

Concentración del monto en colones de ventas diarias de microempresas clientes de entidad financiera costarricense, Gran Área Metropolitana de Costa Rica.

Fiorella Laurito y Fabián Hernández

Abstract—El objetivo de este estudio es analizar la distribución del monto en colones de ventas diarias de microempresas del Grán Área Metropolitana, con el fin de enriquecer la estrategia de negocio de prospectación de nuevos clientes. Además, se partió de la hipótesis de que las ventas diarias de estas empresas están relacionadas a la cantidad de personas que las visitan y por ende a la cercanía a centros urbanos, por lo cual se calculó la distancia al centro de la provincia a la que pertenecen cada una de las empresas. Se obtuvo que la cercanía al centro urbano no es una buena variable para explicar el proceso estocástico de las ventas diarias de estas empresas, puesto que el modelo únicamente con el intercepto presentó mejores resultados. Por último, la interpolación realizada captura bien la distribución de las ventas en la GAM, sin embargo, al no estar usándose ninguna covariable, no es posible predecir valores de ventas para lugares con poca cantidad de datos, como la provincia de Cartago.

Palabras claves: Geoestadística, Estadística espacial, semi-variogram, banca y microempresas.

I. INTRODUCCIÓN

Costa Rica se divide en 482 distritos de los cuales 191 conforman la Gran Área Metropolitana (GAM), donde se concentra la mayoría de la población y el comercio del país.

Esta área corresponde aproximadamente al 19% del total del territorio costarricense, sin embargo, alberga al 54% de la población.

Esta concentración de personas ha hecho que en esta área se lleve acabo la mayor parte del comercio costarricense, por ende concentra la mayor cantidad de pequeñas empresas y micro empresas.

Además, esta misma concentración de personas hace que la GAM abarque los centros urbanos de las 4 provincias más pobladas de Costa Rica: San José, Alajuela, Heredia y Cartago, lo que ha causado que muchas empresas proliferen a las cercanías de estas locaciones.

Dado al auge en el comercio y proliferación de empresas en el GAM, la entidad financiera costarricense considerada en este estudio se le relevante analizar la concentración del monto en colones de ventas diarias de las microempresas a las cuales les otorga servicios financieros. Debido que, para prevenir la morosidad de la cartera, los bancos realizan exhaustivos estudios de crédito para asegurarse el cumplimiento en el pago de los préstamos, donde uno de los factores más importantes a considerar son los activos de la empresa.

En estadística espacial, el análisis geoestadístico usa muestras tomadas de diferentes puntos sobre una superficie y crea interpolaciones en una superficie continuas. [3]. Por lo tanto, obtener una interpolación del monto de ventas diarias en el

espacio, contribuye con una estrategia de negocio puesto que se tendrían indicios de posibles lugares donde prospectar nuevos clientes.

Dado lo anterior, el objetivo general de este estudio es analizar la distribución del monto en colones de ventas diarias de microempresas del Grán Área Metropolitana, con el fin de enriquecer la estrategia de negocio de prospectación de nuevos clientes.

Para este estudio se cuenta con la información de 1404 empresas en la GAM junto al reporte de sus ventas diarias y su locación(latitud y longitud). Por motivos de confidencialidad no se puede comentar explícitamente sobre el segmento de comercio al que pertenecen estas empresas, lo que se puede decir, es que se encuentran en el área de consumo humano, por lo que se tiene la hipótesis de que las ventas diarias de estas empresas están relacionadas a la cantidad de personas que las visitan y por ende a la cercanía a centros urbanos. En la Figura 1 se puede ver la distribución de estas empresas en la GAM.

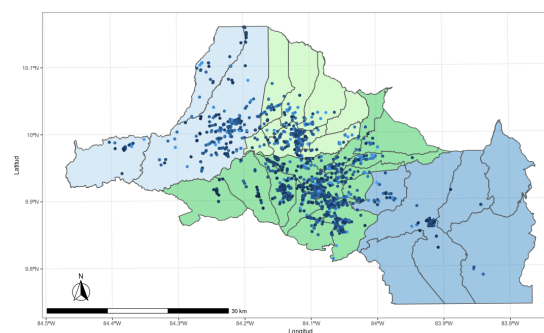


Fig. 1. Distribución de microempresas inscritas según provincia, Gran Área Metropolitana, Costa Rica

Para calcular la cercanía a centro urbano, se calculó la distancia al centro de la provincia a la que pertenecen cada una de las empresas. En la tabla I se pueden ver las latitudes y longitudes tomadas como centros de las provincias.

