Ancianos

Ferran Garcia

2024-01-24

Contents

1 Introducción

1.1 Contexto

El envejecimiento demográfico emerge como un fenómeno geográfico crítico que podría acarrear serias implicaciones macroeconómicas estructurales en nuestra sociedad. En consecuencia, resulta imperativo abordar de manera inmediata políticas orientadas a mejorar la calidad de vida de la población mayor, fundamentadas en estudios que delineen su situación espacial y social, particularmente en aquellos entornos urbanos donde su presencia es más significativa.

1.2 Objetivos

En el actual escenario, nos enfrentamos al desafío de localizar áreas urbanas en el municipio de Sevilla que compartan una estructura demográfica similar, especialmente en lo que respecta a la población mayor. Para abordar este objetivo, emplearemos distintos métodos de análisis estadístico multivariante y espacial. La meta es desarrollar servicios sociales y de asistencia que tomen en cuenta las características específicas de los ancianos y su distribución geográfica en la ciudad. En este contexto, nuestro enfoque consiste en identificar segmentos homogéneos de la población anciana en áreas urbanas. La ejecución de este proceso nos permitirá diseñar servicios sociales y de asistencia que se adapten a las necesidades particulares de los ancianos, considerando tanto su tipología como su ubicación en la ciudad.

1.3 Estructura

2 Metodología

2.1 Técnicas

2.2 Herramientas

3 Reducción dimensional

Disponemos de muchas variables, eso motiva las técnicas de reducción dimensional. Usaremos análisis factorial de componentes pricipales, la idea es sintetizar todas las medidas disponibles en variables latentes. Lo haremos aprovechando la correlación que comparten. Para ello seleccionamos las variables que incluiremos a partir de un breve análisis exploratorio. Luego comprobaremos si es pertinente usar los procedimientos en cuestión en el conjunto de datos resultante. Entonces realizaremos varios ajustes con diferente número de factores hasta alcanzar la descomposición más satisfactoria en términos de representabilidad de las variables iniciales. Finalmente aplicaremos técnicas de rotación con tal de lograr factores ortogonales que sean interpretables. Con éstos extraeremos las puntuaciones factoriales que más adelante usaremos para crear clústers.

3.1 Selección de variables y adecuación de los datos

La selección de variables sirve para subsanar potenciales problemas y afinar en el cálculo de factores. Es importante tanto para que se pueda realizar el análisis factorial como para que éste sea eficiente y se ajuste un modelo robusto con interpretación relativamente simple evitando sobreajuste.

Por ejemplo, si nos fijamos en las variables demográficas veremos que se incluyen múltiples medidas de población quedando algunas determinadas completamente por el resto. En cuanto a la *situación laboral* tenemos población activa, ocupada, inactiva y parados.Entonces el total queda determinado por otra variable, ya sea la población total o u subgrupo de ésta y por tanto son linealmente dependientes. Por tanto la matriz de correlaciones no será inverible y en consecuencia no se podrá realizar el análisis factorial. Lo mismo ocurre con algunas variables del estado de las viviendas. En el annexo se muestra una lista completa de las variables que han sido excluidas por este motivo.

Por otro lado es importante que las variables seleccionadas estén correlacionadas entre sí. Para comprobarlo representamos gráficamente la correlación entre las variables seleccionadas y recogemos en una tabla estadísticos relevantes para la selección de variables como:

- Estimación inicial de comunalidades[1]
- Suma de correlaciones absolutas
- Número de variables no correlacionadas[2]

Las variables que se muestran en la tabla quedan excluidas del análisis factorial.

Table 1: Correlaciones de las variables seleccionadas, se muestran las 15 con menor comunalidad

	Comunalidad	Suma	p.val
CON1HABI	0.5205	6.88	16
AP	0.4817	9.65	28
DEL81AL90	0.3475	4.44	7
CON5OMAS	0.2745	4.26	8
C	0.2607	7.85	25
EA	0.2160	6.15	27
TRAEVEN	0.2101	4.25	2
AFAM	0.1894	5.05	7
COOP	0.1453	2.61	1

[1]Correlación múltiple al cuadrado
[2]Test de correlación de Pearson, nivel de significación del 5%

