cuadrados presentan

un relación positiva con respecto al empleo, sin embargo, la variable principal es la educación, la cual según la literatura se esperaría una relación positiva, misma que se obtuvo en los diferentes modelos aplicados, con lo cual se tiene un guía para toma de decisiones en función de mejorar el nivel educativo de la población de Harare pensando en que en el largo plazo esto permita generar más empleo.

- En la aplicación del modelo, la distancia no resulto significativamente distinta de cero cuando esta presenta una relación negativa con respecto al empleo, lo cual era deseable para efectos del estudio, además, no se tiene certeza si la forma de calcular la distancia entre los distritos y el distrito de mayor concentración de empleo pudo ser un factor en contra de los resultados esperados. Ante lo anterior, la distancia debe estar asociada a variables tiempo de viaje, acceso a transporte público, por ejemplo, que permitan concluir sobre el efecto de la lejanía con respecto a la generación de empleo..

Referencias

- Anselin L. 1988. Spatial econometrics: methods and models. Kluwer Academic Publishers: Dordrecht, Netherlands.
- BivandR, PebesmaE Gómez V. 2008. AppliedSpatialData AnalysiswithR. Springer, New York.
- Carnoy, Martin (1967). "Earnings and schooling in Mexico", Economic Development and Cultural Change, vol. 15 núm. 4, pp. 408-419.
- Keitt T et al. 2002. Accounting for spatial pattern when modelling organism environment interactions. *Ecography* 25: 616-625.
- Kissling W Carl G 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* 17(1): 59-71.

-
- López, M. (2009). Efectos de la correspondencia entre educación y empleo de los jóvenes en el mercado de trabajo mexicano. *Revista Mexicana de Investigación Educativa*. Vol. 14, Núm. 42, pp.863-887
 - McConnell, Campbell y Brue, Stanley (1997). *Economía laboral*, 7^a ed., Madrid: McGraw Hill.
 - Moran P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17-23.