Provincia	Latitud	Longitud
San José	9.932662	-84.078623
Alajuela	10.015008	-84.078623
Heredia	9.998375	-84.117087
Cartago	9.864178	-83.919848

TABLE I

PUNTOS TOMADOS COMO CENTRO DE LAS PROVINCIAS.

II. RESULTADOS

A. Verificación de modelos

Antes de empezar el análisis geoestadístico como tal, se quiere ver la relación entre las ventas diarias de estas empresas y la distancia a los centros urbanos, en la figura Figura 2 se puede ver los resultados.

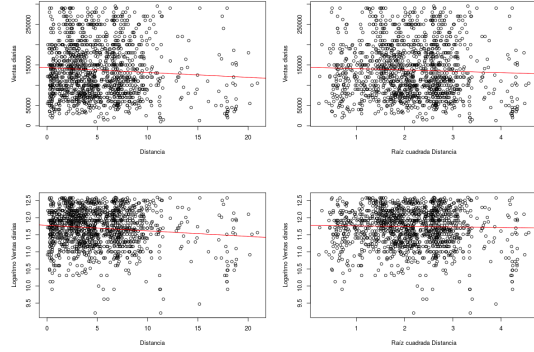


Fig. 2. Relación entre ventas diarias y distancia a centros urbanos.

Además de usar las ventas diarias y la distancia a centros urbanos, también se intentó realizar algunas transformaciones sobre estas variables debido a que la relación lineal no se veía tan claramente. Se le aplicó un logaritmo a las ventas diarias, así como una raíz cuadrada a la distancia, en la misma figura anterior se pueden apreciar estos resultados.

Ninguna de las variantes utilizadas presentan una relación lineal clara, sin embargo, para este trabajo decidimos utilizar la versión del modelo que relaciona el logaritmo de las ventas diarias con la distancia a los centros urbanos. En la figura 3 se puede ver la evaluación de supuestos par este modelo.

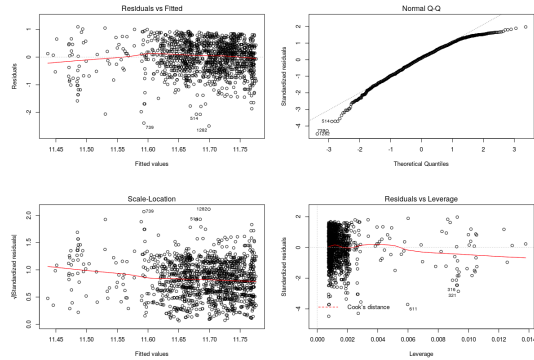


Fig. 3. Revisión de supuestos para el modelo $\log(\text{ventas})$ distancia.

A nivel de supuestos parece no haber una distribución óptima de los residuos, además de que en el gráfico QQ-plot, se puede apreciar que los residuos estandarizados no siguen la línea teórica de normalidad, saliéndose a los extremos, sin embargo, se decidió seguir con el análisis.

Para finalizar con esta sección, se realizó un ANOVA entre este modelo con la covariable de distancia y el modelo sin covariables, dando un p-value de 0.0005 lo que con una significancia del 5% parece ser que agregando la covariable al modelo no hay una mejora importante en el ajuste del

mismo. Al considerarse que el modelo con la covariable de distancia no agrega mejor ajuste que el modelo solo con el intercepto, se decidió realizar el resto del análisis con este modelo solo con el intercepto.

B. Ajuste de variogramas

Primero, para corroborar el supuesto de Isotropía (supuesto que dice que hay independencia de dirección, es decir, que el rezago h se puede expresar como su norma $||h||$ (sin importar la dirección), se ajustó el variograma del modelo con el intercepto para las direcciones de norte, sur, este y oeste, en la Figura 4 se pueden ver los resultados.

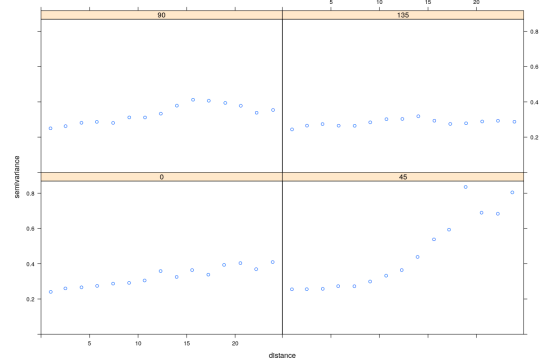


Fig. 4. Variogramas por múltiples direcciones.

Se puede apreciar que los gráficos de la semivariancia contra el rango para las 4 direcciones consultadas parecen ser bastante similares. El único que parece ser un poco diferente es el gráfico de la esquina inferior derecha, donde se puede apreciar que el valor de umbral es más alto que los demás, sin embargo, al considerarse bastantes similares, se decidió continuar con el análisis sin tomar en cuenta la dirección.