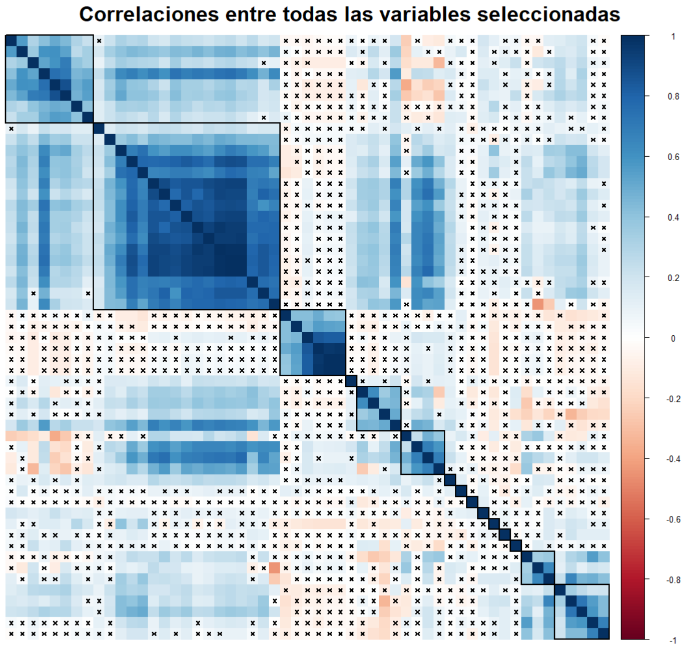

Figure 1: Correlaciones entre todas las variables.

Veamos ahora si el conjunto de datos seleccionados resulta apropiado para el análisis factorial o si debemos seleccionar variables con un criterio más estricto. Para ello realizamos varios tests:

Esfericidad de Barlett:

- Hipóteis nula: La matriz de correlaciones es la identidad (no existe ninguna correlación)
- Bajo H_0 el estadístico de contraste sigue una χ^2, asintóticamente
- Disponemos de más de 400 observaciones, se sostiene la suposición asintótica

```
cortest.bartlett(R1, n = nrow(X1))[["p.value"]] # podemos rechazar H_0
```

```
## [1] 0
```

Kaiser, Meyer, Olkin:

- Medida de adecuación de la muestra para un análisis factorial
- Compara la suma de cuadrados de la "imagen" de la matriz de correlaciones con la original
- Está acotada entre 0 y 1, según Kaiser:
  - Valores mayores a 0,9 son maravillosos
  - Valores mayores a 0,8 son meritorios
  - Valores mayores a 0,7 son medios
  - Valores mayores a 0,6 son mediocres
  - Valores mayores a 0,5 son miserables
  - Valores menores a 0,5 son inaceptables

```
KMO(R1)[[1]] # según Kaiser, la adecuación de la muestra es meritoria
```

```
## [1] 0.8141468
```

3.2 Número, rotación y puntuaciones factoriales

Ahora nuestra intención es identificar una estructura latente dentro del conjunto de datos. Se focaliza el interés en los factores capaces de explicar una proporción significativa de la variabilidad presente en los datos. Para determinar cuántos factores debemos retener, empleamos dos criterios. En primer lugar, se puede dibujar un scree plot para evaluar los *eigenvalues* de los factores estimados. Además, aplicamos la Regla de Kaiser, que sugiere retener aquellos factores cuyos *eigenvalues* superen la unidad. Otra estrategia consiste en seleccionar tantos factores como sean necesarios para explicar alrededor de un 70% de la variabilidad de los datos.

Regla de Kaiser:

```
sum(eigen(R1)[["values"]] > 1) # según el criterio de Kaiser, usaremos 8 factores
```

```
## [1] 8
```

Proporción de la varianza explicada:

4 Clústers

4.1 Método jerárquico

4.2 Método de k-medias

5 Annexo

5.1 Variables excluidas

5.1.1 Variables demográficas

- Categoría de edad: 60-64 años
- Sexo: Población masculina
- Procedencia: Provincial
- Nivel de estudios: Segundo grado
- Situación laboral: Activos
- Situación de actividad: Ocupados
- Categoría profesional: Servicios
- Posición profesional: Empleados
- Temporalidad profesional: Otro
- Residencia: Alojamiento

5.1.2 Variables de la vivienda

- Tamaño en m^2: +120
- Número de habitaciones: 3
- Número de ocupantes: 3
- Año de construcción: 41-60
- Régimen legal: Otro

5.2 Gráficos

5.2.1 Correlaciones

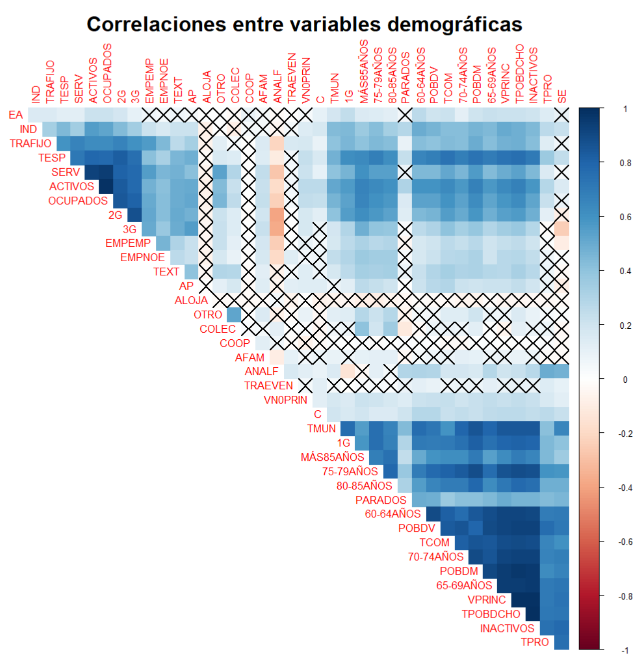

Figure 2: Correlaciones entre variables demográficas.

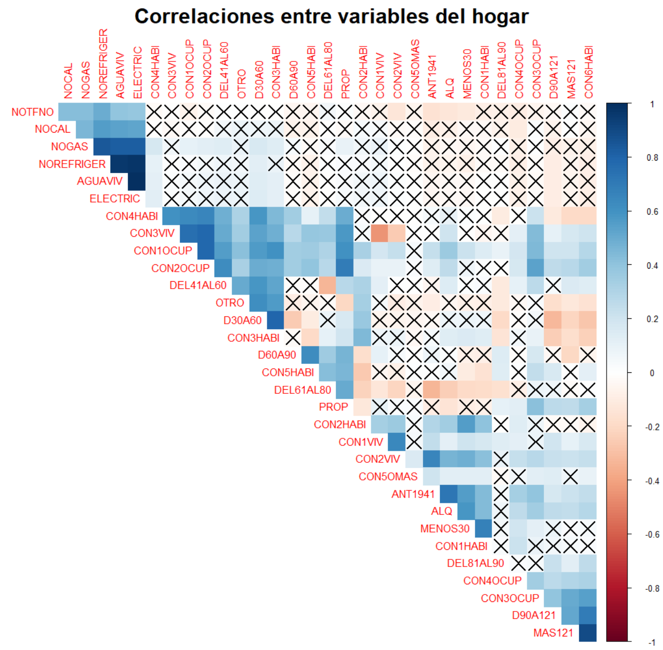

Figure 3: Correlaciones entre variables de las viviendas.

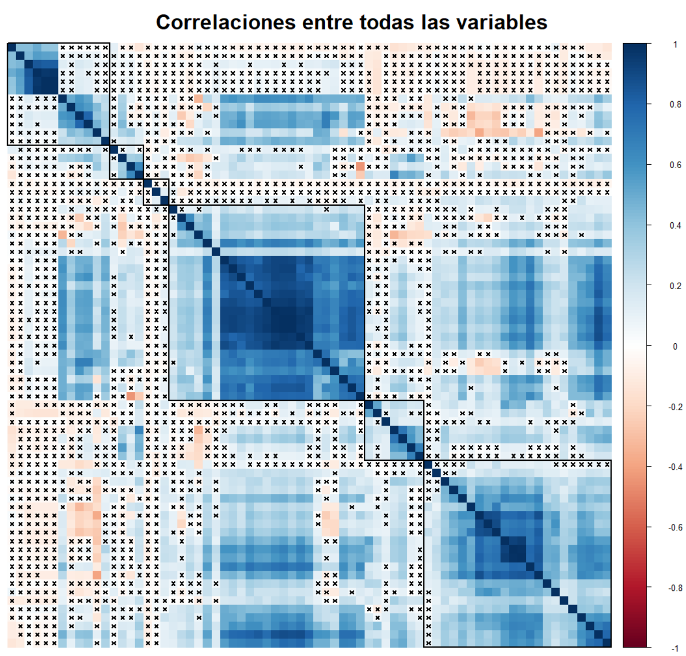

Figure 4: Correlaciones entre todas las variables.

6 Bibliografía

- Datos vectoriales
- Análisis clúster I
- Análisis clúster II
- Análisis clúster III
- ClustGeo
- ClusterR