El semivariograma sin tomar en cuenta la dirección se puede ver en la figura 5.

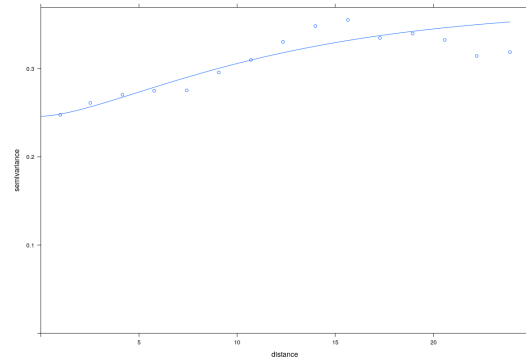


Fig. 5. Variograma sin tomar en cuenta dirección.

Sobre este mismo variograma se ajustó el modelo teórico, considerándose en este caso que el modelo Bes, era el que presentaba mejor ajuste visual. El valor del nugget para este modelo es 0.24 mientras que el rango de 0.36.

Usando el criterio del radio de $\frac{nugget}{rango} (\frac{0.24}{0.36} = 0.66)$ que expresa la dependencia espacial entre los elementos, donde:

- Menos de 25% representa relación espacial fuerte.
- 25% - 75% representa relación espacial moderada.
- Mayor a 75% representa relación espacial débil.

Se puede decir que para este caso no hay relación espacial demasiado marcada entre los elementos analizados [1], [2].

C. Ajuste del modelo

Para ver el ajuste del modelo teórico seleccionado, se procedió a seleccionar de manera aleatoria 100 empresas como valores de pruebas y usando los restantes elementos, se ajustó el modelo teórico Bes. Con este modelo ajustado, se procedió a predecir las ventas para los 100 elementos de prueba y calcular el error cuadrático (RMSE) medio de la predicción y los valores reales. El RMSE dió un valor de 0.64 considerándose un ajuste no tan malo.

D. Interpolación de ventas

Finalmente, usando el modelo con el intercepto se realizó la interpolación de las ventas diarias para toda la GAM. En la Figura 6 se pueden ver los resultados.

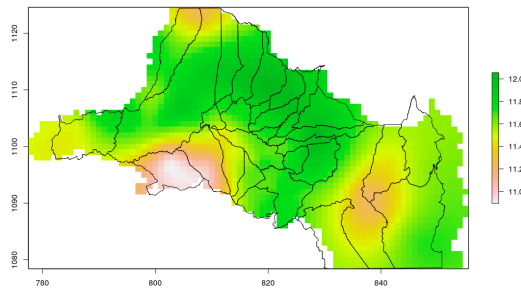


Fig. 6. Interpolación de ventas para la GAM.

Se puede notar que la mayoría de ventas de las empresas se concentran hacia el centro del país, principalmente entre las cabeceras de las provincias de San José, Alajuela y Heredia. El caso de Cartago llama un poco la atención, ya que al tenerse poca muestra para esta zona, pareciera que en el mapa representa una zona de poca concentración en ventas diarias.

III. CONCLUSIONES

- Parece ser que la cercanía al centro urbano no es una buena variable para explicar el proceso estocástico de las ventas diarias de estas empresas, otras variables, como tiempo de permanencia en el mercado o número de empleados podrían revisarse para ver si se logra un mejor grado de explicación.
- Las transformaciones escogidas para las variables utilizadas no permitieron solucionar los problemas de linealidad, la escogencia de otro tipo de transformaciones tal vez puede mejorar estos resultados.
- El valor bajo del RMSE da indicios de que el ajuste del modelo no es tan malo, sin embargo, al no estar tomándose en cuenta ninguna covariable y al tener R^2 del modelo un valor tan bajo (inferior a 10), definitivamente este modelo no es adecuado y nuevas

covariables tienen que investigarse para mejorar estos resultados.

- La interpolación realizada captura bien la distribución de las ventas en la GAM, sin embargo, al no estar usándose ninguna covariable, no es posible predecir valores de ventas para lugares con poca muestra como lo es la provincia de Cartago.

REFERENCES

- [1] Guo HY Houn KH Chien YL, Lee DY. Geostatistical analysis of soil properties of mid-west taiwan soils. *Soil Sc*, (162):291–297, 1997.
- [2] Corbí JMG Matín JAR, Arias ML. Heavy metals contents in agricultural topsoils in the ebro basin (spain). application of the multivariate geostatistical methods to study spatial variations. *Environ Pollut*, (144):1001–1012, 2006.
- [3] Dr. Michael P. Peterson. The principles of geostatistical analysis, 2006.